Chaire de recherche en **évaluation**
économique des **programmes publics**

Using Machine Learning to Predict Nosocomial Infections and Medical Accidents in a NICU

Marc **Beltempo**
Georges **Bresson**
Guy **Lacroix**

Using Machine Learning to Predict Nosocomial Infections and Medical Accidents in a NICU

Frail newborns admitted to a NICU are at high risk of contracting a nosocomial infection and are vulnerable to medical incidents due to the intensity of care they require. In this paper, we use a generalized mixed-effects regression tree model with random effects (GMERT-RI) to elaborate predictions trees for the two outcomes. GMERT-RI is an extension of the standard tree-based methods which is suitable for binary unbalanced panel data with random effects. These account for neonate-specific unobserved variables that may impact their outcomes. The sample includes 4,356 neonates admitted in the CHU de Québec Level-III NICU between 10 April 2008 and 28 March 2013. The algorithm is fed numerous neonates- and NICU-specific daily predictors and manages to unearth non-linear relationships between them as well as identify critical predictor variables thresholds. Interestingly, in addition to usual clinical predictors, important ones are within management's realm. Hence, machine learning algorithms such as GMERT-RI may prove to be a useful decision tool for management and pediatricians alike.

Using Machine Learning to Predict Nosocomial Infections and Medical Accidents in a NICU

Marc Beltempo · **Georges Bresson** · **Guy Lacroix**

Abstract Frail newborns admitted to a NICU are at high risk of contracting a nosocomial infection and are vulnerable to medical incidents due to the intensity of care they require. In this paper, we use a generalized mixed-effects regression tree model with random effects (GMERT-RI) to elaborate predictions trees for the two outcomes. GMERT-RI is an extension of the standard tree-based methods which is suitable for binary unbalanced panel data with random effects. These account for neonate-specific unobserved variables that may impact their outcomes. The sample includes 4,356 neonates admitted in the CHU de Québec Level-III NICU between 10 April 2008 and 28 March 2013. The algorithm is fed numerous neonates- and NICU-specific daily predictors and manages to unearth non-linear relationships between them as well as identify critical predictor variables thresholds. Interestingly, in addition to usual clinical predictors, important ones are within management's realm. Hence, machine learning algorithms such as GMERT-RI may prove to be a useful decision tool for management and pediatricians alike.

Keywords: Generalized mixed-effects regression trees, unbalanced panel data, NICU, nosocomial infections, medical incidents.

JEL: I10, J13, C23

We are grateful to the participants of the research seminar at the Institut de Science Financière et d'Assurances (Institute of Financial Sciences and Insurance), Lyon, France, for their comments and remarks. The usual disclaimer applies.

Marc Beltempo
Department of Pediatrics, McGill University Health Centre, Montreal, QC, Canada E-mail: marc.beltempo@mcgill.ca

Georges Bresson
Department of Economics, Université Paris II, Paris, France E-mail: georges.bresson@u-paris2.fr

Guy Lacroix (Corresponding author)
Department of Economics, Université Laval, Québec Canada E-mail: Guy.Lacroix@ecn.ulaval.ca

Introduction

Neonatal intensive care units (NICUs) must contend with uncertain admissions and fluctuating patient mixes (Tucker et al. [22]). Frail newborns admitted to a NICU are at higher risk of contracting a nosocomial infection which results in increased morbidity and mortality, prolonged lengths of stay, and increased medical costs (Polin et al. [17]). And because they require more care, are more vulnerable to medical incidents such as erroneous medication administration or feeding and equipment malfunctioning (Beltempo et al. [2]). Understanding the mechanisms that lead to these adverse events may help reduce the private and societal costs of adverse events as well as contribute to their cost containment (Russell et al. [18]).

Workforce and resource allocation within a NICU is thus challenging. Management usually turns to nursing overtime to stave off temporary demand surges and to meet required nurse-to-patient ratios (Berney and Needleman [5], Beltempo et al. [3]). While such a policy may prove useful in the very short run, it may not be optimal either from the NICU's nor the neonates' perspectives. Indeed, mounting empirical evidence suggests that nursing overtime may have deleterious health effects due to fatigue and stress (Bae [1], Lin [16], Cimiotti et al. [7], Trinkoff et al. [21], Beltempo et al. [4]). In other words, the short-term cure may have longer-term adverse health (neonate) and financial (NICU) consequences.

Streamlining workload and identifying the most at-risk neonates would allow medical staff to focus on more patient-oriented tasks and allow for more effective use of their time. This paper thus seeks to provide management with a (decision) tool to help identify the circumstances which are likely to yield nosocomial infections or lead to medical incidents. Yet, numerous factors must be considered at once which are both neonate- and NICU-specific. Traditional multivariate regression models, while useful for measuring the marginal impact of a given factor on the conditional expectation of a specific outcome, usually exhibit poor predictive performances. Alternatively, machine learning (ML) is particularly well-suited for the task at hand and is fast permeating into much research on pediatric care, from improving workforce efficiency to improving quality of care (*e.g.*, Clarke et al. [8]).[1]

Our analysis is based upon a generalized mixed-effects regression tree algorithm with random effects (GMERT-RI, Hajjem et al. [12]). This method is at the crossroads between standard econometrics and ML. While the two approaches were developed separately during the 1970s-2000s, they have both benefited from cross-fertilization of methodological innovations (see Hsiao [14], Bresson [6]). GMERT-RI combines a generalized linear mixed model (GLMM) with a mixed-effects regression tree (MERT). It accommodates fixed and random effects as well as unbalanced clusters (neonates). As shown by Hajjem et al. [12], the GMERT method replaces the linear structure used to model the fixed effects component in the GLMM's linear predictor with a regression tree structure, while the random component is still represented through a linear structure as in GLMMs. Basically, it is a penalized quasi-likelihood algorithm used to fit GLMMs where the weighted linear mixed-effects pseudo-model is replaced by a weighted MERT pseudo-model.[2]

The GMERT-RI model is suitable for binary outcomes and count data but has not yet been widely used as a machine learning tool in health applications (see Schultz et al. [19]). This algorithm incorporates observation-level covariates and their potential random effects (Hajjem et al. [12]) and allows

[1] ML has been used for identification of disease onset, classification of disease severity, predicting epileptic seizures, . . . It is fast becoming a hybrid physician-support tools thanks to the vast amount of data generated in healthcare systems. Although machine learning can prove a powerful tool, there is potential for misuse; model performance can be inflated through overfitting and, consequently, will not generalize to the greater population. But a number of recent methods – including the one we use – have been proposed to expand the applicability of machine learning tools and ensure robustness of results for within-subject factors and random effects (see Schultz et al. [19]).

