Using RFID Data to Improve the Identification of Abandonment Behavior in an Emergency Department: Clinical Policy Implications

Yaniv Ravid
University of Toronto
y.ravid@rotman.utoronto.ca

Rouba Ibrahim
University College London
rouba.ibrahim@ucl.ac.uk

Junqi Hu
Old Dominion University
jhu@odu.edu

Kal Pasupathy
University of Illinois Chicago
kap@uic.edu

David Nestler
Mayo Clinic
nestler.david@mayo.edu

Vickie Ernste
Mayo Clinic
ernste.vickie@mayo.edu

Vahid Sarhangian
University of Toronto
sarhangian@mie.utoronto.ca

Philipp Afèche
University of Toronto
philipp.afeche@rotman.utoronto.ca

Abstract

Identifying which Emergency Department (ED) patients are likely to leave without being seen (LWBS) could enable interventions that reduce LWBS rates. Machine Learning (ML) models that updated these predictions as patients wait were developed and validated, to correctly identify more patients who LWBS. Using a dataset of 150,959 patient visits to the ED of an academic medical campus, two types of classification models were developed: (1) a static model that uses patient and ED census information at the time of arrival to predict the risk to LWBS; and (2) a time-dependent model that updates the predictions based on new information after 30 minutes for patients who are still waiting in the ED. Preliminary results show that the time-dependent model reduces the number of missed LWBS cases by approximately 50% as compared to the static model, without incurring any additional false-positives.

Keywords: Machine learning, emergency healthcare, healthcare information technology.

1. Introduction

1.1. The Cost of Patient Abandonment

Patients who leave without being seen (LWBS) after arriving at the emergency department (ED) and could face negative medical outcomes (Geers et al., 2020). In many instances, they may return to the ED with a worsened condition. While awaiting patients are served based on their medical condition, patients of similar condition but different LWBS probability may be better prioritized to avoid LWBS decisions. High rates of patients who LWBS indicate a mismatch between demand for emergency care and the hospital's ability to attend and evaluate patients in a timely fashion. Therefore, the ability to identify patients at high risk to LWBS is critical in improving patients' medical outcomes and efficiently serving all arriving patients.

Some studies have found that patients who LWBS are as likely to return to the ED as those who were seen by a doctor and leave against medical advice (AMA) (Geirsson et al., 2013), and are far more likely to return within 30 days due to injury than those who leave AMA. Among uncompleted visits to the ED, patients who LWBS are therefore likely to require future resources and care from the ED, possibly due to a worse condition than the reason for their original visit. These studies have also shown that mortality risk after decisions to LWBS are also not negligible.

1.2. Predicting LWBS

Past literature has developed machine learning models to predict whether patients will LWBS during their wait (Casey et al., 2018). These models have shown promise as data-driven methods to identify patients at risk to LWBS. These models, however, only utilize information that is available upon patients' arrival and as a result predict a static LWBS risk. This information, which includes patient's acuity level and the number of patients in the waiting room, is computed using the hospital's electronic medical records (EMR). It has been shown that EMR data inflates wait times for patients who LWBS (Geers et al., 2020). Therefore, the static predictions may not be accurate.

URI: https://hdl.handle.net/10125/112105
978-0-9981331-9-5

Static information also prevents models from updating patients' risk to LWBS as their wait time increases. This is even though patients' waiting time was shown to be a factor which strongly impacts the decision to LWBS (Geers et al., 2020). Accurate measurements of the time patients stay in the waiting room were computed in such studies by fitting patients with radio frequency identification (RFID) bracelets (Koenig et al., 2021, Griffin et al., 2020). These bracelets were given to patients at arrival, and in-ceiling readers receive signals from the wristbands and constantly record patients' location. This novel use of RFID data uncovered the gaps between patients' true experience in the waiting room and the information available in EMR data. Past research has also shown strong associations between changes to the state of the waiting area (e.g., observing arrivals and departures) and patients' risk to LWBS (Batt and Terwiesch, 2015). However, the models rely on ex-post changes to the environment when studying patient behavior. Nonetheless, these findings suggest that information available upon patient' arrival may be insufficient to accurately quantify their likelihood to LWBS. Therefore, environmental changes that can only be identified over time during patients' waits, such as observing other patients get seen by a doctor, can lead to an improved ability of identifying patients who LWBS.

