2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)

Machine learning-based modeling of big clinical trials data for adverse outcome prediction: A case study of death events

Ling Tong, Jake Luo[§]
Health Informatics and Administration
University of Wisconsin Milwaukee
Milwaukee, WI, US
{ltong, jakeluo}@uwm.edu

Ron A. Cisler
College of Health and Human Services
Western Michigan University
Kalamazoo, MI, US
ron.cisler@wmich.edu

Michael Cantor
Regeneron Pharmaceuticals Inc
Tarrytown, NY, US
michael.cantor@regeneron.com

***Abstract*—It is known that clinical trials have potential risks for participants, which could result in unexpected adverse events. To quantify and predict the risk of adverse outcomes, we leverage a large amount of clinical reports to build machine learning models to predict adverse outcomes. We focused on death events as the predicting target in this study. From Clinicaltrial.gov, we collected 28,340 reports and transformed the data into vectorized machine learning features. These features were harmonized across studies using semantic mapping and feature selection techniques. The resulting selected clinical trial features were used to build five machine learning models for prediction. We evaluated and compared relative model performances for the prediction task. Results show that the logistic regression algorithm achieved the best overall receiver operating characteristic score at 0.7344. This exploratory study showed that it is feasible to use clinical trial factors to predict adverse outcomes. We demonstrated the approach by focusing on building machine learning models to predict the death outcomes. Predicting adverse outcomes could help clinical trials estimate harmful risks and design better mechanisms to protect participants. We hope by using our models, a clinical trial expert will be able to assess whether serious adverse events are likely to occur in a clinical trial at the early stage and to estimate what potential trial factors could contribute to the potential serious adverse events.**

***Index Terms*—Clinical Trials, Machine Learning, Prediction, Classification, Adverse Event, Big Data**

I. Introduction

A randomized clinical trial is still considered the gold standard for evaluating the safety and efficacy of medical interventions [1]. According to a report from the U.S. Food and Drug Administration (FDA) [2], clinical studies have become more globalized in the recent years. For the FDAs new studies of drugs alone, there were 131,749 participants from 70 countries involved in clinical trials during 2015 and 2016 [2]. Tens of thousands of people participate in clinical trials around the world each year. This participation is crucial for developing and evaluating new intervention methods for controlling diseases. However, given the experimental nature of clinical research, participants in clinical trials have a greater chance of experiencing harmful adverse events. Some of these adverse events are serious, life-threatening events that have a significant impact on the participants and the trials [3]. For example, in 2016, a new cancer treatment trial was temporarily suspended by the FDA for review because of severe side-effects that were the suspected cause of death for a number of participants [4]. Monitoring and reporting adverse events are required by clinical trial regulatory agencies such as the FDA and the European Medicines Agency.

Adverse-event monitoring is a major component of the risk-based monitoring of clinical trials [5]. To reduce harm and protect trial participants, almost all regulatory agencies have guidelines for risk-based monitoring of clinical trials [1]. However, these guidelines are usually described only in general terms, and clinical investigators implement them based on their own experience. In the past decade, a large amount of clinical trial data has been published [6]. We hypothesize that by leveraging such data we could build models to better quantify the adverse-event risks of clinical trials based on the trial and patient characteristics. Therefore, in this study, we explore the potential of harmonizing a large number of trial factors across different clinical trials to construct data-driven machine learning models for predicting adverse event outcomes. This study primarily focuses on the death event, which is the most severe adverse event in a clinical trial [1], [7].

Many factors could be associated with death events in clinical trials. For example, trials targeting acute diseases usually have a higher risk compared to those targeting chronic diseases. Trials targeting cancer conditions also have a higher risk than non-cancer studies. Similarly, the age group of participants and the phase of the clinical trial both are factors influencing the number of serious adverse events. This indicates that clinical trial factors could be used as predictors for outcome analysis. Therefore, one of our goals in this study is to normalize trial factors across a large number of clinical trials to enable the building of cross-trial machine learning models.

