





ORIGINAL RESEARCH

Cardiac Comorbidity Risk Score: Zero-Burden Machine Learning to Improve Prediction of Postoperative Major Adverse Cardiac Events in Hip and Knee Arthroplasty

Dmytro Onishchenko , MSc; Daniel S. Rubin , MD, MS; James R. van Horne, MD; R. Parker Ward , MD; Ishanu Chattopadhyay , PhD

BACKGROUND: In this retrospective, observational study we introduce the Cardiac Comorbidity Risk Score, predicting perioperative major adverse cardiac events (MACE) after elective hip and knee arthroplasty. MACE is a rare but important driver of mortality, and existing tools, eg, the Revised Cardiac Risk Index demonstrate only modest accuracy. We demonstrate an artificial intelligence-based approach to identify patients at high risk of MACE within 4 weeks (primary outcome) of arthroplasty, that imposes zero additional burden of cost/resources.

METHODS AND RESULTS: Cardiac Comorbidity Risk Score calculation uses novel machine learning to estimate MACE risk from patient electronic health records, without requiring blood work or access to any demographic data beyond that of sex and age, and accounts for variable/missing/incomplete information across patient records. Validated on a deidentified cohort (age >45 years, n=445391), performance was evaluated using the area under the receiver operator characteristics curve (AUROC), sensitivity/specificity, positive predictive value, and positive/negative likelihood ratios. In our cohort (age 63.5±10.5 years, 58.2% women, 34.2%/65.8% hip/knee procedures), 0.19% (882) experienced the primary outcome. Cardiac Comorbidity Risk Score achieved area under the receiver operator characteristics curve=80.0±0.4% (95% CI) for women and 80.1±0.5% (95% CI) for males, with 36.4% and 35.1% sensitivities, respectively, at 95% specificity, significantly outperforming Revised Cardiac Risk Index across all studied age-, sex-, risk-, and comorbidity-based subgroups.

CONCLUSIONS: Cardiac Comorbidity Risk Score, a novel artificial intelligence-based screening tool using known and unknown comorbidity patterns, outperforms state-of-the-art in predicting MACE within 4 weeks postarthroplasty, and can identify patients at high risk that do not demonstrate traditional risk factors.

Key Words: hip and knee arthroplasty ■ machine learning ■ Revised Cardiac Risk Index ■ risk of MACE

Major adverse cardiac events (MACE), consisting of myocardial infarction and cardiac arrest, are rare early complications in patients undergoing total hip or total knee arthroplasty.^{1–11} Total hip and knee arthroplasty are frequently performed surgeries in older adults with multiple cardiac comorbidities, and yet the overall frequency of MACE is low, complicating identification of high-risk patients. Risk calculators

used to predict postoperative MACE following lower extremity arthroplasty have shown modest accuracy.^{10,12} The Revised Cardiac Risk Index (RCRI)¹³ is a widely used preoperative risk calculator that uses existing cardiovascular (CVD) comorbidities and surgical procedural risk to determine perioperative risk for MACE. RCRI does not include many other known CVD risk factors or identify patients without a formal diagnosis

Correspondence to: Ishanu Chattopadhyay, PhD, Section of Hospital Medicine, Department of Medicine, University of Chicago, 900 E. 57 St. KCBD Room 10152, Chicago, IL 60637. Email: ishanu@uchicago.edu

Supplemental Material is available at <https://www.ahajournals.org/doi/suppl/10.1161/JAHA.121.023745>

For Sources of Funding and Disclosures, see page 13.

© 2022 The Authors. Published on behalf of the American Heart Association, Inc., by Wiley. This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- The Cardiac Comorbidity Risk Score (CCoR) is introduced to predict perioperative major adverse cardiac events 2–4 weeks after elective hip and knee arthroplasty.
- CCoR is based on known and yet unknown comorbidity signatures extracted via sophisticated processing of individual patient history of medical encounters without any new laboratory tests or blood work, making it applicable at the point of care, and easily integrable with existing electronic healthcare management systems.

What Are the Clinical Implications?

- CCoR outperformed the Revised Cardiac Risk Index to identify patients at high-risk of perioperative major adverse cardiac events in patients undergoing hip and knee arthroplasty.
- The CCoR performed well for patients with no Revised Cardiac Risk Index conditions as it effectively identified patients at high-risk of major adverse cardiac events with no previously diagnosed Revised Cardiac Risk Index conditions.
- Integration of the CCoR into electronic healthcare management systems can improve preoperative risk stratification and reduce healthcare costs as compared with the Revised Cardiac Risk Index over the next 2 decades.

Nonstandard Abbreviations and Acronyms

| | |
|--------------|-------------------------------------|
| CCoR | Cardiac Comorbidity Risk Score |
| MACE | major adverse cardiac events |
| PFSFA | Probabilistic Finite State Automata |
| RCRI | Revised Cardiac Risk Index |

of CVD comorbidities, which may limit its ability to predict perioperative MACE. In this study we sought to develop and validate a more accurate screening tool by systematically incorporating a vastly wider array of comorbidity patterns, including ones that are not already known risk factors for perioperative MACE.

To put in context, machine learning algorithms have been recently applied to assess risk of serious postoperative complications in procedures ranging from liver, pancreatic, colorectal surgeries, gastrectomies, and general in-patient procedures.^{14–18} However, use of such tools for lower extremity arthroplasty has not been reported. Additionally, most reported approaches use manually curated fixed set of input features involving results of specific laboratory tests, patient demographic information to populate the inputs to the

machine learning tools. This requirement imposes additional burden on patients, and caregivers, and can exacerbate healthcare use. We aimed to develop a “zero-burden” tool, that may be applied without any such specific data demands, and be operable with the existing patient history on file.

To accomplish this goal, we built upon our previous work^{19–22} on algorithmic pattern discovery in electronic health records databases. As a part of a broader trend of applying machine learning in medicine,^{23,24} these algorithms identify complex characteristics of comorbidity incidence, timing, sequence, and synchronism, that presage various diagnoses and outcomes,¹⁹ in this case, MACE in the 4 weeks after total hip arthroplasty and total knee arthroplasty. Combining these discovered features with standard machine learning leads to an automated screening tool based only on diagnostic codes already existing in the patient’s medical record. Risk screening using such algorithms therefore entails no additional diagnostic testing or other interventions for the patient, and no completion of risk calculators or other inputs by health professionals. In this study we develop and validate our screening tool for MACE after total hip arthroplasty/total knee arthroplasty, referred to as the Cardiac Comorbidity Risk Score (CCoR). We hypothesized that CCoR would show strong predictive ability on a standalone basis and outperform the RCRI in identifying patients at high risk of myocardial infarction and cardiac arrest following lower extremity arthroplasty.

METHODS

Data and Software Availability

Restrictions apply to some or all the availability of data analyzed during this study because they were used under license. The corresponding author will on request detail the restrictions and any conditions under which access to some data may be provided. A working version of the software tool developed in this study, free for non-commercial evaluation, will be made available by the corresponding author on request. To enable fast execution, some compute intensive features are disabled in this version. Results from this software are for demonstration purposes only, and must not be interpreted as medical advice, or serve as replacement for such.

Data Source, Patient Selection, and Ethics

Our patient data derive from the Truven Health Analytics MarketScan Commercial Claims and Encounters Database²⁵ for the years 2003–2018 (the “Truven data set”). The Truven data set combined deidentified patient records from >150 insurance carriers and large self-insurance companies including Medicare Advantage

and provides comprehensive inpatient and outpatient healthcare data obtained for >87 million patients. The Truven data set contains *International Classification of Diseases, Ninth and Tenth Revisions (ICD-9 and ICD-10)* diagnosis codes, current procedural terminology codes, as well as patients' age at arthroplasty and sex. Our study sample from the Truven data set based on the inclusion and exclusion criteria noted in Table S1 and Figure S1 (CONSORT-artificial intelligence extension diagram). In particular, we included patients with a current procedural terminology code that indicated either a total hip arthroplasty or total knee arthroplasty (Table S2). After the initial cohort selection, we applied the following exclusion criteria to patients: (1) <45 years of age, (2) active enrollment in the insurance plan for <12 months before surgery, and (3) continued enrollment in the insurance plan for <26 weeks following surgery. The University of Chicago Institutional Review Board granted an exemption for informed consent since this is an observational study using deidentified data (ID: IRB21-1272).

We use the RCRI version published in 1999,¹³ which is calculated on a scale from 0 to 6 points, with 1 point assigned for each of the following conditions: history of (1) ischemic heart disease, (2) heart failure, and (3) cerebrovascular disease, (4) insulin therapy for diabetes, and (5) serum creatinine >2.0 mg/dL; 1 to 3

are identified by the presence of any of the diagnosis codes listed in Table S3 in inpatient or outpatient claims data, and insulin therapy was identified based on at least 1 outpatient prescription for insulin in the year before arthroplasty. Since the Truven database contained insufficient laboratory data on preoperative serum creatinine, we used an *ICD* diagnosis code for chronic kidney disease stage III or higher as a surrogate for creatinine >2.0 mg/dL.^{11,26}

The sixth condition contributing to the RCRI is high-risk surgery which includes intraperitoneal surgery, suprainguinal vascular surgery and thoracic surgery. Since our cohort is limited to total hip and knee arthroplasty patients, the maximum number of conditions possible in the context of this study is 5 as enumerated in Table 1 (the contribution from the sixth category is always 0). We map the RCRI score to risk of MACE using published values,²⁷ and use this estimated risk value for evaluating RCRI performance when comparing with CCoR.

Outcome

The primary outcome was a MACE diagnosis, defined as myocardial infarction or cardiac arrest, within the 4 weeks of the date of the elective primary total arthroplasty of the hip or knee. MACE events were included if any of the *ICD* codes shown in Table S4 were

Table 1. Patient Characteristics

| Characteristic, n (%) | All surgeries (N=445391) | MACE within 2 wk (n=722) | MACE within 4 wk (n=882) | No MACE* (n=444509) |
|--|--------------------------|--------------------------|--------------------------|---------------------|
| Mean age at surgery (SD) | 63.5 (10.5) | 70.8 (10.3) | 70.8 (10.3) | 63.4 (10.5) |
| Men | 185992 (41.8%) | 385 (53.4%) | 464 (52.7%) | 185528 (41.8%) |
| Women | 259399 (58.2%) | 337 (46.6%) | 418 (47.3%) | 258981 (58.2%) |
| Knee surgery | 293153 (65.8%) | 451 (62.4%) | 559 (63.4%) | 292594 (65.8%); |
| Hip surgery | 152238 (34.2%) | 271 (37.6%) | 323 (36.6%) | 151915 (34.2%) |
| RCRI score | | | | |
| 0 | 295960 (66.6%) | 213 (29.5%) | 265 (30.0%) | 295695 (66.6%) |
| 1 | 101505 (22.7%) | 238 (32.9%) | 300 (34.0%) | 101205 (22.7%) |
| 2 | 37145 (8.3%) | 186 (25.9%) | 212 (24.1%) | 36933 (8.3%) |
| 3 | 9336 (2.1%) | 70 (9.6%) | 88 (9.9%) | 9247 (2.1%) |
| 4 | 1438 (0.3%) | 15 (2.1%) | 16 (1.8%) | 1422 (0.3%) |
| 5 | 7 (0.002%) | 0 (0%) | 0 (0%) | 7 (0.002%) |
| Chronic kidney disease: stage III or higher† | 18774 (4.2%) | 97 (13.4%) | 111 (12.5%) | 18663 (4.1%) |
| Ischemic heart disease | 98840 (22.1%) | 413 (57.2%) | 497 (56.3%) | 98343 (22.1%) |
| Congestive heart failure | 26309 (5.9%) | 139 (19.2%) | 167 (18.9%) | 26142 (5.8%) |
| Cerebrovascular disease | 64642 (14.5%) | 230 (31.8%) | 279 (31.6%) | 64363 (14.4%) |
| Preoperative treatment with insulin | 783 (0.17%) | 1 (0.13%) | 1 (0.11%) | 782 (0.17%) |

MACE indicates major adverse cardiac events; and RCRI, Revised Cardiac Risk Index.

*The "NO MACE" cohort comprised patients with no recorded MACE in their records during the 26 weeks after arthroplasty.

†Because of insufficient availability of relevant laboratory data in the Truven data set, presence of at least 1 diagnostic code for chronic kidney disease stage III or higher in the medical record in the year before the date of arthroplasty was used as a surrogate for the Revised Cardiac Risk Index condition, serum creatinine concentration >2.0 mg/dL.

documented in the medical record during the 4 weeks after arthroplasty. If a patient experienced multiple events during the 4 weeks postarthroplasty, we considered the first one. The 4-week time frame was chosen to be consistent with previous analyses evaluating the timing, incidence, and prediction of MACE after major noncardiac surgery.⁴ As a secondary outcome, we evaluated MACE prediction within 2 weeks of surgery.

Modeling and Prediction

To predict postsurgical MACE we aimed to classify time-stamped sequences of diagnostic codes into positive and control categories, where the positive category refers to patients experiencing the primary outcome. The control cohort comprised patients without any MACE in their records within the 26 weeks after the surgery. For both groups, we based our predictions on the 52 weeks of medical history before the total joint arthroplasty. We considered altogether >126 million diagnostic codes (>36 000 unique codes) (Table S5) in a sex-stratified analysis. We did not preselect any diagnostic code based on its association with MACE risk, using pattern discovery to find predictive precursors instead.

We proceeded by partitioning the set of all diagnostic codes into 26 broad categories (Tables S6 and S7; Figure 1), referred to as “CCoR phenotypes”. Some of the phenotypes encompass a relatively large number of codes aligning roughly with *ICD* categories.^{17,18,25,28} Other phenotypes include ≥ 1 codes that might have some known or suspected association with MACE. One such phenotype is “frailty” (Table S7), recognized to increase risk of such adverse cardiac outcomes.^{29,30} Applied to the individual timestamped history of diagnoses, each phenotype yields a single time series over weeks. Here each week is assigned a value “0” if no code recorded in that week corresponds to the phenotype, “1” if some code in that phenotype is present, or “2” if a diagnostic code from any other phenotype is present. Ultimately each patient is represented by 26 sparse stochastic event streams, which are compressed into specialized Hidden Markov Models known as Probabilistic Finite State Automata (PFSA).^{20,21} These models are inferred separately for each phenotype, for each sex, and for the control and the positive cohorts, ie, $26 \times 2 \times 2 = 104$ PFSA models are inferred altogether, to identify the distinctive average comorbidity patterns emerging at the population level. Variation in the structure and parameters of these inferred models between the positive and control groups delineated the estimated population-level risk of postoperative MACE. Given these models, and a specific patient’s history, we can quantify the log-likelihood (Table S8) of that history being generated by the control PFSA models as opposed to the positive models. We refer to this difference in likelihood as the sequence likelihood

defect²² (see Data S1) which is one of the key informative features in our approach. Besides the phenotype-specific Markov models, we used a wide range of engineered features that reflect various aspects of the patient-specific diagnostic histories (Table S8). Overall, we computed a total of 380 features for each patient, which were then used to train a standard gradient boosting classifier^{31,32} aiming to map individual patients to a raw risk score. Importantly, these features are derived solely from the information available in the patient electronic health records file, with no additional blood work or tests. It is important to note that the ternary encoding described above leads to $\approx 87\%$ to 89% 0’s, 1% 1’s, and $\approx 9\%$ to 11% 2’s (Table S9). Since the 1’s code for a specific category of disorders, while 2’s for all remaining categories, it is expected that 2’s will be more frequent.

All performance metrics are evaluated on held-back out-of-sample data, ie, on the validation set. We randomly chose 50% of our patients for training the algorithm, with the rest makes up the validation set. Half of the training data set was used for PFSA inference, and the rest for training the gradient boosting classifier. In addition to gradient boosting classifier, we experimented with other prominent classification models including random forests and extremely randomized trees.³³ The gradient boosting classifier emerged as the optimal choice with the highest out-of-sample performance given the initial feature engineering steps discussed above.

Chart Completeness

Records of medical encounters for patients cannot be guaranteed to be over identical time periods, especially when using a large claims database. Thus, we designed our approach to be applicable to patients with varying lengths of medical histories. The underlying models do not need equal-length inputs for computation. Importantly, we use only data within the inference period. If the set of recorded encounters for a patient do not span at least the past 12 months (52 weeks), then that patient is excluded, as described in our inclusion/exclusion criteria. Prospective studies in the future will evaluate applicability in such short-history situations.

Imbalance Between Positive and Control Cohorts

The incidence of MACE in patients undergoing total hip/knee replacement is low (0.3%–0.9%),¹ and $\approx 0.2\%$ in our data set. Such severe imbalances can skew predictors. However, this is not an important issue in our analysis since all reported performance are obtained on out-of-sample data. Nevertheless, we investigated if using a more balanced data set for training would boost

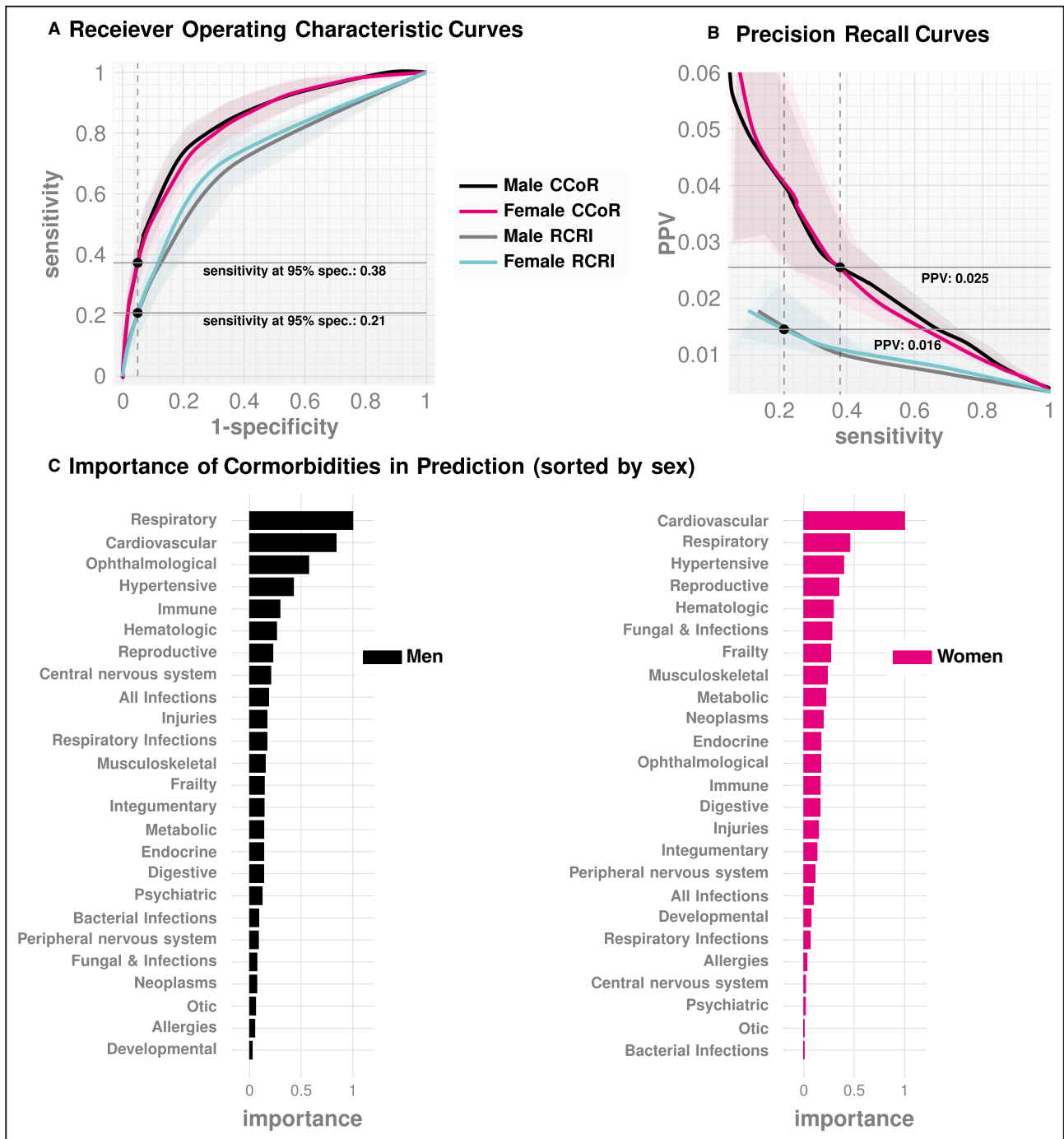


Figure 1. Out-of-sample performance of Cardiac Comorbidity Risk Score (CCoR) and Revised Cardiac Risk Index (N=445391) to predict major adverse cardiac events in the 4 weeks after elective primary lower extremity arthroplasty.

