

ポスター | 医療データ解析

ポスター6

医療データ解析・医療支援

2021年11月20日(土) 09:00 ~ 10:00 P会場 (イベントホール)

[3-P-1-05] 医薬品の臨床開発におけるモニタリングレポートを活用した新しい施設評価手法の試み —機械学習分類モデルによる臨床開発モニターの質的な施設評価の再現—

*辻 真太郎¹、小林 大輔²、荻野 珠樹² (1. IQVIAソリューションズ ジャパン株式会社, 2. IQVIAサービシーズ ジャパン株式会社)

*Shintaro Tsuji¹, Daisuke Kobayashi², Tamaki Ogino² (1. IQVIAソリューションズ ジャパン株式会社, 2. IQVIAサービシーズ ジャパン株式会社)

キーワード : Clinicat trial, Monitoring reports, Machine-learning

背景と目的 近年、製薬企業は、治験を実施する上で候補の医療機関に対する試験実施時のコンプライアンスや、取り扱うデータの質などの「施設の定性的な評価」に強い関心を寄せている。IQVIAではこうしたニーズに応えるため、施設に関する質的な情報を収集するアンケートを臨床開発モニターに対して実施した。しかし、膨大な施設数を対象としたアンケートは、モニターへの負担から頻繁に実施することは容易ではない。そこで本研究では、アンケートで得られた質的評価を回答の得られなかった施設や新規施設に拡張することを目的として、IQVIAに蓄積されているモニタリングレポートとアンケート結果を用いて機械学習の分類モデルを構築し、モニターの質的な施設評価を疑似的に再現した。**方法** アンケートで該当した116試験の2,116レコードを対象とした。教師データは、質的な観点からの評価(該当施設=1, それ以外=0)と、モニタリングレポートに記載されているプロトコル逸脱や手続き期限の超過日数などの25項目、治験の治療領域等とした。全データの75%を教師データに、残りの25%を検証データとした。機械学習の分類アルゴリズムは Random Forestを使用した。モデルによる施設評価は、モニター経験のある複数の関係者から聞き取り調査を実施して妥当性を検討した。**結果と考察** テストデータの検証結果は、Accuracy=97.7%、Precision=0.959、Recall=0.886、F1 score=0.921となった。聞き取り調査の結果、構築したモデルは同程度の症例登録が予想される施設選択時に参考する運用上で十分な精度と判断された。本研究結果により、モニターの質的評価をモニタリングレポートのデータから再現し、アンケートで回答の無かった施設や今後新規に追加される施設にも質的評価を行うことが可能となった。

医薬品の臨床開発におけるモニタリングレポートを活用した 新しい施設評価手法の試み

- 機械学習分類モデルによる臨床開発モニターの質的な施設評価の再現 -

辻 真太郎*1、小林 大輔*2、荻野 珠樹*2

*1 IQVIA ソリューションズ ジャパン株式会社、*2 IQVIA サービスズ ジャパン株式会社

An Attempt of a New Site Evaluation Method Utilizing

Monitoring Reports in Clinical Trial

- Reproduction of Qualitative Site Evaluation by Clinical Research Associate
Leveraging Machine-learning Classification Models -

Shintaro Tsuji*1, Daisuke Kobayashi*2, Tamaki Ogino*2

*1 IQVIA Solutions Japan K.K., *2 IQVIA Services Japan K.K.

Background and Objectives Pharmaceutical companies are paying attention to qualitative site evaluations, including site compliance and smoothness of clinical trial. Against such a background, IQVIA conducted surveillance for CRAs (Clinical Research Associates) to gather insights. However, this does pose a challenge, given this is significantly time consuming for the CRAs to provide their feedback, leading to operational delays. Our research aims to expand CRA's qualitative site evaluations for sites with no feedback, by leveraging machine-learning models on the monitoring reports and questionnaires in IQVIA. **Methods** 2,116 records in 116 studies identified by CRAs were used. Nine categories, including monitoring reports, site management organization availability, and therapeutic area, were used for modeling based on Random Forest. CRA's judgments were defined as binary classification. The models were reviewed by stakeholders, who used the results from this model in actual projects. **Results and Discussion** Performance of models by the test was AUC=0.940, accuracy=97.7%, precision=0.959, recall=0.886, and F1 score=0.921. The reviewer's feedback showed that it is necessary to show evidence-based data to get agreements for utilizing machine-learning-based CRA qualitative evaluation by the users.

