Multi-Stage Predictive Model of Patients Likely to Overstay

It has been observed that there are patients that overstay at hospitals. They do not leave the hospital even though they are deemed medically ready for discharge.

Literature review and medical staff experience indicate that the major reasons for overstay is the need to evaluate the patient's post discharge requirements and prepare the necessary external accommodations; such as home care, nursing home or chronic care. A Transfer Coordinator facilitates the arrangements for home-care, nursing home or chronic care. This process takes time due to consultations with family and allied health care workers, waiting for an opening in a nursing home or chronic care facility, and possible house renovations in home care cases (bathroom fittings, ramps, etc). If this process is delayed, the required accommodations may not be ready when the patient is deemed medically fit for discharge.

Some of these patients that need assistance are discharged on time for various reasons: identified and assessed by a Transition Coordinator in a timely manner, socio-economic reasons, previous home-care assessments, already have nursing-home/chronic care bed waiting, etc.

Others, however, are not identified on time; assessed properly; or their needs are very complex; for some reason, they do not leave on time. We aim to develop a model that can be used to identify these difficult cases, so that a discharge assessment can be initiated and monitored as early by a Transition Coordinator, in order to monitor that the required arrangements are met in a timely manner.

Dataset: January 1, 2007 to March 31, 2012

We have admission data from four hospitals (GRA, HSC, STB, VIC) from Jan. 01/2007 to March 31/2012, for a total of 49,947 records. Patient information includes Admission Date, Transfer Ready Date (when patient deemed medically fit for discharge), Discharge Date, various physiological information that can be obtained with 48 hours of admission (respiratory rate, blood pressure, etc), pre-existing conditions such as diabetes, stroke or heart attack, dementia, activities of daily living score (ADL) and sub categories such assistance needed to dress, feed, use the bathroom. The reason for admission is also recorded (stroke or heart attack), the admission from location and the discharge location to: home (ZZ), various long term care facilities (nursing homes & chronic care) and other wards in various hospitals in the province/city (from there, they may have gone home or long term care).

Overstay is defined as the number of days from the Transfer Ready Date (TRD) to the Discharge Date. Patients that are discharged within 36 hours of the Transfer Ready Date are considered to be discharged "on-time". In order to build the predictive model for potential overstay patients we eventually need to use the TRD. Not all patients have this information recorded. In fact, it stopped being a requirement to record the TRD in the spring of 2010 and resumed in the fall of 2011.

The working dataset consists of MB admissions from January 01/2007 to May 14/2010 and September 15/2011 to March 31/2012. After the palliative patients are removed, we have 35,871 records.

Additional fields need to be calculated from the available data. The new “ZZ” field is 1 for any discharge to ‘ZZ’ in the existing “to” field, 0 otherwise. The new “NH” field is 1 for any discharge to the following: ‘MW’, ‘NW’, ‘DW’, ‘RW’, or ‘AW’, 0 otherwise. The new “XW” field is set to 1 if field “ZZ” and field “NH” are 0. If the “to” field is missing it is labeled as XW, meaning they did not go home nor to a nursing.

The new “AGE” field is calculated from admission date and birth date: AGE=FLOOR((admit-DOB)/365.25,1).

Not all Transfer Dates have the time recorded (some just have a date). In which case, a default of 12 noon was used. The new field “OVERSTAY” uses the “disch” and “transferdate” fields. OVERSTAY=ROUND(disch – transferdate, 1).

An OVERSTAY value of zero (0) means the patient left within the first 12 hours (the same day). If the patient left > 12 hours but < 36 hours after the transfer ready date, then they were considered to have 1 day overstay, etc. Patients are considered overstay if the field OVESTAY has a value of 2 or more days.

The dataset is now divided into 2/3 training set (used to build the models) and 1/3 test set (used to validate the models) based on time of admissions. This will simulate how the model will be used in the pilot study.

The training set consists of 23,858 records from January 1, 2007 to July 31, 2009. The test set has 12,013 admissions from August 1, 2009 to March 31, 2012.

