Risk for Poor Post-Operative Quality of Life Among Wearable Use Subgroups in an All of Us Research Cohort

Nidhi Soley[1], Shanshan Song[1,3,4], Natalie Flaks-Manov[1], Casey Overby Taylor[1,2,3,4]

[1]Institute for Computational Medicine, Whiting School of Engineering, Johns Hopkins University, Baltimore, Maryland, USA, [2]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, USA, [3]Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA, and [4]Biomedical Informatics & Data Science Section, The Johns Hopkins University School of Medicine, Baltimore, Maryland

The objective of this research was to build and assess the performance of a prediction model for post-operative recovery status measured by quality of life among individuals experiencing a variety of surgery types. In addition, we assessed the performance of the model for two subgroups (high and moderately consistent wearable device users). Study variables were derived from the electronic health records, questionnaires, and wearable devices of a cohort of individuals with one of 8 surgery types and that were part of the NIH *All of Us* research program. Through multivariable analysis, high frailty index (OR 1.69, 95% 1.05-7.22, $p<0.006$), and older age (OR 1.76, 95% 1.55-4.08, $p<0.024$) were found to be the driving risk factors of poor recovery post-surgery. Our logistic regression model included 15 variables, 5 of which included wearable device data. In wearable use subgroups, the model had better accuracy for high wearable users (81%). Findings demonstrate the potential for models that use wearable measures to assess frailty to inform clinicians of patients at risk for poor surgical outcomes. Our model performed with high accuracy across multiple surgery types and were robust to variable consistency in wearable use.

Keywords: digital health technologies, wearables, predictive modeling, risk factors

Corresponding Author: Nidhi Soley, MS, Institute for Computational Medicine (ICM) Johns Hopkins Whiting School of Engineering 3101 Wyman Park Drive, Hackerman 318, Baltimore, MD 21218, USA; nsoley1@jh.edu

Introduction

Surgical procedures are becoming more common over the world, with one out of every 10 individuals getting one each year in high-income nations. After discharge, patients have the main responsibility for their recovery, and variance in adherence to this can result in varying outcomes [1].More than 10% of patients over the age of 45 encounter a significant postoperative complication, which is apparent in a variety of surgical groups [1]. Thus, there is a need to better identify patients that are at risk for such poor surgical outcomes with applicability to multiple surgical types.

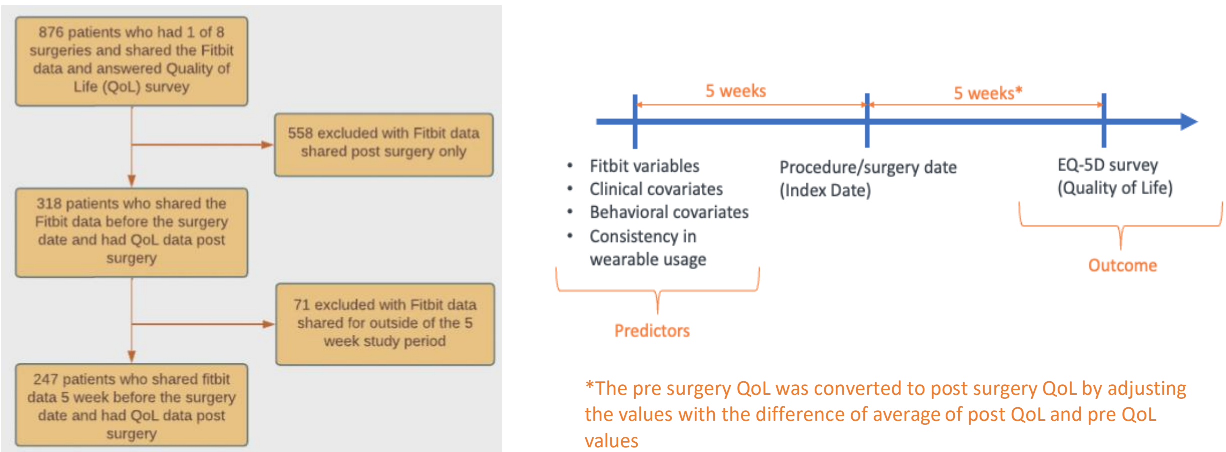
Methods for accurately predicting the probability of post-surgical complications have been studied widely in the past. For predicting surgical morbidity, Copeland proposed the POSSUM (Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity) model in 1991. [2]. Since then, various post-operative morbidity prediction models have been suggested, including the E-POSSUM, Estimation of Physiologic Ability and Surgical Stress (E-PASS) [3], and Barwon Health (BH) 2009 models [4]. However, the predictive capacity of these models beyond the population used to create the model may be limited. Given there are no published models to predict poor post-surgical recovery for different types of surgeries, this work aimed to build a prediction model that uses data types that are accessible across a broad range of surgical patients.