[2] The logistic regression is often used as a baseline model with which to gauge more sophisticated machine learning approaches. It is often appropriate for clinical outcomes when using a small set of variables (see Gao et al. [10]). More sophisticated regularized variants of the logistic regression (*e.g.*, lasso-regularized and ridge logistic regressions) allow to remove uninformative variables and/or identify near-linear relationships between some subsets (see Tibshirani [20]). Yet, the main disadvantage of logistic regression is that it may require large sample sizes to achieve reliable performance, particularly in the presence of high-dimensional variable sets (see Schultz et al. [19]).

observations within clusters to be split when growing trees. This is an important feature for the prediction performance. As stressed by Schultz et al. [19], ML approaches that can account for random effects and nested designs are a promising avenue for more reliable outcomes, particularly so given the high degree of noise in patient data and the fact that these are often collected longitudinally.

In this paper, we focus on the CHU de Quebec NICU, a tertiary referral centre with a 51-bed capacity that tends to a population of 1.7 million over a territory of 452,600 km^2 in Canada. We study the occurrences of health care associated infections and medical incidents/accidents (henceforth HCAI and MA, respectively) among all neonates admitted to the NICU between April 2008 and March 2013. Daily exposure to overtime and regular hours of work, as well as numerous individual and NICU-specific covariates are used to predict the onset of the latter two outcomes.

Data and Clinical Environment

The CHU de Quebec NICU is a Level-III referral center. At the time of the study, there were three daytime board-certified neonatologists each assigned to a different group of patients in the NICU. Two teams, each composed of a staff neonatologist and at least one resident or fellow, managed 18 beds each (12 level-3 and 6 level-2). The other 15 level-2 beds were managed by the third neonatologist. During the weekends, the daytime staff comprised two neonatologists and a resident and/or a fellow. At night, an in-house resident and a consulting neonatologist attended every delivery of at-risk infants. This workforce arrangement prevailed during the entire study period. A total of 165 registered nurses were employed in the NICU, of which 68% (n=112) worked 8-hour shifts and 32% (n = 53) worked on 12-hour shifts. Nurse staffing was determined before each shift or whenever there was a change in the nurse-to-patient ratios or patient acuity. The required number of nurses varied according to patient acuity, planned admissions, and electives procedures/tests. When nurses were deemed in shortage, management initially turned to available off-duty nurses. Next, a pool of floating nurses was relied upon. Finally, it resorted to voluntary and mandatory overtime if necessary.

Overtime is defined as all hours worked beyond the regular work schedule (Fédération Interprofessionnelle de la Santé du Québec [9]).[3] Overtime may also occur if a nurse worked above 37.5 hours per week. Daily nursing overtime was computed starting at midnight each day. Total daily hours worked was computed similarly. Patient characteristics were collected using the hospital clinical database Med-Echo. It included gestational age, birth weight, sex, Apgar score, multiple pregnancies and type of delivery. Daily administrative data on overtime and regular hours of work, daily patient census and number of admissions were collected using the local administrative database Logibec. Information on HCAI was collected using the local infectious disease database TDR. Finally, information on MA was retrieved from the Gesrisk database.[4]

All newborns admitted during the study period were included conditional on having spent at least three days in the NICU. If an infant had more than one episode of bacteremia, these were considered separate events if they occurred more than 14 days apart. The date of the infection was determined as that at which the blood culture was obtained.

Descriptive Statistics

The total number of infants admitted in the NICU over our sample period is 5,466. Of those, 1,110 were omitted since their stay was shorter than three days. The final sample thus includes 4,356 neonates

[3] This occurred whenever a nurse either started her shift earlier than planned or finished later than scheduled. Working beyond 16 consecutive hours per day was prohibited.

[4] Reporting the information on the timing as well as the type of MA is mandatory.

Table 1: Sample Means

Variable	Accident Infection/ No	Accident Yes	Infection Yes
	NEONATES		
Sex (Female=1)(%)	43.84	48.13	43.19
Gestational Age	35.47	32.31	29.81
Weight (Grams)	2556.70	1962.64	1457.69
Apgar $>$ 7 at 5 Min. (%)	88.89	71.78	65.73
C-Section (%)	40.19	54.77	62.91
NICU *vs* Transfer (%)	73.74	78.84	85.45
Twins (%)	15.63	19.91	25.35
Surgical (1)/Medical(0) (%)	6.35	31.95	32.86
DRG Severity Index (1 – 4)			
Low	24.84	5.81	0.94
Medium	42.32	18.26	11.74
High	25.43	33.20	32.86
Very High	7.38	42.74	54.46
Length of stay (Days)	17.54	55.96	70.87
Overtime Reform (%)	16.42	21.83	15.32
	NUMBER OF NEONATES/EVENTS		
Number of Events	3,889	390	272
Neonate Frequency (/4,356)(%)		8.09	6.24
Daily Frequency (/1,822)(%)		21.98	15.03
Daily/Infant Frequency (/85,223) (%)		0.457	0.319
	NICU (1,822 DAYS)		
Daily Admissions	4.32	4.49	4.47
Bed Occupancy	50.19	50.71	50.58
Daily Regular Hours	514.14	538.03	517.13
Daily Overtime Hours	20.80	25.78	26.74

and represents over 85,223 infant/days over the sample period. Table 1 provides the means of the main variables used in the models. Infants who contracted an infection or suffered a medical accident had either, or both, a lower gestational age and birth weight, were fewer to exhibit an Apgar score above 7, were more likely to have been delivered by C-Section, and to have been admitted in the NICU at birth. The next two lines focus on the Diagnosis Related Group (DRG) at admission. Our data contain 113 distinct DRG codes. These are first categorized as surgical or medical. Infants who experienced a MA or a HCAI were much more likely to have been admitted following a surgical intervention. The next line reports the severity index of the DRG code which takes values between 1 and 4 (1=mild, 2=moderate, 3=severe, 4=extreme). Not surprisingly, accident and infection-free neonates had a low or moderately severe condition at entry. Those who did experience a MA or a HCAI were deemed severely or extremely ill at admission. Their poor health translated into lengthy hospitalization spells, particularly so for those who contracted a nosocomial infection.

