An update on statistical modeling for quality risk assessment of clinical trials

Björn Koneswarakantha[1], PhD • Timothé Ménard[1], PharmD

1. F. Hoffmann-La Roche AG - CH-4070, Basel, Switzerland

   corresponding author: timothe.menard@roche.com

Acknowledgement

Content review was performed by Rich Bowling who was employed by Roche/Genentech.

Compliance with Ethical Standards

Funding

Development of the statistical model was funded by Roche.

Conflict of Interest

Björn Koneswarakantha and Timothé Ménard were employed by Roche at the time this research was conducted.

Ethics approval

All human subject data used in this analysis were used in a de-identified format. Ethics approval is not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Code availability

https://github.com/openpharma/quality_risk_assesment_clinical_trials

Authors contributions

TM proposed the research. BK designed the method and drafted the outline. All authors wrote the manuscript. All authors approve the final version of the manuscript for submission and publication, and agree to be accountable for the work presented in the manuscript.

Study highlights / Key points

- What is the current knowledge on the topic?
  - Remote quality monitoring has gained momentum during the COVID-19 pandemic, e.g. for investigator site audits. It required additional efforts on both ends for document retrieval and generating reports. Thus, adapting risk-based strategies to narrow audit focus and to improve audit target selection has become a necessity.

- What question did this study address?
  - We correlated past audit and inspection findings with operational site data to identify risk factors for specific finding categories, using logistic regression.

- What does this study add to our knowledge?
  - We could identify 15 interpretable factors influencing audit finding risk of 4 out of 5 categories. They could be used to realistically predict differences in risk between 25 and 41% for different sites which suffices to rank sites by finding risk.

- How might this change drug discovery, development, and/or therapeutics?
  - Risk factors and resulting risk estimates could be used to complement remote clinical trials quality assurance strategies, and help to manage audit targets and audit focus also in post-pandemic times.

The preprint version of this work is available on MedRXiv:
https://doi.org/10.1101/2021.07.12.21260214

Abstract

Background - As investigator site audits have largely been conducted remotely during the COVID-19 pandemic, remote quality monitoring has gained some momentum. To further facilitate the conduct of remote Quality Assurance (QA) activities for clinical trials, we developed new quality indicators, building on a previously published statistical modelling methodology.

Methods - We modeled the risk of having an audit or inspection finding using historical audits and inspections data from 2011 - 2019. We used logistic regression to model finding risk for 4 clinical impact factor (CIF) categories: Safety Reporting, Data Integrity, Consent and Protecting Endpoints.

Results - We could identify 15 interpretable factors influencing audit finding risk of 4 out of 5 findings. They can be used to realistically predict differences in risk between 25 and 41% for different sites which suffices to rank sites by finding risk.

Conclusion - Continuous surveillance of the identified risk factors and resulting risk estimates could be used to complement remote QA strategies for clinical trials and help to manage audit targets and audit focus also in post-pandemic times.

Background

During the COVID-19 pandemic investigator site audits have largely been conducted remotely relying on digital technology. The need for off-camera support to retrieve documents and prepare reports has increased. To compensate for this additional effort, electronic operational site data could be leveraged to define a risk-based process for the management of audit scope and target. Site-specific quality indicators could be used to identify quality risk of specific sites. To validate potential quality indicators, we could estimate their influence on the probability of past audit findings. We have already shown that such an approach was feasible in the Good Clinical / Pharmacovigilance Practices (GCP/GVP) areas and that we could overcome challenges such as low signal-to-noise ratios and relatively small data sets with high ratios of missing values [1,2]. Having expanded our operational dataset, we have gathered historic investigator site audits and inspection findings from 2011 - 2019 and grouped them into 5 clinical impact factors (CIFs). We then generated a large set of operational features for each audited or inspected site and used them to model the probability of having a finding in a given category. Our goal was to obtain a set of interpretable features which contribute to fairly accurate risk estimates rather than building a classification model for audit findings. In total we could identify 13 site-specific quality indicators that are clearly associated with audit and inspection finding risk for 4 different CIFs. This was a major improvement over our last modelling iteration in which we could only identify 5 site-specific quality indicators. With the help of those indicators, we could improve remote site quality monitoring and ongoing risk assessments for clinical trials.

Methods

Audit finding and inspection data and all clinical trial data were gathered from Roche internal data sources while geographic population data was purchased from https://marketplace.namara.io. It contained 808 audits and inspections from over the course of 9 years. A full description can be found in the supplementary materials and in the code repository. The risk for audit and inspection findings was modelled as previously described [1]. Briefly, audit and inspection findings were assigned to one of five CIF categories. Several operational features (based on Adverse Events (AE), Issues and Deviations and Data Queries) were engineered to reflect the state of the site at the time of the audit or inspection. Operational

Features were complemented by site characteristics such as geographic population and study characteristics such as therapeutic area. To account for the number of patients and the individual study progress of each patient, we normalized features by either number of patient visits or total number of days that have passed for all patients since enrollment. As critical thresholds for quality indicators can be protocol specific, we also calculated the study percent rank for some features (e.g. percentage of missed visits and number of parallel trials) which indicated the percentage of sites in the same study that had a lower value.

To account for non-linear relationships between features and risk, continuous features were normalized using a Yeo Johnson power transformation [3] and binned into 5 groups with the same value range. For missing values all resulting binary bins were set to zero. Finding frequencies for each bin and categorical features were examined and promising candidates were preselected. The set of preselected features was narrowed down by fitting logistic regression models to a training data subset (2011-2015). We then iteratively removed uninterpretable features based on Subject Matter Expert (SME) review, and merged bins to facilitate their interpretation, to obtain a final feature set which was used to fit a model on the entire data set. We checked for multicollinearity between the features of each model by calculating the variance inflation factor (VIF) for each coefficient, and found that the maximum VIF per model was never greater than 2 (see Table. 101), which is well below the accepted critical threshold of 5 [4]. Sites that had missing values for all final features were removed before model fitting, otherwise for missing values all derived bins from that feature were set to zero.

To validate the models and to get a performance estimate, we used time series cross validation [5] in which data from each year would be used as a test set for a model fit with all data from previous years (see Fig. 1). The receiver operator characteristics area under the curve (AUC) and Brier Scores were calculated for each cross-validation test set and mean and standard error were calculated. As we were more interested in accurate risk estimates than in classification, we proceeded to fit a calibration model. The probability risk estimates of the test set predictions were divided into 4 bins of semi-equal range with a minimum of 100 test predictions per bin. For each bin, the predicted risk was averaged and the actual observed risk was calculated. To fit the calibration model, we performed a linear regression on the mean predicted risk versus the observed risk. In order to avoid extreme predictions by combinations of risk factors that are

unvalidated, we calculated a lower and upper range limit using the average observed risk of the 200 lowest and 200 highest risk estimates.

