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

Raf Van Gestel, Tobias Müller,
Johan Bosmans

17-10

February 2017

DISCUSSION PAPERS

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

February 17, 2017

Abstract

Learning curves in health are of interest for a wide range of medical disciplines, healthcare providers and policy makers. In this paper, we distinguish between three types of learning when identifying overall learning curves: economies of scale, learning from cumulative experience and human capital depreciation. In addition, we approach the question of how treating more patients with specific characteristics predicts provider performance. 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. We find that treating an additional TAVI patient is associated with an increase in the probability of 2-year survival by about 0.16%-points. For adverse events like renal failure and stroke, we find that an extra day between procedures is associated with an increase in the probability for these events by 0.12%-points and 0.07%-points respectively. Furthermore, we find evidence for positive learning effects from physicians' experience with defibrillation, treating patients with hypertension and the use of certain types of replacement valves during the TAVI procedure.

[*]Department of Economics, Prinsstraat 13, 2000 Antwerpen, Belgium, email: raf.vangestel@uantwerpen.be

[†]Seminar of Health Economics, Frohburgstrasse 3, 6002 Luzern, Switzerland, email: tobias.mueller@unilu.ch

[‡]Department of Cardiology, Wilrijkstraat 10, 2650 Edegem, Belgium, email: johan.bosmans@uantwerpen.be

[§]On behalf of the Belgian TAVI Registry participants.

1   Introduction

The idea of a link between the volume of treated patients and healthcare provider performance 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 explaining 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 only a subset of these effects. In this paper, we analyze multiple factors simultaneously and we emphasize the potential role of patient subgroups in the learning process. Common policies are volume thresholds for hospitals, report cards and team/provider training (Huesch and Sakakibara, 2009). The approach followed in this paper provides more information on where improvements may be made by healthcare providers and this might improve the quality of team and provider training.

In this paper, we identify three types of learning: economies of scale, cumulative experience and human capital depreciation. Economies of scale refer to total volume effects captured by annual hospital patient volumes. The rationale here is that hospitals with more patients are likely to be better equipped and to have more standardized procedures potentially mapping into better provider performance. Economies of scale have been found to be important for both CABG (coronary artery bypass graft) and PTCA (percutaneous transluminal coronary angioplasty) in the literature (Ho, 2002; Gaynor et al., 2005). Cumulative experience[1] refers to the number of patients that have been treated in the past capturing classical learning by doing effects. Finally, human capital depreciation evaluates the role of time since the last procedure. For CABG and PTCA, cumulative learning has been found to only play a minor role (Ho, 2002; Gaynor et al., 2005). In

---

[1] In what follows, we use the terms "cumulative experience" and "cumulative learning" interchangeably.

contrast, human capital depreciation seems to be important for CABG (Hockenberry and Helmchen, 2014).

In line with the existing literature, we disentangle these overall learning effects in a first step. Subsequently, we follow a data-driven approach to detect learning mechanisms in patient subgroups. Given that performance is predicted by economies of scale, cumulative experience and/or human capital depreciation, these effects may well be driven by particular patient subgroups. More specifically, we assess how treating more patients with certain subgroup characteristics predicts overall health outcomes. Quantifying such information goes beyond the typical volume-outcome relationship and potentially imparts new insights on how to improve health outcomes for new, under-performing or lower volume providers and possibly allows to better identify the sources of learning. For typical outcomes like in-hospital mortality, physicians may have a general feeling on how to improve performance. This is however much less the case for other outcomes like long-term mortality or procedural characteristics.

Existing studies mostly explore patient subgroups from a different point of view, namely how overall volumes influence specific patient subgroups. In contrast to that, in this paper we asses how overall learning effects can be attributed to certain subgroups and therefore providing information for policy makers on how to concentrate procedures in hospitals. An example related to subpopulation effects in trauma care is Matsushima et al. (2014), where larger volumes of geriatric patients were associated with lower mortality and complications among geriatric patients. Larger non-geriatric volumes on the other hand were associated with higher odds of major complications. In Pasquale et al. (2001), higher- (overall) 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[2]. For TAVI, experience has been shown to have an impact on mortality (30-day and 1-year), the duration of procedures, contrast

---

[2]The very first patient underwent TAVI in Belgium in 2007.

volumes and radiation (Möllmann et al., 2015; Alli et al., 2012; Kempfert et al., 2012). Although these findings are based on descriptive analysis, they provide suggestive evidence. In this study, we go beyond simple univariate/descriptive analysis as we isolate the learning effects at hand by controlling for a broad range of patient- and procedure-specific characteristics, as well as hospital fixed-effects[3].

Overall we find that different learning processes apply for different health-related outcomes: While cumulative experience significantly predicts 24-month survival, human capital depreciation has a significant effect on several Major Adverse Cardiac and Cerebrovascular Events (MACE). These events occur as a result of procedural complications. In particular, our results suggest an increase in 24-month survival of 0.16%-points for every extra TAVI patient treated. Furthermore, the likelihood of having a 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 an issue when estimating overall learning curves, it is even more problematic in the estimation of subgroup effects. By treating more patients overall, by definition, also more patients are treated within specific subgroups. Therefore, an 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. 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 improves the information to identify subgroup learning effects. Subgroup learning effects for 24-month survival are found for patients with aortic aneurysm, atrial fibrillation, carotis disease, hypertension, porcelain aorta, NYHA category three and transfemoral access. That is, treating more patients with

---

[3]We do not claim causality regarding our estimated learning effects as the data does not allow us to completely rule out reverse causality and potential patient or provider self-selection.

these characteristics positively or negatively influences 24-month survival. These effects may be attributed to improved knowledge, knowledge transfer or to selection effects in these subgroups.

The remainder of the paper is organized as follows: In section 2, we discuss the data at hand (the Belgian TAVI registry) and the variables used in the analysis. 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 uses the Belgian TAVI registry which contains detailed information on the first 860 patients undergoing TAVI in Belgium in 23 different centers between 2007 and 2012. The data has been collected at the participating centers and has been approved by the institutional ethics committees. The registry holds a wide range of control variables on patient- and procedure-specific characteristics, as well as hospital identifiers. In each hospital, the TAVI procedure is only executed by one specialized team, but we do not have information on physician or hospital characteristics[4]. Whereas the workload with respect to TAVI steadily increased over time, the workload of TAVI during our sample period was relatively limited. We have detailed information on the demographic background of patients, comorbidities, indicators for the severity of the cardiac problem and procedural characteristics (see section 3.1 for more details). The patient outcomes we study are 24-month survival, as well as indicators for major adverse cardiac events (MACE) including renal failure, pacemaker implantation and stroke. Renal failure is known to be related to the use of contrast volumes during the TAVI procedure. Furthermore, stroke and pacemaker implantation are both typical complications in cardiac procedures and surgeries. Pacemaker implantation is known to be strongly related to the type of valve that is used.