The validation cohort comprised the 50% of the study sample not used to develop CCoR. **A** and **B**, The receiver operating characteristics and precision-recall curves, respectively, of the two risk scores by sex. The AUROC of the receiver operating characteristics curve and the precision-recall curve reflect the risk scores' performance in predicting major adverse cardiac events 4 weeks after elective primary total hip or knee joint replacement. CCoR shows a strong prognostic performance that substantially exceeds that of Revised Cardiac Risk Index in both men and women. **C**, The relative importance of phenotypes (comorbidity phenotypes) in CCoR estimation. The 20 top phenotypes by sex are shown in descending order of relative importance in estimating the CCoR. Notably, these importances do not merely sum the predictive contribution from the presence or absence of individual diagnostic codes, a contribution reflected in the comorbidity spectra illustrated in Figure 2. Rather, the relative importance of the phenotypes sums the predictive contribution from all *patterns* (sequence likelihood defect as well as sequence features, see Methods) emerging from all disorders comprising each phenotype. Note that the sex of the patient matters, eg, respiratory disorders are somewhat overrepresented in women and circulatory disorders in men, and only 3 of the top 5 and 5 of the top 10 phenotypes are shared between women and men. CCoR indicates Cardiac Comorbidity Risk Score; RCRI, Revised Cardiac Risk Index; ROC, receiver operating characteristics; and SLD, Sequence Likelihood Defect.



Figure 2. Comorbidity spectra.

Disorders that increase the odds of a “true positive” vs a “true negative”; ie, these disorders, ranked here according to the log-odds ratio in ascending order of frequency in “true positive” versus “true negative” patients, are more likely in patients who are in the positive cohort. ICD-10 indicates *International Classification of Diseases, Tenth Revision*. Panel A and B shows comorbidity spectra for the two sexes, which while similar, have important differences.

our out-of-sample performance. We verify explicitly (Table S10) that such as a strategy confers no advantage, and in fact leads to progressive degradation of performance in out-of-sample data as we use more balanced data for training (a matched one-is-to-one balanced data set yields an out-of-sample AUROC of 75.5% for the female cohort, as opposed to >80% when no balancing is attempted). We suspect that any down-sampling to balance the training data causes the predictor to only focus on a subset of patterns, which then makes it less capable of assessing risk in out-of-sample data.

Feature Importance and Comorbidity Spectra

In addition to estimating risk, our analysis offers insights into the known and unknown comorbid associations of postoperative MACE, via the inferred relative importance of the features used. We computed the relative importance of the features by estimating the mean change in the raw risk via random perturbations of the features (Figure 1C). Additionally, we computed the statistically significant log-odds ratios of specific ICD codes occurring in the true positive versus the true negative patient sets, defining the “comorbidity spectra” (Figure 2). Importantly, the comorbidity spectra are based on individual codes, as opposed to the feature importance shown in Figure 1, which consider the aggregated impact of all features. Every disorder listed in the comorbidity spectra obviously does not appear in a given patient’s records, but codes with high log-odds ratio are significantly more likely in the positive cohort.

Statistical Analysis

When executed just before surgery, our models predict the raw risk score for a MACE within 4 (or 2) weeks of the procedure. Given the raw estimate, a decision threshold is chosen to make a balanced trade-off between Type 1 and Type 2 errors (or, equivalently, between specificity and sensitivity): if the raw risk is greater than this calibrated threshold, then the patient is predicted to be in the positive category, ie, likely to experience MACE. We compared the out-of-sample predictive performance for CCoR and RCRI using standard metrics, including accuracy, the area under the receiver operating characteristic curve (AUROC), sensitivity, specificity, positive and negative predictive values, and positive and negative likelihood ratios (LR+ and LR–, respectively). The 95% CIs on ROC curves and AUROCs were obtained via bootstrapping, and AUROC *P* values were obtained using the Mann–Whitney *U*-test statistic.³⁴

Cost Analysis

The low prevalence of MACE implies that statistically significant predictive advantage alone might not make

a sufficiently convincing case for the clinical relevance of the performance superiority of CCoR over RCRI. Thus, we evaluated the relative impact of the algorithms in a simple cost model, which also accounts for the impact of relatively high false positive rate arising because of the low prevalence of MACE.

Recent publication of the AJRR (American Joint Replacement Registry) puts the number of total hip/knee replacement procedures at 1.7M (2020),³⁵ and is projected to be 1.9M in 2025, 2.8M in 2030, and 4.8M in 2040.³⁶ Mean MACE-related costs incurred for the first event occurring within a few weeks of surgery is high, and was recently estimated to be \$19642.³⁷ The reported 30-day incidence (after total hip/knee arthroplasty) of MACE varies from 0.3% to 0.9%.¹ The additional procedures, testing and increased length of hospital stay because of a positive flag from a chosen risk assessment algorithm is estimated to be ≈\$10624 per patient.³⁸ These estimates inform our cost model designed to compare the clinical value added by CCoR versus RCRI. In particular, we estimate the cost (*C*) per year to the healthcare system as a result of MACE, with perioperative risk assessment performed by any given algorithm to be:

$$C = (t_p + f_p)C_t + f_n C_M$$

where t_p , f_p , f_n are true positives, false positives, and false negatives, and C_t , C_M are the costs per patient for additional testing, and those incurred after experiencing MACE respectively, thus explicitly accounting for the impact of false positives and false negatives. The above equation reduces to:

$$C = \left(s \left(\frac{C_t}{PPV} - C_M \right) + C_M \right) \rho N$$

where *s* is the sensitivity of the chosen algorithm, ρ is the prevalence of MACE among patients undergoing the surgical procedure, and *N* is the estimated number of procedures performed per year.

RESULTS

Our cohort (Table S11) included 445 391 patients undergoing lower extremity arthroplasty, with 0.19% experiencing the primary outcome of MACE within 4 weeks of the procedure (including 0.16% within 2 weeks of surgery). Patient characteristics overall and in the positive and control (negative) cohorts can be seen in Table 1. Rates of MACE within 4 weeks of surgery were 0.21% in patients undergoing total hip arthroplasty and 0.19% in those undergoing total knee arthroplasty, respectively. Table S11 presents the subcohort sizes stratified

by sex and the presence or absence of MACE within 4 weeks of arthroplasty.

CCoR Performance

The key performance metrics for CCoR to predict the primary outcome are presented in Table 2 and Figure 1, which illustrates the ROC curve, the AUROC, and the precision-recall curves, respectively, shown separately for women and men along with 95% CI. Performance for predicting MACE within 2 weeks of surgery is enumerated in Table 3. For predicting MACE 4 weeks postarthroplasty, out-of-sample AUROC was $80.1 \pm 1.9\%$ (95% CI) for women and $80.2 \pm 1.8\%$ (95% CI) for men, with $36.4 \pm 6\%$ and $35.1 \pm 1\%$ sensitivities, respectively, at 95% specificity. CCoR accuracy (the fraction of correct predictions) of 95%, and a PPV (the fraction of true positives among all positive flags) of 2.5% was achieved irrespective of sex. It should be noted that because of the low prevalence of MACE in our sample, even 100% sensitivity for a screening tool at 95% specificity would yield a PPV <5%. Given the low prevalence, the likelihood ratios are more relevant metrics here. For the primary outcome of MACE within 4 weeks of surgery, for women, CCoR achieved a LR+ of 7.28 ± 0.3 at 95% specificity (13.19 ± 2.1 at 99% specificity), implying that the odds of a positive CCoR flag in female patients who do experience MACE is 7.28 times to that in patients who do not, when the tool is operated at 95% specificity. For men, the corresponding LR+ are 7.19 ± 2 and 12.44 ± 0.3 at 95% and 99% specificity, respectively. Similarly, we achieved LR- of 0.67 ± 0.05 for women and 0.68 ± 0.01 for men at 95% specificity, implying that the odds of a negative CCoR result in a patient not experiencing MACE is $1/0.67 = 1.5$ times more than that in patients who do. For the secondary outcome of MACE within 2 weeks of surgery, CCoR achieves an LR+ of 7.5 ± 2.5 (women) and 7.65 ± 0.6 (males) at 95% specificity, and 14.48 ± 0.3 (women) and 13.98 ± 2.1 (men) at 99% specificity. The corresponding LR- estimates for the 2-week horizon are 0.87 ± 0.01 (99% specificity) and 0.66 ± 0.03 (95% specificity) for both sexes. These results show that CCoR is particularly effective when there is a positive flag (high LR+), and moderately so for negative flags.

CCoR performance in predicting the primary outcome in all studied age-, risk-, and comorbidity-defined subgroups of the validation cohort was comparable with the overall out-of-sample performance (Table 2, Tables S12 and S13). The score's performance remained high, albeit slightly decreased, when predictions of 4-week MACE with 99% specificities were considered (Table S14). CCoR performance was slightly stronger (AUROC $80.9 \pm 2.1\%$ in women, $81.3 \pm 1.9\%$ in men) for the shorter time horizon of 2 weeks postoperatively (Table 3 and Table S15).

Performance Comparison With RCRI

Tables 2 and 3 also display performance metrics at 95% specificity for the RCRI to predict the primary and secondary outcomes. In both sexes, (Figure 3) and across all studied age-, risk-, and comorbidity-defined subgroups (Tables S12 and S13), the AUROC for CCoR exceeded that of RCRI, often substantially. This pattern held true when specificity was set to 99% or for either the 2-week or the 4-week time frame (Tables S14 and S15). The differences in positive likelihood ratios are dramatic: CCoR LR+ is $\approx 62\%$ to 78% larger, and LR- is 24% smaller at 95% specificity for the primary outcome. At 99% specificity, the positive and negative likelihood ratios are 110% to 163% larger and 1% to 6% smaller, respectively. At the shorter horizon of 2 weeks, the ratios are similar to that of the primary outcome.

Importantly the CCoR achieved high performance ($76.5 \pm 0.7\%$ in women, $76.6 \pm 1.1\%$ in men, 95% CI) in patients lacking any RCRI condition (RCRI score of 0) for the 4-week prediction horizon, who represented nearly two thirds of the study sample ($n=296539$, 66.5%) and accounted for 29.5% of all patients who suffered a cardiac event (ie, 29.5% of events occurred in RCRI=0 patients). The dramatic positive likelihood ratio in this subcohort (9.4 ± 3.8 and 7.25 ± 1.3 for at 95% specificity, 21.8 ± 1.3 and 5.96 ± 0.9 at 99% specificity for men and women, respectively, see Table 2 and Table S14), demonstrates utility to accurately risk-stratify patients completely missed by current screening tools. Also of interest, CCoR performance significantly exceeds that reported by Harris et al¹⁰ to predict 30-day cardiac complications of elective lower extremity total joint arthroplasty. Harris and colleagues achieved an AUROC of $72\% - 73\% \pm 2\%$, 95% CI, using a limited number of preselected binary comorbidity indicators and patient demographics.

The substantial superiority in achieved out-of-sample AUROC of CCoR over RCRI were found to be statistically significant across different subcohorts that we investigated (Table S16), for both the primary and secondary outcomes.

Cost Comparison Between CCoR and RCRI

With the values of the parameters chosen as described in our cost model (and using a prevalence of 0.3% for MACE), we estimate that the overall cost varies with the chosen sensitivity and the PPV as shown in Figure 3A and 3B, respectively. Under current practice, a common operating point to trigger referral for preoperative tests and corrective measures, is an estimated RCRI risk >6%.²⁷ We estimated that this maps approximately to a sensitivity between 30% and 40% for MACE within 4 weeks of surgery. The

Table 2. Out-of-Sample* Prediction of MACE With 4 Weeks of Hip/Knee Arthroplasty (Primary End Point) at 95% Specificity: CCoR Versus RCRI†

| Sex | Cohort | Model | Sensitivity | PPV | Accuracy | LR+ | LR– | AUROC |
|-------|--------------|-------|-------------|-------------|-------------|----------|-----------|-------------|
| Women | <65 | RCRI | 0.10±0.01 | 0.008±0.000 | 0.947±0.006 | 0.47±0.0 | 0.93±0.00 | 0.639±0.039 |
| Women | <65 | CCoR | 0.31±0.01 | 0.025±0.010 | 0.948±0.006 | 7.31±3.1 | 0.72±0.01 | 0.775±0.035 |
| Men | <65 | RCRI | 0.18±0.02 | 0.015±0.007 | 0.947±0.000 | 4.23±2.3 | 0.87±0.02 | 0.682±0.034 |
| Men | <65 | CCoR | 0.34±0.00 | 0.030±0.014 | 0.948±0.006 | 8.97±4.7 | 0.69±0.01 | 0.783±0.030 |
| Women | 65+ | RCRI | 0.16±0.02 | 0.012±0.002 | 0.947±0.000 | 3.35±0.5 | 0.88±0.02 | 0.664±0.028 |
| Women | 65+ | CCoR | 0.32±0.01 | 0.022±0.010 | 0.948±0.006 | 6.48±3.0 | 0.71±0.01 | 0.771±0.025 |
| Men | 65+ | RCRI | 0.16±0.03 | 0.011±0.001 | 0.947±0.006 | 3.17±0.4 | 0.88±0.03 | 0.661±0.026 |
| Men | 65+ | CCoR | 0.27±0.00 | 0.019±0.006 | 0.948±0.006 | 5.57±1.8 | 0.77±0.00 | 0.762±0.023 |
| Women | All patients | RCRI | 0.18±0.02 | 0.014±0.002 | 0.947±0.000 | 4.13±0.6 | 0.86±0.02 | 0.688±0.023 |
| Women | All patients | CCoR | 0.36±0.06 | 0.025±0.001 | 0.948±0.006 | 7.28±0.3 | 0.67±0.05 | 0.801±0.019 |
| Men | All patients | RCRI | 0.20±0.02 | 0.017±0.002 | 0.947±0.000 | 4.83±0.5 | 0.84±0.02 | 0.705±0.020 |
| Men | All patients | CCoR | 0.35±0.01 | 0.025±0.006 | 0.948±0.006 | 7.19±2.0 | 0.68±0.01 | 0.802±0.018 |
| Women | Frail‡ | RCRI | 0.12±0.03 | 0.009±0.002 | 0.947±0.006 | 2.50±0.6 | 0.92±0.02 | 0.670±0.028 |
| Women | Frail | CCoR | 0.31±0.02 | 0.022±0.007 | 0.948±0.000 | 6.40±2.2 | 0.73±0.02 | 0.791±0.025 |
| Men | Frail | RCRI | 0.23±0.01 | 0.019±0.003 | 0.947±0.006 | 5.46±0.9 | 0.80±0.00 | 0.727±0.027 |
| Men | Frail | CCoR | 0.38±0.02 | 0.026±0.008 | 0.948±0.000 | 7.56±2.6 | 0.66±0.03 | 0.810±0.024 |
| Women | High risk§ | RCRI | 0.11±0.02 | 0.008±0.001 | 0.947±0.006 | 2.19±0.3 | 0.94±0.02 | 0.581±0.029 |
| Women | High risk | CCoR | 0.25±0.02 | 0.017±0.007 | 0.948±0.000 | 5.06±2.2 | 0.79±0.02 | 0.737±0.026 |
| Men | high risk | RCRI | 0.15±0.03 | 0.010±0.001 | 0.947±0.006 | 2.91±0.3 | 0.90±0.02 | 0.617±0.026 |
| Men | High risk | CCoR | 0.30±0.01 | 0.021±0.006 | 0.948±0.000 | 6.02±2.0 | 0.74±0.00 | 0.729±0.024 |
| Women | Low risk¶ | CCoR | 0.33±0.04 | 0.032±0.012 | 0.948±0.000 | 9.40±3.8 | 0.70±0.04 | 0.765±0.036 |
| Men | Low risk | CCoR | 0.33±0.02 | 0.025±0.004 | 0.948±0.000 | 7.25±1.3 | 0.71±0.02 | 0.766±0.032 |

<65 indicates subcohort with patients aged ≤65 years; AUROC, area under the receiver operating characteristic curve; CCoR, Cardiac Comorbidity Risk Score; LR–, negative likelihood ratio; LR+, positive likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; and RCRI, Revised Cardiac Risk Index.

*50% of the study sample (n=445391) used for validation.

†Because of insufficient availability of relevant laboratory data in the Truven data set, presence of at least 1 diagnostic code for chronic kidney disease stage III or higher in the medical record in the year before the date of arthroplasty was used as a surrogate for the Revised Cardiac Risk Index condition, serum creatinine concentration >2.0mg/dL (to convert to micromoles per liter, multiply by 88.4).

‡ICD codes for frailty enumerated in Table S7.

§Low-risk status subcohort comprises patients with Revised Cardiac Risk Index score 0. For RCRI score >0, the patient is deemed to be at high risk.

¶No Revised Cardiac Risk Index performance logged for low-risk patients, since their RCRI score is zero.

estimated cost for achieving a 40% sensitivity in the general population with CCoR is ≈667M US dollars per year currently. With RCRI this estimate to 1.48B US dollars, which approximately reflects current practice (see annotation in panel B).

Additionally, at any fixed yearly cost, CCoR is estimated to deliver substantially higher sensitivity and PPV (Figure 3C and 3D), eg, if the total yearly cost to the healthcare system attributable to preoperative testing+post-MACE costs is fixed at 500M, then CCoR can deliver >150% more sensitivity, and >20% more PPV up to around the year 2025. The sensitivity advantage is estimated to be even wider at later years. If the total estimated cost is set at around current practice, then CCoR can deliver an improvement in either sensitivity or PPV >50% (Figure 3C and 3D).

