

A Statistical Framework to Detect and Quantify Operator-Learning Curves in Medical Device Safety Evaluation

Henry C Ssemaganda¹, Sharon E Davis², Usha S Govindarajulu³, Jejo D Koola⁴, Jialin Mao⁵, Dax Marek Westerman⁶, Amy M Perkins⁶, Theodore Speroff⁷, Craig R Ramsay⁸, Art Sedrakyan⁵, Lucila Ohno-Machado⁹, Michael E Matheny^{10,11,*}, Frederic S Resnic^{1,12,*}

¹Division of Cardiovascular Medicine and Comparative Effectiveness Research Institute, Lahey Hospital and Medical Center, Burlington, MA, USA; ²Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN, USA; ³Center for Biostatistics, Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, NY, USA; ⁴Department of Medicine, University of California San Diego, La Jolla, CA, USA; ⁵Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA; ⁶Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA; ⁷Departments of Medicine and Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA; ⁸Health Services Research Unit, University of Aberdeen, Aberdeen, UK; ⁹Department of Biomedical Informatics and Data Science, Yale School of Medicine, Haven, CT, USA; ¹⁰Departments of Biomedical Informatics, Biostatistics, and Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA; ¹¹Geriatrics Research Education and Clinical Care Service, Tennessee Valley Healthcare System VA, Nashville, TN, USA; ¹²Department of Medicine, University of Massachusetts Chan School of Medicine, Worcester, MA, USA

*These authors contributed equally to this work

Correspondence: Henry C Ssemaganda, Division of Cardiovascular Medicine and Comparative Effectiveness Research Institute, Lahey Hospital & Medical Center, 41 Mall Road, Burlington, MA, 01803, USA, Email semahen@gmail.com

Importance: Safety issues leading to patient harm and significant costs have been identified in several post-market medical devices. Recently, powerful learning effects (LE) have been documented in numerous medical devices. Correctly attributing safety signals to learning or device effects allows for appropriate corrective actions and recommendations to improve patient safety.

Objective: To develop and assess the statistical performance of an analytic framework to detect the presence of LE and quantify the learning curve (LC).

Design and Setting: We generated synthetic datasets based on observed clinical distributions and complex feature correlations among patients hospitalized at US Department of Veterans Affairs facilities. Each dataset represents a hypothetical early experience in the use of high-risk medical devices, with a device of interest and a reference device. The study blinded the analysis team to the data-generation process.

Methods: We developed predictive models using generalized additive models and estimated LC parameters using the Levenberg-Marquardt algorithm. We evaluated the performance using sensitivity, specificity, and likelihood ratio (LR) in detecting the presence of LE and, if present, the goodness-of-fit of the estimated LC based on the root-mean squared error.

Results: Among the 2483 simulated datasets, the median (IQR) number of cases was 218,000 (116,000–353,000). LE were detected in 2065 of the 2291 datasets for which learning was specified (sensitivity: 90%; specificity: 88%; LR: 7). We adequately estimated the LC in 1632 (81%) of the 2013 datasets in which LE was detected and estimated LC.

Discussion: This study demonstrated the framework to be robust in disentangling LE from device safety signals and in estimating LC.

Conclusion: In medical device safety evaluation, the operator-learning effects associated with the safety of medical devices can be effectively modeled and characterized. This study warrants subsequent framework validation by using real-world clinical datasets.

Keywords: post-market surveillance, generalized additive models, Levenberg-Marquardt algorithm, learning curve

Introduction

Medical device design faults or incorrect use may lead to a significant risk of patient injury and represent an important preventable public health risk in the United States.^{1–5} Safety issues have been identified in numerous medical devices



after market release,^{5–11} and the human health and cost burden associated with these device failures is very high.¹² The US Food and Drug Administration (FDA) has recommended active post-market surveillance strategies to address the limitations of voluntary adverse event reporting, which has been the mainstay of medical device safety.^{13–18} Specially designed software platforms, such as the Data Extraction and Longitudinal Time Analysis (DELTA) system that leverages clinical data repositories, have been validated to perform accurate, near real-time, active surveillance and monitoring of adverse outcomes for medical devices.^{19–23}

However, a significant methodological challenge for medical device safety surveillance is the detection and separation of learning effects (LE) from intrinsic device failures, leading to adverse events. Clinical providers frequently experience a learning curve (LC) as they incrementally master the use of novel medical devices and complex procedures that may be required for their successful use. Such learning can have a significant impact on outcomes, particularly during early experience with novel medical devices, representing up to 30–70% of early safety risk.^{24–31} The incremental risk due to LE as providers and institutions become proficient with a novel device may be falsely attributed to the device's intrinsic performance, leading to misattributed warnings for patients, clinicians, institutions, and regulators. Conversely, in cases where the intrinsic device performance is poor, a safety signal may be mistakenly attributed to learning effects, leading to the missed or delayed identification of an intrinsic device safety signal. For example, the adverse effects attributed to Cordis Cypher sirolimus-eluting coronary stents were later investigated and attributed to operator learning. The FDA subsequently addressed this issue through guidance on usage rather than recall.¹⁰ The main mechanism of failure in the Sprint Fidelis lead and ASR hip system recalls was attributed to flaws in device design, rather than operator or surgeon-related factors.^{32–36} The largest systematic review of more than 4500 learning impact studies in healthcare identified a general lack and inconsistency in the methods used to quantify learning.³⁷ The FDA has thus identified the impact of learning effects on procedure performance as an important area of study in post-market surveillance.¹⁸

Learning effects can be defined as a proportional improvement in performance, which may be measured in terms of the clinical outcomes following a medical procedure, with accumulating experience.³⁸

LCs have been established in implantable medical device procedures for example; gastrointestinal stenting, carotid arterial stenting, total coronary occlusions interventions and, vascular closure devices.^{27,29,39–42}

Other surgical procedures include; total hip arthroplasties, stapes surgery, arterial vascular access and, laparoscopic rectal cancer excisions.^{25,26,43,44} In these examples, LCs were quantified using a change in success rates with an increasing number of procedures. The other measures were volume-based for example, the decreasing average number of stents used per procedures with increasing case count. Nallamothe et al demonstrated higher outcome rates among operators with lower annual volumes and, higher outcome rates earlier versus later in a new operator's experience as a LC. Learning effects at both operator and institutional levels have been identified.^{24–26,40,42,45} A strong industrial learning literature describes that learning risk and experience follow a variety of decay curve relationships in which risk declines with increasing experience.³⁸ Using the NCDR (National Cardiovascular Data Registry) CathPCI registry Resnic et al demonstrated a triphasic LC with increasing experience in use of a novel medical device.⁴² Govindarajulu et al used complex simulated datasets to evaluate multiple methods for modeling LE and demonstrated that generalized estimating equations (GEE) performed better than generalized linear mixed-effect (GLME) models when modeling LE in hierarchical data. Additionally, Govindarajulu et al employed and evaluated several smoothing methods to represent the LC.⁴⁶ Using complex simulated datasets, Koola et al developed a machine learning-based framework to account for LE to derive an accurately adjusted intrinsic medical device signal.⁴⁷ These and other previous studies have demonstrated the challenges and value of rigorous modeling of LE and the quantification of LC. In this study, we used generalized additive models (GAM) to develop predictive models. Similar to GEE and GLME, GAMs can model clustered data. GAMs offer greater flexibility than GLMEs by modeling complex nonlinear relationships without prespecified functional forms. This flexibility makes GAMs more versatile than both GEEs and GLMEs for capturing intricate data patterns.⁴⁸

Across previous research, there has been limited exploration of the detection, quantification, and separation of LC effects from device intrinsic performance in medical device evaluation. The detection and quantification of LC are critical precursors during clinical data analysis to achieve the separation of LC from intrinsic device risk and to estimate learning-adjusted device safety profiles. To address this methodological gap, we extended our previously developed active surveillance analytic strategy,^{19–23} into a novel framework to process preliminarily identified safety signals by

determining if LE is present and, if so, subsequently estimating the LC functional form. As a precursor to assessment on real-world application using clinical data registries, we describe the methodological and validation approach used to develop and test the statistical performance of our flexible and robust framework in the setting of an active surveillance strategy.