Using clinical factors to predict medical outcomes has been

[§]Corresponding author: Jake Luo, College of Health Sciences, University of Wisconsin Milwaukee (email: jakeluo@uwm.edu)

978-1-7281-2607-4/19/$31.00 ©2019 IEEE
DOI 10.1109/COMPSAC.2019.10218

proved extremely useful. Clinical factors have been used to successfully predict outcomes for conditions such as Ebola [8], breast cancer [9], heart disease [10], and drug toxicities [11]. Researchers have explored the prediction of mortality for patients who have undergone elective cardiac surgery [12], for patients with sepsis in emergency departments [13], and for hospital patients with pneumonia [14]. All studies published in this area have indicated that building machine learning models based on clinical factors is an effective means of measuring clinical risks and protecting patients. As with other data-driven applications, it should be stressed that data quality and data volume are important in medical outcomes prediction [15]. Randomized clinical trial data is of high quality to be used in data-driven modeling for accurate adverse event prediction.

The main contributions of this study are two: 1) The potential of using clinical trial data for adverse outcome prediction is explored. We develop methods to normalize data across clinical trials. Using the collected data, we evaluate machine-learning based models for outcome prediction. 2) This study complements existing clinical trial risk estimation methods by providing a quantitative approach for analyzing adverse event predictors. Currently, clinical investigators rely on their experience and on risk evaluation guidelines to assess the risks in trials. This study quantifies risk factors using an evidence-based, data-driven approach, which is a novel approach to estimate adverse trial outcomes. We test our hypothesis by focusing on predicting death events, which is the most severe event for a trial participant. To the best of our knowledge, this is the first significant effort in constructing machine learning-based models for adverse outcome prediction in clinical trials.

II. Methods

Overall, our methods consisted of three major stages, as shown in Fig. 1. First, we acquired clinical trial reports from clinicaltrials.gov. We converted the acquired trial reports to a vectorized representation according to the data types and data values. This vectorized representation enabled cross-trial comparison and machine learning-based model building. Second, we trained and evaluated the machine learning model for adverse outcome prediction using five different machine learning algorithms. We also created a baseline for performance analysis for the task of predicting death events. Finally, we explored three approaches to optimize the model performance, including Unified Medical Language System(UMLS)-based synonym mapping, correlation analysis, and feature selection. The third stage also included a comprehensive statistical analysis, which aimed to find potential correlations between trial factors and adverse outcomes.

A. Data Source

The study data was retrieved from clinicaltrials.gov [16]. Clincaltrials.gov collects clinical trial data from more than 180 countries and areas around the world and is the largest public clinical trial repository. Many clinical trial regulation agencies and clinical research journals require that trials publish their protocols and results on clinicaltrials.gov [6], [16], [17]. Therefore, the trial information documented on the repository is comprehensive and representative [17]. The study results are reported when a clinical trial is completed. The collected reports contain a large amount of rich information, such as the clinical trial phase, data on adverse events among participants, and demographic information. In this study, we acquired data for 255,065 clinical trials and filtered in 28,340 clinical trials that reported their final outcomes. Among these studies, a following analysis showed that 1,820 trials had at least 1 death event during the study period. These trials were used as positive cases, and other, non-death trials were used as negative cases for machine learning.

