PhUSE US Connect 2019

Paper AR09

Statistical Models of Risk Flux Reveal Dynamics of RBM-Managed Clinical Trials

Ryan P. McCaffrey, Michael A. Farnum, Aaron J. Mackey PhD, Covance LLC, Princeton, NJ, USA

ABSTRACT

The temporal dynamics of accumulating and mitigated risk, as measured in trials undergoing RBM management, have not been well described. Using 78,000+ historical total risk observations, made serially across 30+ studies, 50+ countries, and 4000+ sites, we have developed a novel statistical framework for inference of parameters that closely model observed risk flux. We have used this framework to a) provide insight into many operational aspects of RBM management, including study-, country-, and time-specific effects on risk flux; b) identify underlying metrics that drive (and/or respond) to changes in site monitoring effort; and c) allow opportunities for predictive risk forecasting. Here we describe the statistical framework for risk flux modeling and risk flow simulation, and summarize our results on operational aspects that significantly affect the dynamics of risk flux. We will also demonstrate how such models may be used to forecast the cost/benefit effects of competing risk-adaptive monitoring plans.

INTRODUCTION

Clinical trial monitoring is critical to protecting patient safety, improving data quality, and ultimately, reducing the overall operational expense of clinical trial management. Organizations such as the US Food and Drug Administration (FDA), TransCelerate, and the International Committee for Harmonization (ICH) have recently advocated for the application of risk-based monitoring (RBM) to clinical trials, which prioritizes resources based on risk and need as opposed to traditional monitoring approaches that distribute resources equally.1–4 Tools used to implement RBM on clinical trials are relatively new and tend to vary widely. As a result, no RBM gold standard approach currently exists.5,6 Covance implements RBM and centrally manages ongoing trials through the use of its Xcellerate Monitoring platform, which allows for the comprehensive review, analysis and visualization of clinical data at various levels (e.g., study, site or subject level), detection of statistical outliers and/or unusual patterns, reallocation of resources, as well as the mitigation of active risks. The platform's hierarchical approach to aggregating operationally-derived risk metrics and indicators into a single, normalized and discretized risk score, ranging from 1 to 4 (where 4 represents the highest risk), has been previously reported.7 Examples of risks that are tracked by the platform include subject screening rates, subject enrollment rates, adverse event reporting rates, protocol deviation rates, site visit scheduling delays, number of patient dropouts, and number of data quality issues.

Despite the industrywide shift from conventional clinical trial monitoring to more risk-based approaches, we find minimal evidence in the literature to suggest risk mitigation is improved through the use of RBM. Previous research has demonstrated some measured success in using RBM to improve clinical trial data quality issues 8,9, but the effectiveness of RBM at mitigating or correcting other operational risk factors (such as habitual protocol deviations, monitoring plan noncompliance, or issue aging) has remained elusive. Here we perform a purely observational study to understand whether or not the structure of risk transitions observed in studies overseen by RBM are materially different than risk transitions associated with various stochastic models. More specifically, this manuscript seeks to quantitatively identify evidence for the efficacy of RBM, by comparing the upward, downward, and static risk score flux from 31 clinical trials managed using Xcellerate Monitoring to random risk flux expectations derived from five candidate random Markov chain models representing different stochastic structure, with parameters derived from fits to the real data. In particular, we compare observed, sequential clinical trial risk score data to simulated risk score data derived from four constrained Markov chain models with 1) asymmetrically constrained transitions (where upward or downward transitions of the same magnitude have the same transition probability), 2) symmetrically constrained transitions (where upward and downward transitions of the same magnitude have the same transition

probability), 3) independently constrained transitions with four different fixed transition probabilities (representing a "regression to the mean" condition), and 4) independently constrained transitions with one fixed transition probability (representing a univariate condition where all end states are equally probable).

METHODS

Data and variable definitions

Clinical trial risk data from 31 studies (both ongoing and completed) from 14 different sponsors carried out across more than 50 different countries was obtained from proprietary relational databases. The pre-filtered data, which spans more than four years, represents all available operationally-derived risk data overseen by Covance. The data represents sequential trial site risk scores (ordinal 1-4, with 4 representing highest risk), calculated at periodic (approximately monthly) assessments, amounting to 78,961 risk observations. A breakdown of number of observations for each data filtering step can be found in Figure 1. Each study in the dataset contained information about its associated trial sites and each trial site contained both historical and current risk score information. To eliminate potential biases resulting from trials that have recently opened, or were smaller in size, trials were filtered such that they contained six or more assessments and 50 or more patients. In addition to Risk Score, associated with each unique study-site-assessment identifier were several other categorical features such as Previous Risk Score, Study, Therapeutic Area, RBM Lead, Phase, Site In USA, Region, Second Half Site Time, Second Half Study Time, Study Site Size, and Study Enrollment Size. A description of these variables can be found in Table 1.

