Does My High Blood Pressure Improve Your Survival? Overall and Subgroup Learning Curves in Health*

Raf Van Gestel †1, Tobias Müller ‡2 and Johan Bosmans§ ¶ 3

[1]University of Antwerp
[2]University of Lucerne
[3]University of Antwerp

June 3, 2016

Abstract

Learning curves in health are of interest for a wide range of medical disciplines, for multiple types of healthcare providers and policy makers. In this paper, we distinguish between three types of learning when identifying overall learning curves: static learning, learning from cumulative experience and human capital depreciation. In addition, we approach the question of how treating more patients with specific characteristics improves provider performance. Information on the role of subgroups has the potential to better inform new or low outcome providers on how to improve. Statistically however, capturing all subgroup experiences in one analysis introduces strong collinearities among regressors. To soften collinearity problems, we explore the use of Lasso regression as a variable selection method and Theil-Goldberger mixed estimation to augment the available information. We use data from the Belgian Transcatheter Aorta Valve Implantation (TAVI) registry, containing information on the first 860 TAVI procedures in Belgium. Ultimately, we find evidence for both overall and subgroup learning effects: for 2-year survival, we find that the probability of survival is increased by about 0.16%-points for each additional patient treated. For adverse events like renal failure and stroke, we find that an extra day between procedures increases the probability for the these events by 0.12%-points and 0.07%-points respectively. These overall effects are then split into subgroup effects where we find evidence for positive learning effects from physicians' experience with defibrillation, hypertension and the use of certain types of replacement valves during the TAVI procedure.

Keywords: *Learning Curves, Lasso, Theil-Goldberger, TAVI*

JEL: I10, C11, C18

* *Acknowledgments*: We are grateful to Diana De Graeve, seminar participants at the Lowlands Health Economics Study group,...

† First author. Prinsstraat 13, 2000 Antwerpen, Belgium, email: raf.vangestel@uantwerpen.be
‡ Corresponding author. Frohburgstrasse 3, 6002 Luzern, Switzerland, email: tobias.mueller@unilu.ch
§ Corresponding author. Wilrijkstraat 10, 2650 Edegem, Belgium, email: johan.bosmans@uantwerpen.be
¶ On behalf of the Belgian TAVI Registry participants.

1 Introduction

The idea of a relationship between volume and performance of healthcare providers is well-known, but uncertainty remains about the source of this relationship. In this respect, the issue of learning by doing or cumulative learning is most frequently encountered in empirical research. Economic research additionally considers economies of scale, human capital depreciation, reverse causality, level of specialization and social learning or "learning by watching" as factors that drive volume-outcome relationships [Ho, 2002, Gaynor et al., 2005, Huesch, 2009, Hockenberry and Helmchen, 2014, Lee et al., 2015, Mesman et al., 2015]. Data collinearities however hamper inference so that theoretical arguments lead most studies to include a subset of these effects. Although unintended, this may lead to ill-guided scientifically inspired drastic policy measures.[1] Common policies are volume thresholds for hospitals, report cards and team/provider training [Huesch and Sakakibara, 2009]. In this paper, we analyze multiple factors simultaneously and we emphasize the potential role of patient subgroups in the learning process. This approach provides extra nuanced and conservative information on where improvements may be made by healthcare providers [Bridgewater et al., 2004].

In this paper we uncover effects from static learning, learning from cumulative experience and learning from recent experience. Cumulative experience refers to the number of patients that have been treated in the past, whereas learning from recent experience evaluates the role of time since the last procedure. For CABG (coronary artery bypass graft) and PTCA (percutaneous transluminal coronary angioplasty), cumulative learning has been found to only play a minor role [Ho, 2002, Gaynor et al., 2005]. In contrast, learning from recent experience, which is also termed human capital depreciation, seems to be important for CABG [Hockenberry and Helmchen, 2014]. Static learning or economies of scale are total volume effects; hospitals with more patients are likely to be better equipped and to have better standardized procedures. Economies of scale have been found to be important for both CABG and PTCA [Ho, 2002, Gaynor et al., 2005].

Following the literature, we first disentangle these overall learning effects. Subsequently we follow a data-driven approach to detect learning mechanisms in patient subgroups. Given that performance is affected by recent and cumulative experience, these effects may well be driven by patient subgroups. More specifically, we assess how treating more patients with certain subgroup characteristics influences overall health outcomes. Quantifying such information goes beyond the typical volume-outcome relationship and improves insight on how to improve health outcomes in lower volume, new or underperforming providers. The information allows to better identify the source of learning and to transfer relevant knowledge to health policy makers and practitioners. It stimulates to pay attention to certain subgroup specific complications and it provides more insights for less intuitive outcomes. For typical outcomes like in-hospital mortality, physicians may have a general feeling on how to improve performance. This is however much less the

[1] See the example of Huesch and Sakakibara [2009] in Bristol where surgical volumes were concentrated in one hospital without complementary care facilities in the same hospital. The resulting disastrous outcomes were fully contributed to learning effects while in fact the organization of care was to a large extent insufficient.

case for other outcome variables like long-term mortality or procedural characteristics. These results are a first step towards efficient knowledge transfer and it should then be investigated if these effects are due to: 1) a change in severity of characteristics over time, or 2) adjustment of physician practice due to these patients.

Existing studies mostly explore subgroup effects from a different point of view, namely how volume impacts on subgroup health. The difference is nuanced but important. In our setup, learning effects in the overall population can be attributed to certain subgroups while other studies look at which subsamples are influenced by volume. Whereas the former has the potential to help improve performance in new, underperforming or small providers, the latter informs policy makers on how to concentrate procedures in hospitals. Some examples related to subpopulation effects in trauma care are Matsushima et al. [2014] where larger volumes of geriatric patients were associated with lower mortality and complications among geriatric patients. Larger non-geriatric volumes were associated with higher odds of major complications. In Pasquale et al. [2001], higher-volume centers were more successful in treating patients in seven out of nine injury types [Caputo et al., 2014].

Our data covers all patients in Belgium that underwent a transcatether aorta valve implantation (TAVI) between 2007 and 2012 – the very first patient with TAVI in Belgium was treated in 2007. For TAVI, experience has been shown to have an impact on mortality (30-day and 1-year), duration of procedures, contrast volume and radiation [Möllmann et al., 2015, Alli et al., 2012, Kempfert et al., 2012]. Although these findings are based on descriptive analyses, they provide suggestive evidence. In this study, the overall learning effects are analyzed while controlling for a broad range of patient- and procedure-specific characteristics, as well as hospital fixed-effects.

Overall we find that different learning processes apply for different outcomes: while cumulative and static learning significantly affect 24-month and 36-month survival, learning from recent experience is significant for several Major Adverse Cardiac and Cerebrovascular Events (MACE). These events occur as a result of situations occurring during the procedure. In particular, our results suggest an increase in 24-month survival of 0.16%-points for every extra TAVI patient treated. Likewise, 36-month survival is increased by 0.30%-points. Furthermore, the likelihood of renal failure or a cerebrovascular stroke is increased by 0.12%-points and 0.07%-points respectively for every additional day since the last TAVI procedure.

While multicollinearity is a well-known statistical issue when attempting to identify overall learning curves, it is even more problematic for subgroup effects. By treating more patients overall, by definition also more patients are treated within specific subgroups (e.g. with renal failure or porcelain aorta). Therefore, another important contribution of this paper to the learning curves literature is that we propose two methodological approaches on how to disentangle these (subgroup) learning effects. In particular, we employ variable selection and data-augmenting methods: *firstly*, we single out relevant predictors for two-year mortality using Lasso regression [Tibshirani, 1996]. *Secondly*, in a Bayesian spirit, we apply Theil-Goldberger mixed estimation to add objective information to the model to soften multicollinearity problems [Theil and Goldberger, 1961]. Theil-Goldberger estimation allows the inclusion of prior information on a sum of coefficients which is central to identify subgroup learning effects. Subgroup learning effects for 24-month survival are found for patients with aortic aneurysm, atrial fibrillation, carotid disease, hypertension, porcelain aorta, NYHA category three and transfemoral access. That is, treating more patients with these characteristics positively or negatively influences 24-month survival. We will see that this can be attributed to increased knowledge from these subgroups or to selection effects.

The remainder of the paper is organized as follows: in section 2 we discuss the data (Belgian TAVI registry) and variables. In section 3 we discuss the identification strategy and substantive questions of the paper. In section 4, we present our main results for the overall and subgroup learning effects. Section 5 continues with robustness checks before we draw final conclusions in section 6.

2 Data

This study makes use of the Belgian TAVI registry. This dataset contains detailed information on the first 860 patients undergoing TAVI in Belgium in 23 different centers. The dataset holds a wide range of control variables on patient- and hospital-specific characteristics and hospital identifiers. Specifically, we have information on the demographic background of patients, different comorbidities, indicators for the severity of the cardiac problem and procedural characteristics (see section 3.1 for more details on the background characteristics). Moreover, there is an extensive follow-up on mortality throughout time. The patient outcomes we study are 24-month and 36-month survival, as well as indicators for major adverse cardiac events (MACE) including renal failure, pacemaker implantations and stroke. Renal failure is known to be related to the use of contrast volumes during the TAVI procedure. Furthermore, stroke can be seen as an procedure induced event and pacemaker implantation is mostly related to the type of valve that is used. In fact, two types of valves are used in Belgian hospitals: CoreValve and SAPIEN replacement valves[2].

Table 1 gives a preliminary descriptive overview of the potential learning effects. Patients are divided in two categories according to physician experience. In the low experience group physicians treated at most 30 patients, in the high experience group more than 30 patients were treated before. On a bivariate level, we find little evidence that experience affects health outcomes. Only for pacemaker implantation we find that the high experience group obtains a pacemaker more frequently.