1.3. LWBS and Information Systems

EDs have increasingly relied on sophisticated information systems (IS) to enhance patient flow, clinical decision-making, and operational efficiency. A classic example comes from Mount Sinai's implementation of an integrated Emergency Department Information System (EDIS) combined with workflow redesign. After adopting this integrated system, overall length of stay reduced by nearly 2 hours—from 6.69 to 4.75 hours (Baumlin, et al., 2010). Such improvements in service times are critical in mitigating LWBS behavior, but do not directly help us identify it before it happens.

Outside of reducing service times, communication tools also play a pivotal role in addressing LWBS behavior. For instance, offering frequent updates on estimated wait times has been shown to increase patients' willingness to wait (Arendt et al., 2003). However, accurately estimating wait times may be a difficult task, and inaccurate estimations may result in patient frustration and possibly higher rates of LWBS behavior. Instead, it may be beneficial to estimate a patient's tendency to LWBS, rather than their wait time.

Lastly, in some instances, tele-intake systems—leveraging remote clinician triage—have reduced both LWBS rates and provider wait times (Joshi et al., 2020). These interventions, nonetheless, require excess capacity in the form of remotely available clinicians, which may be costly. Digital tools that are less costly therefore may be preferable to enlisting additional remote staff.

1.4. Managerial Implications

Updating the predicted risk of each patient's tendency to LWBS over time results in better identification of patients who LWBS. This improved ability can help design interventions and practices that improve service. Introducing wait-time as a feature in the risk calculation can also help create guidelines and targets for how quickly patients should be evaluated. For example, the order in which patients are seen may be adjusted based on accurate predictions of their imminent tendency to LWBS. Such priority decisions are typically based on whether patients have been medically screened and how long they have been waiting for. The addition of time-dependent features in the LWBS risk dimension would prioritize patients that are likely to abandon, reduce inefficiencies and avoid costly outcomes.

1.5. Goals of this Investigation

This study aimed to improve upon traditional static prediction models which classify patients based on their risk to LWBS at any point during their wait. The results also highlight the importance of RFID data which gives accurate records of when patients exit the waiting room. The availability of such precise measurements allows us to track patients' experience in the waiting room and update patients' predicted risk to LWBS. The time-dependent models then predict patients' risk to LWBS at specific times during their wait. An experimental period is then used to evaluate whether these time-dependent prediction models can be more accurate than traditional static models. In addition to the area under the receiver operating characteristic curve (AUROC), another criterion used to compare predictive performance is the number of true-positives and false-positives generated by each model.

2. Materials and Methods

2.1. Setting and EMR Data Collection

A retrospective cohort study was conducted among all patient visits at an ED of a large academic medical center. Patient features used in the analysis were age, gender index (i.e., male or female) and Emergency Severity Index (ESI) score. These were recorded for each patient in the EMR system, which also indicated the time in which patients entered the waiting room. This *entering time* was recorded at the patient's moment of triage. All patients wait in the same room and can see all other waiting patients. The time at which each patient was assigned a bed was also recorded in the EMR data. At the time of bed assignment, a nurse calls for the patient and if present, the patient then exits the waiting room into the treatment area. This is defined as the patient's *exit time*. However, those patients who LWBS are missing when called on by the nurse. The nurse then records in the EMR system that the patient has left.

2.2. RFID Exit Times and Congestion Computation

RFID bracelets recorded the exit time for patients who LWBS. RFID bracelets were given upon arrival. When patients left the waiting room while wearing the RFID bracelet, the precise exit time was recorded in the RFID data. Equipped with these RFID timestamps in addition to the EMR timestamps, the number of patients present in the waiting room at any point in time t can be computed. This number is referred to as the *congestion* at time t. This congestion equals the number of patients whose entering time precedes t, and whose exit time (whether through bed assignment or LWBS) is after t.

2.3. Study Cohort

The studied dataset consisted of 150,959 patient-level entries of all ED visits over a two-year period. A total of 3,270 (2.17%) patients left without being seen. Of these patients, exit time was only available through RFID data for 3,022 (92%) patients. There is no discernible pattern that describes the group of patients who LWBS with missing RFID data. To understand how various factors impact LWBS decisions after a specific waiting time, any such patient whose LWBS time was missing were removed from the analysis. Additionally, two records with missing gender index data were removed. All 2,126 patients of ESI score 1 were also removed for the purpose of analysis. This is because over the two years of data collection, no patient with an ESI score of 1 was observed to LWBS, as expected, since these are usually patients with life-threatening conditions.