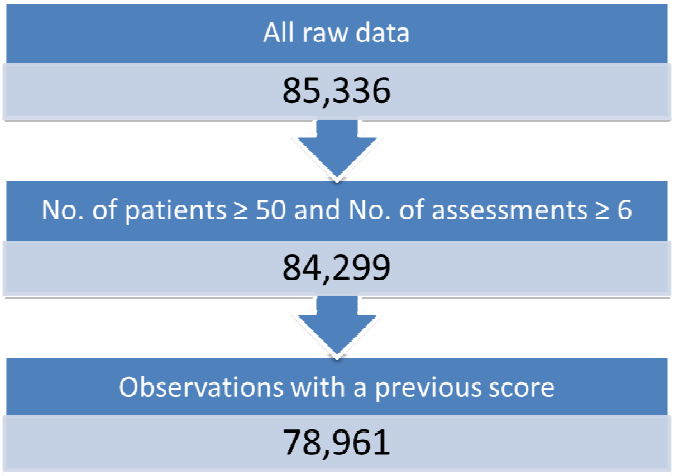

Figure 1 Data filtering steps and resulting number of risk observations

Table 1 Description of clinical trial covariates used in statistical modeling

Covariate	Description
Risk Score	Current risk score. Allowable values were 1, 2, 3, or 4. Higher values represent higher risk. This covariate was the dependent variable.
Previous Risk Score	Risk score from the previous assessment. Allowable values were 1, 2, 3, or 4. Higher values represent higher risk.
Study	Unique numeric identifier corresponding to a clinical trial study.
Therapeutic Area	Categorical variable representing the therapeutic area of a study. Allowable values were 'Endocrinology', 'Oncology' and 'Other'. Less common therapeutic areas were combined into an 'Other' category.
RBM Lead	Categorical variable representing the individual overseeing risk-based monitoring aspects of a study. Allowable values were 'G', 'I' and 'Other'. Less common RBM Leads were combined into an 'Other' category.
Phase	Categorical variable representing the study phase. Allowable values were 'II', 'III' and 'IV'.
Site In USA	Binary variable representing whether or not a particular site is located within the United States.
Region	Categorical variable representing the global region associated with a site. Allowable values were 'Americas', 'Asia Pacific' and ' Europe-Middle East-Africa '.
Second Half Site Time	Binary variable representing whether or not a given assessment took place within the second half of all assessments within a given site.
Second Half Study Time	Binary variable representing whether or not a given assessment took place within the second half of all assessments within a given study.
Study Site Size	Ternary variable of equal-sized buckets representing the number of sites associated with a particular study. Allowable values were 'Small' [35-85], 'Medium' [89-137] and 'Large' [144-643].
Study Enrollment Size	Ternary variable of equal-sized buckets representing the total number of patients enrolled in a particular study. Allowable values were 'Small' [68-342], 'Medium' [348-723] and 'Large' [738-12070].

Each variable in Table 1 was incorporated as a covariate to statistical models used to model risk. For some covariates, less commonly observed labels were binned to reduce cardinality. The covariate term “RBM Lead” refers to a staff position tasked with centrally overseeing clinical trial risk management, which includes reviewing and finalizing system-generated recommendations, entering issues into a risk and issue management system, and prioritizing and communicating potential site interventions. In the context of our analysis, RBM Lead identifies the particular individual who is filling that role. Each study is assigned one RBM Lead. Two temporal features were generated to try to understand if Risk Score was influenced by trial maturity. Second Half Site Time was a binary feature used to identify whether or not a particular assessment occurred in the second half of all assessments associated with a particular site, whereas Second Half Study Time was a similar feature used to identify whether or not a particular assessment occurred in the second half of all assessments associated with a particular study. These two separate temporal features were employed in order to capture the staggered nature of trial sites entering and exiting studies at different times, recognizing that the studies under consideration varied widely in length and some of them were not completed at the time of our analysis. Descriptive information about each of the 31 studies is given in Table 2.

Table 2 Clinical trial metadata. This table outlines several attributes associated with each of 31 studies used in the analysis. With the exception of RBM Lead, which has been de-identified, the data in this table is represented in its raw form. Values from this table were transformed (e.g., binned, categorized, etc.) and mapped to study-specific covariates highlighted in Table 1. Indication is included in the table to give the reader additional context, but was not included as a covariate in our statistical modeling due to its high cardinality

Study No.	Therapeutic Area	Phase	Indication	RBM Lead	No. of Sites	No. of Subjects	No. of Assessments	Last Assessment
1	Endocrinology	III	Diabetes, Type I	E	78	793	22	Feb-17
18	Endocrinology	III	Diabetes, Type I	E	100	782	23	Jun-17
52	Oncology	III	Cancer, leukemia, chronic lymphocytic	J	196	432	38	May-18
104	Cardiovascular	III	Other	E	501	7,214	33	Feb-18
184	Cardiovascular	III	Hypercholesterolemia	A	626	12,070	10	Oct-15
210	Oncology	III	Cancer, prostate	B	48	408	12	May-16
269	Oncology	II	Cancer, breast	E	131	100	36	May-18
271	Oncology	II	Cancer, breast	E	144	283	34	May-18
283	Oncology	III	Cancer, breast	I	113	418	35	May-18
313	Cardiovascular	III	Other	E	643	6,567	33	Jun-18
333	Endocrinology	III	Diabetes, Type I	E	137	1,406	18	Apr-17
360	Endocrinology	III	Diabetes, Type II	D	176	1,701	13	Jan-17
380	Infectious Disease	II/III	Infection, skin and skin structure, complicated	H	77	613	13	Jul-17
384	Oncology	III	Cancer, prostate	C	318	1,288	14	May-18
454	Oncology	III	Cancer, leukemia, acute myelogenous	F	236	308	15	May-18
464	Urology	III	Incontinence	E	107	202	6	Jun-18
465	Urology	III	Incontinence	E	93	207	6	Jun-18
520	Infectious Disease	III	Infection, pneumonia, community-acquired	G	157	738	11	Jan-18
524	Gastroenterology	III	Steatohepatitis	J	279	1,223	22	Jun-18
622	Oncology	III	Cancer, breast	J	67	68	12	Jun-18
628	Pulmonary/Respiratory	IV	Hypertension	C	91	96	10	Jun-18
629	Ophthalmology	III	Glaucoma	I	38	348	10	Mar-18
642	Endocrinology	III	Diabetes, Type II	G	93	518	18	Jun-18
643	Endocrinology	III	Diabetes, Type II	G	61	262	18	Jun-18
654	Reproductive Health	III	Contraception	C	85	366	17	Jun-18
729	Endocrinology	III	Diabetes, Type II	G	92	507	15	May-18

846	Endocrinology	III	Diabetes, Type II	G	84	342	9	May-18
852	Endocrinology	III	Diabetes, Type II	G	122	723	7	Jun-18
853	Endocrinology	III	Diabetes, Type II	G	35	72	9	May-18
856	Endocrinology	III	Diabetes, Type II	G	84	298	9	May-18
864	Endocrinology	III	Diabetes, Type II	G	89	71	7	Jun-18

Markov chain modeling

We chose to model the progression of site risk scores as a stochastic, first-order Markov chain process. This sequential framework is a natural representation to model the changes to risk score that occur at each RBM assessment. To enable the specific design of particular conditional random-walk topologies, we used a Hidden Markov chain model formulation (via the Python pomegranate package 10) with the identity matrix as an a priori fixed emission table parameter (i.e. there are no true hidden states in these models, despite the use of HMM formulation). Four different first-order Markov chain models having constrained (or "tied") transition parameters and one completely unconstrained first-order Markov chain model were fit using the Baum-Welch variant of the Expectation Maximization (EM) algorithm to the complete multi-study dataset, as well as individual data subsets for each study. Convergence was achieved by verifying that the output (log probability scores and final transition matrices) of 10 independent model fits were identical up to 16 significant digits after having initialized all transition matrix row parameters with random multinomial samples from a parent Dirichlet(1, 1, 1, 1) distribution. Model convergence was terminated once a stopping threshold of 1E-9 was reached, as measured by the improvement ratio of the log probability scores associated with consecutive fits. 99.7% and 0.3% of sites had an observed initial risk score of 2 and 3, respectively, as sites are generally assumed to be risk-neutral at the outset of each trial. Constraining the nodes of the Markov chain model transition matrices was necessary to simulate different stochastic conditions. By comparing the unconstrained Markov chain model to its constrained counterparts, we seek to understand whether the observed risk transitions can be truly attributed to RBM or to simpler effects expected to be found in such random-walk stochastic models. To test this, we generated Asymmetrically Constrained, Symmetrically Constrained, Regression to the Mean (RTM), and Univariate Markov chain models, and compared the resulting transition matrices to tabulated transition frequency tables derived from the observed, sequential risk score data.

