HEDG

HEALTH, ECONOMETRICS AND DATA GROUP

WP 17/24

Learning from failure in healthcare: dynamic panel evidence of a physician shock effect

Raf Van Gestel; Tobias Müller and Johan Bosmans

August 2017

http://www.york.ac.uk/economics/postgrad/herc/hedg/wps/

Learning from failure in healthcare: dynamic panel evidence of a physician shock effect

Raf Van Gestel*, Tobias Müller†, Johan Bosmans‡

ABSTRACT

Procedural failures of physicians or teams in interventional healthcare may positively or negatively predict subsequent patient outcomes. We identify this "learning from failure"-effect by applying (non-)linear dynamic panel methods using data from the Belgian Transcatheter Aorta Valve Implantation (TAVI) registry containing information on the first 860 TAVI procedures in Belgium. Using bias-corrected fixed effects linear probability models and the split-panel jackknife estimator proposed by Dhaene and Jochmans (2015), we find that a previous death positively and significantly predicts subsequent survival of the succeeding patient. Moreover, our results also provide evidence for learning from failure for stroke. We find that these learning from failure effects are not long-living and that learning from failure is transmitted across adverse events, e.g., a stroke affects subsequent survival.

Keywords: Physician behavior, Learning, Failure
JEL: I10, I13, I18, C93

PRELIMINARY VERSION

* Corresponding author. Department of Economics, University of Antwerp, Prinsstraat 13, 2000 Antwerp, Belgium. Mail: raf.vangestel@uantwerpen.be, Tel.: +3232654538
† University of Bern, Department of Economics, Schanzeneckstrasse 1, 3001 Bern, Switzerland. Mail: tobias.mueller@vwi.unibe.ch
‡ Department of Cardiology, University of Antwerp, Wilrijkstraat 10, 2560 Edegem, Belgium.

1. INTRODUCTION

The role of the physician is, evidently, widely acknowledged to be important in the production of health. Physician incentives and behavior, but also characteristics, are expected to be predictive of patient health outcomes. In this respect, aspects like education and specialization (Jollis *et al.*, 1996), adherence to guidelines (Ward, 2006) and experience (Hockenberry and Helmchen, 2014) have all been shown to affect patients' health outcomes. At a more aggregated level, physician supply and contract-types (contracted vs. municipality GP's in Aakvik *et al.* (2006)) have been shown to be correlated with mortality rates (Or *et al.*, 2005; Sundmacher *et al.*, 2011; Iizuka, 2016). Also, payment schemes influence physicians' provision behavior. In particular, patients are overserved under fee-for-service and underserved under capitation (Hennig-Schmidt *et al.*, 2011; Brosig-Koch *et al.*, 2016; 2015).

As a result, a number of initiatives aim to improve provider performance. Firstly, these initiatives may raise information through continued medical education (Cervero *et al.*, 2015), traineeships and audiovisual materials (Haynes *et al.*, 1995). Secondly, market based interventions change incentives for patients and physicians which also influences subsequent performance. For example, publicly available report cards (Kolstad, 2013) and pay for performance programs (Li, 2014) target incentives through both intrinsic motivation and market-based stimuli. In this paper, we contribute to this literature by identifying whether learning from failure has an impact on physician performance. More specifically, we explore if previous failures (e.g. adverse events like patient mortality and stroke) affect subsequent physician performance. This research may inform policy makers on the scope to introduce future information or awareness campaigns. In addition, this analysis can also serve as a starting point to further explore underlying reasons for such shock effects.

The literature on learning for physicians focuses on specific case-studies and the identification of different types of learning. Throughout this literature, endogeneity problems hamper inference and mainly arise through reverse causality and risk selection over time (e.g. Gaynor, 2005; Hentschker and Mennicken, 2016). Physician learning is an umbrella term covering multiple types of learning, forgetting and knowledge transfer. Most notably, a distinction is made between physician experience, economies of scale and human capital depreciation (Hockenberry and Helmchen, 2014; Van Gestel *et al.*, 2016). In Van Gestel *et al.* (2016), all three types of learning have been further investigated based on patient subgroups. Specifically, Van Gestel *et al.* (2016) show that overall, treating an extra patient is associated with improved patient outcomes in terms of a reduction in long-term mortality. At the same time, they argue that this overall learning effect can be narrowed down to treating extra patients with certain characteristics (e.g. high blood pressure and renal failure). This provides a more detailed perspective on where exactly these types of learning take place. In this paper, we build on this work by looking at shock effects related to learning from failure. Overall, the performance of a physician or team might improve when more patients are treated or an adverse event may disrupt health provider performance. We coin this effect here as "learning from failure". As such, we study dynamics to look at learning *within* the typical learning effect. [1]