Preliminary Analysis

Preliminary analysis and patient profiles show that the ADL score is one of the major factors in distinguishing patients that go home without the need for additional assistance and the patients in need of a Transition Coordinator. Table 1 shows a clear distinction between the patients that went home (ZZ) on time versus the patients that went to a nursing home or chronic care facility on time (NH). On-time ZZ patients are younger (median age of 62) and a median ADL of 3. On-time NH patients are far older (82 years) and much higher ADL of 30.

Table 1 also shows that ZZ overstay patients have higher ADL than on-time ZZ, but that NH overstay patients do not have such a drastic difference in ADL over on-time NH. The higher

ADL score among overstay ZZ patients would suggest a need for homecare; most NH patients have very large ADL scores. There is an obvious trend, the older the patient with higher the ADL score the more need for post discharge assistance and a Transition Coordinator.

Table 1:

	ON-TIME		OVERSTAY							
MEDIAN	ZZ < 48 hr	NH < 48 hr	NH 3-5	NH 6-9	NH 10-14	NH 15+	ZZ 3-5	ZZ 6-9	ZZ 10-14	ZZ 15+
Trans_LOS (Days)	6	9	10	17	21	24	9	11	12	16
AGE	62	82	82	84	83	83	79	82	81	83
BATH	3	6	6	6	6	6	3	3	3	3
DRESS	0	6	3	3	3	6	3	3	3	3
TOILET	0	6	3	3	3	3	3	3	3	3
TRANSFER	0	6	6	3	3	3	3	3	3	3
CONTINENCE	0	6	3	3	3	3	0	3	3	3
FEED	0	3	3	3	3	3	0	0	0	0
ADL Score	6	30	24	21	21	24	15	15	15	18
ADL 3rd/4th Quart	0 - 15	18-36	15 - 33	12 - 33	12 - 33	15 - 33	6 - 21	9 - 24	9 - 27	12-27

We are most interested in identifying the long overstay patients, ZZ or NH with an overstay of 10 days or more. These patients have the highest impact on hospital resources in terms of wasted bed days. If these patients' overstay can be reduced, it would result in substantial savings.

Using "overstay" vs. "on-time" as the predictive objective did not give adequate results. The "on-time" group includes the lowest ADL patients **and** the highest ADL patients. This confounds classifiers, two cohorts that have dissimilar features yet have the same label.

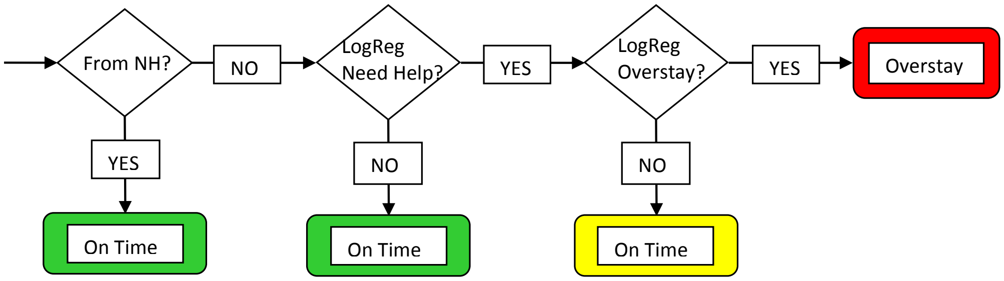
Who are the overstay patients? The majority of these patients come from the cohort that are not able to manage their daily lives on their own and need services such as home care, nursing home or chronic care. These patients become overstay patients because it takes time to determine and arrange their needs from the various Allied Health Care services.

A better approach is to first account for the patients that do not need help to be discharged on time; the younger, low ADL patients with adequate cognitive capabilities. This group can be separated from the group that will need assistance to manage their lives outside the hospital.

Additional model(s) can then focus on the differences between the patients that need post discharge assistance and are likely to overstay if this assistance is not arranged in a timely fashion.