Table 1: Descriptive Learning Effects

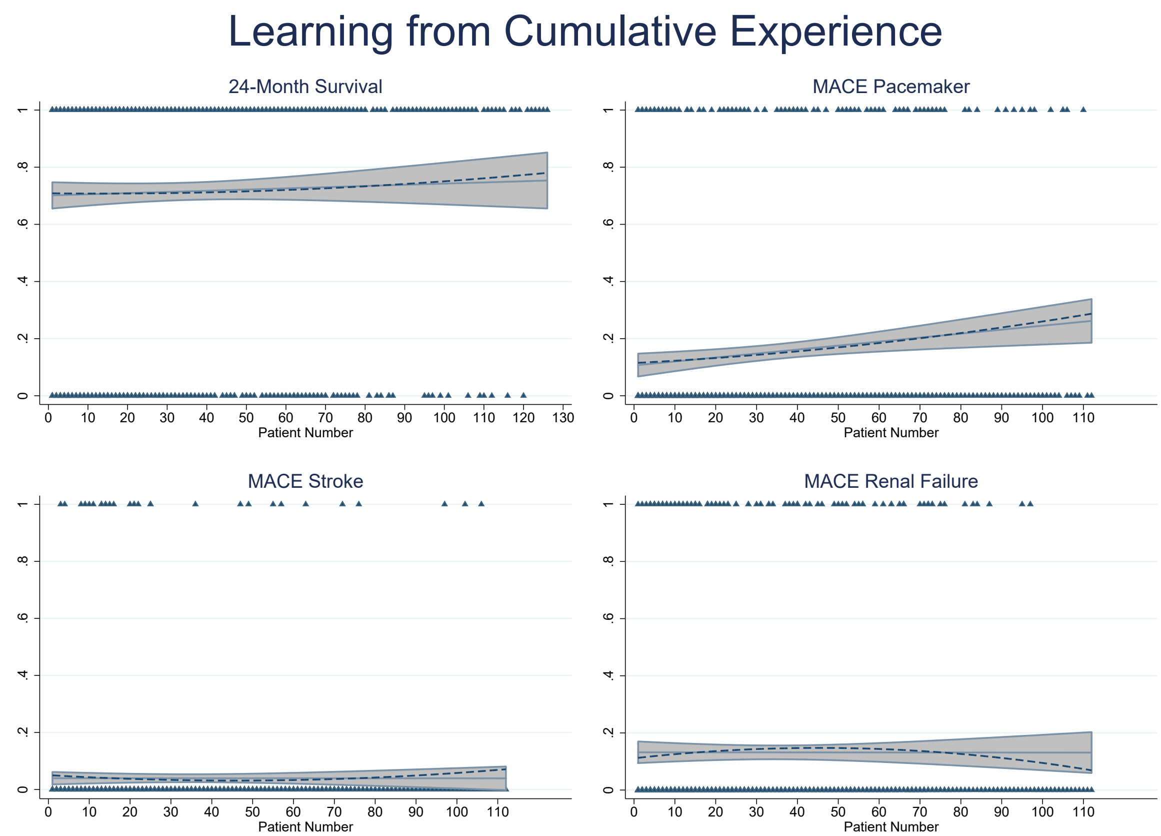
Variable	**Patients**		***z*-score**
	≤ 30	> 30	
Alive after 2 years	0.708	0.720	-0.351
MACE - Pacemaker	0.124	0.189	-2.360
MACE - Renal Failure	0.122	0.142	-0.767
MACE - Stroke	0.046	0.026	1.338

Notes: The test statistics are based on the comparison of proportions for "large samples" and were estimated using the Stata command `prtest`. This procedure is similar to the usual t-test, but it accounts for the binomial distribution of the underlying variable to calculate standard errors [Moore et al., 2009]. The division at 30 patients is based upon the mean, which is close to 30, and also the paper by Alli et al. [2012] which shows a plateau after 30 patients.

[2] CoreValve and SAPIEN are brand names for two types of Transcatheter Heart Valves used for TAVI in Belgium.

Furthermore, in Figure 1 below a positive relationship between experience and 2-year survival provides first evidence for possible learning from cumulative experience effects in the data. Note also that the quadratic and linear fits are nearly identical which points toward a linear relationship between 2-year survival and cumulative experience[3]. These positive learning effects are further reinforced by our parametric coefficient estimates which are all statistically significant different from zero across different model specifications (see section 4).

Figure 1: Preliminary Learning from Cumulative Experience Curves

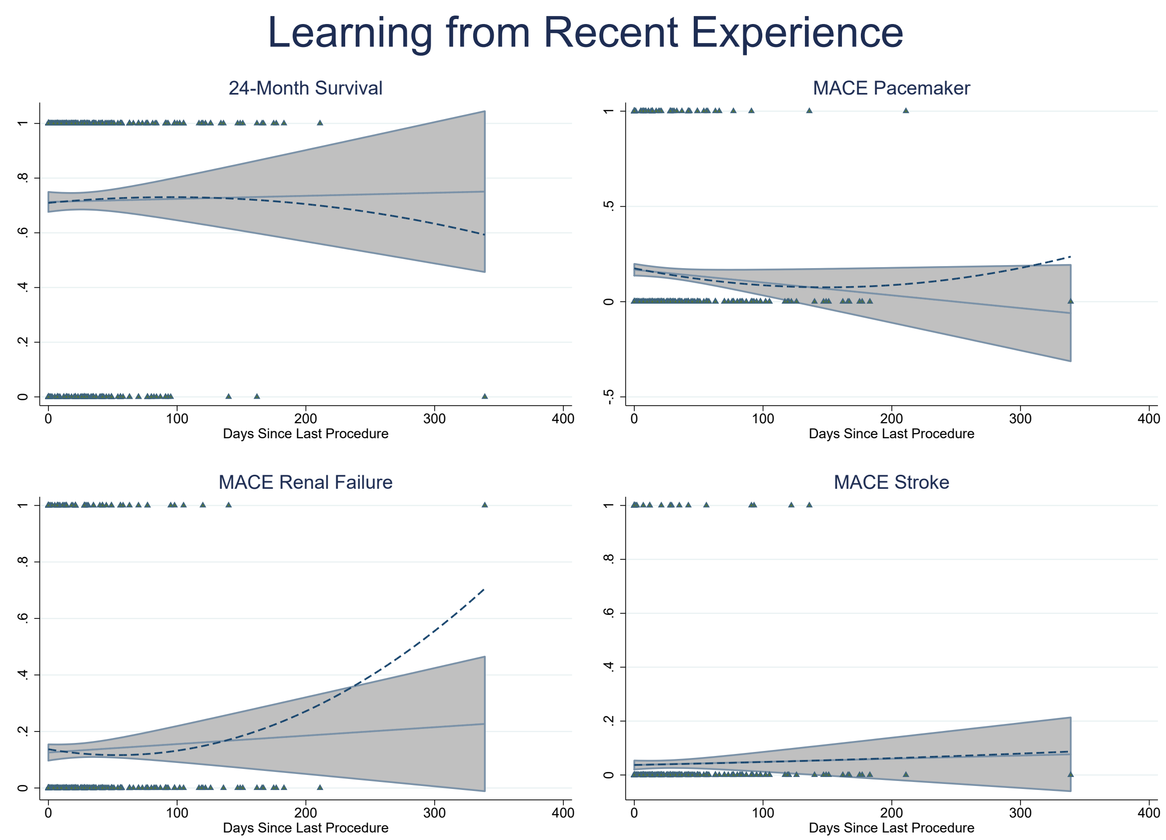


In addition, Figure 1 clearly shows a positive association between pacemaker implantations and cumulative experience. This suggests that as more patients are treated, they are more likely to receive pacemakers during the TAVI procedure. Again, we find positive and highly significant learning from cumulative experience effects regarding pacemaker implantations also in our regression models which condition on a wide range of patient- and procedure-specific covariates. This finding is most likely driven by the use of CoreValve valves in the larger centers because CoreValve valves are known to be associated with pacemaker implantation. In contrast to that, the figures for the adverse events of stroke and renal failure do not provide evidence for the existence of learning from cumulative experience as the rate of these events is roughly constant across all experience levels.

[3]As a logical consequence, we approximate the relationship with linear probability models (LPM).