DISCUSSION

Our study used novel machine learning algorithms to develop and validate a tool to identify comorbidity patterns in past diagnoses to predict postoperative MACE following total hip and knee arthroplasty. We demonstrated 2 key results: (1) our CCoR effectively predicted MACE within 2 to 4 weeks of a total hip or knee arthroplasty, and (2) CCoR is a significantly stronger predictor of perioperative cardiac morbidity after total hip and knee arthroplasty in clinical practice than RCRI. From the inferred relative importance of the features identified we conclude that respiratory and cardiovascular disorders are the most important modulators of risk,³⁹ followed by hematologic, reproductive, nervous system disorders and infections. While the role of past

Table 3. Out-of-Sample* Prediction of MACE With 2 Weeks of Hip/Knee Arthroplasty (Secondary End Point) at 95% Specificity: CCoR Versus RCRI†

| Sex | Cohort | Model | Sensitivity | PPV | Accuracy | LR+ | LR- | AUROC |
|-------|--------------|-------|-------------|-------------|-------------|-----------|-----------|-------------|
| Women | <65 | RCRI | 0.11±0.01 | 0.009±0.000 | 0.947±0.006 | 0.51±0.0 | 0.92±0.00 | 0.647±0.044 |
| Women | <65 | CCoR | 0.32±0.07 | 0.025±0.005 | 0.948±0.006 | 7.30±1.6 | 0.71±0.06 | 0.787±0.039 |
| Men | <65 | RCRI | 0.20±0.03 | 0.017±0.003 | 0.947±0.000 | 4.95±1.0 | 0.84±0.04 | 0.688±0.037 |
| Men | <65 | CCoR | 0.38±0.03 | 0.031±0.014 | 0.948±0.000 | 9.01±4.4 | 0.65±0.03 | 0.797±0.033 |
| Women | 65+ | RCRI | 0.17±0.00 | 0.012±0.002 | 0.947±0.006 | 3.60±0.6 | 0.87±0.00 | 0.671±0.030 |
| Women | 65+ | CCoR | 0.31±0.02 | 0.022±0.012 | 0.948±0.000 | 6.37±3.8 | 0.72±0.02 | 0.787±0.027 |
| Men | 65+ | RCRI | 0.14±0.02 | 0.010±0.001 | 0.947±0.000 | 2.79±0.4 | 0.91±0.02 | 0.667±0.028 |
| Men | 65+ | CCoR | 0.30±0.02 | 0.021±0.007 | 0.948±0.006 | 6.10±2.1 | 0.74±0.02 | 0.780±0.025 |
| Women | All patients | RCRI | 0.19±0.02 | 0.016±0.002 | 0.947±0.000 | 4.49±0.6 | 0.85±0.02 | 0.692±0.025 |
| Women | All patients | CCoR | 0.37±0.01 | 0.026±0.008 | 0.948±0.006 | 7.50±2.5 | 0.66±0.01 | 0.809±0.021 |
| Men | All patients | RCRI | 0.19±0.02 | 0.016±0.002 | 0.947±0.000 | 4.63±0.6 | 0.85±0.02 | 0.710±0.022 |
| Men | All patients | CCoR | 0.37±0.03 | 0.026±0.002 | 0.948±0.006 | 7.65±0.6 | 0.66±0.03 | 0.813±0.019 |
| Women | Frail‡ | RCRI | 0.14±0.02 | 0.011±0.002 | 0.947±0.000 | 3.13±0.7 | 0.90±0.03 | 0.676±0.032 |
| Women | Frail | CCoR | 0.31±0.03 | 0.023±0.003 | 0.948±0.000 | 6.75±0.8 | 0.73±0.04 | 0.807±0.027 |
| Men | Frail | RCRI | 0.23±0.03 | 0.019±0.003 | 0.947±0.000 | 5.40±0.8 | 0.80±0.03 | 0.736±0.029 |
| Men | Frail | CCoR | 0.41±0.04 | 0.028±0.002 | 0.948±0.000 | 8.20±0.7 | 0.62±0.04 | 0.825±0.025 |
| Women | High risk§ | RCRI | 0.12±0.00 | 0.008±0.001 | 0.947±0.006 | 2.35±0.4 | 0.93±0.00 | 0.584±0.032 |
| Women | High risk | CCoR | 0.23±0.05 | 0.017±0.003 | 0.947±0.006 | 4.90±0.8 | 0.81±0.04 | 0.742±0.028 |
| Men | High risk | RCRI | 0.13±0.01 | 0.009±0.001 | 0.947±0.006 | 2.69±0.4 | 0.91±0.00 | 0.628±0.028 |
| Men | High risk | CCoR | 0.33±0.03 | 0.023±0.002 | 0.948±0.000 | 6.57±0.6 | 0.71±0.03 | 0.737±0.026 |
| Women | Low risk¶ | CCoR | 0.43±0.03 | 0.041±0.017 | 0.948±0.006 | 12.20±5.7 | 0.60±0.03 | 0.779±0.040 |
| Men | Low risk | CCoR | 0.39±0.02 | 0.029±0.005 | 0.948±0.000 | 8.57±1.5 | 0.64±0.02 | 0.793±0.035 |

<65 indicates subcohort with patients aged ≤65 years; AUROC, area under the receiver operating characteristic curve; CCoR, Cardiac Comorbidity Risk Score; LR+, positive likelihood ratio; LR-, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; and RCRI, Revised Cardiac Risk Index.

*50% of the study sample (n=445391) used for validation.

†Because of insufficient availability of relevant laboratory data in the Truven data set, presence of at least 1 diagnostic code for chronic kidney disease stage III or higher in the medical record in the year before the date of arthroplasty was used as a surrogate for the Revised Cardiac Risk Index condition, serum creatinine concentration >2.0mg/dL (to convert to micromoles per liter, multiply by 88.4).

‡ICD codes for frailty enumerated in Table S7.

§Low-risk status subcohort comprises patients with Revised Cardiac Risk Index score 0. For Revised Cardiac Risk Index score >0, the patient is deemed to be at high risk.

¶No Revised Cardiac Risk Index performance logged for low-risk patients, since their Revised Cardiac Risk Index score is zero.

cardiovascular diagnoses in postoperative MACE has been recognized,³⁸ prior research^{40–45} has examined binary relationships between the presence or absence of a diagnosis (eg, coronary artery disease). This is a major limitation for the RCRI, as a formal diagnosis of coronary artery disease or heart failure is required to identify patients at risk. Our approach can identify patients with patterns and combinations of comorbidities in past medical encounters associated with increased risk for MACE, as opposed to only those with specific CVD disorders. A key conclusion from our study is that CCoR is markedly better at predicting the occurrence of MACE, than predicting its nonoccurrence. Thus, a positive CCoR flag reliably identifies high-risk patients who may experience postoperative MACE. The reliability is lower for concluding that the absence of a CCoR flag implies no MACE in the postsurgical time frames

considered. In either case, CCoR likelihood ratios are superior to RCRI.

Our approach addresses a major limitation of current comorbidity-based risk calculators that require binary diagnoses (eg, heart failure) to deduce an increased risk. Newer cardiac risk calculators, such as the American College of Surgeons National Surgical Quality Improvement Project⁴⁶ have improved upon this approach by incorporating symptoms (eg, dyspnea with moderate exertion) and functional status, but are still primarily reliant on a limited set of manually curated diagnoses. Additionally, the calculators can be challenging to incorporate into busy clinic workflows. Our approach integrates directly into standard electronic medical record systems to estimate expected cardiac risk before surgical procedures. This could be readily available to clinicians of all specialties at the

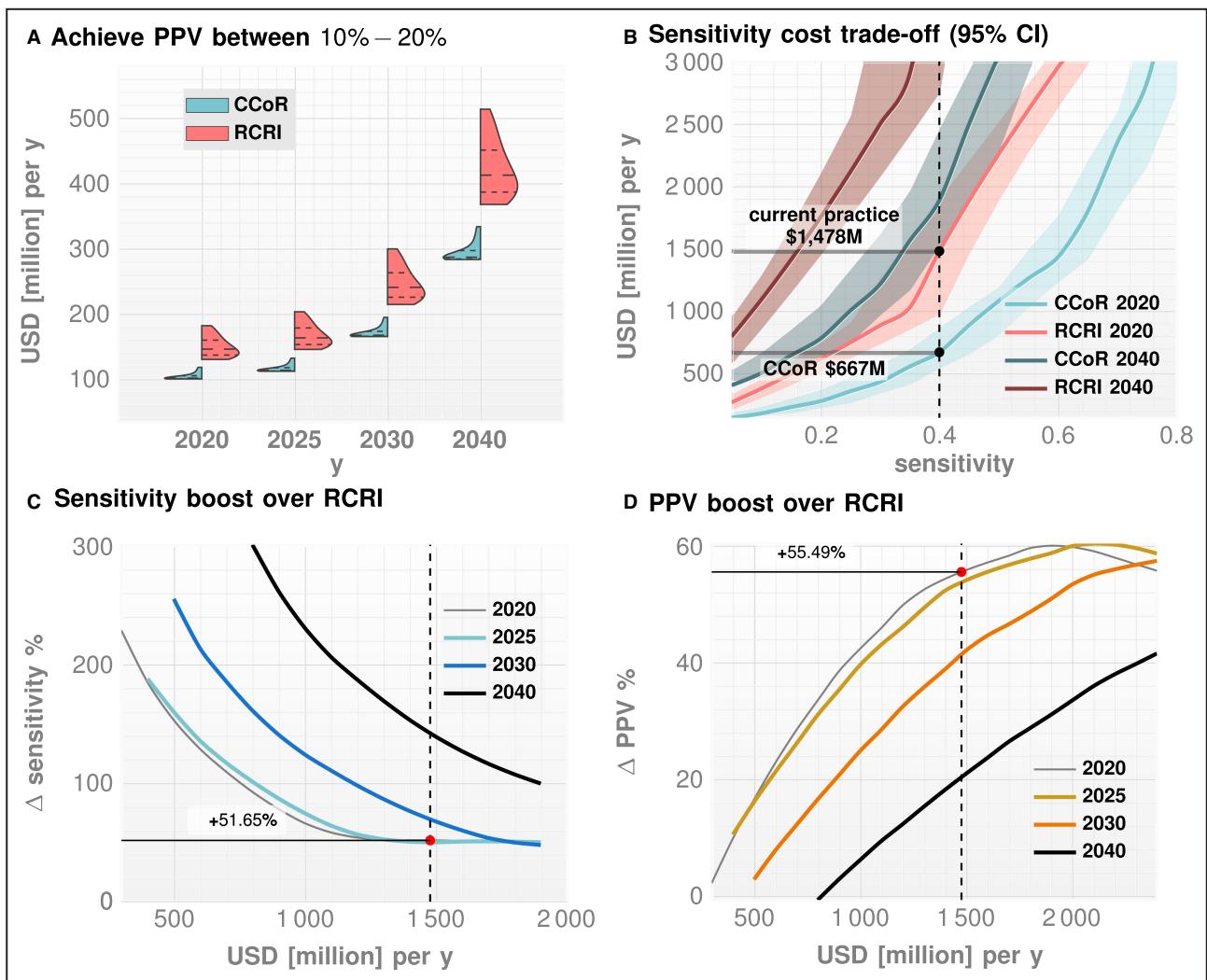


Figure 3. Cost model.

A, The distribution of estimated cost for Cardiac Comorbidity Risk Score (CcoR) and Revised Cardiac Risk Index (RCRI) over the years for achieving positive predictive value between 10% and 20%. **B**, The cost trade-off with sensitivity for 2020 and 2040 along with 95% CI. Note that the estimated cost of achieving a 40% sensitivity in the general population (which approximately corresponds to >6% risk with RCRI and reflects current practice) with CCoR is ≈750M US dollars per year currently, while with RCRI this increases to 1.5B US dollars per year. **C**, The relative boost in sensitivity achieved by CCoR over RCRI as a percentage for a fixed cost. **D**, The relative boost in positive predictive value achieved by CCoR over RCRI as a percentage for a fixed cost. Thus, if the yearly cost is fixed at 500M, then CCoR can deliver >150% more sensitivity, or almost 20% more positive predictive value up to around 2025. If the costs are set to what is estimated for current practice (\$1478M), then CCoR can deliver a >50% improvement in either sensitivity or positive predictive value. The estimated number of surgeries per year, and values for the costs of additional testing on a positive flag, and for the costs incurred upon experiencing major adverse cardiac events are adopted from published estimates of comparable scenarios. CCoR indicates Cardiac Comorbidity Risk Score; PPV, positive predictive value; RCRI, Revised Cardiac Risk Index; and USD, US dollar.

time of office-based perioperative risk assessment, or even guide who should be scheduled for subspecialty risk assessment. Ultimately, patterns of healthcare use before surgical procedures can be leveraged to identify patients that may require additional testing or interventions to modify risk before a surgical procedure. Given that the Truven data set contains healthcare information from roughly 25% of the nation’s population, and that our validation cohort had similar demographic characteristics to those of other published studies undergoing total joint arthroplasties,⁴⁷ CcoR

performance in the present study is expected to be widely generalizable in the United States. Furthermore, the significant superiority of CcoR over RCRI across various subcohorts, and in multiple relevant metrics, points to the potential for substantial improvement in clinical outcomes on deployment.

Some observations, discussed next, related to the reported performance estimates might be of clinical interest. For RCRI, the PPV achieved for the age-related subpopulations (<65 and ≥65 years) were both lower than the PPV of the unrestricted population

Table 4. Performance Comparison of CCoR When Augmented With History of Prescribed Medications

| Sex | Prediction Horizon | AUROC (CCoR) | AUROC (CCoR augmented with prescriptions data) | Sensitivity at 95% specificity (CCoR) | Sensitivity at 95% specificity (CCoR augmented with prescriptions data) |
|-------|--------------------|--------------|--|---------------------------------------|---|
| Men | 2 wk | 81.3% | 81.5% | 34.1% | 34.5% |
| | 4 wk | 80.1% | 80.5% | 31.8% | 32.7% |
| Women | 2 wk | 80.9% | 81.0% | 32.5% | 31.8% |
| | 4 wk | 80.0% | 80.3% | 31.0% | 30.2% |

AUROC indicates area under the receiver operating characteristic curve; and CCoR, Cardiac Comorbidity Risk Score.

(Tables 2, 3 and Tables S14, S15), which is counter-intuitive. However, in this study the PPVs of the different subcohorts are computed separately from that of the unrestricted cohort (indicated as “all patients” in Tables 2, 3 and Tables S14, S15). Since we compute the ROC curves independently in each subcohort, the threshold of binary decision differs from one subcohort to the next, which makes it possible for the estimated PPV to be lower in both younger and older subpopulations compared with the unrestricted cohort.

We also observe that CCoR performance is generally lower in the subcohorts (Tables 2 and 3). We suspect that the unrestricted cohort has a wider range of patient ages, each of which have a different incidence of MACE. This makes them more easily classifiable, on average, using age-at-screening as a feature when considered all together. When we define a subcohort with smaller age-variation, and compute performance within that group, then the predictive advantage of the age-variable disappears, leading to a somewhat lower predictive performance.

In this study, in addition to comorbidity signatures, we investigated how the sex and age of patients modulate MACE risk. This was motivated by the fact that sex and age are both known strong predictors of perioperative MACE.^{48,49} Further, cardiovascular disease presentation, diagnosis, and management have been shown to be different between men and women.⁵⁰ In addition, patient demographics, clinical variables, specific treatments, and prescribed drugs might also be potentially important predictors, and in general, we expect the performance to improve if we add more informative features. However, it was difficult to assess this hypothesis within the constraints of our current data set which lacked patient demographic information such as race and ethnicity, or extensive data on other clinical variables. We will investigate the impact of demographic characteristics and other relevant clinical variables as part of future research. We did have access to information on the history of prescribed medications for the patients, and we explicitly investigated if adding such information improves our model. Considering the presence/absence and timing of prescriptions such as antibiotics, antidepressants, and beta-blockers did not show any significant improvement in performance

(Table 4). We surmised that in the context of the problem at hand, the diagnostic history of patients already captures most of the predictive information that medication history provides, and hence the latter does not improve results.

We also investigated if the different components of our overall model are all indeed necessary. Our results show that, as expected, on their own the individual components, eg, the PFSA models, do substantially worse (Table S17).

Among the 26 studied CCoR phenotypes (Table S7), those encompassing cardiovascular and respiratory conditions unsurprisingly had high importance for predicting early postoperative MACE in both sexes (Figure 1C). Overall, the sexes differed clearly on the relative importance of phenotypes, with only 3 of the top 5 and 5 of the top 10 phenotypes the same in women and men (Figure 1C). Differences between the sexes were even more marked regarding comorbidity spectra (Figure 2). Of the 57 diagnostic codes with the highest relative prevalence in positive versus control patients, only 3 were present among both women and men, and in all cases, at a different rank: skin carcinoma of the lower limb and hip (*ICD-10* code D04.7), nonspecific malignant neoplasm of the lung in a nonspecific part of a bronchus (*ICD-10* code 34.9), and nonspecific diastolic (congestive) heart failure (*ICD-10* code I50.3).

We also estimated that CCoR has the potential to substantially reduce healthcare usage costs. While false positives increase the cost of additional confirmatory tests that must be undertaken for a positive flag, our cost model indicates that even with relatively large number of false positives (lower compared with RCRI), we are more cost-effective.

Finally, we note that a key limitation of this study is the use of administrative codes to ascertain past diagnostic history, which are vulnerable to coding errors, and do not record relevant nuances in diagnostic decisions and uncertainties. Also, while CCoR was derived using a well-validated national database of claims data, local healthcare usage patterns might require some recalibration at the level of an individual local institution. Finally, the CCoR was designed to estimate risk of early MACE following elective lower extremity total joint arthroplasty, and impact on the preoperative workup

(eg, preoperative stress testing¹¹) or treatment (eg, new prescriptions) will require further investigation, as will the effect, if any, on clinical and pharmacoeconomic outcomes. Thus, further research and prospective trials are necessary before CcoR can enter everyday clinical practice.

CONCLUSIONS

We developed and validated CcoR, an automated, potentially widely applicable screening tool to predict myocardial infarction and cardiac arrest in 2 to 4 weeks after primary total hip or knee arthroplasty. CcoR is solely based on patterns of comorbidity incidence, temporality, sequence, and synchronism in data already in the electronic health record, and hence spares patients diagnostic interventions and physicians the need to input and verify data for risk calculators. The impact of CcoR screening on clinical practice and on clinical and pharmacoeconomic outcomes also warrants investigation.

ARTICLE INFORMATION

Received August 25, 2021; accepted May 13, 2022.

Affiliations

Department of Medicine, University of Chicago, IL (D.O., R.P.W., I.C.); Southern Oregon Orthopedics, Medford, OR (J.R.v.H.); Section of Cardiology (R.P.W.), Department of Anesthesia and Critical Care (D.S.R.), Committee on Genetics, Genomics & Systems Biology (I.C.), Committee on Quantitative Methods in Social, Behavioral, and Health Sciences (I.C.), and Section of Hospital Medicine (I.C.), University of Chicago, IL.

Acknowledgments

Robert J. Marlowe, BA from Spencer-Fontayne Corporation, Jersey City, NJ, USA provided editorial support, literature research, and writing assistance. Author contributions: Dr Chattopadhyay had full access to all the data and takes responsibility for the integrity of the data and the accuracy of the data analysis. Mr Onishchenko and Dr Rubin both had complementary but equal contributions in the development and implementation of this study. Concept and design: Onishchenko, Rubin, Chattopadhyay; Acquisition, analysis, or interpretation of data: All authors; Drafting of the manuscript: Onishchenko, Rubin, Chattopadhyay; Critical revision of the manuscript for important intellectual content: All authors; Statistical analysis: Onishchenko, Chattopadhyay; Administrative, technical, or material support: Chattopadhyay; Supervision: Chattopadhyay.

Sources of Funding

This work is funded in part by the Defense Advanced Research Projects Agency project number HR00111890043/P00004. The claims made in this study do not reflect the position or the policy of the US Government.

Disclosures

Dr Chattopadhyay is a founder and shareholder of Zero Burden Laboratories, Inc., a company formed to commercialize biomedical applications of machine learning. He has not taken any salary or money from the company. He has received funding from the United States Department of Defense, the National Institutes of Health, and the Neubauer Collegium for Culture and Society. Dr Rubin is the President of DRDR Mobile Health, a company that creates mobile applications for health care, including functional capacity assessment applications. He has not taken any salary or money from the company. He has engaged in consulting for mobile applications as well. Mr Onishchenko is a founder and shareholder of Zero Burden Laboratories. He has not taken any salary or money from the company. The remaining authors have no disclosures to report.