One data type of particular interest was physical activity data from wearables. Recent studies have shown that utilizing the data from wearables to construct predictive models can help identify surgical complications earlier, improve recovery, and provide safe follow-up. Furthermore, wearables can help patients engage, assist, and care for themselves by bridging the gap between clinical services and their homes [5]. Despite the emergence of numerous digital initiatives in surgery, there has been little or no discussion of wearable use factors on the performance of the prediction models.

To build a model that predicts post-operative outcomes based on the preoperative wearable data, we used candidate risk factors taken from electronic health records (EHR) and a commercial wearable device (Fitbit). In addition, we assessed the impact of wearable usage on model performance. To do this, we assessed the accuracy of the model in cohort stratified by wearable use (high vs moderate/low pre-operative wearable use). We hypothesized that model performance is better for high users when compared to patients with moderate/low wearable usage.

Method

This is a retrospective cohort study based on data collected by the *All of Us* Research Program Dataset v5 (Registered & Controlled Tier) from May 6, 2018, to April 1, 2021 [6]. The cohort includes patients who had gone through one of eight surgeries: general, gynecology, orthopedics, plastic, neuro, vascular, urology, thoracic surgery, shared Fitbit data and completed the survey within 5 weeks since the surgery. Figure 1 (a) shows the flowchart for inclusion and exclusion criteria. 247 participants fulfilled the study criteria. The time range of data (Figure 1 (b)) was defined for a period of 5 weeks, all the variables were averaged for this period before the surgery date. For the study, we required EQ-5D score for Quality of Life (QoL), a self-reported outcome measure for recovery taken within 5 weeks after surgery. For the patients who did not meet this criterion, we adjusted their QoL values by adding the difference of the average QoL post and pre-surgery (0.02) to the pre indices and obtained the post QoL indices for all 247 patients.



a b

Figure 1. a) Inclusion and exclusion criteria flow chart. b) Timeline of the study.

Data Source and Preprocessing

Primary Outcome

The EuroQOL instrument (EQ-5D-5L) was utilized to evaluate QoL. EQ-5D index has been used in several studies to assess the effect of surgery and the difference in the QoL pre- and post-surgery [7][8][9][10]. This is a standardized, proven QoL measurement tool. Mobility, Self-Care, Usual Activities, Pain/Discomfort, and Anxiety/Depression are the five dimensions included in the EQ-5D survey. We included two questions from each category. The responses to the questions were divided into 5 levels, 1 denotes an excellent state of health, and 5 is worse. The 5L profile, the 5-digit number, is generated based on the average of two questions in the five categories, for instance, if you have an excellent state of health your profile would be “11111”. To estimate a single index value depending on the response to this categorization, a broad population-based algorithm was used for US population [11]. The index value is normally distributed and reflects how good or bad the health state is according to the preferences of the general population of a country. The index value for our dataset lies in the range of 0 (worse) to 1(good) [12]. Since we had patients who underwent different kinds of surgeries, we converted the continuous QoL to a status of good and poor recovery using the average QoL of the population as a threshold [9][10][13].

Variables

Fifteen clinicopathological and demographic variables that might affect the postoperative outcome were included. The demographic covariates were age, gender, race, and ethnicity. The clinical covariates included average hemoglobin level in blood (g/dL), average albumin level in blood (g/dL), and average BMI ratio. The values of all variables were observed in the time frame of 5 weeks before the surgery. The behavioral covariates included smoking habits and alcohol consumption habits prior to the surgery. The Fitbit activity data was available in a longitudinal form for each patient. The data from the Fitbit device was in a summarized format for a day and had variables like average calories burned, mean light active minutes, mean of very active minutes, mean of sedentary minutes, and mean of steps count in a day. The characteristics description of the entire cohort is summarized in Table 1.

Frailty is a well-validated predictor of poor postoperative outcomes [14]. We created a frailty index using a standard procedure described by Samuel at el. to assess the impact of frailty on the recovery status post-surgery [15]. The frailty index is frequently stated as a percentage of actual deficits to all deficits considered [15]. For instance, if a person had 10 of the 30 deficiencies that were considered, their frailty index would be 10/30, or 0.33. To create this index, we included 19 variables measured within 5 weeks before the surgery. Function, cognition, co-morbidity, health attitudes and behaviors, and physical performance metrics were all included in the database. The variables included activity data from Fitbit, clinical data, and various comorbid conditions chosen from Charles Comorbid Index’s ICD9 and ICD10 codes for dementia, heart attack, malignancy, and diabetes. Health attitude variables included survey questions that assessed the person's general health like disability in walking/climbing, disability in dressing/bathing, and difficulty in reading/writing. For binary variables "0" denoted the absence of the deficit and "1" the presence of the deficit. To grade survey questions, we used Excellent as 0, Very Good as 0.25, Good as 0.5, Fair as 0.75, and Poor as 1. Similarly, for continuous variables, such as Fitbit activity data [19][20], hemoglobin level [22], known cut-points were applied. An individual’s deficit scores were aggregated to create an index, with 0 denoting no deficit and 1 denoting the presence of all 19 deficits. To assess and validate the variable the slope of a best-fit log of the frailty index in proportion to age was plotted and the association between age and frailty was analyzed.