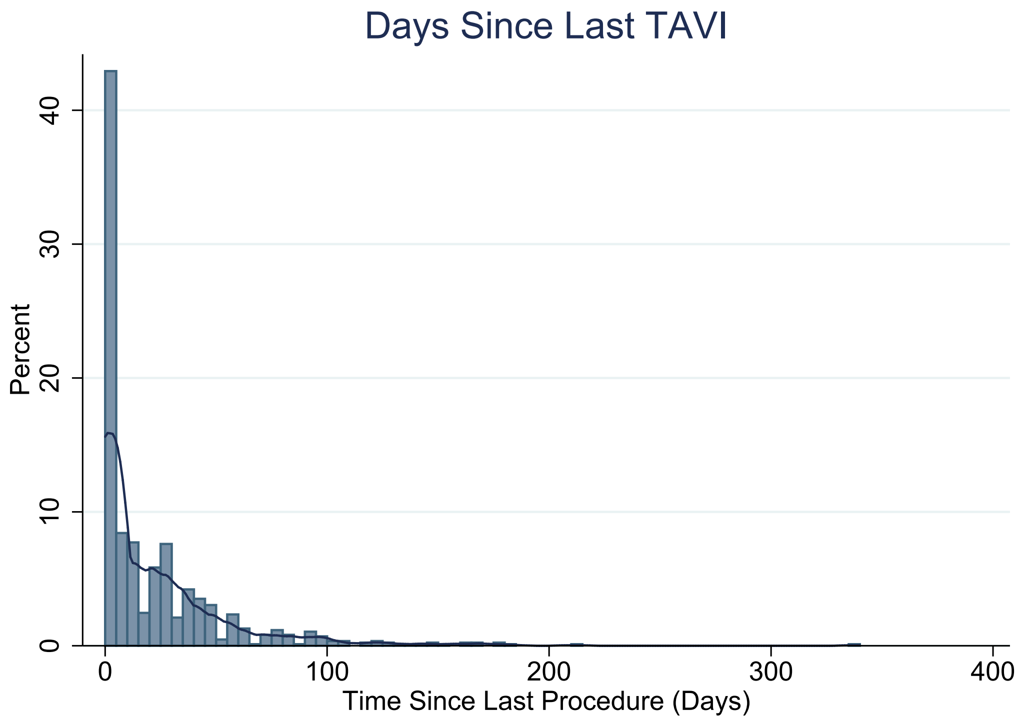
Turning the focus to learning from recent experience, we explore the bivariate relationship of the number of days since the last procedure and different MACE. In line with the findings of Hockenberry and Helmchen [2014], one would expect that the probability of mortality or major cardiovascular events should be increasing with temporal distance to the last procedure. Physicians' skills may suffer from a spell without practice which makes them more likely to make suboptimal decisions or mistakes during procedures (human capital depreciation hypothesis). This line of reasoning is supported for renal failure and stroke as depicted in Figure 2 below. In both cases, we observe a slightly positive relationship between the number of days since the last procedure and the likelihood of renal failure or stroke. This suggests the presence of human capital depreciation regarding adverse events. These findings are again confirmed by our regression models below which are providing evidence for positive and significant learning from recent experience effects. As for 2-year survival and pacemaker implantations the linear and quadratic fits diverge and thus the presence of human capital depreciation effects is unclear in this context and is therefore further explored below.

Figure 2: Preliminary Learning from Recent Experience Curves



Note that the number of days since last performing TAVI ranges from a minimum of zero days to a maximum of 408 days with a mean value of roughly 26 days (median 14 days) in between procedures. The distribution of the temporal distance to the last TAVI procedure in the overall sample is shown in Figure 3 below.

Figure 3: Histogram of Days Since Last TAVI Procedure



Notes: The Figure shows the distribution of the days since the last TAVI procedure and is overlaid with a kernel density estimate (solid dark blue line).

3 Methodology and Substantive Questions

Several overall learning and volume effects are distinguished on mortality and major adverse cardiac events (MACE) including renal failure, stroke and pacemakers. Three key questions arise: *First*, does overall hospital volume, which is related to equipment and facilities, have an impact on patient outcomes (static learning effects)? *Second*, do providers improve long term patient outcomes as more patients are treated in a hospital (learning from cumulative experience)? *Third*, does provider performance erode over time regarding patient outcomes (learning from recent experience)?

Subsequently, we broaden our scope to learning curves for patient subgroups. This again evokes two substantive questions: *First*, when treating more patients, do providers get better at treating subgroups? *Second*, when treating more subgroup patients, do providers get better in their overall care provision? Unlike the existing literature [Matsushima et al., 2014], this paper focuses on the second question as it is most relevant for the general population and because it provides useful information to transfer knowledge to policy makers and practitioners.

3.1 Overall Learning Curves

We estimate linear probability models (LPM) to capture the overall learning effects. Following Huesch and Sakakibara [2009], we distinguish between three types of learning: static learning [Huckman and Pisano, 2006, Gaynor et al., 2005, Ho, 2002], learning from cumulative experience [Bridgewater et al., 2004, Ho, 2002, Karamanoukian et al., 2000] and learning from recent experience [Hockenberry and Helmchen, 2014, Ramanarayanan, 2008, Huckman and Pisano, 2006]. Using patient-level data, we estimate models of the following form:

$$\begin{aligned} Outcome_{i,h,t} = \beta_0 + \beta_1 Static\_Volume_{h,t} + \beta_2 Cum\_Volume_{i,h,t} \\ + \beta_3 Time\_since\_last\_procedure_{i,h,t} + \alpha_1' X_{i,h,t} + \theta_1' H_h + \varepsilon_{i,h,t} \end{aligned} \tag{1}$$

Our outcome variables are binary indicators for the 24-month or 36-month survival rates and MACE indicators for pacemaker implantation, renal failure and stroke for patient i treated in hospital h in year t. $Static\_Volume_{h,t}$ measures the annual number of procedures in hospital h in year t picking up static scale effects. The rationale here is that high-volume hospitals are more likely to be better equipped and that they have improved processes of care and better standardization of procedures [Gaynor et al., 2005, Ho, 2002]. $Cum\_Volume_{i,h,t}$ is the patient number for individual i in hospital h in year t reflecting learning from cumulative experience or hospital-specific learning by doing [Ho, 2002]. This variable indicates how much the treatment of an additional patient improves provider performance. $Time\_since\_last\_procedure_{i,h,t}$ is the amount of days that have passed since the last TAVI procedure for patient i in hospital h and year t and captures the above mentioned learning from recent experience or human capital depreciation effect. It is sensible that the longer the time between procedures, the more skills suffer from absence of practice [Hockenberry and Helmchen, 2014]. Note that β_2 and β_3 can also be interpreted as the impact of practical skills and increased knowledge respectively.

Besides our three main volume and time indicators, we control for a vector of patient- and procedure-specific characteristics $X_{i,h,t}$ which includes information on the demographic background of a patient (age, gender), comorbidities (indicators for various heart diseases, diabetes, renal failure, angina and existing pacemaker), the severity of the cardiac problem (NYHA categories, ejection fraction, aortic valve area, peak and mean gradient), as well as procedure-specific characteristics (type of valve and size of valve). We control for these observable characteristics as they have been identified in the literature to be key determinants of mortality [Holt et al., 2007]. Conditioning on all these factors allows us therefore to isolate the different types of learning effects outlined above. In addition, we include a vector of hospital fixed-effects H_h to account for time-invariant unobserved factors such as quality of care and hospital management quality that potentially differ across hospitals and affect the outcomes of interest. Finally, $\varepsilon_{i,h,t}$ is a classical error term capturing all unobserved time-varying factors such as genetic endowment and health behaviors of patients that also explain our outcomes of interest besides the included explanatory variables.

Time fixed-effects can also be added to the empirical specification to capture "learning-from-watching" and technological improvements. However, similar to Ho [2002], adding year fixed-effects would result in highly collinear effects. Leaving out the time fixed-effects from our models then necessitates interpretation of other learning effects as upper bounds on the true effects because they may pick up part of the positive effect of technological improvements over time.

Typically, endogeneity may arise because of selective referral. In principle, overall outcomes, if publicly known, may cause more patients to select into certain hospitals. There is mixed evidence on the direction of causation. Gaynor et al. [2005] and Ho [2002] found that the causal direction mainly runs from volume to outcome. However, Ramanarayanan [2008] found that sicker patients may select higher volume providers. In the Belgian setting, with very little information on hospital quality, and with even less information on procedure related hospital quality, it is unlikely to find that outcomes cause volume ruling out reverse causality issues. Moreover, by splitting up the learning effect in subgroups, at least part of the selection effect is removed from the overall effect. We return to this statement later in this paper.

3.2 Multicollinearity and Subgroup Learning Curves

In the subgroup analyses all variables from the overall curves are retained and the experience variables are further divided in more detailed groups. For 2-year survival, experience variables are added for all background characteristics. If for example the 30^{th} patient for a provider (hospital) is the 15^{th} patient with hypertension for the same provider, the patient gets patient number 30 and experience for hypertension 15. Statistically these variables are likely to be strongly correlated[4] and this multicollinearity then results in highly insignificant values. By treating more patients overall, also more patients with renal failure, porcelain aorta, etc. will be treated.

To deal with multicollinearity, two general solutions are often proposed: *firstly*, the selection of a subset of variables remedies the consequences of multicollinearity by removing the collinearities. Suppose two variables are highly correlated; removing one of the two variables from the model causes the standard error of the coefficient on the remaining variable to drop significantly. Therefore, we explore the use of the Least absolute shrinkage and selection operator (Lasso) to obtain an optimal subset of experience variables.

Secondly, increasing information provides more evidence to disentangle even collinear effects. This extra information may come from an increase of the sample size or from a restriction on regression coefficients. In this light, we apply the Theil-Goldberger mixed estimation method to introduce (uncertain) information on a sum of coefficients. Although variable selection methods and the use of extra information have very different motivations, they can both be seen as applications of constraints in a regression analysis. This is discussed in more detail throughout the next sections.

[4] Correlations larger than 0.9 are no exception in our sample.