Results

Performance and Model Characteristics

We were able to model audit and inspection finding risk for 4 out of 5 CIFs which could estimate differences in risk between 25 and 41% for different sites with AUCs between 0.59 and 0.63 and Brier Scores from 0.19 to 0.25 (Table 1 and Fig. 2). While classification performance as measured by AUC is poor, the calibration of the native risk predictions as measured by the Brier Score with 0 (perfect accuracy) to 1 (perfect inaccuracy) can be described as adequate. A predictive difference in risk by 25% or more between sites has clear business value and can influence business decisions. We were not able to model the impact factor sponsor oversight properly (Table 1 and Fig. 2) most likely because we were lacking relevant features. The individual adjusted risk factors for each model have been plotted as tabular forest plots which also show the risk factor distribution in the training data (Fig. 3-6).

Features

The identified risk factors could be categorized into 7 groups (AE, Issues and Deviation, Data Queries, Geographical Populations, Parallel Trials and Study Characteristics, see Table 2). AEs, data queries, and issues and deviations represented frequent on-site events that were connected to heavily regulated operational processes and left a coherent data trail. Thus, event frequencies and processing times could serve as proxy measures for operational quality and were likely to influence the risk of findings. We have observed in the past that geographical location influences finding risk. Instead of using country-specific risk factors we chose to analyze characteristics of the population living within a 100km radius of a site as not all sites in a given country draw from a similar patient population and attract similar staff. Also, population-specific features were more descriptive and easier to interpret. For example, a low male to female sex ratio in the younger adult population (18-39 years) was a good surrogate indicator for sites in urban areas (Fig.7), as it tends to be easier for women to find adequate work in urban centers than in rural areas [6,7].

The other risk factor categories parallel trials and study characteristics are less influential and have already been previously identified [1].

Discussion

Interpretation

We were able to identify 15 quality indicators - of which 13 were site specific - that could be linked with an increase or decrease in audit and inspection finding risk. We could show this descriptively by looking at the model coefficients and we could also show the predictive quality of the risk estimates using time series cross validation. In order to get the most realistic average risk estimates, we have trained calibration models that help to balance the risk predictions. The final model features were the result of an hypothesis-driven selection since we involved SMEs in the process. An algorithmic feature selection would have been unrealistic in this low signal-to-noise scenario which would have resulted in uninterpretable false-positive features. Although we could not conclude that there was a causal relationship between the model coefficients and the modeled events, many of them were process-related and would have naturally been identified by SMEs during process monitoring. This made them easy to integrate into a risk-based quality monitoring data product.

Clinical Impact Factor: Consent (Fig. 3)

Informed consent requires paper signatures on forms that include the most up to date study information. This process is not yet fully digitized in most studies [8]. Therefore, a common auditing activity is to verify the paper signatures. If signatures were missing or were obtained too late a consent finding would be raised. The risk for such findings was increased for pediatric studies which require signatures from each parent. Moreover, when patients missed scheduled visits, they might not have been able to reconsented in time. Consequently, risk increases with higher percentages of missed visits.

Site issues and protocol deviations must be captured in an issue log and need to get processed within a specific timeframe. Interestingly, the number of issues that were closed late was indicative of an increase in consent risk, while a low number of minor deviations reduced the risk of consent findings. These risk factors did not seem to be directly connected to the consent

process and seemed to be purely indicative of site quality. They also influenced the risk of findings related to other CIFs (Table 2). An increase in consent risk could be found for sites located in densely populated areas. We could not see how this connected to the consent process. We did not find that population density was influencing any of the other risk models, thus we could only suspect that there was a strong unknown confounding variable that we were not yet capturing.

Clinical Impact Factor: Safety (Fig. 4)

AEs need to be adequately recorded into the medical database followed by a medical seriousness and causality assessment by qualified site staff. For serious AEs (SAEs) accelerated reporting timelines apply. Any detectable failure in this process would trigger safety findings. Accordingly, high rates of AEs increased the risk of safety findings while a low rate of SAEs or the absence of SAEs decreased risk. Independently of AE rates, cancer studies had an increased risk for safety findings as did sites that were located in a region with a high percentage of over 60-year-old. We could speculate that older or terminally ill patients were more likely to suffer from concomitant diseases which added additional complexity to the adequate AE capture and causality assessment.

Clinical Impact Factor: Data Integrity (Fig. 5)

Whenever there were mismatches between source data records and the clinical databases, a data integrity finding was raised. Invalid entries into the clinical database or a site monitor discovering questionable entries could trigger a data query that needed to be addressed by the site staff. We observed that if the normalized number of open queries per site was low, risk for data integrity findings was also decreased. Furthermore, a comparatively low number of active trials in the last year at the site and the number of issues that were due at the time of the audit both decreased data integrity finding risk. These last two risk factors were not directly connected to data integrity and thus should be viewed as generalizable site quality indicators. Interestingly, the ratio of males to females aged 18-39-year-old in the population around the site seemed to influence data integrity risk. A lower ratio was indicative of sites in an urban location while a high ratio seemed to indicate a more rural location as rural labor markets were primarily

populated by men [6,7]. We could speculate that urban site staff was more used to working with medical computer systems or that network connectivity is better at urban sites.

Clinical Impact Factor: Protecting the primary endpoints (Fig. 6)

Root causes for findings that were raised because the primary endpoints were at risk were numerous. Among the most frequent were inadequate study documents, mishandling of samples or the investigational medicinal product (IMP) and mismanaged protocol deviations. In this category we mostly identified risk factors that were indicative of overall site operational quality, some of which also influenced risk for findings for other CIFs such as number late issues, number of minor deviations and query processing time. Furthermore, a speedy reporting of AEs decreased risk as well which was potentially another operative site quality indicator.

Challenges and Limitations

Using more operational site features over the more static study features of our previous iteration [1] was a major improvement. This resulted in more diverse risk predictions for all sites in a given trial with risk estimates that would continuously adjust as sites continued to participate in the trial creating new data points. However, none of the new operational features correlated with sponsor-oversight-related audit and inspection findings. Useful features could probably be engineered from monitoring visits, source data verification and vendor management, but we have not yet been able to obtain that data electronically for a sufficient fraction of previously audited sites. The long observational time span during which data standards and IT systems have changed is the biggest limiting factor. The fact that we could not replicate the sponsor oversight model from our previous iteration [1] showed that our attempt to model risk in this low signal-to-noise ratio environment was not bias free, but we did need to make certain compromises in order to end up with interpretable risk models that could support business decisions. For example, strictly linear relationships between modelled risk and features were rarely detectable. In a lot of cases, risk would only increase or decrease once a threshold crosses a certain threshold (see AE per visit in Fig. 4) and we therefore resorted to binning all numerical features. Moreover, the large fraction of missing values for some features was not ideal. By setting all bins derived from a feature with a missing value to zero, we implicitly imputed missing values as belonging to the

reference bin. This carried some risk for bias which was a trade-off that we were willing to make in favor of including additional interpretable risk factors in our models. It was important to note that despite our effort to validate our models using time-series cross validation and calibration we have merely modelled the risk of historic investigator site audit and inspection findings. The recent pandemic has accelerated the decrease of traditionally conducted audits and inspections and remote quality activities are getting more and more common [9,10]. The number of traditional site audits in 2020 was already too low to include the data in this iteration. Sites audited in the past are often the sites that have recruited the most patients and thus carried the highest risk for the study outcome. Furthermore, findings were restricted to quality issues that could be identified by an auditing team on site. This selection bias limits the generalizability of the models we have created and in practice we display audit finding risk along with additional relevant operational site quality indicators that can be grouped under the same CIF. The risks and the quality indicators are integrated into a clinical analytics dashboard used by quality leads to manage audit focus and audit target selection.