Quantifying Wearable Usage
To quantify wearable use, we calculated the consistency of using the Fitbit device. During the period of 5 weeks prior to the surgery, usage of the Fitbit device varied among the patients and was calculated using equation 1. Consistency and duration of Fitbit usage were used to divide the entire cohort into two subgroups (low/moderate wearable users and high wearable users).

$$\text{Consistency} = \frac{\text{Number of days the patient data was logged}}{\text{Number of days between first date and last date of use (duration)}} \tag{1}$$

Patients with 100% consistency and duration of usage of 5 weeks were classified as high wearable users. The patients with a consistency of less than 1 and a duration of Fitbit usage of fewer than 5 weeks were considered moderate/low users of a wearable device.

Statistical Analysis
Univariate Analysis
To determine the effect of individual risk factors on the binary outcome (good or poor recovery), we applied univariate analysis by chi-square test for categorical variables. For the risk factors like race, ethnicity, and alcohol consumption the small proportion categories were combined to make it a binary variable. Age was divided into three categories 18-49 years, 50-64 years, and 65 years and above. The frailty index was also divided into two categories based on the mean value of the population as non-frail (0-0.54) and frail (0.55-1). A P value of less than 0.05 was considered significant. The statistical approach was applied separately to each risk factor to obtain the odds, odds ratio (OR), and significance of predicting the poor outcome post-surgery. We also implemented these analyses for wearable device use subgroups (see “Quantifying Wearable Usage”).

Multivariable Analysis
To obtain the driving risk factors of poor outcome post recovery, we implemented a multivariable logistic regression model on the entire cohort, on high wearable users, on moderate/low wearable user’s dataset individually. All 15 variables were initially used for the analysis in this model. For collinearity diagnostics, variables with Variance Inflation Factor (VIF) above 5 were regarded as multicollinear. To exclude variables with multi-collinearity, multiple stepwise regression was used to iteratively build regression models that automatically chose independent variables. After removing three collinear variables, the stats model library’s logistic regression model was applied to the remaining twelve independent variables and the binary outcome, recovery status. The statsmodel gives the OR, 95% confidence interval (CI), and p values for each risk factor.

Predictive Modeling
To build a predictive model of post-operative recovery status, we used a supervised machine learning algorithm. The logistic regression model was implemented individually for moderate/low wearable users, high wearable users, and the total population (baseline) datasets with 12 features that were identified non-collinear in multivariable analysis. To improve the model performance, we hyper-tuned the model using the grid search cross-validation technique. Since the outcome, poor recovery, and good recovery classes were imbalanced, we used the stratified K fold cross validation

technique in the grid search cross-validation splitting strategy. After preprocessing, we divided the data into train and test sets, fitted the model on the train set, and then assessed the performance of the model for three separate test datasets (baseline, moderate/low wearable users, high wearable users).

Assessment of Model Performance
To compare the performance of the three models, we used AUC (area under the curve) score, accuracy, sensitivity, and ROC plot. The model with the highest AUC score was considered a better-performing model. The AUC score of the two subgroups was also tested for significance using their confidence intervals (CI). AUC CI calculated using bootstrap sampling method was used to compare the AUCs of models. The comparison of AUC was done using DeLong method [16]. If there was a difference in the two CIs, we concluded that the AUCs were different, and result was significant [17][18].

RESULTS

Study Population
Among a cohort of 247 people, most were female (77%, n=190), White (84%, n=208), and non-Hispanic or Latino (92%, n=228). Ages ranged from 26 to 86 years with an average of 60 years. Before the surgery, 45 % of the cohort had consumed alcohol and the smoking history was largely unknown (95%, n=235). The Fitbit data obtained 5 weeks before the surgery suggested that this was a physically active cohort as per the physical activity standards defined by WHO and CDC [19][20]. The daily average for “light active minutes” in the cohort was 180 minutes which is considered a “healthy lifestyle” according to the WHO [16]. However, the cohort also had average sedentary minutes that was higher than suggested for a healthy lifestyle (948 minutes compared to the suggested 540 minutes) [19][20]. The clinical covariates for the cohort lie in the normal range [21][22]. The average hemoglobin level in blood was 13.03 g/dL and the albumin level in blood was 4.12 g/dL. However, the cohort had an average BMI ratio slightly higher than the normal range [23] with the maximum BMI ratio being 78.3, indicating the presence of highly obese individuals. The smoking habit variable was not included in the study because of its disproportionate division of unknown versus the other categories. The validity of the frailty index was accessed by calculating the slope of the best fit log of the frailty index in proportion to age, the rate of accumulation of deficits was found to be 0.06, prior estimate is 0.03 per year [15]. The pre-surgery QoL adjustment was done for 115 (47%) patients. Characteristics of the cohort are summarized in Table 1.