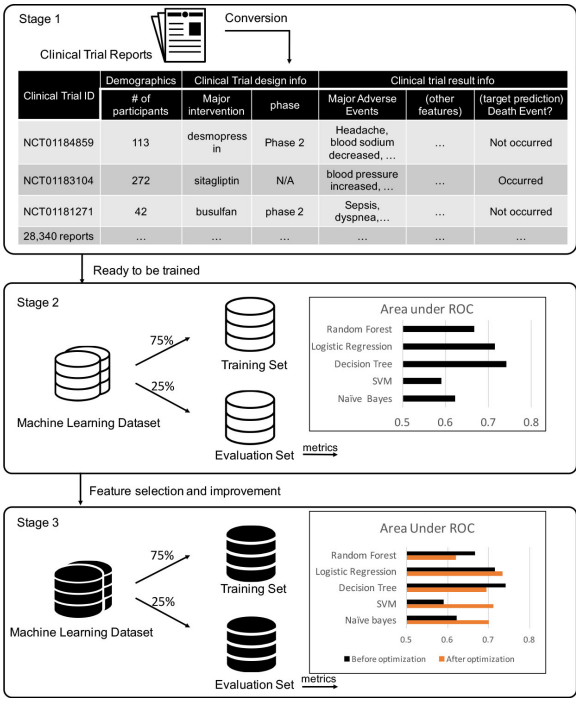
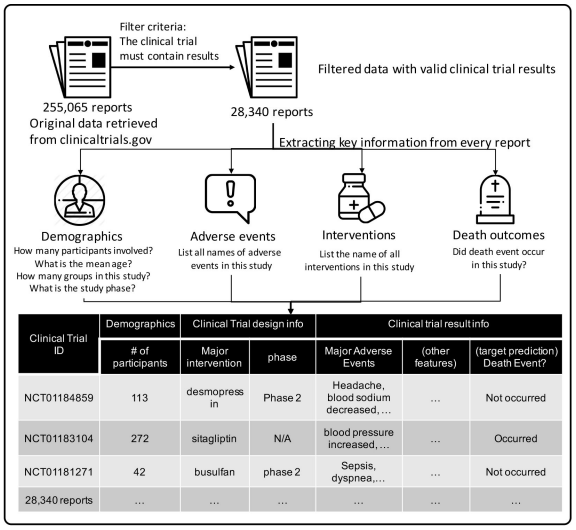


Fig. 1. Diagram of overall study framework. Stage 1: Converting all clinical trial reports into computer readable format. Stage 2: Training and evaluating five different machine learning algorithms. Stage 3: Conducting feature optimization analysis, including semantic normalization. ROC: receiver operating characteristics, SVM: support vector machine.

B. Transforming Trial Reports into a Vectorized Representation for Cross-Trial Modeling

The acquired clinical trial reports were not fully structured and had been originally designed for trial monitoring purposes. The information was not ready for computational modeling. Therefore, the first stage was to transform the collected reports into computable vectorized representations. This representation was a prerequisite for cross-trial analysis and machine learning-based modeling. The goal here was to extract the clinical trial factors from report documents. These extracted factors were categorized and vectorized to ensure consistent representation across trials. From the acquired 28,340 trials, we extracted 122,586 data elements. These data elements can be broadly grouped into seven categories: participant



Clinical Trial ID	Demographics	Clinical Trial design info		Clinical trial result info		
	# of participants	Major intervention	phase	Major Adverse Events	(other features)	(target prediction) Death Event?
NCT01184859	113	desmopressin	Phase 2	Headache, blood sodium decreased,…	…	Not occurred
NCT01183104	272	sitagliptin	N/A	blood pressure increased,…	…	Occurred
NCT01181271	42	busulfan	phase 2	Sepsis, dyspnea,…	…	Not occurred
28,340 reports	…	…	…	…	…	…

Fig. 2. Encoding the text report to matrix notation. First, we filter those clinical trials with valid reports of study results. Second, we extract key information for every report, including these four types: (1) Demographics information of patients and from the clinical trial design. (2) List of names of all adverse events, where every different adverse event has a separate dimension in the dataset. (3) List of names of all study interventions, where every intervention has a separate dimension in the dataset. (4) The outcome (death event) of the study.

information, trial phase, target condition, intervention method, serious event, non-serious event, and the death event. The death event was our prediction target. Other data elements were used as trial features. We hypothesized that the extracted trial features could be used to build machine learning models for predicting death events.