A multi-stage predictive model is proposed. The 1st stage filters for patients that came from a long term care facility. These patients should go back to the same location on time. The 2nd stage identifies the patients that are likely to go home on their own and not need post discharge assistance. The 3rd stage identifies the patients that are most at risk of overstay, the patients with complicated needs that need to be met in order to be discharge in a timely manner. These potential overstay patients need a Transition Coordinator to monitor their progress.

The multi-stage model will label a patient as: GREEN (do not need assistance; will leave on time), YELLOW (needs assistance; may not leave on time) or RED (needs assistance **and** potential long overstay, monitor patient's progress to mitigate length of overstay).



1st Stage

This stage does not require a training set; it is a simple IF-ELSE decision. If the patient was admitted from a nursing home or chronic care facility, the patient will be discharged on time. To validate this stage (determine how many patients actually left on time) we need a transfer date.

In the test set, only 381 patients came from a MB nursing home (field “NH_CC” set to 1) **and** had a transfer date. Of these patients, 275 left within 36 hours. That means this stage is correct 72.2% of the time.

The average Transfer_LOS for on-time NH patients is 7 days. The average Transfer_LOS for the overstay NH patients is 22 days. This means that when a NH patient needs to stay longer in the hospital they are likely to overstay. This could be due to deconditioning; a minimum level of activity is required for a patient to return to the long term care facility.

Comment [EV1]: These patients were not selected based on overstay, but because they came from a nursing home. The assumption is that these will be “on-time” patients because they are in the system and have a place to go, but they did not leave on time…. Why? There must be a reason for such long overstay. For those that left on time, asked to go within 1 week, for overstays, median transfer date is 2 weeks… and then they overstay quite a bit.

2st Stage

Further analysis (Table 2) identified three additional factors in separating "on-time" ZZ from the rest of the hospital population, other than age and ADL. CVA stands for Cerebral Vascular Attack. Dementia, stroke and previous CVA are more prevalent with NH and long overstay ZZ patients than on-time ZZ patients.

Table 2.

	ON-TIME		OVERSTAY									
	ZZ < 48 hr	NH < 48hr	NH 2-4	NH 5-9	NH 10-14	NH 15+	ALL NH	ZZ 2-4	ZZ 5-9	ZZ 10-14	ZZ 15+	ALL ZZ
Dementia	3%	44%	34%	27%	25%	40%	36%	9%	12%	15%	20%	13%
CVA	11%	31%	32%	31%	27%	29%	29%	20%	22%	21%	23%	21%
Stroke	2%	9%	18%	22%	12%	8%	12%	3%	3%	3%	4%	3%

Again, if we consider the on-time groups (|ZZ & NH), patients that go home have a markedly different profile from patients that go to a nursing home. Should these patients be grouped in the same cohort for the purposes of building a predictive model? Not if we use the data available to us, on-time ZZ patients are a different cohort from the on-time NH patients.

When building predictive models it is important to be clear on what is the predictive objective. At this stage the predictive objective is **not** overstay patients vs. on-time patients. Overstay patients are a subset of the patients that need assistance to live outside the hospital. Patients that need home care or nursing home/chronic care are the ones that could overstay patients.

The predictive objective at this stage is the patients that need post-discharge assistance such as home-care or nursing home.

We wish to separate the cohort of patients that go home on time (ZZ < 36 hours), from patients that need assistance. The second group are the patients that go to a nursing home/chronic care and go home late (ZZ >= 36 hours). This second group of patients will then go through STAGE 3 where the predictive objective is long-overstay patients.

Three predictive models were evaluated; Neural Net, Decision Trees and Logistic Regression. All three gave similar performance. In order to optimize the number of patients that need to be evaluated by the Transition Coordinator, a model that allows for the easy adjustment of sensitivity and specificity is preferred. Logistic regression is the easiest model to adjust and is recommended for STAGE 2.