In June 2012, management implemented a series interventions aimed at reducing overtime which we refer to as *Overtime Reform* in what follows. In particular, it hired 15 full-time registered nurses and converted 10% of existing positions from 8-hour shifts to 12-hour shifts (Beltempo et al. [3]). Table 1 shows that 16.42% of infection and accident-free and 15.32% of infected neonates were admitted after the overtime reform had been implemented. On the other hand, proportionately more neonates suffered a medical incident in the post-implementation period (21.83%).

The table also shows that approximately 8.1% (resp. 6.2%) of neonates were victim of a MA (resp. HCAI). From the NICU's perspective, the probability of observing a MA or a HCAI in any given day was 21.9% and 15.0%, respectively. This translates into 0.46% and 0.32% when computed daily and per neonate. The average length of stay is also considerably longer for neonates with either a MA or HCAI.

The table also shows that average daily admissions and occupancy do not vary much across outcomes. On the other hand, regular and overtime hours of work are positively related to the occurrence of both outcomes.

The relationship between hours of work, occupancy and outcomes is further investigated in Figure 1 in which we depict the (smoothed) daily variations in nursing hours, occupancy as well as total total daily of HCAIs and MAs. The figure highlights the inverse relationship between regular hours of work and overtime. Indeed, peaks and troughs almost always mirror one another. The third panel depicts the variations in occupancy rates across time. Occupancy rates are expressed relative to capacity (51 beds). Over our sample window, the NICU operated marginally above capacity 41% of the time (anywhere between 35 to 58 filled beds out of 51). Yet this occurred more frequently after the change in the overtime regime implemented in June 2012. Indeed, prior to the implementation of the overtime reform, the NICU operated above capacity 35% of the time. The proportion increased to 71% in the aftermath. Finally, the last panel depicts the daily total occurrences of HCAIs and MAs. As mentioned above, an MA or a HCAI occurred on average each 3^{rd} day, although this is not readily apparent from the figure due to the compression of the time scale. The figure nevertheless underlines the fact that their occurrences fluctuate considerably from day to day. From management's point of view, thus, linking the outcomes variables of the last panel to the variables depicted above may be quite (too) challenging. Yet, nursing hours, admission policy, and thus unit occupancy, are the main policy variables upon which it may exercise discretionary decisions. Management's task is further complicated by the fact that neonates' characteristics must also be accounted in order to minimize the occurrences of MAs and HCAIs. Finally, it is highly likely that the links between outcomes, neonates' characteristics and work arrangements are fairly complex and non-linear. Fortunately, machine learning is particularly well-suited to analyze such complex interactions.

Machine Learning

Classification and regression trees (CART) are machine-learning methods used to construct decision trees. A decision tree is a is flowchart-like predictive model in which leaves represent classifications (outcomes), non-leaf nodes are predictors, and branches represent conjunctions of features that lead to the classifications. CART models involve selecting the best predictors and determining appropriate split points (nodes) in each of them.

The data at our disposal present a novel and interesting feature. Indeed, the outcomes (MA/HCAI) are observed daily at the neonate level, whereas work arrangements and NICU characteristics are observed daily. Since the time spent in the NICU varies across neonates, our panel dataset is unbalanced. Traditional CART models are ill-suited to analyze such data. Fortunately, Hajjem et al. [12] have proposed the Generalized Mixed-Effects Regression Tree with random effects (GMERT-RI) which is suitable for unbalanced binary outcomes such as ours. In addition, GMERT-RI allows the inclusion of random effects to account for neonate-specific unobserved variables that may impact their outcomes. As with most machine learning methods, the GMERT-RI algorithm splits the sample into three subsets: a learning set, a validation set and a test set. The appropriate number of leaves is determined by cross-validation. We thus use GMERT-RI to predict the binary outcome y_{it} (*MA or HCAI*) of neonate i during its t-th day in the NICU, conditional on a series of predictors.

As mentioned previously, Hajjem et al. [11] have proposed a mixed-effects regression tree (MERT) method. It can appropriately deal with the possible random effects of observation-level covariates and can split observations within clusters since observation-level covariates are candidates in the splitting process. The main idea is to fit a tree after removing the random effects part of the model, update the estimates (or predictions) of the random effect and cycle until convergence. Unfortunately, MERT is designed for Gaussian response data only. More recently, Hajjem et al. [12] have proposed a GMERT-RI, which is suitable for non-Gaussian data (e.g., binary outcomes and count data). Following the steps of the generalized linear mixed models (GLMMs), the GMERT-RI method can handle unbalanced clusters, and can incorporate observation-level covariates and their potential random effects. It allows observations within clusters to be split. For an unbalanced panel data set of neonates ($i = 1, \ldots, N$) observed during T_i days ($t = 1, \ldots, T_i$), let y_{it} be the binary outcome variable and X_{it} the $(1 \times K)$ vector of predictors. The GMERT-RI method incorporates random individual-specific effects and starts with a logistic-mixed model

$$y \mid \beta, u \sim \text{Bernoulli}\left(\text{logit}^{-1}(X\beta + Zu)\right),$$

where the notation $y \sim \text{Bernoulli}(p)$ is shorthand for the entries of y having independent Bernoulli distributions with parameters corresponding to the entries of p and $\text{logit}^{-1}(x)$ is shorthand for the logistic distribution $e^x/(1+e^x)$. y is the $(T \times 1)$ vector of binary outcomes, X is an $(T \times K)$ matrix of covariates, Z is an $(T \times NK)$ block-diagonal matrix of the X_i submatrices where $T = \sum_{i=1}^{N} T_i$. X and Z are called the fixed and random effects design matrices associated with β and u, the $(K \times 1)$ fixed effects and $(NK \times 1)$ random effects vectors. $X_{it,1}$ is the intercept and $X_{it,j}$, $2 \leq j \leq K$ are the other control covariates. The random intercept is defined by the sum $(\beta_1 + u_{i,1})$, the random slope for variable $X_{i,2}$ is the sum $(\beta_2 + u_{i,2})$, etc.