C. Predictive Modeling of Death Events

The vectorized transformation of clinical trial reports enables us to conduct cross-trial studies. The trial factors are represented as features, and the death outcomes are labeled for each trial. We hypothesized that the trial factors could have predictive power with respect to the death outcomes. To verify this hypothesis, we tested and evaluated five machine learning algorithms: logistic regression, support vector machine, nave bayes, random forest, and decision tree. A 10-fold cross validation was performed to estimate the performance of the developed models, and the precision, recall, and F1-scores were evaluated. The evaluation results were used as a baseline model and are reported in the Results section. After building the baseline model, we explored three feature-engineering methods to further improve the clinical trial factor representations.

D. Feature Engineering for Improving Machine Learning Model Performance

As shown in Figure 1, the extracted clinical trial features have a high dimension. This could impede the model building and machine learning optimization. Therefore, we explored three feature engineering methods to improve feature representation. The feature engineering could filter out low-importance or potential duplicate features to reduce data dimension. Reducing the feature dimension could make it easier to identify which clinical trial factors have a stronger association with death events. The three feature engineering methods evaluated included variance selection, semantic mapping, and correlation selection. Finally, to compare the effectiveness of the feature engineering methods, we compared the new models to the baseline model.

E. Optimizing the Model: Variance-Based Feature Selection

The first feature engineering technique applied was variance-based feature selection, which removed trial features that had a low variance score. Low-variance features usually make little contribution to machine learning model building. For example, if a feature had the same value across all data samples, it would make no contribution to model building. We used a minimum threshold of absolute variance value larger than 0.001. Using this threshold, we were able to preserve 99.5% of the overall variance values across all collected data. Features with a variance value less than 0.001 were removed, leaving a total of 16,379 features. We ranked these remaining features based on their variance score, and the results are shown in Figure 3. The machine-learning models were retrained and compared to the baseline models.

F. Semantic Mapping of Clinical Terminologies

Clinical trial reports are not fully standardized across trials. Investigators may use different terms for the same clinical meaning. For example, many phrases were used in the report to indicate a hypertension condition, such as ”hypertension,” ”arterial hypertension,” ”high blood pressure,” and ”blood pressure increased”. we used semantic mapping to address this problem. First, we used the Unified Medical Language System (UMLS) meta-thesaurus to identify clinical terms in the feature set and map them to their semantic concepts. If two or more terms mapped to the same UMLS concept identifier, the corresponding features were merged into one clinical trial feature. This allowed us to group all synonyms for a given clinical meaning and treat them as a single feature. The technique not only reduced the feature dimension but also enhanced the ability for cross-trial analysis by identifying comparable synonymous features across different trials.

G. Correlation-Based Feature Filtering

After merging synonymous features, we used a correlation-based method to measure the association between each remaining individual feature and the death events. Features with low correlation scores were removed. The Pearson coefficient was used to calculate the correlation scores. It is a fast and effective method to measure the correlation between trial feature and outcome variable. The resulting value lies in [-1, 1], with -1 indicating perfect negative correlation, and +1 indicating perfect positive correlation. We ranked these calculated correlation scores to identify trial features that had a strong positive association with death events.

III. Results

Table 1 summarizes the extracted trial features and data element examples. The first type of feature is participant demographics, which includes the number of participants in a trial and their median age. These features are naturally represented as numerical values. The second type of feature is the clinical trial phase. Typically, an intervention tested in human trials moves progressively from phase 1 trials to phase 4 trials, and each later phase involves more participants than the previous. Phases 1, 2, and 3 are pre-marketing trials, and phase 4 represents post-marketing trials conducted after the release of the intervention to the market. The phase of a trial is represented here as a nominal (categorical) feature. All remaining features are represented as binary values. The third feature type includes the target conditions of trials. A target condition is the disease or symptom that a clinical study aims to treat. The intervention is the fourth type of feature, which is the treatment or therapeutic method (e.g., drug or medical device) that was tested to address the target condition. The last two feature types include all non-serious adverse events and all serious, non-death adverse events that occurred during the clinical trials. The death outcome is a boolean (true or false) value, and it serves as the prediction target.