The selection of our stochastic models was deliberate and fundamental to representing specific transition effects. The Asymmetrically Constrained transition matrix (degree of freedom = 7) was used to represent a scenario where the probability of upward or downward risk score transitions of the same magnitude, or step size, and direction are fixed regardless of prior risk. For example, under this scenario, downward transitions of Risk Score from 4→3 and 3→2 are equally likely, and upward transitions from 1→3 and 2→4 are equally likely. The further Symmetrically Constrained transition matrix (dof = 4) was used to represent a scenario where the probability of both upward and downward risk score transitions of the same magnitude are equal. For example, under this scenario, downward transitions of Risk Score from 3→2 and upward transitions 2→3 are equally likely. An RTM transition matrix (dof = 4) was used to represent a scenario where

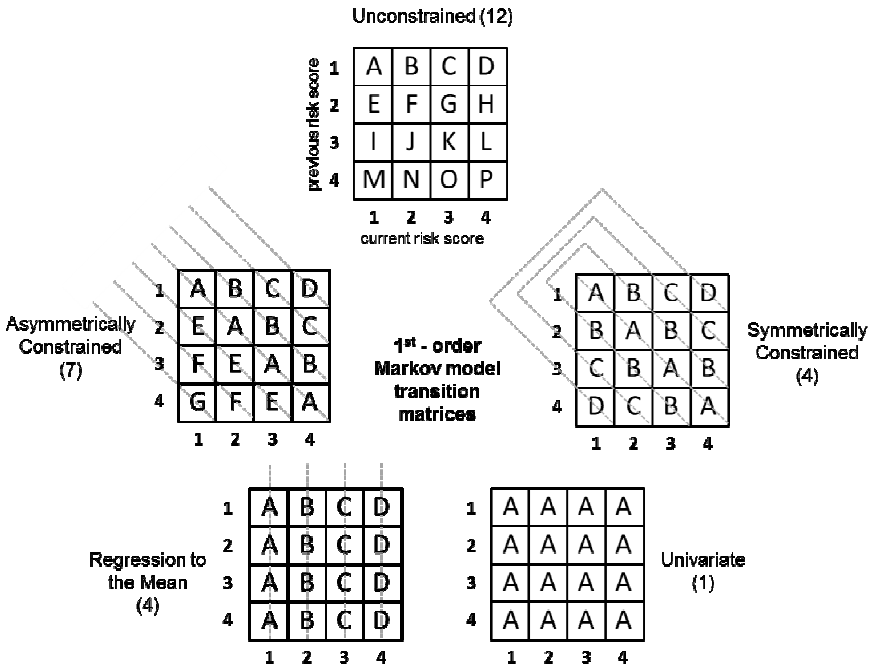

Figure 2 Illustration of first-order Markov chain model transition matrices. The degrees of freedom of each model are parenthetically notated. Each 4□4 Markov chain model transition matrix is represented by one or more letters that correspond to unique risk score transition probabilities. For example, in the Unconstrained transition matrix we see that each cell has a different letter, which indicates that the model contains no additional constraints and all 16 transition probabilities have the potential to be unique. In the case of the Univariate transition matrix we see that each cell has the same letter, which indicates that the model has been constrained such that each risk score transition probability is necessarily the same (i.e., 0.25). Each row of every transition matrix must sum to one

the probability of upward and downward risk score transitions are independent of starting risk score; under this scenario each of the four risk scores have four unique, fixed transition probabilities that do not depend on the starting risk score, equivalent to a zeroth-order Markov chain. The risk score closest to the mean represents the highest probability transition and, therefore, drives all scores toward the mean. A Univariate transition matrix (dof =1) was used to represent random risk score transitions. Under this scenario, each transition probability is equal and fixed, which results in a 0.25 probability for each transition pair. All constrained Markov chain models were compared to their unconstrained counterparts, which contained unique transition probabilities for each transition pair (dof = 12). A visual depiction of each of the Markov chain model transition matrices can be found in Figure 2.

Multi-model inference

Risk Score magnitude and flux was further modeled using an equivalent multinomial logistic regression framework with covariate inputs of Previous Risk Score, Study, Therapeutic Area, RBM Lead, Phase, Site In USA, Region, Second Half Site Time, Second Half Study Time, Study Site Size and Study Enrollment Size. To avoid biases inherent to stepwise multiple covariate testing, covariate importance was evaluated using a coherent multi-model inference approach, where multinomial logistic regression models are exhaustively fit to Previous Risk Score plus all other unique covariate combinations (using the glmulti package in R 11). The top 100 candidate models were ranked according to Akaike Information Criterion (AIC). AIC values were transformed into Akaike weights according to the following formula:

$$w_i(AIC) = \frac{\exp\left(-\frac{1}{2}\Delta_i(AIC)\right)}{\sum_{k=1}^{K}\exp\left(-\frac{1}{2}\Delta_i(AIC)\right)},$$

$$\text{where } \Delta_i(AIC) = AIC_i - \min(AIC)$$

These weights can be directly interpreted as conditional probabilities for each model.12 Covariate importance was calculated by summing the weights of all models containing each covariate. This approach to determining relative covariate importance was favored over other alternatives (such as forward or backward stepwise selection) because it both guarantees that the best model is identified and, more importantly, addresses model selection uncertainty by incorporating a multi-model inference instead of relying upon the parameters of a single model.11

Covariate label importance

Additional analysis was performed to better understand the importance of covariate label values. Risk score transition probability tables were computed for all observed combinations of levels found in statistically significant covariates (each combination of such yielding one subgroup of risk score observations). Jensen-Shannon divergence (JSD) was then calculated between the unconstrained risk score transition matrices for each subgroup and the unconstrained transition matrix of the complete multi-study dataset as a distance metric representing the total change in overall flux specific to that subgroup. This provided insight into how risk transitions associated with a particular set of covariate labels (subgroup) diverged from the overall observed data. The resulting JSD array contained 191 JSD values, one for each observed combination of important covariate labels. Spearman's

