

機械学習による臨床試験データの品質向上

西 基秀 メディデータ・ソリューションズ（株）
Vice President, Professional Services and Customer Care

1 はじめに

臨床試験は新たな医薬品・医療機器や医療技術の開発に欠かすことの出来ないステップである。新規医薬品・医療機器等の安全性と評価は、そこで収集された臨床試験データを用いて評価される。そのため、臨床試験データの信頼性を担保することは、その臨床試験結果の影響を受ける全ての患者の人権保護と安全を担保することに繋がる。

臨床試験を取り巻く環境は大きく変化し、臨床試験の複雑性は年々増している。グローバル臨床試験の増加に伴い、様々な国の医療機関が臨床試験に参加するようになった。臨床試験で扱うデータは、電子データ収集システム、医療画像、臨床検査、電子カルテ記録、ゲノムデータ、バイオメディカルセンサー等が含まれるまでに広がっている。さらには、IT技術の進歩により、データを収集するスピードは桁違いに早くなった。臨床試験データの信頼性を担保していくためには、データのモニタリングとレビューに代表される品質管理活動が増々重要な役割を担うことになるが、既存のシステム・プロセスによるデータのモニタリングとレビューでは、このような環境の変化に対応していくことは難しい。今般改定されたICH-E6（R2）ガイドラインにおいても、品質管理及び品質保証を包括する概念として品質マネジメントの整理が行われ、革新的なデータの信頼性確保への取組が求められている。

Medidata 社がライフサイエンス企業の役員と幹部を対象に実施したアンケート¹⁾では、「臨床試験データの信頼性の確保」が、企業戦略を達成する上で直面している重要な課題の一つとして認識されている（表1）。

表1 企業戦略を達成するうえで直面する重要な課題

1	被験者募集／登録の改善
2	臨床試験データの信頼性の確保
3	被験者参加継続の強化
4	経費の抑制
5	スケジュールの達成

（回答数の順に記載）

規制当局からの新たな医薬品・医療機器の承認を取得するためには、信頼性のある臨床試験データを提供することが不可欠である。しかし、今日のプロセスでは、エラーの発見の遅れや見過ごしが発生し、データの品質が損なわれる可能性がある。結果として、臨床開発の遅延とコストの増加に繋がる。品質管理及び品質保証をより効果的に活用して臨床試験データの品質を担保するためには、適切なIT技術の導入、適切な体制の構築、そして包括的なプロセスのもとでの実施が必要となる。

2 新たなデータ管理パラダイム

従来の臨床試験データのモニタリングとレビューでは、被験者データを一次的にとらえ、違う情報源から得られたデータの多くを症例の別の側面として分けて扱ってきた。しかし、臨床試験で扱われるデータ量やデータソースが増加してきた今日では、被験者データをより包括的にとらえることが重要である。環境の変化に対応し、臨床試験データの品質を確保するためには、従来の静的な臨床試験のデータ管理（データの入力、データレビュー、データロックの連続）からアダプティブなデータ管理に移行する必要がある。新たなデータ取

集技術と人工知能や機械学習のようなデータサイエンス機能を採用することで、臨床試験においてもアダプティブなデータ管理を実現できる。機械学習は、反復型かつ応用型のシステム主導のプロセスであり、膨大なデータセットを組み合わせることで、マニュアルプロセスでは検出が困難な異常や傾向を、より高い確率で検出することができ、臨床試験データの要素間で未だ明らかになっていない関係性を明らかにすることができる。機械学習を利用したアダプティブなデータ管理に移行することで、データの品質が向上するだけでなく、「被験者募集/登録の改善」、「治験実施医療機関の管理」、「有害事象の予測」、「治験薬供給・管理」など、臨床試験の多くの課題に直接的な影響を与えることができる。

3 機械学習による臨床試験データのモニタリングとレビュー

多様な機械学習アルゴリズムを用いることで、臨床試験データのエラーを自動で検出することや異常データへのフラグ付けを行うことが可能となる。例えば、被験者背景、有効性の評価データ、臨床検査値、有害事象、服薬状況を含むすべての収集データを使用して、臨床試験の被験者を複数のクラスターに分類する^{2,4)}。クラスター毎にデータの分布を分析して期待値を特定し、それを基にパターン（すなわち規則）の発見と、そのパターンに適合していないデータへのフラグ付けを行う。パターンは患者ごとおよび患者間など、いくつかのレベルで特定する。このような機械学習アルゴリズムを用いた場合、一つの臨床試験データから、通常3,000を超えるパターンが検出される。すなわち、機械学習により、様々なタイプの異常データと異常傾向を検出することが可能となる。機械学習アルゴリズムが検出した異常データと異常傾向のいくつかを紹介する。

