Development and validation of an EHR-based frailty index for predicting infections, non-operative trauma, cardiac complications, and patient safety indicators

Alex Bokov 1,✉, Sara Espinoza 1, Chandana Tripathy 1, and Kathleen R. Stevens 1

2021-08-08

We conducted a retrospective cohort study on structured data of 14,843 randomly sampled from all adult patients in the Epic EHR system of a large South Texas academic health center and its teaching hospital partner to determine whether our implementation of the Rockwood frailty index (eFI) can predict patient safety indicators (PSIs) using only codes and numeric values routinely collected in EHR systems. According to both the c-statistic and the AIC of the Cox proportional hazard models we used to analyze the data, eFI performs well as a predictor of infections, non-operative trauma, cardiac complications, and to a lesser extent the ocurrence of any PSI.

# Introduction/Background

…

Improving patient safety is a key goal for every health system. A simple and rapid frailty assessment would help clinicians can tailor interventions to avoid poor patient outcomes. Our objective was to evaluate a deficit-accumulation electronic frailty index (eFI) as a predictor of preventable infections, non-operative trauma, cardiac complications, and general patient safety indicators (PSIs) and then validate it against hold-out data.

Frailty is the lifelong erosion of stress resistance and accumulation of impairments across multiple physiological systems. Among older community dwelling adults 32% have been classified as pre-frail and 24% have been classified as frail (Hoover *et al.*, 2013). Frailty predicts disability, falls, and mortality (Pajewski *et al.*, 2019); emergency room visits and hospitalizations (Fried *et al.*, 2001); and long-term care admissions (Rockwood, 2005; Rockwood *et al.*, 2006; Pajewski *et al.*, 2019). The Fried phenotype (Fried *et al.*, 2001) and Rockwood deficit accumulation index (Mitnitski, Mogilner and Rockwood, 2001; Rockwood and Mitnitski, 2007), are two commonly used measures of frailty. There is reasonable convergence between these two approaches (Malmstrom, Miller and Morley, 2014; Li *et al.*, 2015) but the latter is better suited to large-scale deployment in EHR systems because it does not require patient questionnaires nor dedicated physical assessments. Kulminsky Kulminski *et al.* (2008) found evidence for the Rockwood index outperforming the Fried phenotype.

Our version of the eFI improves on previous EHR adaptations of the Rockwood index (Clegg *et al.*, 2016; Pajewski *et al.*, 2019) by eliminating the need for manually curated lab reference ranges and by calculating the eFI over a rolling two-year window for every available patient encounter rather than at one fixed time point. The scripts, code-mappings, and usage notes are freely available online.

# Methods

## Source of data

…

A random 1% sample (14,843 patients, 312,849 visit-days) was drawn from deidentified i2b2 data warehouse records representing the entire patient population in the EHR system of a large academic health center and its teaching hospital partner.

The nominal date of index encounters ranged between 2005-01-20 and 2020-03-05 while that of final encounters ranged between 2005-01-25 and 2020-04-10. The actual dates were obscured by deidentification of the data prior to delivery to the authors but this date-shifting was within a one-year time window of the actual dates and was done in a manner that preserved the chronological order of each patient’s encounters and the duration between them.

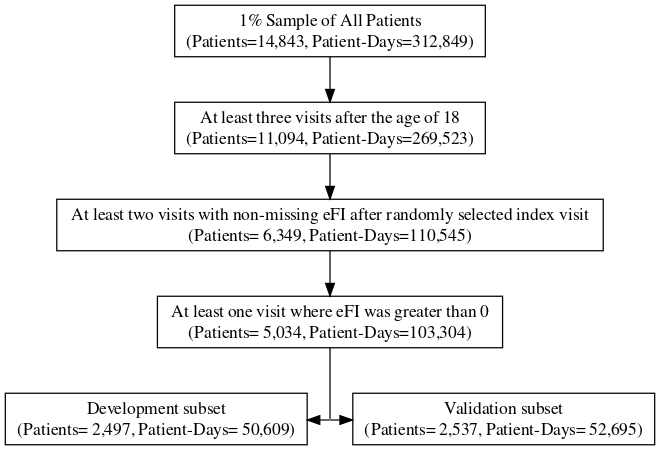
## Participants

…

The patient-visits in the data are a representative sample of the mix of all primary care, specialist, and inpatient encounters of UT Health San Antonio and the University Health System. As summarized in Figure 1, encounters during which the patient was below the age of 18 were excluded and patients who had fewer than three such encounters were excluded. To avoid the bias that would accompany always using the first valid visit as a starting point, an index date was randomly selected for each patient from among all their visit dates and only data recorded on or after that visit was used in analysis (not counting the eFI values which were pre-calculated over for every patient-visit in the EHR system, obtained as a separate file, and linked to the main dataset). After this step, patients who had fewer than two visits remaining were also excluded.