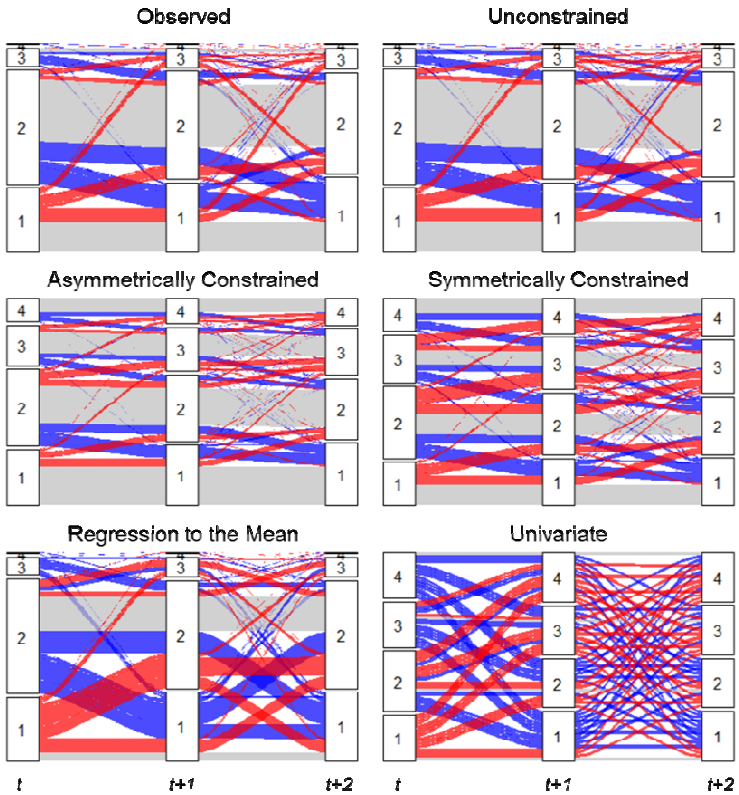

Figure 3 Alluvial diagrams of RBM risk score fluxes over three sequential assessment periods. The alluvial diagrams color code risk score fluxes to illustrate the evolution of Risk Score over three sequential assessment periods using both observed data obtained from clinical trials as well as data simulated from unconstrained and constrained Markov chain models that were previously fit to observed data. Blue represents downward risk fluxes, red represents upward risk fluxes, and grey represents no change in risk. The size of the flux bands is proportional to the frequency of their occurrence. The first alluvial diagram represents Risk Score transitions observed in real data, whereas the remaining five alluvial diagrams represent sampled data from unconstrained or constrained Markov chain models. The Unconstrained and RTM Markov chain models capture the general structure of the Observed risk fluxes, whereas the other models fail to capture this structure.

rank-order correlation was calculated between binary vectors associated with each covariate label value (present or absent) and the JSD array.

Covariate label importance was further explored by calculating the change in the difference between the probabilistic sums of downward and upward transitions (termed “net risk reduction”) for nested sets of important covariate subgroups. By calculating the net risk reduction from transition matrices associated with nested subgroups of important covariate labels we can attribute the change in net risk reduction to the inclusion of a particular covariate label, assuming the inclusion of that covariate label represents the only difference between the subgroup pairs.

RESULTS

Risk flux visualization

Unconstrained, Asymmetrically Constrained, Symmetrically Constrained, RTM, and Univariate Markov chain models were fit to the full multi-study dataset. Observed and model-simulated risk data was visualized using alluvial diagrams, as shown in Figure 3. The first alluvial diagram represents risk score fluxes from observed data, whereas the subsequent diagrams represent simulated data generated from different unconstrained/constrained Markov chain models previously fit to the observed data. Blue fluxes represent cases where risk improves (decreases) in any of the subsequent assessments, grey fluxes represent cases where risk levels remain the same across all assessments, and red fluxes represent cases where risk worsens (increases). A greater presence of blue than red in the alluvial diagrams indicates a greater ability to mitigate risk (i.e., induce downward risk flux). Visually, through these plots we find that the risk flux in the Asymmetrically Constrained, Symmetrically Constrained, and Univariate stochastic models do not recapitulate what we observe in the real data, whereas the Unconstrained and RTM Markov chain models do seem to capture the general structure of these risk transitions. Although we see general agreement in the observed risk transition fluxes between the observed and simulated data from the RTM model, the proportions (frequencies) associated with many of the risk transitions are not in agreement. For example, the proportions of risk scores that remain the same in the observed data (highlighted in grey) are grossly underrepresented by the RTM model. These qualitative visual observations are supported by a Chi-Square goodness of fit test that was applied to the observed risk data and corresponding risk flux associated with each of the Markov chain models over a moving window of three assessments. We report goodness of fit results of 7,044 for the Unconstrained, 8,162 for the Asymmetrically Constrained, 12,842 for the Symmetrically Constrained, 12,993 for the RTM, and 32,413 for the Univariate stochastic models.

Risk transition tables

4x4 transition parameter matrices of each of the probabilistic Markov chain models were visualized using heatmaps, which can be found in Figure 4. The Unconstrained Markov chain model shows that downward transitions become decreasingly likely as risk levels go down and, conversely, upward transitions become decreasingly likely as risk levels go up. We attribute this finding, at least in part, to expected floor/ceiling effects found at the low/high risk score boundaries. In the Unconstrained Markov chain model transition matrix, the sum of all downward transitions (i.e., the lower triangle sum of the transition matrix) is more than four times