Learning from failure is important at all levels of the healthcare sector. Nation-wide failed health reforms should inform future reforms and organizations are expected to foster failure driven learning even though they might face barriers to do so (Oberlander, 2007; Edmondson, 2004). For example, the organization of medical handovers can be improved to reduce the prevalence of incidents (Thomas *et al.*, 2012). At the physician or team level, learning from failure might be encouraged by the information provided by the failure itself (e.g. physicians may learn more from risky patients), by the incentive to avoid malpractice claims (Pänthofer, 2017) or more generally by the aversion for loss. This loss aversion may relate to income, patient break-ups because of low performance and/or adverse events (Rizzo, 2003; Hareli, 2007). However, failure may not only stimulate learning, it could also result in more failures. This result, prevalent in the organizational literature, occurs when for example failed (business) projects negatively influence a team (Shepherd, 2013). Psychologically, personal goal failure may lead to negative affective states and may therefore translate into negative subsequent outcomes (Jones, 2013). Also, physician inertia may contribute to subsequent failure. A failure to respond to adverse outcomes may stem from habit formation and from the reluctance to adjust treatment practice because of sizeable search and learning costs (Janakiraman, 2008). Whereas the idea of learning from failure for physicians is related to research on physician inertia in pharmaceutical prescriptions, we apply the idea of previous experience and inertia to a more interventional setting. Additionally, we stress that learning from failure cannot be considered without taking into account the physicians' learning or innovation pattern.

From an empirical perspective, it is important to disentangle the typical learning curve from physician (or team) learning from failure. The identification of the learning from failure effect on physician performance relies on previous experiences as patient mortality for a physician

[1] Denoted later on and in Van Gestel et al. (2016) as cumulative experience and referring to the effect that as more patients are treated, performance improves.

depends on mortality of the previous patient(s). Estimation of such learning from failure effects imposes two main econometric challenges: First, in the context of binary response fixed effects (FE) panel data models the well-known incidental parameter problem leads to inconsistent coefficient estimates (for an overview see Lancaster, 2000). Second, the inclusion of lagged dependent variables renders the standard FE estimator inconsistent inducing additional bias in form of the classical Nickell bias on the estimated coefficients (Nickell, 1981). As a remedy, difference and system GMM-type estimators are commonly used in the applied literature to address these selection issues. For example, Salge *et al.* (2016) apply dynamic instrumental variable panel methods in relationship with quality of care. They find that infections decrease with better overall cleaning, training on infection control, hand hygiene and a favorable error-reporting environment. We refrain from applying these type of GMM estimators as they have shown to have rather poor small sample properties due to weak instrument problems (Pua, 2015; De Vos *et al.*, 2012; Bruno, 2005). Instead, we apply the bias-corrected FE estimator proposed by De Vos *et al.* (2015) which has shown to have superior small sample properties compared to GMM estimators as most of the bias in the FE estimator is removed. In addition to the bias-correction approach, we estimate the learning from failure effects using the split-panel jackknife FE estimator recently proposed by Dhaene and Jochmans (2015).

In this paper, we provide evidence of substantial short-lived shock effects resulting from a team or provider failure. When the previous patient died, the probability to die within one month is 5 to 11%-points lower for the next patient. Similarly, a previous stroke is correlated with a 6 to 7%-point decrease to have a stroke. Although different non-linear dynamic panel methods provide different results, most specifications provide qualitatively similar results. We also find minor evidence for a transmission mechanism of shocks between adverse events. Previous mortality is correlated with a lower likelihood of a stroke and vice versa.

In the remainder of this paper we discuss the background of our application to TAVI (Transcatheter Aorta Valve Implantation) and our data in section 2. In section three we focus on methods to measure the effect of learning from failure. In section 4 we present our results after which we discuss and conclude in section 5.

2. THE INTRODUCTION OF TRANSCATHETER AORTA VALVE IMPLANTATION

The application in this paper considers the introduction and evolution of Transcatheter Aorta Valve Implantation (TAVI) in Belgium for which the first procedure in a Belgian hospital has taken place in 2007. We use register data for patients from 2007, including the first patient, to the beginning of 2012 in about 20 hospitals. As such, our data is able to describe the performance of physicians for the introduction period of TAVI. In each hospital only one team performs this TAVI procedure and during the sample period total workload for TAVI was limited to about one day a week. Information is available on a wide range of patient specific characteristics (for more information also see Van Gestel *et al.* (2017)) and we have access to hospital identifiers. There is no more information at the hospital level. Table 1 gives an overview on the number of patients treated at each hospital over the different years[2].

Table 1: Descriptive Statistics - Patients undergoing TAVI in Belgium

	2007	2008	2009	2010	2011	2012	Total
Total	**10**	**100**	**163**	**257**	**289**	**36**	**855**

Table 2 provides first descriptive evidence for the learning from failure effect for interventional care. The estimates in the "No Prev." columns contain probabilities for the adverse event if the previous patient did not suffer from the adverse event in contrast to the column "Prev.". For one-month mortality and stroke, the probability of suffering from a stroke or dying within one month is substantially lower if the previous patient exhibited this adverse event. For renal failure and pacemaker, point estimates have the opposite sign and are statistically insignificant. Throughout the next sections, we will thoroughly scrutinize these descriptive findings.

Table 2: Probability of an adverse event after an adverse event

	1-month mortality		Stroke		Renal Failure		Pacemaker	
	No Prev.	**Prev**	**No Prev.**	**Prev**	**No Prev.**	**Prev**	**No Prev.**	**Prev**
Est.	0.099	0.038	0.042	0.000	0.130	0.156	0.145	0.202
P-val.	0.013		0		0.514		0.168	

[2] Note that we only have information on the TAVI procedures until the beginning of 2012 which in turn explains the low number of procedures for the year 2012.