Conclusion

As travel and physical access to sites is getting more restrictive, new complementing strategies for QA based on remote data analytics are emerging. Monitoring quality indicators and audit finding risk assessment could help to manage audit target selection and audit focus. To establish regulatory and industry trust, and to foster adoption of analytics-driven QA , we will continue to focus our effort on the development of novel methods QA [11,12,13,14] and on cross-company collaboration and data sharing [15].

Figures

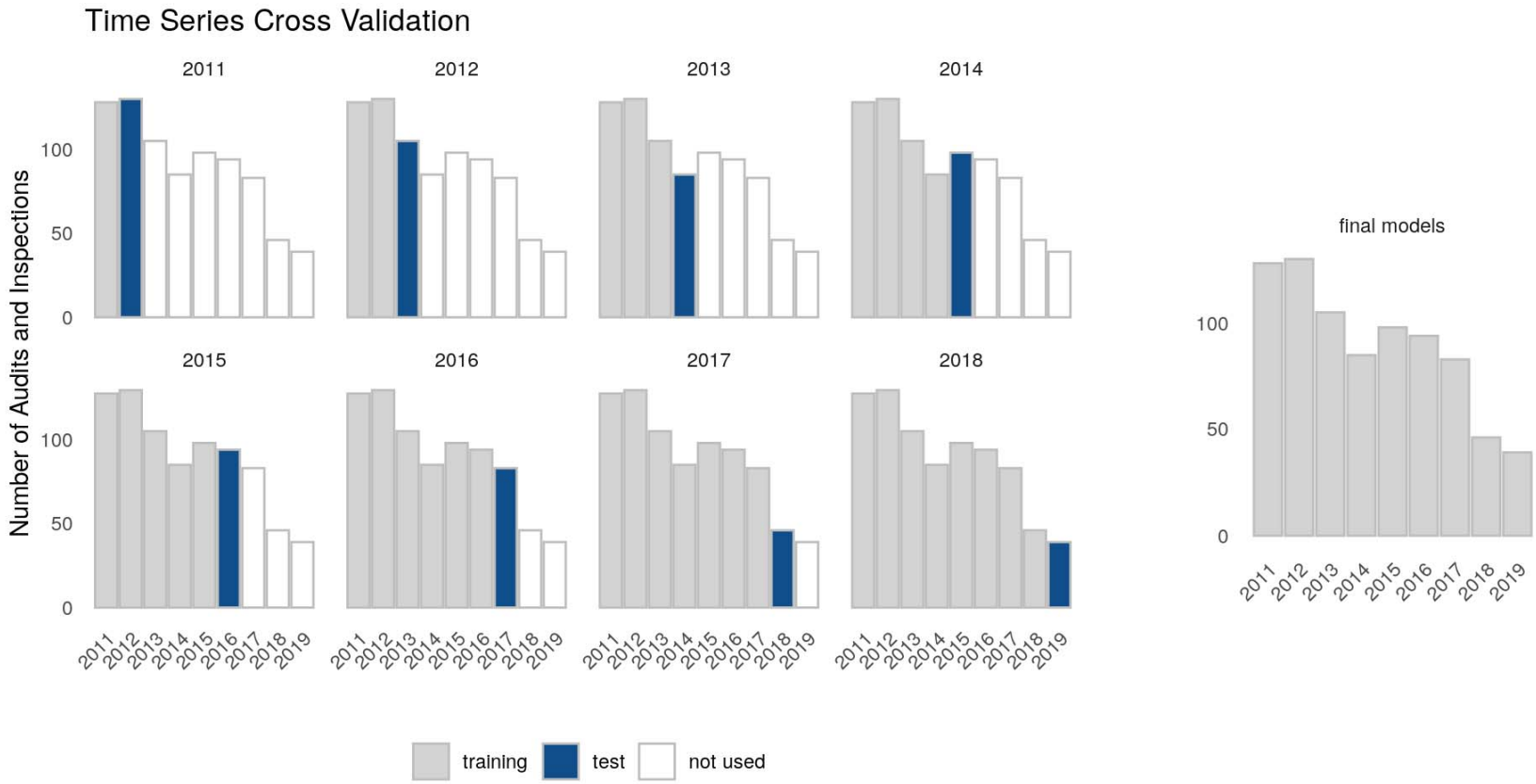


Figure 1 - Time Series Cross Validation

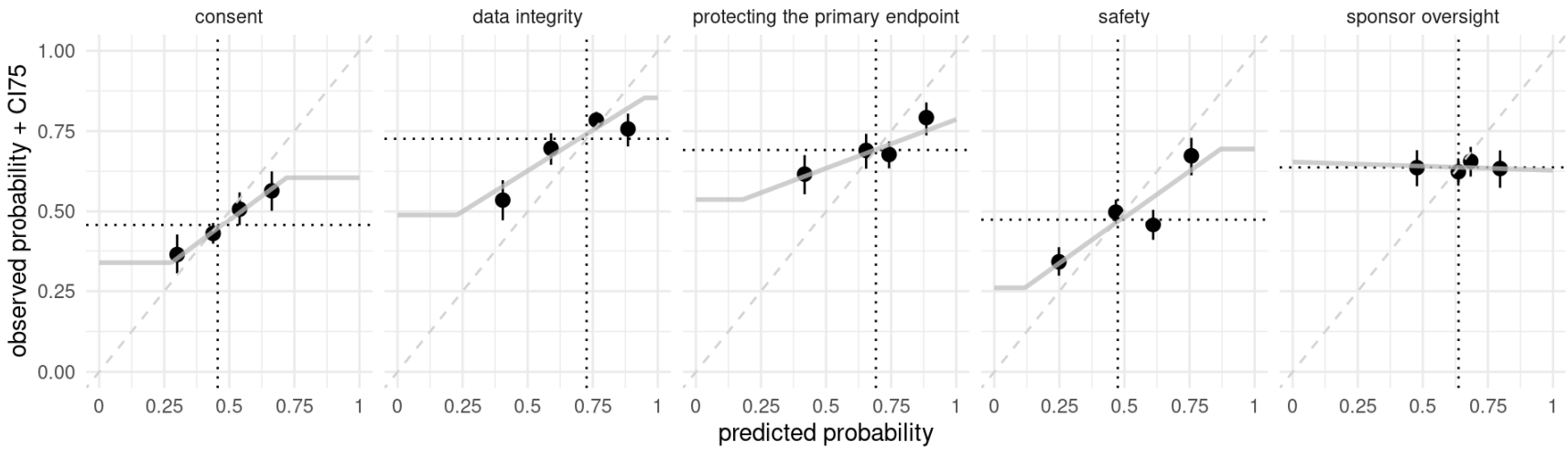


Figure 2 - Calibration

Consent Related Audit and Inspection Findings

Adjusted Risk Factors

	Distribution	N[1]	N with Findings[1]	Coefficient (CI95)[2]	Odds Ratio (CI95)	P	Coefficient Over Time (CI95)[2,3]
# Issues Late							
[0 - 4)		172 (21.2%)	60 (34.9%)		0.6 (0.4-0.9)	*0.01577 **	
[4 - 8)		85 (10.5%)	33 (38.8%)		0.7 (0.5-1.2)	*0.2554*	
[8 - 14)		60 (7.4%)	28 (46.7%)	reference			
[14 - 30)		71 (8.8%)	35 (49.3%)		1.2 (0.7-2.1)	*0.47143*	
[30 - Inf)		47 (5.8%)	27 (57.4%)		1.7 (0.9-3.2)	*0.12573*	
missing		376 (46.4%)	188 (50%)	reference			
# Minor Deviations Per Days on Study[4]							
[0 - 0.0029)		281 (34.6%)	111 (39.5%)		0.7 (0.5-1)	*0.0274 **	
[0.0029 - Inf)		147 (18.1%)	68 (46.3%)	reference			
missing		383 (47.2%)	192 (50.1%)	reference			
Pediatric Study							
Not Pediatric Study		756 (93.2%)	335 (44.3%)	reference			
Pediatric Study		55 (6.8%)	36 (65.5%)		2.4 (1.3-4.4)	*0.00421 ***	
missing		0 (0%)	0 (0%)	reference			
Perc Visits Not Occured (Percent Rank)[5]							
[0 - 0.5)		157 (19.4%)	59 (37.6%)		0.7 (0.5-1)	*0.03882 **	
[0.5 - 0.85)		131 (16.2%)	59 (45%)	reference			
[0.85 - Inf)		67 (8.3%)	33 (49.3%)		1.2 (0.7-2.1)	*0.44656*	
missing		456 (56.2%)	220 (48.2%)	reference			
Total Population[6]							
[8.16 thous. - 8.1 mil.)		436 (53.8%)	180 (41.3%)	reference			
[8.1 mil. - Inf)		205 (25.3%)	117 (57.1%)		2 (1.4-2.8)	*6e-05 ****	
missing		170 (21%)	74 (43.5%)	reference			

[1] Number of Clinical Trial Site Audits and Inspections

[2] Red Reference Line Indicates Zero

[3] Coefficient Value as New Data is Successively Added from 2011-2019

[4] Sum of Days Since Enrollment of All Patients

[5] Percent Rank of Site Within Study

[6] Population Within 100 km Radius of Site

Figure 3 - Clinical Impact Factor: Consent

Safety Related Audit and Inspection Findings

Adjusted Risk Factors