A. Evaluation Results of Baseline Models

As only 7.8% of the clinical studies reported death events (1,820 of 23,374), the classification result is highly imbalanced. We used the area under receiver operating characteristic curve (AUROC) and the area under precision recall curve (AUPRC) to overcome the class imbalance problem. James [18] and Kendrick [19] discussed the meaning and use of the ROC curve and PR curve. They are commonly used in diagnositic problems, especially in binary classification of imbalanced datasets. Therefore, it is a particularly good evaluation for our prediction task.

Table 2 shows the comparison of baseline model and optimized model using five different machine learning algorithms. For baseline models, all five algorithms performed well predicting death events using our dataset with the baseline model. The logistic regression algorithm has the highest precision of 0.7177. The random forest algorithm achieves the highest recall at 0.8916 with a 0.4146 accuracy. The best overall F1-score is 0.6838 when we apply the decision tree algorithm. The baseline model indicates that clinical trial features could be used in machine learning models to predict potential occurrence of death events.

B. Optimization: Variance-Based Feature Selection Results

Fig. 3 shows the top 10 features with the highest variances and frequencies. After variance-based feature selection process, the dimension is reduced from 122,586 to 16,379. First, the two features "number of participants" (σ=38147880.4) and "median age" (σ= 5044.5) have a significantly higher level of variance than other features. This is expected because these two features are the only two columns with numerical values. The other columns are all binary values. As shown in Figure 3, among the rest of the top-ranked features, most of features are common, non-serious adverse events in clinical trials, such as headache, nausea, and vomiting.

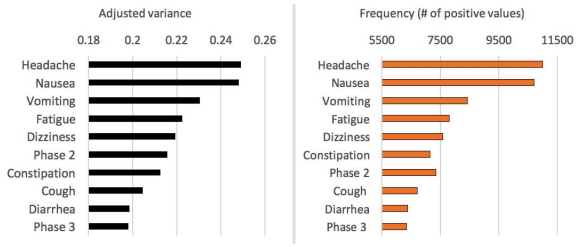


Fig. 3. Top ten features with the highest variance and frequency.

We draw some conclusions from the variance analysis. First, it is difficult to derive any clinical conclusions based on the frequency (or variance) of the features. Since the value of a feature can be only 1 or 0, conducting an analysis of variance would be useless. The top features can reflect only the most common factors in the entire dataset. However, the variance-based feature selection approach is useful in reducing the noise. The reasons are below. First, in the process of machine learning dataset construction, thousands of features are listed as independent factors according to their feature names. There are some types of "bad" features, which almost no one uses in the clinical report. These bad features often have low frequencies for a variety of reasons, whereas "good" features often provide rich information and have higher frequencies.

The high-frequency features are usually higher quality features due to the precision of their terminology and great scalability. The low-frequency features are often problematic in a variety of ways and add noise to the classification model. Therefore, deleting low-variance features can significantly improve the quality of the dataset.

C. Semantic Mapping Results

We adopted the UMLS thesaurus library to conduct semantic mapping. Features with similar semantic meanings were mapped into a single feature. Those features that shared the same Concept Unique Identifier in the library were regarded as having similar semantic meanings. For example, "high blood pressure" and "hypertension" were merged into a single feature. After all synonyms of the feature set were mapped, the dimension decreased from 16,379 to 13,120.

D. Predictive Model Improvement After Feature Engineering

In table 2, the optimized model shows the performance after the three feature selection approaches discussed above. For death events, the support vector machine has the highest F1-score at 0.7075, with precision at 0.8106 and recall at 0.6354. The logistic regression and nave Bayes algorithms also have high F1-scores, reaching 0.6912 and 0.6924, respectively.