3.1 不整合データ

図1の散布図は、腫瘍診断日と腫瘍切除日のデータを示している。腫瘍診断日と腫瘍切除日には高い相関が認められるが、矢印で示した2名の被験者は、データから導き出されたパターンから逸脱していることがわか

る。この2名は、腫瘍と診断される前に腫瘍の切除が行われており、データに誤りがあることはほぼ明らかである。このような異常データは従来のモニタリングとレビューでも検出可能である。しかしながら、機械学習を用いた場合、このようなパターンをユーザーが前もって定義する必要がなくなり、自動的に問題を検出することができる。

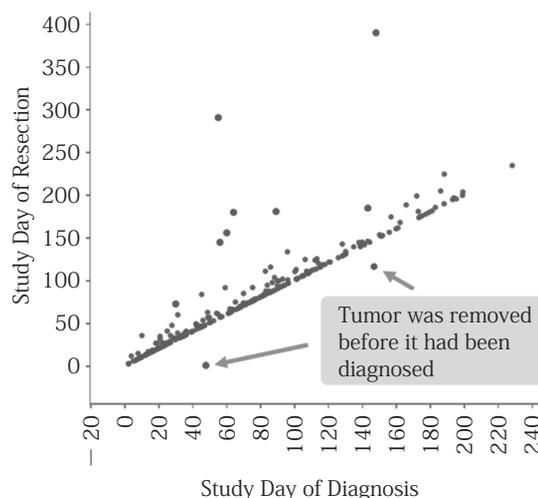


図1 腫瘍診断日と手術日

3.2 ミスコンダクト

図2の散布図は、一つの施設で組み入れられた被験者（13名）の拡張期血圧と収縮期血圧のデータである。データに規則性が見られ不自然なデータであることがわかる。確認の結果、この施設ではデータを四捨五入していた。個々のデータは正常範囲内であり問題としては認識されないため、このような異常パターンを従来のデータのモニタリングとレビューにより検出することは難しく、見逃す可能性が高い。

図3の散布図は、心拍数と呼吸数間隔のデータを示している。機械学習プログラムがこの2つの変数にパターンがあることを検出している。楕円で示した部分はパターンから外れており、全ての異常データは特定の施設からのデータであった。この施設のパターンが他の典型的な施設とは異なっていることが判る。

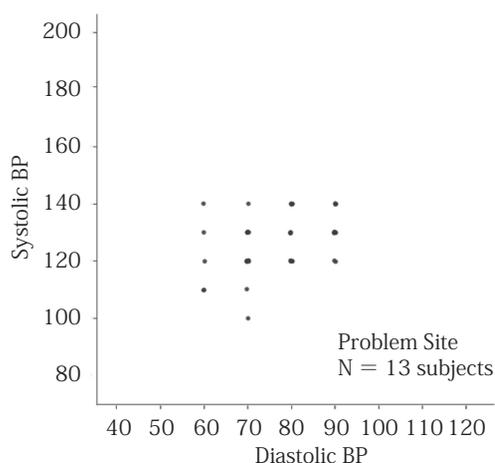


図2 1施設の拡張期血圧と収縮期血圧

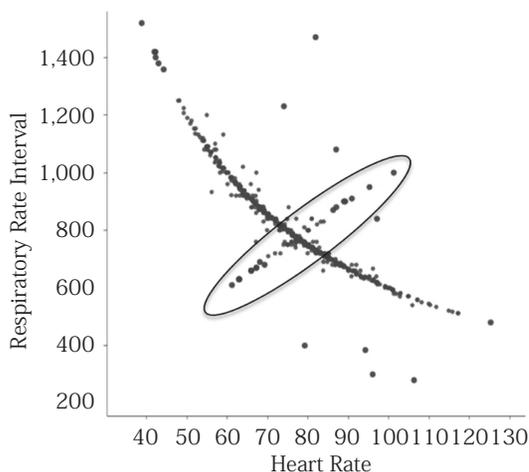


図3 心拍数と呼吸数間隔

3.3 データの偽造

図4は、施設で組み入れられた被検者データの類似性を示している。図上部のバーは、各施設に登録された被検者の数を表し、棒グラフの色違いの部分、機械学習により各施設内で特定された類似性の高い被検者の数である。棒グラフの円で示した施設では、5名の被検者が組み入れられており、機械学習により全ての被験者間でデータの類似性が高いと判断された。類似性が高いかどうかを判断していくためには、被検者間の多くのデータの調査が必要となるため、従来のモニタリングとレビューでは検出が難しい。機械学習では多くのデー