3.2.1 The Least absolute shrinkage and selection operator (Lasso)

To select an optimal subset of regressors, multiple statistical approaches can be considered: subset selection techniques such as forward- and backward-stepwise selection, forward-stagewise regression or shrinkage methods including Ridge regression and the Lasso. Both selection methods improve the statistical analysis in terms of prediction accuracy and interpretation [Hastie et al., 2009]. However, model selection goes at the cost of biased estimates. While the least squares estimator is the best linear unbiased estimator (BLUE), there may exist biased estimators which are more efficient. Shrinkage methods are based on the idea of shrinking single coefficients or sets of coefficients towards zero which trades off lower variance for increased bias. The sparseness of the resulting model also facilitates interpretation. Among all shrinkage methods, Lasso regression introduced by Tibshirani [1996], is favored in this paper because it is more subtle compared to forward- and backward-selection while at the same time it provides sparser results compared to Ridge regression. Technically, the Lasso minimizes the residual sum of squares subject to the constraint that the sum of all absolute values of coefficients is below some constant. Following Hastie et al. [2009] we have:

$$\beta^{lasso} = \underset{\beta}{argmin} \quad \sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p}[x_{ij}\beta_j])^2 \quad s.t. \quad \sum_{j=1}^{p}\beta_j \leq t \tag{2}$$

or alternatively

$$\beta^{lasso} = \underset{\beta}{argmin} \quad \sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p}[x_{ij}\beta_j])^2 + \lambda \sum_{j=1}^{p}|\beta_j|^q \tag{3}$$

where the first part of both equations simply finds the $\beta' s$ for which the sum of squared residuals is lowest. The second part states that the minimization is subject to the condition that the sum of the absolute values of β should be lower than a predefined constant t. Whereas the approach is similar to Ridge regression where a similar constraint is placed on the sum of all squared coefficients, the geometric properties of the lasso sets more coefficients exactly to zero [Tibshirani, 1996]. As such, the benefit of lasso is that it decreases variance while at the same time providing a sparse model. Additionally, because it is not a discrete process in which variables are added one by one, the lasso is less greedy than forward- or backward-variable selection [Efron et al., 2004]. Lasso estimates can be obtained from the Least Angle Regression Selection (LARS) algorithm which also provides more insight in Lasso. In the LARS algorithm, the coefficient of the most correlated variable is increased until the point where a second variable is equally correlated. From this point onwards, the coefficients of both variables are increased until a third variable is equally correlated, etc. This procedure goes on until Mallows' Cp reaches a minimum which in turn provides the subset of variables that best predicts the outcome. Adjusting the LARS algorithm by removing a variable (temporarily) from the active set when the coefficient is set to zero in the LARS, generates the Lasso. Park and Casella [2008] work out the Bayesian Lasso approximation mentioned in Tibshirani [1996]. The constraint can be implemented through Laplace priors and the estimates seem to lie in between the Lasso and Ridge results.

The Lasso singles out the most significant variables that predict health outcomes while other variables are excluded from the model. Next to the standard Lasso, we also employ some modifications and extensions of the Lasso as a robustness exercise. *Firstly*, we use the Lasso to select the subset of regressors and run OLS on this subset after Lasso. This approach is suggested in Efron et al. [2004], Meinshausen [2007], Hastie et al. [2009] to reduce the bias and to allow for a simpler interpretation of the coefficients. *Secondly*, we also restrict the Lasso by adding the "main effects" first. That is, we add all variables except for the volume, learning and experience variables (overall and on the subgroups). Then, controlling for all background information, we search for the most relevant learning predictors using the Lasso[5] [Efron et al., 2004]. *Thirdly*, we run logistic Lasso regressions as our outcomes of interest are binary. Similarly to the Least Squares Lasso, an L_1[6] penalty on the absolute values of coefficients can be introduced to logistic regression [Genkin et al., 2007].

3.2.2 Including prior information with Theil-Goldberger mixed estimation

Intuitively, multicollinearity is the occurrence of "undominated uncertain prior information" [Leamer, 1973]. This definition points out that including extra prior information might soften the multicollinearity problem. Including prior information may increase evidence in favor of a certain hypothesis and it therefore reduces data limitations. Bayesian analysis, where the use of priors is common, is criticized by many for its subjectivity in specifying prior information. However, in this study, prior information from within the data can be used to estimate subgroup effects. Intuitively it is clear that the overall cumulative learning effect, i.e., every time a new patient is treated, the probability for a positive outcome increases, is the sum of all underlying subgroup effects. The prior information we use in this study is thus that the combination of subgroup effects adds up to the overall effect. Interesting in this regard is the Theil and Goldberger [1961] mixed estimation method which uses GLS on an augmented dataset. In this augmented dataset, the data is supplemented by a dummy observation with information on the mean and variance of a (sum of) coefficient(s). This method of data augmentation has widely received credit and gives very similar results as more complex methods using posterior sampling [Discacciati et al., 2015]. In economics, data augmentation is broadly used in the VAR literature under the name of "sum of coefficients prior" [Robertson and Tallman, 1999, Sims, 2005]. As the name already suggests, the method of Theil and Goldberger is similar to a Bayesian analysis with conjugate priors [Theil, 1963]. While it is applied in a frequentist framework, it is definitively Bayesian at heart and adding the dummy observation is a straightforward way of doing it. As it is known to do well in softening multicollinearity on the autoregressive components in the VAR

[5] This is done by using the residuals from a regression of the dependent variable on all main effects and to use these residuals as the dependent variable in a Lasso regression on all subgroup effects.

[6] An L_1 penalty is a constraint on the sum of absolute values as in equations (2) and (3). Alternatively, an L_2 penalty is used in Ridge regression.

literature, we follow the same approach here.

The Theil-Goldberger coefficients and variances, where prior and data information are efficiently weighted in a GLS framework, are given by [Theil and Goldberger, 1961]:

$$\hat{\beta} = [X'\Omega^{-1}X + R'\Psi^{-1}R]^{-1}[X'\Omega^{-1}y + R'\Psi^{-1}r] \tag{4}$$

and

$$V(\hat{\gamma}) = [X'\Omega^{-1}X + R'\Psi^{-1}R]^{-1} \tag{5}$$

X is a $n \times k$ matrix of observations on independent variables; Ω is the $n \times n$ variance-covariance matrix of residuals and Ψ is the variance-covariance matrix of the prior information. For prior information on a sum of coefficients, the $1 \times k$ vector R and the scalar r have to be specified. For example, imposing a constraint on the sum of β_1 and β_2 could be achieved by specifying:

$$R = [1\ 1\ 0 \cdots 0\ 0] \tag{6}$$

$$r = [\beta_1 + \beta_2] \tag{7}$$

As such, equations (4) and (5) are the result of applying GLS to the following two equations:

$$y = X\beta + u \tag{8}$$

and

$$r = R\beta + v \tag{9}$$

Equation (8) holds the relationship for the “real” data. Next, the real data is augmented with an extra observation in the form of the constraint in equation (8). The information in (8) is perfectly similar to the application of constraints in regression analyses. The main difference however is that the constraint is not exact, meaning that there is some uncertainty about the prior information (hence the Ψ matrix). To see how the internal information can be used here as prior information; first consider a consistent (and linear) estimate of the learning curve:

$$m_{ij} = \beta_0 + \beta_1 patnr_{ij} + \beta_2(characteristic\_1)_{ij} + \varepsilon_{ij} \tag{10}$$

m_{ij} stands for two-year mortality for individual i in hospital j and $patnr$ is the patient number of an individual (e.g. patient number 1 in hospital 20). In equation (10), the coefficient on the patient number β_1 is consistently estimated using standard regression techniques. Now let us think of a second model:

$$m_{ij} = \gamma_0 + \gamma_1 patnr_{ij} + \gamma_2(characteristic\_1)_{ij} + \gamma_3(exper\_characteristic\_1)_{ij} + \varepsilon_{ij} \tag{11}$$

In this model $(exper\_characteristic\_1)_{ij}$ is a variable taking the value zero for the first patients until there has been one person with a certain characteristic (say characteristic 1). From then onwards, the variable $(exper\_characteristic\_1)_{ij}$ takes on the value one and is not increased until another patient with the same characteristic is treated. As such, the variable can be seen as an experience variable for characteristic 1. In equation (10), β_1 is a consistent estimate for the increase in the health outcome every time an extra patient is treated in a hospital. In equation (11), the same increase in health outcome for every extra patient is given by γ_1 and γ_3 at the same time. That is, every time an extra patient is treated, health increases by γ_1 and also with approximately the amount $\gamma_3 \times avg(characteristic\_1)$. The increase with γ_1 is obvious while the second part is an increase of γ_3 for every patient with characteristic one and on average only $avg(characteristic\_1)$ of the patient population has the characteristic. As such on average, every time an extra patient is treated, the outcome increases by $\gamma_3 \times avg(characteristic\_1)$. β_1 in equation (10) can therefore be seen as the sum of γ_1 and $\gamma_3 \times avg(characteristic\_1)$. The translation of this prior theoretical knowledge to the matrices that define the constraint on the sum of coefficients is as follows:

$$R = [1 \; avg(char1) \; 0 \cdots 0 \; 0] \tag{12}$$

and

$$r = [\beta_1] \tag{13}$$

4 Results

This section gives an overview of the results for the estimated overall and subgroup learning effects for different model specifications and estimation techniques. We start with the discussion of the overall learning effects, where we distinguish between three types of learning: static learning (economies of scale), learning from cumulative experience and learning from recent experience (section 4.1). In the next step, we explore subgroup learning effects by means of the Lasso and Theil-Goldberger mixed estimation (section 4.2).

4.1 Overall Learning Curves

We estimate the overall learning effects using linear probability models (LPM) for 24-month and 36-month survival, as well as several Major Adverse Cardiac Events (MACE) including pacemaker implantation, renal failure and stroke. We demonstrate robustness of our overall learning effects by estimating three model specifications for each of the outcomes: model one shows the plain overall learning effects for the three learning measures static learning ("Annual Volume"), learning from cumulative experience ("Patient Number") and human capital depreciation ("Days Since Last Procedure"; "Zero Days Since Last Procedure"); model two adds patient- and procedure-specific characteristics as described above in section 3.1; finally, in model three we include hospital fixed-effects. Note that we also replicate our findings using probit/logit specifications to relax the implicit linearity assumption in the marginal probability effects in the LPM. However, the average marginal probability effects are almost perfectly identical to the results shown below reinforcing robustness of our findings.

The estimated overall learning effects on survival can be found in table 2 below: *First*, we find a positive and significant effect on the patient number indicator for 24-month and 36-month survival which points toward learning from cumulative experience across all three model specifications. In fact, our final specification suggests that treating an additional TAVI patient is associated with an increase in 2-year survival of 0.16%-points, ceteris paribus. Likewise, 36-month survival is increased by about 0.30%-points. These cumulative learning effects are sizeable considering that patient volumes were increased on average by more than 10 patients per year in the timespan from 2007-2012. This cumulative learning effect can be interpreted as the result of a learning process in technical skills, but also in the selection process of patients. More experienced teams might be better in selecting patients with a high probability of being alive after 2 years. Also, in the very beginning of the procedure, operators might not yet be aware of the limited potential of the procedure.