3. METHODS

3.1. Empirical specification and learning curves

The literature on learning curves in health distinguishes between three types of learning (see e.g. Hockenberry and Helmchen (2014); Van Gestel *et al.* (2016)): cumulative experience (CE), economies of scale (ESC) and human capital depreciation (HCD). Learning from cumulative experience refers to the idea that treating an additional patient generally improves physician (or team) performance. When referring to economies of scale, we capture the fact that higher volume providers usually have better infrastructure (e.g. equipment, staff) and more standardized procedures. Lastly, the human capital depreciation hypothesis states that provider performance decreases with longer temporal distance to previous procedures. This leads to the specification in equation (1) which is empirically estimated using standard regression techniques:

(1)

Where and denote an individual in hospital at time . In our context, the outcome may be a binary mortality or stroke indicator. Furthermore, we control for a vector of background characteristics and comorbidities contained in and hospital fixed effects . Lastly, is a typical error term which in case for LPM's is not normally distributed. Consequently, we use heteroscedasticity-robust standard errors in all our model specifications.

(2)

In equation (2), we extend (1) by adding a lag of the outcome variable as additional independent variable to the model. Note here that the outcome, say mortality, of patient *i* in hospital *h* in period *t* is regressed on the mortality indicator of the previous patient *i-1* which was treated just before patient *i* in the same hospital *h*. is therefore capturing the learning from failure effect. Furthermore, we include the learning variables from equation (1) to overcome the potential omitted variable bias resulting from a likely correlation between the lagged outcome and the learning indicators. In fact, without the inclusion of the typical learning effects, would likely be an upper bound on the true effect[3]. However, the dynamics in equation (2) require the estimation of a non-linear dynamic panel with fixed

[3] I.e. it would probably overstate the positive effect of a previous failure on survival as e.g. cumulative experience is negatively correlated to both contemporaneous and lagged patient mortality pointing toward an upward bias in .

effects. The following sections discuss and deal with the incidental parameter problem and Nickell bias inherent to this setting.

3.2. Dynamic panels and the identification of shock effects

Since the dynamic effects for the adverse events are of primary interest in this paper, our main goal is to obtain consistent and efficient estimates of the potential learning from failure effect. Specifically, we analyze whether the patient outcome of a team or physician procedure is correlated with previous (negative) experiences. In the context of binary dependent variables and fixed effects, the incidental parameters problem presents a hurdle to obtain consistent and efficient estimates. Because of the incidental parameters problem, the fixed effects estimates are inconsistent and this also translates into inconsistency for all other coefficients (Baltagi, 2008, p. 210; Wooldridge, 2010). The inclusion of incidental parameters is even more problematic with lagged dependent variables because of the Nickell bias[4] (Moon *et al.*, 2015). To address these econometric challenges, we run a series of alternative estimation techniques. As a starting point, we estimate the learning from failure effects using simple fixed-effects Linear Probability Models (FE LPM's). Although widely used in the applied literature, Chernozhukov *et al.* (2013; p. 546) demonstrate that the FE LPM provides inconsistent estimates for average marginal effects in dynamic panel settings. Therefore, we also apply non-linear dynamic panel models. Using probit or logit regression models in combination with lagged dependent variables and fixed effects however calls for corrections because of the abovementioned incidental parameters and Nickell bias. These corrections often rely on instrumental variables methods to remove the correlations with the error term. However, as shown in Pua (2015), the particularly popular difference and system GMM estimators often tend to suffer from substantial finite sample bias due to weak instrument problems.

To address these challenges, we apply the bias-corrected FE estimator proposed by De Vos *et al.* (2015) based on Everaert and Pozzi (2007) which has been shown to have superior small sample properties than the classical GMM-type estimators as it removes most of the bias in the classical FE estimator.

Finally, we also apply the split-panel jackknife recently suggested by Dhaene and Jochmans (2015) which addresses the incidental parameters problem, while at the same time accounting for the dynamics. We apply the jackknife-based approach because we do not

[4] As shown in Nickell (1981), the inclusion of lagged dependent variables leads to a direct correlation of the lagged outcome with the error term rendering the FE estimator inconsistent.

include the typical time fixed effects[5] as in Fernández-Val and Weidner (2016) where the analytical bias correction is preferred. Whereas most solutions to the abovementioned difficulties are computationally complex and require analytical solutions, the jackknife is relatively straightforward to implement and performs similarly or even better than (p.995) other approaches. The drawback however is the difficulty to include time trends in non-linear fixed effects models which makes it hard/impossible to correct for the overall learning curve. With different dynamics in the subpanels, the estimates of all coefficients may differ. Consequently, we provide a range of estimates and test for the sensitivity to these differential trends in the robustness section below.

The intuition underlying the split-panel jackknife is to divide each panel in smaller panels of consequent observations to split the panel in subpanels.[6] Because the bias depends on the length T of the panel, using different panel lengths by generating subpanels and comparing the coefficient estimates of the subpanels with the estimate for the complete panel provides an estimate of the bias. Subtracting the estimated bias from the complete panel estimate generates the split-panel jackknife estimator (Dhaene and Jochmans, 2015, p. 998). One simple choice to determine the subpanel lengths also suggested in Dhaene and Jochmans (2015) is the half-panel jackknife where the panels are simply divided in two. In a dynamic setting (p. 1007) the jackknife may perform suboptimally when the dynamics are very different in the half-panels. Tests based on the difference in estimates across subpanels can be applied to test for sensitivity to differential dynamics.[7]

4. RESULTS: EVIDENCE ON LEARNING FROM FAILURE

4.1. Fixed effects linear probability estimates

As a starting point, we estimate simple fixed effects linear probability models (henceforth FE LPM's) before presenting the bias-corrected FE and the split-panel jackknife estimates. Table 3 below presents the FE LPM estimates of our learning from failure effects for the outcomes of one- and 24-month mortality and having a stroke during the TAVI procedure for two

[5] This is simply because our constructed "panel" does not have a strict time dimension which makes it difficult to interpret and because we include a range of time trends (HCD, CE, ESC) as independent variables of interest.