Training Set

Patients that expired were removed from the training set along with patients that were admitted from nursing home/chronic care. The field "hsurv" has S if patient survived, E if the patient died.

Only MB patients were used in the training set.

The patients that went home (or another ward XW) on-time are assumed to be the patients that do not need home-care or nursing home. It must be noted that some ZZ patients left on time but needed home care. Since we do not know which ZZ patients were assessed for home care, we are using on-time ZZ discharge as a proxy for "do not need home care". Using on-time discharge as a proxy will give us some false positives (the ones that required home care) when testing the model, this should be kept in mind when evaluating the model's performance.

Since we can only use ZZ & XW patients with an overstay value, ZZ & XW patients without a transfer date were removed from the training set.

All patients that did not go to a nursing home and left within 36 hours of transfer ready date were labeled as NO-ASSISTANCE (ZZ=9,258, XW=787).

All patients that went to a nursing home or chronic care (NH) were labeled as ASSISTANCE (2,303). All patients that went ZZ or XW 36 hours or more past their transfer date were labeled as ASSISTANCE (ZZ=2,004, XW=204) on the assumption that they were delayed because home care needs were not met on time.

At this point we have 14,556 records; 4,511 admissions labeled ASSISTANCE and 10,045 admissions that were labeled NO-ASSISTANCE.

Logistic Regression Model

The data mining software WEKA was used to build the various predictive models. http://www.cs.waikato.ac.nz/ml/weka/

Logistic Regression results were calculated using the formula found here, http://en.wikipedia.org/wiki/Logistic_regression, with the coefficients returned by the WEKA LogReg model.

$$f(z) = \frac{e^z}{e^z+1} = \frac{1}{1+e^{-z}} \qquad z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k,$$

Using a probability threshold of 0.5, WEKA returned the following results on the training set using 2/3 build and 1/3 validation: sen=56.4% spec=89.9% PPV=71.5

	ASSIST VS. NO-ASSIST			
	CONDITION	ASSIST	NO-ASSIST	
TEST	ASSIST	2542	1011	
	NO-ASSIST	1969	9034	
		Sensitivit	Specifity	PPV
	OVERSTAY	56.4%	89.9%	71.5%

The data elements in Table 3 gave the best results. Weka returns the coefficients of the model when the entire training set is used. The above results are due to splitting the ***training set***. We have a separate test set which simulates the model in operation (consecutive admissions in subsequent years). The results of the test set in the next section is the expected performance during the pilot study.

The coefficients for the Logistic Regression are listed in the table.

Table 3:

AGE	0.0517
CVA	0.133
Dementia	1.4224
CNSI_Any	0.9095
BATH	0.1054
DRESS	0.1633
TOILET	0.1265
TRANSFER	0.1819
CONTINENCE	0.0663
ADLScore	-0.0469
Intercept	-5.6882

Test Set

We want to simulate how the model would perform during normal hospital admissions. Palliative care patients and the ones that went ZZ or XW without a transfer date were removed from the test set. For ZZ/XW patients we are using overstay and on-time discharge as proxies for ASSISTANCE vs. NO-ASSISTANCE. With no transfer date we do not have an overstay value and can't label a patient one or the other, so they were removed. We now have a total of 8,095 admissions in the test set.

All the ZZ/XW patients with a transfer date were tested. Those discharged on time (< 36 hours) were labeled as NO-ASSISTANCE (ZZ=4,131, XW=301). Those with overstay >= 36 hours were labeled as ASSISTANCE (ZZ=1,120, XW=135).

Patients that were admitted from a NH were removed from the test set as they were handled in STAGE 1. The remaining NH patients were labeled as ASSISTANCE and tested (NH=1,113)

All the patients that expired were tested (EXP=1,295). This is done because patients that died will be admitted and we want to know how much of an impact they would have in terms of potential resources being assigned to them.

They were labeled as ASSISTANCE even though they may or not be (some could have gone ZZ on-time or NH if not died). Since the assignment of the test label is arbitrary, expired patients cannot be used in calculating specificity, sensitivity and PPV. Expired patients are used only to calculate the total % of patients that would be flagged as needing post discharge assistance (in stage 2) and overstay (in stage 3).