Table 2: Overall Learning Effects:
Survival

Outcome Variable	**24-Month Survival**			**36-Month Survival**		
Specification	(1)	(2)	(3)	(1)	(2)	(3)
Annual Volume	**-0.0040****	**-0.0053****	-0.0044	**-0.0059*****	**-0.0071*****	-0.0032
	(0.0020)	(0.0023)	(0.0031)	(0.0021)	(0.0023)	(0.0032)
Patient Number	**0.0013***	**0.0013***	**0.0016****	**0.0019*****	**0.0021*****	**0.0030*****
	(0.0006)	(0.0007)	(0.0008)	(0.0007)	(0.0007)	(0.0008)
Time Since Last Procedure (Days)	-0.0001	-0.0004	-0.0005	0.0001	-0.0003	-0.0005
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Zero Days Since Last Procedure	-0.0134	-0.0342	-0.0396	0.0240	0.0034	-0.0053
	(0.0379)	(0.0390)	(0.0400)	(0.0393)	(0.0406)	(0.0413)
Patient- and procedure-specific characteristics	No	Yes	Yes	No	Yes	Yes
Hospital fixed-effects	No	No	Yes	No	No	Yes
Number of Observations	854	780	780	854	780	780

Notes: Heteroscedasticity robust standard errors in parentheses: *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$.

Second, table 2 provides evidence for static learning as models one and two show negative and highly significant coefficients on the annual volume measure. Note that "Annual Volume" is only insignificant when hospital fixed-effects are included. The hospital fixed-effects pick up volume effects persisting over time and therefore the yearly volume effects get insignificant when including hospital indicators. *Third*, our estimates do not indicate that skill depreciation has an effect on survival as our time difference indicators ("Time Since Last Procedure"; "Zero Days Since Last Procedure") do not show any statistically significant coefficients.

In addition to the survival outcomes, we analyze learning curves regarding adverse cardiac events. The results are summarized in table 3 below. While cumulative and static learning showed to be significant for the survival outcomes above, skill depreciation is significant for several MACE. We find that the likelihood of renal failure and stroke are both significantly increased as time passes since the last TAVI procedure. To be more precise, our estimates suggest that the likelihood of renal failure after TAVI is increased by 0.12%-points for every additional day since the last procedure, ceteris paribus. Less drastic is the human capital depreciation in case of stroke. Here an additional day since the last procedure increases the probability of suffering a stroke by about 0.07%-points. Again these skill depreciation effects can be considered sizeable as the average number of days between procedures is more than 10 days across all hospitals and time periods. Regarding stroke, we also find that patients treated on the same day ("Zero Days Since Last Procedure") have a higher probability of getting a stroke which may point out that the team loses concentration during the course of a given day. However, this effect is not found for the other adverse events. As can be seen in the robustness section, the results on MACE should be interpreted with care. These results are sometimes driven by only a few extreme observations.

Table 3: Overall Learning Effects:
Major Adverse Cardiac Events (MACE)

Outcome Variable	**MACE Pacemaker**			**MACE Renal Failure**			**MACE Stroke**		
Specification	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Annual Volume	0.0011	0.0018	-0.0018	-0.0009	-0.0004	0.0007	-0.0007	-0.0004	-0.0012
	(0.0017)	(0.0018)	(0.0031)	(0.0015)	(0.0017)	(0.0025)	(0.0009)	(0.0008)	(0.0011)
Patient Number	**0.0012***	**0.0013***	0.0009	0.0004	0.0004	-0.0001	0.0002	0.0003	0.0002
	(0.0007)	(0.0007)	(0.0008)	(0.0005)	(0.0006)	(0.0007)	(0.0003)	(0.0003)	(0.0004)
Time Since Last Procedure (Days)	0.0000	0.0002	0.0003	0.0006	**0.0009***	**0.0012****	0.0002	0.0005	**0.0007****
	(0.0004)	(0.0004)	(0.0004)	(0.0006)	(0.0005)	(0.0005)	(0.0003)	(0.0003)	(0.0004)
Zero Days Since Last Procedure	**0.0591***	0.0394	0.0497	0.0418	0.0482	0.0518	0.0165	**0.0315***	**0.0403****
	(0.0327)	(0.0325)	(0.0331)	(0.0314)	(0.0316)	(0.0336)	(0.0174)	(0.0182)	(0.0194)
Patient- and procedure-specific characteristics	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Hospital fixed-effects	No	No	Yes	No	No	Yes	No	No	Yes
Number of Observations	854	780	780	854	780	780	854	780	780

Notes: Heteroscedasticity robust standard errors in parentheses: *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$.

Summing up, we find that different types of learning apply for different outcomes. Cumulative learning which might also be interpreted as increased knowledge is relevant for 24-month and 36-month survival. For adverse events, more frequent practice plays a more important role because skills required for preventing the events may depreciate over time as illustrated for the occurrence of renal failure and stroke.

4.2 Subgroup Learning Curves

Knowing that there are different types of learning for different outcomes, it is interesting to investigate to what extent patient subgroups with certain characteristics account for these overall effects. Table 4 below shows the estimates from a linear probability model (LPM) using 24-month survival as the dependent variable while controlling for all patient- and procedure-specific characteristics and hospital fixed-effects. In contrast to the models estimated above, the specification below also includes all the experience variables for each of the background characteristics. The issue of multicollinearity is immediately apparent in the columns VIF which

Table 4: Subgroup Learning Effects:
OLS Estimates

Outcome Variable: 24-Month Survival

Explanatory Variable	**Coefficient**	**VIF**	**Explanatory Variable**	**Coefficient**	**VIF**
Annual Volume	**-0.0084***	8.26	Experience_mydrocardial infarction	-0.0115	-0.011
	(0.0047)			(0.0188)	
Patient Number	0.0388	**9138.33**	Experience_porcelain aorta	-0.0125	**23.07**
	(0.0488)			(0.0261)	
Time Since Last Procedure (Days)	-0.0005	1.73	Experience_renal failure	0.0107	**72.67**
	(0.0006)			(0.0171)	
Zero Days Since Last Procedure	-0.0427	1.59	Experience_ percutaneous coronary intervention	-0.0055	**142.91**
	(0.0399)			(0.0203)	
Experience_female	-0.0179	**277.92**	Experience_pacemaker	-0.0085	**25.32**
	(0.0172)			(0.0221)	
Experience_angina	-0.0062	**186.7**	Experience_stroke	0.0362	**71.05**
	(0.0181)			(0.0272)	
Experience_aortic aneurism	**-0.0750***	**81.68**	Experience_mediastinal radiation	-0.0062	**33.54**
	(0.0398)			(0.0470)	
Experience_atrial fibrillation	-0.0197	**172.28**	Experience_defibrillator	-0.1262	7.34
	(0.0224)			(0.1181)	
Experience_carotis disease	-0.0155	**62.65**	Experience_CABG	0.0097	**235.17**
	(0.0182)			(0.0259)	
Experience_coronary artery disease	0.0085	**649.49**	Experience_valvesurg	0.0428	**28.25**
	(0.0217)			(0.0543)	
Experience_chronic obstructive disease	-0.0127	**96.16**	Experience_CoreValve	0.0190	**887.82**
	(0.0185)			(0.0198)	
Experience_chronic heart failure	-0.0053	**378.21**	Experience_transfemoral access	-0.0258	**631.81**
	(0.0116)			(0.0171)	
Experience_diabetes	0.0050	**123.44**	Experience_smallest valve	0.0138	**456.25**
	(0.0234)			(0.0218)	
Experience_hypertension	0.0241	**549.23**	Experience_medium valve	-0.0117	**912.86**
	(0.0156)			(0.0336)	
Experience_nyhacat2	-0.0276	**14.21**	Experience_large valve	-0.0838	5.32
	(0.0392)			(0.0632)	
Experience_nyhacat3	-0.0313	**22.63**	Experience_highef	-0.0063	**526.92**
	(0.0334)			(0.0216)	
Experience_nyhacat4	-0.0047	**472.23**	Experience_pulmonary hypertension	0.0043	**108.04**
	(0.0393)			(0.0113)	
Patient- and procedure-specific characteristics	Yes		Number of Observations	780	
Hospital fixed-effects	Yes		R^2	0.1419	

Notes: For the sake of brevity, the coefficient estimates on the patient- and procedure-specific characteristics and the hospital indicators are not shown in the table above. VIFs above 10 in bold. Heteroscedasticity robust standard errors in parentheses: *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$.

display the Variance Inflation Factors[7] for all variables. In fact, almost all experience variables show VIF values above ten indicating serious multicollinearity issues. As a direct consequence of the strong correlation among the regressors, standard errors are inflated leading to potentially insignificant subgroup learning effects. Overall the results indicate both positive and negative experience effects: for example, treating more patients with diabetes, renal failure or stroke is associated with an increase in 24-month survival, though statistically insignificant. On the other hand, treating patients for example with an aortic aneurism has a significant negative effect on 2-year survival indicating that this subgroup might be carefully looked at for future health improvements. Finding negative effects can be due to two facts: *Firstly*, "experience" for a certain characteristic may be transferred from some subgroups to others where it actually should not be transferred. *Secondly*, over time, "worse patients" with stronger manifestations of the characteristic tend to be treated. In addition, the above results are consistent with our findings in table 2 as we find statistically significant evidence for static learning even when including all the experience variables. In sharp contrast to that, the cumulative learning effect is no longer significant due of the strong correlation to all the experience variables as indicated by the largest VIF on the patient number indicator.