Lastly, pediatric patients (≤19 years) were served by nurses, doctors, and beds that were designated for pediatric care. Additionally, pediatric patients LWBS at a far lower rates than the general population. Therefore, pediatric patients from the dataset were also removed for the purpose of analysis. The remaining general population of patients choose to LWBS and much higher rates and are therefore a more critical group of patients whose behavior is crucial to understand.

2.4. Panel Data Creation

To study how the risk to LWBS evolved with patients' wait time, each patient visit was split into intervals, and patients' tendency to LWBS in each interval was studied. The number of intervals chosen impacts predictive accuracy. More than two intervals would result in decreasing predictive accuracy, for reasons explained below. Therefore, patient visits are split into two intervals. The first consists of the initial 30 minutes of the patient's waiting time. The second interval begins at the 30-minute mark of waiting and ends at the patient's exit time. The choice of 30 minutes for the length of the first interval balances the number of patients who LWBS in the first and second interval. If the first interval were too short, then a very small number of patients who LWBS will be observed in it. If the first interval were too long, then a very small number of patients who LWBS will be observed during the second waiting intervals.

RFID data also allows for accurate description each patient with a correct number of intervals. Consider a patient who LWBS within 30 minutes of arrival but was first noticed to be missing more than 30 minutes after their arrival. EMR records, owing to the approach described above, incorrectly associate two waiting intervals with this patient. However, using RFID data, only one waiting interval for the patient is correctly recorded.

Characterizing each patient by at most two intervals, the dataset is transformed into a panel dataset, where each observation is a patient-interval pair. For patients whose exit time is prior to the 30-minute mark of waiting, only one interval exists in the panel dataset. For those who wait longer than 30 minutes, two intervals exist in the panel dataset. Note that by choosing a larger number of intervals to split patients waits into, the panel dataset would grow. However, the number of observed LWBS cases in the dataset would not change. Therefore, the task of predicting LWBS behavior becomes more difficult.

2.5. Independent Variables

Patient features used for prediction were age, gender index and ESI level. The entering timestamp of each patient was used to determine the day of the week and the hour of day in which they entered the waiting room. Congestion upon arrival was also used as a feature for prediction. Note that these features are static and do not change as patients' wait continues. No other features such as time of symptom onset are available in the data.

Each patient-interval pair is also characterized by the *sojourn* time the patient has been waiting for at its onset (0 and 30 minutes for the first and second interval, respectively). Patient-interval pairs representing the second waiting intervals are also characterized by the congestion at the onset of the second interval. Additionally, one of the study's goals is to understand how events that patients saw in the past affect their future decisions. Therefore, the number of events that patients saw during their first waiting interval are also used to describe their second waiting interval. Specifically, these event counts include the number of arrivals, departures into service, and abandonments. Table 1 lists all the independent variables used in the analysis. We note that since all variables must be defined for all intervals, we impute the value of the event counts using their mean value for the first intervals. This way, second interval entries in the data with no observed event count at their onset are distinct in their event-count from first interval entries in the data.

2.6. Dependent Variables

The dependent variable associated with each patient-interval pair is a binary indicator which took value 1 if the patient decided to LWBS during the interval. If the patient was otherwise served during the interval, the indicator variable took value 0. Alternatively, for the first intervals of waiting, the indicator variable took value 0 if the patient continued to wait into a second waiting interval.

2.7. Predictive Models

The two years over which data was collected were split into a model *estimation period*, consisting of the first 20 months of data, and an *experimentation period* consisting of the final four months of data. Three predictive models, described below, were trained and tested using only the first 20 months of collected data. Patients who arrived during this 20-month estimation period were randomly divided into training (80%) and testing (20%) datasets. This testing set was used to validate the models and avoid overfitting. This was done by early stopping the training process when performance on the testing set began to worsen. The hyperparameter selection process, as described below, was also guided by performance on the testing population. During these procedures, the final four months of data were not used in any way. The final four months of data were instead used for the *experimentation* period during which the estimated models were applied. During this period, the models' ability to identify who, among patients that were held out from the estimation process, chose to LWBS was measured.