According to the comparison between baseline models and optimized models, most algorithms show some level of improvement, although the decision tree and random forest algorithms shows decreases in performance. The nave bayes

TABLE I
CLINICAL TRIAL FEATURE SUMMARY

Type	Name	Example	Value Example	Value Type	Total Count
Predictor variable	Participants' information	Main age	45.8	Numeral	23,374
		# of participants	543	Numeral	23,374
Predictor variable	Phase	Phase number	Phase 3	Nominal 1-4	23,374
Predictor variable	Medical conditions	Renal Failure	1 or 0	Binary	10,371(distinct)
Predictor variable	Trial interventions	Prograf(tacrolimus)	1 or 0	Binary	27,316(distinct)
Predictor variable	Serious adverse events	Anemia	1 or 0	Binary	36,227(distinct)
Outcome variable	Death Events	Death	True	Boolean	1,820 true values

TABLE II
COMPARISON OF FOR FIVE MACHINE LEARNING ALGORITHMS AT PREDICTING DEATH EVENTS

Algorithms	Tags	Baseline Model					Optimized Model				
		Precision	Recall	F1-score	AUROC	AUPRC	Precision	Recall	F1-score	AUROC	AUPRC
Logistic Regression	Non-death	0.5843	0.8314	0.6815	0.7153	0.6258	0.6577	0.7865	0.7115	0.7344	0.8194
	Death Event	**0.7177***	0.4202	0.5254			0.7697	0.6358	0.6912		
SVM(linear)	Non-death	0.6615	0.3686	0.4688	0.5918	0.7384	0.6706	0.8333	0.7382	0.7110	0.799
	Death Event	0.5684	0.8151	0.6649			**0.8106***	0.6354	**0.7075***		
Naive Bayes	Non-death	0.5843	0.8314	0.6815	0.6218	0.7153	0.6590	0.7573	0.6998	0.7042	0.7932
	Death Event	0.7162	0.4201	0.5209			0.7508	0.6510	0.6924		
Decision Tree	Non-death	0.8286	0.6776	0.7406	0.7416	0.7313	0.7982	0.6276	0.6978	0.6952	0.718
	Death Event	0.6134	0.7849	**0.6838***			0.5781	0.7629	0.6529		
Random Forest	Non-death	0.9486	0.6137	0.7405	0.6658	0.7526	0.9678	0.5554	0.7012	0.6202	0.7354
	Death Event	0.4146	**0.8916***	0.5617			0.3099	**0.9154***	0.6623		

Fig. 4. Change in performance from before optimization (Table 2 baseline) to after feature selection

model has the highest level of improvement with a 0.1670 increase in F1-score. In terms of F1-score, the support vector machine algorithm is the best model, with an improvement of 4.2%. This F1-score is high enough to be valuable for estimating the risk for a critical event such as patients death. The result also indicates that the methods used for feature selection are reliable approaches to improving the performance of predictive models. Overall, the five machine learning algorithms deliver impressive performances in predicting clinical outcomes. Fig. 4 provides the measurements of five classifiers in area under ROC curve, and area under PR curve.

IV. DISCUSSION

This study explored the potential of leveraging clinical trial reports to build models to predict adverse events. To identify clinical factors that contribute to the risk of death, we built novel machine learning models to predict death outcomes in clinical trials. Our model could be used to improve trial design and safety. By leveraging clinical trial big data, we developed a method to normalize and evaluate clinical trial factors. We analyzed these clinical trial factors and optimized the model using statistical methods. The selected factors were then used to train machine learning classification algorithms to predict death events. The best algorithm (the support vector machine algorithm) achieved a precision of 81.06% and a recall of 63.54% in predicting death events.

Our study focuses on clinical trial adverse outcomes [20]. Clinical trials are designed to study and evaluate the safety and efficacy of medical interventions. Due to the experimental nature of clinical trials and the uncertainty of risk factors, participants in clinical trials are often affected by adverse events [6], [21], [22]. Some of these adverse events may have lethal effects [23], [24]. Our study shows the potential to leverage data-driven techniques to improve clinical trial design and monitoring, which could lower participants risk of harmful events. This study also demonstrates the possibility of using clinical trial data to predict the risk of life-threatening events.