Logistic Regression Results on Test Set

We need to evaluate Stage 2 in order to determine what threshold level to use for the regression model to obtain adequate sensitivity and specificity. We also need to know the potential work load generated (% of patients flagged). This gives us an upper limit and we can adjust the regression threshold not only for sensitivity and specificity but also for expected workload.

Table 4:

Build Model: Jan 2007 to Apr 2009 / Test Model: Apr 2009 to May 2010													
				total %	< 36 hours flagged			>= 36 hours flagged			No Tdate	
	sen.	spec.	ppv	flagged	ZZ	XW	NH/CC	3-5 days	6-9 days	10+ days	NH/CC	expired
ASSIST vs. NO-ASSIST LogReg (0.40)	70.4%	83.2%	69.1%	41.1%	8.0%	1.2%	2.9%	4.5%	2.7%	7.4%	3.1%	11.3%
ASSIST vs. NO-ASSIST LogReg (0.45)	64.5%	86.2%	71.5%	36.7%	6.5%	1.0%	2.9%	3.8%	2.4%	6.9%	2.9%	10.3%
ASSIST vs. NO-ASSIST LogReg (0.50)	58.4%	88.8%	73.1%	32.5%	5.3%	0.9%	2.8%	3.3%	1.7%	6.2%	2.8%	9.5%

Table 4 gives the sensitivity/specificity/PPV and % of patients that will pass on to STAGE 3 for overstay prediction. A threshold of 0.45 gives the best sensitivity at the cost of specificity. Flagging on-time ZZ patients does not mean wasted resources as this group is not guaranteed to be composed exclusively of patients not needing assistance. A number of on-time ZZ patients will need home care assistance, but receive it on time thus labeled as NO-ASSISTANCE.

Double the number of long overstay (10+ days) patients are flagged as opposed to short overstays (3-5 days), which is one of the desired outcomes, to not miss long overstays. The largest percentage of flagged admissions is the expired patients group.

Evaluation of the model's performance can also give us more information than just sensitivity/specificity/PPV. For example, the sensitivity is for the correct classification of all patients labeled as ASSISTANCE. This includes short overstay and long overstay patients. We are interested in the long overstay patients. We need to break down the performance on a per group basis as shown in Table 5. For LogReg threshold of 0.45, 72.8% of the 10+ day overstays are flagged and will pass on to Stage 3.

Table 5:

Build Model: Jan 2007 to Apr 2009 / Test Model: Apr 2009 to May 2010 (Green=NO-ASSISTANCE; Yellow=ASSISTANCE)														
	< 36 hours			3-5 Days			6-9 Days			10+ Days			No Tdate	
ASSIST vs. NO-ASSIST LogReg (0.40)	ZZ	XW	NH/CC	ZZ	XW	NH/CC	ZZ	XW	NH/CC	ZZ	XW	NH/CC	NH/CC	Expired
Correctly Pred.	3482	204	238	274	27	65	150	16	52	218	37	340	249	911
Incorrectly Pred.	649	97	37	292	30	9	83	9	8	103	16	48	67	384
% Correctly Predicted	84.3%	67.8%	86.5%	48.4%	47.4%	87.8%	64.4%	64.0%	86.7%	67.9%	69.8%	87.6%	78.8%	70.3%
	83.2%		86.5%	52.5%			68.6%			78.1%			78.8%	70.3%
ASSIST vs. NO-ASSIST LogReg (0.45)	ZZ	XW	NH/CC	ZZ	XW	NH/CC	ZZ	XW	NH/CC	ZZ	XW	NH/CC	NH/CC	Expired
Correctly Pred.	3601	221	233	229	21	59	132	13	50	194	32	329	236	835
Incorrectly Pred.	530	80	42	337	36	15	101	12	10	127	21	59	80	460
% Correctly Predicted	87.2%	73.4%	84.7%	40.5%	36.8%	79.7%	56.7%	52.0%	83.3%	60.4%	60.4%	84.8%	74.7%	64.5%
	86.2%		84.7%	44.3%			61.3%			72.8%			74.7%	64.5%
ASSIST vs. NO-ASSIST LogReg (0.45)	ZZ	XW	NH/CC	ZZ	XW	NH/CC	ZZ	XW	NH/CC	ZZ	XW	NH/CC	NH/CC	Expired
Correctly Pred.	3706	228	223	198	17	54	82	12	46	158	31	309	222	763
Incorrectly Pred.	425	73	52	368	40	20	100	13	14	163	22	79	94	532
% Correctly Predicted	89.7%	75.7%	81.1%	35.0%	29.8%	73.0%	45.1%	48.0%	76.7%	49.2%	58.5%	79.6%	70.3%	58.9%
	88.8%		81.1%	38.6%			52.4%			65.4%			70.3%	58.9%