Supplemental Material

Data S1
Figure S1
Tables S1–S17
References 20, 21, 51–64

REFERENCES

- Petersen PB, Kehlet H, Jorgensen CC; Lundbeck Foundation Center for Fast-Track Hip and Knee Replacement Collaborative Group. Myocardial infarction following fast-track total hip and knee arthroplasty-incidence, time course, and risk factors: a prospective cohort study of 24,862 procedures. *Acta Orthop*. 2018;89:603–609. doi: [10.1080/17453674.2018.1517487](https://doi.org/10.1080/17453674.2018.1517487)
- Bemenderfer TB, Rozario NL, Moore CG, Karunakar MA. Morbidity and mortality in elective total hip arthroplasty following surgical care improvement project guidelines. *J Arthroplasty*. 2017;32:2359–2362. doi: [10.1016/j.arth.2017.02.080](https://doi.org/10.1016/j.arth.2017.02.080)
- Menendez ME, Memtsoudis SG, Opperer M, Boettner F, Gonzalez Della Valle A. A nationwide analysis of risk factors for in-hospital myocardial infarction after total joint arthroplasty. *Int Orthop*. 2015;39:777–786. doi: [10.1007/s00264-014-2502-z](https://doi.org/10.1007/s00264-014-2502-z)
- Belmont PJ Jr, Goodman GP, Kusnezov NA, Magee C, Bader JO, Waterman BR, Schoenfeld AJ. Postoperative myocardial infarction and cardiac arrest following primary total knee and hip arthroplasty: rates, risk factors, and time of occurrence. *J Bone Joint Surg Am*. 2014;96:2025–2031. doi: [10.2106/jbjs.n.00153](https://doi.org/10.2106/jbjs.n.00153)
- Malviya A, Martin K, Harper I, Muller SD, Emmerson KP, Partington PF, Reed MR. Enhanced recovery program for hip and knee replacement reduces death rate. *Acta Orthop*. 2011;82:577–581. doi: [10.3109/17453674.2011.618911](https://doi.org/10.3109/17453674.2011.618911)
- Kirksey M, Chiu YL, Ma Y, Della Valle AG, Poultsides L, Gerner P, Memtsoudis SG. Trends in in-hospital major morbidity and mortality after total joint arthroplasty: United States 1998–2008. *Anesth Analg*. 2012;115:321–327. doi: [10.1213/ane.0b013e31825b6824](https://doi.org/10.1213/ane.0b013e31825b6824)
- Jorgensen CC, Kehlet H; Lundbeck Foundation Centre for Fast-Track Hip and Knee Replacement Collaborative Group, Soeballe K, Hansen TB, Husted H, Laursen MB, Hansen LT, Kjærsgaard-Andersen P, Solgaard S, Krarup NH, et al. Time course and reasons for 90-day mortality in fast-track hip and knee arthroplasty. *Acta Anaesthesiol Scand*. 2017;61:436–444. doi: [10.1111/aas.12860](https://doi.org/10.1111/aas.12860)
- Zmistowski B, Restrepo C, Hess J, Adibi D, Cangoz S, Parvizi J. Unplanned readmission after total joint arthroplasty: rates, reasons, and risk factors. *J Bone Joint Surg Am*. 2013;95:1869–1876. doi: [10.2106/jbjs.l.00679](https://doi.org/10.2106/jbjs.l.00679)
- Phillips JL, Rondon AJ, Vannello C, Fillingham YA, Austin MS, Courtney PM. How much does a readmission cost the bundle following primary hip and knee arthroplasty? *J Arthroplasty*. 2019;34:819–823. doi: [10.1016/j.arth.2019.01.029](https://doi.org/10.1016/j.arth.2019.01.029)
- Harris AH, Kuo AC, Weng Y, Trickey AW, Bowe T, Giori NJ. Can machine learning methods produce accurate and easy-to-use prediction models of 30-day complications and mortality after knee or hip arthroplasty? *Clin Orthop Relat Res*. 2019;477:452–460. doi: [10.1097/corr.0000000000000601](https://doi.org/10.1097/corr.0000000000000601)
- Rubin DS, Hughey R, Gerlach RM, Ham SA, Ward RP, Nagele P. Frequency and outcomes of preoperative stress testing in total hip and knee arthroplasty from 2004 to 2017. *JAMA Cardiol*. 2021;6:13–20. doi: [10.1001/jamacardio.2020.4311](https://doi.org/10.1001/jamacardio.2020.4311)
- Manning DW, Edelstein AI, Alvi HM. Risk prediction tools for hip and knee arthroplasty. *J Am Acad Orthop Surg*. 2016;24:19–27. doi: [10.5435/jaaos-d-15-00072](https://doi.org/10.5435/jaaos-d-15-00072)
- Lee TH, Marcantonio ER, Mangione CM, Thomas EJ, Polanczyk CA, Cook EF, Sugarbaker DJ, Donaldson MC, Poss R, Ho KK, et al. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. *Circulation*. 1999;100:1043–1049. doi: [10.1161/01.cir.100.10.1043](https://doi.org/10.1161/01.cir.100.10.1043)
- Merath K, Hyer JM, Mehta R, Farooq A, Bagante F, Sahara K, Tsilimigras DI, Beal E, Paredes AZ, Wu L, et al. Use of machine learning for prediction of patient risk of postoperative complications after liver, pancreatic, and colorectal surgery. *J Gastrointest Surg*. 2020;24:1843–1851. doi: [10.1007/s11605-019-04338-2](https://doi.org/10.1007/s11605-019-04338-2)
- Xue B, Li D, Lu C, King CR, Wildes T, Avidan MS, Kannampallil T, Abraham J. Use of machine learning to develop and evaluate models

- using preoperative and intraoperative data to identify risks of postoperative complications. *JAMA Netw Open*. 2021;4:e212240. doi: 10.1001/jamanetworkopen.2021.2240
16. Lu S, Yan M, Li C, Yan C, Zhu Z, Lu W. Machine-learning-assisted prediction of surgical outcomes in patients undergoing gastrectomy. *Chin J Cancer Res*. 2019;31:797–805. doi: 10.21147/j.issn.1000-9604.2019.05.09
 17. Fritz BA, Cui Z, Zhang M, He Y, Chen Y, Kronzer A, Abdallah AB, King CR, Avidan MS. Deep-learning model for predicting 30-day postoperative mortality. *Br J Anaesth*. 2019;123:688–695. doi: 10.1016/j.bja.2019.07.025
 18. Sofo L, Caprino P, Schena C, Sacchetti F, Potenza A, Ciociola A. New perspectives in the prediction of postoperative complications for high-risk ulcerative colitis patients: machine learning preliminary approach. *Eur Rev Med Pharmacol Sci*. 2020;24:12781–12787. doi: 10.26355/eurrev_202012_24178
 19. Onishchenko D, Chattopadhyay I, Huang Y, van Horne J, Smith PJ, Msall MM. Reduced false positives in autism screening via digital bio-markers inferred from deep co-morbidity patterns. *Sci Adv*. 2021;7:eabf0354. doi: 10.1126/sciadv.abf0354
 20. Chattopadhyay I, Lipson H. Data smashing: uncovering lurking order in data. *J R Soc Interface*. 2014;11:20140826. doi: 10.1098/rsif.2014.0826
 21. Chattopadhyay I, Lipson H. Abductive learning of quantized stochastic processes with probabilistic finite automata. *Philos Trans A Math Phys Eng Sci*. 2013;371:20110543. doi: 10.1098/rsta.2011.0543
 22. Huang Y, Chattopadhyay I. Data smashing 2.0: sequence likelihood (sl) divergence for fast time series comparison. *arXiv*. 2019;1909:12243. doi: 10.48550/ARXIV.1909.12243
 23. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Sci Rep*. 2016;6:1–10. doi: 10.1038/srep26094
 24. Davenport T, Kalakota R. The potential for artificial intelligence in health-care. *Future Healthc J*. 2019;6:94–98. doi: 10.7861/futurehosp.6-2-94
 25. International Business Machines. MarketScan Research Databases, 2003–2021. 2022. Available at: <https://www.ibm.com/products/market-scan-research-databases>. Accessed January 1, 2021.
 26. Davis C, Tait G, Carroll J, Wijeyesundera DN, Beattie WS. The Revised Cardiac Risk Index in the new millennium: a single-Centre prospective cohort re-evaluation of the original variables in 9,519 consecutive elective surgical patients. *Can J Anaesth*. 2013;60:855–863. doi: 10.1007/s12630-013-9988-5
 27. Ford MK, Beattie WS, Wijeyesundera DN. Systematic review: prediction of perioperative cardiac complications and mortality by the Revised Cardiac Risk Index. *Ann Intern Med*. 2010;152:26–35. doi: 10.3410/f.1844956.1386054
 28. International Classification of Diseases, Ninth revision, Clinical Modification. *Definitions*. 2020. doi: 10.32388/ggn38l
 29. El Haddad K, Rolland Y, Gérard S, Mourey L, Sourdet S, Vellas B, Stephan E, Abellan Van Kan G, De Souto BP, Balardy L. No difference in the phenotypic expression of frailty among elderly patients recently diagnosed with cancer vs cancer free patients. *J Nutr Health Aging*. 2020;24:147–151. doi: 10.1007/s12603-019-1293-8
 30. Rubin DS, Peden CJ. Preoperative frailty and cognitive dysfunction assessment. *Anesthesiology*. 2020;133:1164–1166. doi: 10.1097/aln.0000000000003579
 31. Cover TM, Thomas JA. *Elements of Information Theory*. Wiley-Interscience; 2006. doi: 10.1002/047174882x
 32. Ke G, Meng Q, Finlay T, Wang T, Chen W, Chen W, Ma W, Ye Q, Liu TY. Lightgbm: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst*. 2017;30:3146–3154. doi: 10.1109/iccse.2019.8845529
 33. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011;12:2825–2830. doi: 10.3389/jmlr.2014.00014
 34. Fay MP, Michael AP. Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Stat Surv*. 2010;4:1–39. doi: 10.1214/09-ss051
 35. Levine BR, Springer BD, Golladay GJ. Highlights of the 2019 American joint replacement registry annual report. *Arthroplasty Today*. 2020;6:998–1000. doi: 10.1016/j.artd.2020.09.010
 36. Singh JA, Yu S, Chen L, Cleveland JD. Rates of total joint replacement in the United States: future projections to 2020–2040 using the national inpatient sample. *J Rheumatol*. 2019;46:1134–1140. doi: 10.3899/jrheum.170990
 37. Korsnes JS, Davis KL, Arieli R, Bell CF, Mitra D. Health care resource utilization and costs associated with nonfatal major adverse cardiovascular events. *J Manag Care Spec Pharm*. 2015;21:443–450. doi: 10.18553/jmcp.2015.21.6.443
 38. Coffman J, Tran T, Quast T, Berlowitz MS, Chae SH. Cost conscious care: preoperative evaluation by a cardiologist prior to low-risk procedures. *BMJ Open Qual*. 2019;8:e000481. doi: 10.1136/bmjopen-2018-000481
 39. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97:1837–1847. doi: 10.1161/01.cir.97.18.1837
 40. Stürmer T, Dreinhöfer K, Gröber-Grätz D, Brenner H, Dieppe P, Puhl W, Günther KP. Differences in the views of orthopaedic surgeons and referring practitioners on the determinants of outcome after total hip replacement. *J Bone Joint Surg Br*. 2005;87:1416–1419. doi: 10.1302/0301-620x.87b10.16702
 41. Hofstede SN, Gademán MG, Vliet Vlieland TP, Nelissen RG, Marangvan de Mheen PJ. Preoperative predictors for outcomes after total hip replacement in patients with osteoarthritis: a systematic review. *BMC Musculoskelet Disord*. 2016;17:212. doi: 10.1186/s12891-016-1070-3
 42. Hilton ME, Gioe T, Noorbaloochi S, Singh JA. Increasing comorbidity is associated with worsening physical function and pain after primary total knee arthroplasty. *BMC Musculoskelet Disord*. 2016;17:421. doi: 10.1186/s12891-016-1261-y
 43. Tsang ST, Gaston P. Adverse peri-operative outcomes following elective total hip replacement in diabetes mellitus: a systematic review and meta-analysis of cohort studies. *Bone Joint J*. 2013;95-B:1474–1479. doi: 10.1302/0301-620x.95b11.31716
 44. Ethgen O, Bruyere O, Richy F, Dardennes C, Reginster JY. Health-related quality of life in total hip and total knee arthroplasty. A qualitative and systematic review of the literature. *J Bone Joint Surg Am*. 2004;86:963–974. doi: 10.2106/00004623-200405000-00012
 45. Olthof M, Stevens M, Bulstra SK, van den Akker-Scheek I. The association between comorbidity and length of hospital stay and costs in total hip arthroplasty patients: a systematic review. *J Arthroplasty*. 2014;29:1009–1014. doi: 10.1016/j.arth.2013.10.008
 46. Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmieciak TE, Ko CY, Cohen ME. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. *J Am Coll Surg*. 2013;217:833–842. doi: 10.1016/j.jamcollsurg.2013.07.385
 47. Lauren RP, Kayee I. Projected volume of primary and revision total joint replacement in the US 2030 to 2060. *Annual meeting of American Academy of Orthopaedic Surgeons*. 2018. Available at: https://aaos-annualmeeting-presskit.org/2018/research-news/sloan_tjr/. Accessed January 20, 2022.
 48. Sorrentino R, Santoro C, Bardi L, Rigolin V, Gentile F. Non-cardiac surgery in patients with valvular heart disease. *Heart*. 2021;heartjnl-2021-319160. doi: 10.1136/heartjnl-2021-319160
 49. Banco D, Dodson JA, Berger JS, Smilowitz NR. Perioperative cardiovascular outcomes among older adults undergoing in-hospital non-cardiac surgery. *J Am Geriatr Soc*. 2021;69:2821–2830. doi: 10.1111/jgs.17320
 50. Leening MJ, Ferket BS, Steyerberg EW, Kavousi M, Deckers JW, Nieboer D, Heeringa J, Portegies ML, Hofman A, Ikram MA, et al. Sex differences in lifetime risk and first manifestation of cardiovascular disease: prospective population-based cohort study. *BMJ*. 2014;17:349. doi: 10.1136/bmj.g5992
 51. Granger CW, Joyeux R. An introduction to long-memory time series models and fractional differencing. *J Time Ser Anal*. 1980;1:15–29. doi: 10.1111/j.1467-9892.1980.tb00297.x
 52. General equivalence mappings. Available at: https://www.cms.gov/medicare/coding/icd10/downloads/icd-10_gem_fact_sheet.pdf. Accessed January 1, 2022.
 53. Kullback S, Leibler RA. On information and sufficiency. *Ann Math Stat*. 1951;22:79–86. doi: 10.1214/aoms/1177729694
 54. Doob JL. *Stochastic Processes*. Wiley Publications in Statistics. John Wiley & Sons; 1953. doi: 10.4236/jmf.2012.22022
 55. Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal*. 2002;38:367–378. doi: 10.1016/S0167-9473(01)00065-2
 56. Hopcroft JE, Motwani R, Ullman JD. Introduction to automata theory, languages, and computation. *ACM SIGACT News*. 2001;32:60–65. doi: 10.1145/568438.568455

-
57. Klenke A. *Probability Theory: A Comprehensive Course*. Springer Science & Business Media; 2013. doi: [10.1007/978-1-4471-5361-0](https://doi.org/10.1007/978-1-4471-5361-0)
 58. Chattopadhyay I, Ray A. Structural transformations of probabilistic finite state machines. *Int J Control*. 2008;81:820–835. doi: [10.1080/00207170701704746](https://doi.org/10.1080/00207170701704746)
 59. Bondy JA, Murty US. *Graph Theory*. Springer Publishing Company, Incorporated; 2008. doi: [10.5555/1481153](https://doi.org/10.5555/1481153)
 60. Vidyasagar M. *Hidden Markov Processes*. Princeton University Press; 2014. doi: [10.1515/9781400850518](https://doi.org/10.1515/9781400850518)
 61. Chung KL. *Markov Chains*. Springer-Verlag; 1967. doi: [10.1007/978-3-642-49686-8](https://doi.org/10.1007/978-3-642-49686-8)
 62. Trahtman AN. An algorithm for road coloring. *J Discrete Algorithms*. 2012;16:213–223. doi: [10.1016/j.jda.2012.05.003](https://doi.org/10.1016/j.jda.2012.05.003)
 63. Matthews AGG, Hensman J, Turner R, Ghahramani Z. Proceedings of the 19th International Conference on Artificial Intelligence and Statistics. PMLR. 2016; 51:231–239.
 64. Hardy GH. Note on a divergent series. *Proc Camb Philos Soc*. 1941;37:1–8. doi: [10.1017/s0305004100021472](https://doi.org/10.1017/s0305004100021472)

Supplemental Materials

DATA S1. SUPPLEMENTAL METHODS

A. Time-series Modeling of Diagnostic History

Individual diagnostic histories can have long-term memory [52], implying that the order, frequency, and comorbid interactions between diseases are important for assessing the future risk of our target phenotype. We analyze patient-specific diagnostic code sequences by first representing the medical history of each patient as a set of stochastic categorical time-series — one each for a specific group of related disorders — followed by the inference of stochastic models for these individual data streams. These inferred generators are from a special class of Hidden Markov Models (HMMs), referred to as Probabilistic Finite State Automata (PFSA) [53]. The inference algorithm we use is distinct from classical HMM learning, and has important advantages related to its ability to infer structure, and its sample complexity (See Supplementary text, Section VI). We infer a separate class of models for the positive and control cohorts, and then the problem reduces to determining the probability that the short diagnostic history from a new patient arises from the positive as opposed to the control category of the inferred models.

B. Inference & Event Periods

We train our predictive pipeline with all diagnostic codes that are recorded in the past 26years from the point at which a prediction is made. This period from which we use data to train our pipeline is called the “inference window”. Our aim is to make predictions on the occurrence of the target diagnostic codes at 2year from the end of the inference window. For patients in the control cohort, we make sure that no target code appears for 26years after the end of the inference window. Additionally, when making predictions further into the future (upto 4 years, as described in the main text), we always make sure that the control group has no target codes for 1 year after the predicted time of diagnosis, i.e., if we are making a prediction of a diagnosis 4 years in future, then control group patients are chosen to have no diagnosis in at least next 5 years.

C. Step 1: Partitioning The Human Disease Spectrum

We begin by partitioning the human disease spectrum into 26non-overlapping categories. Each category is defined by a set of diagnostic codes from the International Classification of Diseases, Ninth Revision (ICD9) (See Table SI-S7 for description of the categories used in this study). For this study, we ended up using 4879398 and 7753318 diagnostic codes for males and females respectively (17554 and 19209 unique codes) spanning both ICD9 and ICD10 protocols (using ICD10 General Equivalence Mappings (GEMS) [54] equivalents where necessary), from a total 445391 patients. Transforming the diagnostic histories to report only the broad categories reduces the number of distinct codes that the pipeline needs to handle, thus improving statistical power. Our categories largely align with the top-level ICD9 categories, with small adjustments, e.g. bringing all infections under one category irrespective of the pathogen or the target organ. We do not pre-select the phenotypes; we want our algorithm to seek out the important patterns without any manual curation of the input data. For each patient, the past medical history is a sequence $(t_1, x_1), \dots, (t_m, x_m)$, where t_i are timestamps and x_i are ICD9 codes diagnosed at time t_i . We map individual patient history to a three-alphabet categorical time series z^k corresponding to the disease category k , as follows. For each week i , we have:

$$z_i^k = \begin{cases} 0 & \text{if no diagnosis codes in week } i \\ 1 & \text{if there exists a diagnosis of category } k \text{ in week } i \\ 2 & \text{otherwise} \end{cases} \quad (1)$$

The time-series z^k is observed in the inference period. Thus, each patient is represented by 43 mapped trinary series.

D. Step 2: Model Inference & The Sequence Likelihood Defect Δ

The mapped series, disease-category, and perioperative cardiac event diagnosis-status are considered to be independent sample paths, and we want to explicitly model these systems as specialized HMMs (PFSAs). We model the positive and the control cohorts and each disease category separately, ending up with a total of 104 HMMs at the population level (26 categories, 2 perioperative cardiac event status categories: positive and control, and 2 sexes). Each of these inferred models is a PFSA; a directed graph with probability-weighted edges, and acts as an optimal generator of the stochastic process driving the sequential appearance of the three letters

(as defined by Eq. (1)) corresponding to disease category, and perioperative cardiac event status-type (See Section VI in the Supplementary text for background on PFSA inference).

To reliably infer the perioperative cardiac event status-type of a new patient, i.e, the likelihood of a diagnostic sequence being generated by the corresponding perioperative cardiac event status-type model, we generalize the notion of Kullbeck-Leibler (KL) divergence [55] between probability distributions to a divergence $\mathcal{D}_{\text{KL}}(G||H)$ between ergodic stationary categorical stochastic processes [56] G, H as:

$$\mathcal{D}_{\text{KL}}(G||H) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{x:|x|=n} p_G(x) \log \frac{p_G(x)}{p_H(x)} \quad (2)$$

where $|x|$ is the sequence length, and $p_G(x), p_H(x)$ are the probabilities of sequence x being generated by the processes G, H respectively. Defining the log-likelihood of x being generated by a process G as :

$$L(x, G) = -\frac{1}{|x|} \log p_G(x) \quad (3)$$

The cohort-type for an observed sequence x — which is actually generated by the hidden process G — can be formally inferred from observations based on the following provable relationships (See Supplementary Text Section VI, Theorem 6 and 7):

$$\lim_{|x| \rightarrow \infty} L(x, G) = \mathcal{H}(G) \quad (4a)$$

$$\lim_{|x| \rightarrow \infty} L(x, H) = \mathcal{H}(G) + \mathcal{D}_{\text{KL}}(G||H) \quad (4b)$$

where $\mathcal{H}(\cdot)$ is the entropy rate of a process [32]. Importantly, Eq. (4) shows that the computed likelihood has an additional non-negative contribution from the divergence term when we choose the incorrect generative process. Thus, if a patient is eventually going to be diagnosed with perioperative cardiac event, then we expect that the disease-specific mapped series corresponding to her diagnostic history be modeled by the PFSA in the positive cohort. Denoting the PFSA corresponding to disease category j for positive and control cohorts as G_+^j, G_0^j respectively, we can compute the *sequence likelihood defect* (SLD, Δ^j) as:

$$\Delta^j \triangleq L(G_0^j, x) - L(G_+^j, x) \rightarrow \mathcal{D}_{\text{KL}}(G_0^j||G_+^j) \quad (5)$$

With the inferred PFSA models and the individual diagnostic history, we estimate the SLD measure on the right-hand side of Eqn. (5). The higher this likelihood defect, the higher the similarity of diagnosis history to that of women with perioperative cardiac event.