Model 1: A static benchmark model that resembles past LWBS prediction models (Casey et al., 2018) is first fit. The training and testing panel datasets were therefore reduced to only represent each patient encounter once. The model only used information that was available upon the patient's arrival at the waiting room. Any time-dependent features such as sojourn time and congestion after 30 minutes were therefore not used for prediction. The exact list of static features is shown in Table 1. The binary outcome these static features were used to predict took value 1 if the patient LWBS at any point during their wait, and 0 otherwise.

Model 2: Next, the full panel dataset is utilized and the task of predicting whether a patient will LWBS during each specific waiting interval is considered. The same static features used in Model 1 are relied on, with the addition of sojourn time. Sojourn time equaled 0 for the first intervals in the panel dataset, and 30 minutes for the second intervals. The binary outcome was now interval-specific and equaled 1 for patient-interval pairs whenever the patient chose to LWBS during the interval. The binary outcome equaled 0 otherwise. Unlike the static model, this model was able to update patients' risk to LWBS as their wait went on.

Model 3: The final statistical model, in addition to the stipulations in Model 2, further allowed for time-varying congestion effects and introduced more time-dependent features. For example, this model included the number of arrivals that a patient saw during their first 30 minutes of waiting to predict their risk to LWBS during the second interval. Table 1 indicates the exact variables used in the model. The binary outcome variable was identical to that used in Model 2.

Table 1. Defining prediction features, and their usage in the three models. (*These variables are only defined for the 2nd waiting intervals in the panel dataset, and are imputed with their mean value for the 1st intervals.)

Feature	Type	Description	Present in Models
Age	Binary	Focal patient's age	1,2,3
Gender Index	Numerical	Focal patient's sex, 1 (0) for male (female)	1,2,3
Day of week	Categorical	Six dummy variables for day of the week in which the patient arrived (Friday as reference class)	1,2,3
Hour of day	Categorical	Five dummy variables for the four-hour window in which the patient arrived (0am to 4am, etc., with noon to 4pm as the reference class)	1,2,3
ESI	Categorical	Three dummy variables for the ESI level of the patient (ESI 2 as reference class, ESI 1 not included in analysis)	1,2,3
Congestion at arrival	Numerical	Number of non-pediatric patients who are waiting at the onset of the focal patient's wait (i.e., initial congestion fixed effect)	1,2,3
Sojourn	Numerical	Number of hours the focal patient has been waiting for at the beginning of the waiting interval	2,3
Congestion_30*	Numerical	Number of non-pediatric patients who are waiting at the onset of the focal patient's 2nd waiting interval.	3
Departures_30*	Numerical	Number of non-pediatric departures into service observed by the focal patients in the 30 minutes prior to the onset of the 2nd waiting interval.	3
Arrivals_30*	Numerical	Number of non-pediatric arrivals observed by the focal patients in the 30 minutes prior to the onset of the 2nd waiting interval	3
Abands_30*	Numerical	Number of non-pediatric LWBS observed by the focal patients in the 30 minutes prior to the onset of the 2nd waiting interval	3

For all three models, the XGBoost classifier (Chen and Guestrin, 2016) was used to predict the risk to LWBS using the corresponding features. This ensemble model combines multiple non-linear classification algorithms, and as a result captures complex relationships between features and the decision to LWBS. Past literature has also shown that XGBoost specifically outperforms other learning algorithms in various ED contexts (Casey et al., 2018, Morey et al., 2025). The XGBoost Python package was used in the analysis. The objective used in the estimation period was the logistic probability of observing the LWBS decisions in the dataset. A high-performing model would therefore predict high probabilities for observations in the data that chose to LWBS. The metric for evaluating predictive performance and preventing overfitting was the AUROC.

For Model 1, the binary classification was made once for every patient and represented learning each patient's likelihood to LWBS at any point during their wait. For Models 2 and 3, the binary classification was made for each patient-interval pair and represented learning the time-dependent likelihood that the patient will LWBS during the interval.

The training population was used to estimate each model, and the testing population was used to ensure this estimation did not result in overfitting. A tree-specific booster was used for all models. The model hyperparameters were tuned for each model separately to maximize the testing performance. For each model, the set of hyperparameters that gave it the best AUROC on the testing dataset was chosen.