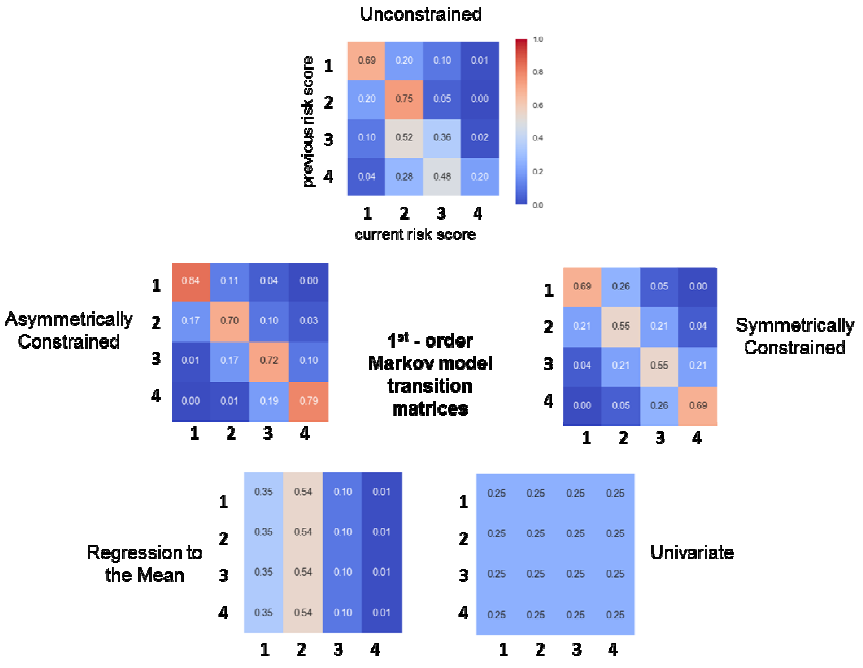

Figure 4 Heatmap visualizations of first-order Markov chain model transition matrices. The heatmap plots represent risk score transition probabilities (colored by magnitude) for various Markov chain models fit to observed, multi-study data. Heatmap annotations represent corresponding transition probabilities. Concentrations along the diagonals represent no change in risk, concentrations in the lower triangle represent downward changes in risk (risk improvement/mitigation), and concentrations in the upper triangle represent upward changes in risk (risk worsening/deterioration). The rows of each transition matrix must sum to one. This added constraint sometimes prevents linked transitions from having identical probabilities. For example, values along the main diagonal of the Asymmetrically Constrained transition matrix should be identical, but are not.

greater than the sum of all upward transitions (i.e., the upper triangle sum of the transition matrix). This indicates that RBM exhibits a strong risk reduction preference on Risk Score. For each of the Markov chain models, the net risk reduction was plotted against the probabilistic sum of achieving no change in risk (termed "residual static risk") and can be found in Figure 5(a). We note that although the RTM and Unconstrained Markov chain models exhibit the greatest net risk reduction, all models exhibit a non-negative net risk reduction, which highlights an overall preferential tendency toward risk improvement. The RTM Markov chain model explains much of the net risk reduction observed in the Unconstrained Markov chain model, however, it fails to explain the observed residual static risk. A corresponding analysis of net risk reduction in the observed data was performed on covariate subgroups (refined to include only the most important covariates) and can be viewed in Figure 6.

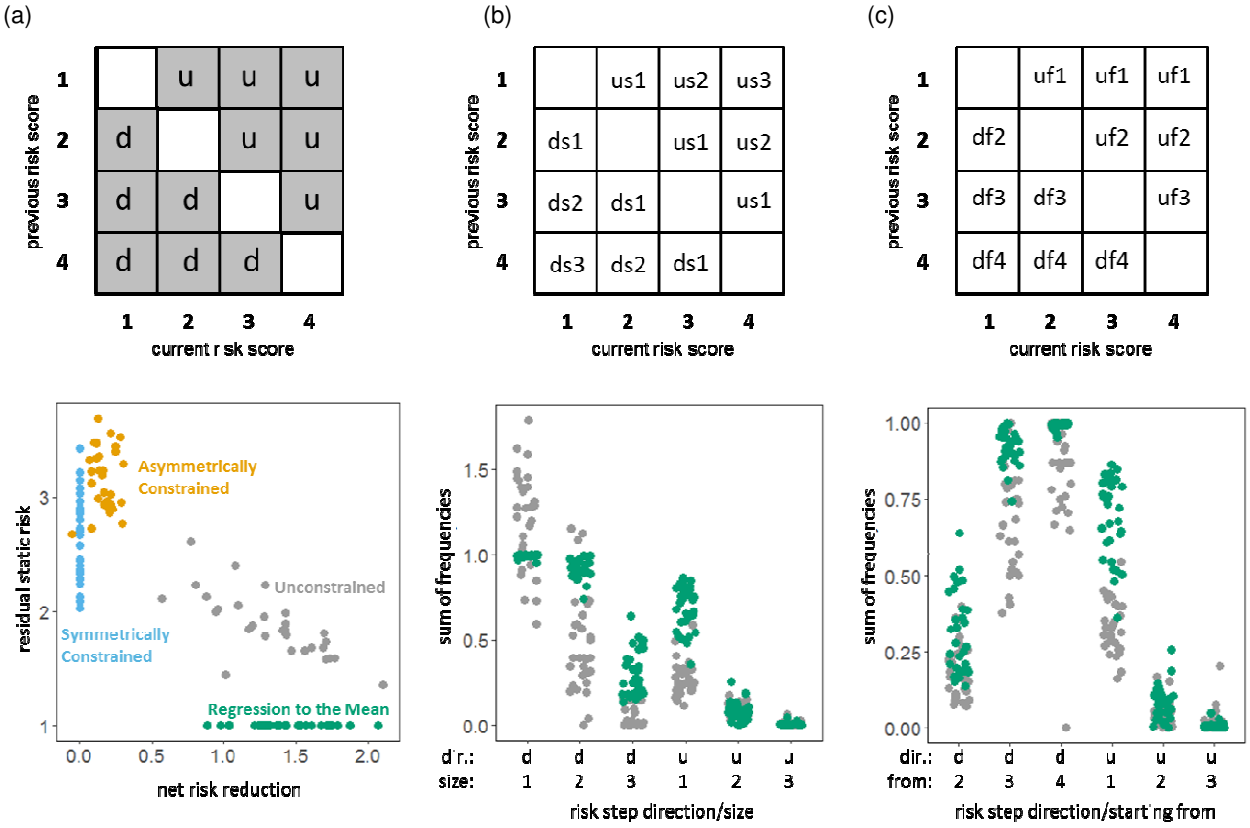

Figure 5 (a) Schematic of possible risk transitions (top) and study-specific net risk reduction, colored by Markov chain model type (bottom): net risk reduction (x-axis) is defined as the difference between the sums of the lower (decreasing risk transition, "d") and upper (increasing risk transition, "u") triangles of the transition frequency matrix. A positive net risk reduction represents a greater tendency to mitigate risk. Residual static risk (y-axis) is defined as the sum of frequencies with no change in risk. (b) Schematic of risk transition steps of varying direction and magnitude (top) and frequencies of such steps (bottom); "ds1" represents a downward transition of size 1, "us1" represents an upward transition of size 1, etc. (c) Schematic of risk transition steps of varying direction and starting risk score (top) and frequencies of such steps (bottom); "df2" represents a downward transition from a starting risk score of 2, "uf1" represents an upward transition from a starting risk score of 1, etc.