タを短時間で扱えるため、機械学習から得られたパターンに対し、被検者プロファイルを調査し、類似していると特定された理由を確認していくことができる。この場合、5名の被検者の一つの臨床検査値に類似性が見られた(図4)。さらには、稀な有害事象(試験全体では2%程度の発現率)が、この施設の5名の被験者全てに発現しており、被検者間のVisitのタイミングや間隔が非常に類似していたことも確認された。

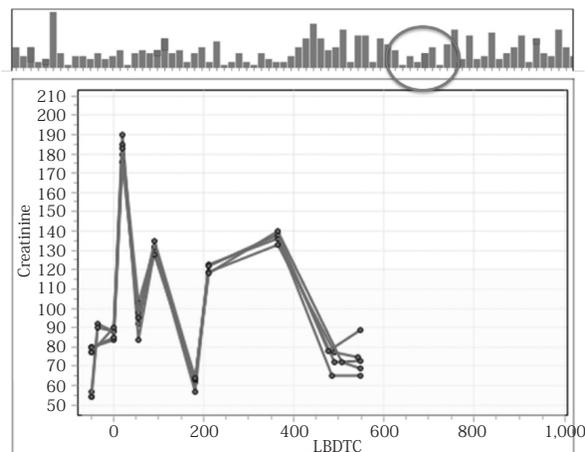


図4 被検者データの類似性

機械学習による異常データおよび異常パターンの検出により、収集されたデータのモニタリングとレビューが効率化されるだけでなく、リスクの大きい被験者および施設、リスクの大きいプロセスやオペレーションの特定ができるようになり、将来起こり得る問題に早期に対応していくことが可能となる。

4 機械学習を用いた意思決定能力の向上

ライフサイエンス企業は、臨床開発段階でタイムリーかつ適切な意思決定を行っていく必要がある。機械学習は異常データおよび異常パターンの検出をするだけでなく、企業の意思決定をサポートするツールとしても有用である。

有害事象、併用薬、既往歴などの様々な報告用語は、スペルミスなどが良く見られ、同件事象や薬剤であっても、報告用語にばらつきが出る。報告用語をコーディン

グすることでデータに一貫性が生まれ、安全性や効果を評価することが可能となる。しかしながら、このプロセスには時間も費用もかかるため、機械学習による効率化が有効である。

例えば、有害事象の報告用語を自動で MedDRA 用語にコーディングをする場合、複数の自然言語処理技術を適切に組み合わせることで、処理量を最小限に抑え、且つ報告用語に最も一致する MedDRA 用語を探し出すことができる。自然言語処理技術に加え、過去に報告された用語を参照し、マッチング率を上げることができるが、スペルミスなどへの対応は難しい。ここまでは従来の自動コーディング機能である。そこで、過去に報告された用語から学習アルゴリズム (Naive Bayes Classifier⁵⁾) を最後のステップに加えることで、報告用語と過去に MedDRA 用語にコーディングされた報告用語の類似性を考慮させることができる。Medidata 社の調査では、自然言語処理のステップのみを利用した自動コーディングのマッチング率は 93.1% (マッチングカバー率: 95.8% × 精度: 97.2%) であった。Naive Bayes Classifier を追加した場合、マッチング率は 95.0% (マッチングカバー率: 100% × 精度: 95.0%) となり、全体のエラー率が約 2% 減少した。

このように、機械学習を利用して過去の臨床試験データからの学習を行い、一貫性のある均質なコード化を自動で行うことができる。正確な有害事象名のコード化は、シグナル検出の改善に繋がるため、安全性のリスクをより早期に特定することが可能となる。さらに、関連する臨床的特徴と有害事象の重篤度を機械学習のアルゴリズムで学習させることにより、規制当局への有害事象の報告時間を短縮できる。