In our specific case, we only consider random intercept model (*i.e.*, $u_{i,j} = 0$, $\forall j > 1$). Hence, Z is restricted to an $(T \times N)$ block-diagonal matrix of N subvectors $(T \times 1)$ of ones. The GMERT-RI thus simplifies to a generalized mixed-effects regression tree with random intercept (GMERT-RI). The linear fixed part $X\beta$ is replaced by a function $f(X)$ which is estimated using a standard regression tree design.[5] The overall data set is divided into three subsets: a learning subset, a validation subset and a test subset (corresponding respectively to 40%, 40% and 20% of the initial dataset).

A standard tree design — that does not account for random specific effects — is generally adjusted using the dataset resulting from merging the learning and validation subsets. To choose the right number of leaves, one proceeds by cross validation. We estimate the tree's performance by 10-block cross-validation for each level of relevant simplification. The complexity of the decision tree is defined as the number of splits in the tree and the complexity parameter (CP) is used to control the size of the decision tree and to select the optimal tree size. A good choice of CP for pruning the tree is often the leftmost value for which the cross-validation error (*i.e.*, the rate of misclassifications relative to the original score) lies below the "horizontal line".[6] Here, GMERT-RI is adjusted using the learning subset, and then validated using the validation subset, *i.e.* the best GMERT-RI is selected based on minimum misclassification rate (MCR) observed on the validation subset.[7] The misclassification rate depends on the cutpoint value used to classify the observations, specifically when the data have a nested structure — as in our case — with clusters having different sizes in the training and the test data sets. The cutpoint is selected as the value such that, in the training set, the proportion of observations assigned to infections (or accidents) is closest to the actual proportion of infections (or accidents).[8]

To select variables and reduce dimensionality, we can rank the predictors by some measure of importance and remove variables with low rank. Variable importance (VIMP) was originally defined using a

[5] The supplementary material provides additional details on the GMERT-RI algorithm of Hajjem et al. [12]. The R codes are available in the supplementary material of Hajjem et al. [12]

[6] It represents the highest cross validation error less than the sum of the minimum cross validation error and the standard deviation of the error on that tree.

[7] The misclassification rate (MCR) is given by $MCR = \left(\sum_{i=1}^{N^{(v)}} \sum_{t=1}^{T_i^{(v)}} \mid y_{it} - \widehat{y}_{it} \mid\right) / T^{(v)}$ where $\widehat{y}_{it}$ is the predicted class of observation t in cluster i: $\widehat{y}_{it} = \text{Bernoulli}(\widehat{\mu}_{it})$ with $\widehat{\mu}_{it} = \left(1 + \exp\left(-\widehat{f}(X'_{it}) - Z'_{it}\widehat{u}_i\right)\right)^{-1}$. $\widehat{f}(X'_{it})$ is the predicted fixed component that results from the tree and $Z'_{it}\widehat{u}_i$ is its predicted random part corresponding to its cluster. $N^{(v)}$ is the number of clusters in the validation set, $T_i^{(v)}$ is the size of cluster i and $T^{(v)}$ is the total number of observations in the validation set.

[8] In a Monte Carlo simulation study with random effects, Hajjem et al. [12] have shown that the mixed-effects classification trees give better results than the usual classification trees even with a misspecified random component part.

measure involving surrogate variables (see for instance Hastie et al. [13]). VIMP is calculated for each variable individually and its value is calculated as the sum of the decrease in impurity.[9] It counts both when the variable appear as a primary split and when it appears as a surrogate. The relative importance of a given variable is a number between 0 and 100 and corresponds to the its VIMP divided by the maximum VIMP among all the variables.

Once the optimal tree structure is defined, and conditional on the different cutoffs, the true (resp. false) positive rate (sensitivity (resp. specificity)), the false positive (negative) rates, *etc.* are estimated. Using these, we can compute the usual goodness-of-fit performance measures such as ROC curve and associated AUC, sensitivity/specificity curve, *etc.*[10]

Results

We analyze the two outcomes (MA and HCAI) separately. The GMERT-RI algorithm is fed the following set of neonate-specific characteristics: `Sex`, `Gestational Age`, `Weight`, `Apgar < 7`, `C-section`, `Admission at birth`, `Twins`, `Surgical/Medical Birth`, `Severity of Illness Index (1-4)`. (`NICU characteristics) Occupancy`, `# admissions`, `Reform` dummy if event occurs after June 2012. The analysis of MA further includes daily regular and overtime hours in the NICU as well as a dummy variable equal to one if the infant incurred a HCAI prior to the MA. The analysis of HCAI assumes the infection occurred when results from the blood culture were obtained. Although there is no consensus on the lead time and sequence of events that may trigger the onset of a HCAI, we include the daily prior three-day moving average number of regular and overtime hours in the analysis (`Regular.Hours.MA3, Overtime.Hours.MA3`) (see Hugonnet et al. [15], Polin et al. [17]). Finally, we include a dummy variable to account for the occurrence of a MA prior to the HCAI.

