APPENDIX A: QUALITATIVE METHODS

Our mixed-methods research emerged as what is sometimes labeled a "holistic triangulation" design, in which "the objective is to obtain fuller understanding through unique insights / perspectives gained from the different research strategies (Turner et al. 2017: 250)." We began our study by gathering quantitative data, assuming that we were dealing with a mature theoretical problem (Edmondson and McManus 2007). We modeled our analysis after quantitative research to study learning from a firm's own experience (Haunschild and Sullivan 2002) and from the experience of others (Baum and Dahlin 2007). In our quantitative analysis, it appeared that firms did learn from the adverse event reports of others and that they identified the reports that were likely to lead to learning through a combination of focusing on reporting delays and volumes of reports.

Two aspects of our study presented particular challenges. Both were a consequence of the diversity of the firms contributing to our data set. First, it was difficult to analyze our data following the theoretical assumptions of existing research, which typically follow the "Carnegie School" in assuming that firms learn from referent others when they fail to meet their performance targets (Greve 2003). Our approach relaxed the standard assumption in the literature that firms would be most likely to learn from similar and superior-performing referent others. Second, it was not possible to follow the methodological practices of existing research, which focus on observable performance changes as evidence of learning (Madsen and Desai 2010). In response to each of these challenges, we chose to follow the learning more directly by focusing on the adverse events as evidence of learning. While this allowed us to get closer to the actual knowledge transfer, it also made interpretation of the findings more difficult.

Our qualitative methods were a direct response to these problems of interpretation (Greene et al. 1989). This sequence in which modeling quantitative relationships is followed by qualitative analysis is a common approach for researchers (Gibson 2017). There is one notable difference with this study, however, in that our presentation of our qualitative data precedes the quantitative data. This is because our study combined both exploratory and explanatory purposes (Creswell 2003). The exploratory elements of our study were not entirely intentional. Though we thought we were following mature theoretical and

methodological approaches in gathering our quantitative data, our data set turned out to have unique features that were only apparent after we gathered our qualitative data. As a result, our findings are explorations in both setting and measures. The explanatory functions of our data followed from the unique features of our data and of the resulting methods. We used our qualitative methods to lend validity to both the theoretical approach to learning and the interpretive claims we make from that data.

Data Gathering

Our qualitative data gathering was a direct response to these exploratory and explanatory purposes. Our qualitative evidence consisted of three broad components. First, the first author was deeply immersed in the medical device industry throughout the research process. Over the course of the research, he engaged in a variety of site visits to manufacturers, hospitals, and research labs. He also built longstanding relationships with a range of medical device executives. One output of that engagement was a 150-person conference on medical device innovation organized by the first author. These experiences led to a range of encounters that informed our quantitative approach (c.f. Kim and Miner 2007).

We used purposeful sampling to gather our qualitative data. We first sent a survey to all device manufacturers registered with the FDA. We selected our sample of medical device manufacturers using email addresses gathered from the FDA Registration and Listing website. We sent an email to the 4702 device manufacturers registered with the FDA in 2009 and 2016. Because the FDA only publishes the emails of active registrants, we built this list to capture those that exited from the industry. The email asked potential respondents if they would like to participate in a short survey about behavioral outcomes in the industry. Of these, 202 firms agreed to participate and from that sample we were able to obtain 190 complete responses. From that list we purposefully sampled the firms that had failure experience (at least 1000 adverse events) and no failure experience, and those that had innovation experience (that is, firms with 510(k) applications for a new device) and those with no recent innovation experience. We also purposefully sought informants from a variety of company sizes and structures, including small family-owned firms and large, public multinationals, and from a variety product markets, including medical imaging, ear, nose and throat, neurology, and other device markets. From our sample of firms we chose

31 informants to represent a diverse range of adverse event experience, occupation, and roles, including CEO or President, middle management, regulatory analyst, sales, device design and consultant roles. Of those 31, 12 agreed to participate.

As shown in Table 3, the final 12 informants represent a diverse range of adverse event experience, occupation, and roles, including CEO or President, middle management, regulatory analyst, sales, device design and consultant roles. Our semi-structured interviews varied in length from 45 to 120 minutes. We also conducted follow-up interviews with two informants for a total sample of 14 interviews. All interviews were recorded and transcribed verbatim. Our interview protocol used a range of open-ended and directed questions that would avoid demand effects while still eliciting evidence on learning from adverse event reports in general and the MAUDE data set specifically.

Our interview process proceeded in two phases. In the first phase, we tested our interview protocol. We began our interviews with a series of questions on medical device innovation both to gather data on the learning process and to establish a comfortable relationship with the respondent, and a section on medical device failure. In the latter section, we sought to induce a hypothetical device-failure story. When this did not lead to direct evidence, we revised our interview protocol. Rather than inquire indirectly about how they might respond to failure, we chose to address the learning directly by asking how they scanned for experience and how they converted that knowledge to organizational change. In order to avoid demand effects around the MAUDE data set, we structured the interviews to address how they scanned the industry, how they learned from others, and how they used databases like the MAUDE in general.

Third, we gathered media evidence of device failure to supplement our experiential and interview data. We focused on both vivid examples that garnered media attention, and specific reports as evidence of learning from the MAUDE data. This data helped us understand adverse event reporting as it went into the MAUDE data, including how firms made sense of compliance requirements, how they grappled with complex problems, and how they learned throughout the process. These examples informed our understandings of how delays work.

Data Analysis. Our data gathering resulted in 17.5 hours of interview data and approximately 250 pages

of qualitative text. From this data, the first author initially gathered all instances of vicarious learning from others' failure, focusing in particular on how firms learned from the MAUDE data and collecting terms and insights discussed by the informants. This first round resulted in 150 pages of raw text of various themes. The research team then discussed the themes in that data and compared the themes against existing literature. From that discussion, we produced a 25-page document that addressed the major themes in the data: the uniqueness and complexity of the medical device industry, the noisiness of the MAUDE dataset, finding valid signals in the MAUDE data, the MAUDE data as a source of learning.