3rd Stage

The main purpose of the 2nd stage of the model is to identify the patients who do NOT need home care or nursing home after being discharged from the hospital and thus are not likely to overstay. These are the majority of the patients admitted to the hospitals and will be discharged on time; these patients will be labeled GREEN. The patients in need of some sort of post discharge assistance are flagged, labeled YELLOW, and passed on to the 3rd stage to determine if they are very long overstays (+10 days).

The 3rd stage of the model is meant to reduce the patients flagged as in need of home care/nursing home to a smaller set, the ones that are most at risk of very long overstay, labeled RED. The patients that overstay 10+ days past the transfer date should be closely monitored to mediate potential delays.

Audits of very long overstay case studies indicate the one of the major reasons for the overstays is that these patients' ability to take care of themselves is not so low that nursing home or chronic care are obvious choices. These patients could be discharged home, but would need substantial family assistance. The families need time to decide where to place the patient. If these consultations are not started early enough, delays in discharge are very likely.

Analysis of patient data corroborates the above observation. Figure 1 shows the box plot of the major factor for patient placement, ADL score. The on-time NH patients have the highest ADL score yet leave on time. The placement of these patients is obvious early on and arrangements can start early. The same is true for the short ZZ overstay patients (2-4 days). They have the lowest ADL scores and discharging them back home with home-care is probably a quick decision.

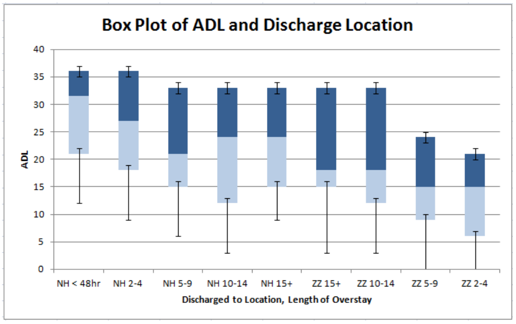


Figure 1.

The long overstay ZZ patients have ADL score comparable to most of the long overstay NH patients. The 3rd stage aims to identify the "bubble" patients, the ones that could go ZZ with home-care or to a NH. These patients are the most challenging for the discharge team as they require the most family consultations and compromises by the families.

Logistic Regression

Another logistic regression model was used for stage 3. The challenge at this stage is to find a boundary between short/long overstay that will separate the 10+ day overstay from the 0-4 days overstay. The closer the overstay threshold is to 10+ days, the more difficult it is to separate the 10+ days. A threshold of medium overstay (5+ days) was used. It is better to flag 5-9 days as long overstay than to miss too many 10+ days.