Materials and Methods

Synthetic Data Development

To evaluate and calibrate our framework, selected team members (SED, MEM, and FSR) developed and validated the Data Generation Process (DGP). The methodology development team (including HCS, JDK, JM, AMP, and USG) was blinded to the underlying assumptions and specifications of each simulated dataset, which provided the ground truth. The DGP produced complex, synthetic datasets that mimic real-world clinical datasets representing high-risk medical device exposures and clinical outcomes.⁴⁹ The datasets incorporated a wide range of patient features, device safety signals and LE, including no LE (Refer to [Figure 1](#) for an illustration of the difference between device and learning effects). Each dataset was modeled using two medical devices: one device was identified as the device of interest (DOI), whereas the other served as the reference comparator device. The accruing operator experience for DOI and reference devices was represented by separate case orders.

Specification of the Data Generation Process

Patient Features

Patient features were simulated based on clinical data from the US Department of Veterans Affairs (VA). We collected information on a sample of inpatient admissions that lasted at least 48 hours, involved patients aged ≥ 18 years, and during which patients did not receive hospice care. We excluded admissions at VA facilities with fewer than 100 admissions per year and those that did not report key data to the central data warehouse. Our dataset included demographics, vital signs, medications, laboratory values, diagnoses, admission characteristics, and healthcare utilization (See [Table 1](#)).

Devices

We assigned hypothetical patients to one of two devices, considering one device to be an established treatment and the other to be a novel DOI for safety evaluation. The DOI was assigned to a mean of 20% (range: 15–25%; Standard Deviation (SD): 5%) or 50% (range: 45–55%; SD 5%) of patients, with one of 5 patterns of association between patient features and device assignment randomly selected to generate each dataset. Safety signals for the DOI were specified at

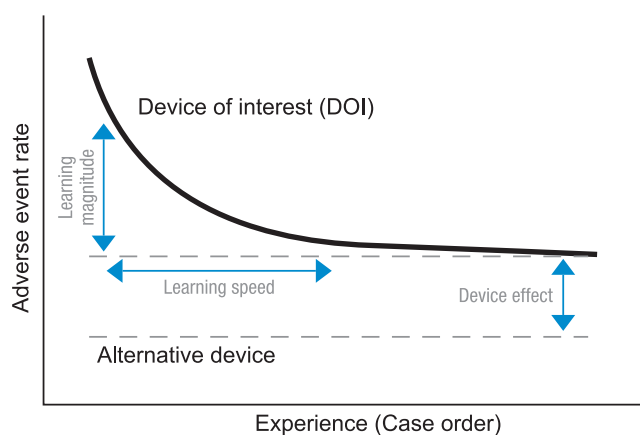


Figure 1 Illustration of the difference between device and learning effects. *Learning effect*-The device of interest (DOI) demonstrates a decrease in the adverse event rate with accruing operator experience; *Learning magnitude*-The increment in the base outcome rate attributable to learning effects with a peak at the earliest experience level; *Learning speed*-A measure of how much experience is needed for the peak learning magnitude to decrease by 95%; *Device effect*-The difference in adverse event rate between the alternative (reference) device and DOI, after adjusting for patient factors and learning effects.

Table 1 Overview of Select Patient Features in Reference Population Used by the Data Generation Process

Feature	Distribution in Reference Population
Age in years, median [IQR]	64 [56, 75]
Sex (% male)	96
Race (%)	
White	73
Black	22
Other	2
Unknown	3
Chronic kidney disease (%)	16
Chronic obstructive pulmonary disease (%)	30
Hypertension (%)	76
Myocardial infarction (%)	20
Insulin use in prior 90 days (%)	19
Statin use in prior 90 days (%)	49
Hemoglobin at admission, median [IQR]	12.4 [10.7, 14]
White blood cell count at admission, median [IQR]	7.7 [6, 10]
Encounters in prior year, median [IQR]	34 [18, 61]

Abbreviation: IQR, Interquartile Range.

four levels with a mean odds ratio of 1.0 (no signal, all 1.0), 1.25 (range: 1.2–1.3; SD: 0.1), 1.75 (range: 1.7–1.8; SD: 0.1), or 2.5 (range: 2.45–2.55; SD: 0.1).

Institutions

Patients in each dataset were assigned to an institution, with each dataset including either a mean of 25 (range: 20–30; SD: 9) or 75 (range: 65–85; SD: 9) institutions. Half of the institutions in each dataset were assigned a high volume (mean: 50 providers; range: 40–60; SD: 3) and half a low volume (mean: 15 providers; range: 10–20; SD: 3). No institutional-level learning effects were generated in this simulation.

Operators

Patients in each dataset were assigned to an operator. Each operator was assigned to patients from a single institution, and operators entered the time series of cases over the course of 3 years, with 50% of an institution's operators beginning at the start of the series, 25% entering in year 2, and 25% entering in year 3. The operators were assigned a high or low volume based on the annual number of cases. High-volume operators were assigned a mean of 100 patients (range: 85–115; SD: 4) per year, and low-volume operators were assigned a mean of 30 patients (range: 20–40; SD: 3) per year. The learning effects for patients receiving DOI were specified at the operator level in some datasets and not in others. When the learning effects were specified, they took a power-law, exponential, or Weibull form. The initial learning magnitudes were set to 25% (range: 20–30%; SD: 5%) or 50% (range: 45–55%; SD: 5%) of the mean probability of an adverse outcome due to patient features and device assignment. The learning effects dissipated as simulated operators achieved proficiency over their first 25 (range: 20–30; SD: 8) or 75 (range: 70–80; SD: 8) cases with DOI.

Outcomes

The association between patient features and outcomes was randomly selected from five patterns of association. The overall outcome rate in each dataset was set to a mean of 5% (range: 3–7%; SD: 2%) or 20% (range: 17–23%; SD: 2%). The conversion from probabilities to binomial outcomes generated noise; however, no additional noise was generated.

Other

No missing data were created, and all variables used in the assignment of the device and/or outcome were included in the final dataset available to the analysis team.

Statistical Framework

We developed a process to assess each synthetic clinical dataset for the evidence of LE. If LE was present, the functional form and magnitude of the LC were estimated (Refer to Figure 2 for the overall steps in the analytic workflow). The overall framework was implemented as a data analysis pipeline using the R programming language version 4.3⁵⁰ which was designed with customizable options to allow substantial configuration at each step.

Outcome Measures

The main outcome measures were the sensitivity, specificity, and likelihood ratio (LR) of correctly detecting the presence or absence of learning effects, and the proportion of datasets with an adequately quantified LC.