4.3 Lasso Regressions

The Lasso selects those variables most correlated to the dependent variable and finds the subset which has the lowest Mallows C_p. Similar to the Akaike information criterion, this criterion chooses a model with the lowest sum of squared residuals but simultaneously punishes the inclusion of extra variables. Whereas the Lasso succeeds in finding the most relevant subset of variables, it simultaneously ignores the fact that the variables may typically pick up effects from other variables. In this sense, some of the variables may in fact not be relevant, but only become statistically significant because of their relationship with other characteristics. Nevertheless, singling out the most relevant experience predictors might inform where to start looking to improve health outcomes.

Table 5 below contains a range of Lasso-related specifications: In the first specification (1), the usual Lasso is applied on 2-year survival to detect the subset of optimal predictors. Here the Lasso reveals that the overall annual TAVI volume and three subgroup experience variables are particularly relevant for survival: *First*, treating more patients with a porcelain aorta[8] is negatively associated with 24-month survival. *Second*, experience with defibrillation and using CoreValve (Typevalve=1, 0 for Sapiens) replacement valves on the other hand seems to positively affect patient survival. Whereas the CoreValve indicator is the most robust experience variable across all specifications, valve types are constant within hospitals and might therefore pick-up part of the learning differences between hospitals. Besides the subgroup experience variables, the Lasso also identifies the annual TAVI patient volumes as a key negative predictor of 2-year

[7] The VIF for regressor k is defined as: $\frac{1}{1-R_k^2}$, where R_k^2 is the R-squared from a regression of x_k on all other explanatory variables

[8] A porcelain aorta is a heavily calcified ascending thoracic aorta which may obviate usual aortic valve replacement through that approach.

survival which is again consistent with the findings presented above.

In the second column, we use a simple linear regression model on the optimal subset of predictors identified in column (1) to reduce the bias in the estimated Lasso parameters. A comparison of specifications (1) and (2) clearly shows that Lasso shrinks' the regression coefficients towards zero as all the estimated parameters in (1) are smaller in absolute value than in (2). More importantly however is that specification (2) reinforces the existence of subgroup learning effects as now the coefficients on experience with defibrillation and CoreValve replacement valves becomes statistically significant. In fact, both coefficients are strictly positive and thus indicate that using a CoreValve during the TAVI procedure and being treated by a practitioner with a lot of defibrillator experience has the potential to increase a patients likelihood of long-term survival.

The third specification adds in a first step all non-experience and non-learning related variables and then finds that adding experience variables with the Lasso does not provide any supplementary significant effects. Overall this indicates that we are not able to find substantial and robust subgroup learning effects controlling for all basic variables by applying the Lasso. This may indicate that the experience variables pick up parts of the effects of the basic variables in specifications (1) and (2).

The results in columns (4) and (5) show the selection of optimal subsets of predictors for two logistic Lasso specifications using different lambda parameters (these are the Lagrangian penalties on the likelihood function)[9]. Model (4) uses the global optimum (λ=232) and (5) uses a smaller value ($\lambda = 100$) to obtain more regressors because of a smaller penalty. In model (5), we find again evidence for both positive and negative overall and subgroup learning effects as the logistic Lasso exclusively selects learning-related variables: The effect on the static annual volume shows the usual negative sign which can also found in specifications (1), (2) and tables 2 and 4. Moreover, both specifications (4) and (5) provide again evidence for a positive subgroup learning effect of using CoreValve replacement valves. Furthermore, specification (5) adds two additional experience variables to the set of optimal predictors of 2-year survival, namely experience with patients with a chronic heart failure, as well as using smallest size valves in the TAVI procedure.

To summarize, the standard Lasso variables selection procedure suggests that especially using more CoreValve replacement valves during TAVI procedures and increased physicians experience with defibrillators play the most important role for 2-year patient survival. An interesting point to stress here is that the positive subgroup learning effects on survival are driven by experience with procedural-characteristics (type of valve implanted) and physician specific knowledge rather than by the type of patient that is treated.

[9] See appendix A for a plot of optimal λ's.

Table 5: Subgroup Learning Effects:
Lasso Estimates

Outcome Variable: 24-Month Survival					
Specification	**Lasso** (1)	**OLS on Lasso** (2)	**Main Effects first** (3)	**Logistic Lasso** (4)	**Logistic Lasso** (5)
Annual Volume	-0.0014	**-0.004***	-	-	-0.012
Time Since Last Procedure	-	-	-	-	0.0002
Female	0.0125	0.028	0.028	-	-
Atrial fibrillation	-0.0379	-0.054	-0.052	-	-
Carotis disease	0.0676	0.113***	0.107**	-	-
Chronic obstructive pulmonary disease	-0.0908	-0.115***	-0.112***	-	-
Pulmonary hypertension	-0.0392	-0.069**	-0.095**	-	-
Porcelain aorta	-0.0642	-0.107*	-0.118*	-	-
Renal failure	-0.0494	-0.074*	-0.073*	-	-
Stroke	-0.062	-0.101**	-0.101**	-	-
nyhacat4	-0.0176	-0.038	-0.042	-	-
CoreValve	0.012	0.035	0.324***	-	-
Transfemoral access	0.0714	0.109***	0.110***	-	-
Medium valve	-0.043	-0.112**	-0.160**	-	-
Exper_porcelain aorta	-0.002	-0.006	-	-	-
Exper_defibrillator	0.0343	**0.074***	-	-	-
Exper_CoreValve	0.001	**0.002***	-	0.006	0.006
Exper_chronic heart failure	-	-	-	-	-0.006
Exper_smallest valve	-	-	-	-	0.015
Hospital fixed-effects	No	No	Yes	No	No
Number of Observations	780	780	780	780	780

4.4 Theil-Goldberger Mixed Estimation

Whereas computation of the Lasso is relatively straightforward, the Theil-Goldberger (TG) method is not standardly available in statistical software[10]. Computation requires the calculation of the formulas in equation (4) and (5) (see section 3.2.2 above). Results of this computation with a stochastic constraint are provided in specifications (4)-(6) in table 6 next to the OLS results (specification (1)), constrained regression with an exact constraint (specification (2)) and GLS instead of the LPM with robust standard errors (specification (3)). For the Theil-Goldberger estimation we also implement robust standard errors (specification (5)) and a GLS form (specification (6)) for the non-augmented part. The latter refers to the use of weighted least squares on the non-augmented data, i.e., the matrix Ω^{-1} is estimated to obtain a feasible GLS estimate.

Overall, the results show that the jump from OLS to GLS contributes more in efficiency terms than TG estimation in itself. As expected, the standard errors for TG are in between standard errors of OLS and constrained regression (compare specifications (1), (2) and (4)). The main effect of implementing GLS is that more variables turn out to be significant regressors due to substantial reductions in standard errors. In contrast, the effect of implementing the stochastic constraint on coefficients generates a limited impact on standard errors and by extension on significance. However, there are some significant differences between models (3) and (6). Intuitively, the additional information that is added in the TG estimation seems rather limited. Whereas there are no qualitative differences between the specifications without GLS, we clearly observe differences in significance between the third and sixth specification (for an illustration see grey shaded bars in table 6): On top of experience with aortic aneurysm ("Experience_aortic aneurism"), carotis disease ("Experience_carotis disease"), hypertension ("Experience_hypertension"), porcelain aorta ("Experience_porcelain aorta") and transfemoral access ("Experience_transfemoral access"), we find additional significant effects for atrial fibrillation ("Experience_atrial fibrillation") and New York Heart Association category 3 ("Experience_NYHAcat3") when using TG mixed estimation.

These findings should be closely scrutinized to find how experience translates in better outcomes. In particular, we find evidence for positive learning effects on 2-year survival for treating more patients overall (learning from cumulative experience), as well as treating more patients with hypertension (subgroup learning). On the other hand, treating additional patients with an aortic aneurism, atrial fibrillation, carotis disease, porcelain aorta or using the transfemoral access route is associated with lower patient survival and thus indicating negative subgroup learning effects. This evidence for negative subgroup learning strongly suggests the presence of selection effects. Typically, patients with these adverse characteristics become more vulnerable over time and thus more likely to die in two years' time. In a sense, the division of the overall

[10] In Stata, the `tgmixed` command implements a limited version of the Theil-Goldberger mixed estimation method. There is no option to include robust standard errors for the "real" data (which is a priori essential for a Linear Probability Model) and there is no possibility to insert prior information on the sum of coefficients. Mata program code is provided in appendix B to include these possibilities. The `tgmixed` command ado file was used as a guide while writing this code.