[6] We provide a brief and intuitive summary of the split-panel jackknife. For more detailed technical information, please see Dhaene and Jochmans (2015). For more information on the technical implementation in Stata, consult the help files on *probitfe.* https://ideas.repec.org/c/boc/bocode/s458278.html

[7] However, the tests are not standardly available in Stata or other statistical software. Do-files for the non-stationarity test as implemented in the robustness section are available upon request.

different model specifications. While specification (1) includes patient- and procedure specific characteristics and hospital fixed effects, specification (2) additionally controls for the different learning effects described above.

Overall, we find highly significant negative coefficients on our lagged outcome variables for both the mortality outcomes and having a stroke. In fact, our FE LPM estimates indicate that the predicted likelihood of dying in the first month after the TAVI procedure is associated to decrease by about 7.8%-points if the last patients past away. Likewise, the probability of dying 2-years after the procedure is predicted to decrease by about 7.4%-points pointing again towards strong learning from failure effects. Furthermore, in line with the findings in Van Gestel *et al.* (2016), our estimates provide evidence for a significant positive learning from cumulative experience effect as treating an additional patient is associated with a decrease in 2-year mortality of about 0.2%-points. Although seemingly a small effect, this quickly becomes large for sizeable patient samples.

Table 3: FE LPM estimates of the learning from failure effects

Outcome	**1-m mortality** (1)	**1-m mortality** (2)	**24-m mortality** (1)	**24-m mortality** (2)	**Stroke** (1)	**Stroke** (2)
Lagged outcome	-0.079***	-0.078***	-0.069*	-0.074**	-0.059***	-0.068***
	(0.026)	(0.025)	(0.036)	(0.036)	(0.021)	(0.023)
CE		0.000		-0.002**		0.000
		(0.001)		(0.001)		(0.000)
ESC		-0.001		0.005		-0.001
		(0.002)		(0.003)		(0.001)
HCD		0.000		0.001		0.001*
		(0.000)		(0.001)		(0.000)
Same day		0.039		0.040		0.039*
		(0.025)		(0.041)		(0.020)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	762	762	762	762	680	680
R-squared	0.088	0.093	0.106	0.116	0.073	0.085

The table shows the FE LPM estimates of the learning from failure effects for one- and 24-month mortality, as well as having a stroke for two model specifications. The average baseline probabilities for one- month mortality is at around 8.8%, for 24-month mortality at 28.9% and that for having a stroke at 4% across all hospitals and years. Heteroscedasticity-robust standard errors in parenthesis: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

As for the likelihood of suffering a stroke, we find that if the last patient suffered a stroke during the procedure the likelihood that the next patient also suffers a stroke is associated to decrease by approximately 6.8%-points. These findings imply that the likelihood of an adverse event strongly diminishes with a previous failure. The dynamic effects suggest that there is substantial room for self-correction and personal improvement after failure. This might also be interpreted as the suboptimal use of production capacity because the production lies at a point below the production possibility frontier. As such, policies may be invoked to make efficient use of all production capacities. In addition, for stroke we find a significant effect for human capital depreciation (see also Van Gestel *et al.* (2016)). With each additional day between procedures the probability for a stroke is predicted to increase by about 0.1%-points. Moreover, having a second TAVI procedure on the same day is associated with an increase in the likelihood of suffering a stroke by approximately 3.9%-points, ceteris paribus.

4.2. Bias-corrected FE LPM and split-panel jackknife FE probit estimates

As discussed in the methodology section, the dynamics in the model specifications above cause the errors to be correlated with the lagged dependent variables inducing a Nickell Bias on all FE LPM coefficient estimates shown in table 3. The preferable strategy therefore is to consider the dynamics and simultaneously address the incidental parameter problem. To this end, we apply the bias-corrected FE estimator suggested by De Vos *et al.* (2015) and the split-panel jackknife for probit models proposed by Dhaene and Jochmans (2015). We apply bias corrections also to LPM's because, as shown in table 2, a previous stroke perfectly predicts subsequent survival. As such, in a multiplicative model the coefficient equals infinity which makes it impossible to estimate. As a result, the estimation of probit models is impossible as opposed to LPM's where coefficients are interpreted additively. For one and 24-month mortality however we show results for both the LPM bias correction and the split-panel jackknife.[8]

Table 4 below shows the bias-corrected FE estimates of the learning from failure effects for two different specifications for each of the outcomes. Both specifications are based on 250 bootstrap samples and we use the burn-in initialization scheme to set the initial values of the lagged dependent variables. However, while in specifications (1) we allow for general heteroscedasticity of the error term using the wild bootstrap suggested by Liu (1988) and Mammen (1993), in specifications (2) we impose a pure cross-sectional heteroscedasticity error sampling scheme used in the FE bias-correction algorithm proposed by De Vos *et al.* (2015).