It it conceivable that individual characteristics that are unfavourable to the occurrence of a MA/HCAI may be compensated for by detrimental environmental factors, and *vice versa*. In addition, the levels at which the favourable/detrimental factors operate are not know *a priori* and may in fact interact in a highly non-linear fashion. Figures 2a and 2b report the main variables found to influence the occurrence of both outcomes. The predictors are arranged in descending order of importance. While GMERT-RI includes as many as 15 predictors in each outcome, the algorithm retains only eleven of them when predicting either a MA or a HCAI. Importantly from a policy perspective, there are almost as many NICU-related factors as there are individual ones that matter in predicting the occurrence of either outcome. In fact, regular and overtime hours of work are two of the three most influential predictors. Regular hours ranks first in predicting MA and second for HCAI. Not surprisingly, the algorithm identifies birth weight and gestational age as the main infant-specific drivers of the two outcomes. Predictors that have little predictive power are omitted in the pruning process as stressed above and do not appear in the figures. The average relative variable importance is equal to 25.34% in the MA model (dashed line) and 25.78% in the HCAI model. Figures 2a and 2b identify the main predictor variables in the entire MA and HCAI trees. Although `regular hours` and `overtime hours` stand out as the main variables, it does not necessarily follow that the main split nodes occur along these predictors. We investigate this issue further by focusing on the prediction trees for MA and HCAI as exhibited in Figures 3 and 4, respectively. The number inside each node corresponds to the predicted probability of the outcome.[11] The number of observations (n) in

[9] Gini impurity is a measure of how often a randomly chosen element from the set would be incorrectly labeled if it was randomly labeled according to the distribution of labels in the subset. The Gini impurity can be computed by summing the probability p_i of an item with label i being chosen times the probability $(1-p_i)$ of a mistake in categorizing that item. To compute Gini impurity I_G for a set of items with J classes, suppose $i \in \{1,2,...,J\}$ and let p_i be the fraction of items labeled with class i, then $I_G = 1 - \sum_{i=1}^{J} p_i^2$. In our case, $J = 2$ for accident (1) and no accident (0).

[10] We use the recent R package `ROCR` which allows to create cutoff-parameterized 2D performance curves by freely combining any two from over 25 performance measures.

[11] To save on space, probabilities smaller than 1% appear as "0" and those above 99% appear as "1" inside the nodes.

appears just below the node. Finally, the splitting values appear in bold characters below the nodes. To ease reading, paths that lead to high probabilities of adverse outcomes are emphasized (> 75%).

The MA tree is illustrated in Figure 3. It contains 40 leaves.[12] The primary split occurs with respect to the DRG *Severity Index*. Thus neonates who were deemed as having a "High" severity index or below at admission appear on the main left-hand side branch. Those with a "Very High" index are located on the right-hand side. The second split involves Surgical/Medical interventions (left) and prior infection (right). Thus the first two levels involve neonate-specific factors. The ensuing paths involve a mixture of neonates- and NICU-specific factors. The tree contains six left-hand paths which likely lead to a MA and as many as seven on the right-hand side.

Decision trees such as Figure 3 are easy to interpret, understand and visualize. The pathways that lead to the highest probability of accidents identify critical thresholds and combinations of predictor variables that must be avoided to the extent possible in order to minimize the occurrences of MAs. For instance, out of a total of 85,223 infant/days, the primary split between very high (right) and high or below (left) severity index allows to discriminate between 126 accidents on the left-hand-side with a probability of 100% and as many as 152 accidents on the right-hand-side with a probability greater than 97%.[13] NICU-specific factors on the left side of the decision tree include occupancy, regular hours and overtime hours. For example, 22 accidents occurred among neonates admitted for medical reasons, whose weight was above 645 g, while occupancy was below 57.5%, but while regular hours were above 609 hours, but whose gestational age was below 27.5 weeks. Likewise, 28 accidents occurred among those admitted for surgical reasons, whose weight was above 2195 g and whose gestational age was below 33.5 weeks. On the right-hand-side of the tree (very high severity index), occurrences of MAs can be predicted from several combinations of prior infections, regular and overtime hours of work, gestational age and number of daily admissions.

The HCAI tree is depicted in Figure 4. As show, it contains 34 leaves. The primary split occurs with respect to birth weight at a threshold of 1195 grams. The second level splits occur on the left-hand side with respect to the severity index (Low/Medium), and with respect of regular hours of work (threshold equal to 618) on the right-hand side. The tree identifies thirteen paths which likely lead to the onset of a nosocomial infection, seven on the left-hand side and six on the right-hand of the primary split, respectively. As with the previous figure, one simply needs to follow the 13 pathways to identify the neonates- and NICU-specific predictor variables and thresholds that best predict the onset of a HCAI. As expected, the clinical conditions of the neonates at admissions are undoubtedly important predictors. The GMERT-RI algorithm also clearly identifies the NICU-specific factors as important determinants. In addition, it manages to unearth the fact that the latter interact in a highly non-linear fashion with the former.[14]

GMERT-RI for infections has a predictive minimum classification rate (MCR) of 1.23% (resp. 1.05%) for the validation set (resp. the test set). These very low MCRs confirm the goodness-of-fit obtained with the GMERT-RI method when applied to rare events such as infections and accidents amongst neonates. The estimated area under the ROC curve (AUC) for accidents (resp. infection) is 0.66 (resp. 0.73) when computed for the whole set. With extremely rare events, though, these measures should be viewed with caution.[15]

While MA and HCAI are relatively rare events at the neonate/day level, the classification trees depicted in Figures 3 and 4 manage to identify the routes through which these are likely to occur. In addition,

[12] Figures 3 and 4 only give nodes and ancestors for predicted probabilities > 75%. The entire trees are given in the supplementary material.

[13] These numbers correspond to the sum of the observations in each cell of the terminal nodes, or leaves. On the left-hand-side this gives 126 = 22+13+37+13+13+28, while on the right-hand-side is corresponds to 152 = 20+22+33+17+28+17+15.

[14] Note that the model slightly overestimates the true number of infections, *i.e.* 372 instead of 272. On the other hand, if we focus on events with probabilities strictly larger than 80% then over-estimation is reduced significantly, *i.e.* from 372 to 289.

[15] The supplementary material provides additional comments on the ROC curves and various performance measures.

the split nodes and their thresholds values are sensical and provide useful guidance for management when designing policies to mitigate these deleterious outcomes. The GMERT-RI algorithm manages to unearth two important features of the data. First, surprisingly, it shows that institutional features are just as important drivers as neonate-specific medical conditions in predicting MA and HCAI, and hence within the grasp of management. Second, traditional statistical analyses are assuredly incapable of identifying these highly non-linear relations between these predictors and their threshold values.