We took two clear lessons from this first round of analysis. First, though firms were very attentive to their competitive set, they did not use the MAUDE data to follow all of the actions of their direct competitors, particularly for senior managers in the organization. Instead, they had a much broader range of mechanisms to track direct competitors. Second, they did not use the MAUDE data to learn about major adverse events in these direct competitors. Instead, that information clearly travelled through other channels (ie. through personal networks and through sales representatives who talk to physicians and hospitals). With these insights in mind, we focused more directly on how they used the MAUDE data. Given our informants' concerns about the noisy data, we focused on the process by which they identified valid signals from the MAUDE data.

We began by focusing on how they thought about the MAUDE data. One clear theme was how our informants used the MAUDE data to understand the inherent risks of the medical devices they produced. A second, and related, theme was that they used the MAUDE data to develop responses to those risks. We explored these themes by collecting all instances in which they described the MAUDE data as particularly useful. From this collection, we started to assemble the instances around these themes and began to see patterns. The instances that followed the first theme were all assembled around failure modes, something our informants talked about. The instances around the second theme were somewhat more diffuse, but related to the protocols that they followed either in reporting to the FDA or in addressing the use of their devices. We then constructed a structure that focused on the various ways that the informants used each of these failure modes to identify adverse events from which they could learn.

TABLE 3
Descriptive Data of Interview Respondents

Organization Description	Role in Org	Job Title	Industry Codes*	Revenues*	No. of Employees*	Owner-Ship	Firm Age*	HQ Location	Interview Length (min)
< 1000 Adverse Events in MAUDE									
Technology Diversified MNE	Middle Manager	Director QA/RA	28	68	56	Public	30	Japan	1st: 73 2nd: 70
Healthcare- focused Diversified MNE	Middle Manager	Executive Director	14	Blank	Blank	Public	Blank	Canada	74
Healthcare Focused Family Business SME	Senior Manager	President	37	63	20	Private	52	USA	80
Small Biotech Company	Senior Manager	VP US Regulatory Affairs	13	Blank	2	Public	30	USA	99
Medical Device-focused Firm	Senior Manager	VP Sales	31	Blank	100	Private	28	USA	120
Medical Device Consulting Firm	Senior Manager	Senior VP for Regulatory, Quality and Development	31	43	22	Public		USA	1st: 65 2nd: 65
> 1000 Adverse Events in MAUDE									
Healthcare-focused Diversified MNE	Analyst	Regulatory Specialist	14	Blank	Blank	Public	Blank	Canada	71
Medical Device-focused MNE	Middle Manager	Senior Manager of Pre-clinical Affairs		Blank	100	Public	Blank	USA	60
Healthcare-focused Diversified MNE, Medical Device Division	Middle Manager	Senior Manager, US Regulatory Operations	14	Blank	Blank	Public	Blank	Canada	74
Technology Diversified MNE	Senior Manager	VP of Healthcare	18	77	77	Public	96	USA	56
Technology Diversified MNE	Senior Manager	VP, US Head of Technology and Quality	2	Blank	100	Public	100	Germany	43
Medical Device MNE	Senior Manager	Retired Director / Consultant	31	Blank	100	Public	Blank	USA	108

*Numbers disguised to protect confidentiality. **SME = Small Medium-Sized Enterprise, MNE = Multinational Enterprise, *** N=14 interviews

APPENDIX B: TRADE-OFFS IN INCLUSIVE FAILURE REPOSITORY DATA

As mentioned in the main body of the text, the data in a repository like the MAUDE data set includes data from a variety of sources that have not been vetted thoroughly from a third-party overseer. As a result, the data does not meet the standards of data from official reports that is often used to study learning (Haunschild and Sullivan 2002) and vicarious learning (Baum and Dahlin 2007). For analytic purposes, this presents challenges both in matching reports to firms and in treating missing and under-reported data.

Dataset Indexing and Matching

Every study drawing on multiple data sets needs to have indexes to match records across these datasets. Some of the data sets that we used were not indexed with each other, so matching reports to products across all firms was difficult. Such matching was especially challenging for the small firms that manufactured single medical devices, because their adverse events might appear as only a single observation amidst millions of adverse event reports. To deal with this challenge, we continually refined our matching algorithm across the course of our research. We traded off two possible approaches in creating our data set. An earlier approach that we have used matched as many records as possible across the data sets. This approach risked poorly matching the data by combining multiple records with similar firm names to the same report, thereby producing duplicate reports for individual firms.

The alternative algorithm, and the one we used here, provides a cleaner data set with distinct records. Our final algorithm involved a series of steps to generate unique IDs for each firm. These IDs were recorded by cycling through all possible matches from various combinations of fields in the data. We also 'cleaned' the data, and matched firm names based on a custom dataset of firm names. A check at the end of this procedure indicates that there were no duplicate records. Compared with other matching algorithms, this approach produced a 'cleaner' dataset that was more conservative in selecting only adverse event reports that were clearly unique. While this algorithm produced more reliably and consistently independent reports, it did limit the possible matches with records across the data sets.

Treatment of Missing and Under-reported of Data

A second challenge was dealing with missing or under-reported adverse events. When firms received possible evidence of an adverse event, they needed to make a judgment as to whether the event was worth reporting. Such judgments required dealing with the ambiguity of an adverse event. Manufacturers track reported incidents to them, but they only need to report adverse events to the CDRH when they believe there is probable cause that the device was the reason for the adverse event. If they believed the incident was due to user error they did not need to report the incident. Our interview respondents generally suggested they erred on the side of reporting, but medical device firms do not report failures. Not all adverse events were clearly attributed to a firm and its products and adjudicating whether reporting was necessary or valuable was a dynamic process as firms learned about events and their causes. As a result, some adverse events remain unobserved. These unobserved cases increased the difficulty of analyzing the data. Where most vicarious learning studies rely on data that has no missing or unobserved data, in our case both the dependent and independent variables are incomplete. This meant that we needed to deal with three combinations of unobserved dependent and independent variables.

1. Unobserved dependent variable and unobserved independent variable events

Given that some firms do not report adverse events, it is entirely possible that there are many events from which learning could occur but for which both the dependent and independent variables are unobserved. It is clearly impossible to make meaningful inferences about these cases from quantitative evidence because it is impossible to know if they exist and are missing. It is notable, however, that our qualitative evidence suggests that managers are actively trying to learn from others' failures—or absence of failures—even when they are not reporting adverse events. As we noted in our qualitative findings, managers noticed when they had failures that others did not report and tried to understand the reasons for the difference.