[8] Note that we use the probitfe command here, we tested robustness with the logitfe command. Using logitfe, the log likelihood function does not converge. The difference with the probitfe might be in…

The results in table 4 below are largely in line with our findings from table 3 reinforcing the presence of strong learning from failure effects: the likelihood that the current patient dies after or during the TAVI procedure is predicted to decrease if the last patient has died within one- respectively 24-months after the procedure. Similarly, if the last patient suffered a stroke during the intervention, the probability that the next patient also suffers a stroke is associated to decrease significantly by about 7.5%-points, ceteris paribus. In addition, we again find evidence for significant learning from cumulative experience effects for 2-year mortality, human capital depreciation for the likelihood of suffering a stroke and having more than one procedure on a single day significantly increases the likelihood of 1-month mortality. In contrast to the FE LPM estimates above, the bias-corrected FE estimates tend to be smaller in absolute value for the mortality indicators and larger for stroke thus pointing toward substantial Nickell bias in the FE LPM estimates above.

Table 4: Bias-corrected FE LPM estimates of the learning from failure effects

Outcome	**1-m mortality** (1)	**1-m mortality** (2)	**24-m mortality** (1)	**24-m mortality** (2)	**Stroke** (1)	**Stroke** (2)
Lagged outcome	-0.046	-0.045	-0.060*	-0.060	-0.075**	-0.077
	(0.033)	(0.037)	(0.033)	(0.042)	(0.032)	(0.053)
CE	-0.001	-0.001	-0.003***	-0.003*	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
ESC	0.007**	0.007**	0.008*	0.008	-0.002	-0.002
	(0.003)	(0.003)	(0.005)	(0.005)	(0.003)	(0.003)
HCD	0.000	0.000	0.000	0.000	0.001*	0.001**
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Same day	0.060*	0.060	0.010	0.009	0.042	0.038
	(0.032)	(0.039)	(0.043)	(0.045)	(0.029)	(0.027)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	550	550	550	550	477	477

The table shows the estimated learning from failure effects using the bias-corrected FE LPM proposed by De-Vos et al. (2015) for the outcomes of one- and 24-month mortality and having a stroke. Bootstrapped standard errors in parenthesis: *** p<0.01, ** p<0.05, * p<0.1.

In a next step, we estimate the learning from failure effects using the split-panel jackknife FE probit estimator recently proposed by Daene and Jochmans (2015). Table 5 above shows the estimated learning from failure effects for one-month and 2-year mortality while controlling for the usual patient- and procedure-specific characteristics and also including hospital fixed effects. Overall, we again find significant evidence for the presence of a learning from failure effect. Specifically, our split-panel jackknife estimate suggest that the likelihood of survival is associated to increase by about 11%-points (resp. 7%-points) if the last patient past away within 1-month (24-months) after the TAVI procedure.

In conclusion, our different estimation approaches all lead to the same qualitative conclusion. Physician (or team) performance tends to positively respond to previous failures or shock events as we observe both a significant decrease in the likelihood of short- and long-term mortality for the next patient, as well as a reduced probability for complications (here: stroke) during the procedure for the succeeding patient.

Table 5: Split-panel jackknife FE probit estimates of the learning from failure effects[9]

Outcome	**1-m mortality**	**24-m mortality**
Lagged outcome	-0.112***	-0.070**
	(.020)	(0.034)
CE	0.001	0.000
	(.001)	(0.001)
ESC	-0.005	-0.001
	(.002)	(0.003)
HCD		-0.000
	-	(0.001)
Same day	0.12	0.016
	(.022)	(0.041)
Controls	Yes	Yes
Hospital FE	Yes	Yes
Observations	730	761

The table shows the estimates of the learning from failure effects using the Split-panel jackknife FE probit proposed by Dhaene and Jochmans (2015). Standard errors in parenthesis: *** p<0.01, ** p<0.05, * p<0.1.
For one-month mortality, the estimates do not converge when HCD is included. Overall, the coefficient of the Lagged outcome is robust to leaving out the learning variables (i.e. leaving out all learning variables for one-month mortality and leaving out HCD for 2-year mortality).

4.3. Nature and interpretation of the shock effect

The significant lagged effect may be interpreted in several ways. Firstly, we might expect that more can be learned from a failure than a success. In this respect, we would expect that the failure would impact on the outcome through better learning which would empirically translate in a (slope-) shift of the "typical" learning curve. This slope-shift is expected because it is likely that more is learned from the first failure as opposed to later failures. However, in figure 1 below, the same downward effect of a previous failure on

[9] Note that we use the *probitfe* command here, we tested robustness with the *logitfe* command. Using logitfe, the log likelihood function does not converge.

mortality is found across all patient numbers. This points out that the effect of an adverse event is constant at all levels of experience and we could therefore conclude that instead of influencing the learning curve, an adverse event has a constant impact, which could point towards a sort of concentration effect.