Some patients whose eFI is zero are non-frail, but many have zero eFIs for artifactual reasons– e.g. incomplete records and sparse encounter histories. Furthermore, at the tail ends of their histories most patients have date-only records during which no diagnoses, procedures, nor lab results were recorded. These may have been missed appointments or accesses to patient files without actual encounters. We therefore excluded patients who never had an eFI greater than zero and for those that remained we excluded all visits after the final procedure, diagnosis, lab, or eFI (whichever happened last) was documented.

Finally, the patients were split into development ( 2,497 patients, 50,609 visit-days) and validation ( 2,537 patients, 52,695 visit-days) subsets. The incidence of the outcomes is reported in the [Results](#results) section, below.

 Figure 1: CONSORT diagram.

This study was performed using only retrospective EHR data with no treatment, intervention or any other interaction with patients. The exposure groups were determined by patient eFI as will be discussed below.

## Outcomes

*[TRIPOD 6.a] Clearly define the outcome that is predicted by the prediction model, including how and when assessed.*

The four outcomes we predicted with eFI were infections, non-operative trauma, cardiac complications, and the occurrence of one or more of these or one or more PSIs in general as defined by Southern *et al.* (2017). Following principles of implementation science (Meissner *et al.*, 2020), we drew on the expertise of the clinical stakeholders on our team (CT and SE) to select outcomes that are meaningful, relevant, and useful in guiding clinical care planning.

All decisions about processing and statistical analysis were made using only the development data, blinded to the validation data. After the analysis scripts were finalized for publication, they were re-run on the validation data to generate the results reported here, where appropriate showing the development data for comparison. The development version of each table and figure is available in the supplemental materials.

## eFI, a deficit accumulation index

The standard procedure for creating a deficit accumulation frailty index (Mitnitski, Mogilner and Rockwood, 2001; Searle *et al.*, 2008; Song, Mitnitski and Rockwood, 2010) is to select at least 30-40 functional deficits (e.g. diagnoses or abnormal lab results) that span multiple physiological systems, dichotomize or rescale them such that 0 indicates absence and 1 indicates presence, and then for a given time window, calculate an unweighted proportion of deficits that are present in a patient. This proportion is the eFI. Our selection of functional deficits is based on that of Clegg *et al.* (2016) for UK health systems as modified for ICD10 codes by Pajewski *et al.* (2019). We built upon their work by mapping laboratory tests to LOINC codes and omitting data elements which might not be available in some EHR systems (e.g. patient questionnaires). The resulting crosswalk tables map systolic and diastolic blood pressure, smoking status, BMI, 32 LOINC codes, 626 ICD10 codes, and 503 ICD9 codes (for compatibility with older data sources) to 48 functional deficits. These tables are available in S5 and online (Bokov, 2020). For every visit date of every patient in our site’s i2b2 data warehouse, the above tables were used to aggregate all matching codes documented during the preceding two-year window to a count of distinct functional deficits and divided by the total number of defined deficits (48). Thus, if a patient had six distinct functional deficits during the two-year period preceding a given visit, their eFI for that visit would be 0.125. Though eFI is a continuous variable, for purposes of visualization and interpretation we use a cutpoint of 0.19, established by Stow *et al.* (2018). Patients above this cut-point are described as frail while those below as non-frail.. The above deployment of the eFI was done prior to and separately from the design of this study. The eFI scores were provided to the authors as a separate file which the authors linked to the rest of the study data using shared but irreversibly de-identified keys. For this reason, when we omitted all visits prior to a randomly select index visit for each patient (see [Participants](#participants-1) section, above) this did not interfere with eFI calculation which was the only piece of information retained about a patient’s history prior to the index visit.

## Statistical analysis methods

For each of the four outcomes, we used a Cox proportional hazard model (Cox, 1972) with eFI as the predictor of the first occurrences of the respective outcomes after the patients’ index visits. Unlike most previous studies, did not select a single time point to represent each patient. In real-world scenarios patient data updates at every encounter so we designed eFI as a time-varying numeric predictor updated at every visit (encompassing all relevant codes documented during a two-year rolling window preceding that visit). Follow-ups for each patient ended at the first occurrence of the outcome being modeled. Harrell’s c-statistic (Harrell *et al.*, 1984; Harrell, Lee and Mark, 1996) was used to assess model performance and the proportionality assumption was tested for each model as per Grambsch and Therneau (1994) (Table 2).

We used the R statistical language (R Core Team, 2021) and the Survival package for R (Terry M. Therneau and Patricia M. Grambsch, 2000) for all analysis and visualizations. We fit simple univariate Cox models that have the following specification, in the R syntax:

coxph( Surv(t0,t1,xx) ~ I(10\*efi), data=DATA)

Here t0 and t1 are the start and stop of an observation period (i.e. the time elapsed from one encounter and the next). These are integers, representing the number of days relative to each patient’s index visit. The xx is an indicator variable for the outcome with a value of TRUE if that outcome occurred during the observation period and FALSE if it did not. The predictor eFI is represented by efi, a continuous variable with a range of 0 to 1. The full expression, I(10\*efi) means we scaled efi by 10 so that the coefficient could be interpreted as the change in hazard ratio per 0.1 change in eFI. All model coefficients are shown below the corresponding Kaplan-Meier curves in Figures 2-5.