Risk transition steps were further compared between the Unconstrained and RTM models, both from a perspective of transition step magnitude and starting risk score, as seen in Figure 5(b-c). The probabilistic sums of risk transition steps of various magnitudes and directions showed smaller transition steps were more likely, which was in line with our expectation. The RTM model overestimates the probabilities of achieving downward transitions of size 2, downward transitions of size 3, and upward transitions of size 1, as compared to the Unconstrained model. The mean probabilistic sums of downward transitions of size 1 are not apparently different between the RTM and Unconstrained models, however, the RTM model failed to capture the high variance observed in these transitions.

The probabilistic sums of risk transitions with varying direction and starting risk values were explored in an effort to understand how starting risk values impact risk transitions in the observed data and whether or not these effects were captured by the RTM model. Individually visualizing the risk transition steps of all 31 studies again highlights the influence of floor/ceiling boundary effects in the Unconstrained model (i.e., downward/upward risk transition probabilities decrease as Risk Score decreases/increases), which can be viewed in Figure 6(c). This is in line with our intuition about the role Risk Score boundary conditions play. In comparing the Unconstrained and RTM models

we find that the RTM model generally overestimates the probability of transitioning down when at a Risk Score of 3 and transitioning up when at a Risk Score of 1. We attribute this to the fact that the mean Risk Score across all studies is 1.8, so sites that have a Risk Score of 3 or 1 will generally see their risk levels driven downward or upward, respectively, toward the mean. The RTM model also overestimates the probability of transitioning down when at a Risk Score of 4 and, by extension, underestimates the probability of transitioning up when at a Risk Score of 3. Transition probabilities seem to be in reasonable agreement between the Unconstrained and RTM models at starting Risk Scores of 2.

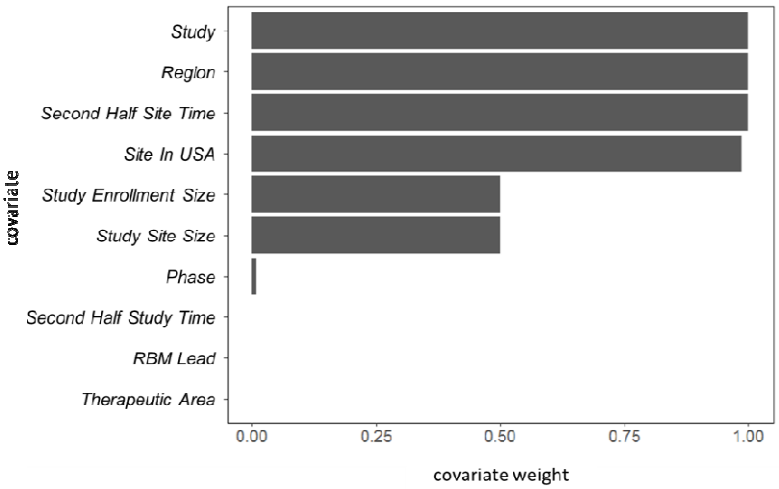

Figure 6 Ranking of weights of covariates used in multinomial logistic regression modeling of Risk Score. The covariate weights were determined using multi-model inference with a multinomial logistic regression fitting function. Previous Risk Score was omitted from bar chart as it was included in every candidate model and is, therefore, assumed to have a weight of 1. The bar chart displays the relative importance of the covariates used to model Risk Score, which includes main effects only. In addition to Previous Risk Score, the covariates with the highest importance were Study, Region, Second Half Site Time and Site In USA. Further modeling demonstrated all of these highly important features also exhibited significant interaction effects with Previous Risk Score.

Multi-model inference

Multi-model inference with a multinomial logistic regression fitting function was used to determine the relative importance of each covariate. The multinomial logistic regression modeling incorporated scores from the previous assessment in addition to other covariates. As shown in Figure 6, the covariates with the highest importance to Risk Score were Previous Risk Score, Study, Second Half Site Time, Region and Site In USA. Covariates were then subset based on importance and the most important covariates were incorporated into a second multi-model inference analysis that included covariate interactions with Previous Risk Score (analysis not shown). The analysis demonstrated that in addition to their main effects, each of the four most important covariates exhibit additional influence that is both significant and dependent on particular values of Previous Risk Score. These covariates were used in subsequent plots to analyze covariate subgroup populations. Figure 7 shows the differences in net risk reduction among each important covariate subgroup, colored by Study. Having found that these covariates were important we sought to understand the extent of their influence on net risk reduction. The magnitude of the influence Study, Second Half Site Time, Region and Site In USA had on net risk reduction was no greater than that of Study itself.

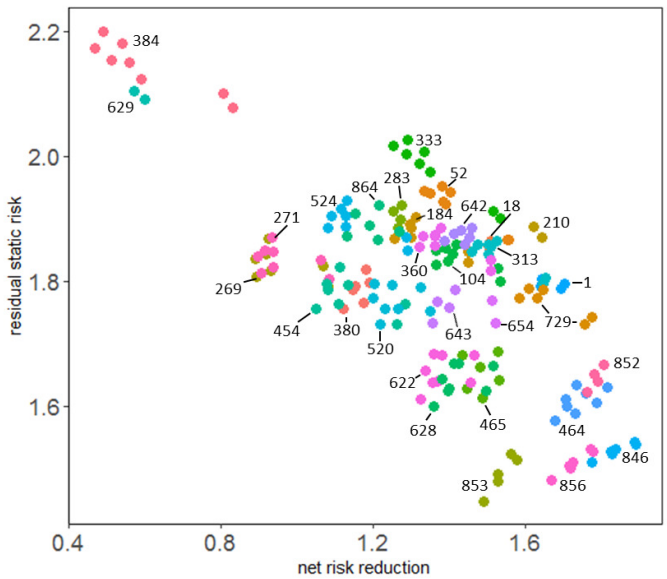

Figure 7 Net risk reduction plotted for each important covariate subgroup, colored by Study. Subgroups were generated from observed combinations of important covariate labels. Each point in the plot corresponds to the net risk reduction derived from a transition matrix associated with a particular subgroup. Each subgroup consists of one label value from each of the important covariates (i.e., Study, Region, Second Half Site Time, and Site In USA). A single study may contain multiple points on the plot because studies are contained within many different subgroups.