2. Unobserved dependent variable events and observed independent variable events

It is possible to address situations where the independent variables are observed but the dependent variable is only observed occasionally. Zero-inflated negative binomial regression, and rare events logit models have been implemented in Stata to estimate these events (Allison 2012a, b; King and Zeng 2001;

Winkelmann 2000). Rare-events logit models predict the occurrence of a binary event given excess zeros in the dependent variable. Zero-inflated negative binomial models predict the count of events given excess zeros in the dependent variable. This is a very active area of research with debate about whether such models are warranted and what the rules of thumb are for suggesting a variable has excess zeros. These models account for the rarity of observed values in the dependent variable (as manifest in an excess of zeroes) by creating a mixture model (or a penalty) combining the probability of those who report and those who do not report with the count of adverse events from those who report. We modeled the likelihood of the dependent variable being unobserved by using this two-stage model (ie. zero-inflated negative binomial model). The results were not substantially different (these are available upon request), but their validity depends upon the assumptions made about the first-stage predictors (see Allison 2012). Moreover, interpretation of the coefficients of such two-stage procedures is complex because the second-stage results are conditioned upon the significance and inclusion of covariates in the first-stage.

Another way to model a count variable with unobserved data is by imputing zeros in place of unobserved data and to simply run the regression within a negative binomial model. This can be done because the negative binomial model already adds sufficient parameter flexibility to account for problems like overdispersion from the excess zeros. As Allison (2012) discusses, the added complexity of these zero-inflated models may not be justified given that "the difference in fit [between zero-inflated and conventional negative binomial models] is usually trivial." Even so, we ran this simpler negative binomial panel data model (xtnbreg), including zeros for all non-reported observations. This model had both significant multicollinearity problems and led to model convergence problems.

A third approach to addressing problems with unobserved data is to analyze a sample of the population that meets a minimum threshold for reported events. We chose this method for our paper. Our first sample focused on all firms that experienced at least one adverse event in the 15-year window. This technique allowed us to generalize to major players in the medical-device industry (e.g., 3M, Abbott, B.Braun, Ethicon, Medtronic, Philips, Zimmer). However, like the larger sample, this sample produced multicollinearity problems. Most firms have long spells during which they report no failures because

failure reports cluster across time. Firms often report a large number of adverse events, followed by long periods of almost no adverse events. As a result, it is hard to address this variance with a larger sample.

In response, we restricted the sample even further. To obtain a sample that reported consistently, we focused on the sample of all firms (13) that reported adverse events in every month for 15 years from 1997-2012 (Alcon, Abbott, Boston Scientific, Cordis, CR.Bard, Depuy, Ethicon, Invacare, Medtronic, Smith & Nephew, St. Jude, Teleflex, Zimmer). This sample is the most conservative. These firms should *not* learn from each other because they manufacture different types of devices. If it appears that they do, then it is a strong case for our arguments.

We ruled out sample selection bias by conducting Heckman models (Wooldridge 2000) for all observations in our population (44,213 firm-month observations), including those that did and did not report adverse event reports. We first calculate the inverse mills ratio (i.e., the selection parameter) by regressing firm age, the number of products in that firm's portfolio, and whether health professionals submitted the reports on being selected into our sample. These variables predict selection because firms with better reporting capabilities based on age, larger size and professionalization are more likely to report and be included in the sample. We use a probit model for this calculation. In the second step, we regress all of the variables of interest and the inverse mills ratio on failure rates using random effects negative binomial regression. While the inverse mills ratio is significant, suggesting presence of selection bias, the model replicates our findings. Only the control variables change noticeably. Time, autocorrelation, others' new product introductions, and firm size become significantly negative. Regulations become significantly positive. The effect size of vicarious learning decreases from -0.84 to -0.30, but is still significant ($p<0.001$).

Observed dependent variable events and unobserved independent variable events

One way to deal with cases where the dependent variable is observed but the independent variable is not is to create some proxy indicator for the missing or zero observation. This indicator-type of model is preferred for step-function or "if, then" questions. For example, in a single-competitor setting, the

indicator provides coefficients for the impact of the missing observation and the effect of referent other's adverse events contingent on controlling for missing observations.

Our piecewise 'threshold' model is a more parsimonious solution that avoids the interpretive complexity of this kind of model. The proxy-indicator model treats the missing observations as conceptually different than the observed events. In our qualitative data we found that our respondents did not treat unobserved events as distinct from events with few observations. Instead, the managers we talked to were skeptical about missing observations and used those events to look into whether firms were reporting events as they should have.

Disclosure

The timing of disclosing one's own failures generally affected learning from one's own experience. A reporting delay of one day yields a 29% decrease in a firm's own adverse events in the following month (Model 8 of Table 5). Additionally, firms that postponed disclosure learned more from their own experience. A firm's delays in reporting significantly moderated its learning. As experience accumulated, the adverse event rate increased for fast reporters, but not slow reporters. This result suggests a tradeoff between a firm's own experience and the timing of disclosure, consistent with idea that delays could lead to insight (March, et al., 1991).

Learning from Injuries

Some failures are more consequential, and firms are more motivated to reduce failures that lead to injury. We used random effects negative binomial models to estimate the number of reported injuries. Models 9 and 10 in Table 5 confirm that the coefficient for vicarious learning is in the predicted negative direction (H1), even given reporting delays (H3). This result is consistent with our predictions.

APPENDIX C: CALCULATION OF MEASURES

Vicarious Learning Measure

We calculated vicarious learning as follows:

i. First, we calculate the firm's failure trajectory using the equation, $R_{i,t} = a_{i0}E_{it-1}^{b_{it-1}}$ (Argote 2013). The parameter a is the initial number of failures, E is cumulative failure experience at time *t-1*, and b is the slope for firm *i* based on experience until *t-1*. We modeled that the slopes of these trajectories change in each period by using a vector of recursively updating estimates for firm *i* in each period until *t-1*. This vector is estimated using an ordinary least squares regression with the equation $lnR_i = lna_{i0} + b_{ik} \cdot lnE_{ik}$ and for observations of *i* until *k* and $t_0 \leq k \leq t-1$. For example, the slope estimates for period 21 is based on experience from periods 1 thru 20, and the slope estimates for period 25 is based on experience from periods 1 thru 24[7].

ii. Second, we recursively calculate the correlation between a firm's slope and each other's slopes. We then take the mean of these correlations in each period.