---

[4]Physician and hospital characteristics have shown to be significant predictors of health outcomes (see e.g. Ho (2002)) and would therefore ideally be included in the analysis to further address patient and provider selection issues.

Two brands of valves are used in Belgian hospitals: CoreValve and SAPIEN and each Belgian hospital only uses one of the two brands. More detailed descriptive statistics are shown in the appendix.

Figure 1 below shows a positive relationship between the number of TAVI patients treated and 2-year survival providing first evidence for learning from cumulative experience (see top left graph). The raw data is overlaid with linear and quadratic prediction plots (dashed and solid dark blue lines) and the gray areas show the 95% confidence intervals for the quadratic fits. Note also that the quadratic and linear fits are nearly identical pointing toward a linear relationship between 2-year survival and cumulative experience. In addition, there is clearly a positive association between pacemaker implantations and cumulative experience (top right graph). This suggests that as more patients are treated, they are more likely to receive a pacemaker during the TAVI procedure. This finding is most likely driven by the use of CoreValve valves in the larger centers because this type of valve is known to be associated with pacemaker implantation. In contrast to that, the rate of adverse events for stroke and renal failure is roughly constant across all experience levels suggesting no volume-outcome relationship.

— Insert Figure 1 here —

In line with the findings of Hockenberry and Helmchen (2014), it is expected that the probability of mortality or major adverse cardiac events increases 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 during procedures. We find evidence for such negative effects of human capital depreciation 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 having a stroke or suffering from renal failure (see bottom left and right graphs). 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[5].

— Insert Figure 2 here —

3 Methodology and substantive questions

In a first step, we focus on overall learning curves in long-term patient survival and major adverse cardiac events (MACE) including renal failure, stroke and pacemaker implantations. Subsequently, we broaden our scope to learning curves for specific patient subgroups. This evokes two substantive questions: First, if more patients are treated overall, are specific subgroups heatlhier on average? Second, when treating more patients in specific subgroups, do providers get better in their overall care provision? Unlike the existing literature [e.g. Matsushima et al. (2014)], this paper focuses on the second question as it provides useful information to transfer knowledge between practitioners and possibly imparts new knowledge on how to improve health outcomes for under-performing or lower volume providers.

3.1 Overall Learning Curves

We estimate linear probability models (LPM) to identify overall learning effects. Following Huesch and Sakakibara (2009), we distinguish between three types of learning: economies of scale (Huckman and Pisano, 2006; Gaynor et al., 2005; Ho, 2002), learning from cumulative experience (Ho, 2002; Karamanoukian et al., 2000) and human capital depreciation (Hockenberry and Helmchen, 2014; Ramanarayanan, 2008; Huckman and Pisano, 2006). Using patient-level data, we estimate models of the following form:

$$Outcome_{i,h,t} = \beta_0 + \beta_1 Economies\_of\_Scale_{h,t} + \beta_2 Cumulative\_Experience_{i,h,t}$$
$$+ \beta_3 Human\_Capital\_Depreciation_{i,h,t} + \alpha'_1 X_{i,h,t} + \theta'_1 H_h + \varepsilon_{i,h,t}$$
$$(1)$$

---

[5]See the appendixes A, B and C for more descriptive statistics of the data at hand.

Our outcome variables are binary indicators for 24-month survival and MACE indicators for pacemaker implantation, renal failure and stroke for patient i treated in hospital h in year t. $Economies\_of\_Scale_{h,t}$ 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 possibly have improved processes of care and more standardized procedures (Gaynor et al., 2005; Ho, 2002). $Cumulative\_Experience_{i,h,t}$ is the patient number for individual i in hospital h in year t reflecting learning from cumulative experience (Ho, 2002). This variable indicates if the treatment of additional patients predicts provider performance. $Human\_Capital\_Depreciation_{i,h,t}$ 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 human capital depreciation effect. It is sensible that the longer the time between procedures, the more skills suffer from the absence of practice (Hockenberry and Helmchen, 2014).

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[6] categories, ejection fraction, aortic valve area, peak and mean gradient), as well as procedure-specific characteristics (type of valve and size of valve implanted). 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 health-related outcomes of interest. Finally, $\varepsilon_{i,h,t}$ is a classical error term capturing all unobserved factors such as genetic endowment and health behaviors of patients that also explain our outcomes of

---

[6]The New York Heart Association (NYHA) Classification is a frequently used measure to indicate the severeness of cardiac disease and the extent of heart failure.

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. Similar to Ho (2002), adding year fixed-effects likely results 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.

3.2 Multicollinearity and Subgroup Learning Curves

In the subgroup analysis, all variables from the overall analysis are retained and experience variables are introduced for all background characteristics. That is, if for example the 30th patient for a provider (hospital) is the 15th patient with hypertension for the same provider, the patient gets patient number 30 (cumulative experience) and experience for hypertension 15. Statistically these variables are likely to be strongly correlated[7]. 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. Therefore, we explore the use of the Least absolute shrinkage and selection operator (Lasso) to obtain an optimal subset of experience variables predicting our health-related outcomes of interest.

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 paper, 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

---

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

motivations, they are both applications of constraints in a regression analysis. This is discussed in more detail throughout the next sections.

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. These methods improve prediction accuracy and interpretation but come at the cost of biased estimates (Hastie et al., 2009). 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. 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 and Elastic Net 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:

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

where the first part of equation 2 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 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 set more coefficients exactly equal to zero (Tibshirani, 1996). 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 are obtained from the Least Angle Regression Selection (LARS) algorithm in which the optimal set of coefficients is the one

where Mallows' Cp reaches a minimum.

The Lasso singles out the most significant variables that predict health outcomes. 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 a subset of optimal predictors and use them in standard OLS regressions. 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. Thirdly, we run logistic Lasso regressions as our outcomes of interest are binary. Similarly to the Least Squares Lasso, an L_1[8] penalty on the absolute values of coefficients can be introduced to logistic regression (Genkin et al., 2007). Fourthly, we also apply Elastic Net regression which is a hybrid between Lasso and Ridge regression and uses weighted L_1 and L_2 penalties. The benefit of Elastic Net is that it copes better with highly correlated regressors, i.e., among groups of highly correlated variables it singles out multiple variables whereas the Lasso only includes one variable.

3.2.2 Including prior information with Theil-Goldberger mixed estimation

Multicollinearity can be interpreted as the occurrence of "undominated uncertain prior information" (Leamer, 1973). This definition points out that including extra prior information might soften the multicollinearity problem. 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 is the sum of all underlying subgroup effects. 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).