Our missing data strategy approximates a complete-case analysis in that we excluded patients with too few visits or missing eFI. No imputation was done on the remaining records; absence of ICD10 codes for the respective outcomes was interpreted as non-occurrence.

We limited ourselves to a 1% initial sample for two reasons. First, given anticipated utility of the eFI for predicting other clinically important outcomes in future studies, and given that our approach uses information from a large time-span for each sampled patient, we need to avoid depleting the un-analyzed patient population in these early stages. Secondly, we needed to stay within processing-time and file-size constraints of our research data warehouse team. Extrapolating from the effect sizes in Pajewski *et al.* (2019), events with an overall prevalence of 5% and a hazard ratio of 1.33 between frail and non-frail groups should be detectable with a power well over 0.80.

## Development vs. validation

We used the following validation methods recommended by Royston and Altman (2013): regressing out-of-sample linear predictors against the in-sample linear predictors (Table 3), re-fitting the validations model using the developmental predictors as the sole explanatory variables (Table 4), and comparing the cumulative distributions of the out-of-sample and in-sample linear predictors (supplement). There were no differences in the choice of predictors, outcome, inclusion criteria, or clinical setting between the development and validation data. Both were mutually exclusive, randomly assigned subgroups of patients from the same sample. No updates or model recalibration were needed after analyzing the validation data.

# Results

## Participants

…

There were 1,450 (58.1%) females and 1,046 (41.9%) males in the development data and 1,461 (58.5%) females and 1,034 (41.4%) in the validation data. In the development data 1,960 (78.7%) of the patients were white and in the validation data, 1,974 (79.2%) were. As expected, most patients had an eFI below 0.19 those with a higher eFI were moderately older (56 vs 48 in the development data and 56 vs 48 in the validation data). Patients in the high eFI groups had a mean (SD) eFI of 0.28 (0.072) and 0.28 (0.072) in the development and validation data respectively. Overall, the development and validation cohorts were very similar to each other in both demographics and outcome incidence.

Table 1: Cohort demographics

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Frail (eFI > 0.19) | | | Non-Frail (eFI <= 0.19) | | | Total | |
|  | Development (N=373) | | Validation (N=373) | Development (N=2,124) | | Validation (N=2,124) | Development (N=2,497) | | Validation (N=2,497) |
| **Sex**: N(%) |  | |  |  | |  |  | |  |
| Female | 195 (52.3%) | | 216 (57.9%) | 1,255 (59.1%) | | 1,245 (58.6%) | 1,450 (58.1%) | | 1,461 (58.5%) |
| Male | 178 (47.7%) | | 156 (41.8%) | 868 (40.9%) | | 878 (41.3%) | 1,046 (41.9%) | | 1,034 (41.4%) |
| Unknown | 0 (0%) | | 1 (0.268%) | 1 (0.0471%) | | 1 (0.0471%) | 1 (0.0400%) | | 2 (0.0801%) |
| **Race**: N(%) |  | |  |  | |  |  | |  |
| Asian | 6 (1.61%) | | 8 (2.14%) | 47 (2.22%) | | 43 (2.03%) | 53 (2.13%) | | 51 (2.05%) |
| Black | 18 (4.83%) | | 28 (7.51%) | 160 (7.55%) | | 146 (6.89%) | 178 (7.14%) | | 174 (6.98%) |
| Other | 14 (3.75%) | | 25 (6.70%) | 138 (6.51%) | | 125 (5.90%) | 152 (6.10%) | | 150 (6.02%) |
| Unknown | 11 (2.95%) | | 21 (5.63%) | 138 (6.51%) | | 122 (5.76%) | 149 (5.98%) | | 143 (5.74%) |
| White | 324 (86.9%) | | 291 (78.0%) | 1,636 (77.2%) | | 1,683 (79.4%) | 1,960 (78.7%) | | 1,974 (79.2%) |
|  |  | |  |  | |  |  | |  |
| **Hospital Acquired Infections**: N(%) | 62 (16.6%) | | 62 (16.6%) | 100 (4.71%) | | 100 (4.71%) | 162 (6.49%) | | 162 (6.49%) |
| **Hospital Acquired Trauma**: N(%) | 50 (13.4%) | | 50 (13.4%) | 95 (4.47%) | | 95 (4.47%) | 145 (5.81%) | | 145 (5.81%) |
| **Cardiac Complications**: N(%) | 58 (15.5%) | | 58 (15.5%) | 50 (2.35%) | | 50 (2.35%) | 108 (4.33%) | | 108 (4.33%) |
| **Any Patient Safety Event**: N(%) | 150 (40.2%) | | 150 (40.2%) | 254 (12.0%) | | 254 (12.0%) | 404 (16.2%) | | 404 (16.2%) |
|  |  | |  |  | |  |  | |  |
| **Frailty**: Mean(SD) | 0.28 (0.072) | | 0.28 (0.072) | 0.077 (0.05) | | 0.077 (0.05) | 0.11 (0.091) | | 0.11 (0.091) |
| **Patient age (years)**: Mean(SD) | 56 (17) | | 56 (17) | 48 (17) | | 48 (17) | 49 (18) | | 49 (18) |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| . |  |  |  |  |  |  |  |