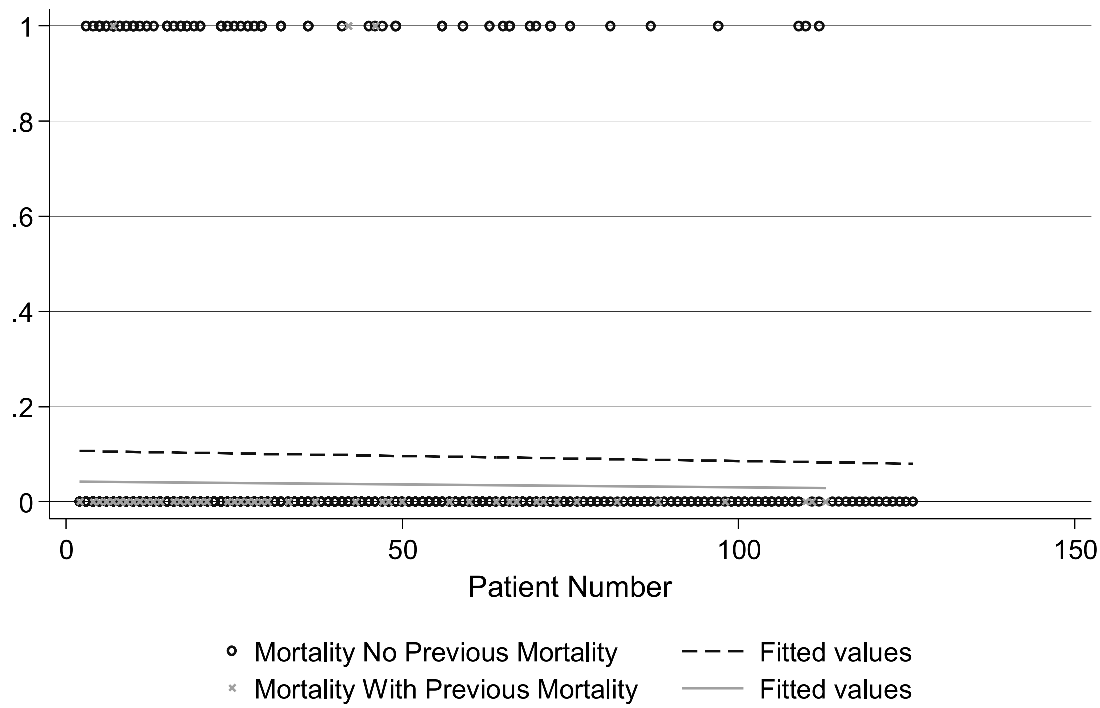

Figure 1: Learning pattern with and without previous failure

Secondly, we are interested in the persistence of the shock effect, i.e. whether longer lags on the dependent variable are still significant. However, in none of our analyses, the second lag is statistically significant which suggests that the shock effect is not long-living. Additionally, it would also be of interest to investigate whether the second lag effect differs according to the value for the first lag. However, because there are almost no subsequent failures, the interaction would hold approximately no information. In fact, in our data there are only three cases for which the two previous patients deceased.

4.4. Transmission of Risks

The LPM's below (columns one and three), show evidence of a transmission of shocks between adverse events. Firstly, we find that when a patient has a stroke in a hospitalization, (s)he is also more likely to die. As such having a stroke is strongly correlated with a procedural failure. Secondly, provided that the previous patient had a stroke, the probability of mortality is lower for the next patient. The same intuition holds for the effect of a previous mortality on a subsequent stroke. However, when correcting for incidental parameters with the bias correction of De Vos *et al.* (2015), the results become insignificant.

Table 6: Transmission of shocks

Outcome	Mortality LPM	Mortality Bias Corr.	Stroke LPM	Stroke Bias Corr.
Mortality	-	-	0.116**	0.123***
	-	-	(0.048)	(0.037)
Lagged Mortality	-0.052*	-0.003	-0.024**	-0.018
	(0.028)	(0.056)	(0.012)	(0.042)
Stroke	0.240***	0.218***	-	-
	(0.091)	(0.059)	-	-
Lagged Stroke	-0.047*	-0.053	-0.053**	-0.059
	(0.026)	(0.057)	(0.025)	(0.050)
Controls	Yes	Yes	Yes	Yes
Observations	671	479	671	477
R-squared	0.137	-	0.107	-

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

5. ROBUSTNESS

To further test for the robustness of our results, we provide several additional tests and specifications throughout this section. Firstly, we test for the robustness of non-stationarity of regressors in the split-panel jackknife estimation. Secondly, the learning-related regressors are likely to have a non-linear relationship with patient health outcomes. We show that adding learning variables in different specifications does not qualitatively alter our results.

5.1. Non-stationarity

As discussed in the methodology section, the split-panel jackknife estimator assumes stationary data. This assumption is violated in our application because of the non-stationarity of learning effect regressors. In this section we implement the stationarity test suggested by Dhaene and Jochmans (2015, p. 1007) to assess the adequacy of the coefficient on the lagged adverse events. A Wald test on the difference of coefficients for subpanels guides intuition on the robustness to non-stationary data and is defined as follows:

Where is the number of panel units, is the (average) number of time periods and corrects for variance inflation resulting from the use of subpanels (Dhaene and Jochmans, 2015, p.1007). is a metric to compare subpanel estimates and is the information matrix or, equivalently, the inverse of the variance-covariance matrix or the negative of the expected Hessian matrix.[10] Finally, the Wald test statistic is Chi-square distributed with the dimension of the parameter vector (which equals one in a one-by-one parameter comparison).