Table 6: Subgroup Learning Effects: Theil-Goldberger Mixed Estimation

Outcome Variable: 24-Month Survival	**OLS**						**Theil-Goldberger**					
Specification	**Robust OLS** (1)		**Constrained Regression** (2)		**GLS** (3)		**Normal** (4)		**Robust** (5)		**GLS** (6)	
	Coef.	R.S.E	Coef.	R.S.E	Coef.	S.E.	Coef.	S.E.	Coef.	R.S.E	Coef.	S.E.
Annual Volume	-0,008*	0,005	-0,008*	0,005	-0,004	0,003	-0,008*	0,005	-0,008*	0,005	-0,004	0,003
Patient Number	0,039	0,049	0,034	0,043	**0,071***	**0,04**	0,05	0,05	0,05	0,047	**0,086****	**0,036**
Time Since Last Procedure (Days)	-0,001	0,001	-0,001	0,001	0	0,001	-0,001	0,001	-0,001	0,001	0	0
Zero Days Since Last Procedure	-0,043	0,04	-0,042	0,04	-0,013	0,035	-0,043	0,041	-0,043	0,04	-0,015	0,033
Exper_female	-0,018	0,017	-0,018	0,017	-0,017	0,015	-0,019	0,017	-0,019	0,017	-0,019	0,014
Exper_angina	-0,006	0,018	-0,007	0,018	-0,006	0,016	-0,003	0,018	-0,003	0,018	-0,002	0,015
Exper_aortic aneurism	-0,075*	0,04	-0,074*	0,04	-0,085**	0,034	-0,067*	0,038	-0,067*	0,039	-0,073**	0,03
Exper_atrial fibrillation	-0,02	0,022	-0,018	0,021	**-0,028**	**0,02**	-0,027	0,021	-0,027	0,021	**-0,038****	**0,017**
Exper_carotis disease	-0,015	0,018	-0,016	0,018	-0,029*	0,016	-0,017	0,018	-0,017	0,018	-0,03*	0,015
Exper_coronary artery disease	0,008	0,022	0,009	0,021	0,025	0,018	0,008	0,022	0,008	0,022	0,023	0,017
Exper_chronic obstructive pulmonary disease	-0,013	0,019	-0,013	0,019	-0,014	0,016	-0,015	0,019	-0,015	0,018	-0,015	0,015
Exper_chronic heart failure	-0,005	0,012	-0,005	0,011	0,001	0,01	-0,004	0,012	-0,004	0,011	0,003	0,01
Exper_diabetes	0,005	0,023	0,004	0,023	0,013	0,02	0,007	0,023	0,007	0,023	0,016	0,018
Exper_hypertension	0,024	0,016	0,024*	0,016	**0,024**	**0,014**	0,021	0,016	0,021	0,015	**0,022***	**0,013**
Exper_pulmonary hypertension	0,004	0,011	0,004	0,01	0	0,009	0,006	0,011	0,006	0,01	0,002	0,008
Exper_myocardial infarction	-0,011	0,019	-0,012	0,019	-0,013	0,016	-0,012	0,019	-0,012	0,019	-0,013	0,015
Exper_porcelain aorta	-0,013	0,026	-0,011	0,025	-0,038	0,021	-0,01	0,026	-0,01	0,026	-0,034	0,02
Exper_renal failure	0,011	0,017	0,012	0,017	-0,003	0,015	0,008	0,017	0,008	0,017	-0,006	0,014
Exper_percutaneous coronary intervention	-0,006	0,02	-0,006	0,02	-0,017	0,018	-0,008	0,02	-0,008	0,02	-0,02	0,017
Exper_pacemaker	-0,008	0,022	-0,009	0,022	0,011	0,018	-0,012	0,022	-0,012	0,022	0,005	0,017
Exper_stroke	0,036	0,027	0,037	0,027	0,01	0,024	0,035	0,027	0,035	0,027	0,008	0,022
Exper_nyhacat2	-0,028	0,039	-0,027	0,039	-0,029	0,031	-0,036	0,043	-0,036	0,038	-0,04	0,028
Exper_nyhacat3	-0,031	0,033	-0,031	0,033	**-0,031**	**0,026**	-0,039	0,039	-0,039	0,032	**-0,042***	**0,023**
Exper_nyhacat4	-0,005	0,039	-0,003	0,038	-0,011	0,03	-0,012	0,043	-0,012	0,038	-0,02	0,028
Exper_mediastinal radiation	-0,006	0,047	-0,004	0,046	-0,002	0,04	-0,007	0,046	-0,007	0,047	-0,004	0,037
Exper_defibrillator	-0,126	0,118	-0,126	0,118	-0,038	0,088	-0,112	0,106	-0,112	0,117	-0,026	0,082
Exper_CABG	0,01	0,026	0,011	0,026	-0,012	0,021	0,013	0,026	0,013	0,026	-0,006	0,02
Exper_valvesurg	0,043	0,054	0,043	0,054	0,05	0,051	0,056	0,054	0,056	0,052	0,067	0,046
Exper_CoreValve	0,019	0,02	0,018	0,019	0,016	0,016	0,019	0,019	0,019	0,02	0,019	0,015
Exper_transfemoral access	-0,026	0,017	-0,026	0,017	-0,042***	0,014	-0,027	0,017	-0,027	0,017	-0,044***	0,013
Patient- and procedure-specific characteristics	Yes		Yes		Yes		Yes		Yes		Yes	
Hospital fixed-effects	Yes		Yes		Yes		Yes		Yes		Yes	

Notes: For the sake of brevity, the coefficient estimates on the patient- and procedure-specific characteristics and the hospital indicators are not shown in the table above.

effect in all subgroups also splits up the selection effect of patients, providing more evidence of where the selection mainly takes place. The overall learning effect may indicate improved skills and knowledge as treating more patients improves patient outcomes.

Comparing these results with the Lasso results in Table 5 provides rather mixed evidence. While both methods single out subgroup effects as important factors for survival, there is no agreement on which subgroups are more relevant. In fact, the Lasso stresses the importance of experience with procedural-characteristics (type of valve implanted) and physician specific knowledge (defibrillation) for patient survival, whereas TG estimation identifies types of patients that are relevant for learning. This finding may result from the fact that in the Lasso, some of the variables pick up effects from others that are truly controlled for in the Theil-Goldberger method or from the inappropriateness of the summation constraint on the subgroups. As a consequence, we suggest to compare both Lasso and Theil-Goldberger mixed estimation results and to interpret them with care. These results should then be further discussed and investigated by policy makers and practitioners to improve survival.

5 Robustness checks

Throughout the text, a range of methods is used to assess the quality of different results. In this subsection we verify the consequences of changes in the sample size. In table 7 we remove the four largest hospitals one by one from the regressions with 2-year survival as dependent variable. The results from these removals are to a large extent similar to the original regressions shown above in table 2. In particular, for most regressions we find a significant positive effect of cumulative experience on 2-year survival. At the same time, there is no evidence for learning from recent experience and static learning on 24-month survival as essentially none of the coefficients is statistically significant different from zero.

Table 7: Robustness Checks I:
Exclusion of the Largest Hospitals

Outcome Variable: 24-Month Survival				
Excluding	**Hospital 12**	**Hospital 17**	**Hospital 19**	**Hospital 20**
Annual Volume	-0.0063*	-0.0047	-0.0033	-0.0048
	(0.0033)	(0.0032)	(0.0033)	(0.0036)
Patient Number	0.0020**	0.0019**	0.0016	0.0016*
	(0.0008)	(0.0008)	(0.0010)	(0.0009)
Time Since Last Procedure (Days)	-0.0003	-0.0002	-0.0007	-0.0005
	(0.0007)	(0.0007)	(0.0007)	(0.0007)
Zero Days Since Last Procedure	-0.0271	-0.0134	-0.0556	-0.0448
	(0.0422)	(0.0422)	(0.0441)	(0.0434)
Patient- and Procedure-specific characteristics	Yes	Yes	Yes	Yes
Hospital fixed-effects	Yes	Yes	Yes	Yes
Number of Observations	708	707	659	680

Notes: For the sake of brevity, the coefficient estimates on the patient- and procedure-specific characteristics and the hospital indicators are not shown in the table above. Heteroscedasticity robust standard errors in parentheses: *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$.

Compared to 24-month survival, major adverse cardiac events regressions are more vulnerable to minor changes in the sample size (table 8). We find that for renal failure, the effect disappears when cases that were more than 100 or 150 days apart were removed. This implies that the overall result is largely driven by very strong outliers that were more than 150 days apart in treatment. For stroke, the result of human capital depreciation is not found for observations with fewer than 100 days' difference. However, including observations with less than 150 days between procedures, the effect is sizeable. This means that once there are about 100 days between procedures, the probability of suffering from a stroke during the operation is strongly increased. These two events strongly point towards non-linearity of human capital depreciation on MACE. As Theil-Goldberger mixed estimation is not fully appropriate for this kind of effects, we solely apply TG to 24-months survival.

Table 8: Robustness Checks II:
Varying Time Since Last Procedure

Outcome Variable	**MACE Renal Failure**			**MACE Stroke**		
	Overall	**Timediff < 100**	**Timediff < 150**	**Overall**	**Timediff < 100**	**Timediff < 150**
Specification	(1)	(2)	(3)	(1)	(2)	(3)
Annual Volume	0.0007	0.0008	0.0005	-0.0012	-0.0015	-0.0012
	(0.0025)	(0.0025)	(0.0025)	(0.0011)	(0.0011)	(0.0011)
Patient Number	-0.0001	-0.0001	-0.0001	0.0002	0.0003	0.0002
	(0.0007)	(0.0007)	(0.0007)	(0.0004)	(0.0004)	(0.0004)
Time Since Last Procedure (Days)	0.0012**	0.0009	0.0012*	0.0007**	0.0008	0.0012**
	(0.0005)	(0.0008)	(0.0007)	(0.0004)	(0.0005)	(0.0006)
Zero Days Since Last Procedure	0.0518	0.0460	0.0492	0.0403**	0.0430**	0.0545**
	(0.0336)	(0.0359)	(0.0344)	(0.0194)	(0.0218)	(0.0224)
Patient- and Procedure-specific characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Hospital fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	696	677	688	696	677	688

Notes: For the sake of brevity, the coefficient estimates on the patient- and procedure-specific characteristics characteristics and the hospital indicators are not shown in the table above. Heteroscedasticity robust standard errors in parentheses: *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$.

Whereas a broad range of results is provided, several concerns remain. *First*, while in the Belgian case there is little evidence for a causal relationship from outcome to volume, it would have been better to include this in the analysis. To remove the endogeneity bias from selective referrals, the literature employs instrumental variable methods. However, as our focus lies on the subgroup analysis which includes a lot of experience variables, the number of instruments required would make IV methods practically infeasible. *Second*, assuming effects to be linear imposes a heavy strain on the analysis. If in fact the overall learning curve would be non-linear, the specific structure of the experience variables may pick up these non-linearities. Graphical intuition on this argument is provided in Appendix C. The combination of the Theil-Goldberger and logistic models or a lot of quadratic terms is practically infeasible and there-fore this limitation remains. Nevertheless, a linear approach provides a useful first insight in the decomposition of learning curves.