2.8. Out of Sample Performance

The final four months of data representing the experimentation period were then used to apply the models and measure their performance. The number of patients who LWBS that were caught and missed by each model in this period was measured. The number of false positives that were raised by each model was also calculated. A false-positive indicates a patient who is predicted to LWBS but did not LWBS. Specifically, two levels of prediction thresholds that result in different false-positive rates were considered. By choosing a high (low) threshold for prediction, the models predicted fewer (more) patients as likely to LWBS. If patient risk scores vary from 0 to 1, a high (low) threshold for prediction could mean only classifying patients with a risk score above 0.8 (0.2) as patients who will LWBS. As a result, a lower (higher) rate of false positives will be achieved. The number of patients who LWBS that were missed by each model when only 10% of all panel-dataset observations were allowed to be falsely predicted to LWBS was measured. Then, the number of patients who LWBS that were missed by each model were computed when this rate is increased to 50%. The two thresholds correspond to one set of sparing LWBS predictions, and one set of liberal LWBS predictions. While the former threshold will make few false-positive predictions, the latter will separate patients into two roughly equal groups: one group will be of high LWBS risk and the other of low risk. This equal separation of patients makes interpretation of LWBS predictions easier for priority purposes. We note that this predicted risk need not change how priority decisions depend on patients' medical risk, but simply enriches them by adding an additional dimension to the decision.

Additionally, the Shapley values (SHAP) of important features (Lundberg and Lee, 2017) were reported to indicate which features strongly impact the models' prediction output. These values measure the contribution that each feature makes toward the predicted risk to LWBS. When the Shapley value of a specific patient feature is very positive (negative), the feature then strongly increases (decreases) the predicted risk to LWBS. The SHAP package in Python was used for this purpose, with a prediction explainer specifically built for tree-based models.

3. Results

3.1. Cohort Characteristics

The cohort consisted of 119,326 patients. Table 2 describes this cohort, as well as the estimation period cohort used to develop the three predictive models, and the experimentation period cohort used to measure the models' performance.

Over the entire study cohort, 566 of the 2,911 patients who chose to LWBS did so within 30 minutes of arrival. Of those who LWBS in the first 30 minutes, 142 were detected as LWBS by a nurse within 30 minutes of their entering time. Therefore, a panel dataset built without RFID data would incorrectly include a second waiting interval for 424 patients.

On average, patients who LWBS are first noticed to be gone after 3 hours and 5 minutes. However, accurate exit time measurements using RFID data show that the average true wait time for these patients is 2 hours and 17 minutes. The average inflation in observed waiting time for LWBS patients is therefore 48 minutes.

3.2. Predictive Performance

The three models were used to predict the risk to LWBS for each patient who visited the ED in the four-month experimental period. Using the panel dataset defined by these visits, all three models were used to predict the risk to LWBS for all patient-interval pairs in the dataset. Model 1 achieved an AUROC of 0.801, while Models 2 and 3 both achieved an AUROC of 0.869. This gives an improvement in the AUROC of 0.068.

Table 2. Cohort statistics. Data is reported as n (percent), mean (SD), or percent.

	Entire Cohort	Model Estimation Cohort	Model Experimentation Cohort
Unique Visits (% LWBS)	119,326 (2.4%)	99,085 (2.5%)	20,241 (2.2%)
ESI 2	28,248 (0.7%)	23,959 (0.7%)	4,289 (0.6%)
ESI 3	71,714 (2.3%)	59,584 (2.4%)	12,130 (2.2%)
ESI 4	18,609 (5.2%)	14,905 (5.5%)	3,704 (4.2%)
ESI 5	755 (7.2%)	637 (7.5%)	118 (5.1%)
Age	54.89 (20.05)	54.82 (20.06)	55.25 (19.98)
% Male	48%	48%	48%
Exit time (hours)	0.70 (1.11)	0.69 (1.10)	0.73 (1.12)
ESI 2	0.27 (0.58)	0.28 (0.60)	0.25 (0.52)
ESI 3	0.76 (1.14)	0.76 (1.14)	0.78 (1.16)
ESI 4	1.09 (1.32)	1.08 (1.33)	1.11 (1.28)
ESI 5	0.91 (1.28)	0.93 (1.32)	0.80 (1.02)
Congestion upon arrival	6.18 (6.95)	6.08 (6.89)	6.67 (7.26)
Hours to LWBS	1.58 (1.15)	1.57 (1.15)	1.63 (1.11)
Offered wait time for patients who LWBS (hours)	2.50 (1.52)	2.51 (1.54)	2.44 (1.39)