The Theil-Goldberger coefficients and variances, where prior and data information are

---

[8]An L_1 penalty is a constraint on the sum of absolute coefficient values as in equation (2). Alternatively, an L_2 penalty is a constraint on squared coefficient values used in Ridge regression.

efficiently weighted in a GLS framework, are given by (Theil and Goldberger, 1961):

$$\hat{\beta} = [X'\Omega^{-1}X + R'\Psi^{-1}R]^{-1}[X'\Omega^{-1}y + R'\Psi^{-1}r] \qquad (3)$$

and

$$V(\hat{\gamma}) = [X'\Omega^{-1}X + R'\Psi^{-1}R]^{-1} \qquad (4)$$

X is a $n \times k$ matrix of observations on independent variables; Ω is the $n \times n$ variance-covariance matrix of error terms 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:

$$R = [1\ 1\ 0 \cdots 0\ 0] \qquad (5)$$

$$r = [\widehat{\beta_1 + \beta_2}] \qquad (6)$$

As such, equations (3) and (4) are the result of applying GLS to the following two equations:

$$y = X\beta + u \qquad (7)$$

and

$$r = R\beta + v \qquad (8)$$

Equation (7) 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 main dif-

ference with applying exact constraints is 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:

$$s_{ij} = \beta_0 + \beta_1 Cumulative\_Experience_{ij} + \beta_2 Characteristic\_1_{ij} + \varepsilon_{ij} \qquad (9)$$

s_{ij} stands for two-year survival for individual i in hospital j and $Cumulative\_Experience$ is the patient number of an individual (e.g. patient number 1 in hospital 20). In equation (9), the coefficient on $Cumulative\_Experience$, β_1, is consistently estimated using standard regression techniques. Now let us think of a second model:

$$\begin{aligned} s_{ij} = {} & \gamma_0 + \gamma_1 Cumulative\_Experience_{ij} + \gamma_2 Characteristic\_1_{ij} \\ & + \gamma_3 Exper\_Characteristic\_1_{ij} + \varepsilon_{ij} \end{aligned} \qquad (10)$$

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. In equation (9), 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 (10), the same increase in health outcome for every extra patient is given by γ_1 and γ_3. 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 (9) 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:

$$R = [1 \ avg(char1) \ 0 \cdots 0 \ 0] \tag{11}$$

and

$$r = [\hat{\beta}_1] \tag{12}$$

3.3 Endogeneity

In our context, endogeneity issues regarding the estimated learning effects may typically arise because of (*i*) selective referral or (*ii*) reverse causality. In principle, more experienced providers might be able to select desirable patients with a higher likelihood of survival and refer others to their colleagues thus creating a classical omitted variable bias problem (patient selection). In addition, if overall health outcomes for different health care providers are publicly known, then this might cause certain patients to select into specific hospitals (provider selection). We cannot perfectly address these selection issues in our analysis due to the lack of detailed physician and hospital characteristics ruling out a causal interpretation of the estimated learning effects. However, by splitting up the learning effect in subgroups, as illustrated in section 3.2, we capture part of the patient selection (and selective referral effect) in the subgroup analyses. Subgroup experience effects represent either a true learning effect or a selection effect regarding a specific subgroup, e.g., over time patients with a less severe degree of a characteristic could be more likely to be treated than others. In interpreting the results, it is important to stress that both effects are interesting in their own right, but that they cannot be empirically disentangled. Furthermore, because subgroup experiences are included, the overall cumulative experience effect is measured more accurately because the selection in these subgroups is no longer captured by the overall effect.

Regarding reverse causality, the existing literature provides mixed evidence on the di-

14

rection of causation: Gaynor et al. (2005) and Ho (2002) find that causality mainly runs from volume to outcome whereas Ramanarayanan (2008) finds that sicker patients may select higher volume providers. In the Belgian setting, there is very little public information on hospital quality, and even less on procedure related hospital quality. Moreover, mortality rates across hospitals are only available to practitioners and are anonymized.

4 Results

4.1 Overall Learning Curves

We estimate the overall learning effects using linear probability models (LPM) for 24-month survival, as well as several Major Adverse Cardiac Events (MACE) including pacemaker implantation, renal failure and stroke. Model one shows the plain overall learning effects for the three learning measures economies of scale, learning from cumulative experience and human capital depreciation; model two adds patient- and procedure-specific characteristics as described above in section 3.1; finally, in model three we also include hospital fixed-effects. In addition, we include a binary indicator for "Zero Days Since Last Procedure" in all model specifications because there are regularly two TAVI procedures scheduled on one day. We also replicated our findings using probit/logit specifications to relax the implicit linearity assumption in the marginal effects[9]

The estimated overall learning effects on survival can be found in table I below: First, we find a positive and significant effect on the cumulative experience indicator for 24-month survival pointing toward learning from doing across all three model specifications. Specification (3) suggests that treating an additional TAVI patient is associated with an increase in the probability of 2-year survival of about 0.16%-points, ceteris paribus. This effect is sizeable considering that patient volumes were increased on average by more than 10 patients per year in the timespan from 2007 to 2012. The cumulative learning effect

---

[9]We refrain from showing the probit/logit estimates as the average marginal probability effects are almost perfectly identical to the LPM estimates shown below.

can be interpreted as the result of a learning process in technical skills, but also in the selection process of patients by physicians. 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, providers might not yet be aware of the limited potential of the procedure.

Second, table I also provides weak evidence for economies of scale as models one and two show negative and significant coefficients on the annual patient volume measure. However, the effects is no longer significant when hospital fixed-effects are included in specification three. The hospital fixed-effects pick up volume effects persisting over time and therefore the yearly volume effects become insignificant when including hospital indicators. Third, our estimates do not indicate that human capital depreciation is associated with survival as our time difference indicators do not show any statistically significant coefficients.

— Insert Table I here —

In addition to the effects on patient survival, we also analyze adverse cardiac events. The results are summarized in table I above. While cumulative learning showed to be significant for survival, human capital depreciation is significant for several MACE. The likelihood of suffering a renal failure or a stroke during the procedure is significantly higher when more days have passed since the last TAVI procedure. The estimates suggest that the likelihood of renal failure after TAVI is 0.12%-points higher for every additional day since the last procedure. For stroke, an additional day since the last procedure is associated with an increase in the probability of suffering a stroke of 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

16

the team loses concentration during the course of a given day. As can be seen in the robustness section, the results on MACE should be interpreted with care as the results are sometimes driven by only a few extreme observations.

Summing up, we find that different types of learning apply for different outcomes: Learning from cumulative experience is relevant for 24-month survival and more frequent practice plays a key role for adverse events. Skills required for preventing these events may depreciate over time as illustrated for renal failure and stroke.

4.2   Subgroup Learning Curves

Knowing that different types of learning apply for different outcomes, it is interesting to investigate to what extent patient subgroups account for these overall learning effects. Table II below shows the estimates of a linear probability model (LPM) using 24-month survival as the dependent variable while controlling for all patient- and procedure-specific characteristics, as well as hospital fixed-effects. In addition, the model also includes all the experience variables for each of the background characteristics. Overall the results indicate both positive and negative experience effects: For example, treating more patients with diabetes is associated with an increase in 24-month survival, though statistically insignificant. Positive effects may exist because of actual learning within specific subgroups or because of positive patient selection over time. 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 examined to improve future health improvements. Finding negative subgroup 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 and thus per se lower probability of survival may be treated.