## Prediction of Outcomes

Kaplan-Meier plots of the validation data are shown in Figures 2-5 , split on eFI (Frail if above 0.19, non-Frail otherwise). Each plot is followed by Cox model results with p-values adjusted for multiple comparisons (Holm, 1979) because we are testing four separate hypotheses on the same data. Below that, each figure has a risk table for a sample of time points (30 days, 60 days, 90 days, one year, and 3 years). The count of censored events is the number of patient-visits during which the outcome has not yet been observed at that time point. The estimated survival function is also shown. Kaplan-Meier plots from the development data are superimposed as dashed lines on each plot. For each of the outcomes, eFI was found to be a statistically and practically significant predictor. For each 0.1 increase in EFI, we found approximately a doubling of the hazard. Roughly approximating these hazards as probabilities, if a hypothetical patient with an eFI of 0 has a 0.5% baseline chance of experiencing a PSI during a 30-day period, then if their eFI were 0.1 their chance would rise to 1.2% and if their eFI were 0.2, just above our operational definition of frailty, it would be 2.8%

## Chart Description automatically generated

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Hazard Ratio | β (95% CI) | SE | Wald | P |
| 2.351 | 0.855 (0.772, 0.937) | 0.042 | 20.331 | 2.063× 10-91 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Frail | t (Days) | At Risk | Events | Censored | S(t) | SE | Lower CI | Upper CI |
| FALSE | 0 | 1,854 | 0 | 0 | 1 | 0 | 1 | 1 |
|  | 30 | 1,647 | 21 | 924 | 0.99 | 0 | 0.98 | 0.99 |
|  | 60 | 1,586 | 10 | 452 | 0.98 | 0 | 0.98 | 0.99 |
|  | 90 | 1,461 | 6 | 381 | 0.98 | 0 | 0.97 | 0.99 |
|  | 365 | 985 | 50 | 1,504 | 0.94 | 0.01 | 0.93 | 0.95 |
|  | 1,095 | 484 | 44 | 1,739 | 0.88 | 0.01 | 0.86 | 0.9 |
| TRUE | 0 | 220 | 0 | 0 | 1 | 0 | 1 | 1 |
|  | 30 | 212 | 10 | 380 | 0.95 | 0.01 | 0.93 | 0.98 |
|  | 60 | 199 | 11 | 202 | 0.9 | 0.02 | 0.86 | 0.94 |
|  | 90 | 185 | 11 | 174 | 0.85 | 0.02 | 0.81 | 0.9 |
|  | 365 | 158 | 31 | 803 | 0.71 | 0.03 | 0.65 | 0.77 |
|  | 1,095 | 25 | 36 | 922 | 0.4 | 0.05 | 0.32 | 0.52 |

Figure 2: Any PSI

Figure 2: Any PSI

## Chart, line chart Description automatically generated

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Hazard Ratio | β (95% CI) | SE | Wald | P |
| 2.04 | 0.713 (0.585, 0.841) | 0.065 | 10.914 | 2.949× 10-27 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Frail | t (Days) | At Risk | Events | Censored | S(t) | SE | Lower CI | Upper CI |
| FALSE | 0 | 1,938 | 0 | 0 | 1 | 0 | 1 | 1 |
|  | 30 | 1,738 | 5 | 842 | 1 | 0 | 0.99 | 1 |
|  | 60 | 1,665 | 3 | 438 | 1 | 0 | 0.99 | 1 |
|  | 90 | 1,534 | 1 | 349 | 0.99 | 0 | 0.99 | 1 |
|  | 365 | 1,037 | 22 | 1,411 | 0.98 | 0 | 0.97 | 0.99 |
|  | 1,095 | 522 | 18 | 1,750 | 0.95 | 0.01 | 0.94 | 0.97 |
| TRUE | 0 | 267 | 0 | 0 | 1 | 0 | 1 | 1 |
|  | 30 | 253 | 3 | 374 | 0.99 | 0.01 | 0.98 | 1 |
|  | 60 | 228 | 9 | 172 | 0.95 | 0.01 | 0.92 | 0.98 |
|  | 90 | 221 | 2 | 185 | 0.94 | 0.02 | 0.91 | 0.97 |
|  | 365 | 195 | 10 | 973 | 0.9 | 0.02 | 0.86 | 0.94 |
|  | 1,095 | 37 | 13 | 1,296 | 0.79 | 0.03 | 0.73 | 0.86 |

Figure 3: Hospital acquired infections.