A. Feature Analysis

The feature analysis presents little potential in interpreting the clinical meaning, and further professional analysis may be needed. In terms of performance, the unoptimized dataset and optimized dataset each have advantages, depending on the adopted machine learning algorithms. Another point is that a binary classifier is not the best option for predicting the occurrence of death events, especially when the result is used for clinical analysis. Instead, developing targeted probability models could be a better approach for estimating the risk of death events.

B. Limitations and Future Work

There are two major limitations of this study. First, we evaluated each clinical factor independently, overlooking the potential internal correlations between clinical factors. Research shows that adverse event combinations could be key predictors of severe adverse events [25]. Applying association rule mining algorithms and generating feature combinations could be helpful in uncovering strong combinational predictors for death events and could improve the performance of machine learning models. Second, we used only structured data, such as target conditions, interventions, and adverse events. Some unstructured data, such as plain-text descriptions of clinical trials, may contain key information that is more important than the structured data in Table 1. Finally, the predictive outcome of this study is a boolean value instead of a probability value. The reason is most trials reported only a small number of deaths, which makes it difficult to develop probability models to predict the incident rate of death events.

V. Conclusion

By leveraging clinical trial big data, we developed machine learning classifiers for predicting death events in clinical trials. This study indicates that routine clinical trial reports contain key clinical factors that can be used to build predictive models to estimate the risk of death events. Five algorithms were evaluated in this study. The best model is based on the support vector machine algorithm, and it has a precision of 81.06%, a recall of 63.54%, and an F1-score of 70.75%. Our model could be used to predict death events and to improve the safety of patients who participate in clinical studies.

Our methodology has two innovative aspects: 1) We propose a machine learning-based approach to convert a text clinical report to a matrix dataset that is friendly for machine learning training algorithms. 2) We use novel statistical analysis to optimize the dataset and produce better performance. This predicting classifier can serve as an adjunct tool of risk control for clinical experts. Here is our vision: At the start of a clinical study, a clinical trial expert is able to assess whether serious adverse events are likely to occur in the upcoming clinical trial and to see what factors contribute most to those potential serious adverse events.

References

[1] E. H. Shortliffe, Biomedical informatics: computer applications in health care and biomedicine. London: Springer, 2013.

[2] "2015-2016 Global Participation in Clinical Trials Report." [Online]. Available: https://www.fda.gov/downloads/Drugs/InformationOnDrugs/UCM570195.pdf. [Accessed: 20-Jan-2018].

[3] W. F. Rosenberger and J. M. Lachin, Randomization in clinical trials: theory and practice. Hoboken: Wiley, 2016.

[4] "FDA suspends (and promptly restarts) clinical trial for new cancer therapy," Science in the News, 19-Jul-2016. [Online]. Available: http://sitn.hms.harvard.edu/flash/2016/fda-suspends-promptly-restarts-clinical-trial-new-cancer-therapy/. [Accessed: 12-Dec-2018].

[5] S. R. Ahmad, "Adverse drug event monitoring at the food and drug administration," Journal of General Internal Medicine, vol. 18, no. 1, pp. 57-60, 2003.

[6] L. M. Friedman, C. Furberg, and D. L. DeMets, Fundamentals of clinical trials. New York: Springer, 2010.

[7] Z. Obermeyer and E. J. Emanuel, "Predicting the Future Big Data, Machine Learning, and Clinical Medicine," New England Journal of Medicine, vol. 375, no. 13, pp. 1216-1219, 2016.

[8] A. Colubri, T. Silver, T. Fradet, K. Retzepi, B. Fry, and P. Sabeti, "Transforming Clinical Data into Actionable Prognosis Models: Machine-Learning Framework and Field-Deployable App to Predict Outcome of Ebola Patients," PLOS Neglected Tropical Diseases, vol. 10, no. 3, 2016.