Keywords: Clinical trial, Monitoring reports, Machine-learning

1. 緒論

製薬企業にとって治験に参加する施設の選定は、治験の成功やその効率的な実施を左右する重要な工程である^{1,2)}。IQVIA では、候補に挙がった医療機関の過去の治験実績に複数の医療データを組み合わせ、いわゆる「定量値に基づく施設の評価」により候補施設を製薬企業に提案している。

その一方で近年製薬企業は、企業治験を実施する上で候補の医療機関に対し、試験実施時のコンプライアンスや治験の実施の円滑さなどの定量データでは測れない「施設の定性的な評価」に強い関心を寄せている。このような定性的評価は、上記の定量値にはない情報を含んでおり、施設の治験実施の可否を評価する上で無視できない。

IQVIA ではこうした製薬企業のニーズに応えるため、施設に関する質的な情報を収集するアンケートを臨床開発モニター(以下、モニター)に対して実施した。しかし、定性評価を得るアンケートには以下の問題がある。(1)膨大な施設数・試験数を対象としたアンケートの実施は、モニターへの負担から頻繁に実施することは容易ではない。(2)モニター全員からのアンケートを回収することは容易ではない。従って、アンケート実施に対するモニターの負荷や、回答率を考慮すると、定性的な評価を何らかの手法を用いて拡張して補完する必要がある。

2. 目的

機械学習モデルによるクラス分類は、限られたモニターによる施設の定性評価を補完する手法として最適と考えられる。そこで本研究では、アンケートで得られた医療機関の質的評価を回答の得られなかった施設や新規施設に拡張することを目的として、IQVIA に蓄積されているモニタリングレポートとア

ンケート結果を用いて機械学習の分類モデルを構築し、モニターの質的な施設評価を疑似的に再現した。

3. 方法

本研究ではモニターの定性的な施設評価を判定する機械学習モデルの構築を行うこととし、施設評価アンケート及び施設におけるモニターの活動の記録であるモニタリングレポートを用いた。

3.1 臨床開発モニターへの施設評価アンケート

IQVIA に所属する約 1,000 名のモニター及び、現在はモニターとして勤務していないが過去にモニター経験のある社員を対象としたアンケートを実施した。また、施設の定性評価は、医療機関の体制によって大きく変化する可能性があることから、過去 5 年以内にモニターとして施設を訪問した際の評価を取集した。具体的には、「同程度に患者登録数が見込める候補施設が複数ある場合、自身の担当した施設は候補から除外すべきか」という質問に対して「はい」または「いいえ」での回答形式とした。また、除外すべきと判断した根拠として次に示す 6 つの回答項目を作成し回答を得た。(1)治験実施計画書からの逸脱に関する要因、(2)原資料及び Electronic data capture(以下、EDC)に関する要因、(3)コミュニケーションに関する要因、(4)人的な要因、(5)施設固有の手続きに関する要因、(6)その他の要因(自由記載)。(6)以外の項目は 3 段階の選択式(やや該当する:1、該当する:2、非常に該当する:3)とし、それぞれの数値をスコアとして利用した。モニタリングレポートの概要を表 1 に示す。

3.2 モデリングに使用したモニタリングレポート

本研究では表 1 のモニタリングレポートに加えて、(1)Site

Management Organization (以下、SMO)による施設に対する治験業務支援の有無、(2)試験の治療領域を加えてモデリングを行った。(1)については SMO の支援により施設内の手順が大きく影響を受ける理由から、そして(2)に関しては、アンケートの回答が試験毎に回答されることから追加を行った。

表1 モニタリングレポートの概要

項目名	概要
患者登録数	治験に登録された患者数
治験実施計画書からの逸脱数	治験実施計画書に規定された試験デザインや手順からの全ての変更、相違、乖離の数
モニターへの訪問数	モニターへの施設への訪問数
施設の立ち上げ期間	施設選定実施日～試験開始日までの日数
モニターのアクションアイテムの数	モニターが解決する必要がある、当該施設に関する積み残し事項の数
アクションアイテムの解決期日を超えた日数	区分(31～60日、61～90日、91～120日、121日以上)別の数
EDCの入力期間	入力開始日から実際に施設が入力するまでの日数