The table also clearly shows the underlying multicollinearity issue in the regression

when looking at column VIF which displays the Variance Inflation Factors[10]. Nearly all experience variables have VIF values higher than ten, indicating strongly inflated standard errors and rendering most subgroup learning effects statistically insignificant.

— Insert Table II here —

4.3   Lasso Regressions

The Lasso selects those variables most correlated to the dependent variable. Whereas it succeeds in singling out the most relevant 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 III below contains a range of Lasso-related specifications: In the first specification (1), we apply OLS on the Lasso selection of best predictors. This is done to reduce the bias in the estimated parameters and to facilitate interpretation. 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[11] is negatively associated with 24-month survival, though statistically insignificant. This negative effect may be caused by irrelevant knowledge transfer between subgroups or negative selection. Second, experience with defibrillation and using CoreValve (Typevalve=1, 0 for Sapiens) replacement valves on the other hand seems to be positively associated with patient survival. This positive relationship reflects again either a direct learning or a selection effect. Positive selection might also be interpreted as a form of learning as physicians become better at selecting patients with certain characteristics which have a

---

[10]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

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

higher probability of being alive after two years. 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. One possible explanation for this effect is the practice of clinical proctoring which is more extensive for Corevalve. In practice this means that Corevalve users receive a longer guidance from an experienced physician. Besides the subgroup experience variables, the Lasso also identifies the annual TAVI patient volumes (economies of scale) as a key negative predictor of 2-year survival. In column (2), the Elastic Net singles out more subgroup experience variables. Nevertheless, the original variables from Lasso are retained and they practically have the highest effects on 2-year survival[12]. Finally, the results in column (3) show the selection of optimal subsets of predictors for a logistic Lasso specification. The results again provide evidence for a positive subgroup learning effect of using CoreValve replacement valves. However, the logistic Lasso includes substantially fewer variables and adds several variables in comparison with the non-logistic Lasso. Whereas without shrinking, results between logistic and LPM's are mostly similar, this does not seem to hold when an L1 penalty is applied. The divergence may be due to different penalties in the ordinary and logistic lasso combined with the lasso adjustment to the lars algorithm (when a variable is changing signs, the coefficient is temporarily set to 0).

— Insert Table III here —

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[13]. Computation requires

---

[12]Results from including quadratic terms to capture non-linearities are qualitatively similar to results in table 3.

[13]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 available upon request, the `tgmixed` command ado file was used as a guide.

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 (3)-(4) in table IV below next to OLS where specification (1) shows LPM estimates using heteroscedasticity-robust standard errors and specification (2) shows generalized least squares (GLS) estimates. For the Theil-Goldberger estimation we implement robust standard errors (specification (3)) and a GLS form (specification (4)) 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 change from OLS to GLS contributes more in efficiency terms than TG estimation. In contrast, the effect of implementing the stochastic constraint on coefficients only has a limited impact on standard errors and significance. However, there are some differences in significance between models (2) and (4). Intuitively, the additional information that is added in the TG estimation seems rather limited. Nevertheless, although there are no qualitative differences between the OLS and TG specifications without GLS, we clearly observe differences in significance between the second and fourth specification (see grey shaded bars in table IV below): 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. 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

20

subgroup learning effects. This evidence for negative subgroup learning strongly suggests the presence of selection effects. For patients with these characteristics the severity of the characteristic changes over time such that they are more likely to die in two years' time.

— Insert Table IV here —

Summing up, the comparison of the the TG mixed estimation with the Lasso results in Table III provides mixed evidence. While both methods single out subgroup effects as important factors for survival, there is no agreement on which subgroups are more relevant. 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

While we have already assessed the robustness of our findings to different model specifications and estimation techniques, in this subsection, we check the robustness of our findings to changes in the sample size. In table V below, we remove the four largest hospitals one by one from the regressions with 2-year survival as the dependent variable. The results from these removals are almost perfectly identical to the original regressions shown above in table I. In particular, we find a positive and significant effect of cumulative experience on 2-year survival[14]. As above, there is no evidence for an effect of economies of scale or human capital depreciation on 24-month survival as essentially none of the coefficients is statistically significant different from zero.

---

[14]Only for hospital 19 the significance disappears. This is however caused by the drop in degrees of freedom rather than a change in coefficients as can be seen from the regression without hospital 20.

— Insert Table V here —

For adverse cardiac events displayed in table VI, we find that the significant effect of human capital depreciation for renal failure disappears when cases that were more than 100 days apart were removed mainly because of a sizeable drop in the coefficient size. However, the effect reappears if the sample is increased to procedures with less than 150 days time difference (see specification 2). The results for renal failure are therefore mainly driven by observations for which more than 100 days have passed since the team has last performed the TAVI procedure indicating that the negative effects of human capital depreciation only manifest themselves for relatively long time periods between procedures. Interestingly, the same pattern emerges when analyzing the effects on having a stroke: The coefficient on human capital depreciation becomes insignificant for observations with fewer than 100 days difference but turns significant once including all observations with less than 150 days between procedures. The insignificance can be attributed to a drop in degrees of freedom because the coefficient increases in size. In line with the results above, when the sample is restricted, we find a highly significant and positive effect on the "Zero days since last procedure" indicator in all subsamples pointing towards that the cardiology team loses concentration during the course of a day and therefore is more likely to cause medical errors when performing more than one procedure on a given day.

— Insert Table VI here —

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 explicitly address this issue in the analysis. To remove the endogeneity bias caused by selective referrals in the overall effect, the literature usually employs instrumental variable methods. However, we did not have any sensible instrument at our disposal. Because our focus lies on the subgroup analysis, also selection is informative to obtain knowledge to improve performance. The drawback of our method is

that we are unable to disentangle true learning from selection effects. Second, assuming effects to be linear imposes a heavy strain on the analysis. Including squared terms in the Lasso regressions did not qualitatively influence the results. For the Theil-Goldberger method, 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 mixed estimation method and logistic models is practically infeasible and therefore 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 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 economies of scale, learning from cumulative experience and human capital depreciation 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 survival, more frequent practice plays a key role for adverse events like renal failure and stroke.

In addition, we extend the existing literature by exploring subgroup learning effects which provides 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 to improve clinical practice. We apply both Lasso regression as a variable selection method and Theil-Goldberger mixed estimation to augment the data. Underlying the overall effects of treating more patients are subgroup learning effects for experience with using CoreValves

23

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 by both practitioners and policy makers.

Acknowledgments

References

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

Caputo, L. M., K. M. Salottolo, D. S. Slone, C. W. Mains, and D. Bar-Or
2014. The relationship between patient volume and mortality in american trauma centres: a systematic review of the evidence. *Injury*, 45(3):478–486.

Efron, B., T. Hastie, I. Johnstone, R. Tibshirani, et al.
2004. Least angle regression. *The Annals of statistics*, 32(2):407–499.

Gaynor, M., H. Seider, and W. B. Vogt
2005. The volume-outcome effect, scale economies, and learning-by-doing. *American Economic Review*, Pp. 243–247.