The Chi-square test as applied to the coefficient of interest, the lagged mortality, has a p-value of 0.00, pointing out that the dynamics indeed strongly influence the coefficient.[11]

5.2. Quadratic learning curve specifications

As an additional robustness check, we allow for non-linearities in the relationship between the typical learning variables (cumulative experience, economies of scale and human capital depreciation) and patient mortality or having a stroke. The resulting FE LPM estimates of learning from failure effects can be found in table 7 below. Overall, the estimated effects are in line with our previous findings: previous failures are both negatively associated with short- and long-term mortality of the next patient and having complications during the procedure. Non-linearities play only a minor role in explaining these patient outcomes as all the coefficients on the squared learning variables are near to zero and therefore not economically significant.

Table 7: Inclusion of squared learning variables

Outcomes	**1-m mortality**	**24-m mortality**	**Stroke**
	(1)	(1)	(1)
Lagged outcome	-0.080***	-0.076**	-0.068***
	(0.025)	(0.036)	(0.023)
CE	0.002	-0.001	-0.001
	(0.002)	(0.003)	(0.001)
ESC	0.010	0.013	0.003

[10] — — , where and are the average subpanel lengths.

— — , where and are the subpanel estimates and the full panel estimate.

[11] This result is robust to the inclusion of different subsets of learning variables because of non-convergence issues in the split-panel jackknife.

	(0.007)	(0.010)	(0.005)
HCD	0.000	-0.000	0.001*
	(0.001)	(0.001)	(0.001)
CE squared	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)
ESC squared	-0.000*	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
HCD squared	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)
Same day	0.047	0.021	0.049**
	(0.030)	(0.047)	(0.024)
Controls	Yes	Yes	Yes
Hospital FE	Yes	Yes	Yes
Observations	762	762	680
R-squared	0.098	0.117	0.089

The table shows the FE LPM estimates of the learning from failure effects including quadratic terms of the typical learning variables (cumulative experience, economies of scale and human capital depreciation). Heteroscedasticity-robust standard errors in parenthesis: *** p<0.01, ** p<0.05, * p<0.1

6. CONCLUSION

Identifying different channels through which physicians affect patients' health may help policy makers to efficiently allocate resources to policy interventions. In this paper, we shed light on the question how procedural failures of physicians (or teams) affect subsequent patient outcomes. We show that this "learning from failure effect" is an important source of physician learning besides the commonly identified factors such as economies of scale, learning from cumulative experience and human capital depreciation. To identify such learning from failure effects, we apply the recently developed bias-corrected fixed effects estimator by De Vos *et al.* (2015) and the split-panel jackknife estimator proposed by Dhaene and Jochmans (2015) to address the econometric challenges inherent to non-linear dynamic panel data settings.

Our findings for TAVI heart valve replacements provide evidence for a significant and sizeable negative effect from a previous failure on subsequent patient mortality. We find that a previous death significantly decreases the probability of a subsequent patient death with about 6 to 12 percentage points. Moreover, we find that if the last patient suffered a stroke during the procedure the likelihood that the next patient also suffers a stroke is associated to significantly decrease. However, our results suggest that these effects are only short-lived and they do not shift the slope of the cumulative learning effects. Possible short-term reasons underlying this learning from failure effect are increased concentration or short-term improved knowledge of the physicians or teams. Finally, we have illustrated the robustness of our results by showing a range of estimation techniques, model specifications and non-stationarity tests.

7. LITERATURE

Aakvik, A., Holmas, T.H., (2006). Access to primary health care and health outcomes: The relationships between GP characteristics and mortality rates. Journal of Health Economics, 25, 1139-1153.

Baltagi, B. (2005). Econometric analysis of panel data. John Wiley & Sons Ltd.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., & Wiesen, D. (2016). Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision. *Journal of Economic Behavior & Organization*, *131*, 17-23.

Brosig Koch, J., Hennig Schmidt, H., Kairies Schwarz, N., & Wiesen, D. (2015). The effects of introducing mixed payment systems for physicians: Experimental evidence. *Health economics*.

Bruno, G.S.F. (2005). Estimation and inference in dynamic unbalanced panel-data models with a small number of individuals. *The Stata Journal*, 5(4), 473-500.

Bun, M. J., & Kiviet, J. F. (2006). The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models. *Journal of econometrics*, *132*(2), 409-444.

Bun, M. J., & Windmeijer, F. (2010). The weak instrument problem of the system GMM estimator in dynamic panel data models. *The Econometrics Journal*, *13*(1), 95-126.

Cervero, R.M., Gaines, J.K. (2015). The impact of CME on physician performance and patient health outcomes: an updates synthesis of systematic reviews. *Journal of Continuing Education in the Health Professions,* 35(2),131-138.

Chernozhukov, V., Fernández-Val, I., Hahn, J., Newey, W. (2013). Average and quantile effects in nonseparable panel models. *Econometrica*, 81(2), 535-580.

De Vos, I., Everaert, G., & Ruyssen, I. (2015). Bootstrap-based bias correction and inference for dynamic panels with fixed effects. *Stata Journal (forthcoming)*, 1-31.

Dhaene, G., Jochmans, K. (2015). Split-panel jackknife estimation of fixed-effect models. *Review of Economics Studies*, 82, 991-1030.

Edmondson, A.C. (2004). Learning from failure in health care: frequent opportunities, pervasive barriers. Qaulity and Safety in Health Care, 13(Sup. II), ii3-ii9.