Conclusion

Predicting health care associated infections or medical accidents in a NICU is a complex task. Recent machine learning algorithms are well-suited to unearth potential correlations as they now allow to account for unbalanced panel data and discrete health outcomes, two frequent features of clinical data. Prediction trees are now relatively widespread in empirical research and are integrated in many software suites (Polin et al. [17], Schultz et al. [19]). From an operational point of view, prediction trees can complement traditional management tools in preventing undesirable health outcomes in the NICU. In addition to permeating into much research on pediatric care, ML algorithms can also be used as hybrid physician-support tools. As stressed by Clarke et al. [8], while ML approaches have the potential to profoundly modify pediatric care from "bench to bedside", and in particular in the NICU, it is by no means a substitute to the clinical judgment of pediatricians.

References

1. Sung-Heui Bae. Presence of nurse mandatory overtime regulations and nurse and patient outcomes. *Nursing Economic$*, 31(2):59–89, 2013.
2. M Beltempo, G Lacroix, M Cabot, R Blais, and B Piedboeuf. Association of nursing overtime, nurse staffing and unit occupancy with medical incidents and outcomes of very preterm infants. *Journal Of Perinatology*, 38:175 EP –, 09 2017. URL `https://doi.org/10.1038/jp.2017.146`.
3. Marc Beltempo, Guy Lacroix, Michele Cabot, and Bruno Piedboeuf. Factors and costs associated with the use of registered nurse overtime in the neonatal intensive care unit. *Pediatrics and neonatal nursing Open Journal*, 4:17–23, 08 2016.
4. Marc Beltempo, Georges Bresson, Jean-Michel Étienne, and Guy Lacroix. Infections, accidents and nursing overtime in a neonatal intensive care unit. *The European Journal of Health Economics*, 2021.
5. B. Berney and J. Needleman. Trends in nurse overtime, 1995-2002. *Policy Polit Nurs Pract*, 6: 183–90, 2005.
6. Georges Bresson. Comments on "an Econometrician's perspective on Big Data" by Cheng Hsiao. In T. Li, M.H. Pesaran, and D. Terrell, editors, *Essays in Honor of Cheng Hsiao*, pages 431–443. Emerald Publishing Limited, 2020.
7. Jeannie P. Cimiotti, Linda H. Aiken, Douglas M. Sloane, and Evan S. Wu. Nurse staffing, burnout, and health care-associated infection. *American Journal of Infection Control*, 40(6):486–490, 2012.
8. Sarah LN Clarke, Kevon Parmesar, Moin A Saleem, and Athimalaipet V Ramanan. Future of machine learning in paediatrics. *Archives of Disease in Childhood*, pages 1–6, 2021.
9. Fédération Interprofessionnelle de la Santé du Québec. Convention collective 2011-2015, article 19.01. 2011.
10. Chao Gao, Hanbo Sun, Tuo Wang, Ming Tang, Nicolaas I Bohnen, Martijn LTM Müller, Talia Herman, Nir Giladi, Alexandr Kalinin, Cathie Spino, et al. Model-based and model-free machine learning techniques for diagnostic prediction and classification of clinical outcomes in parkinson's disease. *Scientific reports*, 8(1):1–21, 2018.

11. Ahlem Hajjem, François Bellavance, and Denis Larocque. Mixed-effects random forest for clustered data. *Journal of Statistical Computation and Simulation*, 84(6):1313–1328, 2014.
12. Ahlem Hajjem, Denis Larocque, and François Bellavance. Generalized mixed effects regression trees. *Statistics & Probability Letters*, 126:114 – 118, 2017.
13. T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Prediction, Inference and Data Mining*. Springer-Verlag, New York, 2009.
14. Cheng Hsiao. An Econometrician's perspective on Big Data. In T. Li, M.H. Pesaran, and D Terrell, editors, *Essays in Honor of Cheng Hsiao*, pages 413–423. Emerald Publishing Limited, 2020.
15. Stéphane Hugonnet, Jean-Claude Chevrolet, and Didier Pittet. The effect of workload on infection risk in critically ill patients. *Critical Care Medicine*, 35(1):76–81, 2007.
16. Haizhen Lin. Revisiting the relationship between nurse staffing and quality of care in nursing homes: An instrumental variables approach. *Journal of Health Economics*, 37:13 – 24, 2014.
17. Richard A. Polin, Susan Denson, and Michael T. Brady. Strategies for prevention of health care–associated infections in the NICU. *Pediatrics*, 129(4):e1085–e1093, 2012.
18. Rebecca B. Russell, Nancy S. Green, Claudia A. Steiner, Susan Meikle, Jennifer L. Howse, Karalee Poschman, Todd Dias, Lisa Potetz, Michael J. Davidoff, Karla Damus, and Joann R. Petrini. Cost of hospitalization for preterm and low birth weight infants in the United States. *Pediatrics*, 120(1):1–9, 2007.
19. Benjamin G Schultz, Zaher Joukhadar, Usha Nattala, Maria del Mar Quiroga, Francesca Bolk, and Adam P Vogel. Best practices for supervised machine learning when examining biomarkers in clinical populations. In Ahmed A Moustafa, editor, *Big Data in Psychiatry & Neurology*, pages 1–34. Elsevier, 2021.
20. Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.
21. Alison M. Trinkoff, Meg Johantgen, Carla L. Storr, Ayse P. Gurses, Yulan Liang, and Kihye Han. Nurses' work schedule characteristics, nurse staffing, and patient mortality. *Nursing Research*, 60 (1):1–8, 2011.
22. J. Tucker, W. Tarnow-Mordi, C. Gould, G. Parry, and N. Marlow. On behalf of the UK neonatal staffing study collaborative group. uk neonatal intensive care services in 1996. *Child Fetal Neonatal Ed*, 80:F233–34, 1999.