[9] Chaurasia, Vikas and Pal, Saurabh, A Novel Approach for Breast Cancer Detection Using Data Mining Techniques (June 29, 2017). International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) vol. 2, Issue 1, January 2014. Available at SSRN: https://ssrn.com/abstract=2994932

[10] C. S.dangare and S. S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques," International Journal of Computer Applications, vol. 47, no. 10, pp. 44-48, 2012.

[11] J. Luo, C. Eldredge, C. C. Cho, and R. A. Cisler, "Population Analysis of Adverse Events in Different Age Groups Using Big Clinical Trials Data," JMIR Medical Informatics, vol. 4, no. 4, 2016.

[12] J. Allyn, N. Allou, P. Augustin, I. Philip, O. Martinet, M. Belghiti, S. Provenchere, P. Montravers, and C. Ferdynus, "A Comparison of a Machine Learning Model with EuroSCORE II in Predicting Mortality after Elective Cardiac Surgery: A Decision Curve Analysis," Plos One, vol. 12, no. 1, 2017.

[13] R. A. Taylor, J. R. Pare, A. K. Venkatesh, H. Mowafi, E. R. Melnick, W. Fleischman, and M. K. Hall, "Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data-Driven, Machine Learning Approach," Academic Emergency Medicine, vol. 23, no. 3, pp. 269-278, 2016.

[14] G.F. Cooper, C.F. Aliferis, R. Ambrosino, J. Aronis, B.G. Buchanan, R. Caruana, et al. "An evaluation of machine-learning methods for predicting pneumonia mortality," Artificial Intelligence in Medicine, vol. 9 no. 2, pp. 107-138, 1997.

[15] Shukla, D. P., Shamsher Bahadur Patel, and Ashish Kumar Sen. "A literature review in health informatics using data mining techniques." International Journal of Software and Hardware Research in Engineering, vol. 2, no. 2. pp. 123-129, 2014.

[16] "For Researchers," ClinicalTrials.gov. [Online]. Available: https://clinicaltrials.gov/ct2/help/for-researcher. [Accessed: 12-Apr-2019].

[17] "Home," Home - ClinicalTrials.gov. [Online]. Available: https://clinicaltrials.gov/ct2/home. [Accessed: 15-Apr-2019].

[18] J. A. Hanley and B. J. Mcneil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," Radiology, vol. 143, no. 1, pp. 29-36, 1982.

[19] K. Boyd, K. H. Eng, and C. D. Page, "Area under the Precision-Recall Curve: Point Estimates and Confidence Intervals," Advanced Information Systems Engineering Lecture Notes in Computer Science, pp. 451-466, 2013.

[20] I. R. Edwards and J. K. Aronson, "Adverse drug reactions: definitions, diagnosis, and management," The Lancet, vol. 356, no. 9237, pp. 1255-1259, 2000.

[21] P. Olivier, A. Gimbert, A.-L. Colin, F. Salvo, M. Becker, V. Marty, J.-L. Montastruc, and N. Petitpain, "Serious adverse drug events related to non-investigational drugs in academic clinical trials: another source of safety data for risk assessment?," British Journal of Clinical Pharmacology, vol. 82, no. 4, pp. 1069-1075, 2016.

[22] M. Unnikrishnan, "Eminence or evidence? The volatility, uncertainty, complexity, and ambiguity in healthcare," Journal of Pharmacology and Pharmacotherapeutics, vol. 8, no. 1, p. 1, 2017.

[23] B. S. Cooper, M. F. Boni, W. Pan-Ngum, N. P. J. Day, P. W. Horby, P. Olliaro, T. Lang, N. J. White, L. J. White, and J. Whitehead, "Evaluating Clinical Trial Designs for Investigational Treatments of Ebola Virus Disease," PLOS Medicine, vol. 12, no. 4, 2015.

[24] C. Heneghan, B. Goldacre, and K. R. Mahtani, "Why clinical trial outcomes fail to translate into benefits for patients," Trials, vol. 18, no. 1, 2017.

[25] Luo, Zhihui et al. "Mining patterns of adverse events using aggregated clinical trial results." AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science. pp. 112-116, 2013.