	Distribution	N[1]	N with Findings[1]	Coefficient (CI95)[2]	Odds Ratio (CI95)	P	Coefficient Over Time (CI95)[2,3]
% Population with Age over 60[4]							
[0.043 - 0.14)		75 (9.2%)	33 (44%)		1 (0.6-1.6)	*0.96856*	
[0.14 - 0.28)		485 (59.8%)	234 (48.2%)	reference			
[0.28 - Inf)		37 (4.6%)	26 (70.3%)		2.3 (1.1-5)	*0.02759 **	
missing		214 (26.4%)	98 (45.8%)	reference			
AE Per Visit							
[0 - 0.52)		325 (40.1%)	144 (44.3%)	reference			
[0.52 - Inf)		82 (10.1%)	50 (61%)		1.7 (1.1-2.9)	*0.02567 **	
missing		404 (49.8%)	197 (48.8%)	reference			
SAE Per Visit							
[0 - 0.0075)		171 (21.1%)	62 (36.3%)		0.6 (0.4-0.8)	*0.00362 ***	
[0.0075 - Inf)		236 (29.1%)	132 (55.9%)	reference			
missing		404 (49.8%)	197 (48.8%)	reference			
Therapeutic Area							
Autoimmune Disease Study		109 (13.4%)	41 (37.6%)		0.9 (0.5-1.5)	*0.72133*	
Cancer Study		460 (56.7%)	262 (57%)		1.7 (1.2-2.6)	*0.00815 **	
Neuro- Psychiatric Study		112 (13.8%)	36 (32.1%)		0.6 (0.4-1.1)	*0.1131*	
other Therapeutic Area Study		130 (16%)	52 (40%)	reference			
missing		0 (0%)	0 (0%)	reference			

[1] Number of Clinical Trial Site Audits and Inspections

[2] Red Reference Line Indicates Zero

[3] Coefficient Value as New Data is Successively Added from 2011-2019

[4] Population Within 100 km Radius of Site

Figure 4 - Clinical Impact Factor: Safety

Data Integrity Related Audit and Inspection Findings

Adjusted Risk Factors

	Distribution	N[1]	N with Findings[1]	Coefficient (CI95)[2]	Odds Ratio (CI95)	P	Coefficient Over Time (CI95)[2,3]
# Active Trials At Site In TA Past Year (Percent Rank)[4,5]							
[0 - 0.4)		313 (39.5%)	210 (67.1%)		0.6 (0.5-0.9)	*0.00521 **	
[0.4 - Inf)		391 (49.4%)	302 (77.2%)	reference			
missing		88 (11.1%)	63 (71.6%)	reference			
# Issues Due							
[0 - 2)		288 (36.4%)	189 (65.6%)		0.6 (0.4-0.8)	*0.00076 ****	
[2 - Inf)		147 (18.6%)	109 (74.1%)	reference			
missing		357 (45.1%)	277 (77.6%)	reference			
# Open Queries Per Days On Study[6]							
[0 - 0.0055)		179 (22.6%)	109 (60.9%)		0.5 (0.4-0.8)	*0.00066 ****	
[0.0055 - Inf)		162 (20.5%)	123 (75.9%)	reference			
missing		451 (56.9%)	343 (76.1%)	reference			
Male to Female Ratio in Population with Age 18-39[7]							
[0.46 - 0.49)		34 (4.3%)	20 (58.8%)		0.6 (0.3-1.3)	*0.20156*	
[0.49 - 0.5)		175 (22.1%)	121 (69.1%)		0.8 (0.5-1.2)	*0.29585*	
[0.5 - 0.51)		249 (31.4%)	183 (73.5%)	reference			
[0.51 - 0.52)		149 (18.8%)	123 (82.6%)		1.8 (1.1-2.9)	*0.01796 **	
[0.52 - Inf)		34 (4.3%)	28 (82.4%)		2 (0.9-5.6)	*0.12908*	
missing		151 (19.1%)	100 (66.2%)	reference			