Figure 3: Hospital acquired infections.

## Chart Description automatically generated

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Hazard Ratio | β (95% CI) | SE | Wald | P |
| 1.931 | 0.658 (0.513, 0.803) | 0.074 | 8.884 | 1.93× 10-18 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Frail | t (Days) | At Risk | Events | Censored | S(t) | SE | Lower CI | Upper CI |
| FALSE | 0 | 1,945 | 0 | 0 | 1 | 0 | 1 | 1 |
|  | 30 | 1,745 | 11 | 833 | 0.99 | 0 | 0.99 | 1 |
|  | 60 | 1,681 | 5 | 382 | 0.99 | 0 | 0.99 | 1 |
|  | 90 | 1,539 | 2 | 307 | 0.99 | 0 | 0.99 | 0.99 |
|  | 365 | 1,042 | 15 | 1,454 | 0.98 | 0 | 0.97 | 0.99 |
|  | 1,095 | 519 | 13 | 1,764 | 0.96 | 0.01 | 0.95 | 0.97 |
| TRUE | 0 | 269 | 0 | 0 | 1 | 0 | 1 | 1 |
|  | 30 | 251 | 5 | 340 | 0.98 | 0.01 | 0.97 | 1 |
|  | 60 | 238 | 2 | 194 | 0.97 | 0.01 | 0.95 | 0.99 |
|  | 90 | 238 | 3 | 176 | 0.96 | 0.01 | 0.94 | 0.99 |
|  | 365 | 204 | 8 | 872 | 0.93 | 0.02 | 0.9 | 0.96 |
|  | 1,095 | 40 | 7 | 1,240 | 0.88 | 0.03 | 0.83 | 0.93 |

Figure 4: Non-operative trauma.

Figure 4: Non-operative trauma.

## Chart Description automatically generated

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Hazard Ratio | β (95% CI) | SE | Wald | P |
| 2.726 | 1.003 (0.862, 1.144) | 0.072 | 13.912 | 1.614× 10-43 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Frail | t (Days) | At Risk | Events | Censored | S(t) | SE | Lower CI | Upper CI |
| FALSE | 0 | 1,951 | 0 | 0 | 1 | 0 | 1 | 1 |
|  | 30 | 1,755 | 3 | 811 | 1 | 0 | 1 | 1 |
|  | 60 | 1,697 | 1 | 404 | 1 | 0 | 1 | 1 |
|  | 90 | 1,551 | 1 | 329 | 1 | 0 | 0.99 | 1 |
|  | 365 | 1,037 | 12 | 1,488 | 0.99 | 0 | 0.98 | 0.99 |
|  | 1,095 | 530 | 12 | 1,862 | 0.97 | 0.01 | 0.96 | 0.98 |
| TRUE | 0 | 256 | 0 | 0 | 1 | 0 | 1 | 1 |
|  | 30 | 238 | 2 | 323 | 0.99 | 0.01 | 0.98 | 1 |
|  | 60 | 228 | 6 | 186 | 0.96 | 0.01 | 0.94 | 0.99 |
|  | 90 | 223 | 4 | 155 | 0.95 | 0.01 | 0.92 | 0.98 |
|  | 365 | 197 | 14 | 889 | 0.88 | 0.02 | 0.84 | 0.93 |
|  | 1,095 | 35 | 15 | 1,146 | 0.72 | 0.05 | 0.63 | 0.82 |

Figure 5: Cardiac complications.

Figure 5: Cardiac complications.

## Model performance

…

All models had a C-statistic above 0.7, with cardiac complications exceeding 0.8. The proportional hazard assumption was satisfied for all outcomes except for all patient safety indicators combined (Table 2). Though statistically significant, this time-dependent deviation from proportionality is modest and is in the direction of increasing divergence with time (Figure S1).

Table 2: Performance of EFI models. Concordance shows Harrell’s c-statistic for each model. The remaining two columns are the χ2 statistic and p-value the proportional hazard test. A low p-value indicates a violation of the proportionality assumption.

|  |  |  |  |
| --- | --- | --- | --- |
| Outcome | Concordance (95% CI) | PH: χ2df=1 | PH: P |
| Infections | 0.73 (0.67, 0.78) | 2.28 | 0.131 |
| Trauma | 0.7 (0.64, 0.76) | 2.87 | 0.09 |
| Cardiac | 0.81 (0.75, 0.86) | 2.63 | 0.105 |
| Any PSI | 0.74 (0.71, 0.78) | 9.69 | 0.002 |

In Table 3 we report the calibration of the models by using the development models to make out-of-sample predictions on the validation data and using those predictions as offsets in Cox models fitted to the validation data (Royston and Altman, 2013). The “Mismatch (CI)” column represents the amount of remaining variation explained by the updated (validation) model but not explained by the out-of-sample prediction of the development model. Here a non-significant result means that the model predicted out-of-sample results without there being any benefit to updating on the new data. The models for all the outcomes passed this test with the exception of the all-PSI model.