Vicarious Learning Given Delays Measure

For each adverse event, the FDA requires a firm to report both the date of occurrence and the date of reporting. We calculated the number of days between occurrence and reporting for each social referent. The fields are the minimum between "Date Report (B4)" (date initial reporter provided information to the manufacturer) and "Date of Event (B3)" (actual or best estimate of the onset of the event) minus the "Date Received" (date received by FDA) in the "MDRFOI" dataset in MAUDE database.

We screened dates that occurred before the beginning of the dataset (January 1, 1997) because firms are not accountable for data entry errors. In reports with no event date, we classified the event time as either the earliest date by which the manufacturer was aware of the incident, the date by which the

[7] The rolling window specification is an alternative to recursive updating (Baum 2006), in which a fixed length of experience is used to generate estimates. We chose not to present this method because it assumes that experience terminates at an arbitrary point in the past. Nonetheless, in a robustness check using a rolling window specification, the results were similar for every window moving up by 12 month increments (ie. 12, 24, 36 months…), however, larger windows are more significant.

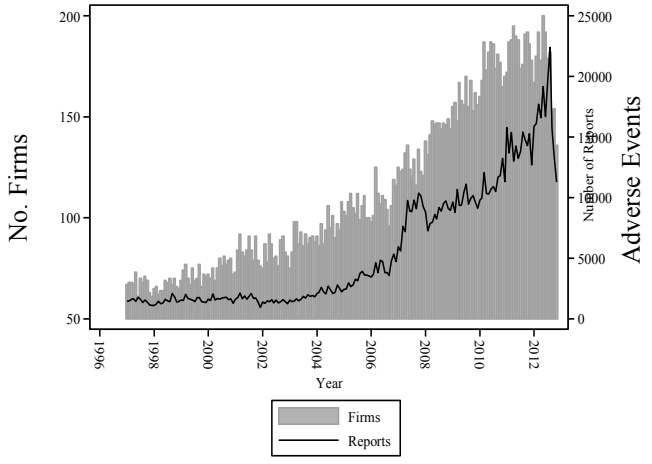
manufacturer received the faulty device, or the date by which the device was returned to the manufacturer for inspection after an incident. An audit of reporting between 2003 and 2008 by the Department of Health and Human Services found that 11% of adverse events reported by manufacturers exceeded 30 days (Levinson 2009). Our estimates of delays are longer because the audit used only the "Date of Event (B3)" field in initial reports, whereas we include both initial and follow-up reports, and use the minimum between "Date Report (B4)" and "Date of Event (B3)."

In our sample, the mean delay was 70.7 days with a standard deviation of 86.2 days. As shown in Figure 4, the average reporting delay was highest in 2004 (102.0 days) and occurred as a result of Ethicon laparosonic shears and absorbable sutures failures. This also had the maximum delay of 3.5 years. The month with the highest standard deviation in reporting also occurred in 2004 (sd= 166 days).

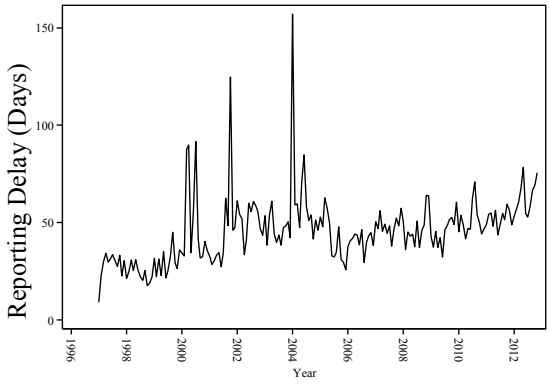
There are penalties when firms are caught delaying too long. For example, Guidant, a division of Boston Scientific, was fined the maximum civil penalty for intentional delays of $16,500 per offense of its faulty ICD leads, and paid a total of $296 million for its three-year delay in 2002. While enforcement is rare, firms are allowed to backdate experiences that they previously classified as a non-incident if new information suggests this is warranted (respondents also told us they did so on a regular basis).

FIGURE 4
(a) Number of adverse events and number of firms reporting adverse events, and (b) reporting delays (days) for firms in sample, 1997-2012

(a)



(b)



TECHNICAL DETAILS OF THESE MEASURES

We calculate vicarious learning given delays using a two-step estimator. The technique allows us to estimate the non-linear effects of delays, and the dependence between delays and reporting, while not directly estimating point estimates for a delay. The effect is observed by plotting a vector of estimates for $\widehat{\boldsymbol{\alpha}}_{\boldsymbol{j}} = (\hat{\alpha}_{j0}, \hat{\alpha}_{j10}, \hat{\alpha}_{j20} \ldots \hat{\alpha}_{jT})$ for all values of $\tau = 0{,}10{,}20 \ldots T$ days and lag of j. The effects of volume are estimated in the same fashion. We model two levels of learning; first accounting for firms' learning from their own experience and then accounting for vicarious learning.

Level 1: Firm-level Learning

In this level, firms learning from their own experience follow the learning curve, $R_{it} = a_{i0} E_{it-j}^{\boldsymbol{b}_{it-j}}$. The notation of this model is:

$R_{it} =$ No. adverse events reported by firm *i* at time *t*: $R_{it} \geq 0$.

$a_{i0} =$ Estimated initial No. adverse events reported by firm *i* at time t_0.

$E_{it-j} = \sum_{m=t_o}^{t-j} y_{im}$= experience is accumulated for firm *i* from t_o updated until $t - j$.

To model depreciation of knowledge, we include a depreciation parameter: $0 \leq \lambda \leq 1$. In this specification, experience is calculated as $E_{it-j} = \lambda E_{it-j-1} + R_{it}$. This becomes $E_{it-j} = \sum_{m=t_o}^{t-j} \lambda^{t-m} R_{im}$. λ is found using a grid-search algorithm, maximizing the log-likelihood.

$b_{it-j} =$ The slope of a failure trajectory for firm *i* based on experience until $t - j$. b_{it-j} is predictive of practices within the firm: $b_{it-j} < 0$ indicates that firms are reducing the rate of failure with experience; $b_{it-j} \geq 0$ indicates that a firm's number of failures is not changing or increasing with experience.