[1] Number of Clinical Trial Site Audits and Inspections

[2] Red Reference Line Indicates Zero

[3] Coefficient Value as New Data is Successively Added from 2011-2019

[4] Percent Rank of Site Within Study

[5] Active Trials in the Past Year in the Same Theraputic Area

[6] Sum of Days Since Enrollment of All Patients

[7] Population Within 100 km Radius of Site

Figure 5 - Clinical Impact Factor: Data Integrity

Protecting The Primary Endpoint Related Audit and Inspection Findings

Adjusted Risk Factors

	Distribution	N[1]	N with Findings[1]	Coefficient (CI95)[2]	Odds Ratio (CI95)	P	Coefficient Over Time (CI95)[2,3]
# Issues Late							
[0 - 4)		172 (27.6%)	116 (67.4%)		0.9 (0.6-1.3)	*0.60298*	
[4 - 30)		216 (34.6%)	152 (70.4%)	reference			
[30 - Inf)		47 (7.5%)	37 (78.7%)		1.5 (0.7-3.4)	*0.27818*	
missing		189 (30.3%)	134 (70.9%)	reference			
# Minor Deviations Per Days on Study[4]							
[0 - 0.011)		355 (56.9%)	247 (69.6%)	reference			
[0.011 - Inf)		73 (11.7%)	58 (79.5%)		1.9 (1.1-3.7)	*0.03814 **	
missing		196 (31.4%)	134 (68.4%)	reference			
Median AE Reporting Delay (days)							
[0 - 27)		206 (33%)	131 (63.6%)		0.6 (0.4-0.8)	*0.00541 **	
[27 - Inf)		179 (28.7%)	136 (76%)	reference			
missing		239 (38.3%)	172 (72%)	reference			
Median Query Processing Time (days)							
[0 - 2)		207 (33.2%)	140 (67.6%)	reference			
[2 - 5)		52 (8.3%)	40 (76.9%)		1.6 (0.8-3.3)	*0.18389*	
[5 - Inf)		19 (3%)	16 (84.2%)		2.2 (0.7-9.9)	*0.21083*	
missing		346 (55.4%)	243 (70.2%)	reference			

[1] Number of Clinical Trial Site Audits and Inspections

[2] Red Reference Line Indicates Zero

[3] Coefficient Value as New Data is Successively Added from 2011-2019

[4] Sum of Days Since Enrollment of All Patients

Figure 6 - Clinical Impact Factor: Protecting the primary endpoint

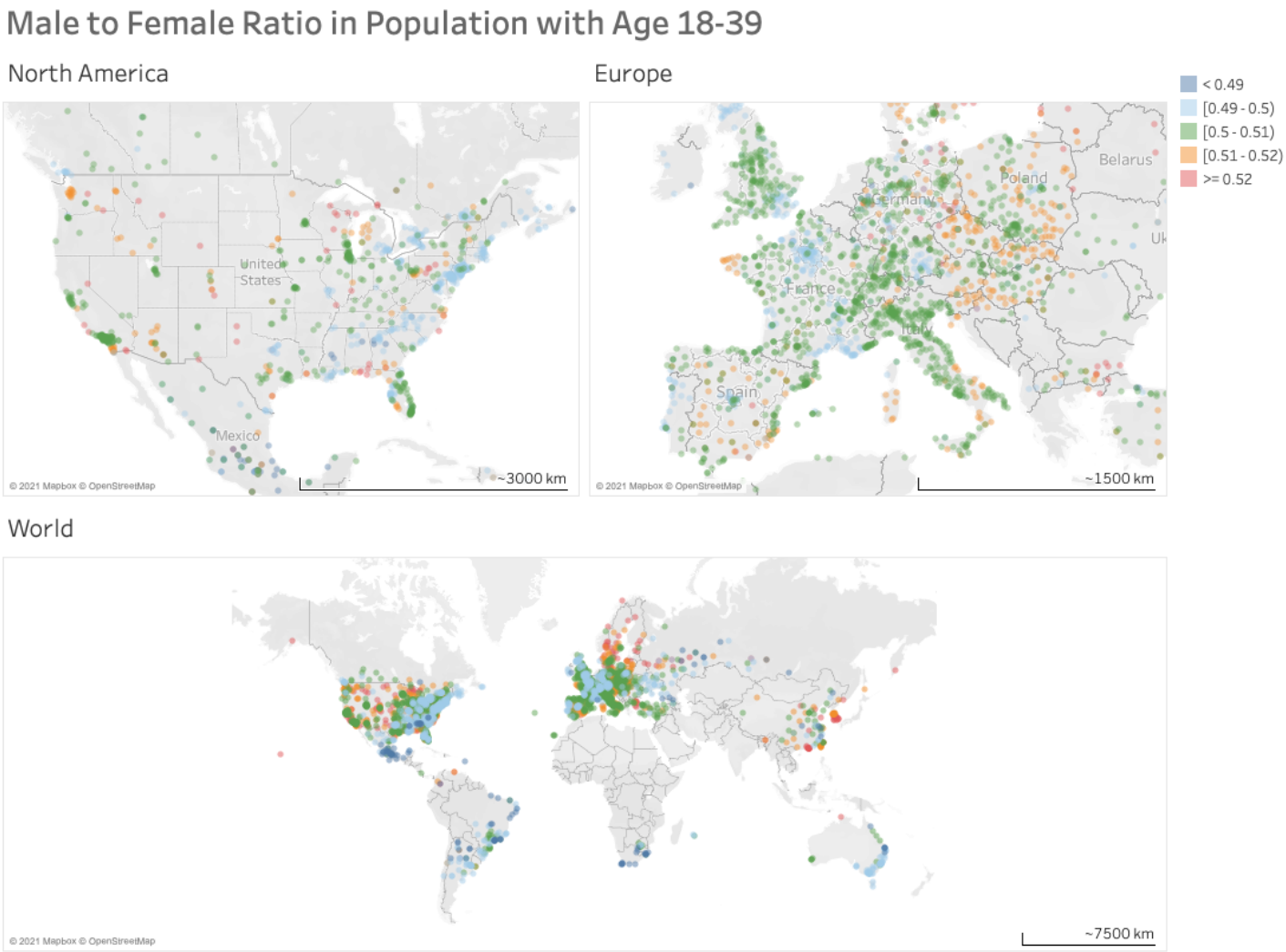


Figure 7 - Male to Female Ratio in Population with Age 18 - 39

Tables

Clinical Impact Factor	Maximum VIF	Mean AUC ±SE	Mean Brier Score ±SE	Calibrated Prediction Range %	Base Rate Probability %
Consent	1.25	0.61 ± 0.15	0.24 ± 0.01	34-61(∆ 27)	46
Data Integrity	1.11	0.60 ± 0.1	0.19 ± 0.02	49-85(∆ 36)	73
Protecting Endpoints	1.05	0.59 ± 0.06	0.23 ± 0.01	54-79(∆ 25)	69
Safety	1.97	0.63 ± 0.7	0.25 ± 0.01	26-69 (∆ 43)	47
Sponsor Oversight	1.05	0.53 ± 0.06	0.24 ± 0.01	63-65 (∆ 2)	64