Learning Effect Detection

We used GAM models implemented in the “mgcv” R-package to build two models.⁴⁸ The first model was specified with a binary safety outcome (device failure) regressed on patient features and device used as the covariates, but without including operator experience or device order in the model. The second GAM model is similar to the first model but now includes the case order for the DOI (Device B). To specify no expectation of learning for the reference device (Device A), we set the case order for all reference device observations to the maximum case order for the DOI, presuming that the

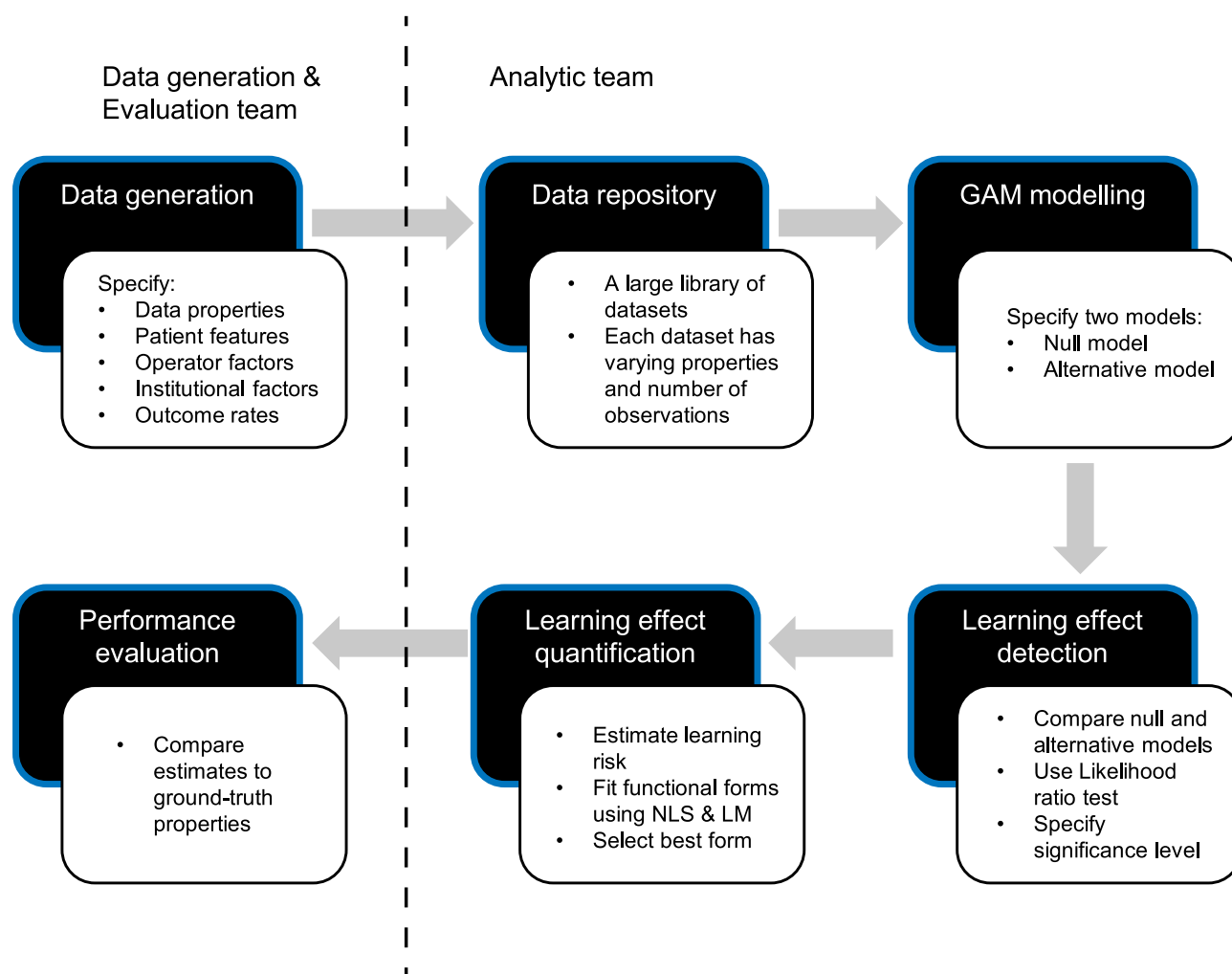


Figure 2 Overall steps in the analytic workflow to detect the presence of learning effects and quantify a learning curve. The analytic team was separated and blinded from the data generation process and performance evaluation.

Abbreviations: GAM, Generalized additive model; NLS, Non-linear Least Squares; LM, Levenberg-Marquardt algorithm.

highest observed case order for the DOI was at a point after proficiency was achieved, and any learning effects were extinguished. Both the models include operator clustering. To detect the presence of LE, we compared the two models using the Likelihood ratio test (LRT) to assess the goodness-of-fit where, the first model without case order was the null model.^{51–53} We tested the null hypothesis that both models fit the data equally well. A significance level of 5% and below was used to reject the null hypothesis and to conclude that the second model, which incorporated the device use case order, had a significantly better fit, indicating the presence of LE. Refer to [Appendix 1](#) for details on our considerations and specification of GAM.

Learning Effect Quantification

In cases where learning was detected, LE quantification was performed in which we sought to estimate the functional form of LC. We defined LC as the mathematical functional form that represents the decay relationship as LE decreases towards 0, as proficiency with the novel device is achieved. Firstly, we estimated the “LE adjusted risk” by using the second GAM model to derive predictions for each case where the DOI was used. Second, we estimated the “non-LE-adjusted risk” by deriving predictions from the same model and DOI dataset, setting the case order for each observation to be equal to the maximum case order value when the effect of learning on the probability of an outcome would be minimized. Third, we derived the “learning-associated risk” for each case as the difference between “LE adjusted risk” and “non-LE adjusted risk”.

Finally, we sought to characterize the relationship between case order (a covariate) and “learning-associated risk” (the dependent variable) by fitting decay functions to all DOI data. We assumed that if learning effects exist, they are represented by functional forms whose parameters can be estimated from one of several well-established learning-curve functional forms (see [Table 2](#)). We relied on the Non-linear Least Squares (NLS) approach using the Levenberg-Marquardt (LM) algorithm, as implemented in the software,^{54,55} to estimate the parameters for three pre-selected LC functional forms: Exponential, power law, and the four-parameter Weibull. The LM algorithm was implemented using an R-package-*minpack.lm* that provides an initial estimation for the starter values for each parameter.⁵⁴ When a form fit failed to converge, no recovery from failure was implemented. Instead, the next form was fit until all three forms were tried before proceeding with the form selection. We implemented a mechanism to select the best form based on multiple criteria, which included the speed of model convergence measured by the number of iterations required for the model to converge, significance of the parameter estimates (p-value<0.05), and lowest residual sum of squares. If the criteria were similar among functional forms, we chose the power law as a default form because of its wide usage in representing learning as well as the fact that this form has the least number of parameters.³⁷ We generated a 95% confidence interval of the estimated learning curve by performing the same NLS fit of the selected form on 10,000 bootstrap samples of the DOI dataset.

Evaluation of System Performance

The evaluation process was designed to assess the performance of the framework in terms of the statistical measures. Using synthetic datasets with known specified characteristics, our estimates were compared with the actual conditions (ground truth) specified by DGP. The performance of our LE detection approach was evaluated using the sensitivity, specificity, and likelihood ratio (LR). The learning curve quantification performance was measured as the proportion of datasets with adequate goodness-of-fit, defined as the root-mean-squared error (RMSE) between the estimated and

Table 2 Functional Forms and Statistical Characteristics Among Potential Learning Rate Equations

	Functional Form	Learning Rate $\partial f(x)/\partial x$	Initial Performance $f(1)$	Asymptote $\lim_{x \rightarrow \infty} f(x)$
Weibull	$a + b \exp(-cx^d)$	$-bcdx^{d-1} \exp(-cx^d)$	$a + b \exp(-c)$	a
Power law	$a + bx^{-c}$	$-cbx^{-(c+1)}$	$a + b$	a
Exponential	$a + b \exp(-cx)$	$-bc \exp(-cx)$	$a + b \exp(-c)$	a

ground truth (specified) LC. We categorized the RMSE values into three levels of adequate goodness-of-fit: a) highly conservative (≤ 1), b) moderately conservative (≤ 3.0), and c) less conservative (≤ 5.0).