Covariate label importance

Calculating Spearman's rank-order correlation between JSD values and binary vectors representing the presence or absence of individual covariate labels provided insight into covariate label importance. High JSD values indicate greater dissimilarity, or higher divergence, between the subgroup transition matrix and the full multi-study transition matrix. The correlation analysis demonstrates that covariate label values associated with Study dominate the low and high ends of the rank-order correlation spectrum, which is highlighted in Figure 8.

Negative Spearman's rank-order correlations indicate that the presence of a particular covariate label decreases the divergence (and is closer to the marginal mean of the full model), which we interpret to mean that the characteristics of that covariate label are both important and similar to the overall covariate set. Conversely, positive Spearman's rank correlations indicate that the presence of a particular covariate label increases the divergence, which we interpret to mean that the characteristics of that covariate label are both important and dissimilar to the overall covariate set. We interpret Study to be an important covariate given that certain studies (namely 846, 464, 628, 856, etc.) exhibit large deviations from the marginal mean values of the full model, whereas other studies are more closely represented by the full model.

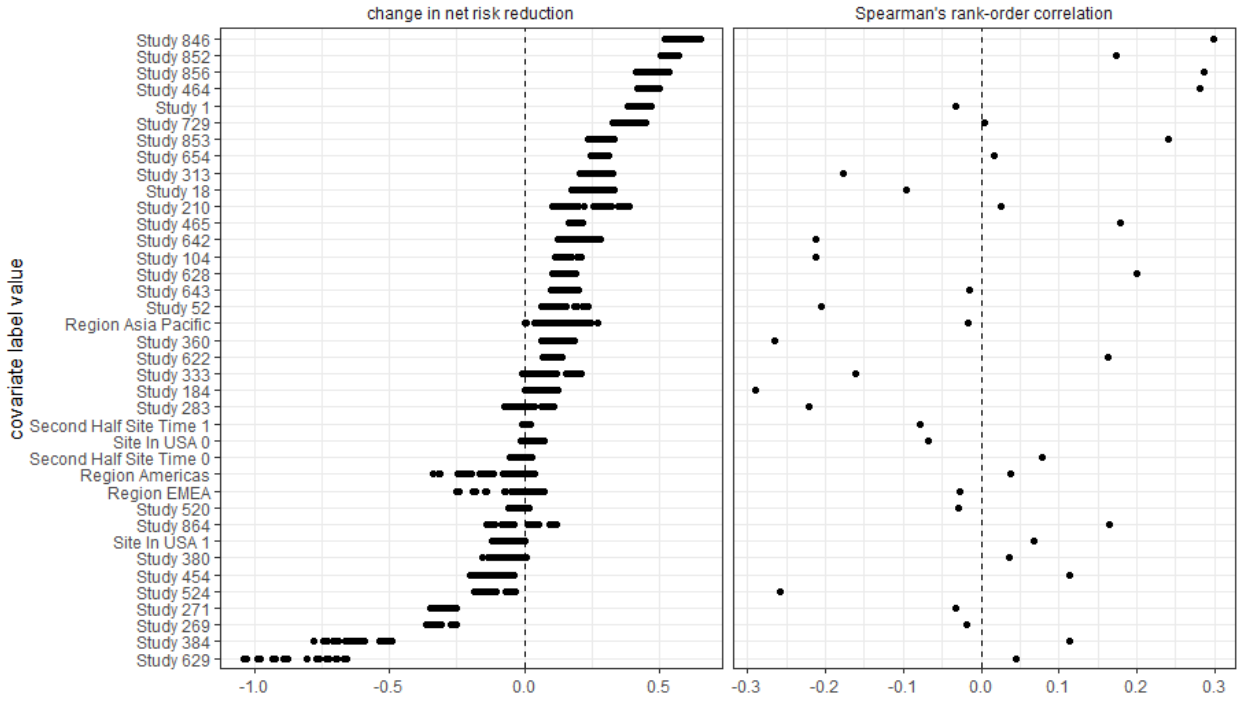

Figure 8 a) Change in net risk reduction plotted for each important covariate label. Multinomial logistic regression candidate models were generated for all possible combinations of important covariates. In order to assess the magnitude and directionality that a covariate label had on risk reduction, we calculate the change in net risk reduction between two transition matrices, where one is associated with a given set of covariate labels and another is associated with same set of covariate labels plus an additional covariate label of interest. b) Pairwise JSD calculations were performed between transition matrices of each important covariate subgroup and the full multi-study transition matrix. Spearman's rank-order correlation was performed between the resulting JSD vector and a binary vector representing the presence or absence of individual covariate labels. Higher correlations represent a greater covariate label importance.

In addition, by analyzing the distribution of net risk reduction changes that are attributed to specific covariate labels we gain insight into the effect size and influence each covariate label has on the ability to mitigate risk. A plot detailing the net risk reduction changes associated with each important covariate label can be found in Figure 8. Covariate labels associated with Study exhibit the largest source of variation, whereas the effect sizes of covariate labels associated with Site in USA, Second Half Site Time, and Region were less pronounced. Study size (by either Study Enrollment Size or Study Site Size) was found to be inversely correlated with net downward risk flux, though with great variation (data not shown).

DISCUSSION

A Chi-Square goodness of fit analysis of tabulated risk score transitions derived from observed clinical trial data and transition matrices resulting from the four Markov chain models fit to the observed risk data reveal that risk fluxes and mitigations in RBM-governed trials cannot alone be explained by any of the stochastic models. The Asymmetrically Constrained, Symmetrically Constrained and Univariate Markov chain models fail to appropriately represent the structure of the observed sequential risk transitions. The RTM model captures much of the net risk reduction observed in the actual data, but grossly underestimates the residual static risk. Additionally, a deeper exploration of risk transitions with respect to step size and starting risk value reveals that the RTM model does not agree with the observed data. Moreover, a preferential bias toward net risk reduction was observed in the risk data which suggests clinical trials governed by RBM are, on aggregate, more than four times more likely to mitigate than worsen risk. This finding is based on a comparison of the sums of the upward and downward risk transition frequencies derived from the tabulated multi-study risk transition frequency table. We conclude that the structure observed in risk score transitions as well as the net downward influence on risk score is non-stochastic and postulate that it is intrinsic to the RBM process.