When the entire cohort was divided into moderate/low (n=109) and high users (n=138) the distribution of the population changed and is summarized in Table 1. The proportion of individuals represented in different demographic and social factor groups were similar among subgroups. The clinical covariates for the two cohorts were also similar and lie within the normal range of albumin and hemoglobin level in the blood for a healthy adult [21][22]. The average frailty index appears to be higher for moderate/low wearable users (0.570) with respect to the entire cohort (0.549). The average frailty index for high users (0.541) was slightly lower than the average of the entire cohort. The Fitbit activity data for the two populations suggests that people who used the device consistently were more active as compared to those who used the device moderately. The patients using the device regularly on average had 35 minutes more light active minutes than the population using the device irregularly, and on an average burned 150 calories more than the moderate wearable users.

Table 1: Characteristics of study participants

		Total		Moderate/Low users		High users	
		N	%	N	%	N	%
Number of patients		247		109		138	
Categorical Variables							
Gender	Female	190	77%	83	76%	108	78%
	Male	57	23%	26	24%	30	22%
Race	White	208	84%	90	83%	117	85%
	Black or African American	13	5%	6	6%	7	5%
	Asian	5	2%	1	1%	5	4%
	None of these	21	9%	8	7%	9	7%
Ethnicity	Not Hispanic or Latino	228	92%	99	91%	129	93%
	Hispanic or Latino	14	6%	8	7%	6	4%
	None Of These	5	2%	2	2%	3	2%
Smoking Habit	Unknown	235	95%	105	96%	130	94%
	Past or Current Smoker	5	2%	3	3%	4	3%
	Never Smoked	7	3%	1	1%	4	3%
Alcohol consumer	Yes	112	45%	49	45%	61	44%
	No	3	1%	2	2%	3	2%
	Unknown	117	47%	58	53%	74	54%
Recovery (measured by QoL)	Good	153	62%	67	62%	94	68%
	Poor	94	38%	42	38%	43	32%
Continuous Variables	**Mean [SD]**	**Min**	**Max**	**Mean [SD]**		**Mean [SD]**	
Age *(years)*	60 [13.45]	26	86	57 [13.23]		62 [13.3]	
Frailty index*	0.5493 [0.082]	0.29	0.86	0.571 [0.07]		0.541 [0.08]	
Mean calories burnt in a day	802.19 [403.05]	573.87	2608.29	715.35 [389.80]		869.8[406.04]	
Mean light active minutes in a day	180.31 [73.59]	143.76	379.71	159.24 [77.62]		195.66 [67.41]	
Mean sedentary minutes in a day	947.97 [241.28]	329.29	1440	1044.08[249.9]		874.14[210.24]	
Mean very active minutes in a day	14.49 [17.96]	2.08	125.86	11.44 [13.92]		16.85 [20.29]	
Mean steps count in a day	6440 [3360.60]	4230	17543	5624 [3298]		7066 [3288]	
Albumin level	4.12 [0.361]	9.91	78.3	4.13 [0.45]		4.11 [0.26]	
Hemoglobin level	13.03 [1.280]	8.51	15.9	13.02[1.32]		13.04 [1.23]	
BMI ratio	32.4 [8.546]	9.91	78.3	33.8 [8.69]		31.3 [8.34]	

*Created using 19 variables including 5 wearable device variables.

Univariate Analysis

The primary risk factors of poor recovery from the univariate analysis for the entire population of 247 were gender, age, and frailty index. Findings from the univariate analysis of the entire cohort are summarized in Table 2. Females are at twice as high risk for having poor recovery post-surgery as compared to males (OR=2.22, p<0.025). People 65 years and over are at a threefold greater risk of having poor recovery after surgery (OR=3.11, p<0.001) as compared to people 18-49 years old. The frail population above an average frailty index (0.54) had a higher risk of having poor recovery as compared to the non-frail population (OR=2.72, p<0.001). Whites (OR= 1.68) and non-Hispanic or Latino (OR= 1.06) were not statistically significant.

On performing the univariate analysis (Table 2) for people who used the Fitbit device regularly, the significant risk factors were age and frailty index. High wearable users of Fitbit devices who are in the age range of 50-64 were associated with an increased risk for poor recovery post-surgery (OR 1.98, p<0.048) compared to young population (18-49 years). However, most of the elderly people (65 years and over) are in the category of good recovery post-surgery and have a lower risk of having poor recovery (OR 0.74, p<0.048).

Table 2: Univariate analysis of association between recovery status and risk factors.