Table 3: Goodness of fit and calibration.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Mismatch (CI) | SE | t | P |
| Infections | 0.016 (-0.112, 0.144) | 0.065 | 0.245 | 0.8068 |
| Trauma | 0.052 (-0.093, 0.197) | 0.074 | 0.703 | 0.4821 |
| Cardiac | 0.088 (-0.054, 0.229) | 0.072 | 1.215 | 0.2245 |
| Any PSI | 0.122 (0.04, 0.204) | 0.042 | 2.9 | 0.0037 |

In Table 4, the agreement between development and validation models with each other is evaluated by regressing their linear predictors against each other (Royston and Altman, 2013). Exact agreement would result in an intercept of zero and a slope of one. A slope greater than one indicates that the validation model is actually more sensitive than implied by the development results, a slope less than indicates lowered sensitivity and a slope of zero would imply no correlation between the development and validation results. For all outcomes the slopes were significantly lower than 1 with very small standard errors, so relying on the development data alone would have over-esimated the sensitivity of eFI as a predictor. However, except for all-PSI, the slopes were above 0.9.

Table 4: Validation against out-of sample predictions from the development model.

|  |  |  |  |
| --- | --- | --- | --- |
| Outcome | Coefficient | SE | t |
| (Intercept) | 0.011 | 2.55× 10-16 | 4.40× 1013 |
| Infections | 0.978 | 2.77× 10-16 | 3.52× 1015 |
| (Intercept) | -0.015 | 8.67× 10-16 | -1.72× 1013 |
| Trauma | 0.921 | 1.04× 10-15 | 8.81× 1014 |
| (Intercept) | 0.017 | 3.10× 10-17 | 5.38× 1014 |
| Cardiac | 0.913 | 2.47× 10-17 | 3.68× 1016 |
| (Intercept) | -0.009 | 4.81× 10-17 | -1.85× 1014 |
| Any PSI | 0.857 | 4.63× 10-17 | 1.85× 1016 |

For all outcomes, distributions of their linear predictors were similar whether they were generated out-of-sample by the development models or in-sample by the validation models (Figure S2).

# Discussion

We implemented an electronic Frailty Index using the Rockwood deficit-accumulation framework and used it to predict PSIs in real-world data from an EHR system. Our results add to the evidence that Rockwood indexes are a rapid, low-effort method of risk assessment that scales to large patient populations.

## Limitations

*[TRIPOD 18] Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).*   
Our study sample is representative of patients at one health system, but not of the general population. Furthermore, since patients with zero eFIs were excluded our findings may not apply to patients who completely lack functional deficits. Our data also shares the fundamental limitation of the EHR system from which it was obtained: like all EHR systems, it only has information that providers and coders put into it. Events taking place outside the health system or at un-connected health systems are not visible to our analysis. On the other hand, providers who rely on EHR systems at point of care are also working under these limitations and the patients whose outcomes they most need to predict are those who do in fact have functional deficits. The data we used is representative of this scenario, and despite the limitations eFI provides accurate predictions of poor patient outcomes. Since it is based on real EHR data, eFI is more directly transferable to clinical use than implementations based on curated registries and is more immediate than claims data. eFI is most accurate for patients who have accumulated a reasonable in-system visit-history and further work is needed to find a more precise relationship between the length of a patient’s visit-history and the accuracy of eFI.

## Interpretation

*[TRIPOD 19.b] Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.*

This study only used outcome data for each patient after a randomly selected index date between that patient’s first and last available visit. This design mimics a scenario where an existing patient is seen for the first time after clinical deployment of eFI and the goal is to determine how well eFI by itself will predict the outcomes without any additional information about prior occurrences. eFI predicted infections, trauma, cardiac complications, and overall PSI with high statistical significance. In all cases the effect size was large enough to be clinically significant and the models had high concordances (Table 2). PSI was the one outcome for which the Cox model’s proportionality assumption was not met. This may be due to PSI encompassing a more heterogeneous set of diagnoses than the other outcomes and some of these diagnoses may be more prevalent with age thus introducing a time interaction which the univariate model does not include. This may also be why PSI was the only outcome for which there was a significant amount of variation in the validation data which was not explained by out-of-sample predictions (Table 3). For infections, trauma, and cardiac complications the out-of-sample prediction was not significantly worse than the in-sample prediction. There was generally a high correlation between the out-of-sample and in-sample predictions ( 0.91 - 0.98 , Table 4) except PSI which had a moderate correlation ( 0.86 ).

## Implications

*[TRIPOD 20] Discuss the potential clinical use of the model and implications for future research.*

Assessing frailty helps clinicians identify high risk patients and tailor interventions to prevent health decline and poor outcomes. There is no gold standard method to assess Frailty in clinical practice. Currently available frailty assessment tools used in geriatric practice have good validity (for example Fried *et al.* (2001)) but these are time intensive and often difficult to implement in a busy general practice. Because of the simplicity of the Rockwood Frailty Index, it is more likely to be adopted by clinicians in a busy practice.