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

Hastie, T., R. Tibshirani, and J. Friedman
2009. *Unsupervised learning.* Springer.

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

Hockenberry, J. M. and L. A. Helmchen
2014. The nature of surgeon human capital depreciation. *Journal of Health Economics*, 37:70–80.

Holt, P., J. Poloniecki, I. Loftus, J. Michaels, and M. Thompson
2007. 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.

Huckman, R. S. and G. P. Pisano

2006. The firm specificity of individual performance: Evidence from cardiac surgery. *Management Science*, 52(4):473–488.

Huesch, M. D.

2009. Learning by doing, scale effects, or neither? cardiac surgeons after residency. *Health Services Research*, 44(6):1960–1982.

Huesch, M. D. and M. Sakakibara

2009. Forgetting the learning curve for a moment: how much performance is unrelated to own experience? *Health Economics*, 18(7):855–862.

Karamanoukian, H. L., A. L. Panos, J. Bergsland, and T. A. Salerno

2000. Perspectives of a cardiac surgery resident in-training on off-pump coronary bypass operation. *The Annals of Thoracic Surgery*, 69(1):42–45.

Kempfert, J., A. Rastan, D. Holzhey, A. Linke, G. Schuler, F. W. Mohr, and T. Walther

2012. The learning curve associated with transapical aortic valve implantation. *Annals of Cardiothoracic Surgery*, 1(2):165–171.

Leamer, E. E.

1973. Multicollinearity: a bayesian interpretation. *The Review of Economics and Statistics*, Pp. 371–380.

Lee, K. C., K. Sethuraman, and J. Yong

2015. On the hospital volume and outcome relationship: Does specialization matter more than volume? *Health Services Research*, 50(6):2019–2036.

Matsushima, K., E. W. Schaefer, E. J. Won, S. B. Armen, M. C. Indeck, and D. I. Soybel

2014. Positive and negative volume-outcome relationships in the geriatric trauma population. *JAMA Surgery*, 149(4):319–326.

Meinshausen, N.

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

Mesman, R., G. P. Westert, B. J. Berden, and M. J. Faber

    2015. Why do high-volume hospitals achieve better outcomes? a systematic review about intermediate factors in volume–outcome relationships. *Health Policy*, 119(8):1055–1067.

Möllmann, H., W.-K. Kim, J. Kempfert, T. Walther, and C. Hamm

    2015. Complications of transcatheter aortic valve implantation (tavi): how to avoid and treat them. *Heart*, 101(11):900–908.

Pasquale, M. D., A. B. Peitzman, J. Bednarski, and T. E. Wasser

    2001. Outcome analysis of pennsylvania trauma centers: factors predictive of non-survival in seriously injured patients. *Journal of Trauma and Acute Care Surgery*, 50(3):465–474.

Ramanarayanan, S.

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

Theil, H. and A. S. Goldberger

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

Tibshirani, R.

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

TABLES

Table I: Overall Learning Effects:

Survival & Major Adverse Cardiac Events (MACE)

Outcome Variable	24-Month Survival			MACE Pacemaker			MACE Renal Failure			MACE Stroke		
Specification	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Economies of Scale	-0.0040**	-0.0053**	-0.0044	0.0011	0.0018	-0.0018	-0.0009	-0.0004	0.0007	-0.0007	-0.0004	-0.0012
	(0.0020)	(0.0023)	(0.0031)	(0.0017)	(0.0018)	(0.0031)	(0.0015)	(0.0017)	(0.0025)	(0.0009)	(0.0008)	(0.0011)
Cumulative Experience	0.0013*	0.0013*	0.0016**	0.0012*	0.0013*	0.0009	0.0004	0.0004	-0.0001	0.0002	0.0003	0.0002
	(0.0006)	(0.0007)	(0.0008)	(0.0007)	(0.0007)	(0.0008)	(0.0005)	(0.0006)	(0.0007)	(0.0003)	(0.0003)	(0.0004)
Human Capital Depreciation	-0.0001	-0.0004	-0.0005	0.0000	0.0002	0.0003	0.0006	0.0009*	0.0012**	0.0002	0.0005	0.0007**
	(0.0006)	(0.0006)	(0.0006)	(0.0004)	(0.0004)	(0.0004)	(0.0006)	(0.0005)	(0.0005)	(0.0003)	(0.0003)	(0.0004)
Zero Days Since Last Procedure	-0.0134	-0.0342	-0.0396	0.0591*	0.0394	0.0497	0.0418	0.0482	0.0518	0.0165	0.0315*	0.0403**
	(0.0379)	(0.0390)	(0.0400)	(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	No	Yes	Yes
Hospital fixed-effects	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Number of Observations	854	780	780	754	692	691	758	697	696	759	697	696

Notes: Linear probability model estimates of the overall learning curves for 2-year survival and MACE indicators for pacemaker implantation, renal failure and stroke using three different model specifications: Model 1 includes no covariates; Model 2 then adds controls for several patient- and procedure-specific characteristics; Model 3 adds hospital fixed-effects. Heteroscedasticity-robust standard errors in parentheses. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$.

Table II: Subgroup Learning Effects:

LPM Estimates and Variane Inflation Factors

Outcome Variable: 24-Month Survival	Coefficient	VIF
Economies of Scale	-0.0084*	8.263
	(0.0047)	
Cumulative Experience	0.0388	9138.332
	(0.0488)	
Human Capital Depreciation	-0.0005	1.729
	(0.0006)	
Zero Days Since Last Procedure	-0.0427	1.586
	(0.0399)	
Experience_female	-0.0179	277.921
	(0.0172)	
Experience_angina	-0.0062	186.697
	(0.0181)	
Experience_aortic aneurism	-0.0750*	81.684
	(0.0398)	
Experience_artial fibrillation	-0.0197	172.285
	(0.0224)	
Experience_carotis disease	-0.0155	62.654
	(0.0182)	
Experience_coronary artery disease	0.0085	649.885
	(0.0217)	
Experience_coronary obstructive disease	-0.0127	96.161
	(0.0185)	
Experience_chronic heart failure	-0.0053	378.212
	(0.0116)	
Experience_diabetes	0.0050	123.441
	(0.0234)	
Experience_hypertension	0.0241	549.234
	(0.0156)	
Patient- and procedure specific characteristics	Yes	
Hospital fixed-effects	Yes	
Number of Observations	780	

Notes: For the sake of brevity, the table shows only a selection of the coefficients on the experience variables. Heteroscedasticity-robust standard errors in parentheses: *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$.