		Total*							High usage of wearable*							Moderate/Low usage of wearable*						
		Good Recovery		Poor Recovery		Odds	OR	P value	Good Recovery		Poor Recovery		Odds	OR	P value	Good Recovery		Poor Recovery		Odds	OR	P value
		N	Rates per 100 patients	N	Rates per 100 patients				N	Rates per 100 patients	N	Rates per 100 patients				N	Rates per 100 patients	N	Rates per 100 patients			
Characteristics		153		94					95		43					70		39				
Gender	Female	110	57.9	80	42.1	0.73	2.22	0.025	72	67	36	33	0.50	1.62	0.410	49	59	34	41	0.70	2.92	0.074
	Male	43	75.4	14	24.6	0.33	1.00		23	77	7	23	0.31	1.00		21	81	5	19	0.24	1.00	
Race#	White	125	60.1	83	39.9	0.67	1.68	0.229	77	66	40	34	0.52	3.06	0.114	59	66	31	34	0.53	0.73	0.711
	Non-White	28	71.8	11	28.2	0.40	1.00		18	86	3	14	0.17	1.00		11	58	8	42	0.73	1.00	
Ethnicity#	Not Hispanic or Latino	141	61.8	87	38.2	0.62	1.06	1.000	88	68	41	32	0.47	1.63	0.821	65	66	34	34	0.53	0.53	0.523
	Other	12	63.2	7	36.8	0.59	1.00		7	78	2	22	0.29	1.00		5	50	5	50	1.00	1.00	
Alcohol consumer#	Yes	73	65.2	39	34.8	0.54	1.08	0.092	43	69	19	31	0.45	0.96	0.990	32	65	17	35	0.54	0.94	0.989
	Other	80	66.7	40	33.3	0.50	1.00		52	68	24	32	0.47	1.00		38	63	22	37	0.58	1.00	
Age	18-49	31	57.0	23	43.0	0.75	1.00		16	73	6	27	0.38	1.00		20	63	12	38	0.60	1.00	
	50-64	42	53.0	38	48.0	0.91	1.22	0.001	24	57	18	43	0.75	1.98	0.048	24	62	15	38	0.63	1.05	0.010
	65 years and over	34	30.0	79	70.0	2.33	3.11		58	78	16	22	0.28	0.74		12	32	26	68	2.17	3.62	
Frailty	0-0.54	90	71.0	36	29.0	0.40	1.00	0.001	58	78	16	22	0.28	1.00	0.015	30	75	10	25	0.34	1.00	0.114
	0.55-1	58	48.0	63	52.0	1.09	2.72		37	58	27	42	0.73	2.61		40	58	29	42	0.73	2.15	

*The division for good or poor recovery was done based on the average QoL of the population, for the entire cohort mean QoL is 0.663, for high wearable users the mean QoL is 0.67, for moderate/low wearable users the mean QoL is 0.65

#Combined all small proportion categories as other.

In moderate/low users of Fitbit, age was the only significant risk factor (Table 2). People who are 65 years and over are threefold higher risk of poor recovery post-surgery (OR 3.62, p<0.010) compared to young people (18-49 years).

Multivariable Analysis

Findings from the multivariable analysis of the entire cohort are summarized in Table 3. Among 247 patients, three covariates were observed to be significant risk factors of poor recovery status. Among sociodemographic variables, age and race were significant risk factors. The elderly population is more likely to have a poor recovery as compared to the population below that age group (OR 1.76, 95% 1.55-4.08, p<0.024). The frailty index was also a statistically significant risk factor. Population with a higher frailty index was at increased risk of poor recovery as compared to individuals whose frailty index was lower than 0.54 (OR 1.69, 95% 1.05-7.22, p<0.006).

The multivariable analysis on high wearable users shows that people with frailty index over 0.54 (frail) have higher risk of having poor recovery (OR 1.73, 95%CI 1.08-9.62, p<0.007). In moderate wearable users' frailty index is not a statistically significant risk factor (Table 3).

Table 3: Multivariable analysis for quality of life (QoL).