The three models' precisions on the experimental cohort are illustrated in Table 3. Model 1 gave the highest number of missed LWBS cases both at 10% and 50% false-positive rate. Models 2 and 3 gave the lowest number of missed LWBS cases under the same false-positive rate. At the 50% false-positive rate specifically, Model 1 missed nearly twice as many LWBS cases as Models 2 and 3 (34 missed LWBS cases compared to 19, over a period of four months).

3.3. Feature Importance

Lastly, Figure 1 illustrates the mean absolute Shapley values of the features in the time-dependent Models 2 and 3 among the experimentation cohort. Congestion observed upon arrival has the largest average absolute impact on the predicted risk in both models. Sojourn time is the second most impactful feature in both models as well. The event count features in Model 3 are not as significant for prediction as basic patient characteristics and sojourn waiting time.

Table 3. Comparing the three models during the four-month experimentation period. The middle (rightmost) column corresponds to a high (low) prediction threshold which results in lower (higher) number of false-positives and true-positives. The total number of patients who LWBS in the four-month period was 452.

	LWBS missed at 10% false-positive rate	LWBS missed at 50% false-positive rate
Model 1	280	34
Model 2	184	19
Model 3	184	19

(Note: since many observations shared predicted risk values, exact false-positive rates cannot always be achieved. For each model, the number of missed cases that most closely achieves the desired rate is therefore reported.)

4. Discussion

Patients' decision to LWBS while waiting in the ED presents a major operational hurdle to many hospitals. Recent research has shown the capability to identify patients with high tendency to LWBS. However, this study shows that more cases can be identified, and fewer cases can be missed by also predicting when during the patients' wait is the tendency to LWBS the highest. Instead of treating LWBS as a single binary outcome for every patient, it was treated as a time-dependent risk level that changed with patients' experience in the waiting room. RFID data allowed for accurate measurements of the length of time patients spent in the waiting room, which would otherwise be inflated as the true moment at which patients LWBS would be unknown. RFID data also allows for the incorporation of time-dependent features in prediction models. Compared to models that ignore these features, the results showed that a static binary prediction performed poorly and missed more LWBS cases than a time-dependent risk level can otherwise identify.

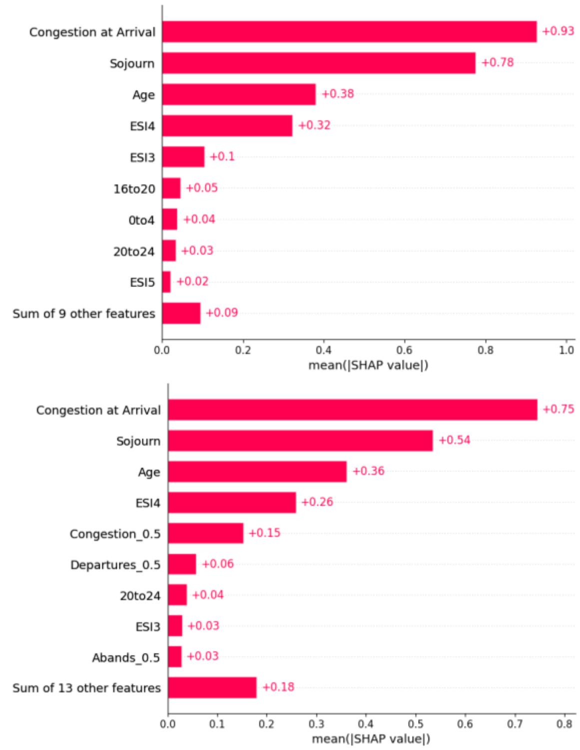

Figure 1. Shapley values of the features in Model 2 (top) and Model 3 (bottom), ranked by average absolute impact on the final predicted value.

Our initial results show an improvement in AUROC from the static score of Model 1 to the time-dependent score of Model 2. This time-dependent score was only based on the amount of time patients have waited. During the experimental cohort representing four months of ED arrivals, the time-dependent predictor always found more LWBS cases than its static counterpart. This result highlighted the importance of updating LWBS risk predictions over time.