This work was approved as exempt by the Tennessee Valley Healthcare System Veteran Administration (VA) Institutional Review Board (IRB) and the Vanderbilt University Medical Center IRB. The exemption was granted based on a determination that the study did not constitute human subject research.

Results

We generated 2483 synthetic datasets using combinations of the DGP specifications described above. Table 3 presents a summary of the characteristics of the synthetic datasets specified during DGP. These datasets include a median of 218,000 observations and 1500 operators. No learning effects were specified for 192 (7.7%) of the datasets. Among the 2291 datasets with specified learning effects, the functional forms assigned were evenly distributed among the Weibull (n: 767), power law (n: 761), and Exponential (n: 763) distributions. Approximately half of the datasets with specified LE had a large magnitude (n: 1146), and half incorporated fast learning speed (n: 1145).

Table 3 Characteristics of the 2483 Synthetic Datasets Created by the Data Generating Process

All Datasets (n=2483)	
Number of Operators, median (IQR)	1500 (792–2420)
Number of Observations, median (IQR)	218,000 (116,000–353,000)
Operator Learning – Form	
Absent	192 (7.7%)
Weibull	767 (30.9%)
Power law	761 (30.6%)
Exponential	763 (30.7%)
Operator Learning – Initial Magnitude^a	
Absent	192 (7.7%)
Small	1145 (46.1%)
Large	1146 (46.2%)
Operator Learning – Speed^b	
Absent	192 (7.7%)
Slow	1146 (46.2%)
Fast	1145 (46.1%)
Strength of device signal	
Absent	623 (25.1%)
Low (OR ~ 1.25)	622 (25.1%)
Medium (OR ~ 1.75)	618 (24.9%)
High (OR ~ 2.5)	620 (25.0%)
Base outcome rate^c	
p ~ 0.05	1241 (50.0%)
p ~ 0.20	1242 (50.0%)
Number of patient variables	
25 (n=1058)	1267 (51.0%)
50 (n=1020)	1216 (49.0%)

Notes: ^aLearning magnitude, defined as the absolute increase in adverse event rate in the first case involving B (small and large), corresponds to 25% and 50% of the mean probability of an adverse outcome due to patient features and device, respectively. ^bLearning speed is defined as the case order by which 95% of learning has occurred; slow and fast correspond to 75 and 25 cases, respectively. ^cBase outcome rate defined as adverse event rate when learning has been saturated.

Abbreviations: IQR, Interquartile range; OR, Odds ratio.

Learning Effect Detection

The LRT technique detected learning in 2065 of the 2291 datasets where learning was specified and in 24 of the 192 datasets where no learning effect was specified, yielding an overall sensitivity of 90.1% with an overall specificity of 87.5% (likelihood ratio: 7.2). The performance of the learning detection method was better when the magnitude of learning was high (sensitivity: 99.6%), the speed of learning was lower (sensitivity: 94.1%), and the specified learning effect functional form was exponential (sensitivity: 94.8%). Conversely, the method performed poorly when the outcome rate was low (sensitivity: 83.1%) and the magnitude of the learning effect was low (sensitivity: 74.9%), as shown in Table 4.

Learning Effect Quantification

The LM algorithm successfully estimated the learning curve in 2013 (98.1%) of the 2065 datasets in which learning was specified and detected. We correctly selected the specified functional form of LC in 50% of these datasets (see Table 5). None of the three LC functional form models converged in 52 of the 2065 datasets (convergence failure rate: 2.5%) where learning was specified and detected. Non-convergence was slightly higher in datasets with Weibull as a specified form (3.5%) compared with exponential (2.1%) and power law (2%).

Table 4 Performance Evaluation for Learning Effect Detection Stratified by Outcome Rate, Functional Form, Learning Magnitude and Speed

Dimension	Levels	Datasets (n)	Specified (n)	Detected (n)	Accuracy (%)	Specificity (%)	Sensitivity (%)	PPV (%)	NPV (%)	LR
Overall		2483	2291	2089	89.9	87.5	90.1	98.9	42.6	7.2
Specified Learning	Absent	192	0	24	87.5	87.5	–	–	100	
Learning	Present	2291	2291	2065	90.1	–	90.1	100	–	
Outcome Rate (Mean)	5	1241	1145	965	83.4	86.5	83.1	98.7	30.1	6.2
	20	1242	1146	1124	96.5	88.5	97.1	99	72	8.4
Functional Form	Power law	761	761	662	87	–	87	100		
	Weibull	767	767	680	88.7	–	88.7	100		
	Exponential	763	763	723	94.8		94.8	100		
Learning Magnitude ^a	Low	764	764	572			74.9	100		
	Medium	764	764	733			95.9	100		
	High	763	763	760			99.6	100		
Learning Speed ^b	Low	763	763	718			94.1	100		
	Medium	764	764	683			89.4	100		
	High	764	764	664			86.9	100		

Notes: ^aTertiles based on maximum event rate before experience is accrued. ^bTertiles based on the specified experience level when 50% learning was achieved.

Abbreviations: PPV, Positive predictive value; NPV, Negative predictive value; LR, Likelihood ratio.

Table 5 Comparison Between the Specified Functional Form and Form Selected

Learning Effects Specified by DGP as Present and Detected (N=2065)			
Specified Form	Datasets, n (%)	Fitted Form	Form Selected, n (%)
Exponential	723 (35)	Exponential	354 (49)
		Power law	43 (5.9)
		Weibull	311 (43)
		No convergence	15 (2.1)

(Continued)

Table 5 (Continued).

Learning Effects Specified by DGP as Present and Detected (N=2065)			
Specified Form	Datasets, n (%)	Fitted Form	Form Selected, n (%)
Power law	662(32.1)	Exponential	73 (11)
		Power law	359 (54.2)
		Weibull	217 (32.8)
		No convergence	13 (2)
Weibull	680 (32.9)	Exponential	290 (42.6)
		Power law	70 (10.3)
		Weibull	296 (43.5)
		No convergence	24 (3.5)

Abbreviation: DGP, Data generation process.

Table 6 provides an example of the output generated during the LC quantification stage for a synthetic dataset. The NLS models were fitted using all three pre-specified mathematical functional forms. However, for this dataset, the power law model failed to converge, whereas the Weibull and exponential models converged. We then applied our pre-specified criteria to identify the best fitting form, selecting the exponential form owing to the low number of iterations needed to converge, lower residual sum squared, and fewer parameters compared to the four-parameter Weibull. During evaluation, the functional form for the dataset in this example was correctly identified as exponential, and the estimated learning curve had an RMSE of 0.13 when compared to the ground-truth curve specified in DGP, representing a high goodness-of-fit.

Figure 3, we provide two illustrative examples of the best-fitting curve versus the underlying specified decay function. In one example (**Figure 3A**), the DGP specified an exponential function form and, while we estimated a power law curve, our estimated LC closely followed the specified curve, with an RMSE of 0.01. In the second example (**Figure 3B**), we estimated an exponential form for the LC that did not perform well compared to the specified Weibull LC (RMSE: 4.07).

Overall, as shown in **Table 7**, based on a highly conservative cutoff ($RMSE \leq 1.0$), 1632 (81.1%) of the 2013 datasets for which LE was estimated had an adequate fit compared to the specified LC. LC estimation performance was better with moderate learning speeds (93.4%) and a larger volume of DOI observations (91.2%). Conversely, the method performed poorly when the specified learning magnitude was high (59.5%), and when there were fewer DOI observations (69.4%).