E. Step 3: Risk Estimation Pipeline With Semi-supervised & Supervised Learning Modules

The risk estimation pipeline operates on patient specific information limited to the available diagnostic history in the inference period, and produces an estimate of the relative risk of perioperative cardiac event, with an associated confidence value. To learn the parameters and associated model structures of this pipeline, we transform the patient specific data to a set of engineered features, and the feature vectors realized on the positive and control sets are used to train a gradient-boosting classifier [57]. The complete list of 380 features used is provided in Table 6.

We need two training sets: one to infer the models, and one to train the classifier with features derived from the inferred models. Thus, we do a random 3-way split of the set of unique patients into *feature-engineering* (25%), *training* (25%) and *test* (50%) sets. We use the feature-engineering set of ids first to infer our PFSA models (*unsupervised model inference in each category*), which then allows us to train the gradient-boosting classifier using the training set and PFSA models (*classical supervised learning*), and we finally execute out-of-sample validation on the test set. Fig. 1c in the main text shows the top 20 features ranked in order of their relative importance (relative loss in performance when dropped out of the analysis).

I. THRESHOLD SELECTION ON ROC CURVE

Once the ROC curve has been computed, we must choose a decision threshold to trade-off true positive rate and false positive rate. In situations where the number of negatives vastly outnumber the number of positives (which is the case in our problem), it is better to base this trade-off on a measure that is independent of the number of true negatives. The two popular measures considered in the literature are accuracy and the F1-score:

$$\text{accuracy} = \frac{t_p + t_n}{t_p + f_p + f_n + t_n} \quad (6)$$

$$\text{F1} = \frac{2t_p}{2t_p + f_p + f_n} \quad (7)$$

The F1-score is the same as accuracy where the number of true negatives is the same as the number of true positives, thus partially correcting for the class imbalance.

The selection of the threshold may also be dictated by the current practice of ensuring high specificities in screening tests. Thus, a relevant clinically operating point is the one corresponding to 95% specificity, which is highlighted in Fig. 1a.

II. NOTE ON RECIEVER OPERATING CHARACTERISTICS (ROC) AND PRECISION-RECALL CURVES

The ROC curve is a plot between the False Positive rate (FPR) and the True Positive Rate (TPR), and the area under the ROC curve (AUC) is often used as a measure of classifier performance. For the same of completeness, we introduce the relevant definitions:

In the following P denotes the total number of positive samples (number of patients who are eventually diagnosed), and N denotes the total number of negative samples (number of patients in the control group).

Definition 1. *True positive rate, true negative rate, false positive rate, positive predictive value (PPV), and prevalence (ρ) are defined as:*

$$TPR = \frac{t_p}{P} = \frac{t_p}{t_p + f_n} \quad (8)$$

$$TNR = \frac{t_n}{N} = \frac{t_n}{t_n + f_p} \quad (9)$$

$$FPR = 1 - TNR \quad (10)$$

$$PPV = \frac{t_p}{t_p + f_p} \quad (11)$$

$$\rho = \frac{P}{N + P} \quad (12)$$

where as before t_p, t_n, f_p, f_n are true positives, true negatives, false positives, and false negatives respectively.

Note that TPR is also referred to as **recall** or **sensitivity**, and PPV is also referred to as **precision**. True negative rate is also known as **specificity**.

A **precision-recall curve**, or a PPV-sensitivity curve is a plot between PPV and TPR.

Denoting sensitivity by s , and specificity by c , it follows that:

$$PPV = \frac{t_p/P}{t_p/P + (f_p/N)(N/P)} = \frac{TPR}{TPR + ((N - t_n)/N)(N/P)} \quad (13)$$

$$\Rightarrow PPV = \frac{s}{s + (1 - c)(\frac{1}{\rho} - 1)} \quad (14)$$

Thus, we note that for a fixed specificity and sensitivity, the PPV depends on prevalence. Indeed, it is clear from the above argument that PPV decreases with decreasing prevalence, and vice versa.

III. EFFECT OF CLASS IMBALANCE

ROC curves are generally assumed to be robust to class imbalance. Note that if we assume that patient outcomes are independent (which is well-justified in the case of a non-communicable condition, particularly in large databases), then t_p should scale linearly with the total number of positives P , implying:

$$TPR = \frac{t_p}{P} = \frac{t'_p}{P'} \quad (15)$$

implying that with different sizes of the set of positive samples (or negative samples), the ROC curve remains unchanged. In particular, note that even if the prevalence is very small (say 0.01%), we cannot cheat to boost the AUC by labeling all predictions as negative, or stating that risk is always zero: in that case, our P is very small, but our $t_p = 0$ strictly, implying that our $TPR = 0$, thus leading to a zero AUC. We can cheat to boost the accuracy (See the previous section), but not the AUC.

Note that while relative class sizes or imbalance does not affect the ROC (under the assumption that true positives and true negatives scale with the number of positives and negatives), very small absolute sample sizes might still result in poor performance of the model.

The precision-recall curves do get affected by class imbalance, or the prevalence, as shown by Eq (14). However, in diagnostic analysis, they are important since we are generally less interested in the number of true negatives; the ratio of false positives to the total number of positive recommendations by the algorithm is much more relevant, i.e., the PPV or the precision.

IV. GENERATING PFSA MODELS FROM SET OF INPUT STREAMS WITH VARIABLE INPUT LENGTHS

Our PFSA reconstruction algorithm [53] is distinct from standard HMM learning. We do not need to pre-specify structures, or the number of states in the algorithm, and all model parameters are inferred directly from data. Additionally, we can operate either with 1) a single input stream, or 2) a set of input streams of possibly varying lengths which are assumed to be different and independent sample paths from the unknown stochastic generator we are trying to infer. At an intuitive level, we use the input data to infer the length of histories one must remember to estimate the current state, and predict futures for the process being modeled. Thus, we do not step through the symbol streams with a pre-specified model structure, and avoid the need to have equal-length inputs. More details of the algorithm are provided in the next section.

The ability to model a set of input streams of varying lengths is particularly important, since medical histories of different patients are typically of different lengths.

V. PROBABILISTIC FINITE STATE AUTOMATA INFERENCE

A. Probabilistic Finite-State Automaton

Let Σ be a finite alphabet of symbols with size $|\Sigma|$. The set of sequences of length d over Σ is denoted by Σ^d . The set of finite but unbounded sequences over Σ is denoted by Σ^* , the Kleene star operation [58], i.e. $\Sigma^* = \bigcup_{d=0}^{\infty} \Sigma^d$. We use lower case Greek, for example σ or τ , for symbols in Σ , and lower case Latin, for example x or y , for sequences of symbols, i.e. $x = \sigma_1 \sigma_2 \dots \sigma_n$. We use $|x|$ to denote the length of x . The empty sequence is denoted by λ .

We denote the set of strictly infinite sequences over Σ by Σ^ω , and the set of strictly infinite sequences having x as prefix by $x\Sigma^\omega$. Let $\mathcal{S} = \{x\Sigma^\omega : x \in \Sigma^*\} \cup \{\emptyset\}$, we can verify that \mathcal{S} is a semiring [59] over Σ^ω . We use \mathcal{F} to denote the sigma algebra generated by \mathcal{S} .

Definition 2 (Stochastic Process over Σ). *A stochastic process over a finite alphabet Σ is a collection of Σ -valued random variables $\{X_t\}_{t \in \mathbb{N}}$ indexed by positive integers [56].*

We are specifically interested in processes in which the X_t s are not necessarily independently distributed.

Definition 3 (Sequence-Induced Measure and Derivative). *For a process \mathcal{P} , let $\Pr_{\mathcal{P}}(x)$ or simply $\Pr(x)$ denote the probability \mathcal{P} producing a sample path prefixed by x . The **measure** μ_x **induced by a sequence** $x \in \Sigma^*$ is the extension [59] to \mathcal{F} of the premeasure defined on the semiring \mathcal{S} given by*

$$\forall x, y \in \Sigma^*, \mu_x(y\Sigma^\omega) \triangleq \frac{\Pr(xy)}{\Pr(x)}, \text{ if } \Pr(x) > 0 \quad (16)$$

*For any $d \in \mathbb{N}$, the **d-th order derivative** of a sequence x , written as ϕ_x^d , is defined to be the marginal distribution of μ_x on Σ^d , with the entry indexed by y denoted by $\phi_x^d(y)$. The first-order derivative is called the **symbolic derivative** and is denoted by ϕ_x for short.*

Definition 4 (Probabilistic Nerode Equivalence and Causal States [60]). *For any pair of sequences $x, y \in \Sigma^*$, x is equivalent to y , written as $x \sim y$, if and only if either $\Pr(x) = \Pr(y) = 0$, or $\mu_x = \mu_y$. The equivalence class of a sequence x is denoted by $[x]$ and is called a **causal state** [61]. The cardinality of the set of causal states is called the **probabilistic Nerode index**, or the Nerode index for simplicity.*

We can see from the definition that causal states captures how the history of a process influences its future. Since the probabilistic Nerode equivalence is right invariant, it gives rise naturally to a automaton structure introduced below.

Definition 5 (Probabilistic Finite-State Automaton (PFSA)). *A PFSA G is defined by a quadruple $(Q, \Sigma, \delta, \tilde{\pi})$, where Q is a finite set, Σ is a finite alphabet, $\delta : Q \times \Sigma \rightarrow \Sigma$ is called the transition map, and $\tilde{\pi} : Q \rightarrow \mathbf{P}_\Sigma$, where*

\mathbf{P}_Σ is the space of probability distributions over Σ , is called the transition probability. The entry of $\tilde{\pi}(q)$ indexed by σ is denoted by $\tilde{\pi}(q, \sigma)$.

Definition 6 (Transition and Observation Matrices). The transition matrix Π is the $|Q| \times |Q|$ matrix with the entry indexed by q, q' , written as $\pi_{q,q'}$, satisfying

$$\pi_{q,q'} \triangleq \sum_{\{\sigma \in \Sigma | \delta(q, \sigma) = q'\}} \tilde{\pi}(q, \sigma) \quad (17)$$

and the observation matrix $\tilde{\Pi}$ is a $|Q| \times |\Sigma|$ matrix with the entry indexed by q, σ equaling $\tilde{\pi}(q, \sigma)$.

We note that both Π and $\tilde{\Pi}$ are stochastic, i.e. non-negative with rows summing up to 1.

Definition 7 (Extension of δ and $\tilde{\pi}$ to Σ^*). For any $x = \sigma_1 \dots \sigma_k$, $\delta(q, x)$ is defined recursively by

$$\delta(q, x) \triangleq \delta(\delta(q, \sigma_1 \dots \sigma_{k-1}), \sigma_k) \quad (18)$$

with $\delta(q, \lambda) = q$, and $\tilde{\pi}(q, x)$ is defined recursively by

$$\tilde{\pi}(q, x) \triangleq \prod_{i=1}^k \tilde{\pi}(\delta(q, \sigma_1 \dots \sigma_{i-1}), \sigma_i) \quad (19)$$

with $\tilde{\pi}(q, \lambda) = 1$.

Definition 8 (Strongly Connected PFSA). We say a PFSA is strongly connected if the underlying directed graph is strongly connected [62]. More precisely, a PFSA $G = (Q, \Sigma, \delta, \tilde{\pi})$ is strongly connected if for any pair of distinct states q and $q' \in Q$, there is an $x \in \Sigma^*$ such that $\delta(q, x) = q'$.

We assume all PFSA in the discussions in the sequel are strongly connected if not specified otherwise. For strongly connected PFSA G , there is a unique probability distribution over Q that satisfies $\mathbf{v}^\top \Pi = \mathbf{v}^\top$. This is the **stationary distribution** [63], [64] of G and is denoted as \wp_G , or \wp if G is understood.

Definition 9 (Γ -Expression). We can encode the information contained in δ and $\tilde{\pi}$ by a set of $|Q| \times |Q|$ matrices $\Gamma = \{\Gamma_\sigma | \sigma \in \Sigma\}$, where

$$\Gamma_\sigma |_{q,q'} \triangleq \begin{cases} \tilde{\pi}(q, \sigma) & \text{if } \delta(q, \sigma) = q', \\ 0 & \text{if otherwise.} \end{cases} \quad (20)$$

Γ_σ is called **event-specific transition matrix**, with the event being that σ is current the output. Γ_σ can also be extended to arbitrary $x \in \Sigma^*$ by defining $\Gamma_x = \prod_{i=1}^k \Gamma_{\sigma_i}$ with $\Gamma_\lambda = I$.

Definition 10 (Sequence-Induced Distribution on States). For a PFSA $G = (Q, \Sigma, \delta, \tilde{\pi})$ and a distribution \wp_0 on Q , the **distribution on Q induced by a sequence x** is given by $\wp_{G, \wp_0}^\top(x) = \left[\wp_0^\top \Gamma_x \right]$ with $\wp_{G, \wp_0}(\lambda) = \wp_0$. The entry indexed by $q \in Q$ of the vector $\wp_{G, \wp_0}(x)$ is written as $\wp_{G, \wp_0}(x, q)$. When $\wp_0 = \wp_G$, the stationary distribution of G , we write $\wp_{G, \wp_0}(x)$ as $\wp_G(x)$, or simply as $\wp(x)$, if G is understood.

Definition 11 (Stochastic Process Generated by a PFSA). Let $G = (Q, \Sigma, \delta, \tilde{\pi})$ be a PFSA and let \wp_0 be a distribution on Q , the Σ -valued stochastic process $\{X_t\}_{t \in \mathbb{N}}$ generated by G and \wp_0 satisfies that X_1 follows the distribution \wp_0 and X_{t+1} follows the distribution $\wp_{G, \wp_0}(X_1 \dots X_t)$ for $t \in \mathbb{N}$.

For the rest of this paper, we will assume $\wp_0 = \wp_G$ if not specified otherwise. We can show that, when initialized with \wp_G , the process generated by a PFSA G is stationary and ergodic. We also note the, for the process generate by G , we have $\phi_x = \wp_G(x)^\top \tilde{\Pi}$. Since $\wp_G(\lambda) = \wp_G$, the symbolic derivative of the empty sequence ϕ_λ is the stationary distribution on the symbols.

Definition 12 (Synchronizable PFSA and Synchronizing Sequence). A **synchronizing sequence** is a finite sequence that sends an arbitrary state of the PFSA to a fixed state [65]. To be more precise, let $G = (Q, \Sigma, \delta, \tilde{\pi})$ be a PFSA, we say a sequence $x \in \Sigma^*$ is a synchronizing sequence to a state $q \in Q$ if $\delta(q', x) = q$ for all $q' \in Q$. A PFSA is **synchronizable** if it has at least one synchronizing sequence. Given a sample path generated by a PFSA, we say the PFSA is **synchronized** if a synchronizing sequence transpires in the sample path.

Definition 13 (Equivalence and Irreducibility). Two PFSA G and H are **equivalent** if they generate the same stochastic process. A PFSA G is said to be **irreducible**, if there is not another PFSA with smaller state set that is equivalent to G .

Definition 14. Consider a PFSA G over state set Q . For a give $\varepsilon > 0$, we say a sequence x is a ε -synchronizing sequence to a state $q \in Q$ if

$$\|\wp_G(x) - \mathbf{e}_q\|_\infty \leq \varepsilon. \quad (21)$$

Algorithm 1: GenESeSS

Data: A sequence x over alphabet Σ , $0 < \varepsilon < 1$
Result: State set Q , transition map δ , and transition probability $\tilde{\pi}$
/* **Step One: Approximate ε -synchronizing sequence** */

- 1 Let $L = \lceil \log_{|\Sigma|} 1/\varepsilon \rceil$;
- 2 Calculate the **derivative heap** $\mathcal{D}_\varepsilon^x$ equaling $\{\hat{\phi}_y^x : y \text{ is a sub-sequence of } x \text{ with } |y| \leq L\}$;
- 3 Let \mathcal{C} be the convex hull of $\mathcal{D}_\varepsilon^x$;
- 4 Select x_0 with $\hat{\phi}_{x_0}^x$ being a vertex of \mathcal{C} and has the highest frequency in x ;
- /* **Step Two: Identify transition structure** */
- 5 Initialize $Q = \{q_0\}$;
- 6 Associate to q_0 the **sequence identifier** $x_{q_0}^{\text{id}} = x_0$ and the probability vector $d_{q_0} = \hat{\phi}_{x_0}^x$;
- 7 Let \tilde{Q} be the set of states that are just added and initialize it to be Q ;
- 8 **while** $\tilde{Q} \neq \emptyset$ **do**
- 9 Let $Q_{\text{new}} = \emptyset$ be the set of new states;
- 10 **for** $(q, \sigma) \in \tilde{Q} \times \Sigma$ **do**
- 11 Let $x = x_q^{\text{id}}$ and $d = \hat{\phi}_{x\sigma}^x$;
- 12 **if** $\|d - d_{q'}\|_\infty < \varepsilon$ **for some** $q' \in Q$ **then**
- 13 Let $\delta(q, \sigma) = q'$;
- 14 **else**
- 15 Let $Q_{\text{new}} = Q_{\text{new}} \cup \{q_{\text{new}}\}$ and $Q = Q \cup \{q_{\text{new}}\}$;
- 16 Associate to q_{new} the sequence identifier $x_{q_{\text{new}}}^{\text{id}} = x\sigma$ and the probability vector $d_{q_{\text{new}}} = d$;
- 17 Let $\delta(q, \sigma) = q_{\text{new}}$;
- 18 Let $\tilde{Q} = Q_{\text{new}}$;
- 19 Take a strongly connected subgraph of the labeled directed graph defined by Q and δ , and denote the vertex set of the subgraph again by Q ;
- /* **Step Three: Identify transition probability** */
- 20 Initialize counter $N[q, \sigma]$ for each pair $(q, \sigma) \in Q \times \Sigma$;
- 21 Choose a random starting state $q \in Q$;
- 22 **for** $\sigma \in \Sigma$ **do**
- 23 Let $N[q, \sigma] = N[q, \sigma] + 1$;
- 24 Let $q = \delta(q, \sigma)$;
- 25 Let $\tilde{\pi}(q) = \llbracket (N[q, \sigma])_{\sigma \in \Sigma} \rrbracket$;
- 26 **return** $Q, \delta, \tilde{\pi}$;

While there exists PFSA that is not synchronizable, we can show that an irreducible PFSA always has an ε -synchronizing sequence for some state q for arbitrarily small $\varepsilon > 0$. Moreover, we can show that as length increases, sequences produced by PFSA become uniformly ε -synchronizing. These two are the underpinning properties for the inference algorithm of PFSA (See Alg. 1), because they imply that ϕ_x can be used to approximate $\tilde{\pi}(q)$ if x are properly prefixed and long enough.