An extension to the time-dependent model which accounted for events that occurred during patients' waits in addition to the time they have spent waiting was studied in Model 3. These events were used when computing patients' risk to LWBS. However, this extension did not outperform the basic time-dependent predictor, despite the additional features, and in fact resulted in a worse AUROC. Including more granular information about what patients observe during their wait does not improve our ability to predict their tendency to LWBS. A simple model that only tracks how long the patient has been waiting can perform just as well without the additional time-dependent features, despite their sometimes-significant impact to predicted risk. Among these features, the Shapley value of the congestion present at the 30-minute mark significantly impacts patients' risk to LWBS in their second waiting interval. This could be since the risk to LWBS is only predicted twice, and so both sojourn time and congestion at 30 minutes take value zero for patients in their first waiting interval. As a result, both features indicate that the patient is in their second waiting interval, which could be the underlying source to LWBS risk. Across all models, initial congestion strongly impacted the predicted risk. Sojourn time also had a large effect on the predicted risk to LWBS.

The XGboost model has been used to predict static LWBS risks in the past (Casey et al., 2018), but rarely do these studies report model precision (Hunter et al., 2024). These precision measurements indicate the level of incorrect noise in the form of false-positive predictions that a model produces. While identifying LWBS cases is a critical objective, highly noisy predictions would make it difficult to separate the true LWBS cases from those that are falsely predicted to LWBS. To shed light onto the precision at which LWBS cases can be identified, the analysis illustrates the precision of both the traditional static binary predictor and the time-dependent risk models. The results showed that under similar false-positive rates, the time-dependent predictors always outperform the static ones. When the models are allowed to falsely predict nearly half of all observations in the experimental period as likely to LWBS, Models 2 and 3 correctly identified 95.8% of all patients who LWBS. In effect, the time-dependent prediction models can be used to separate the population of waiting patients into two equal groups, those with high risk to LWBS and those of low LWBS risk. Based on the experimentation cohort, nearly all of patients who LWBS will be correctly placed in the high-risk group. Across a period of four months, with more than 100 daily arrivals, this equates to only 19 missed LWBS cases, approximately one per week.

4.1. Policy Implications

The ability to find patients who will abandon during their wait, as well as the intervals during their wait in which they will likely LWBS, also allows ED staff to intervene and respond to the risk. Possible interventions may take the form of adjusted scheduling decisions and prioritizing high LWBS risk patients.

Specifically, our finding supports policies that emphasize continuous data capture and real-time analytics rather than one-time risk assessment. As decisions to LWBS are time-dependent due to instantaneously changing willingness to wait, so should constantly change the digital information representing patients in the hospital's electronic records. ED information systems and EMR modules would need to log time-stamped waiting room events and automatically trigger recalculation of LWBS predictions at predefined intervals. Dashboards used by the ED staff could display color-coded, continuously updated risk indicators, creating a shared situational awareness.

4.2. Limitations

A limitation of the study is the lack of patient-level clinical variables (e.g., chief complaint, medical history, lab, and imaging results) that may further improve the ability to identify LWBS behavior.

Additionally, predictive models cannot reliably evaluate counterfactual scenarios. If, for instance, the medical staff was increased to reduce congestion, it cannot be claimed that LWBS rates will decline. Therefore, designing interventions based on predicted risk to LWBS should be done carefully, accounting for such counterfactual scenarios.

Our time-dependent predictions of patients' tendency to LWBS also require real-time tracing of patients' location, which may not be physically or financially for many EDs.

4.3. Future Work

Our findings open the door to research that could further expand on how data and information can be used to address patients' choices to LWBS.

First, incorporating specific patient conditions and personal characteristics such as major complaint and insurance status could significantly improve the identification of patients who will LWBS as compared to the predictive models we explore using limited patient characteristics. Measuring the importance of such features would also allow us to compare how operational factors, such as congestion and time spent waiting which we explore, compare with medical and personal information. Understanding the interaction between these two, and how medical and operational factors together affect patients' tendency to LWBS, would greatly expand our knowledge of how to assign risks of LWBS to patients. For example, two patients of similar medical characteristics who observe similar congestion and events would have similar predicted risk to LWBS under our formulation, but significantly different insurances statuses might make one highly more likely to LWBS than the other. More specifically, insurance status might make the patient far for sensitive to observing some event, such as further arrivals. As a result, understanding how such features affect the tendency to LWBS are critical in better identifying the behavior before it occurs.