Estimation. The slope is estimated by doing the standard approach of taking the log of both sides and estimating using ordinary least squares regression: $lnR_t = lna_0 + b_{t-j} \cdot lnE_{t-j}$. To calculate the slope specific to firm *i* until time *t-j*, the sample is restricted to only those incidents reported by firm *i*. Because

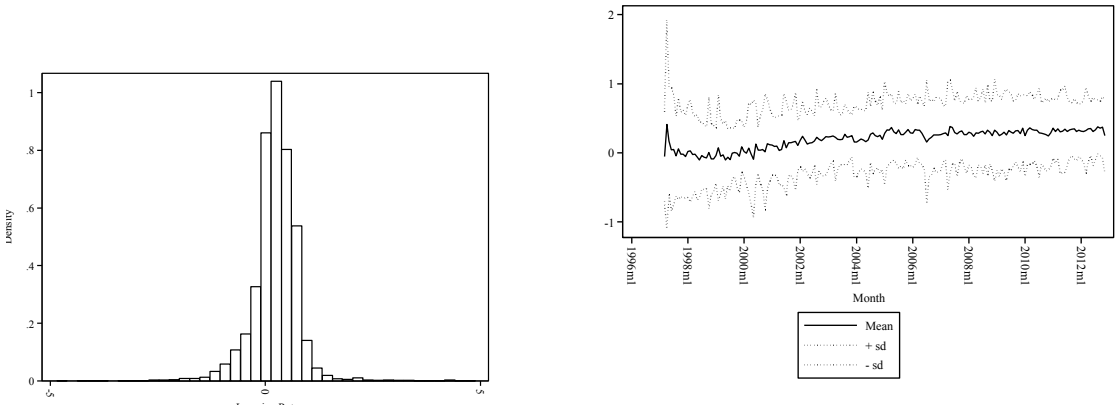
we wanted to observe how the slope changes in each period based on updating of experiences, we define the following:

$\boldsymbol{b}_{it-j} = \left(\hat{b}_{it_0}, \hat{b}_{it_0+1}, \hat{b}_{it_0+2}, \ldots, \hat{b}_{it-j}\right)$ = a vector of slope estimates for firm *i* in each period until *t-j*. Where, $\hat{b}_{ik}$ is estimated for observations of *i* until *k* and $t_0 \leq k \leq t-j$.

Figure 5 shows that the histogram and slopes of failure trajectory over time for all firms. The mean slope is generally centered around 0 (i.e. no learning is occurring).

FIGURE 5
(a) Histogram of slopes and (b) slopes over time from 1997-2012.

(a) (b)



Level 2: Vicarious Learning

We model vicarious learning by looking at the correlation of failure trajectories between firm *i* and all firms except *i* until period *t-j*. Using the estimates of slopes (level 1), we estimate the correlation between slopes on R_{it}. The estimation equation is:

$$R_{it} = X_{it}\gamma + \sum_{j=1}^{J}\left(\alpha_j r_{it-j}|_{D_{i't-j}>\tau}\right) + v_i + \varepsilon_{it} \qquad \text{(B1)}$$

$r_{it-j}|_{D_{i't-j}>\tau}$ = the correlation between the failure slope of firm i and the failure slope for all other firms except i from t_0 until $t-j$ with $D_{i't-j} > \tau$.

$= \frac{1}{n}\sum_{l=1}^{n} corr\left(\boldsymbol{b}_{it-j}, \boldsymbol{b}_{lt-j}\right)|_{D_{lt-j}>\tau}$, $\boldsymbol{b}_{it-j}$ is a vector of failure slope estimates for firm *i* in each period until *t-j* and $\boldsymbol{b}_{lt-j}$ is a vector of slope estimates for all other firms except *i* until *t-j* with reporting delays longer than τ.

$D_{i't-j}$ = Delay in reporting for all firms except i at $t-j$.

τ = threshold delay (ie. 10, 20, 30… days).

In other words, $r_{it-j}|_{D_{i't-j}>\tau}$ an estimate of the similarity of failure patterns until time *t-j*. The graph of the vector of point estimates of experience given delay tells us how vicarious learning is affected by delays. For example, if $\hat{\alpha}_{1,10}$ is negative and significant, we can infer that firms with positively correlated failure trajectories with social referents with reporting delays greater than 10 days are more likely to reduce their reporting of incidents. And, this effect takes one month to occur for firm *i*. This measure needs at least three consecutive months of failure observations to estimate the slope of the failure trajectory. There is no overlap in our calculation of past similarity and the rate of adverse events, as i) the correlation in *t-1* are calculated based on experience until period *t-2*, and then ii) the correlation in *t-1* predicts performance in period *t*.

Implementation in Stata

Level 1: Firm-level Learning. To estimate the slope of the failure trajectory without forgetting, the specific code in Stata is:

```
**************************************************************************
1       tsset f tid
2       sort f tid
3       gen a=.
4       gen b=.
5       qui forval i = 1/`nof' {
6               qui forval jt=`mint'/`maxt'{
7                       qui sum f if tid > `mint' & tid <= `jt' & f==`i'
8                       loc tot = r(N)
9                       if `tot'>2{
10                              reg lny l1.lnx l1.lny if tid > `mint' & tid <= `jt' & f==`i'
11                              replace b = _b[l1.lnx] if tid == `jt' & f==`i'
12                              replace a = _b[_cons] if tid == `jt' & f==`i'
13                      }
14              }
15      }
**************************************************************************
```

The first line of command, ‘tsset’ declares the data to be a panel for ‘f’ firms with the time variable being ‘tid.’ This allows us to use the lag function (l1.). Line 2 sorts that data in terms of firms and time. Line 3 and 4 initiate the a and b variables, where a is the initial number of failures per month and b is the slope of the failure trajectory.

Lines 5-15 specifies three loops. Line 5 initiates the first loop for firms 1 through the total number of firms (`nof’). Line 6 initiates the second loop for the window of experience. It specifies that the ending window starts at the minimum time (`mint’) and ends at the maximum time (`maxt’).