Group 1: LONG (overstay >= 5 days)
Group 2: SHORT (overstay < 5 days)

The training set used for stage 2 was used for stage 3 with the following adjustments:

- cannot use ZZ/XW with overstay < 36 hrs (we don't know which of these patients belong to the cohort that needed home-care, stage 3 focuses on the home-care/nursing home patients)
- only use ZZ/XW >= 36 hours (we assume these patients needed home-care)
- only use NH patients with a transfer date (need overstay value to label discharge as LONG/NOT-LONG)
- patients discharged 5+ days were labeled LONG (2,304 admissions)
- patients discharged < 5 days were labeled NOT-LONG (1,527 admissions).
- two new fields are needed. The ADL for short NH overstay are high, and the ADL short ZZ overstay are low. To account for this in the regression model, we will use the quadratic term of the ADL center. The ADL center is ADL-average ADL. The average ADL of the training and test sets used in stage 2 (22,651 total admissions) is 12.97. Include the new fields ADL_center and ADL_center squared.

We wish to minimize the number of short overstay admissions (< 5 days) flagged and maximize the number of 10+ overstays flagged. Discharges from 5-9 days are not as important if flagged or not. If not flagged the overstay is manageable, if flagged then the overstay would be reduced.

Admissions with overstays of 5-9 days were omitted from the training set. The final training set has 3,139 admissions; 1,612 LONG (10+ days) and 1,527 NOT-LONG (0-4 days).

The Logistic Regression coefficients are:

```
AGE            0.0183
CVA           -0.047
Dementia       0.0674
CNSI_Any       0.4862
Stroke        -0.7647
BATH           0.1879
DRESS          0.132
TOILET         0.2423
TRANSFER       0.1678
CONTINENCE     0.1751
ADL center    -0.1386
ADL square    -0.0008
Intercept     -3.5055
```

For evaluation purposes, **patients that passed Stage 2 are used** to calculate % flagged, but only the ones that survived and had a transfer date are used to calculate the sens/spec/ppv.

A Logistic Regression threshold of 0.45 was used in Stage 2. From 8,095 test-set admissions, 2,973 passed the threshold of 0.45 (this includes expired and NH without transfer date).

To evaluate sens/spec/ppv, the following test subset was used. Of the patients with a transfer date (1,902 admissions), all patients discharged 5+ days were labeled LONG. All patients discharged < 5 days were labeled NOT-LONG.

Expired patients were labeled LONG. NH admissions without a transfer date were labeled LONG. Expired and no-transfer-date NH patients were not considered when calculating the overall sensitivity/specificity/ppv, but they were used to calculate the percentage of stage 3 patients flagged as long overstay.

Table 6 shows the sen/spec/ppv for stage 3 patients (only the 1,902 admissions) and percentage of stage 3 patients flagged as long overstay (all 2,973 admissions). It shows that the model will flag most stage 3 patients as long overstay. This tells us that using only physiological information if not enough to separate the long overstay patients from the short overstay. Socio, economic and discharge process information is also needed.

Table 6:

Build Model: Jan 2007 to July 2009 / Test Model: Aug 2009 to March 2012												
				total %	< 36 hours flagged			> 36 hours flagged			No Tdate	
	sen.	spec.	ppv	flagged	ZZ	XW	NH/CC	2-5 days	6-9 days	10+ days	NH/CC	expired
LONG vs. NOT-LONG LogReg (0.50)	79.5%	27.6%	41.7%	71.5%	13.8%	1.5%	4.7%	8.0%	5.0%	15.1%	5.3%	18.0%
LONG vs. NOT-LONG LogReg (0.525)	70.7%	39.9%	43.4%	60.7%	11.9%	1.3%	3.4%	6.7%	4.6%	13.2%	4.7%	15.0%
LONG vs. NOT-LONG LogReg (0.55)	61.5%	50.5%	44.9%	50.9%	9.7%	1.3%	2.7%	5.5%	4.0%	11.6%	3.7%	12.5%

Table 7 gives a better breakdown and shows that with Logistic Regression the majority of the 10+ overstay patients (70.8% for a threshold of 0.525) are flagged as RED.