Future work may also expand on how to accurately identify patients who LWBS without RFID data. Such tracking information unlikely to be available in many clinical settings would make our results hard to act on, and so finding alternative prediction models that do not utilize such data would have a large impact on the ability to address LWBS in data-poor settings.

Additional work can also implement prediction models of patients' risk to LWBS. Analyzing the resulting rates of patients who LWBS would shed light on the true power of strong predictive performance. If by better identifying patients who will LWBS, the overall rates of abandonment are reduced, then a strong link between predictive power and improved performance can be made. Alternatively, showing that accurate LWBS prediction results in a change to patient behavior and a corresponding inefficiency of risk-based identification would point to the need for stronger interventions. Simply identifying high risk patients would not be enough and should be accompanied by alternative workflows or increased capacity.

5. References

Arendt, K. W., Sadosty, A. T., Weaver, A. L., Brent, C. R., & Boie, E.T. (2003). The left-without-being-seen patients: what would keep them from leaving? *Annals of Emergency Medicine*, 42 (3), 317 - 323.

Batt, R. J., & Terwiesch, C. (2015). Waiting Patiently: An empirical study of queue abandonment in an emergency department. *Management Science*, 61(1), 39-59.

Baumlin, K. M., Shapiro, J. S., Weiner, C., Gottlieb, B., Chawla, N., & Richardson, L. D. (2010). Clinical information system and process redesign improves emergency department efficiency. *Joint Commission Journal on Quality and Patient Safety*, 36 (4), 179 - 185.

Casey, P., Zolfaghar, K., Eckert, C., Waters, L., Sonntag, H., McKelvey, T., & Mark, N. M. (2018). Predicting patients at risk for leaving without being seen using machine learning. *Annals of Emergency Medicine*, 72 (4), S5-S6.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining,* 785-794.

Geers, J., Pasupathy, K., Lovik, K., Finley, J., Hellmich, T., Marisamy, G., Nestler, D., Sadosty, A., Sir, M., & Heaton, H. (2020). Characterization of emergency department abandonment using a real-time location system. *The American Journal of Emergency Medicine*, 38 (4), 759-762.

Geirsson, O. P., Gunnarsdottir, O. S., Baldursson, J., Hrafnkelsson, B., & Rafnsson, V. (2013). Risk of repeat visits, hospitalisation and death after uncompleted and completed visits to the emergency department: a prospective observation study. *Emergency Medicine Journal*, *30*(8), 662-668.

Griffin, J. M., Hellmich, T. R., Pasupathy, K. S., Funni, S. A., Pagel, S. M., Srinivasan, S. S., Heaton, H. A., Sir, M. Y., Nestler, D. M., Blocker, R. C., Hawthorne, H. J., Koenig, K. R., Herbst, K. M., & Hallbeck, M. S. (2020). Attitudes and behavior of health care workers before, during, and after implementation of real-time location system technology. *Mayo Clinic Proceedings: Innovations, Quality & Outcomes*, 4 (1), 90-98.

Hunter, D., Carr, B., Morey, J., & Jones, D. (2024). 242 Machine learning model to predict emergency department patients who left without being seen. *Annals of Emergency Medicine*, 84 (4), S112 - S113.

Joshi, A. U., Randolph, F. T., Chang, A. M., Slovis, B. H., Rising, K. L., Sabonjian, M., Sites, F. D., & Hollander, J. E. (2020). Impact of Emergency Department Tele-intake on Left Without Being Seen and Throughput Metrics. *Academic Emergency Medicine,* 27 (2), 139-147.

Koenig, K. R., Pasupathy, K. S., Hellmich, T. R., Hawthorne, H. J., Karalius, V. P., Sir, M., Das, D., Heaton, H. A., & Nestler, D. M. (2021). Measuring sensitivity and precision of real-time Location systems (RTLS): Definition, protocol and demonstration for clinical relevance. *Journal of Medical Systems*, 45 (1), 15.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processes Systems*, 30.

Morey, J., Winters, R., & Jones, D. (2025). Artificial intelligence to predict billing code levels of emergency department encounters. *Annals of Emergency Medicine*, 85 (1), 63 - 73.