Lines 7 and 8 counts the number of observations in the window of experience for each firm. Line 9 specifies that there needs to be at least 3 observations to calculate the coefficients. Line 10-12 calculate the coefficients. Line 10 specifies that the natural log of y be regressed on the one month lag of the log of x, and the log of lagged y to account for autocorrelation. `reg’ specifies ordinary least squares regression. Where y is unit performance for the xth unit, a is the initial performance level to produce the first unit, x is the cumulative amount of experience, and b measures the slope of the failure trajectory. Line 11 and 12 record the parameters (a and b) for each window of experience in the a and b vectors.

Level 2 Vicarious Learning. The Stata code is:

```
**************************************************************************
16      gen mr=.
17      sort tid f
18      qui forval i = 1/`nof' {
19              gen ib1 = b if f== `i'
20              by tid: egen ib = sum(ib)
21              gen rbase = .
22              forval j=1/`nof'{
23              if `j' != `i' {
24                      forval jt=`mint'/`maxt'{
25      qui sum tid if tid > `mint' & tid <= `jt' & f == `j' & !missing(ib) & !missing(b)
26                              loc tot = r(N)
27                              if `tot'>2{
30                                      corr ib b if tid > `mint' & tid <= `jt' & f == `j'
31                                      replace rbase = r(rho) if tid == `jt' & f == `j'
32                              }
33                      }
34              }
35      }
36      by tid: egen mr1 = mean(rbase)
37      replace mr=mr1 if f == `i'
**************************************************************************
```

Line 16 declares the variable ‘mr’ as the mean correlation. Line 17 sorts the data based on time (tid) and the firm (f). Line 18 is the first of three loops, starting with looping through firm 1 to the total number of firms (`nof’). Line 19 and 20 generates a variable that compares slopes of firm *i* to all other firms. Line 21 declares the ‘rbase’ variable, which is the correlation of one specific firm to all other firms.

Lines 22-24 declare two more loops in which we compare to all other firms (*j*) which are not firm *i* and only for a window of experience from ‘mint’ to ‘maxt.’ Line 25-26 makes sure there is at least 2 observations to make the correlation. Lines 30 and 31 specify that the correlation for slopes between firm *i* and *j* is calculated and saved in the vector ‘rbase’ in period ‘jt’. Lines 36-37 takes the mean of the correlation for time period. To the calculate correlation given a delay greater than some threshold of ‘tau’, we include a fourth loop for increments of delays. Then 30 and 31 become:

```
corr ib b if tid > `mint' & tid <= `jt' & f == `j' & Delays > `tau'
replace rbase = r(rho) if tid == `jt' & f == `j' & Delays > `tau'
```

APPENDIX D: ADDITIONAL CONTROL VARIABLE DEFINITIONS, PAIRWISE CORRELATIONS, AND FULL RESULTS

Month. Our analyses include a time trend that captured month-to-month increases in the rate of adverse events. This time trend captures the effects of technological progress (Argote 2013) and other omitted variables (Thompson 2012), which may affect the stock of industry experience.

Diversification. Some firms operate in multiple product sectors. Depending on the number of products the firm has in each sector it will be exposed to different rates of failure experience. We use an entropy measure of each firm's diversification ($\text{DIVERSE}_{i,t-1}$) to accounts for the proportion of products a firm has in each medical sector.

Competition. Competition can potentially influence the learning rate of firms. We proxy competition as the firm's number of competitors in its main industry sector ($\text{COMPETITOR}_{i,t-1}$). Sectors are determined by a board of scientific advisors that evaluates devices for FDA approval.

Complexity. Firms also learn differently from simple versus complex failures (Haunschild and Sullivan 2002). In our study, firms may have one or many products that fail within or across product sectors. To capture this difference, we use an entropy-based measure of complexity that accounts for the proportion of adverse events, *p*, added each year by referents in each medical sector, designated *r*, and observed by firm *i*. We define $\text{COMPLEXITY}_{i,t-1}$ as $\sum_{i=1}^{n}\sum_{r=1}^{m} p_{i,r}\ln\,(p_{i,r})$.

Publicly Owned. Because public firms may be more closely scrutinized and/or incur greater reputational risks, we include a dummy variable to indicate whether each firm is publicly owned ($\text{PUBLIC}_{i,t-1}$). We obtained data on public ownership from COMPUSTAT and CRSP databases.

Failure-Related Controls

Size. A single product failure may be less relevant for firms with larger product portfolios than for firms with smaller product portfolios. We controlled for the log of the number of products each firm *i* manufactured in the past ($\text{SIZE}_{i,t-1}$) and for the portfolio of the average other firm ($\text{SIZE}_{j,t-1}$).

New Product Introductions ($\text{NPI}_{i,t-1}$). New products introduced by a firm may influence adverse events experienced by that firm. This measure controls for each firm's ability to anticipate and resolve potential

flaws by sequencing changes in product design. It is operationalized by the number of a firm's new product introductions within 365 days preceding the start of period *t*. We also controlled for the average number of new product introductions by other firms ($\mathrm{NPI}_{j,t-1}$).

***Severity*.** The incidence of death rates varies depending on the use of a product. Because firms pay attention to failures with more severe consequence, we included the number of deaths reported by each firm *i* ($\mathrm{SEVERITY}_{i,t-1}$) and its average other firm ($\mathrm{SEVERITY}_{j,t-1}$) from the past period *t-1*.

Regulatory Controls

Regulation ($\mathrm{REGULATION}_{i,t-1}$). FDA regulation influences firm behavior. This measure was operationalized as the number of independent FDA regulations experienced by each firm. We extracted regulatory information from the FDA's database on the number of product regulations each firm had to comply with in its main sector. The greater the number of regulations each firm had to comply with, the less latitude it had to vary the timing of its reporting. We also controlled for the Medical Device User Fee and Modernization Act (MDUFMA) of 2002 using a dummy variable.