Definition 15 (Joint ε -Synchronizing Sequence). *Let G and H be two PFSA over state sets Q_G and Q_H , respectively. For a fixed ε , a sequence x is said to be **jointly ε -synchronizing** to $(q, r) \in Q_G \times Q_H$ if x is ε -synchronizing to q and to r simultaneously. We define*

$$\Sigma_{\varepsilon, (q, r)}^d \triangleq \{x \in \Sigma^d : x \text{ jointly } \varepsilon\text{-synchronizing to } (q, r)\} \quad (22)$$

Definition 16 (Joint Pair of States). *Let G and H be two PFSA over state sets Q_G and Q_H , respectively. Define*

$$p_G(q, r) \triangleq \lim_{d \rightarrow \infty} p_G \left(\Sigma_{\varepsilon, (q, r)}^d \right) \quad (23)$$

*A pair of states $(q, r) \in Q_G \times Q_H$ is called a **G-joint pair** of states if $p_G(q, r) > 0$. We also define*

$$Q_G \triangleq \{(q, r) \in Q_G \times Q_H : (q, r) \text{ is a G-joint pair}\} \quad (24)$$

The inference algorithm for PFSA is called **GenESeSS** for **Generator Extraction Using Self-similar Semantics**. With an input sequence x and a hyperparameter ε , **GenESeSS** outputs a PFSA in the following three steps: 1) approximate an almost synchronizing sequence; 2) identify the transition structure of the PFSA; 3) calculate the transition probabilities of the PFSA. See Alg. 1 [53] for details.

Algorithm 2: Log-likelihood

Data: A PFSA $G = (\Sigma, Q, \delta, \tilde{\pi})$ and a sequence x over alphabet Σ

Result: Log-likelihood $L(x, G)$ of G generating x

- 1 Calculate the state transition matrix Π and observation $\tilde{\Pi}$;
 - 2 Calculate the stationary distribution over states \wp_G of G from Π ;
 - 3 Calculate the stationary distribution of alphabet $\phi_\lambda^\top = \wp_G^\top \tilde{\Pi}$;
 - 4 Initialize \mathbf{p} by \wp_G and \mathbf{q} by ϕ_λ ;
 - 5 Let $L = 0$;
 - 6 **for** i from 1 to $|x|$ **do**
 - 7 Let σ be the i -th entry of x ;
 - 8 Let $L = L - \log \mathbf{q}|_\sigma$;
 - 9 Let $\mathbf{p}^\top = \llbracket \mathbf{p}^\top \Gamma_\sigma \rrbracket$ where Γ_σ is defined in 9;
 - 10 Let $\mathbf{q}^\top = \mathbf{p}^\top \tilde{\Pi}$;
 - 11 **return** $L/|x|$;
-

VI. THEORETICAL DEVELOPMENT OF SEQUENCE LIKELIHOOD DEFECT

Definition 17 (Entropy Rate and KL Divergence). *By entropy rate of a PFSA, we mean the entropy rate of the stochastic process generated by the PFSA [32]. Similarly, by KL divergence of two PFSA, we mean the KL divergence between the two processes generated by them [66]. More precisely, we have*

$$\mathcal{H}(G) = - \lim_{d \rightarrow \infty} \frac{1}{d} \sum_{x \in \Sigma^d} p(x) \log p(x) \quad (25)$$

and the KL divergence

$$\mathcal{D}_{KL}(G \parallel H) = \lim_{d \rightarrow \infty} \frac{1}{d} \sum_{x \in \Sigma^d} p_G(x) \log \frac{p_G(x)}{p_H(x)} \quad (26)$$

whenever the limits exist.

Theorem 1 (Closed-form Formula for Entropy Rate and KL Divergence). *The entropy rate of a PFSA $G = (\Sigma, Q, \delta, \tilde{\pi})$ is given by*

$$\mathcal{H}(G) = \sum_{q \in Q} \wp_G(q) \cdot h(\tilde{\pi}(q)) \quad (27)$$

where $h(\mathbf{v})$ is the based-2 entropy of the probability vector \mathbf{v} .

Consider two PFSA $G = (Q_G, \Sigma, \delta_G, \tilde{\pi}_G)$ and $H = (Q_H, \Sigma, \delta_H, \tilde{\pi}_H)$ with μ_G being absolutely continuous with respect to μ_H . Let Q_c be the set of G -joint pairs of states, we have

$$\mathcal{D}_{KL}(G \parallel H) = \sum_{(q,r) \in Q_c} p_G(q, r) \mathcal{D}_{KL}(\tilde{\pi}_G(q) \parallel \tilde{\pi}_H(r)) \quad (28)$$

Definition 18 (Log-likelihood). *Let $x \in \Sigma^d$, the log-likelihood [32] of a PFSA G generating x is given by*

$$L(x, G) = -\frac{1}{d} \log p_G(x) \quad (29)$$

The calculation of log-likelihood is detailed in Alg. 2.

Theorem 2 (Convergence of log-likelihood). *Let G and H be two reduced PFSA, and let $x \in \Sigma^d$ be a sequence generated by G . Then we have*

$$L(x, H) \rightarrow \mathcal{H}(G) + \mathcal{D}_{KL}(G \parallel H) \quad (30)$$

in probability as $d \rightarrow \infty$.

Proof. We first notice that

$$\sum_{x \in \Sigma^d} p_G(x) \log \frac{p_G(x)}{p_H(x)} = \sum_{x \in \Sigma^{d-1}} \sum_{\sigma \in \Sigma} p_G(x) \wp_G(x) \tilde{\Pi}_G \Big|_\sigma \log \frac{p_G(x) \wp_G(x) \tilde{\Pi}_G \Big|_\sigma}{p_H(x) \wp_H(x) \tilde{\Pi}_H \Big|_\sigma} \quad (31)$$

$$= \sum_{x \in \Sigma^{d-1}} p_G(x) \log \frac{p_G(x)}{p_H(x)} + \underbrace{\sum_{x \in \Sigma^{d-1}} p_G(x) \sum_{\sigma \in \Sigma} \wp_G(x) \tilde{\Pi}_G \Big|_\sigma \log \frac{\wp_G(x) \tilde{\Pi}_G \Big|_\sigma}{\wp_H(x) \tilde{\Pi}_H \Big|_\sigma}}_{\mathcal{D}_d} \quad (32)$$

By induction, we have $\mathcal{D}_{\text{KL}}(G \parallel H) = \lim_{d \rightarrow \infty} \frac{1}{d} \sum_{i=1}^d D_i$, and hence by Cesàro summation theorem [67], we have $\mathcal{D}_{\text{KL}}(G \parallel H) = \lim_{d \rightarrow \infty} D_d$. Let $x = \sigma_1 \sigma_2 \dots \sigma_n$ be a sequence generated by G . Let $x^{[i-1]}$ is the truncation of x at the $(i-1)$ -th symbols, we have

$$-\frac{1}{n} \sum_{i=1}^n \log \varrho_H \left(x^{[i-1]} \right) \tilde{\Pi}_H \Big|_{\sigma_i} = \underbrace{\frac{1}{n} \sum_{i=1}^n \log \frac{\varrho_G \left(x^{[i-1]} \right) \tilde{\Pi}_G \Big|_{\sigma_i}}{\varrho_H \left(x^{[i-1]} \right) \tilde{\Pi}_H \Big|_{\sigma_i}}}_{A_{x,n}} - \underbrace{\frac{1}{n} \sum_{i=1}^n \log \varrho_G \left(x^{[i-1]} \right) \tilde{\Pi}_G \Big|_{\sigma_i}}_{B_{x,n}} \quad (33)$$

Since the stochastic process G generates is ergodic, we have

$$\lim_{n \rightarrow \infty} A_{x,n} = \lim_{d \rightarrow \infty} D_d = \mathcal{D}_{\text{KL}}(G \parallel H) \quad (34)$$

and $\lim_{n \rightarrow \infty} B_{x,n} = \mathcal{H}(G)$. \square

VII. PIPELINE OPTIMIZATION: HYPER-TRAINING, TRAINING, & VALIDATION

Our pipeline comprises a network of individually trained light gradient boosting machine (LGBM) [32] classifiers that focus on complementary aspects of the problem, and operate on different categories of input features as described next. Importantly, some of these features need to be generated non-trivially from the raw data, and these *feature generators* have parameters that need to be trained as well (or comprise models that need to be inferred). We call this inference of the feature-generators as **hyper-training**. Importantly, this is different from the more common notion of hyper-parameters. Hyper-parameters are one or more variables whose scalar values are commonly tuned by grid-search or via some meta-heuristics to optimize classifiers, whereas hyper-training produces generators of features, not simply a set of numbers.

Hyper-training & Training

Trinary Quantization of Medical Histories: The medical histories are mapped into trinary disease-phenotype-specific data-streams to enable generation of some of the features described below, as outlined in Section -C (Step 1).

Feature Categories: The features used in the pipeline maybe categorized as follows:

PFSA scores: The PFSA scores are computed on the basis of the inferred PFSA models as described in the previous sections. The generation of the PFSA models from the trinary data-streams is the first hyper-training step. These scores consist of the negative and positive log-likelihood of a phenotype-specific quantized medical history being generated by the PFSA models for the positive cohort and the control cohort of sex-stratified patients, and the corresponding sequence likelihood defects (See SI-Section -D). Recall that PFSAs are specialized HMMs, and these measures encode the dynamics of the underlying processes, and are sensitive to the ordering, and frequency of the codes at the resolution of the disease phenotypes. Also, recall that diseases phenotypes are broad categories of diagnostic codes, and that we generate PFSA models for each category, and separately for the sexes and the positive cohort and the control cohorts).

Prevalence scores (p-scores): The p-scores focus on individual diagnostic codes, and we create a dictionary of the ratio of relative prevalence of each code (relative to the set of all codes present) in the positive category (for each sex) to the control category. This is the second hyper-training step. In the later steps of the pipeline, we use dictionary look ups to map codes to their p-scores, and also their aggregate measures such as mean, median, and variance to train a downstream LGBM.

Rare scores: These scores consist of a subset of p-scores which correspond to codes with particularly high and low relative prevalences ($p\text{-score} > 2$ or $< .5$). Thus, this feature category depends on the p-score dictionary generated in the second hyper-training step.

Sequence scores: Sequence scores are relatively straight-forward statistical measures such as mean, median, variance, time since last occurrence etc.. on the trinary phenotype-specific sex-stratified histories. No hyper-training is required for the generation of the sequence features.

Thus we require three splits of the training dataset. The first split is used to carry out hyper-training of the PFSA models and the p-score dictionary. The second split is used to train the score-category specific LGBMs, one for each feature category. And the third split is used to train the final LGBM that takes inputs from the outputs of the four LGBMs in the previous layer.

Validation

In validation, or actual prediction of patient fate, we use the trinary mapping, generate the features using the PFSA models and the p-score dictionary, and calculate the raw-risk via the trained LGBM network. The relative score is then obtained by a choice of the operating point reflecting the specificity/sensitivity trade-off discussed before.

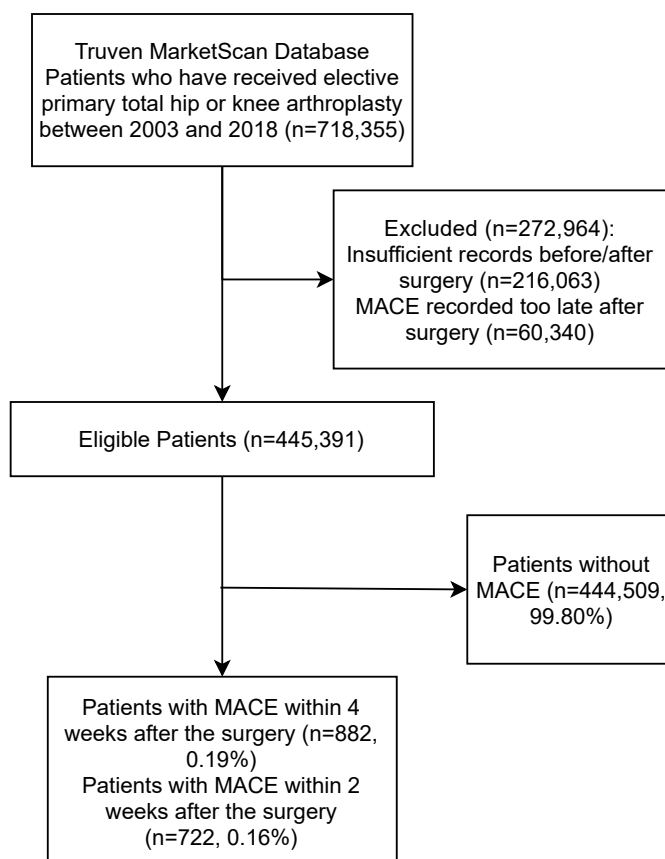


Figure S1: CONSORT diagram conforming to the CONSORT-AI Extension guidelines stated at [https://doi.org/10.1016/S2589-7500\(20\)30218-1](https://doi.org/10.1016/S2589-7500(20)30218-1)

Table S1: Inclusion/Exclusion, Positive/Control Criteria & Cohort Definitions

| | Definitions |
|------------------------------|--|
| Inclusion/Exclusion Criteria | Age 45 - 95 |
| | Has total hip/knee CPT codes (See Table S2) in medical history and length of history available before cardiac event spans ≥ 1 year |
| | Has a myocardial infarction or a cardiac arrest [‡] (See Table S4 for list of target codes used to identify cardiac event in diagnostic history) 4 weeks (2 weeks considered in secondary analysis) after surgery (positive cohort) |
| | Has 0.5 yr of medical history available after surgery (control) |
| Positive & Control Cohorts | Positive Cohort: At least one code for cardiac event (Table S4) |
| | Control Cohort: No code on cardiac event within 26 weeks of surgery |

Table S2: Current Procedural Terminology (CPT) codes for total hip/knee replacement used

| CPT code | description |
|----------|---|
| 27130 | Total Hip Replacement/Resurfacing |
| 27132 | Total Hip Replacement/Resurfacing |
| 81.51 | Total hip replacement |
| 0SR9 | Replacement: Hip Joint, Right |
| 0SRB | Replacement: Hip Joint, Left |
| 27442 | Knee Total Replacement - (Arthroplasty) |
| 27443 | Knee Total Replacement - (Arthroplasty) |
| 27445 | Knee Total Replacement - (Arthroplasty) |
| 27446 | Knee Total Replacement - (Arthroplasty) |
| 27447 | Knee Total Replacement - (Arthroplasty) |
| 81.54 | Total knee replacement |
| 0SRC | Replacement: Knee Joint, Right |
| 0SRD | Replacement: Knee Joint, Left |