Table 1. Mean Modelling Performance per CIF Model - Mean AUC and Brier Score Including Standard Error (SE) were Calculated Based on Test Set Predictions Derived from Time Series Cross-Validation Strategy with One Value per Year from 2011 - 2018

Clinical Impact Factor				
Feature Group	**Consent**	**Data Integrity**	**Protecting Endpoints**	**Safety**
Issues and Deviations	▲ # late issues ▲ # minor deviations per days on study	▲ # overdue issues at time of audit	▲ # late issues ▲ # minor deviations per days on study	
Geographical Population[1]	▲ total population	▲ male to female ratio in 18-39 age group		▲ % over 60 years
Patient AEs and Visits	▲ % visits not done		▲ AE reporting time	▲ AE per visit ▲ SAE per visit
Data Queries[2]		▲ open queries per days on study	▲ query processing time	
Parallel Trials[3]		▲		
Study Characteristics	▲ pediatric			▲ cancer ▼ neuro-psychiatric ▼ autoimmune

[1]Living Within a 100km Radius Around Site
[2]Automatic or Manual Data Queries directed towards Site by Sponsor
[3]Active Roche Trials in the Past Year in the Same Therapeutic Area

Table 2. Features Contributing to each CIF Model - Features Generally Correlate Positively with Audit Finding Risk Unless Indicated Otherwise (Green Downward Arrow).

References

1. Koneswarakantha, B., Ménard, T., Rolo, D. et al. Harnessing the Power of Quality Assurance Data: Can We Use Statistical Modeling for Quality Risk Assessment of Clinical Trials?. Ther Innov Regul Sci 54, 1227–1235 (2020). https://doi.org/10.1007/s43441-020-00147-x

2. Zou, M., Barmaz, Y., Preovolos, M. et al. Using Statistical Modeling for Enhanced and Flexible Pharmacovigilance Audit Risk Assessment and Planning. Ther Innov Regul Sci 55, 190–196 (2021). https://doi.org/10.1007/s43441-020-00205-4

3. Yeo, IK., Johnson, RA. A new family of power transformations to improve normality or symmetry. Biometrika, 87(4):95 (2000). https://doi.org/10.1093/biomet/87.4.954

4. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. 1st ed. New York: Springer (2013). https://doi.org/10.1007/978-1-4614-7138-7

5. Bergmeir, C., Benítez, JM. On the use of cross-validation for time series predictor evaluation. Information Sciences 191, 192-213. (2012) https://doi.org/10.1016/j.ins.2011.12.028

6. Rodríguez-Vignoli, J., Rowe, F. How is internal migration reshaping metropolitan populations in Latin America? A new method and new evidence, Population Studies, 72:2, 253-273 (2018). https://doi.org/10.1080/00324728.2017.1416155

7. Kröhnert, S., Vollmer, S. Gender-Specific Migration from Eastern to Western Germany: Where Have All the Young Women Gone?. International Migration, 50: 95-112 (2012). https://doi.org/10.1111/j.1468-2435.2012.00750.x

8. Bleiberg H, Decoster G, de Gramont A, et al. A need to simplify informed consent documents in cancer clinical trials. A position paper of the ARCAD Group. Ann Oncol. 28(5):922-930 (2018). https://doi.org/10.1093/annonc/mdx050

9. Ménard, T., Bowling R., Mehta P., et al. Leveraging analytics to assure quality during the Covid-19 pandemic - The COVACTA clinical study example. Contemporary Clinical Trials Communications 20, 100662 (2020). https://doi.org/10.1016/j.conctc.2020.100662

10. Ménard, T. Letter to the Editor: New Approaches to Regulatory Innovation Emerging During the Crucible of COVID-19. Ther Innov Regul Sci 55, 631–632 (2021). https://doi.org/10.1007/s43441-021-00281-0

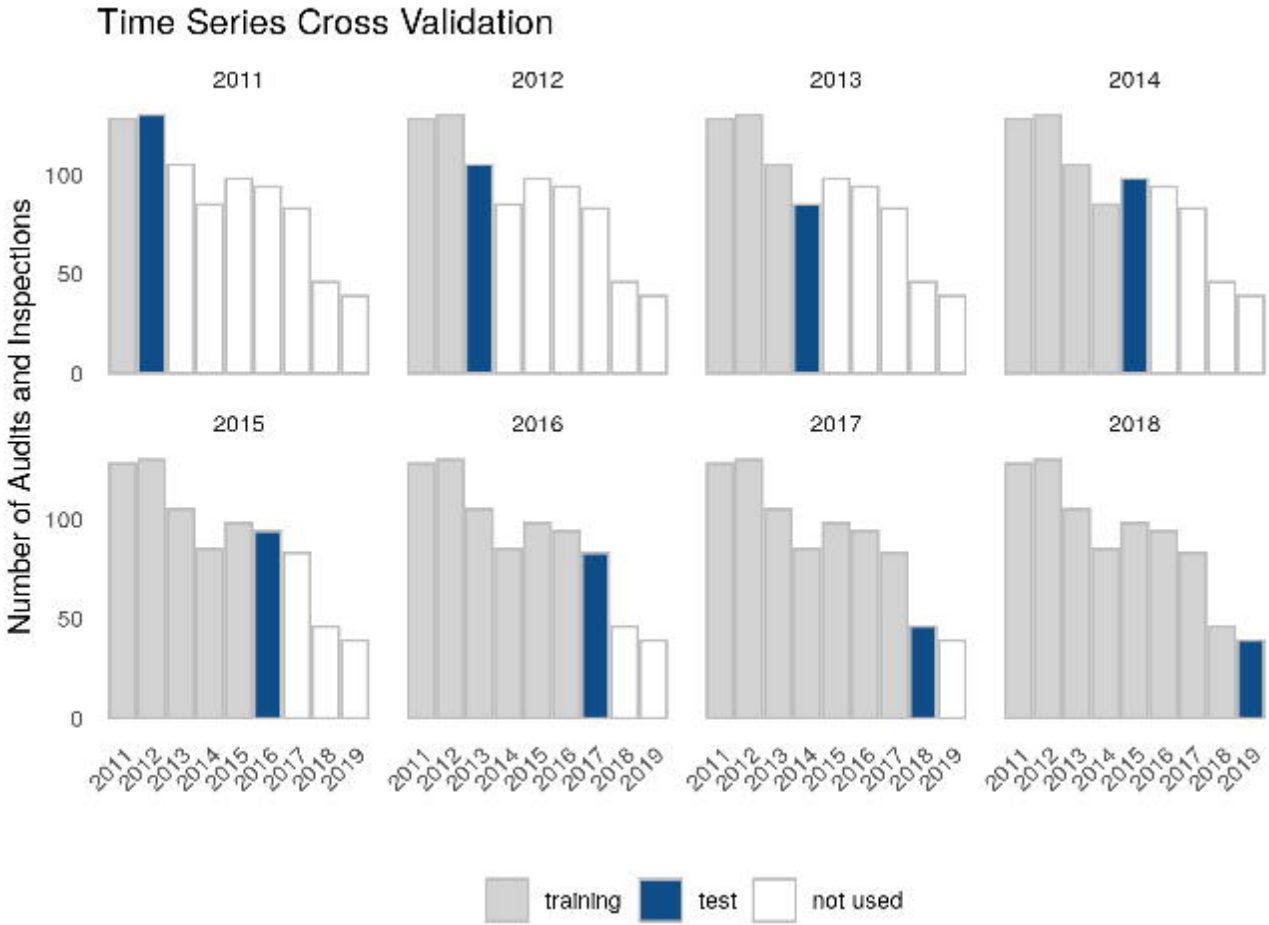
11. Ménard, T., Barmaz, Y., Koneswarakantha, B. et al. Enabling Data-Driven Clinical Quality Assurance: Predicting Adverse Event Reporting in Clinical Trials Using Machine Learning. Drug Saf 42, 1045–1053 (2019). https://doi.org/10.1007/s40264-019-00831-4