Risk Factors	**All users (N=247)**		**Moderate/Low wearable users (N=109)**		**High wearable users (N=138)**	
	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value
Gender (Female, ref. Male)	2.67 (0.98-6.08)	0.055	3.05 (1.66-7.89)	0.012	1.31(0.26-5.43)	0.680
Race (White, ref. Non-White)	1.65 (1.25-4.05)	0.024	0.89 (0.27-6.99)	0.702	2.32 (0.26-8.09)	0.191
Ethnicity (Non-Hispanics, ref. others)	1.06 (0.27-8.88)	0.477	0.16 (0.36-2.58)	0.103	1.28(0.50-17.77)	0.660
Alcohol Consumer (Yes, ref. others)	0.68 (0.55-4.97)	0.838	0.08 (0.01-1.10)	0.056	0.98(0.15-5.39)	0.984
Age (over 65, ref. less than 65)	1.76 (1.55-4.08)	0.024	1.98 (0.29-2.24)	0.783	1.09(0.12-5.67)	0.089
Frailty Index (over 0.54, ref. less than 0.54)	1.69 (1.05-7.22)	0.006	2.08 (0.01-7.75)	0.071	1.73(1.08-9.62)	0.007
Mean light active minutes in a day	1.00 (0.99-1.05)	0.923	1.00 (1.00-1.01)	0.410	1.00(0.99-1.87)	0.560
Mean sedentary minutes in a day	1.00 (0.99-1.00)	0.691	0.99 (0.98-1.01)	0.078	1.00(1.00-1.02)	0.056
Mean very active minutes in a day	1.02 (0.99-1.03)	0.166	1.00(0.99-1.04)	0.720	1.08(0.89-1.09)	0.895

Albumin level	1.96 (0.95-5.56)	0.310	2.12 (1.96-6.98)	0.003	1.44(0.44-1.47)	0.542
BMI ratio	0.94 (0.92-.1.00)	0.051	0.98 (0.90-1.05)	0.520	0.95(0.89-1.20)	0.172
Hemoglobin level	0.89 (0.67-1.35)	0.507	0.82 (0.54-1.44)	0.224	0.96(0.61-1.69)	0.870
Model evaluation metrics						
Accuracy	0.79		0.73		0.81	
Misclassification	0.21		0.27		0.19	
Sensitivity	0.92		0.93		0.95	
Specificity	0.77		0.5		0.62	
AUC score (95%CI)	0.759 (0.652-0.772)		0.721 (0.610-0.733)		0.792 (0.741-0.879)	

Logistic Regression Model performance and wearable usage

The comparison of the Logistic Regression model performance on three datasets is summarized in Table 3. The model performance for all the participants in the baseline dataset (247) was intermediate between the subgroup datasets with high wearable and moderate wearable users. When we focused on the participants with consistent wearable usage, the accuracy of the model increased by 2% from the baseline dataset. The misclassification rate also reduced. The AUC score was highest for high wearable users (0.792, 95%CI 0.741-0.879) as compared to the other two datasets. The model performance decreased when we focused on a population that was moderate in using the device prior to surgery, the accuracy of the model dropped to 0.73 from the baseline 0.79, and the AUC score (0.721, 95%CI 0.610-0.733) was also reduced by 3 units. The ROC (Receiver Operating Characteristics) curve for the comparison between the models for the two-subgroup population is shown in Figure 2. The CI for the AUC score for high wearable users is different from the CI of moderate/low wearable users AUC score which suggests the difference between the scores obtained from the two datasets is significant.

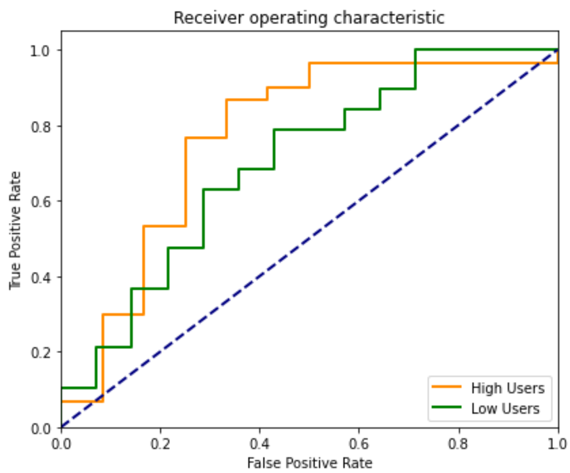


Figure 2. ROC curve for high vs moderate users of wearable device.

Discussion

In this retrospective study of AoU study participants who underwent 1 of 8 types of surgeries, we created a logistic regression model to predict poor QoL after surgery. We identified 15 risk factors to predict the recovery status post-surgery in terms of QoL. Out of which 5 risk factors were obtained from a wearable device. We examined the association between individual risk factors and QoL post-surgery using the chi-square test and multivariable logistic regression. In addition to analyzing the full cohort, we also conducted a separate analysis for the patients who were consistent in using the wearable device and patients who were moderately consistent in using the device. The model built with high wearable usage dataset had the best performance, outperforming the model implemented on the baseline dataset (see Table 3).