Table III: Subgroup Learning:

Variable Selection Methods

Outcome Variable: 24-Month Survival

Specification	(1) OLS on Lasso	(2) OLS on Elastic Net	(3) Logistic Lasso
Economies of Scale	-0.004*	-0.005**	-0.012
Human Capital Depreciation	-	-0.020	0.0002
Female	0.028	0.030	-
Atrial fibrillation	-0.054	-0.052	-
Carotis disease	0.113***	0.115***	-
Chronic obstructive pulmonary disease	-0.115***	-0.113***	-
Pulmonary hypertension	-0.069**	-0.077**	-
Porcelain aorta	-0.107*	-0.094	-
Renal failure	-0.074*	-0.080**	-
PCI	-	-0.026	-
Stroke	-0.101**	-0.099**	-
NYHAcat2	-	0.016	-
NYHAcat4	-0.038	-0.052	-
CoreValve	0.035	0.042	-
Transfemoral access	0.109***	0.103**	-
Medium valve	-0.112**	-0.096**	-
exp_pulmonary hypertension	-	0.003	-
exp_porcelain aorta	-0.006	-0.006	-
exp_renal failure	-	0.008*	-
exp_medrad	-	-0.004	-
exp_defibrillator	0.074*	0.039	-
exp_CoreValve	0.002*	0.007	0.006
exp_chronic heart failure	-	-	-0.006
exp_small valve	-	-0.008	0.015
exp_highef	-	-0.005	
Hospital Fixed Effects	No	No	No
Observations	780	780	780

Notes: Columns (1) and (2) show the OLS coefficient estimates on the Lasso and Elastic Net selection of optimal predictors for 2-year survival. Column (3) displays the Logistic Lasso selection of predictors: *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$.

Table IV: Subgroup Learning:

Theil-Goldberger Mixed Estimation

Outcome Variable: 2-Year Survival	OLS				Theil-Goldberger			
	Robust (1)		GLS (2)		Robust (3)		GLS (4)	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Economies of Scale	-0.008*	0.005	-0.004	0.003	-0.008*	0.005	-0.004	0.003
Cumulative Experience	0.039	0.049	0.071*	0.040	0.050	0.047	0.086**	0.036
Human Capital Depreciation	-0.001	0.001	0.000	0.001	-0.001	0.001	0.000	0.000
Zero Days Since Last Procedure	-0.043	0.040	-0.013	0.035	-0.043	0.040	-0.015	0.033
Exper_female	-0.018	0.017	-0.017	0.015	-0.019	0.017	-0.019	0.014
Exper_angina	-0.006	0.018	-0.006	0.016	-0.003	0.018	-0.002	0.015
Exper_aortic aneurism	-0.075*	0.040	-0.085**	0.034	-0.067*	0.039	-0.073**	0.030
Exper_atrial fibrillation	-0.020	0.022	-0.028	0.020	-0.027	0.021	-0.038**	0.017
Exper_carotis disease	-0.015	0.018	-0.029*	0.016	-0.017	0.018	-0.030*	0.015
Exper_coronary artery disease	0.008	0.022	0.025	0.018	0.008	0.022	0.023	0.017
Exper_chronic obstructive pulmonary disease	-0.013	0.019	-0.014	0.016	-0.015	0.018	-0.015	0.015
Exper_chronic heart failure	-0.005	0.012	0.001	0.010	-0.004	0.011	0.003	0.010
Exper_diabetes	0.005	0.023	0.013	0.020	0.007	0.023	0.016	0.018
Exper_hypertension	0.024	0.016	0.024*	0.014	0.021	0.015	0.022*	0.013
Exper_pulmonary hypertension	0.004	0.011	0.000	0.009	0.006	0.010	0.002	0.008
Exper_myocardial infarction	-0.011	0.019	-0.013	0.016	-0.012	0.019	-0.013	0.015
Exper_porcelain aorta	-0.013	0.026	-0.038*	0.021	-0.010	0.026	-0.034*	0.020
Exper_renal failure	0.011	0.017	-0.003	0.015	0.008	0.017	-0.006	0.014
Exper_pci	-0.006	0.020	-0.017	0.018	-0.008	0.020	-0.020	0.017
Exper_pacemaker	-0.008	0.022	0.011	0.018	-0.012	0.022	0.005	0.017
Exper_stroke	0.036	0.027	0.010	0.024	0.035	0.027	0.008	0.022
Exper_nyhacat2	-0.028	0.039	-0.029	0.031	-0.036	0.038	-0.040	0.028
Exper_nyhacat3	-0.031	0.033	-0.031	0.026	-0.039	0.032	-0.042*	0.023
Exper_nyhacat4	-0.005	0.039	-0.011	0.030	-0.012	0.038	-0.020	0.028
Exper_mediastinal radiation	-0.006	0.047	-0.002	0.040	-0.007	0.047	-0.004	0.037
Exper_defibriallator	-0.126	0.118	-0.038	0.088	-0.112	0.117	-0.026	0.082
Exper_CABG	0.010	0.026	-0.012	0.021	0.013	0.026	-0.006	0.020
Exper_valve surgery	0.043	0.054	0.050	0.051	0.056	0.052	0.067	0.046
Exper_CoreValve	0.019	0.020	0.016	0.016	0.019	0.020	0.019	0.015
Exper_transfemoral access	-0.026	0.017	-0.042***	0.014	-0.027	0.017	-0.044***	0.013
Patient- and procedure-specific characteristics	Yes		Yes		Yes		Yes	
Hospital fixed-effects	Yes		Yes		Yes		Yes	
Number of Observations	780		780		780		780	

Notes: OLS and Theil-Goldberger Mixed Estimation estimates of the overall and subgroup learning curves. 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: *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$.

Table V: Robustness Checks:

Exclusion of the Largest Hospitals

Outcome Variable: 24-Month Survival

Excluding	**Hospital 12**	**Hospital 17**	**Hospital 19**	**Hospital 20**
Economies of Scale	-0.0063*	-0.0047	-0.0033	-0.0048
	(0.0033)	(0.0032)	(0.0033)	(0.0036)
Cumulative Experience	0.0020**	0.0019**	0.0016	0.0016*
	(0.0008)	(0.0008)	(0.0010)	(0.0009)
Human Capital Depreciation	-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: *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$.

Table VI: Robustness Checks:

Varying Time Since Last Procedure

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

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: *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$.