Table S3: Codes used to determine RCRI

| Description | Constituent Codes (*NDC: National Drug Code) |
|---|--|
| History of Heart Failure | ICD9 codes: 428.0, 428.1, 428.20, 428.21, 428.22, 428.23, 428.30, 428.31, 428.32, 428.33, 428.40, 428.41, 428.42, 428.43, 428.9, 428.2, 428.3, 428.4; ICD10 codes: I50.1, I50.20, I50.21, I50.22, I50.23, I50.30, I50.31, I50.32, I50.33, I50.40, I50.41, I50.42, I50.43, I50.810, I50.811, I50.812, I50.813, I50.814, I50.82, I50.83, I50.84, I50.89, I50.9, I50.2, I50.3, I50.4, I50.8, I50.81 |
| History of Cerebrovascular Disease | ICD9 codes: 430, 431, 432.0, 432.1, 432.9, 433.00, 433.01, 433.10, 433.11, 433.20, 433.21, 433.30, 433.31, 433.80, 433.81, 433.90, 433.91, 434.00, 434.01, 434.10, 434.11, 434.90, 434.91, 435.0, 435.1, 435.2, 435.3, 435.8, 435.9, 436, 437.0, 437.1, 437.2, 437.3, 437.4, 437.5, 437.6, 437.7, 437.8, 437.9, 438.0, 438.10, 438.11, 438.12, 438.13, 438.14, 438.19, 438.20, 438.21, 438.22, 438.30, 438.31, 438.32, 438.40, 438.41, 438.42, 438.50, 438.51, 438.52, 438.53, 438.6, 438.7, 438.81, 438.82, 438.83, 438.84, 438.85, 438.89, 438.9, 432, 433, 434, 435, 437, 438, 433.0, 433.1, 433.2, 433.3, 433.8, 433.9, 434.0, 434.1, 434.9, 438.1, 438.2, 438.3, 438.4, 438.5, 438.8; ICD10 codes: I60.00, I60.01, I60.02, I60.10, I60.11, I60.12, I60.2, I60.30, I60.31, I60.32, I60.4, I60.50, I60.51, I60.52, I60.6, I60.7, I60.8, I60.9, I61.0, I61.1, I61.2, I61.3, I61.4, I61.5, I61.6, I61.8, I61.9, I62.00, I62.01, I62.02, I62.03, I62.1, I62.9, I63.00, I63.011, I63.012, I63.013, I63.019, I63.02, I63.031, I63.032, I63.033, I63.039, I63.09, I63.10, I63.111, I63.112, I63.113, I63.119, I63.12, I63.131, I63.132, I63.133, I63.139, I63.19, I63.20, I63.211, I63.212, I63.213, I63.219, I63.22, I63.231, I63.232, I63.233, I63.239, I63.29, I63.30, I63.311, I63.312, I63.313, I63.319, I63.321, I63.322, I63.323, I63.329, I63.331, I63.332, I63.333, I63.339, I63.341, I63.342, I63.343, I63.349, I63.39, I63.40, I63.411, I63.412, I63.413, I63.419, I63.421, I63.422, I63.423, I63.429, I63.431, I63.432, I63.433, I63.439, I63.441, I63.442, I63.443, I63.449, I63.49, I63.50, I63.511, I63.512, I63.513, I63.519, I63.521, I63.522, I63.523, I63.529, I63.531, I63.532, I63.533, I63.539, I63.541, I63.542, I63.543, I63.549, I63.59, I63.6, I63.81, I63.89, I63.9, I65.01, I65.02, I65.03, I65.09, I65.1, I65.21, I65.22, I65.23, I65.29, I65.9, I65.99, I66.01, I66.02, I66.03, I66.09, I66.11, I66.12, I66.13, I66.19, I66.21, I66.22, I66.23, I66.29, I66.3, I66.8, I66.9, I67.0, I67.1, I67.2, I67.3, I67.4, I67.5, I67.6, I67.7, I67.81, I67.82, I67.83, I67.841, I67.848, I67.850, I67.858, I67.89, I67.9, I68.0, I68.2, I68.8, I69.00, I69.010, I69.011, I69.012, I69.013, I69.014, I69.015, I69.018, I69.019, I69.020, I69.021, I69.022, I69.023, I69.028, I69.031, I69.032, I69.033, I69.034, I69.039, I69.041, I69.042, I69.043, I69.044, I69.049, I69.051, I69.052, I69.053, I69.054, I69.059, I69.061, I69.062, I69.063, I69.064, I69.065, I69.069, I69.090, I69.091, I69.092, I69.099, I69.093, I69.098, I69.10, I69.110, I69.111, I69.112, I69.113, I69.114, I69.115, I69.118, I69.119, I69.120, I69.121, I69.122, I69.123, I69.128, I69.131, I69.132, I69.133, I69.134, I69.139, I69.141, I69.142, I69.143, I69.144, I69.149, I69.151, I69.152, I69.153, I69.154, I69.159, I69.161, I69.162, I69.163, I69.164, I69.165, I69.169, I69.190, I69.191, I69.192, I69.193, I69.198, I69.20, I69.210, I69.211, I69.212, I69.213, I69.214, I69.215, I69.218, I69.219, I69.220, I69.221, I69.222, I69.223, I69.228, I69.231, I69.232, I69.233, I69.234, I69.239, I69.241, I69.242, I69.243, I69.244, I69.249, I69.251, I69.252, I69.253, I69.254, I69.259, I69.261, I69.262, I69.263, I69.264, I69.265, I69.269, I69.290, I69.291, I69.292, I69.293, I69.298, I69.30, I69.310, I69.311, I69.312, I69.313, I69.314, I69.315, I69.318, I69.319, I69.320, I69.321, I69.322, I69.323, I69.328, I69.331, I69.332, I69.333, I69.334, I69.339, I69.341, I69.342, I69.343, I69.344, I69.349, I69.351, I69.352, I69.353, I69.354, I69.359, I69.361, I69.362, I69.363, I69.364, I69.365, I69.369, I69.390, I69.391, I69.392, I69.393, I69.398, I69.80, I69.810, I69.811, I69.812, I69.813, I69.814, I69.815, I69.818, I69.819, I69.820, I69.821, I69.822, I69.823, I69.828, I69.831, I69.832, I69.833, I69.834, I69.839, I69.841, I69.842, I69.843, I69.844, I69.849, I69.851, I69.852, I69.853, I69.854, I69.859, I69.861, I69.862, I69.863, I69.864, I69.865, I69.869, I69.890, I69.891, I69.892, I69.893, I69.898, I69.90, I69.910, I69.911, I69.912, I69.913, I69.914, I69.915, I69.918, I69.919, I69.920, I69.921, I69.922, I69.923, I69.928, I69.931, I69.932, I69.933, I69.934, I69.939, I69.941, I69.942, I69.943, I69.944, I69.949, I69.951, I69.952, I69.953, I69.954, I69.959, I69.961, I69.962, I69.963, I69.964, I69.965, I69.969, I69.990, I69.991, I69.992, I69.993, I69.998, I60, I61, I62, I63, I65, I66, I67, I68, I69, I60.0, I60.1, I60.3, I60.5, I62.0, I63.0, I63.1, I63.2, I63.3, I63.4, I63.5, I63.8, I65.0, I65.2, I66.0, I66.1, I66.2, I67.8, I69.0, I69.1, I69.2, I69.3, I69.8, I69.9, I63.01, I63.03, I63.11, I63.13, I63.21, I63.23, I63.31, I63.32, I63.33, I63.34, I63.41, I63.42, I63.43, I63.44, I63.51, I63.52, I63.53, I63.54, I67.84, I67.85, I69.01, I69.02, I69.03, I69.04, I69.05, I69.06, I69.09, I69.11, I69.12, I69.13, I69.14, I69.15, I69.16, I69.19, I69.21, I69.22, I69.23, I69.24, I69.25, I69.26, I69.29, I69.31, I69.32, I69.33, I69.34, I69.35, I69.36, I69.39, I69.81, I69.82, I69.83, I69.84, I69.85, I69.86, I69.89, I69.91, I69.92, I69.93, I69.94, I69.95, I69.96, I69.99 |
| History of Ischemic Heart Disease | ICD9 codes: 410.00, 410.01, 410.02, 410.10, 410.11, 410.12, 410.20, 410.21, 410.22, 410.30, 410.31, 410.32, 410.40, 410.41, 410.42, 410.50, 410.51, 410.52, 410.60, 410.61, 410.62, 410.70, 410.71, 410.72, 410.80, 410.81, 410.82, 410.90, 410.91, 410.92, 411.0, 411.1, 411.81, 411.89, 412, 413.0, 413.1, 413.9, 414.00, 414.01, 414.02, 414.03, 414.04, 414.05, 414.06, 414.07, 414.10, 414.11, 414.12, 414.19, 414.2, 414.3, 414.4, 414.8, 414.9, 410, 411, 413, 414, 410.0, 410.1, 410.2, 410.3, 410.4, 410.5, 410.6, 410.7, 410.8, 410.9, 411.8, 414.0, 414.1; ICD10 codes: I20.0, I20.1, I20.8, I20.9, I21.01, I21.02, I21.09, I21.11, I21.19, I21.21, I21.29, I21.3, I21.4, I21.9, I21.A1, I21.A9, I22.0, I22.1, I22.2, I22.8, I22.9, I23.0, I23.1, I23.2, I23.3, I23.4, I23.5, I23.6, I23.7, I23.8, I24.0, I24.1, I24.8, I24.9, I25.10, I25.110, I25.111, I25.118, I25.119, I25.2, I25.3, I25.41, I25.42, I25.5, I25.6, I25.700, I25.701, I25.708, I25.709, I25.710, I25.711, I25.718, I25.719, I25.720, I25.721, I25.728, I25.729, I25.730, I25.731, I25.738, I25.739, I25.750, I25.751, I25.758, I25.759, I25.760, I25.761, I25.768, I25.769, I25.790, I25.791, I25.798, I25.799, I25.810, I25.811, I25.812, I25.82, I25.83, I25.84, I25.89, I25.9, I20, I21, I22, I23, I24, I25, I21.0, I21.1, I21.2, I21.A, I25.1, I25.4, I25.7, I25.8, I25.11, I25.70, I25.71, I25.72, I25.73, I25.75, I25.76, I25.79, I25.81 |
| Pre-operative creatinine > 2 mg/dL / 176.8 µmol/L - Approximated by History of Chronic Kidney Disease | ICD9 codes: 585.3, 585.5, 585.6, 585.4; ICD10 codes: N18.30, N18.31, N18.32, N18.4, N18.5, N18.6, N18.3 |
| Pre-operative treatment with Insulin | NDC* codes: 08881242112, 08881250305, 54868582400, 08881750023, 08881242120, 08881250313, 38396043277, 08881250321, 00169369619, 08881520665, 08881242138, 36652040218, 56151171101, 08881520673, 08496275501, 08881250354, 08496275511, 08881250362, 08080810055, 36652040276, 00002831101, 08080040028, 08080040029, 08080040030, 08396800100, 38396043377, 08881750130, 36652040318, 00002831517, 56151171201, 08881750155, 08881512597, 36652400801, 36652400802, 36652400803, 36652400804, 36652400805, 36652400806, 36652400807, 36652400808, 68258889903, 08881103025, 36652040376, 96295010494, 96295010495, 96295010496, 96295010497, 96295010498, 08881512647, 08396800200, 38396043477, 57515008218, 08881750239, 56151171301, 08881750254, 55948009710, 08881250545, 57515008258, 36652400901, 36652400902, 36652400903, 36652400904, 36652400905, 08080032010, 36652400906, 36652400907, 36652400908, 59060183302, 08881512738, 08881512746, 08396800300, 54569165101, 08881701166, 54569165102, 38396076339, 08881701174, 54274048310, 08290328888, 38396043577, 08222073150, 08881750338, 08222032195, 96295010629, 36652040518, 08881676624, 96295010643, 08881906005, 96295010645, 08881676632, 08881701216, 00002821001, 08881701224, 59060183402, 08881512811, 08080032110, 08222073198, 36652040576, 89134072202, 00002831501, 08396800400, 54569165200, 54569165202, 08881512852, 54274048410, 08881512860, 38396043677, 51927368100, 36652040618, 08881512878, 08881906104, 57515082180, 08080327114, 08080032210, 36652040676, 08881750510, 52297086578, 08881512944, 08287126003, 08287126004, |

Continued on next page

| | |
|---|--|
| <p style="text-align: center;">Pre-operative treatment with Insulin</p> | <p>NDC* codes: 08290843801, 08881601689, 08290843803, 56151170201, 08881601697, 08881200292, 08881601705, 08881700010, 08881601713, 08881601721, 08881200318, 08881200326, 08496315601, 08881601747, 08881200342, 08881601762, 00536993001, 08881601770, 59060231404, 38396042490, 54569295100, 54569295101, 56151170301, 08881200383, 08881716503, 54868131100, 08881716511, 87701445930, 08881200433, 08881716529, 08881601846, 08881200441, 08881716537, 08881601853, 08881601861, 08881020233, 54868589900, 08881200466, 08881601879, 08290008410, 08290008411, 00182312285, 38396042590, 00182312288, 08290844001, 08290008430, 08290008431, 87701044593, 08881200508, 08222072191, 00904396160, 36652400001, 36652400002, 36652400003, 08881200516, 36652400004, 36652400005, 36652400006, 36652400007, 36652400008, 08290008465, 08290008466, 08881200573, 00839802306, 38396042690, 00069070737, 08881135068, 54868532700, 54868532701, 36652400101, 36652400102, 36652400103, 36652400104, 36652400105, 36652400106, 36652400107, 36652400108, 08881135084, 08214502901, 00002841101, 59060231704, 00839802406, 38396042790, 08080621100, 00002821601, 08881520178, 08080621112, 08881700408, 08881520186, 00002824001, 00002821517, 08881200714, 36652400202, 36652400203, 36652400205, 36652400206, 36652400207, 36652400208, 08222072399, 00169033301, 00002879359, 08214503001, 00169352815, 08881200755, 08881520251, 00839802506, 08290328203, 38396042890, 08080826012, 00003183715, 38396042912, 08881200805, 08290328233, 08496291501, 54868238001, 08881716917, 08496291511, 08881847993, 08290328278, 08290328279, 08290328280, 08290328281, 08290328282, 08290328283, 08290320096, 08290328289, 08290328290, 08290328291, 08881250016, 08881250024, 08881676012, 08290320109, 00839802606, 38396042990, 08881250032, 08290320119, 08881250040, 54868623100, 08080826112, 38396043012, 08214355719, 08881250057, 00002811201, 08881250065, 08222072597, 08881250073, 08881250081, 68258897701, 08881250099, 08881250107, 08881512258, 08881250115, 08881250123, 38396043090, 08881250131, 08290328410, 08290328411, 08290328412, 08290328418, 08881250149, 38396043112, 08290328430, 08290328431, 08881250164, 08290328438, 08290328440, 08881250172, 08881250180, 08881053577, 08290320271, 08290328465, 08290328466, 08290328468, 08881250198, 08290328471, 00002854001, 08881250206, 08881250214, 38396043177, 08881250222, 08881070001, 08080818100, 08080818101, 08881250230, 38396043190, 68115070905, 08080818112, 08080818113, 08881250248, 08881250255, 08080220112, 08881250263, 08881160157, 08881250271, 08881242088, 08881250289, 08881750007</p> |
| | |
| | |

Table S4: ICD codes for myocardial infarction used to identify positive cohort

| ICD code | description |
|-----------------|---|
| I46.8 | Cardiac arrest due to other underlying condition |
| I21 | ST elevation (STEMI) myocardial infarction involving left main coronary artery |
| 410.72 | Subendo infarct subseq |
| I21.01 | ST elevation (STEMI) myocardial infarction involving left main coronary artery |
| 410.01 | AMI anterolateral init |
| I21.4 | Non-ST elevation (NSTEMI) myocardial infarction |
| I21.A9 | Other myocardial infarction type |
| I21.A1 | Myocardial infarction type 2 |
| 410.61 | True post infarct init |
| I21.3 | ST elevation (STEMI) myocardial infarction of unspecified site |
| 410.8 | AMI NEC unspecified |
| 410.42 | AMI inferior wall subseq |
| 427.5 | Cardiac arrest |
| I46 | Cardiac arrest due to underlying cardiac condition |
| I46.2 | Cardiac arrest due to underlying cardiac condition |
| 410 | AMI anterolateral unspec |
| 410.71 | Subendo infarct initial |
| 410.11 | AMI anterior wall init |
| 410.12 | AMI anterior wall subseq |
| 410.7 | Subendo infarct unspec |
| I21.21 | ST elevation (STEMI) myocardial infarction involving left circumflex coronary artery |
| 410.4 | AMI inferior wall unspec |
| 410.21 | AMI inferolateral init |
| 410.82 | AMI NEC subsequent |
| 410.9 | AMI NOS unspecified |
| 410.2 | AMI inferolateral unspec |
| I21.9 | Acute myocardial infarction unspecified |
| I46.9 | Cardiac arrest cause unspecified |
| 410.1 | AMI anterior wall unspec |
| 410.02 | AMI anterolateral subseq |
| 410.51 | AMI lateral NEC initial |
| 410.52 | AMI lateral NEC subseq |
| 410.92 | AMI NOS subsequent |
| I21.02 | ST elevation (STEMI) myocardial infarction involving left anterior descending coronary artery |
| I21.19 | ST elevation (STEMI) myocardial infarction involving other coronary artery of inferior wall |
| 410.81 | AMI NEC initial |
| 410.41 | AMI inferior wall init |
| 410.31 | AMI inferopost initial |
| 410.62 | True post infarct subseq |
| I21.09 | ST elevation (STEMI) myocardial infarction involving other coronary artery of anterior wall |
| 410.0 | AMI anterolateral unspec |
| 410.5 | AMI lateral NEC unspec |
| 410.6 | True post infarct unspec |
| 410.3 | AMI inferopost unspec |
| 410.91 | AMI NOS initial |
| 410.32 | AMI inferopost subseq |
| I21.11 | ST elevation (STEMI) myocardial infarction involving right coronary artery |
| 410.22 | AMI inferolateral subseq |
| I21.29 | ST elevation (STEMI) myocardial infarction involving other sites |

Table S5: Number of diagnostic codes encountered in dataset

| gender | Number of codes | Number of unique codes |
|--------|-----------------|------------------------|
| M | 4879398 | 17554 |
| F | 7753318 | 19209 |
| Total | 12632716 | 36763 |

Table S6: CCoR phenotypes and maximum number of unique ICD codes defining CCoR phenotypes

| CCoR phenotype | count of ICD codes in definition (Table S7) |
|-----------------------------|--|
| Allergic | 191 |
| Cardiovascular | 2017 |
| CNS | 765 |
| Development | 820 |
| Digestive | 1317 |
| Endocrine | 237 |
| Frailty | 557 |
| Health-Services | 501 |
| Hematologic | 429 |
| Hypertension | 80 |
| Immune | 1546 |
| Infections-Bacterial | 409 |
| Infections-Fungal-and-Other | 784 |
| Infections-General | 3612 |
| Infections-Respiratory | 712 |
| Injuries | 53265 |
| Integumentary | 1457 |
| Metabolic | 373 |
| Musculoskeletal | 7533 |
| Neoplastic | 3022 |
| Ophthalmological | 3401 |
| Otic | 856 |
| PNS | 394 |
| Psychiatric | 1478 |
| Reproductive | 2675 |
| Respiratory | 724 |

Table S8: Feature Definitions (Total number of features used: 380)

| Feature name | Explanation | T ₁ -features |
|---|---|--------------------------|
| feature scores relative to phenotype score | Mean p-score of feature codes within sequence divided by general p-score of feature | 26 |
| feature scores relative to whole score | Mean p-score of feature codes within sequence divided by mean p-score of all codes in the record | 26 |
| aggregation score | aggregation of the p-scores in the record | 13 |
| high scores proportion | proportion of codes with very high p-scores among all codes in the record | 1 |
| low scores proportion | proportion of codes with very low p-scores among all codes in the record | 1 |
| dynamics of mean score | mean p-score of second half of the record divided by mean p-score of first half of the record | 1 |
| dynamics of geometric mean score | geometric mean p-score of second half of the record divided by mean p-score of first half of the record | 1 |
| dynamics of st.dev score | standard deviation of p-scores of second half of the record divided by standard deviation of p-scores of first half of the record | 1 |
| dynamics of score range | range of p-scores of second half of the record divided by range of p-scores of first half of the record | 1 |
| dynamics of score skew | skew of p-scores of second half of the record divided by skew of p-scores of first half of the record | 1 |
| aggregation relative to phn score | aggregation of all feature 's mean scores divided by corresponding general p-score of feature | 9 |
| aggregation relative to whole score | aggregation of all feature 's mean scores divided by mean p-score of all codes in the record | 9 |
| predicted risk from pfsa model | predicted risk from pfsa model | 1 |
| predicted risk from seq model | predicted risk from seq model | 1 |
| predicted risk from pscore model | predicted risk from pscore model | 1 |
| predicted risk from rare model | predicted risk from rare model | 1 |
| age at screening | Patient age at the moment of the screening | 1 |
| feature proportion | Ratio of number of weeks with the codes of a given phenotype to the total number of weeks in sequence | 26 |
| feature prevalence | Ratio of number of weeks with the codes of a given phenotype to the number of weeks with any diagnosis code recorded | 26 |
| feature first incident | Time interval from observation date to the first phenotype code, normalized by record length | 26 |
| feature last incident | Time interval from observation date to the last phenotype code, normalized by record length | 26 |
| feature mean position | Mean time position of phenotype codes in the record, normalized by record length | 26 |
| feature streak | Length of the longest uninterrupted subsequence of weeks with the codes of a given phenotype recorded | 26 |
| Max/Mean/Std/Range intermission | Maximum/Mean/Standard Deviation/Range of the lengths of subsequences of consequent weeks with codes | 4 |
| Max/Mean/Std cluster | Maximum/Mean/Standard Deviation of the lengths of subsequences of consequent weeks without codes | 3 |
| Max/Std/Range prevalence | Maximum/Standard Deviation/Range of the phenotype prevalences | 3 |
| Density of DX Record | Proportion of weeks in a record observed where at least one DX code was recorded | 1 |
| feature | Sequence Likelihood Defect for a given phenotype | 26 |
| feature neg llk | Negative LogLikelihood score for a given phenotype | 26 |
| feature pos llk | Positive LogLikelihood score for a given phenotype | 26 |
| feature llk ratio | Ratio of Positive to Negative LogLikelihood score for a given phenotype | 26 |
| Mean Δ | Mean Sequence Likelihood Defect | 1 |
| Std. deviation Δ | Range of Sequence Likelihood Defects | 1 |
| Range Δ | Standard Deviation of Sequence Likelihood Defects | 1 |
| Mean neg llk | Mean Negative LogLikelihood score | 1 |
| Range neg llk | Range of Negative LogLikelihood score | 1 |
| Std. deviation neg llk | Standard Deviation of Negative LogLikelihood score | 1 |
| Mean pos llk | Mean Positive LogLikelihood score | 1 |
| Range pos llk | Range of Positive LogLikelihood score | 1 |
| Std. deviation pos llk | Standard Deviation of Positive LogLikelihood score | 1 |
| Mean llk ratio | Mean LogLikelihood score ratio | 1 |
| Range llk ratio | Range of LogLikelihood score ratio | 1 |
| Std. deviation llk ratio | Standard Deviation of LogLikelihood score ratio | 1 |
| high scores proportion | proportion of codes with very high p-scores among all codes in the record | 1 |
| low scores proportion | proportion of codes with very low p-scores among all codes in the record | 1 |

* Δ : Sequence Likelihood Defect (See Methods)

† neg llk: loglikelihood of observed sequence being generated by the model inferred from control (See Methods)

‡ pos llk: loglikelihood of observed sequence being generated by the model inferred from positive (See Methods)

Table S9: Proportion of 0's, 1's and 2's on average in trinary encodings with 95% CI

| cohort | sex | proportion of 0 | proportion of 1 | proportion of 2 |
|----------|--------|-----------------|-----------------|-----------------|
| control | Female | 0.879 ± 0.002 | 0.013 ± 0.001 | 0.106 ± 0.002 |
| control | Male | 0.891 ± 0.003 | 0.012 ± 0.001 | 0.095 ± 0.003 |
| positive | Female | 0.858 ± 0.040 | 0.015 ± 0.013 | 0.126 ± 0.038 |
| positive | Male | 0.873 ± 0.037 | 0.014 ± 0.012 | 0.111 ± 0.035 |
| TOTAL | Female | 0.879 ± 0.002 | 0.013 ± 0.001 | 0.106 ± 0.002 |
| TOTAL | Male | 0.891 ± 0.002 | 0.012 ± 0.001 | 0.095 ± 0.002 |

Table S10: Out-of-sample performance achieved (mean AUC) when training dataset is balanced (Note: performance degrades as we attempt to train with more balanced data, e.g., downsampling ratio of 1 is the case where we sample the control cohort to use only as many patients as in the positive cohort)

| sex | downsampling ratio | all patients | age 65+ years | age < 65 years | frail |
|--------|--------------------|--------------|---------------|----------------|-------|
| Female | 1 | 0.755 | 0.715 | 0.694 | 0.732 |
| Female | 2 | 0.756 | 0.723 | 0.700 | 0.736 |
| Female | 5 | 0.768 | 0.735 | 0.727 | 0.752 |
| Female | 10 | 0.781 | 0.750 | 0.737 | 0.769 |
| Female | 20 | 0.772 | 0.750 | 0.728 | 0.759 |
| Female | 40 | 0.790 | 0.760 | 0.743 | 0.772 |
| Male | 1 | 0.724 | 0.665 | 0.690 | 0.721 |
| Male | 2 | 0.743 | 0.701 | 0.698 | 0.746 |
| Male | 5 | 0.754 | 0.708 | 0.722 | 0.761 |
| Male | 10 | 0.751 | 0.711 | 0.718 | 0.757 |
| Male | 20 | 0.759 | 0.718 | 0.725 | 0.759 |
| Male | 40 | 0.759 | 0.714 | 0.729 | 0.759 |