At a moderately conservative cutoff ($RMSE \leq 3.0$), 1926 (95.7%) datasets had an adequate LC fit. A better performance was observed when there was a larger volume of DOI observations (99.7%). Conversely, slightly lower performance was observed in datasets with fast learning speeds (92.1%) and in datasets with a low volume of DOI observations (91.4%). At a less conservative cut-off ($RMSE \leq 5.0$), 1969 (97.8%) datasets had an adequate LC fit. Better

Table 6 Example of the Output From Learning Curve Functional Form Parameter Estimation

Functional Form	Parameter	Estimate	SE	p-value	95% LCI	95% UCI	Converged	Iterations	RSS
Weibull	a	0.0662	0.0004	<0.001	0.0654	0.067	Yes	29	54.34
	b	-0.1083	0.0041	<0.001	-0.1163	-0.1003			
	c	7.9988	0.3807	<0.001	7.2526	8.745			
	d	-0.6272	0.0241	<0.001	-0.6744	-0.58			
Exponential	a	-0.0017	0.0005	0.0014	-0.0027	-0.0007	Yes	23	54.32
	b	0.0725	0.0004	<0.001	0.0717	0.0733			
	c	0.0345	0.0006	<0.001	0.0333	0.0357			

Abbreviations: SE, Standard Error; LCI, Lower confidence interval; UCI, Upper confidence interval; RSS, Residual Sum Squared.

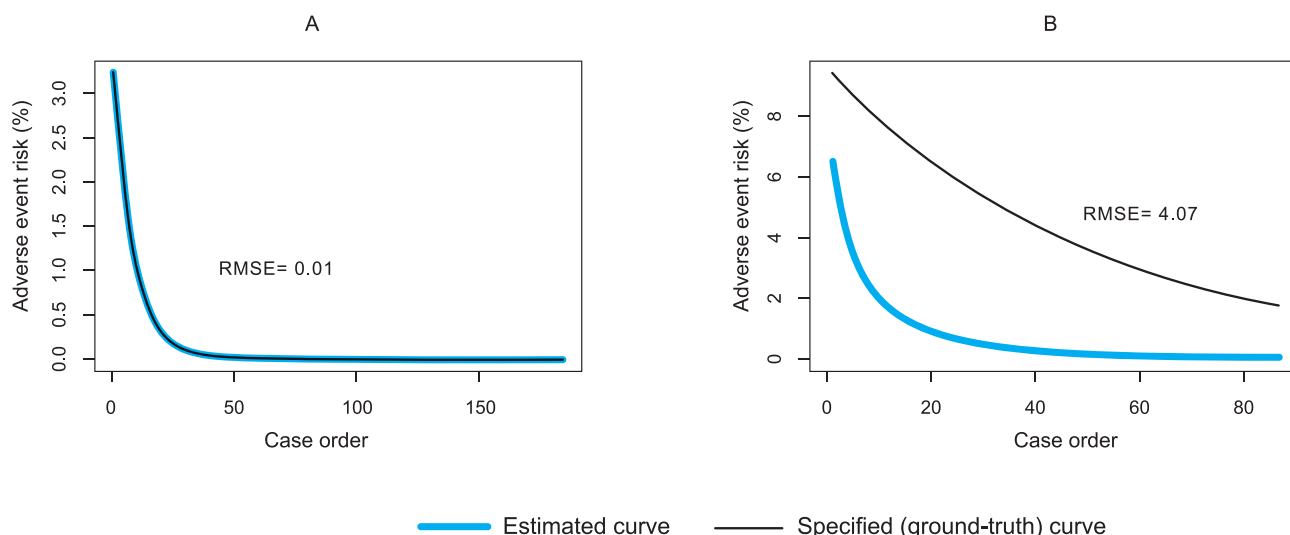


Figure 3 Illustrative examples. Illustrative examples using two synthetic datasets with specified learning effects that were detected and estimated by our framework. **(A)** DGP specified an exponential functional form, and we estimated the Power-Law with an RMSE of 0.01. **Notes:** The estimated superimposes the specified form. **(B)** DGP specified a Weibull functional form; we incorrectly estimated an exponential form with an RMSE of 4.07.

performance was observed in datasets with a larger volume of DOI observations (100%), slow learning speeds (99.1%), and when the specified functional form was power law (99.1%). Conversely, slightly lower performance was observed in datasets with fewer DOI observations (94.6%) and faster learning speeds (95.4%). Among the 2013 datasets for which we estimated LCs, only 44 (2.2%) had an RMSE > 5.0. These datasets were most commonly specified in the exponential form (63.3%), had a low volume of DOI observations (81.8%), and were specified with fast learning speeds (70.5%).

Discussion and Outlook

In this study, we addressed an important methodological gap in medical device safety evaluation by providing a framework for LE detection and LC quantification in clinical data analyses. This builds on prior post-marketing surveillance methods,^{42,46,56,57} by seeking to disentangle LE from intrinsic device safety risks. We pre-specified three functional forms of learning, developed criteria to assess and determine the best functional form based on observed data,

Table 7 Performance Assessment of the Estimated Learning Curve Using RMSE Stratified by Specified Functional Form, Patient Features, Number of DOI Observations, Outcome Rates Learning Speed and Magnitude

RMSE Cutoffs			<=1.0	<=2.0	<=3.0	<=4.0	<=5.0	>5
Dimension	Level	Total	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Overall		2013	1632 (81.1)	1849 (91.9)	1926 (95.7)	1962 (97.5)	1969 (97.8)	44 (2.2)
Specified functional form	Power-Law	649	464 (71.5)	587 (90.4)	628 (96.8)	642 (98.9)	643 (99.1)	6 (0.9)
	Weibull	656	560 (85.4)	615 (93.8)	632 (96.3)	643 (98)	646 (98.5)	10 (1.5)
	Exponential	708	608 (85.9)	647 (91.4)	666 (94.1)	677 (95.6)	680 (96)	28 (4)
Number of patient features ^a	Low	671	542 (80.8)	623 (92.8)	644 (96)	660 (98.4)	660 (98.4)	11 (1.6)
	Medium	671	543 (80.9)	608 (90.6)	638 (95.1)	651 (97)	653 (97.3)	18 (2.7)
	High	671	547 (81.5)	618 (92.1)	644 (96)	651 (97)	656 (97.8)	15 (2.2)
DOI Number of observations ^a	Low	671	466 (69.4)	567 (84.5)	613 (91.4)	630 (93.9)	635 (94.6)	36 (5.4)
	Medium	671	554 (82.6)	625 (93.1)	644 (96)	661 (98.5)	663 (98.8)	8 (1.2)
	High	671	612 (91.2)	657 (97.9)	669 (99.7)	671 (100)	671 (100)	0 (0)
Outcome rates ^a	Low	671	542 (80.8)	623 (92.8)	644 (96)	660 (98.4)	660 (98.4)	11 (1.6)
	Medium	671	543 (80.9)	608 (90.6)	638 (95.1)	651 (97)	653 (97.3)	18 (2.7)
	High	671	547 (81.5)	618 (92.1)	644 (96)	651 (97)	656 (97.8)	15 (2.2)

(Continued)

Table 7 (Continued).