The statistical analysis we used to *validate* eFI as a predictor of infections, trauma, and cardiac complications and PSI is not a prerequisite for *deploying* eFI in the clinic. Though Cox models can be useful initially to find optimal thresholds tuned to specific populations and outcomes instead of relying on the 0.19 cutoff we used here though it was effective at partitioning patients into risk groups for very different types of events. Nor is eFI limited to any specific EHR system. As long as it is possible to extract a table of dates, patient IDs, ICD-10 codes, and LOINC codes for labs with accompanying value-flags, such a result-set can be cross-walked to eFI values at patient-visit granularity.

# Conclusions

This contributes to a growing body of evidence that risk scores built using the Rockwood framework (Mitnitski, Mogilner and Rockwood, 2001) will be a valuable tool for clinical decision support not restricted to any one illness or specialty. The variables used to calculate EFI are ones that are in some form available in every EMR system. If there are EMR systems where a few of these variables are not available, Rockwood deficit accumulation indexes have been shown to continue giving consistent results despite variations in what individual codes are available (Rockwood *et al.*, 2006) as long as there is a sufficiently large and representative collection of deficits into which these variables can be binned (Mitnitski *et al.*, 2005; Searle *et al.*, 2008).

To facilitate adoption and refinement of EFI, in addition to the crosswalk tables (S5 and S6) we released mapping tables and example SQL scripts (Bokov, 2020) as a starting point for researchers or enterprise analytics teams that wish to deploy eFI at their own health systems. With input from colleagues at other learning health systems we intend to evolve these scripts into a self-contained app that interfaces with EMR systems (via FHIR) to provide real-time frailty assessment at the clinical point of care and assist clinicians in developing care plans to mitigate the risks of frailty.

# Acknowledgments

This work was supported by NIH/NCATS UL1TR001120 and UL1TR002645 (AFB, KRS), the Long School of Medicine KL2 Award (AFB), the Claude D. Pepper Older Americans Independence Center P30 AG044271 (SER), and the Berneice Castella Distinguished Professorship in Aging Research in Nursing (KRS). The funders of this study had no influence on the design, data collection, analysis, interpretation, or manuscript preparation and submission.

We would also like to thank Dr. Meredith Zozus, director of the Informatics core at the Institute for Integration of Medicine and Science and division chief for research informatics for helping us and countless other researchers leverage clinical data in their studies.

# References

Bokov, A. (2020) ‘Bokov/FreeFI: First Release’. Zenodo. doi: [10.5281/ZENODO.4053088](https://doi.org/10.5281/ZENODO.4053088).

Clegg, A. *et al.* (2016) ‘Development and validation of an electronic frailty index using routine primary care electronic health record data’, *Age and Ageing*, 45(3), pp. 353–360. doi: [10.1093/ageing/afw039](https://doi.org/10.1093/ageing/afw039).

Cox, D. (1972) ‘Regression Models and Life-Tables’, *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 34(2), pp. 187–220.

Fried, L. P. *et al.* (2001) ‘Frailty in older adults: Evidence for a phenotype’, *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 56(3), pp. M146–M157. doi: [10.1093/gerona/56.3.M146](https://doi.org/10.1093/gerona/56.3.M146).

Grambsch, P. M. and Therneau, T. M. (1994) ‘Proportional hazards tests and diagnostics based on weighted residuals’, *Biometrika*, 81(3), pp. 515–526. doi: [10.1093/biomet/81.3.515](https://doi.org/10.1093/biomet/81.3.515).

Harrell, F. E. *et al.* (1984) ‘Regression modelling strategies for improved prognostic prediction’, *Statistics in Medicine*, 3(2), pp. 143–152. doi: [10.1002/sim.4780030207](https://doi.org/10.1002/sim.4780030207).

Harrell, F. E., Lee, K. L. and Mark, D. B. (1996) ‘Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors’, *Statistics in Medicine*, 15(4), pp. 361–387.

Holm, S. (1979) ‘A simple sequentially rejective multiple test procedure.’, *Scandinavian Journal of Statistics*, 6, pp. 65–70.

Hoover, M. *et al.* (2013) ‘Validation of an index to estimate the prevalence of frailty among community-dwelling seniors’, *Health Reports*, 24(9), pp. 10–17.

Kulminski, A. M. *et al.* (2008) ‘Cumulative Deficits Better Characterize Susceptibility to Death in Elderly People than Phenotypic Frailty: Lessons from the Cardiovascular Health Study: FRAILTY, CUMULATIVE DEFICITS, AND SURVIVAL’, *Journal of the American Geriatrics Society*, 56(5), pp. 898–903. doi: [10.1111/j.1532-5415.2008.01656.x](https://doi.org/10.1111/j.1532-5415.2008.01656.x).