FIGURES

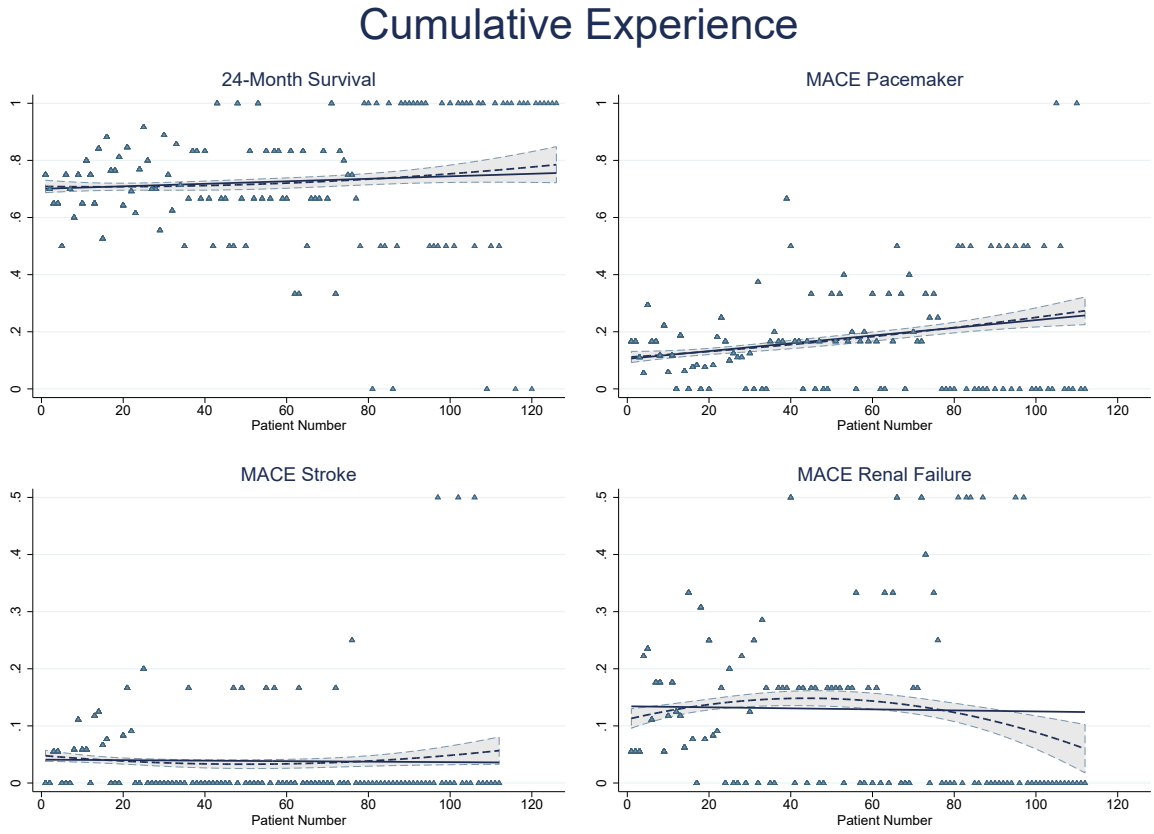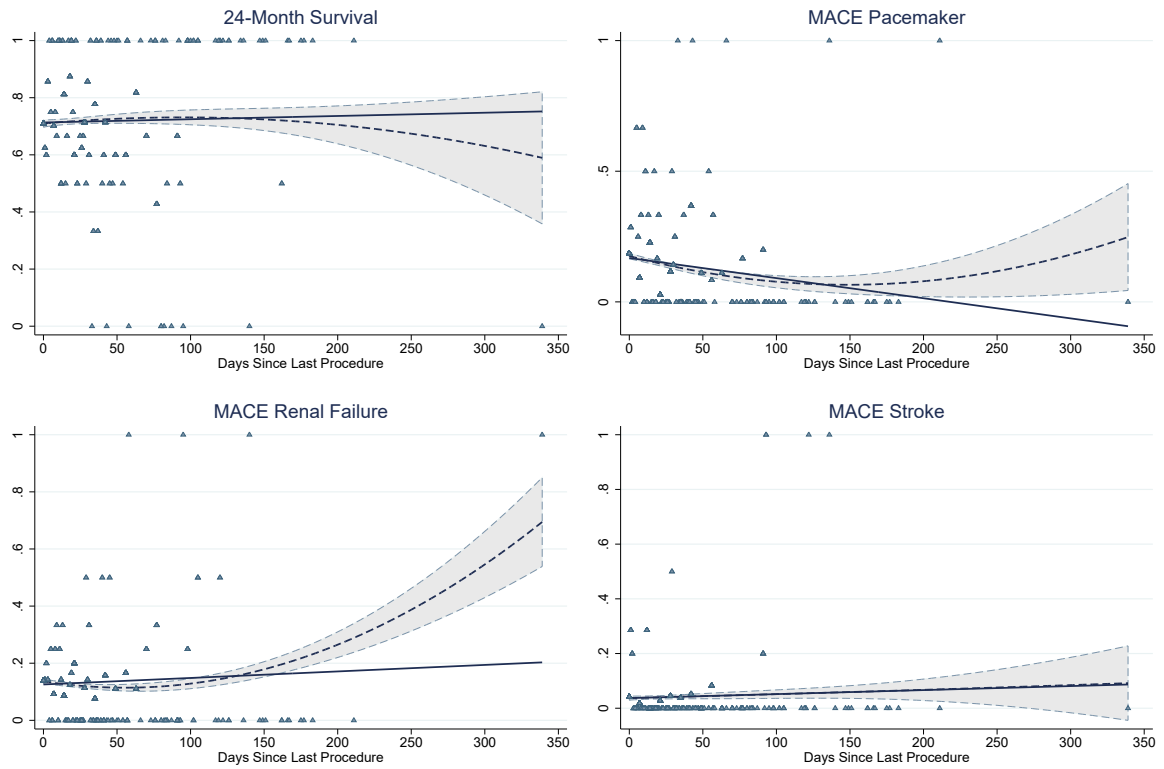Figure 1: Preliminary Learning from Cumulative Experience Curves

Figure 2: Preliminary Human Capital Depreciation Effects

# Human Capital Depreciation

Appendix A: Descriptive Statistics

Table VII: Summary Statistics

	Mean	S.D.	Min	Max	N
24-Month Survival Rate	0.71	0.45	0	1	780
Annual Volume	18.66	9.85	1	37	780
Time Since Last Procedure (Days)	21.4	32.82	0	339	780
Age	82.55	6.10	50	98	780
Female	0.53	0.5	0	1	780
Angina	0.29	0.45	0	1	780
Aortic Aneurism	0.05	0.22	0	1	780
Atrial Fibrillation	0.3	0.46	0	1	780
Carotid Artery Disease	0.2	0.4	0	1	780
Coronary Artery Disease	0.61	0.49	0	1	780
Chronic Obstructive pulmonary disease	0.29	0.45	0	1	780
Chronic Heart Failure	0.68	0.47	0	1	780
Diabetes	0.25	0.43	0	1	780
Hypertension	0.76	0.43	0	1	780
Pulmonary Hypertension	0.55	0.5	0	1	780
Heart Attack	0.22	0.42	0	1	780
Porcelain Aorta	0.08	0.27	0	1	780
Renal Failure	0.25	0.43	0	1	780
Pacemaker	0.13	0.34	0	1	780
Stroke	0.16	0.37	0	1	780
NYHA cat2	0.18	0.39	0	1	780
NYHA cat3	0.6	0.49	0	1	780
NYHA cat4	0.2	0.4	0	1	780
Mediastinal Radiation	0.05	0.21	0	1	780
Defibrillator	0	0.06	0	1	780
CABG	0.26	0.44	0	1	780
Valve Surgery	0.03	0.16	0	1	780
CoreValve	0.45	0.5	0	1	780
Transfemoral Access	0.76	0.43	0	1	780

Notes: Summary statistics for the full set of outcome variables and all patient- and procedure specific characteristics.