RMSE Cutoffs			<=1.0	<=2.0	<=3.0	<=4.0	<=5.0	>5
Dimension	Level	Total	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Specified learning speed ^b	Slow	671	483 (72)	608 (90.6)	649 (96.7)	664 (99)	665 (99.1)	6 (0.9)
	Medium	671	627 (93.4)	653 (97.3)	659 (98.2)	663 (98.8)	664 (99)	7 (1)
	Fast	671	522 (77.8)	588 (87.6)	618 (92.1)	635 (94.6)	640 (95.4)	31 (4.6)
Learning magnitude ^c	Low	671	614 (91.5)	632 (94.2)	644 (96)	649 (96.7)	654 (97.5)	17 (2.5)
	Medium	671	619 (92.3)	645 (96.1)	654 (97.5)	658 (98.1)	659 (98.2)	12 (1.8)
	High	671	399 (59.5)	572 (85.2)	628 (93.6)	655 (97.6)	656 (97.8)	15 (2.2)

Notes: ^aDerived as tertiles. ^bDerived as tertiles of the number of cases at 50% of asymptote (a high number of cases means a slow learning speed). ^cTertiles based on maximum event rate before experience is accrued.

Abbreviations: RMSE, root-mean-squared error; DOI, Device-of-interest.

and demonstrated that the LM algorithm could be applied to estimate the parameters of an LC functional form. In addition, we introduced GAMs as a robust alternative modeling technique to the GEE and GLME models when modeling complex nonlinear relationships between adverse events and accruing experience. One of the strengths of this framework is that it is flexible and allows the inclusion of other mathematical LC functional forms as needed.

Overall, the performance of the framework was robust in detecting the presence of a learning effect and estimating a learning curve that reflected the specified curve when learning was determined to be present. We achieved good performance for the outcome measures in detecting the presence of learning and adequately estimated LC in datasets for which a curve could be estimated, depending on the conservativeness of the threshold used.

It is important to note that this framework employs parametric methods that require pre-specifications of the level of significance and the functional forms to consider. We used an alpha level of 5% during learning detection to achieve a higher sensitivity than specificity. The LE detection stage is a critical step in our active surveillance strategy, because the decision to seek the separation of learning from device effects relies on our ability to detect the presence of learning with high reliability. Therefore, a more conservative alpha level can be considered if the specificity is more critical than the sensitivity for a particular case.

There were several notable findings from the LE quantification stage, during which the algorithm estimated the LC by approximating the parameters of the pre-specified mathematical functional forms. As may be expected, based on a highly conservative threshold ($RMSE \leq 1.0$), the performance was better in datasets with high volumes of DOI observations and medium learning speeds. Lower performance in the presence of fast or slow learning speeds is an assumed limitation in curve fitting when the slope change over time is too rapid or too slow at fast and slow learning speeds, respectively. Similarly, the algorithm's performance may be limited in low-data-volume scenarios with relatively few observations of the DOI at each case order, which reflects the early adoption of a novel medical device.

We observed a slightly higher rate of convergence failure during LE quantification in the datasets with a specified Weibull form. We attribute this to the larger parameter space for the four-parameter Weibull form; it may be more difficult to identify suitable starter values to achieve convergence. The function may have multiple minima and becomes sensitive to meta-parameter starter values, causing a lack of LM algorithm convergence if meta-parameters tend towards local minima.⁵⁸ While we mitigated this possible issue by using all observations without summarization and, the randomness observed in noisy data may help the algorithm escape local minima, a systematic grid search for meta-parameters could be considered in future implementations of our analytic framework. Alternatively, in our future work we plan to implement and evaluate a new process for handling non-converging datasets. This approach will utilize a non-parametric pipeline, recently developed by Koola et al, to reprocess the datasets. By integrating the machine-learning pipeline, we aim to complement our approach by leveraging its flexibility and freedom from parametric assumptions to improve overall performance. Although power law is the most common functional form of LC, we found that the performance of the Weibull form was high. We interpret this as indicating that the algorithm may struggle in scenarios with heavy-tailed LC distributions (seen in slow learning) such as power law, compared with light-tailed distributions

(seen in fast learning) such as exponential. A generalized Weibull distribution is light-tailed and can be heavy-tailed when the shape parameter is between zero and one, making it versatile.

Applications on Real-World Data

Running experiments on data with known properties are an invaluable tool in health research and validation of novel methods.⁵⁹ Here we used synthetic datasets with properties that are unknown in real-world datasets, allowing us to validate and benchmark the performance of our approach under a variety of data properties and potential circumstances affecting real-world healthcare data. We also employed best practices to blind analytic team members to the underlying parameters and characteristics of the synthetic data in order to minimize experimenter bias. While this approach establishes the feasibility of our approach, synthetic data cannot fully reflect the real world, and as a result, the performance may differ in real-world data. We observed lower performance in datasets with limited observations, suggesting that low device adoption in real-world scenarios may lead to increased uncertainty in assessing learning curves. In future work, we will apply our statistical framework to real-world clinical device registry data to validate our novel LC detection tools against previously reported device safety concerns leading to recall, redesign, or updated training requirements. In real-world scenarios, causal ambiguity poses a significant challenge, making it difficult to discern whether device-related issues stem from inherent design flaws, operator factors such as learning curves, or a complex interplay between both. This ambiguity is further compounded by limitations like sparse data, unknown confounders, and intricate interactions.

Study Limitations

First, as noted above, synthetic datasets require some simplifying assumptions and cannot fully reflect the complexity of clinical observational data, which may influence the performance of our framework when applied to real-world scenarios. Multiple unknown operator, institutional, and patient factors, as well as device originality are challenging to simulate yet they contribute to real-world cause ambiguity. To address the limitations of synthetic datasets as part of our framework development strategy, we are conducting subsequent studies with real-world clinical data for further validation. Second, our approach assumes that LE is represented by only three pre-specified functional forms, which do not fully represent all potential learning curve mathematical forms or deal with unknown complexities of the experience-risk relationship. To address this limitation, the framework was built with flexibility to allow other functional forms to be considered in the LE quantification stage. Additionally, we included the Weibull functional form, which is highly versatile for modeling various forms and may be able to capture patterns similar to diverse LC forms. Third, the study evaluated operator-level learning effects but did not evaluate potential coexisting institutional learning effects. Finally, we relied on parametric methods that may not identify interaction terms and nonlinear relationships between risk covariates if not declared a priori.

Conclusion

We developed and validated a comprehensive analytical framework that demonstrates the ability to detect learning effects and estimate an accurate LC that reflects observed learning. This approach represents an important step towards fully disentangling the intrinsic device safety risk from the risk associated with the learning process. This study contributes to a broad interest in understanding the learning effects of active surveillance efforts on medical devices. The framework developed here is flexible and can be adapted outside medical device safety evaluations, including procedural outcome evaluations, where LE have also been observed. This study contributes towards our larger post-marketing active surveillance strategy with the goal of reducing the misinterpretation of device signals by correctly attributing devices and learning effect risks. In practice, this framework could improve signal interpretation during early adoption of novel devices, reducing false alarms that lead to costly recalls and better informing training programs and regulatory guidance. Estimating the magnitude and rate of learning effects could also help inform physician training protocols to optimize outcomes as new devices are adopted into clinical practice. The framework offers a data-driven method to identify potential safety risks related to learning effects, allowing regulatory agencies to assess these risks more effectively under

different statistical conditions. Additionally, the framework can be used for ongoing monitoring and evaluation, tracking the effectiveness of risk mitigation strategies over time as new data becomes available, and enabling adjustments to be made as needed. The framework's ability to quantify learning curves for a program or individual operators can inform the development of targeted training programs. In the case of Cordis Cypher stents, our framework would have analyzed eight years of data on Cordis Cypher stent usage to detect learning effects early on and understand the shape and characteristics of the operator learning curves. Conversely, in the Sprint Fidelis lead and ASR failure cases, the framework would demonstrate a possible lack of statistically significant correlation between operator experience and adverse event rates, suggesting that device-related factors were the primary contributors to the failures rather than operator issues.