Table 7:

Build Model: Jan 2007 to July 2009 / Test Model: Aug 2009 to March 2012																
OVERSTAY	< 36 hours			2-5 Days			6-9 Days			10+ Days			No Tdate	
LONG vs. NOT-LONG LogReg (0.50)	ZZ	XW	NH/CC	ZZ	XW	NH/CC	ZZ	XW	NH/CC	ZZ	XW	NH/CC	NH/CC	Expired
Predicted LONG	**410**	**45**	**141**	**186**	**16**	**36**	**151**	**9**	**31**	**151**	**27**	**270**	**159**	**536**
Predicted NOT-LONG	**120**	**35**	**92**	**43**	**5**	**23**	**43**	**4**	**19**	**43**	**5**	**59**	**77**	**299**
% Predicted LONG	**77.4%**	**56.3%**	**60.5%**	**81.2%**	**76.2%**	**61.0%**	**77.8%**	**69.2%**	**62.0%**	**77.8%**	**84.4%**	**82.1%**	**67.4%**	**64.2%**
	74.6%		**60.5%**	**77.0%**			**74.3%**			**80.7%**			**67.4%**	**64.2%**
LONG vs. NOT-LONG LogReg (0.525)	ZZ	XW	NH/CC	ZZ	XW	NH/CC	ZZ	XW	NH/CC	ZZ	XW	NH/CC	NH/CC	Expired
Predicted LONG	**353**	**39**	**100**	**156**	**15**	**29**	**133**	**9**	**27**	**133**	**23**	**237**	**139**	**445**
Predicted NOT-LONG	**177**	**41**	**133**	**73**	**6**	**30**	**61**	**4**	**23**	**61**	**9**	**92**	**97**	**390**
% Predicted LONG	**66.6%**	**48.8%**	**42.9%**	**68.1%**	**71.4%**	**49.2%**	**68.6%**	**69.2%**	**54.0%**	**68.6%**	**71.9%**	**72.0%**	**58.9%**	**53.3%**
	64.3%		**42.9%**	**64.7%**			**65.8%**			**70.8%**			**58.9%**	**53.3%**
LONG vs. NOT-LONG LogReg (0.55)	ZZ	XW	NH/CC	ZZ	XW	NH/CC	ZZ	XW	NH/CC	ZZ	XW	NH/CC	NH/CC	Expired
Predicted LONG	**288**	**38**	**81**	**127**	**13**	**23**	**120**	**8**	**22**	**120**	**23**	**203**	**109**	**371**
Predicted NOT-LONG	**242**	**42**	**152**	**102**	**8**	**36**	**74**	**9**	**28**	**74**	**9**	**126**	**127**	**464**
% Predicted LONG	**54.3%**	**47.5%**	**34.8%**	**55.5%**	**61.9%**	**39.0%**	**61.9%**	**47.1%**	**44.0%**	**61.9%**	**71.9%**	**61.7%**	**46.2%**	**44.4%**
	53.4%		**34.8%**	**52.8%**			**57.5%**			**62.3%**			**46.2%**	**44.4%**

We are most interested in identifying the 10+ overstay patients without flagging too many patients for special attention. The overall performance for the 8,476 admissions tested with the multi-stage predictive model (a patient going through stages 1 to 3) is given on Table 8. Adjusting the stage 2 and stage 3 thresholds will result in different spec/sen/ppv and % flagged. During the pilot study, the thresholds can be adjusted depending on the workload that is being generated by the model.

Table 8:

Build Model: Jan 2007 to July 2009 / Test Model: Aug 2009 to March 2012				
	sen.	spec.	ppv	total % flagged
10+ Days vs. NOT 10+ Days LogReg (stage 2 -> 0.45; stage 3 -> 0.50)	58.8%	86.1%	26.1%	25.1%
10+ Days vs. NOT 10+ Days LogReg (stage 2 -> 0.45; stage 3 -> 0.525)	51.6%	88.0%	32.2%	21.3%
10+ Days vs. NOT 10+ Days LogReg (stage 2 -> 0.45; stage 3 -> 0.55)	45.4%	89.9%	33.4%	17.9%