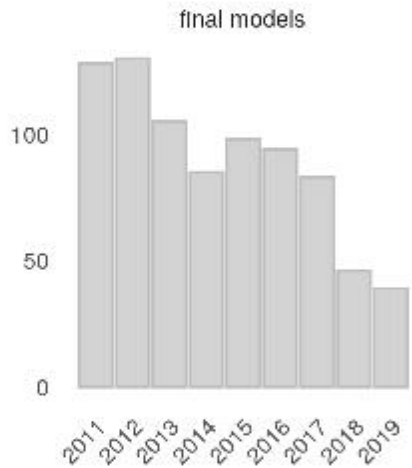
12. Ménard, T., Koneswarakantha, B., Rolo, D. et al. Follow-Up on the Use of Machine Learning in Clinical Quality Assurance: Can We Detect Adverse Event Under-Reporting in Oncology Trials?. Drug Saf 43, 295–296 (2020). https://doi.org/10.1007/s40264-019-00894-3

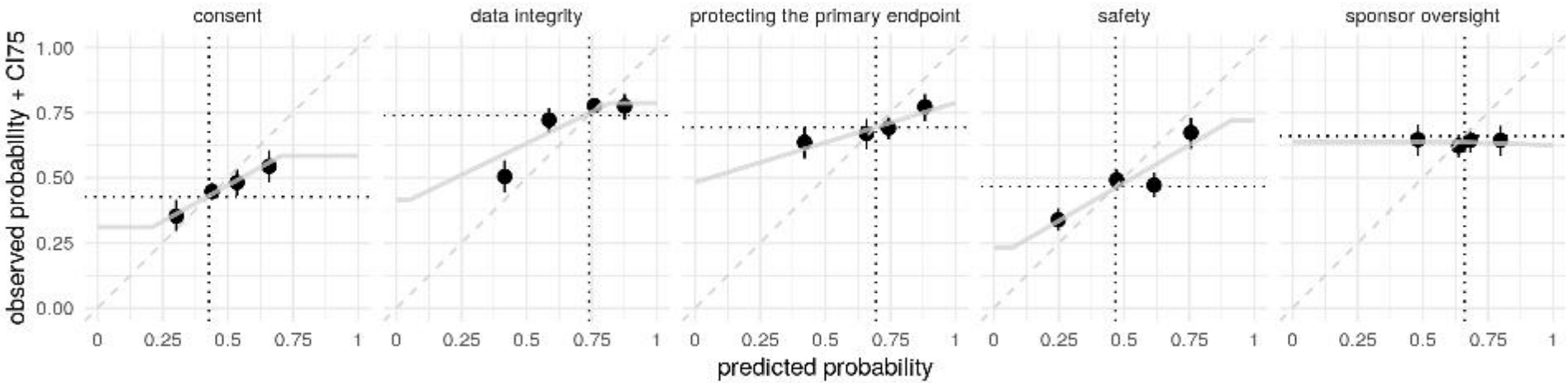
13. Koneswarakantha, B., Barmaz, Y., Ménard, T. et al. Follow-up on the Use of Advanced Analytics for Clinical Quality Assurance: Bootstrap Resampling to Enhance Detection of Adverse Event Under-Reporting. Drug Saf 44, 121–123 (2021). https://doi.org/10.1007/s40264-020-01011-5

14. Barmaz Y., Ménard, T. Bayesian Modeling for the Detection of Adverse Events Underreporting in Clinical Trials. Drug Saf (2021). https://doi.org/10.1007/s40264-021-01094-8

15. Ménard, T., Young, K., Emerson, J., et al. Cross-company collaboration to leverage analytics for clinical quality and accelerate drug development - the IMPALA industry group. CPT Pharmacometrics Syst Pharmacol., 00:1–5 (2021). https://doi.org/10.1002/psp4.12677


Time Series Cross Validation
2011
2012
2013
2014
2015
2016
2017
2018
Number of Audits and Inspections
0
50
100
2011 2012 2013 2014 2015 2016 2017 2018 2019
training
test
not used


final models
0
50
100
2011 2012 2013 2014 2015 2016 2017 2018 2019


consent
data integrity
protecting the primary endpoint
safety
sponsor oversight
observed probability + CI75
1.00
0.75
0.50
0.25
0.00
0
0.25
0.5
0.75
1
predicted probability

Consent Related Audit and Inspection Findings

Adjusted Risk Factors

	Distribution	N[1]	N with Findings[1]	Coefficient (CI95)[2]	Odds Ratio (CI95)	P	Coefficient Over Time (CI95)[2,3]
# Issues Late							
[0 - 4)		172 (21.2%)	60 (34.9%)		0.6 (0.4-0.9)	*0.0189 **	
[4 - 8)		85 (10.5%)	33 (38.8%)		0.7 (0.4-1.2)	*0.21467*	
[8 - 14)		60 (7.4%)	28 (46.7%)	reference			
[14 - 30)		71 (8.8%)	35 (49.3%)		1.2 (0.7-2.1)	*0.4762*	
[30 - 146)		46 (5.7%)	27 (58.7%)		1.7 (0.9-3.4)	*0.09289*	
missing		376 (46.4%)	188 (50%)	reference			
# Minor Deviations Per Days on Study[4]							
[0 - 0.0029)		282 (34.8%)	111 (39.4%)		0.7 (0.5-0.9)	*0.02002 **	
[0.0029 - 0.082)		146 (18%)	68 (46.6%)	reference			
missing		382 (47.1%)	192 (50.3%)	reference			
Pediatric Study							
Not Pediatric Study		756 (93.2%)	335 (44.3%)	reference			
Pediatric Study		55 (6.8%)	36 (65.5%)		2.3 (1.3-4.3)	*0.00478 ***	
missing		0 (0%)	0 (0%)	reference			
Perc Visits Not Occured (Percent Rank)[5]							
[0 - 0.5)		157 (19.4%)	59 (37.6%)		0.7 (0.5-1)	*0.04115 **	
[0.5 - 0.85)		131 (16.2%)	59 (45%)	reference			
[0.85 - 1)		66 (8.1%)	33 (50%)		1.2 (0.7-2.1)	*0.43918*	
missing		456 (56.2%)	220 (48.2%)	reference			
Total Population[6]							
[8.16 thous. - 8.3 mil.)		421 (51.9%)	175 (41.6%)	reference			
[8.3 mil. - 33.53 mil.)		190 (23.4%)	107 (56.3%)		1.9 (1.3-2.6)	*0.00031 ****	
missing		199 (24.5%)	88 (44.2%)	reference			