3.3 モデリングの概要

本研究におけるモデリングの開発環境は Cloudera Data Science Workbench を利用し、開発言語は Python、機械学習モデルの構築には Scikit-learn を利用した。アルゴリズムは Random Forest とし、全データの 75% を教師データに、残りの 25% を検証データとした。モデルのパラメータはグリッドサーチを実施して最適な組み合わせを決定した。モデルの検証ではクロスバリデーションを実施した。モデルを評価する指標として AUC(Area Under the Curve)、Accuracy、Precision、Recall、F1 score を算出した。

3.4 モデルの実運用に向けた聞き取り調査

構築したモデルの社内での実用を視野に入れた検証として、モニタリング責任者などの利用者に対して機械学習モデルの判定に対する効果的な利用方法について聞き取り調査を実施した。

4. 結果

4.1 回収したアンケートの概要

アンケートにより延べ 283 施設に対する施設評価を得た。純施設数は 157 であった。アンケートで回答された施設の回答毎のスコアの平均値を表 2 に示す。最も影響の大きかった項目は「人的な要因」と「施設固有の手続きに関する要因」であり、モニターの施設の定性評価に強く影響を与えていることが明らかとなった。

表2 アンケートの回答施設とその平均スコア

候補施設から除外すべきと考えた要因	スコア平均値
(1)プロトコルからの逸脱に関する要因	0.77
(2)原資料/EDCに関する要因	1.00
(3)コミュニケーションに関する要因	1.51
(4)人的な要因	1.70
(5)施設固有の手続きに関する要因	1.69

4.2 モニタリングレポート

社内データベースからアンケートで該当した試験のデータを抽出したところ、116 試験が該当し、これらの試験に参加している施設は延べ 2,116 施設であった。

4.3 機械学習モデルの構築

テストデータの検証結果は、AUC=0.94、Accuracy=98.2%、Precision=0.99、Recall=0.89、F1 score=0.94 となった。モデルによるパラメータの重み付けは、「モニターの訪問数」が最も高く 39%、次いで「アクションアイテム数」が 20% となった。

4.4 施設選定関係者に対する聞き取り調査

社内の関係者 50 名の聞き取り調査により、機械学習モデルの判定結果を効果的に用いるために、主に以下の 2 つの要望が挙げられた。

1. モデリングに使用したデータの数値を参照したい。
2. モデルによる判定の基準を参照したい。

5. 考察

アンケートの結果から、モニターの施設評価の傾向として「人的な要因」と「施設固有の手続きに関する要因」という要素が候補施設には適さないと考える主な要因となっていることが明らかとなった。これに対して、構築したモデルは「モニターの訪問数」と「アクションアイテム数」を重視して分類を行っていることが示された。今回構築したモデルがモニターの評価を再現できた理由の 1 つとしては、「施設の固有の手続きが増える」と、「モニターの訪問数・アクションアイテムが増加する」という一般的に認められる関係性がアンケートとモニタリングレポート間に認められたためだと考える。

一方、「人的な要因」もモニターにとって重要な指標であることが示されたが、今回利用したモニタリングレポート等に直接関連性のある項目が存在しなかった。そこで、今後アンケートの自由記載欄からモニターの判断の背景を自然言語処理等で精査する必要があり、今後の課題としたい。

機械学習モデルをビジネスで運用する際には、利用者の納得感を得た上で導入していくことは、社内での活用促進という点で必要不可欠なプロセスである。本研究では、利用者に対する聞き取り調査により、機械学習による判断の根拠となるデータの提示とその可視化が期待されていることを明らかにした。今後はモデルの判定結果に使用されたデータの可視化を行い、実運用にて利用者参照してもらう仕組みを構築する予定である。

6. 結論

本研究結果により、モニターの質的評価をモニタリングレポートのデータ等から機械学習モデルを用いて疑似的に再現し、アンケートで回答の無かった施設や今後新規に追加される施設にも質的評価の提供を行うことが可能となった。

参考文献

- 1) 稲野 彰洋, 内田 英二, 井上 悠輔. 多施設共同治験における倫理審査集約化の現状と課題. 臨床薬理 2018 ; 49 (4) :159-167.
- 2) 楠岡 英雄. 総説 治験・臨床研究の歩みについて. 医療と社会 2018 ; 28(1) : 3-16.