The findings of this study warrant future research and validation by applying our approach to real-world examples, in which understanding device safety and learning are critical for protecting patient safety while ensuring access to successful novel medical advances.

Data Sharing Statement

The underlying code, a sample dataset simulated and analyzed during the current study is available in a public repository on GitHub accessed via this link-<https://github.com/CPHI-TVHS/learning-curve-quantification>.

Acknowledgments

Susan Robbins and Joshua Osmanski coordinated team efforts.

Funding

Research reported in this publication was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under award number (NHLBI; grant number R01HL149948). The funder played no role in the study design, data collection, analysis and interpretation of data, or writing of the manuscript.

Disclosure

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial or non-financial interests in the subject matter or materials discussed in this manuscript.

References

1. Samore MH, Evans RS, Lassen A, et al. Surveillance of medical device-related hazards and adverse events in hospitalized patients. *JAMA*. 2004;291(3):325–334. doi:10.1001/jama.291.3.325
2. Banerjee S, Campbell B, Rising J, et al. Long-term active surveillance of implantable medical devices: an analysis of factors determining whether current registries are adequate to expose safety and efficacy problems. *BMJ Surg Interv Health Technol*. 2019;1(1):e000011. doi:10.1136/bmjst-2019-000011
3. Garber AM. Modernizing device regulation. *N Engl J Med*. 2010;362(13):1161–1163. doi:10.1056/NEJMp1000447
4. Maisel WH. Unanswered questions—drug-eluting stents and the risk of late thrombosis. *N Engl J Med*. 2007;356(10):981–984. doi:10.1056/NEJMp068305
5. Hauser RG, Kallinen LM, Almquist AK, et al. Early failure of a small-diameter high-voltage implantable cardioverter-defibrillator lead. *Heart Rhythm*. 2007;4(7):892–896. doi:10.1016/j.hrthm.2007.03.041
6. Maisel WH. Semper fidelis—consumer protection for patients with implanted medical devices. *N Engl J Med*. 2008;358(10):985–987. doi:10.1056/NEJMp0800495
7. Resnic FS, Majithia A, Marinac-Dabic D, et al. Registry-based prospective, active surveillance of medical-device safety. *N Engl J Med*. 2017;376(6):526–535. doi:10.1056/NEJMoa1516333
8. Wykrzykowska JJ, Kraak RP, Hofma SH, et al. Bioresorbable scaffolds versus metallic stents in routine PCI. *N Engl J Med*. 2017;376(24):2319–2328. doi:10.1056/NEJMoa1614954
9. Ali ZA, Serruys PW, Kimura T, et al. 2-year outcomes with the Absorb bioresorbable scaffold for treatment of coronary artery disease: a systematic review and meta-analysis of seven randomised trials with an individual patient data substudy. *Lancet*. 2017;390(10096):760–772. doi:10.1016/S0140-6736(17)31470-8
10. Muni NI, Gross TP. Problems with drug-eluting coronary stents—the FDA perspective. *N Engl J Med*. 2004;351(16):1593–1595. doi:10.1056/NEJMp048262
11. U.S Food and Drug Administration. Medical device recalls [Internet]. Available from: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfres.cfm>. Access on November 17, 2024.

12. Schulte F, Christina J. Replacing faulty heart devices costs medicare \$1.5 billion in 10 years. The New York Times. Available from: <https://www.center4research.org/replacing-faulty-heart-devices-costs-medicare-1-5-billion-10-years/>. Accessed January 10, 2024.
13. U.S. Food and Drug Administration, HHS. Medical devices; device tracking. Final rule. *Fed Regist*. 2002;67(27):5943–5952.
14. Brown SL, Morrison AE, Parmentier CM, et al. Infusion pump adverse events: experience from medical device reports. *J Intraven Nurs*. 1997;20(1):41–49.
15. Fuller J, Parmentier C. Dental device-associated problems: an analysis of FDA postmarket surveillance data. *J Am Dent Assoc*. 2001;132(11):1540–1548. doi:10.14219/jada.archive.2001.0087
16. O’Shea JC, Kramer JM, Califf RM, et al. Part I: identifying holes in the safety net. *Am Heart J*. 2004;147(6):977–984. doi:10.1016/j.ahj.2004.03.001
17. Gross TP, Kessler LG. Medical device vigilance at FDA. *Stud Health Technol Inform*. 1996;28:17–24.
18. Strengthening our national system for medical device postmarket surveillance. silver spring. MD: food and drug administration. 2013. Available from: <https://www.fda.gov/media/84409/download>. Accessed January 2024.
19. Matheny ME, Ohno-Machado L, Resnic FS. Monitoring device safety in interventional cardiology. *J Am Med Inform Assoc*. 2006;13(2):180–187. doi:10.1197/jamia.M1908
20. Vidi VD, Matheny ME, Donnelly S, et al. An evaluation of a distributed medical device safety surveillance system: the DELTA network study. *Contemp Clin Trials*. 2011;32(3):309–317. doi:10.1016/j.cct.2011.02.001
21. Resnic FS, Gross TP, Marinac-Dabic D, et al. Automated surveillance to detect postprocedure safety signals of approved cardiovascular devices. *JAMA*. 2010;304(18):2019–2027. doi:10.1001/jama.2010.1633
22. Kumar A, Matheny ME, Ho KK, et al. The data extraction and longitudinal trend analysis network study of distributed automated postmarket cardiovascular device safety surveillance. *Circ Cardiovasc Qual Outcomes*. 2015;8(1):38–46. doi:10.1161/CIRCOUTCOMES.114.001123
23. Majithia A, Matheny ME, Dani SS, et al. Impact of early surveillance on safety signal identification in the CathPCI DELTA study. *BMJ Surg Interv Health Technol*. 2020;2(1):e000047. doi:10.1136/bmjst-2020-000047
24. Cook JA, Ramsay CR, Fayers P. Statistical evaluation of learning curve effects in surgical trials. *Clin Trials*. 2004;1(5):421–427. doi:10.1191/1740774504cn042oa
25. Masonis J, Thompson C, Odum S. Safe and accurate: learning the direct anterior total Hip arthroplasty. *Orthopedics*. 2008;31(12 Suppl 2):orthosupersite.com/view.asp?rID=37187.
26. Yung MW, Oates J. The learning curve in stapes surgery and its implication for training. *Adv Otorhinolaryngol*. 2007;65:361–369. doi:10.1159/000098861
27. Williams D, Law R, Pullyblank AM. Colorectal stenting in malignant large bowel obstruction: the learning curve. *Int J Surg Oncol*. 2011;2011:917848. doi:10.1155/2011/917848
28. Fok WY, Chan LY, Chung TK. The effect of learning curve on the outcome of caesarean section. *BJOG*. 2006;113(11):1259–1263. doi:10.1111/j.1471-0528.2006.01060.x
29. Thompson CA, Jayne JE, Robb JF, et al. Retrograde techniques and the impact of operator volume on percutaneous intervention for coronary chronic total occlusions an early U.S. experience. *JACC: Cardiovasc Interv*. 2009;2(9):834–842. doi:10.1016/j.jcin.2009.05.022
30. Weiss ES, Meguid RA, Patel ND, et al. Increased mortality at low-volume orthotopic heart transplantation centers: should current standards change? *Ann Thorac Surg*. 2008;86(4):1250–1260. doi:10.1016/j.athoracsur.2008.06.071
31. Hannan EL, Racz M, Ryan TJ, et al. Coronary angioplasty volume-outcome relationships for hospitals and cardiologists. *JAMA*. 1997;277(11):892–898. doi:10.1001/jama.1997.03540350042031
32. Maurer-Ertl W, Friesenbichler J, Holzer LA, et al. Recall of the ASR XL head and hip resurfacing systems. *Orthopedics*. 2017;40(2):e340–e347. doi:10.3928/01477447-20161213-04
33. Bernthal NM, Celestre PC, Stavrakis AI, et al. Disappointing short-term results with the DePuy ASR XL metal-on-metal total Hip arthroplasty. *J Arthroplasty*. 2012;27(4):539–544. doi:10.1016/j.arth.2011.08.022
34. Hauser RG, Kallinen LM, Almquist AK, et al. Early failure of a small-diameter high-voltage implantable cardioverter-defibrillator lead. *Heart Rhythm*. 2008;4(7):892–896. doi:10.1016/j.hrthm.2007.03.041
35. Birnie D, Farwell D, Green MS, et al. Accelerating risks of Fidelis lead fracture. *Heart Rhythm*. 2008;5(10):1375–1379. doi:10.1016/j.hrthm.2008.06.024
36. Ellenbogen KA, Wood MA, Swerdlow CD. The sprint fidelis lead fracture story: what do we really know and where do we go from here? *Heart Rhythm*. 2008;5(10):1380–1381. doi:10.1016/j.hrthm.2008.08.001
37. Ramsay CR, Grant AM, Wallace SA, et al. Statistical assessment of the learning curves of health technologies. *Health Technol Assess*. 2001;5(12):1–79. doi:10.3310/hta5120
38. Ramsay CR, Wallace SA, Garthwaite PH, et al. Assessing the learning curve effect in health technologies. Lessons from the nonclinical literature. *Int J Technol Assess Health Care*. 2002;18(1):1–10.
39. Nallamothu BK, Gurm HS, Ting HH, et al. Operator experience and carotid stenting outcomes in medicare beneficiaries. *JAMA*. 2011;306(12):1338–1343. doi:10.1001/jama.2011.1357
40. Verzini F, Cao P, De Rango P, et al. Appropriateness of learning curve for carotid artery stenting: an analysis of periprocedural complications. *J Vasc Surg*. 2006;44(6):1205–1212. doi:10.1016/j.jvs.2006.08.027
41. Warren BS, Warren SG, Miller SD. Predictors of complications and learning curve using the Angio-Seal closure device following interventional and diagnostic catheterization. *Catheterization Cardiovasc Interventions*. 1999;48(2):162–166. doi:10.1002/(SICI)1522-726X(199910)48:2<162::AID-CCD8>3.0.CO;2-2
42. Resnic FS, Wang TY, Arora N, et al. Quantifying the learning curve in the use of a novel vascular closure device: an analysis of the NCDR (national cardiovascular data registry) CathPCI registry. *JACC: Cardiovasc Interv*. 2012;5(1):82–89. doi:10.1016/j.jcin.2011.09.017
43. Goldberg SL, Renslo R, Sinow R, et al. Learning curve in the use of the radial artery as vascular access in the performance of percutaneous transluminal coronary angioplasty. *Catheterization Cardiovasc Diagn*. 1998;44(2):147–152. doi:10.1002/(SICI)1097-0304(199806)44:2<147::AID-CCD5>3.0.CO;2-6
44. Bege T, Lelong B, Esterni B, et al. The learning curve for the laparoscopic approach to conservative mesorectal excision for rectal cancer: lessons drawn from a single institution’s experience. *Ann Surg*. 2010;251(2):249–253. doi:10.1097/SLA.0b013e3181b7fdb0