Correlations

The pairwise correlations are reported in Table 4. Some of the variables exhibit high levels of collinearity. For example, industry average experience and industry experience stock have a correlation of 0.85 (See Thompson 2012: 214 for similar issues). High correlations are often observed in comparable research (e.g. Haunschild and Sullivan 2002; Kim and Miner 2007; Madsen and Desai 2010; Posen and Chen 2013; Strang and Patterson 2013) with correlations ranging as high as 0.97 in these studies. This is not surprising, as collinearity is a symptom of vicarious learning: firms are replicating or avoiding others' actions. However, in our setting, some of this multicollinearity is caused by the rare nature of adverse events. Recall that many of the covariates are properties of reported adverse events. This issue becomes particularly acute in larger samples where the lack of adverse events means that the different variables mostly take on zero values because the reported adverse events are treated as outliers relative to these zeros in a least squares regression. We addressed this through both sample selection and variable design:

1) Sample Selection. Multicollinearity is reduced by focusing on the 13 firms that had no zeros over the 15-year span. *2) Variable Design and Selection.* We carefully constructed our variables and iterative procedure to match our qualitative results. The variables of interest are never in the same regression model. In Table 4 we include the correlation (ρ=0.85) between vicarious learning and vicarious learning given delays greater than the mean (46 days) for illustrative purposes only.

We also performed an analysis which excludes the controls (not shown) to illustrate the effects are independent of multicollinearity in the controls. The main results persist independent of all other control variables used in the analyses, once accounting for the time trend and lagged dependent variable because of the path-dependent and cumulative nature of adverse events. This suggests that the results appear to be independent of other confounding factors and alternative specifications of control variables.

TABLE 4
Pairwise Correlations

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1. No. of Failures	$R_{i,t}$																						
2. Vicarious Learning	$E_{j,t-1}$	**.34**																					
3. Vicarious Learn | Delays	$E_{j,t-1,>46\ Days}$	**.33**	**.90**																				
4. Vicarious Learn | Volume	$E_{j,t-1,>69\ Reports}$	**.29**	**.85**	**.73**																			
Vicarious controls																							
5. Ind. Avg. Exp.	$IF_{j,t-1}$	**.24**	**.38**	**.43**	**.31**																		
6. Ind. Exp. Stock	$IS_{j,t-1}$	**.32**	**.47**	**.50**	**.45**	**.85**																	
Firm controls																							
7. Month	MONTH	**.38**	**.48**	**.52**	**.45**	**.92**	**.97**																
8. No. of Failures	$R_{i,t-1}$	**.89**	**.34**	**.33**	**.29**	**.23**	**.32**	**.38**															
9. Experience	$E_{i,t-1}$	**.69**	**.45**	**.42**	**.44**	**.33**	**.47**	**.54**	**.70**														
10. Disclosure	$d_{i,t-1}$	**-.04**	-.03	**-.05**	-.03	**.18**	**.16**	**.13**	.00	**-.13**													
11. Experience X Disclosure	$E_{i,t-1}$ X $d_{i,t-1}$	**.44**	**.30**	**.26**	**.29**	**.36**	**.45**	**.48**	**.49**	**.60**	**.70**												
12. Diversification	$DIVERSE_{i,t-1}$	**.12**	**.12**	**.17**	**.05**	**.21**	**.24**	**.27**	**.13**	**.37**	**.05**	**.32**											
13. Competition	$COMPETITOR_{i,t-}$	**-.07**	-.02	.02	-.03	**.18**	**.07**	**.07**	**-.08**	**-.28**	**.26**	-.03	**.11**										
14. Complexity	$COMPLEXITY_{i,t-1}$	**.20**	**.15**	**.17**	**.13**	**.10**	**.15**	**.17**	**.20**	**.34**	**.12**	**.33**	**.71**	**.11**									
15. Publicly Owned	$PUBLIC_{i,t-1}$	**-.10**	**.11**	**.09**	.00	**-.22**	**-.19**	**-.24**	**-.10**	**-.22**	**-.08**	**-.21**	**.19**	**.10**	**.31**								
Failure controls																							
16. Size, Firm	$SIZE_{i,t-1}$	-.03	**.10**	**.14**	**.06**	.00	.00	.01	-.03	.02	.02	.04	**.68**	**.34**	**.78**	**.59**							
17. Size, Referent	$SIZE_{j,t-1}$	**-.08**	**-.27**	**-.27**	**-.23**	**.05**	.02	.02	**-.08**	**-.10**	**.05**	-.04	**-.42**	**-.25**	**-.60**	**-.51**	**-.86**						
18. New Product Intro., Firm	$NPI_{i,t-1}$	**-.07**	.02	.02	**.06**	**-.17**	**-.14**	**-.15**	**-.07**	-.01	**-.24**	**-.20**	**.04**	**-.11**	**-.05**	**.08**	**.07**	**-.12**					
19. New Product Intro., Comp.	$NPI_{j,t-1}$	**-.19**	**-.29**	**-.29**	**-.31**	**-.45**	**-.43**	**-.45**	**-.19**	**-.32**	.03	**-.20**	**-.19**	**.05**	**-.07**	**.04**	-.04	**.06**	**-.32**				
20. Severity, Firm	$SEVERITY_{i,t-1}$	**.38**	**.30**	**.29**	**.24**	**.07**	**.19**	**.20**	**.40**	**.51**	.00	**.37**	**.46**	**-.24**	**.53**	**.17**	**.28**	**-.22**	**-.07**	**-.10**			
21. Severity, Referent	$SEVERITY_{j,t-1}$	**.18**	**.36**	**.35**	**.40**	**.72**	**.77**	**.76**	**.18**	**.30**	**.14**	**.31**	**.11**	**.05**	.04	**-.14**	**-.05**	**.06**	**-.14**	**-.49**	.04		
Regulatory controls																							
22. Regulation	$REGULATION_{i,t-1}$	**-.05**	**.06**	**.11**	.03	.01	.00	.00	**-.05**	-.01	.00	.01	**.68**	**.32**	**.77**	**.58**	**.99**	**-.84**	**.07**	-.04	**.27**	**-.04**	
23. MDUFMA	MDUFMA	**.27**	**.43**	**.44**	**.46**	**.74**	**.84**	**.83**	**.27**	**.43**	**.14**	**.40**	**.22**	.01	**.14**	**-.12**	.01	.01	**-.15**	**-.47**	**.21**	**.87**	.00

N=2,170 firm-month observations, significant correlations in bold ($p<0.05$)