Fig. 1: Smoothed Daily Variations in Nursing Hours, Occupancy and # of Infections and Medical Incidents

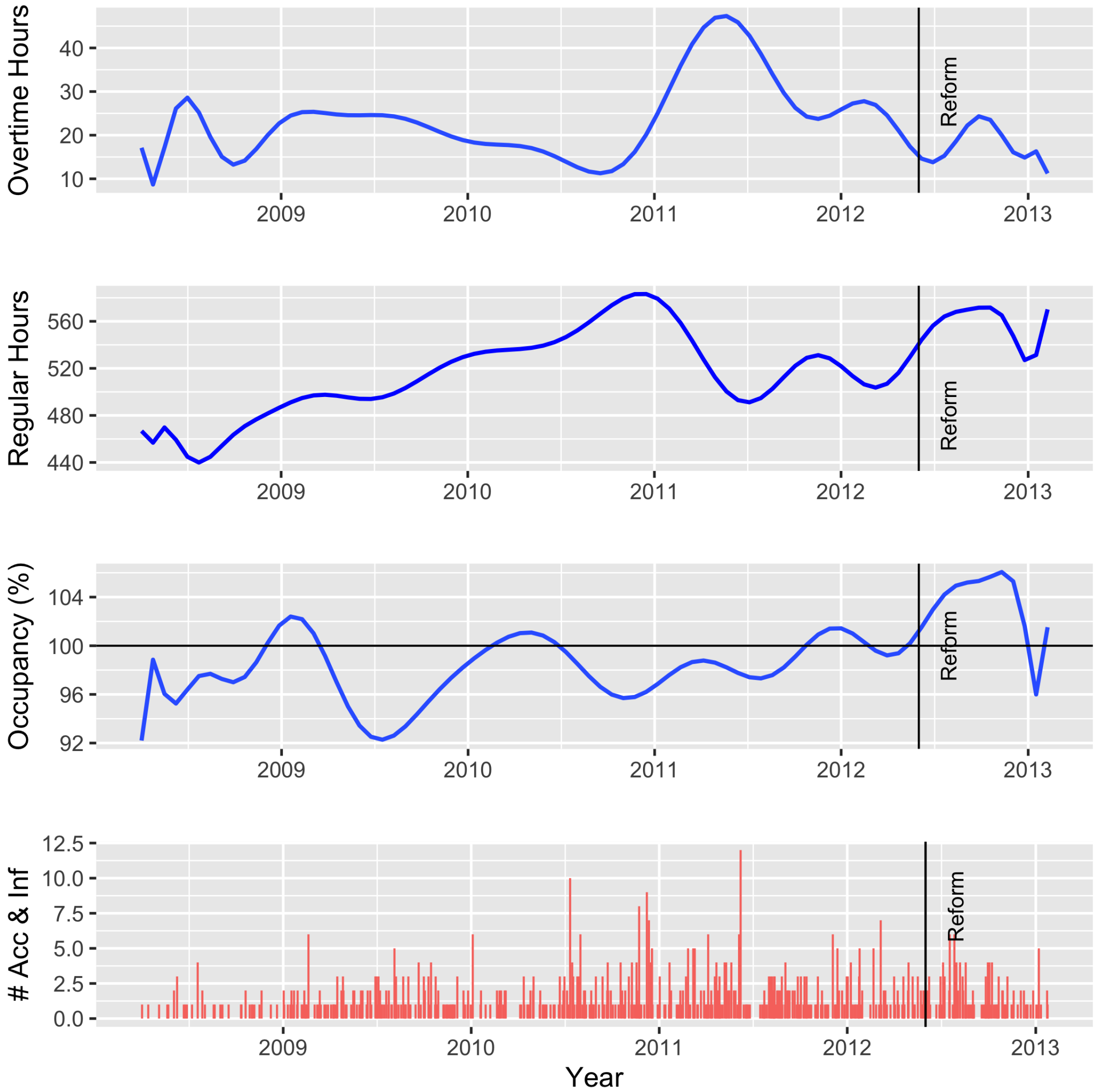
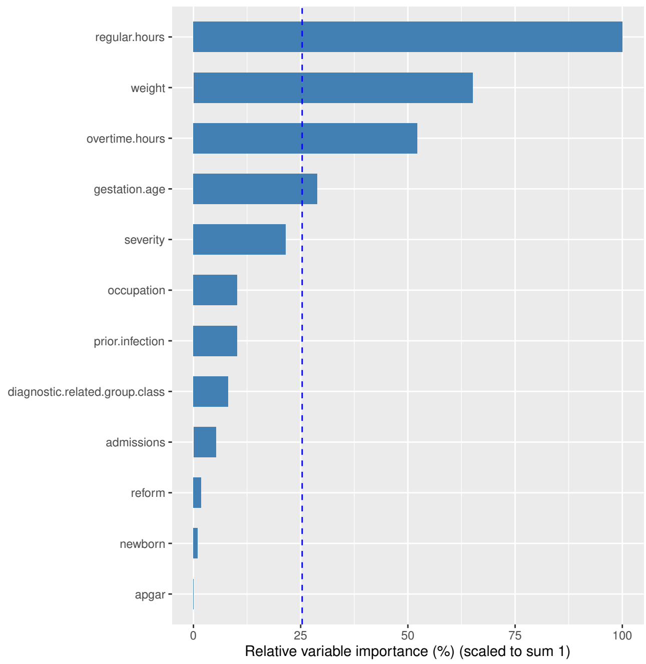
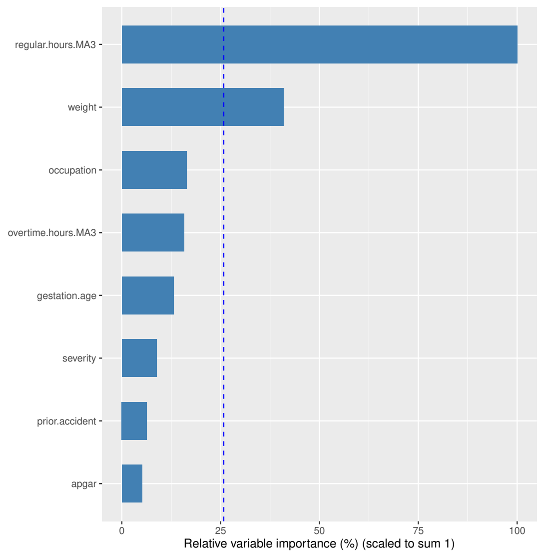


Fig. 2: Relative Variable importance (%)
Generalized Mixed-Effect Regression Trees