45. Schillinger W, Athanasiou T, Weicken N, et al. Impact of the learning curve on outcomes after percutaneous mitral valve repair with MitraClip and lessons learned after the first 75 consecutive patients [published correction appears in *Eur J Heart Fail.* 2012 Jun;14(6):679]. *Eur J Heart Fail.* 2011;13(12):1331–1339. doi:10.1093/eurjhf/hfr141
46. Govindarajulu US, Stillo M, Goldfarb D, et al. Learning curve estimation in medical devices and procedures: hierarchical modeling [published correction appears in *Stat Med.* 2017 Nov 30;36(27):4420. doi: 10.1002/sim.7487]. *Stat Med.* 2017;36(17):2764–2785. doi:10.1002/sim.7309
47. Koola JD, Ramesh K, Mao J, et al. A machine learning framework to adjust for learning effects in medical device safety evaluation. *J Am Med Inform Assoc.* 2025;32(1):206–217. doi:10.1093/jamia/ocae273
48. Wood SN. *Generalized Additive Models: An Introduction with R.* 2nd ed. Chapman and Hall/CRC; 2017.
49. Davis SE, Ssemaganda H, Koola JD, et al. Simulating complex patient populations with hierarchical learning effects to support methods development for post-market surveillance. *BMC Med Res Methodol.* 2023;23(1):89. doi:10.1186/s12874-023-01913-9
50. R Core Team. *R: A Language and Environment for Statistical Computing 4.3.1.* Vienna, Austria: R Foundation for Statistical Computing; 2023.
51. Huang L, Zalkikar J, Tiwari R. Likelihood-ratio-test methods for drug safety signal detection from multiple clinical datasets. *Comput Math Methods Med.* 2019;2019:1526290. doi:10.1155/2019/1526290
52. Nam K, Henderson NC, Rohan P, Woo EJ, Russek-Cohen E. Logistic regression likelihood ratio test analysis for detecting signals of adverse events in post-market safety surveillance. *J Biopharm Stat.* 2017;27(6):990–1008. doi:10.1080/10543406.2017.1295250
53. Huang L, Zalkikar J, Tiwari R. Likelihood ratio based tests for longitudinal drug safety data. *Stat Med.* 2014;33(14):2408–2424. doi:10.1002/sim.6103
54. Elzhov TV, Mullen KM, Spiess A, et al. `_minpack.lm`: r interface to the levenberg-marquardt nonlinear least-squares algorithm found in MINPACK, plus support for bounds. R package version 1.2-4. Available from: <https://CRAN.R-project.org/package=minpack.lm>. Accessed January 2024.
55. More JJ. The Levenberg-Marquardt algorithm: implementation and theory. In: Watson GA, editor. *Lecture Notes in Mathematics 630: Numerical Analysis.* Berlin: Springer-Verlag; 1978:105–116.
56. Suri RM, Minha S, Alli O, et al. Learning curves for transapical transcatheter aortic valve replacement in the PARTNER-I trial: technical performance, success, and safety. *J Thorac Cardiovasc Surg.* 2016;152(3):773–780.e14. doi:10.1016/j.jtcvs.2016.04.028
57. Charland PJ, Robbins T, Rodriguez E, et al. Learning curve analysis of mitral valve repair using telemanipulative technology. *J Thorac Cardiovasc Surg.* 2011;142(2):404–410. doi:10.1016/j.jtcvs.2010.10.029
58. Gavin HP. The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems. Available from: <https://people.duke.edu/~hpgavin/lm.pdf>. Accessed November 2024.
59. Boulesteix AL, Groenwold RH, Abrahamowicz M, et al. Introduction to statistical simulations in health research. *BMJ Open.* 2020;10(12):e039921. doi:10.1136/bmjopen-2020-039921

Medical Devices: Evidence and Research

Publish your work in this journal

Medical Devices: Evidence and Research is an international, peer-reviewed, open access journal that focuses on the evidence, technology, research, and expert opinion supporting the use and application of medical devices in the diagnosis, monitoring, treatment and management of clinical conditions and physiological processes. The identification of novel devices and optimal use of existing devices which will lead to improved clinical outcomes and more effective patient management and safety is a key feature of the journal. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/medical-devices-evidence-and-research-journal>

Dovepress
Taylor & Francis Group