Li, G. *et al.* (2015) ‘Comparison between Frailty Index of Deficit Accumulation and Phenotypic Model to Predict Risk of Falls: Data from the Global Longitudinal Study of Osteoporosis in Women (GLOW) Hamilton Cohort’, *PLOS ONE*. Edited by A. M. Chamberlain, 10(3), p. e0120144. doi: [10.1371/journal.pone.0120144](https://doi.org/10.1371/journal.pone.0120144).

Malmstrom, T. K., Miller, D. K. and Morley, J. E. (2014) ‘A Comparison of Four Frailty Models’, *Journal of the American Geriatrics Society*, 62(4), pp. 721–726. doi: [10.1111/jgs.12735](https://doi.org/10.1111/jgs.12735).

Meissner, P. *et al.* (2020) ‘Engagement science: The core of dissemination, implementation, and translational research science’, *Journal of Clinical and Translational Science*, 4(3), pp. 216–218. doi: [10.1017/cts.2020.8](https://doi.org/10.1017/cts.2020.8).

Mitnitski, A. *et al.* (2005) ‘Relative Fitness and Frailty of Elderly Men and Women in Developed Countries and Their Relationship with Mortality: FITNESS AND FRAILTY IN RELATION TO MORTALITY’, *Journal of the American Geriatrics Society*, 53(12), pp. 2184–2189. doi: [10.1111/j.1532-5415.2005.00506.x](https://doi.org/10.1111/j.1532-5415.2005.00506.x).

Mitnitski, A. B., Mogilner, A. J. and Rockwood, K. (2001) ‘Accumulation of Deficits as a Proxy Measure of Aging’, *The Scientific World JOURNAL*, 1, pp. 323–336. doi: [10.1100/tsw.2001.58](https://doi.org/10.1100/tsw.2001.58).

Pajewski, N. M. *et al.* (2019) ‘Frailty Screening Using the Electronic Health Record Within a Medicare Accountable Care Organization’, *The Journals of Gerontology: Series A*. Edited by A. Newman, 74(11), pp. 1771–1777. doi: [10.1093/gerona/glz017](https://doi.org/10.1093/gerona/glz017).

R Core Team (2021) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: <https://www.R-project.org/>.

Rockwood, K. (2005) ‘A global clinical measure of fitness and frailty in elderly people’, *Canadian Medical Association Journal*, 173(5), pp. 489–495. doi: [10.1503/cmaj.050051](https://doi.org/10.1503/cmaj.050051).

Rockwood, K. *et al.* (2006) ‘Long-Term Risks of Death and Institutionalization of Elderly People in Relation to Deficit Accumulation at Age 70: LONG-TERM RISK OF DEATH DEFINED BY AGE 70’, *Journal of the American Geriatrics Society*, 54(6), pp. 975–979. doi: [10.1111/j.1532-5415.2006.00738.x](https://doi.org/10.1111/j.1532-5415.2006.00738.x).

Rockwood, K. and Mitnitski, A. (2007) ‘Frailty in Relation to the Accumulation of Deficits’, *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 62(7), pp. 722–727. doi: [10.1093/gerona/62.7.722](https://doi.org/10.1093/gerona/62.7.722).

Royston, P. and Altman, D. G. (2013) ‘External validation of a Cox prognostic model: Principles and methods’, *BMC Medical Research Methodology*, 13(1), p. 33. doi: [10.1186/1471-2288-13-33](https://doi.org/10.1186/1471-2288-13-33).

Searle, S. D. *et al.* (2008) ‘A standard procedure for creating a frailty index’, *BMC Geriatrics*, 8(1). doi: [10.1186/1471-2318-8-24](https://doi.org/10.1186/1471-2318-8-24).

Song, X., Mitnitski, A. and Rockwood, K. (2010) ‘Prevalence and 10-year outcomes of frailty in older adults in relation to deficit accumulation’, *Journal of the American Geriatrics Society*, 58(4), pp. 681–687. doi: [10.1111/j.1532-5415.2010.02764.x](https://doi.org/10.1111/j.1532-5415.2010.02764.x).

Southern, D. A. *et al.* (2017) ‘Deriving ICD-10 Codes for Patient Safety Indicators for Large-scale Surveillance Using Administrative Hospital Data:’, *Medical Care*, 55(3), pp. 252–260. doi: [10.1097/MLR.0000000000000649](https://doi.org/10.1097/MLR.0000000000000649).

Stow, D. *et al.* (2018) ‘Evaluating frailty scores to predict mortality in older adults using data from population based electronic health records: Case control study’, *Age and Ageing*, 47(4), pp. 564–569. doi: [10.1093/ageing/afy022](https://doi.org/10.1093/ageing/afy022).

Terry M. Therneau and Patricia M. Grambsch (2000) *Modeling survival data: Extending the Cox model*. New York: Springer.