[1] Number of Clinical Trial Site Audits and Inspections

[2] Red Reference Line Indicates Zero

[3] Coefficient Value as New Data is Successively Added from 2011-2019

[4] Sum of Days Since Enrollment of All Patients

[5] Percent Rank of Site Within Study

[6] Population Within 100 km Radius of Site

Safety Related Audit and Inspection Findings

Adjusted Risk Factors

	Distribution	N[1]	N with Findings[1]	Coefficient (CI95)[2]	Odds Ratio (CI95)	P	Coefficient Over Time (CI95)[2,3]
% Population with Age over 60[4]							
[0.043 - 0.12)		48 (5.9%)	18 (37.5%)		0.7 (0.4-1.3)	*0.2418*	
[0.12 - 0.27)		485 (59.8%)	234 (48.2%)	reference			
[0.27 - 0.34)		36 (4.4%)	25 (69.4%)		2.1 (1-4.5)	*0.04244 **	
missing		241 (29.7%)	114 (47.3%)	reference			
AE Per Visit							
[0 - 0.52)		326 (40.2%)	144 (44.2%)	reference			
[0.52 - 5.64)		80 (9.9%)	49 (61.3%)		1.8 (1.1-3)	*0.01788 **	
missing		404 (49.8%)	197 (48.8%)	reference			
SAE Per Visit							
[0 - 0.0075)		171 (21.1%)	62 (36.3%)		0.6 (0.4-0.9)	*0.0048 ***	
[0.0075 - 0.67)		235 (29%)	131 (55.7%)	reference			
missing		404 (49.8%)	197 (48.8%)	reference			
Therapeutic Area							
Autoimmune Disease Study		109 (13.4%)	41 (37.6%)		0.9 (0.5-1.6)	*0.76089*	
Cancer Study		460 (56.7%)	262 (57%)		1.7 (1.2-2.6)	*0.00684 **	
Neuro- Psychiatric Study		112 (13.8%)	36 (32.1%)		0.6 (0.4-1.1)	*0.1173*	
other Therapeutic Area Study		130 (16%)	52 (40%)	reference			
missing		0 (0%)	0 (0%)	reference			

[1] Number of Clinical Trial Site Audits and Inspections

[2] Red Reference Line Indicates Zero

[3] Coefficient Value as New Data is Successively Added from 2011-2019

[4] Population Within 100 km Radius of Site

Data Integrity Related Audit and Inspection Findings

Adjusted Risk Factors

	Distribution	N[1]	N with Findings[1]	Coefficient (CI95)[2]	Odds Ratio (CI95)	P	Coefficient Over Time (CI95)[2,3]
# Active Trials At Site In TA Past Year (Percent Rank)[4,5]							
[0 - 0.41)		300 (39%)	200 (66.7%)		0.6 (0.5-0.9)	*0.00612 **	
[0.41 - 1)		352 (45.7%)	275 (78.1%)	reference			
missing		101 (13.1%)	74 (73.3%)	reference			
# Issues Due							
[0 - 2)		288 (37.4%)	189 (65.6%)		0.6 (0.4-0.8)	*0.00102 ***	
[2 - 125)		146 (19%)	108 (74%)	reference			
missing		335 (43.5%)	262 (78.2%)	reference			
# Open Queries Per Days On Study[6]							
[0 - 0.0056)		178 (23.1%)	109 (61.2%)		0.5 (0.4-0.8)	*0.00117 ***	
[0.0056 - 0.87)		151 (19.6%)	116 (76.8%)	reference			
missing		440 (57.1%)	334 (75.9%)	reference			
Male to Female Ratio in Population with Age 18-39[7]							
[0.46 - 0.49)		27 (3.5%)	17 (63%)		0.7 (0.3-1.7)	*0.45269*	
[0.49 - 0.5)		170 (22.1%)	116 (68.2%)		0.8 (0.6-1.2)	*0.37012*	
[0.5 - 0.51)		237 (30.8%)	177 (74.7%)	reference			
[0.51 - 0.52)		143 (18.6%)	118 (82.5%)		1.8 (1.1-3)	*0.01416 **	
[0.52 - 0.55)		34 (4.4%)	28 (82.4%)		2.1 (0.9-5.7)	*0.1189*	
missing		158 (20.5%)	103 (65.2%)	reference			

[1] Number of Clinical Trial Site Audits and Inspections

[2] Red Reference Line Indicates Zero

[3] Coefficient Value as New Data is Successively Added from 2011-2019

[4] Percent Rank of Site Within Study

[5] Active Trials in the Past Year in the Same Theraputic Area

[6] Sum of Days Since Enrollment of All Patients

[7] Population Within 100 km Radius of Site

Protecting The Primary Endpoint Related Audit and Inspection Findings

Adjusted Risk Factors

	Distribution	N[1]	N with Findings[1]	Coefficient (CI95)[2]	Odds Ratio (CI95)	P	Coefficient Over Time (CI95)[2,3]
# Issues Late							
[0 - 4)		172 (27.6%)	116 (67.4%)		0.8 (0.6-1.2)	*0.32242*	
[4 - 30)		216 (34.7%)	152 (70.4%)	reference			
[30 - 146)		46 (7.4%)	37 (80.4%)		1.4 (0.7-3.1)	*0.35444*	
missing		188 (30.2%)	133 (70.7%)	reference			
# Minor Deviations Per Days on Study[4]							
[0 - 0.011)		356 (57.1%)	248 (69.7%)	reference			
[0.011 - 0.082)		72 (11.6%)	57 (79.2%)		1.8 (1-3.5)	*0.04892 **	
missing		194 (31.1%)	132 (68%)	reference			
Median AE Reporting Delay (days)							
[0 - 27)		206 (33.1%)	131 (63.6%)		0.6 (0.4-0.8)	*0.00259 ***	
[27 - 1.33 thous.)		178 (28.6%)	135 (75.8%)	reference			
missing		238 (38.2%)	171 (71.8%)	reference			
Median Query Processing Time (days)							
[0 - 2)		202 (32.4%)	138 (68.3%)	reference			
[2 - 5)		48 (7.7%)	37 (77.1%)		1.5 (0.8-3.2)	*0.26426*	
[5 - 14)		16 (2.6%)	13 (81.2%)		1.9 (0.6-8.2)	*0.33011*	
missing		356 (57.1%)	249 (69.9%)	reference			

[1] Number of Clinical Trial Site Audits and Inspections

[2] Red Reference Line Indicates Zero

[3] Coefficient Value as New Data is Successively Added from 2011-2019

[4] Sum of Days Since Enrollment of All Patients

Male to Female Ratio in Population with Age 18-39

North America

Europe

World