Table S11: Cohort Sizes

| sex | cardiac event within week | n _{positive} | n _{control} | n _{high risk} | n |
|-----|---------------------------|-----------------------|----------------------|------------------------|--------|
| M | 2 | 385 | 185528 | 146782 | 185913 |
| M | 4 | 464 | 185528 | 146782 | 185992 |
| F | 2 | 337 | 258981 | 204170 | 259318 |
| F | 4 | 418 | 258981 | 204170 | 259399 |
| | Total | 882 | 444509 | 350952 | 445391 |

Table S12: Out-of-sample predictive performance to predict MACE 4 weeks after surgery in sub-cohorts with pre-existing conditions

| Pre-existing phenotype | Female CCoR | Female RCRI | Male CCoR | Male RCRI |
|-----------------------------|-------------|-------------|-----------|-----------|
| Allergic | 0.77 | 0.71 | 0.81 | 0.78 |
| CNS | 0.80 | 0.67 | 0.89 | 0.75 |
| Cardiovascular | 0.78 | 0.69 | 0.80 | 0.67 |
| Development | 0.77 | 0.83 | 0.86 | 0.68 |
| Digestive | 0.81 | 0.73 | 0.80 | 0.71 |
| Endocrine | 0.80 | 0.69 | 0.80 | 0.67 |
| Frailty | 0.78 | 0.69 | 0.85 | 0.73 |
| Health Services | 0.81 | 0.71 | 0.83 | 0.71 |
| Hematologic | 0.80 | 0.72 | 0.85 | 0.74 |
| Hypertension | 0.77 | 0.68 | 0.80 | 0.66 |
| Immune | 0.81 | 0.71 | 0.82 | 0.70 |
| Infections Fungal and Other | 0.80 | 0.76 | 0.84 | 0.68 |
| Infections General | 0.80 | 0.75 | 0.84 | 0.68 |
| Infections Respiratory | 0.76 | 0.64 | 0.83 | 0.63 |
| Injuries | 0.79 | 0.73 | 0.84 | 0.69 |
| Integumentary | 0.78 | 0.69 | 0.80 | 0.72 |
| Metabolic | 0.80 | 0.70 | 0.82 | 0.70 |
| Musculoskeletal | 0.81 | 0.71 | 0.83 | 0.70 |
| Neoplastic | 0.87 | 0.75 | 0.78 | 0.68 |
| Ophthalmological | 0.79 | 0.69 | 0.76 | 0.65 |
| Otic | 0.79 | 0.75 | 0.84 | 0.62 |
| PNS | 0.80 | 0.70 | 0.84 | 0.73 |
| Psychiatric | 0.85 | 0.74 | 0.89 | 0.72 |
| Reproductive | 0.80 | 0.70 | 0.82 | 0.79 |
| Respiratory | 0.81 | 0.70 | 0.83 | 0.67 |

Table S13: Out-of-sample predictive performance in sub-cohorts stratified by age

| age | gender | auc CCoR | auc RCRI | n _{positive} | n _{control} |
|---------|--------|----------|----------|-----------------------|----------------------|
| 45 - 55 | F | 0.59 | 0.58 | 3 | 9056 |
| 45 - 55 | M | 0.89 | 0.79 | 5 | 7027 |
| 55 - 65 | F | 0.80 | 0.66 | 31 | 27256 |
| 55 - 65 | M | 0.78 | 0.63 | 39 | 20244 |
| 65 - 75 | F | 0.81 | 0.73 | 34 | 14235 |
| 65 - 75 | M | 0.73 | 0.58 | 31 | 9635 |
| 75 - 85 | F | 0.70 | 0.65 | 25 | 8515 |
| 75 - 85 | M | 0.79 | 0.71 | 36 | 5164 |
| 85 - 95 | F | 0.80 | 0.67 | 8 | 1578 |
| 85 - 95 | M | 0.75 | 0.48 | 12 | 847 |

Table S14: Out-of-sample* performance for predicting MACE with 4 weeks of Hip or Knee Arthroplasty (Primary Endpoint) at 99% Specificity: CCoR vs. RCRI**

| sex | cohort | model | sensitivity | PPV | acc | LR+ | LR- | AUC |
|--------|---------------|-------|-------------|-------------|-------------|-----------|-----------|-------------|
| Female | < 65 | RCRI | 0.01±0.02 | 0.008±0.000 | 0.987±0.006 | 0.04±0.1 | 1.00±0.02 | 0.639±0.039 |
| Female | < 65 | CCoR | 0.14±0.06 | 0.047±0.004 | 0.987±0.006 | 14.12±1.1 | 0.87±0.06 | 0.775±0.035 |
| Male | < 65 | RCRI | 0.07±0.03 | 0.025±0.003 | 0.987±0.000 | 7.42±0.8 | 0.94±0.03 | 0.682±0.034 |
| Male | < 65 | CCoR | 0.15±0.06 | 0.065±0.003 | 0.987±0.006 | 19.94±0.9 | 0.85±0.06 | 0.783±0.030 |
| Female | 65+ | RCRI | 0.03±0.01 | 0.012±0.002 | 0.987±0.000 | 3.39±0.5 | 0.97±0.01 | 0.664±0.028 |
| Female | 65+ | CCoR | 0.09±0.06 | 0.036±0.004 | 0.987±0.006 | 10.70±1.1 | 0.92±0.06 | 0.771±0.025 |
| Male | 65+ | RCRI | 0.03±0.01 | 0.011±0.001 | 0.987±0.006 | 3.17±0.4 | 0.98±0.00 | 0.661±0.026 |
| Male | 65+ | CCoR | 0.09±0.05 | 0.031±0.002 | 0.987±0.006 | 9.09±0.6 | 0.92±0.05 | 0.762±0.023 |
| Female | all patients | RCRI | 0.05±0.01 | 0.016±0.003 | 0.987±0.000 | 4.67±0.8 | 0.96±0.01 | 0.688±0.023 |
| Female | all patients | CCoR | 0.13±0.01 | 0.044±0.007 | 0.987±0.006 | 13.19±2.1 | 0.88±0.01 | 0.801±0.019 |
| Male | all patients | RCRI | 0.05±0.01 | 0.019±0.002 | 0.987±0.000 | 5.44±0.7 | 0.95±0.01 | 0.705±0.020 |
| Male | all patients | CCoR | 0.12±0.05 | 0.042±0.001 | 0.987±0.006 | 12.44±0.3 | 0.89±0.05 | 0.802±0.018 |
| Female | frail*** | RCRI | 0.03±0.00 | 0.009±0.002 | 0.987±0.006 | 2.59±0.6 | 0.98±0.00 | 0.670±0.028 |
| Female | frail | CCoR | 0.11±0.03 | 0.036±0.002 | 0.987±0.000 | 10.36±0.7 | 0.90±0.03 | 0.791±0.025 |
| Male | frail | RCRI | 0.06±0.04 | 0.020±0.002 | 0.987±0.006 | 5.91±0.7 | 0.95±0.04 | 0.727±0.027 |
| Male | frail | CCoR | 0.15±0.03 | 0.050±0.002 | 0.987±0.000 | 15.02±0.7 | 0.86±0.03 | 0.810±0.024 |
| Female | high risk**** | RCRI | 0.02±0.00 | 0.008±0.001 | 0.987±0.006 | 2.19±0.3 | 0.99±0.00 | 0.581±0.029 |
| Female | high risk | CCoR | 0.07±0.03 | 0.026±0.002 | 0.987±0.000 | 7.73±0.6 | 0.94±0.03 | 0.737±0.026 |
| Male | high risk | RCRI | 0.03±0.01 | 0.010±0.001 | 0.987±0.006 | 2.91±0.3 | 0.98±0.00 | 0.617±0.026 |
| Male | high risk | CCoR | 0.09±0.05 | 0.033±0.002 | 0.987±0.006 | 9.61±0.7 | 0.92±0.05 | 0.729±0.024 |
| Female | low risk† | CCoR | 0.22±0.04 | 0.071±0.004 | 0.987±0.000 | 21.81±1.3 | 0.79±0.04 | 0.765±0.036 |
| Male | low risk | CCoR | 0.11±0.04 | 0.042±0.003 | 0.987±0.000 | 5.96±0.9 | 0.90±0.04 | 0.766±0.032 |

Abbreviations, AUC, area under the receiver operating characteristic curve; CCoR, Cardiac Co-Morbidity Risk Score; LR+, positive likelihood ratio; LR-, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; acc, accuracy; RCRI, Revised Cardiac Risk Index

*50% (n=445391) of cohort used for validation

**Because of insufficient availability of relevant laboratory data in the Truven dataset, presence of at least one diagnostic code for chronic kidney disease stage III or higher in the medical record in the year before the date of arthroplasty was used as a surrogate for the RCRI condition, serum creatinine concentration > 2.0mg/dL (to convert to micromoles per liter, multiply by 88.4).

***Frail subcategory was defined by codes specified in Table S7

****Low risk subcohort comprises patients with RCRI score 0. High risk patients have RCRI score > 0.

†No RCRI performance logged for low-risk patients, since their RCRI score is zero.

Table S15: Out-of-sample* performance for predicting MACE with 2 weeks of Hip or Knee Arthroplasty (Secondary Endpoint) at 99% Specificity: CCoR vs. RCRI**

| sex | cohort | model | sensitivity | PPV | acc | LR+ | LR- | AUC |
|--------|---------------|-------|-------------|---------------|---------------|-------------|-------------|---------------|
| Female | < 65 | RCRI | 0.01 ± 0.02 | 0.009 ± 0.000 | 0.987 ± 0.006 | 0.04 ± 0.1 | 1.00 ± 0.02 | 0.647 ± 0.044 |
| Female | < 65 | CCoR | 0.11 ± 0.01 | 0.073 ± 0.023 | 0.987 ± 0.006 | 22.87 ± 8.5 | 0.90 ± 0.01 | 0.787 ± 0.039 |
| Male | < 65 | RCRI | 0.09 ± 0.02 | 0.032 ± 0.009 | 0.987 ± 0.000 | 9.33 ± 2.8 | 0.92 ± 0.02 | 0.688 ± 0.037 |
| Male | < 65 | CCoR | 0.14 ± 0.04 | 0.056 ± 0.004 | 0.987 ± 0.000 | 16.83 ± 1.3 | 0.87 ± 0.04 | 0.797 ± 0.033 |
| Female | 65+ | RCRI | 0.04 ± 0.04 | 0.013 ± 0.002 | 0.987 ± 0.006 | 3.66 ± 0.6 | 0.97 ± 0.03 | 0.671 ± 0.030 |
| Female | 65+ | CCoR | 0.10 ± 0.03 | 0.041 ± 0.002 | 0.987 ± 0.000 | 12.08 ± 0.7 | 0.91 ± 0.04 | 0.787 ± 0.027 |
| Male | 65+ | RCRI | 0.03 ± 0.00 | 0.010 ± 0.001 | 0.987 ± 0.000 | 2.79 ± 0.4 | 0.98 ± 0.00 | 0.667 ± 0.028 |
| Male | 65+ | CCoR | 0.09 ± 0.03 | 0.032 ± 0.002 | 0.987 ± 0.000 | 9.32 ± 0.6 | 0.92 ± 0.03 | 0.780 ± 0.025 |
| Female | all patients | RCRI | 0.05 ± 0.01 | 0.018 ± 0.003 | 0.987 ± 0.000 | 5.17 ± 0.9 | 0.96 ± 0.01 | 0.692 ± 0.025 |
| Female | all patients | CCoR | 0.14 ± 0.06 | 0.048 ± 0.001 | 0.987 ± 0.006 | 14.48 ± 0.3 | 0.87 ± 0.06 | 0.809 ± 0.021 |
| Male | all patients | RCRI | 0.05 ± 0.01 | 0.018 ± 0.002 | 0.987 ± 0.000 | 5.15 ± 0.7 | 0.96 ± 0.01 | 0.710 ± 0.022 |
| Male | all patients | CCoR | 0.14 ± 0.02 | 0.047 ± 0.007 | 0.987 ± 0.000 | 13.98 ± 2.1 | 0.87 ± 0.02 | 0.813 ± 0.019 |
| Female | frail*** | RCRI | 0.03 ± 0.01 | 0.012 ± 0.003 | 0.987 ± 0.000 | 3.43 ± 0.9 | 0.98 ± 0.01 | 0.676 ± 0.032 |
| Female | frail | CCoR | 0.12 ± 0.03 | 0.041 ± 0.009 | 0.987 ± 0.000 | 11.78 ± 2.7 | 0.89 ± 0.03 | 0.807 ± 0.027 |
| Male | frail | RCRI | 0.06 ± 0.01 | 0.020 ± 0.003 | 0.987 ± 0.000 | 5.82 ± 0.9 | 0.95 ± 0.01 | 0.736 ± 0.029 |
| Male | frail | CCoR | 0.17 ± 0.03 | 0.059 ± 0.009 | 0.987 ± 0.000 | 17.76 ± 3.0 | 0.84 ± 0.03 | 0.825 ± 0.025 |
| Female | high risk**** | RCRI | 0.02 ± 0.03 | 0.008 ± 0.001 | 0.987 ± 0.006 | 2.35 ± 0.4 | 0.99 ± 0.02 | 0.584 ± 0.032 |
| Female | high risk | CCoR | 0.08 ± 0.00 | 0.029 ± 0.006 | 0.987 ± 0.006 | 8.45 ± 1.9 | 0.93 ± 0.01 | 0.742 ± 0.028 |
| Male | high risk | RCRI | 0.03 ± 0.03 | 0.009 ± 0.001 | 0.987 ± 0.006 | 2.69 ± 0.4 | 0.98 ± 0.02 | 0.628 ± 0.028 |
| Male | high risk | CCoR | 0.09 ± 0.02 | 0.033 ± 0.007 | 0.987 ± 0.000 | 9.66 ± 2.3 | 0.92 ± 0.02 | 0.737 ± 0.026 |
| Female | low risk† | CCoR | 0.28 ± 0.07 | 0.092 ± 0.002 | 0.988 ± 0.006 | 28.78 ± 0.7 | 0.72 ± 0.07 | 0.779 ± 0.040 |
| Male | low risk | CCoR | 0.12 ± 0.04 | 0.047 ± 0.004 | 0.987 ± 0.000 | 6.78 ± 1.1 | 0.89 ± 0.05 | 0.793 ± 0.035 |

Abbreviations, AUC, area under the receiver operating characteristic curve; CCoR, Cardiac Co-Morbidity Risk Score; LR+, positive likelihood ratio; LR-, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; acc, accuracy; RCRI, Revised Cardiac Risk Index

*50% (n=445391) of cohort used for validation

**Because of insufficient availability of relevant laboratory data in the Truven dataset, presence of at least one diagnostic code for chronic kidney disease stage III or higher in the medical record in the year before the date of arthroplasty was used as a surrogate for the RCRI condition, serum creatinine concentration > 2.0mg/dL (to convert to micromoles per liter, multiply by 88.4).

***Frail subcategory was defined by codes specified in Table S7

****Low risk subcohort comprises patients with RCRI score 0. High risk patients have RCRI score > 0.

†No RCRI performance logged for low-risk patients, since their RCRI score is zero.

Table S16: Statistical significance of CCoR AUC > RCRI AUC (* denotes significance at 95% level, ** denotes significance at 99% level)

| sex | model | weeks after surgery | RCRI auc | CCoR auc | p value | significance |
|--------|------------------------|---------------------|----------|----------|---------|--------------|
| Female | all patients | 2 | 0.692 | 0.809 | 0.010 | ** |
| | | 4 | 0.688 | 0.801 | 0.007 | ** |
| | frail [†] | 2 | 0.676 | 0.807 | 0.017 | * |
| | | 4 | 0.670 | 0.791 | 0.014 | * |
| | high risk [‡] | 2 | 0.584 | 0.742 | 0.005 | ** |
| | | 4 | 0.581 | 0.737 | 0.003 | ** |
| | 65+ | 2 | 0.671 | 0.787 | 0.020 | * |
| | | 4 | 0.664 | 0.771 | 0.020 | * |
| | < 65 | 2 | 0.647 | 0.787 | 0.036 | * |
| | | 4 | 0.639 | 0.775 | 0.028 | * |
| Male | all patients | 2 | 0.710 | 0.813 | 0.010 | * |
| | | 4 | 0.705 | 0.802 | 0.009 | ** |
| | frail | 2 | 0.736 | 0.825 | 0.045 | * |
| | | 4 | 0.727 | 0.810 | 0.042 | * |
| | high risk | 2 | 0.628 | 0.737 | 0.017 | * |
| | | 4 | 0.617 | 0.729 | 0.011 | * |
| | 65+ | 2 | 0.667 | 0.780 | 0.017 | * |
| | | 4 | 0.661 | 0.762 | 0.019 | * |
| | < 65 | 2 | 0.688 | 0.797 | 0.047 | * |
| | | 4 | 0.682 | 0.783 | 0.042 | * |

[†]Frail subcategory was defined by codes specified in Table S7

[‡]Low risk subcohort comprises patients with RCRI score 0. High risk patients have RCRI score > 0.

Table S17: Out-of-sample performance achieved using only PFSA component of the CCoR model (Note: the performance is significantly degraded, with all p-values < 0.01.)

| sex | prediction horizon | AUC |
|--------|--------------------|---------------|
| Female | 2 weeks | 0.696 ± 0.082 |
| Female | 4 weeks | 0.698 ± 0.074 |
| Male | 2 weeks | 0.679 ± 0.074 |
| Male | 4 weeks | 0.656 ± 0.067 |

REFERENCES

- [52] C. W. J. Granger and R. Joyeux, "An introduction to long-memory time series models and fractional differencing," *Journal of Time Series Analysis*, vol. 1, no. 1, pp. 15–29.
- [53] I. Chattopadhyay and H. Lipson, "Abductive learning of quantized stochastic processes with probabilistic finite automata," *Philos Trans A*, vol. 371, no. 1984, p. 20110543, Feb 2013.
- [54] "General equivalence mappings." [Online]. Available: https://www.cms.gov/Medicare/Coding/ICD10/downloads/ICD-10_GEM_fact_sheet.pdf
- [55] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 03 1951. [Online]. Available: <https://doi.org/10.1214/aoms/1177729694>
- [56] J. Doob, *Stochastic Processes*, ser. Wiley Publications in Statistics. John Wiley & Sons, 1953. [Online]. Available: <https://books.google.com/books?id=KvJQAAAAMAAJ>
- [57] J. H. Friedman, "Stochastic gradient boosting," *Comput. Stat. Data Anal.*, vol. 38, no. 4, pp. 367–378, Feb. 2002. [Online]. Available: [http://dx.doi.org/10.1016/S0167-9473\(01\)00065-2](http://dx.doi.org/10.1016/S0167-9473(01)00065-2)
- [58] J. E. Hopcroft, *Introduction to automata theory, languages, and computation*. Pearson Education India, 2008.
- [59] A. Klenke, *Probability theory: a comprehensive course*. Springer Science & Business Media, 2013.
- [60] I. Chattopadhyay and A. Ray, "Structural transformations of probabilistic finite state machines," *International Journal of Control*, vol. 81, no. 5, pp. 820–835, 2008.
- [61] I. Chattopadhyay and H. Lipson, "Data smashing: uncovering lurking order in data," *Journal of The Royal Society Interface*, vol. 11, no. 101, p. 20140826, 2014.
- [62] J. Bondy and U. Murty, "Graph theory (2008)," *Grad. Texts in Math*, 2008.
- [63] M. Vidyasagar, *Hidden markov processes: Theory and applications to biology*. Princeton University Press, 2014, vol. 44.
- [64] L. C. Kai, *Markov Chains: With Stationary Transition Probabilities*. Springer-Verlag, 1967.
- [65] A. N. Trahtman, "The road coloring and Černý conjecture," in *Proc. of Prague stringology conference*, vol. 1. Citeseer, 2008, p. 12.
- [66] A. G. d. G. Matthews, J. Hensman, R. Turner, and Z. Ghahramani, "On sparse variational methods and the kullback-leibler divergence between stochastic processes," *Journal of Machine Learning Research*, vol. 51, pp. 231–239, 2016.
- [67] G. Hardy, "Divergent series, with a preface by je littlewood and a note by ls bosanquet, reprint of the revised (1963) edition," *Éditions Jacques Gabay, Sceaux*, 1992.