(a) Accidents



(b) Infections

Fig. 3: Generalized Mixed-Effects Regression Tree - Accidents

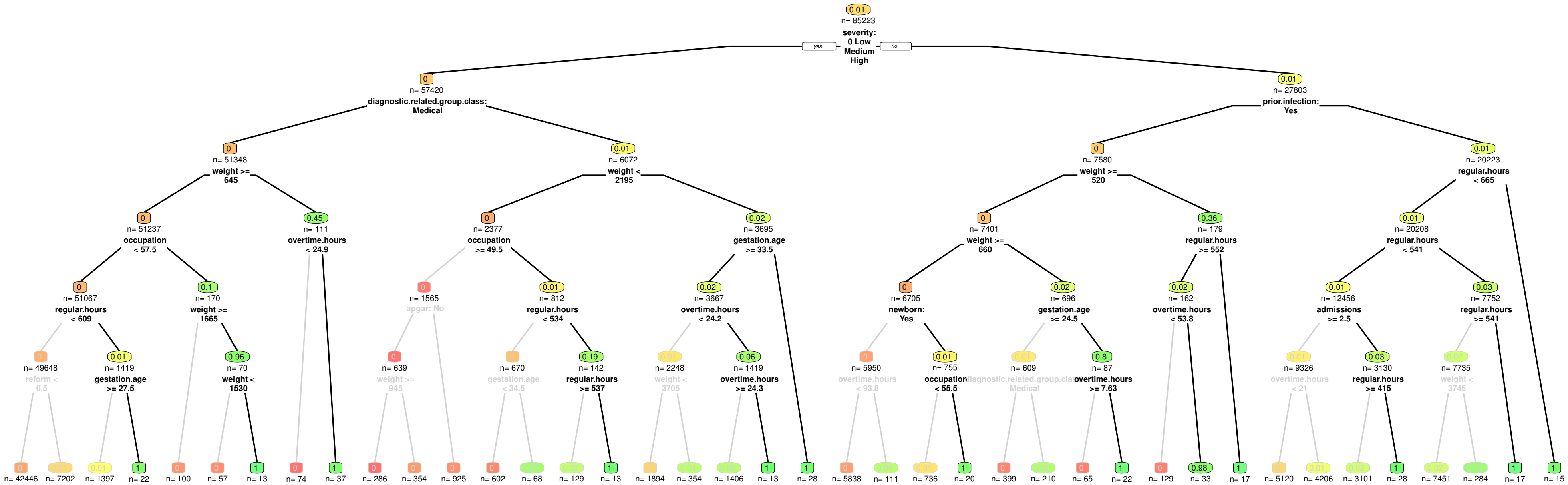
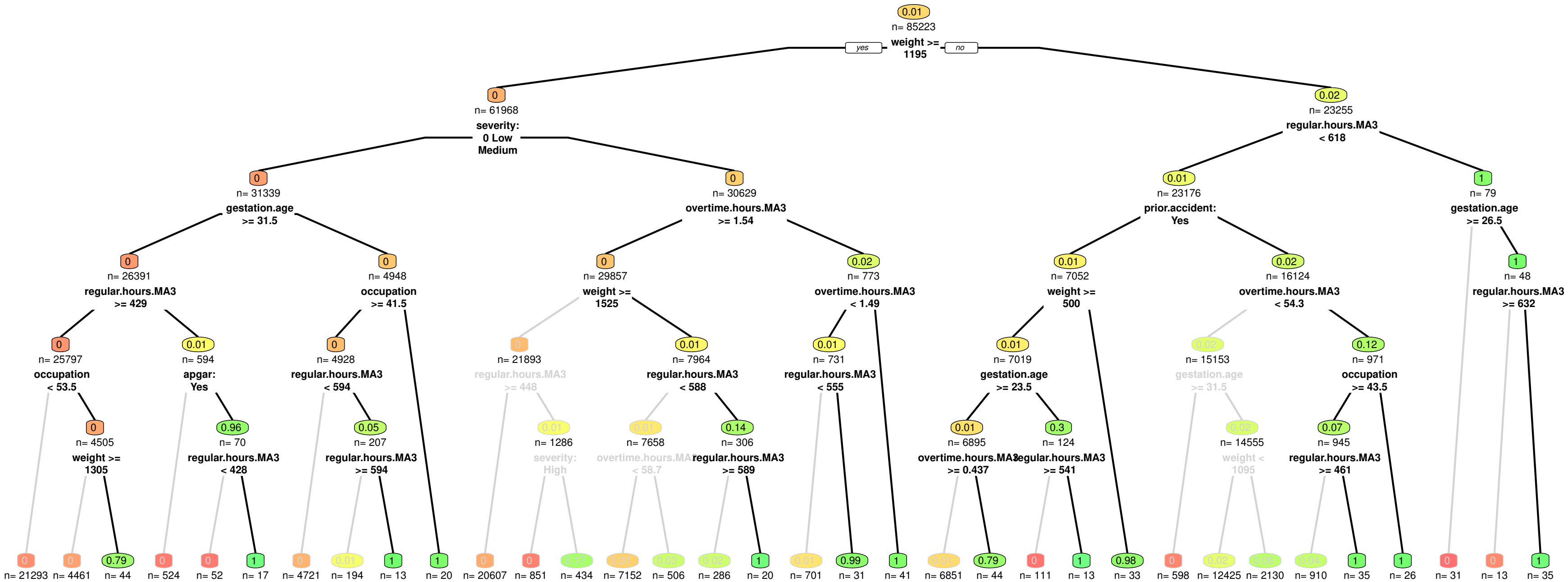


Fig. 4: Generalized Mixed-Effects Regression Tree - Infections



Using Machine Learning to Predict Nosocomial Infections and Medical Accidents in a NICU

Marc **Beltempo,** Department of Pediatrics, McGill University Health Centre, Montreal, QC, Canada
Georges **Bresson,** Département de science économique, Université Paris-II, France
Guy **Lacroix,** Département d'économique, Université Laval

We are grateful to the participants of the research seminar at the Institut de Science Financière et d'Assurances (Institute of Financial Sciences and Insurance), Lyon, France, for their comments and remarks. The usual disclaimer applies.

Chaire de recherche en **évaluation économique** des **programmes publics**

UNIVERSITÉ LAVAL

Secrétariat du Conseil du trésor Québec