6 Conclusion

In the last decades, a whole strand of the literature has contributed to learning, volume and scale effects in healthcare provision. In this paper we explore both overall, as well as subgroup learning curves using information on the first 860 Transcatheter Aorta Valve Implantations (TAVI) in

Belgium. Considering overall learning, we distinguish between static learning, learning from cumulative experience and learning from recent experience and assess their role for patient survival and adverse cardiac events during the TAVI procedure. Overall, our analysis shows that different types of learning apply for different outcomes: while cumulative experience is of great importance for 24-month and 36-month survival, more frequent practice plays a key role for adverse events like renal failure and stroke.

In addition, we attempt to extend the existing literature by exploring subgroup learning effects which provide an extra instrument to potentially improve and explain provider performance. Knowing that certain groups of patients contribute to the learning process gives more detailed information for both policy makers and healthcare providers. The extra information should be considered as a preliminary step between no volume-oriented policy and harsher measures such as concentration of procedures in only a limited amount of hospitals. Underlying the overall effects of an increasing number of patients on outcomes are subgroup learning effects for experience with using CoreValves replacement valves, hypertension, aortic aneurysm and physicians experience with defibrillators – to name a few. Trying to improve processes of care, these groups or techniques should be closely investigated. Increased knowledge from subgroups and the transfer of this knowledge to other patients or a change in patient selection for TAVI may drive these results.

References

Oluseun O Alli, Jeffrey D Booker, Ryan J Lennon, Kevin L Greason, Charanjit S Rihal, and David R Holmes. Transcatheter aortic valve implantation: assessing the learning curve. *JACC: Cardiovascular Interventions*, 5(1):72–79, 2012.

Ben Bridgewater, Antony D Grayson, John Au, Ragheb Hassan, Walid C Dihmis, Chris Munsch, and Paul Waterworth. Improving mortality of coronary surgery over first four years of independent practice: retrospective examination of prospectively collected data from 15 surgeons. *British Medical Journal*, 329(7463):421, 2004.

Lisa M Caputo, Kristin M Salottolo, Denetta Sue Slone, Charles W Mains, and David Bar-Or. The relationship between patient volume and mortality in american trauma centres: a systematic review of the evidence. *Injury*, 45(3):478–486, 2014.

Andrea Discacciati, Nicola Orsini, Sander Greenland, et al. Approximate bayesian logistic regression via penalized likelihood by data augmentation. *Stata Journal*, 15(3):712–736, 2015.

Bradley Efron, Trevor Hastie, Iain Johnstone, Robert Tibshirani, et al. Least angle regression. *The Annals of statistics*, 32(2):407–499, 2004.

Martin Gaynor, Harald Seider, and William B Vogt. The volume-outcome effect, scale economies, and learning-by-doing. *American Economic Review*, pages 243–247, 2005.

Alexander Genkin, David D Lewis, and David Madigan. Large-scale bayesian logistic regression for text categorization. *Technometrics*, 49(3):291–304, 2007.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *Unsupervised learning*. Springer, 2009.

Vivian Ho. Learning and the evolution of medical technologies: the diffusion of coronary angioplasty. *Journal of Health Economics*, 21(5):873–885, 2002.

Jason M Hockenberry and Lorens A Helmchen. The nature of surgeon human capital depreciation. *Journal of health economics*, 37:70–80, 2014.

PJE Holt, JD Poloniecki, IM Loftus, JA Michaels, and MM Thompson. Epidemiological study of the relationship between volume and outcome after abdominal aortic aneurysm surgery in the uk from 2000 to 2005. *British journal of surgery*, 94(4):441–448, 2007.

Robert S Huckman and Gary P Pisano. The firm specificity of individual performance: Evidence from cardiac surgery. *Management Science*, 52(4):473–488, 2006.

Marco D Huesch. Learning by doing, scale effects, or neither? cardiac surgeons after residency. *Health services research*, 44(6):1960–1982, 2009.

Marco D Huesch and Mariko Sakakibara. Forgetting the learning curve for a moment: how much performance is unrelated to own experience? *Health economics*, 18(7):855–862, 2009.

Hratch L Karamanoukian, Anthony L Panos, Jacob Bergsland, and Tomas A Salerno. Perspectives of a cardiac surgery resident in-training on off-pump coronary bypass operation. *The Annals of thoracic surgery*, 69(1):42–45, 2000.

Jörg Kempfert, Ardawan Rastan, David Holzhey, Axel Linke, Gerhard Schuler, Friedrich Wilhelm Mohr, and Thomas Walther. The learning curve associated with transapical aortic valve implantation. *Annals of Cardiothoracic Surgery*, 1(2):165–171, 2012.

Edward E Leamer. Multicollinearity: a bayesian interpretation. *The review of economics and statistics*, pages 371–380, 1973.

Kris CL Lee, Kannan Sethuraman, and Jongsay Yong. On the hospital volume and outcome relationship: Does specialization matter more than volume? *Health services research*, 50(6): 2019–2036, 2015.

Kazuhide Matsushima, Eric W Schaefer, Eugene J Won, Scott B Armen, Matthew C Indeck, and David I Soybel. Positive and negative volume-outcome relationships in the geriatric trauma population. *JAMA surgery*, 149(4):319–326, 2014.

Nicolai Meinshausen. Relaxed lasso. *Computational Statistics & Data Analysis*, 52(1):374–393, 2007.

Roos Mesman, Gert P Westert, Bart JMM Berden, and Marjan J Faber. Why do high-volume hospitals achieve better outcomes? a systematic review about intermediate factors in volume–outcome relationships. *Health Policy*, 119(8):1055–1067, 2015.

Helge Möllmann, Won-Keun Kim, Jörg Kempfert, Thomas Walther, and Christian Hamm. Complications of transcatheter aortic valve implantation (tavi): how to avoid and treat them. *Heart*, 101(11):900–908, 2015.

DS Moore, GP McCabe, and BA Craig. Table b: random digits. *Introduction to the practice of statistics*, 2009.

Trevor Park and George Casella. The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686, 2008.

Michael D Pasquale, Andrew B Peitzman, Jeffrey Bednarski, and Thomas E Wasser. Outcome analysis of pennsylvania trauma centers: factors predictive of nonsurvival in seriously injured patients. *Journal of Trauma and Acute Care Surgery*, 50(3):465–474, 2001.

Subbu Ramanarayanan. Does practice make perfect: An empirical analysis of learning-by-doing in cardiac surgery. *Mimeo*, 2008.

John C Robertson and Ellis W Tallman. Vector autoregressions: forecasting and reality. *Economic Review-Federal Reserve Bank of Atlanta*, 84(1):4, 1999.

Christopher A Sims. Dummy observation priors revisited. *Manuscript, Princeton University*, 2005.

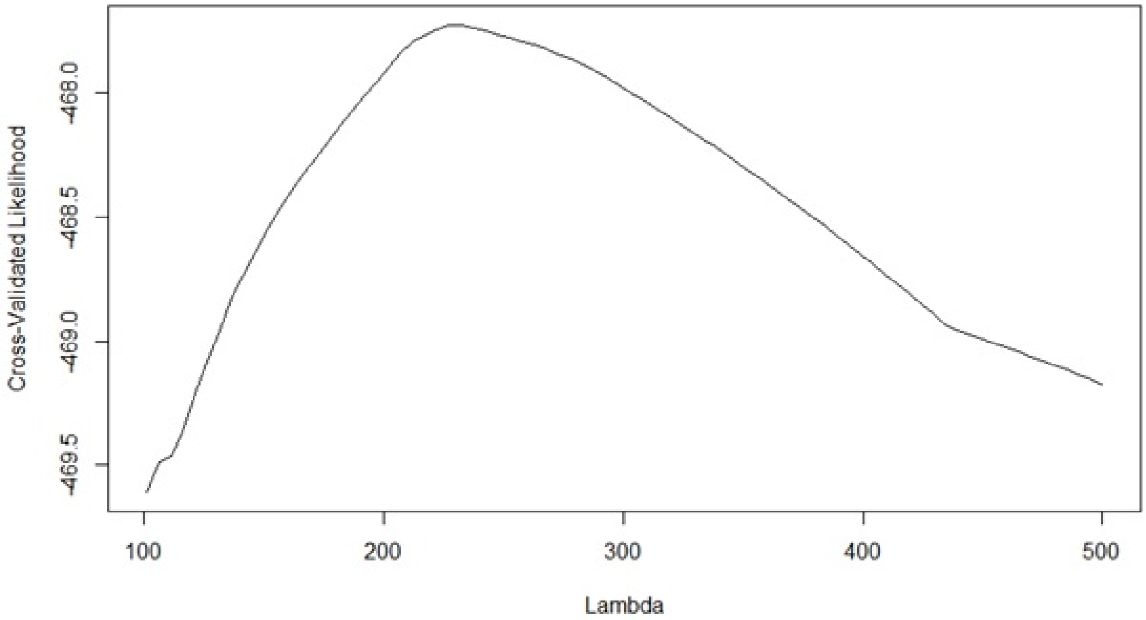
Henri Theil. On the use of incomplete prior information in regression analysis. *Journal of the American Statistical Association*, 58(302):401–414, 1963.

Henry Theil and Arthur S Goldberger. On pure and mixed statistical estimation in economics. *International Economic Review*, 2(1):65–78, 1961.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

Appendix A: Logistic Lasso Plot

The plot is obtained from the R-function `profL`. The likelihood is maximized for a value of $\lambda = 232$.



Appendix B: Stata/Mata code implementing the Theil-Goldberger mixed estimation method

The codes are available upon request.

Appendix C: Visual Intuition on Non-linear Effects

The left plot is an added variable plot that shows the curvature in the effect of patient number on 2-year mortality. The right graph shows the cubic-like relationship between, e.g., patient number and the experience variable for renal failure. As such, the experience on renal failure may pick up the non-linearity of the overall patient number when assuming linearity.