Appendix B: Economies of Scale

Figure 3: Preliminary Economies of Scale Effects

Economies of Scale
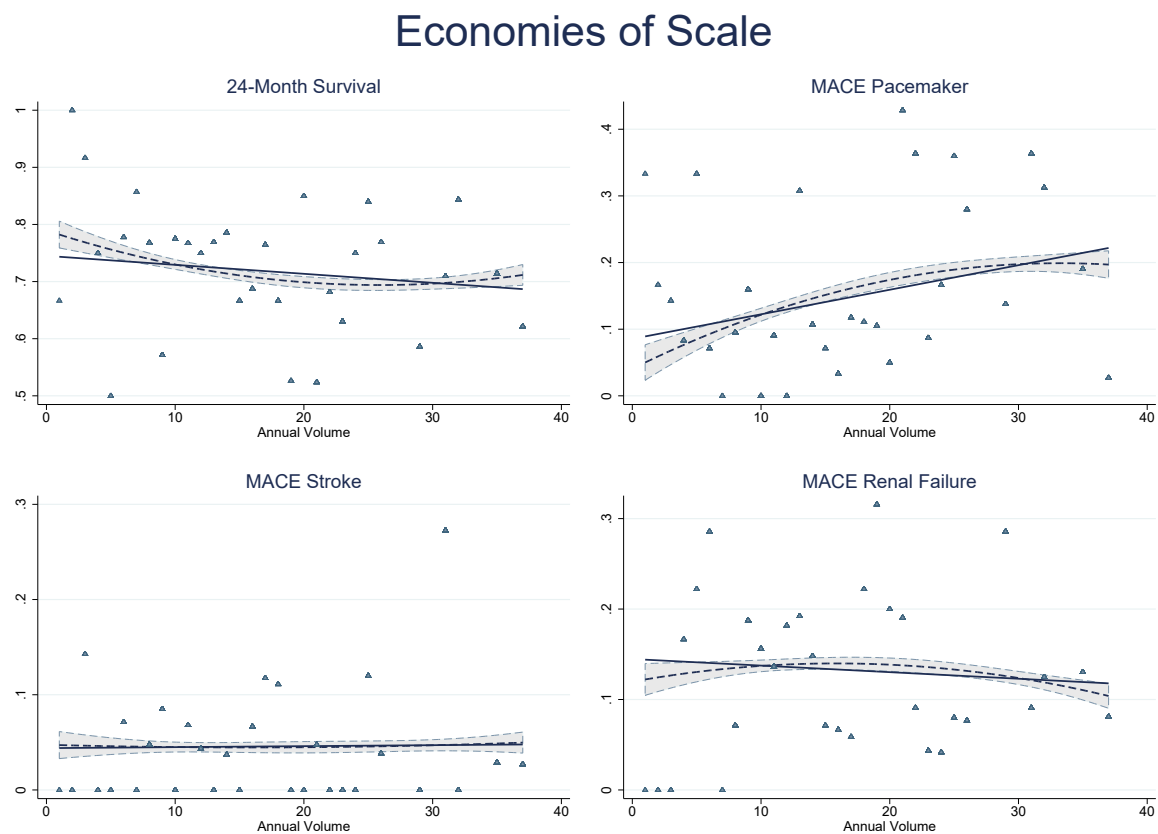
Appendix C: Days Since Last Procedure

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 4 below.

Figure 4: Histogram of Days Since Last TAVI Procedure
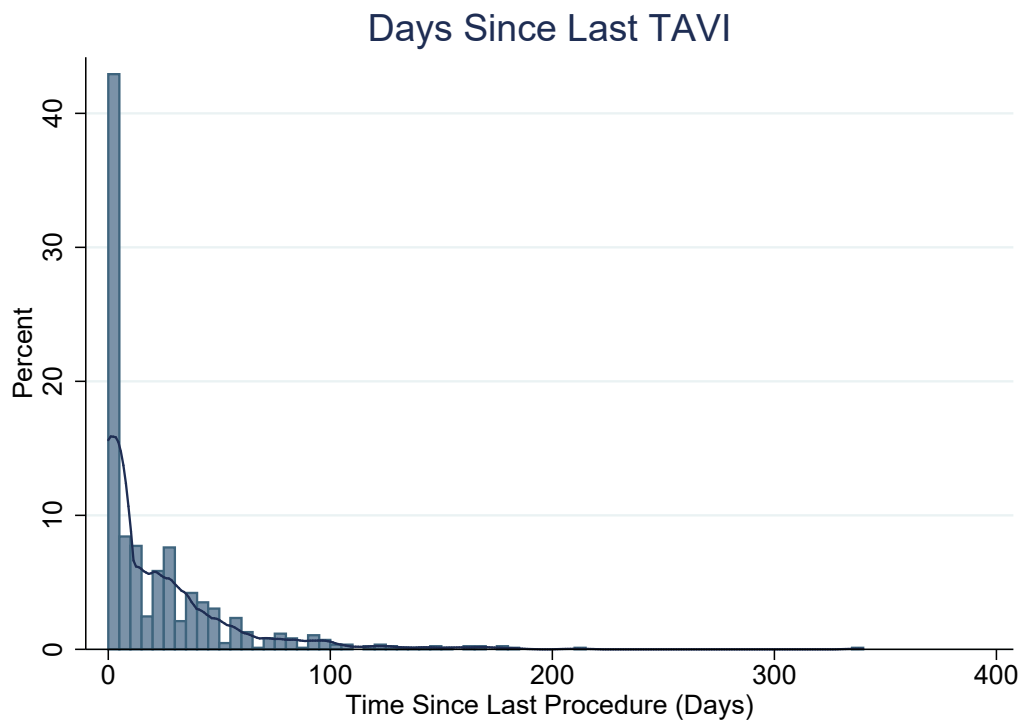
Notes: The Figure shows the distribution of the days since the last TAVI procedure and is overlaid with a kernel density estimate (solid dark blue line).

Appendix D: Visual Intuition on Non-linear Effects

The first plot below is an added variable plot showing the curvature in the effect of patient number on 2-year mortality. The added variable plot displays two clouds of observations. The clustering of observations in these two clouds signals a poor fit which results from the use of linear probability models. However, the use of LPM's instead of more sophisticated models is warranted because of the assumption of linear effects in determining the constraint. The matrix of graphs shows the relationship between patient number and the significant experience variables identified by the Theil-Goldberger estimation. If in reality the effect of patient number (or cumulative learning) would be non-linear, it might be that the other experience variables become significant depending on their relationship to the patient number. If the pattern of this relationship exhibits the same pattern as the first figure, the other experience variables may pick up the underlying non-linearity of the patient number. Visually we are not able to detect strong similarities between the graphs suggesting that signicant results in Theil-Goldberger are not fully driven by non-linearities in the patient number.