Multi-model inference yielded Previous Risk Score, Study, Second Half Site Time, Region and Site In USA as the most important covariates to our Risk Score candidate models. Previous Risk Score was expected to be an important contributor to Risk Score as Previous Risk Score is auto-correlated with Risk Score (analysis not shown). Study was also expected to exhibit high importance to Risk Score since protocol complexity (as well as many of its derived operational challenges) is dependent upon the trial itself. Covariate labels associated with Study also exhibited the greatest impact (both negative and positive) on net risk reduction changes. In addition, the variation among the study-level transition matrices of the Unconstrained Markov chain model further supports the finding that Study is a highly important contributor to Risk Score (see Supplemental Figure 1). Furthermore, a closer examination of covariate label values using the aforementioned JSD-Spearman's rank order correlation analysis shows that highly important labels tend to be associated with Study, which suggests the largest source of variability is attributed to the study itself. Covariate label values with the highest Spearman's rank-order correlation also tend to have lower values of residual static risk, as observed in Figure 7. The importance of the remaining covariates – namely Second Half Site Time, Region and Site In USA – were not immediately obvious and highlight the dependence of Risk Score on the temporal and geographic aspects of trials.

Although Second Half Site Time exhibited high covariate importance, its impact on net risk reduction was minimal. Trial sites located outside the United States demonstrated a tendency to improve net risk reductions while sites located within the United States demonstrated a tendency to worsen net risk reduction. Moreover, sites located within the Asia Pacific region were found to be correlated with risk flux improvements, whereas sites located within the Americas and Europe-Middle East-Africa regions exhibited influences on net risk reduction that ranged from neutral to slight risk worsening. The net risk reduction findings related to sites within the United States and Americas region were consistent, which was expected given that sites located within the United States are a subset of those located within the Americas region. It should also be noted that RBM Lead was of little covariate importance, which demonstrates that the efficacy of the RBM methodology is not personnel dependent.

An ideal way to directly measure the influence of RBM on risk flux transitions and mitigations compared to traditional, non-risk-based clinical trial monitoring approaches would be through a blinded study, where RBM is applied to a fraction of sites within a single study or a fraction of studies within multiple trials, an idea that we are currently exploring with a number of sponsors. In the meantime, and in the absence of such a comparative study, we sought to compare observed risk scores with simulated data generated from stochastic Markov chain models to address the question of whether or not the observed risk fluxes were random. Because it is impossible to compare the observed risk fluxes with the span of all possible stochastic processes, we have limited our comparison to include five of the most probable. Therefore, this work cannot determine definitively that the observed risk fluxes of RBM-managed trials are non-stochastic. In addition, the list of risk-impacting covariates that we considered was non-exhaustive. There may be additional sources of variation – such as our inclusion of risk data from both on-going and completed studies; or the exclusion of study sponsor-related covariates (due to high cardinality and concern for possible re-identification) – that could potentially help explain the observed risk fluxes. Furthermore, this analysis does not address or measure the extent of potential cost-benefit improvements that RBM provides over traditional monitoring approaches. As future work, the authors intend to collect and study identically-measured risk data from trials

receiving traditional (non-RBM) monitoring approaches to compare directly to RBM. Such a comparative study has so far been elusive because traditional monitoring approaches typically use different risk metrics and indicators, which makes a direct comparison between the two monitoring methods challenging. Additionally, the authors also intend to explore the application of machine learning techniques to the forecasting of clinical trial risk scores one or more assessments into the future, taking into account both in-flight and historical site performance. 13

CONCLUSION

Tabulated transition frequencies derived from multi-study, RBM-managed observed risk score data demonstrate a net preference toward risk reduction. The observed risk flux could not be explained by any of the four representative first-order Markov chain models, which suggests that the risk mitigation achieved by RBM management is beyond what is expected from simple random variations in risk. Multi-model inference using a multinomial logistic regression framework revealed the most important covariates to risk were those associated with temporal and geographic aspects of trials, scores from the previous risk assessment, and the individual trial itself. A technique combining Jensen-Shannon divergence and Spearman's rank-order correlation found the covariate labels with highest importance were those related to individual trials.

REFERENCES

1. US Food and Drug Administration. Guidance for industry. Oversight of clinical investigations – A risk-based approach to monitoring. Available at: https://www.fda.gov/downloads/Drugs/guidanceComplianceRegulatoryInformation/Guidances/UCM269919.pdf. Accessed June 20, 2018.

2. Gill, D. Re-inventing clinical trials through TransCelerate. Nat. Rev. Drug Discov. 13, 787–788 (2014).

3. International Committee for Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH harmonized guideline. Integrated addendum to ICH E6(R1): Guideline for good clinical practice. E6(R2). Available at: https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R2__Addendum_Step2.pdf . Accessed June 20, 2018.

4. European Medicines Agency. Reflection paper on risk-based quality management in clinical trials. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/11/WC500155491.pdf. Accessed June 20, 2018.

5. Hurley, C. et al. Perceived barriers and facilitators to Risk Based Monitoring in academic-led clinical trials: a mixed methods study. 11 (2017).

6. Hurley, C. et al. Risk based monitoring (RBM) tools for clinical trials: A systematic review. Contemp. Clin. Trials 51, 15–27 (2016).

7. Agrafiotis, D. K. et al. Risk-Based Monitoring of Clinical Trials: An Integrative Approach. Clin. Ther. (2018). doi:10.1016/j.clinthera.2018.04.020

8. Diani, C. A., Rock, A. & Moll, P. An evaluation of the effectiveness of a risk-based monitoring approach implemented with clinical trials involving implantable cardiac medical devices. Clin. Trials Lond. Engl. 14, 575–583 (2017).

9. Tantsyura, V. et al. Extended Risk-Based Monitoring Model, On-Demand Query-Driven Source Data Verification, and Their Economic Impact on Clinical Trial Operations. Ther. Innov. Regul. Sci. 50, 115–122 (2016).

10. Schreiber, J. pomegranate: Fast and Flexible Probabilistic Modeling in Python. J. Mach. Learning Res.18. 1–6 (2018)

11. Calcagno, V. glmulti: An R Package for Easy Automated Model Selection with (Generalized) Linear Models. J. Stat. Softw. 29

12. Burnham, K. P. & Anderson, D. R. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. (Springer-Verlag, 2002).

13. Yang, E. et al. Quantifying and visualizing site performance in clinical trials. Contemp. Clin. Trials Commun. 9, 108–114 (2018).

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Michael Farnum
Covance, LLC
206 Carnegie Center
Princeton, NJ 08540
Email: Michael.Farnum@covance.com