The findings from univariate and multivariable analyses of the entire cohort suggests that high frailty index, older age, and female gender are the driving risk factors of poor recovery post-surgery. The frailty index was the most significant risk factor which is a composition of data obtained from wearable device, survey questions and clinicopathological measures. Numerous studies have suggested that measurements from wearable sensors are related to clinical outcomes, such as complications, length of hospital stay, and readmission [24]. Adding to this evidence, we found that frailty (a measure created using the activity data obtained from the wearable device) was the most significant risk factor of poor recovery post-surgery across different datasets. Previous research also showed that patients with frailty had worse postoperative results across surgical specialties, including a greater incidence of morbidity, death, and ICU admission [25][26][27][28]. In our study, we also found a significant difference between the non-frail and frail patients in their risk associated with poor post-operative recovery. However, for the subgroup that used the device inconsistently and for a lower duration, there was no significant difference between frail and non-frail patients (Table 2 and 3). We believe that this could be because frailty is associated with older age and the population distribution in the moderate wearable users was uniform hence there is no difference in the frail and non-frail groups (Table 2). However, we did not find the 5 physical activity variables measured from wearable device to be a significant risk factor when considered independently in the univariate or multivariate analysis.

The logistic regression model with 12 features used to classify patients into poor or good recovery status gives the highest accuracy on high wearable use subgroup (provided Fitbit data continuously for 5 weeks prior to surgery) (Table 3). The good performance could be associated with the completeness, correctness, and homogeneity of the activity data obtained from the Fitbit device. Since we had observations for each day the average values of the variables for 5 weeks were non-null. The wearable usage measure defines the adherence to the device and our findings suggest that if the patient used the device more frequently to monitor themselves before the surgery, then it is more likely to accurately predict their recovery status post-surgery and readiness for the surgery.

Our findings from the logistic regression model comply with the findings of others that suggest that people at higher risk of poor recovery post-surgery could benefit the most from continuous preoperative monitoring using a wearable device [29]. In our study, the performance of the model is best on the high user dataset that includes more than half elderly population (over 65 years) and have a lower risk of poor postoperative recovery (Table 2) which could be associated with good pre-operative monitoring done through the wearable device.

The prospect of using wearable device technology for postoperative monitoring in both the hospital and the home will increase patient safety and promote continuity of care. Wearable technologies may ease early discharge and thereby minimize the length of hospital stay by continuously monitoring several health parameters [29]. Postoperative monitoring using wearable devices can also be extended before surgery to give baselines for comparison and as part of a prehabilitation approach, improving perioperative care holistically. From our findings, there is an opportunity for better guidance on wearable use to improve perioperative care. Additionally, there could be potential to integrate wearable activity data with other EHR measures. Frailty index was a good example and was one of most important risk factors for poor post operative recovery status that we identified. Another way to improve perioperative care could be to promote proper use of wearable device to monitor the patient including their vitals, and then using that data to predict the recovery status. If the patient is at high risk of poor recovery, then the surgery might be postponed, or the physician could take preventive measures to ensure better outcomes.

The major shortcoming of our work is the small sample size and QoL as the single post-operative outcome to study across multiple surgery types. Even so, the accuracy and other metrics of our model performance were good. Future work seeks to validate findings in larger datasets derived from a variety of hospital settings.

Acknowledgment
We thank Dr. Ryan Roemmich for discussion of experimental techniques and advice on experimental design and statistical interpretation. N.S., S.S., C.T. were supported by NIH NHGRI R35 HG010714.

References

1.Knight, Stephen R et al. "Mobile devices and wearable technology for measuring patient outcomes after surgery: a systematic review." NPJ digital medicine vol. 4,1 157. 12 Nov. 2021, doi:10.1038/s41746-021-00525-1
2. Copeland, G P et al. "POSSUM: a scoring system for surgical audit." *The British journal of surgery* vol. 78,3 (1991): 355-60.
3. Haga, Y et al. "Estimation of Physiologic Ability and Surgical Stress (E-PASS) as a new prediction scoring system for postoperative morbidity and mortality following elective gastrointestinal surgery." *Surgery today* vol. 29,3 (1999): 219-25.
4. Kong, Cherng H et al. "Recalibration and validation of a preoperative risk prediction model for mortality in major colorectal surgery." *Diseases of the colon and rectum* vol. 56,7 (2013): 844-9.
5. Cho, Sylvia, et al. "Factors Affecting the Quality of Person-Generated Wearable Device Data and Associated Challenges: Rapid Systematic Review." *JMIR mHealth and uHealth* vol. 9,3 e20738. 19 Mar. 2021.
6. The All of Us Research Program is supported by the National Institutes of Health, Office of the Director: Regional Medical Centers: 1 OT2 OD026549; 1 OT2 OD026554; 1 OT2 OD026557; 1 OT2 OD026556; 1 OT2 OD026550; 1 OT2 OD 026552; 1 OT2 OD026553; 1 OT2 OD026548; 1 OT2 OD026551; 1 OT2 OD026555; IAA #: AOD 16037; Federally Qualified Health Centers: HHSN 263201600085U; Data and Research Center: 5 U2C OD023196; Biobank: 1 U24 OD023121; The Participant Center: U24 OD023176; Participant Technology Systems Center: 1 U24 OD023163; Communications and Engagement: 3 OT2 OD023205; 3 OT2 OD023206; and Community Partners: 1 OT2 OD025277; 3 OT2 OD025315; 1 OT2 OD025337; 1 OT2 OD025276. In addition, the All of Us Research Program would not be possible without the partnership of its participants.
7. Jansson, Karl-Åke, et al. "Health-related quality of life (EQ-5D) before and after orthopedic surgery." *Acta orthopaedica* vol. 82,1 (2011): 82-9. doi:10.3109/17453674.2010.548026
8. Fermont, Jilles M et al. "The EQ-5D-5L is a valid approach to measure health related quality of life in patients undergoing bariatric surgery." *PloS one* vol. 12,12 e0189190. 18 Dec. 2017, doi:10.1371/journal.pone.0189190
9. Noyez, Luc. "Is quality of life post cardiac surgery overestimated?." *Health and quality of life outcomes* vol. 12 62. 29 Apr. 2014, doi:10.1186/1477-7525-12-62