Everaert, G., Pozzi, L. (2007). Boostrap-based bias correction for dynamic panels. *Journal of Economic Dynamics & Control,* 31, 1160-1184.

Fernández-Val, I., Weidner, M. (2016). Individual and time effects in nonlinear panel models with large N, T. *Journal of Econometrics,* 192, 291-312.

Gaynor M., Seider H., Vogt W.B. (2005). The Volume-Outcome Effect, Scale Economies, and Learning-by-Doing. *The American Economic Review: Papers and Proceedings,* **95(2),** 243-247.

Hareli, S., Karnieli-Miller, O., Hermoni, D., Eidelman, S. (2007). Factors in the doctor-patient relationship that accentuate physicians‵ hurt feelings when patients terminate the relationship with them. *Patient Education and Counseling*, 67, 169-175.

Hennig-Schmidt, H., Selten, R., & Wiesen, D. (2011). How payment systems affect physicians' provision behaviour - an experimental investigation. *Journal of Health Economics*, *30*(4), 637-646.

Hentschker, C., Mennicken, R. (2015), The Volume-outcome Relationship and Minimum Volume Standards - Empirical Evidence for Germany. *Health Economics*, 24(6), 644-658.

Hockenberry, JM., Helmchen L.A., (2014). The nature of surgeon human capital depreciation. *Health Economics,* 37, 70-80.

Iizuka, T., Watanabe, Y., (2016). The impact of physician supply on the healthcare system: evidence from Japan's new residency program. *Health Economics,* 25,1433-1447.

Janakiraman, R., Dutta, S., Sismeiro, C., Stern, P. (2008). Physisians' persistence and its implications for their response to promotion of prescription drugs. *Management Science,* 54(6), 1080-1093.

Jollis, J.G., Delong, E.R., Peterson, E.D., Muhlbaier, L.H., Fortin, D.F., Califf, R.M. Mark, D.B. (1996). Outcome of acute myocardial infarction according to the specialty of the admitting physician. The New England Journal of Medicine, 335, 1880-1887.

Jones, N.P., Papadakis, A.A., Orr, C.A., Strauman, T.J. (2013). Cognitive processes in response to goal failure: a study of ruminative thought and its affective consequences. *Journal of Social and Clinical Psychology*, 23(5), 1-19.

Kolstad, J.T (2013), Information and Quality when Motivation is Intrinsic: Evidence from Surgeon Report Cards. *American Economic Review*, 103(7), 2875-2910.

Lancaster, T. (2000). The incidental parameter problem since 1948. Journal of econometrics, 95(2), 391-413.

Li, J., Hurley, J. Decicca, P., Buckley, G. (2014). Physician response to pay-for-performance: evidence from a natural experiment. *Health Economics,* 23, 963-978.

Liu, R. Y. (1988). Bootstrap procedures under some non-iid models. The Annals of Statistics, 16(4), 1696-1708.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. The annals of statistics, 255-285.

Moon, H.R., Perron, B., Phillips, P.C.B. (2015). Incidental parameters and dynamic panel modelling in "The Oxford Handbook of Panel Data" (ed. Badi H. Baltagi)

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, 1417-1426.

Oberlander, J. (2007). Learning from Failure in Health Care Reform. The New England Journal of Medicine, 357(17), 1677-1679.

Or, Z., Wang, J., Jamison, D. 2005. International differences in the impact of doctors on health: a multilevel analysis of OECD countries. *Journal of Health Economics,* 24,531-560.

Panthöfer, S. (2017). Do Doctors Prescribe Antibiotics Out of Fear of Malpractice? Mimeo.

Pua, A. A. Y. (2015). On IV estimation of a dynamic linear probability model with fixed effects (No. 15-01). Universiteit van Amsterdam, Department of Econometrics.

Rizzo, J.A., Zeckhauser, R.J. (2003). Reference incomes, loss aversion, and physician behavior. *The Review of Economics and Statistics,* 85(4), 909-922.

Salge, T.O., Vera, A., Antons, D., Cimiotti, J.P. (2016). Fighting MRSA infections in hospital care: how organizational Factors Matter. Health Services Research, 52(3), 959-983.

Shepherd, D.A., Haynie, M., Patzelt, H. (2013). Project failures arising from corporate entrepreneurship: impact of multiple project failures on employees' accumulated emotions, learning, and motivation. *Journal of Product Innovation Management,* 30(5), 880-895.

Sundmacher, L., Busse, R., (2011). The impact of physician supply on avoidable cancer deaths in Germany. A spatial analysis. *Health Policy,* 103, 53-62.

Thomas, M.J.W., Schultz, T.J., Hannaford, N., Runciman, W.B. (2013). Failures in transition: learning from incidents relating to clinical handover in acute care. *Journal for Healthcare Quality*, 35(3), 49-56.

Van Gestel, R., Müller, T., Bosmans, J. Forthcoming. Does My High Blood Pressure Improve Your Survival? Overall and Subgroup Learning in Health. *Health Economics.*

Ward, M.M., Yankey, J.W., Vaughn, T.E., BootsMiller, B.J., Flach, S.D., Welke, K.F., Pendergast, J.F., Perlin, J., Doebbeling, B.N. (2004). Physician Process and Patient Outcome Measures for Diabetes care: Relationships to Organizational Characteristics. *Medical Care*, 42(9), 840-850.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. MIT press.