TABLE 5
Adverse event rates for medical device firms, 1997-2012

		Adverse Events								Injuries	
		(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
H1: Vicarious Learning	$E_{j,t-1}$			-0.82***						-1.04***	
				(0.10)						(0.15)	
H2: Vicarious Learning | Volume	$E_{j,t-1,>69\ Reports}$				-0.40***						
					(0.06)						
H3: Vicarious Learning | Delays	$E_{j,t-1,>46\ Days}$					-0.30***			-0.28***		-0.22**
						(0.07)			(0.07)		(0.09)
H4: Vicarious Learning | Volume, Delays	$E_{j,t-1,>46\ Days\ >69\ Reports}$						-0.11*				
							(0.05)				
H4: Vicarious Learning | Volume, Delays	$E_{j,t-1,>46\ Days\ >200\ Reports}$							-0.13*			
								(0.06)			
Vicarious controls											
Ind. Avg. Exp.	$IF_{j,t-1}$		-0.05	-0.04	-0.08+	-0.05	-0.06	-0.03	-0.04	-0.04	-0.06
			(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.06)	(0.06)
Ind. Exp. Stock	$IS_{j,t-1}$		0.03	-0.08	0.02	-0.00	-0.10	0.18	0.03	0.02	0.09
			(0.13)	(0.05)	(0.13)	(0.13)	(0.16)	(0.23)	(0.13)	(0.18)	(0.17)
Firm controls											
No. of Failures	$R_{i,t-1}$	0.05	0.04	0.03	0.02	0.03	0.02	-0.02	0.03	-0.25***	-0.24***
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.04)
Experience	$E_{i,t-1}$	0.57***	0.57***	0.64***	0.63***	0.58***	0.59***	0.67***	0.48***	0.85***	0.81***
		(0.05)	(0.05)	(0.05)	(0.05)	(0.03)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)
Disclosure	$d_{i,t-1}$	-0.04***	-0.04***	-0.07***	-0.05***	-0.06***	-0.04*	-0.02+	-0.28***	-0.01	0.00
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.06)	(0.02)	(0.01)
Experience X Disclosure	$E_{i,t-1}$ X $d_{i,t-1}$								0.04***		
									(0.01)		
Month	MONTH	0.00**	0.00	0.00*	0.00+	0.00	0.00	0.09	0.00	0.01***	0.00*

		(0.00)	(0.00)	(0.00)	(0.00)	(0.06)	(0.07)	(0.09)	(0.00)	(0.00)	(0.00)
Diversification	$DIVERSE_{i,t-1}$	0.39***	0.40***	0.37***	0.37***	0.40***	0.41***	0.38***	0.36***	0.17***	0.17***
		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.03)	(0.03)
Competition	$COMPETITOR_{i,t-1}$	0.35***	0.35***	0.39***	0.39***	0.36***	0.37***	0.39***	0.38***	0.29***	0.26***
		(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Complexity	$COMPLEXITY_{i,t-1}$	-0.30***	-0.30***	-0.27***	-0.28***	-0.28***	-0.30***	-0.39***	-0.29***	-0.19***	-0.23***
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Publicly Owned	$PUBLIC_{i,t-1}$	0.04	0.04	0.04	0.03	0.01	0.02	0.01	0.01	0.03	-0.02
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.05)	(0.05)
Failure controls											
Size, Firm	$SIZE_{i,t-1}$	0.37***	0.38***	1.00***	0.68***	0.56***	0.72**	1.38***	0.75***	3.22***	2.90***
		(0.27)	(0.28)	(0.28)	(0.27)	(0.03)	(0.29)	(0.33)	(0.27)	(0.33)	(0.34)
Size, Referent	$SIZE_{j,t-1}$	-1.67***	-1.67***	-2.00***	-2.12***	-1.73***	-1.87**	-1.55**	-1.81***	-0.69	0.00
		(0.45)	(0.45)	(0.44)	(0.45)	(0.45)	(0.49)	(0.57)	(0.44)	(0.43)	(0.42)
New Prod., Firm	$NPI_{i,t-1}$	-0.06	-0.06	-0.01	-0.00	-0.03	-0.04	-0.15	-0.02	-0.08	0.09
		(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.07)	(0.06)	(0.05)	(0.06)	(0.07)
New Prod., Referent	$NPI_{j,t-1}$	-0.08	-0.09	-0.16+	-0.10	-0.13	-0.11	-0.11	-0.07	-0.56**	-0.49
		(0.13)	(0.14)	(0.14)	(0.14)	(0.14)	(0.15)	(0.18)	(0.14)	(0.19)	(0.19)
Severity, Firm	$SEVERITY_{i,t-1}$	0.05**	0.05**	0.04**	0.04**	0.04**	0.05**	0.04*	0.04*	0.18***	0.19***
		(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)
Severity, Referent	$SEVERITY_{j,t-1}$	0.03	0.03	0.04	0.04	0.03	0.02	0.01	0.03	0.02	0.02
		(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
Regulatory controls											
Regulation	$REGULATION_{i,t-1}$	-0.79**	-0.80***	-1.59***	-1.24***	-1.01***	-1.23***	-1.93***	-1.25***	-3.96***	-3.42***
		(0.30)	(0.31)	(0.31)	(0.30)	(0.30)	(0.32)	(0.37)	(0.31)	(0.38)	(0.38)
MDUFMA	MDUFMA	0.07	0.06	0.12*	0.14	0.07	0.12+	0.09	0.09	-0.03	-0.07
		(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.10)	(0.06)	(0.07)	(0.07)
	Constant	2.43	2.37	3.48+	3.97*	2.28	3.38	1.06	3.41+	-3.94*	-6.76***
		(1.86)	(1.86)	(1.83)	(1.85)	(1.89)	(2.04)	(2.52)	(1.84)	(1.87)	(1.85)
	Log-Likelihood	-11,367	-11,366	-11,331	-11,274	-11,351	-10,311	-7,893	-11,344	-9,674	-9,608
	Wald Test		0.52	71.6***	50.1***	20.4***	5.3*	4.7*	15.8***	50.9***	6.1*

N=2,170 firm-month observations. N=1,951 for model 6. N=1,445 for model 7. N=2,025 firm-month observations for injury models. ***, **, *, + represents $p<0.001$, $p<0.01$, $p<0.05$, and $p<0.1$. Wald Test based on the 'test' post-estimation command on variable of interest in Stata, where we test whether the coefficient differs from zero.