10. Brooks R: EuroQoL: the current state of play. Health Policy 1996, 37: 53–72. 10.1016/0168-8510(96)00822-6
11. Dolan P: Modeling valuations for EuroQol health states. Med Care 1997, 35: 1094–1108
12. *EQ-5D-5L User Guide - University of Nebraska Medical Center*. (https://www.unmc.edu/centric/_documents/EQ-5D-5L.pdf)
13. Koide, Ryo et al. "Quality assessment using EQ-5D-5L after lung surgery for non-small cell lung cancer (NSCLC) patients." *General thoracic and cardiovascular surgery* vol. 67,12 (2019): 1056-1061. doi:10.1007/s11748-019-01136-0
14. Panayi, A C et al. "Impact of frailty on outcomes in surgical patients: A systematic review and meta-analysis." American journal of surgery vol. 218,2 (2019): 393-400. doi:10.1016/j.amjsurg.2018.11.020.
15. Searle, Samuel D et al. "A standard procedure for creating a frailty index." *BMC geriatrics* vol. 8 24. 30 Sep. 2008, doi:10.1186/1471-2318-8-24
16. DeLong, E R et al. "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach." Biometrics vol. 44,3 (1988): 837-45
17. Tian, Lili et al. "Confidence interval estimation of the difference between paired AUCs based on combined biomarkers." Journal of statistical planning and inference vol. 139,10 (2009): 3725-3732. doi:10.1016/j.jspi.2009.05.001
18. Comparing Two ROC Curves – Independent Groups Design. Retrieved from https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Comparing_Two_ROC_Curves-Independent_Groups_Design.pdf
19. "Physical Activity." *World Health Organization*, World Health Organization, https://www.who.int/news-room/fact-sheets/detail/physical-activity.
20. "How Much Physical Activity Do Adults Need?" *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 2 June 2022, https://www.cdc.gov/physicalactivity/basics/adults/index.htm.
21. *BioPro: I Met Creatinine - Centers for Disease Control and Prevention*. https://wwwn.cdc.gov/nchs/data/nhanes/2015-2016/labmethods/BIOPRO_I_MET_CREATININE_DXC800.pdf.
22. "Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS)." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 1 June 2022, https://www.cdc.gov/me-cfs/pdfs/wichita-data-access/lab-tests-code.pdf
23. "Defining Adult Overweight & Obesity." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 3 June 2022, https://www.cdc.gov/obesity/basics/adult-defining.html#:~:text=If%20your%20BMI%20is%2018.5,falls%20within%20the%20obesity%20range.
24. Wells, Cameron I et al. "Wearable devices to monitor recovery after abdominal surgery: scoping review." *BJS open* vol. 6,2 (2022): zrac031. doi:10.1093/bjsopen/zrac031
25. McIsaac, Daniel I et al. "Frailty as a Predictor of Death or New Disability After Surgery: A Prospective Cohort Study." *Annals of surgery* vol. 271,2 (2020): 283-289. doi:10.1097/SLA.0000000000002967
26. Shen, Yanjiao et al. "The impact of frailty and sarcopenia on postoperative outcomes in older patients undergoing gastrectomy surgery: a systematic review and meta-analysis." *BMC geriatrics* vol. 17,1 188. 21 Aug. 2017, doi:10.1186/s12877-017-0569-2
27. Panayi, A C et al. "Impact of frailty on outcomes in surgical patients: A systematic review and meta-analysis." American journal of surgery vol. 218,2 (2019): 393-400. doi:10.1016/j.amjsurg.2018.11.020.
28. Saxton A, Velanovich V. Preoperative frailty and quality of life as predictors of postoperative complications. Ann Surg. 2011;253(6):1223–9.
29. Amin, Tajrian et al. "Wearable devices for patient monitoring in the early postoperative period: a literature review." *mHealth* vol. 7 50. 20 Jul. 2021, doi:10.21037/mhealth-20-13