5 コスト オブ クオリティ

従来の臨床試験データのモニタリングとレビューでは、エディットチェックの作成やマニュアルデータレビューに多額の投資をしている。そこで、米国で実施された臨床試験のマニュアルデータレビューとエディットチェックについての分析を行った (Medidata 社による分析)。分析の結果、約 50% 程度のマニュアルデータレ

ビューを機械学習により自動化することが可能であり、マニュアルデータレビューに必要な時間を短縮できることが判った。また、74% のエディットチェックは、データの問い合わせ (クエリ) につながっていないことが示され、これらのほとんどがプログラミングを必要とする発生頻度の少ないイベントを検出するためのチェックであった。これらのエディットチェックの多くは、機械学習が得意する異常データやパターンの検出に該当する。

この分析から、第三相試験で用いたデータのモニタリングとレビューに機械学習を用いた場合、約 5,000 ドル/月、試験全体で約 150,000 ドルのデータレビューに掛かる人件費を削減することが期待できる (臨床試験のデータレビュー期間を 30 ヶ月で計算)。一方で、機械学習の利用には追加のシステム費用が必要となるが、データの信頼性の確保による将来のレギュラトリーリスクの軽減等を考慮すれば、合理的な投資と言える。

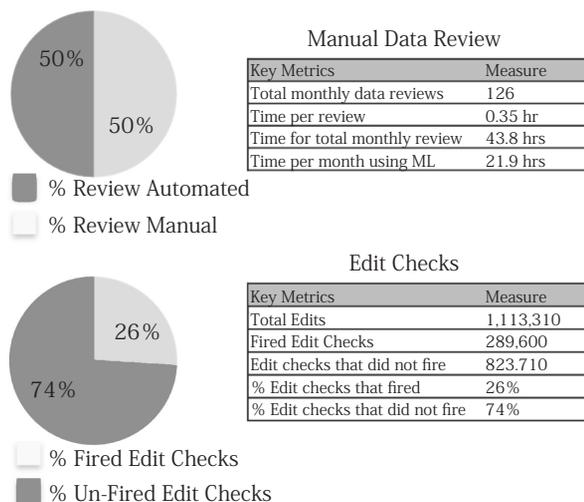


図5 マニュアルレビューの割合とエディットチェックによる検出率

医薬品開発は、規制当局の承認まで約 10 年以上もの長い開発期間と 200 ~ 300 億円もの費用がかかる。このような中で、申請時や上市後に、臨床試験データの信頼性に起因した不祥事が発生した場合、企業の利益と成長に大きく影響するだけでなく、医師や患者に与えるマイナスイメージは非常に深刻である。すなわち、臨床試験データの信頼性の問題を長期間放置した場合、掛かる

コスト（失敗コスト）は増加し、膨大になる。機械学習機能を利用して、将来起こり得る問題に早期に対応していくことで、長期的な費用対効果を向上させることができる。

6 機械学習の利用に対する課題

データ量が少ない場合、機械学習では実用上十分な精度が出せないため、適用範囲が限定される。ある程度の症例データおよびデータポイントが入力されるまでは、機械学習を用いた異常データおよびパターンの検出は難しい。すなわち、臨床試験の開始段階で早期に対応を要する問題の検出には、従来のデータモニタリングとレビューが有効である。

リアルタイムに継続的、且つ、データの転記や転送エラーがない状態で、臨床試験データを収集することができれば、機械学習の効果を早期に享受することができる。モバイルヘルスを利用したデータ収集、等 eSource データ利活用（電子カルテ等）に加え、データを集約する柔軟なプラットフォームを導入していくことも有効な手段である。これらの技術を利用してデータの収集、管理、分析をしていくことは、将来の医療技術・臨床開発に必要なエビデンスを提供していくための基盤となる。これらの新たな技術を適用していく能力、さらには、適切な機械学習アルゴリズムや統計モデリングを適用するためのデータサイエンスの能力、機械学習が検出した異常データやパターンおよび洞察に対し、適切な判断と対応をしていく能力が必要となる。

7 まとめ

臨床試験データのモニタリングとレビューに機械学習を利用することで、個々のデータの品質向上だけでなく、従来検出することが困難であった、データを包括的にとらえた異常パターンを検出し、適切な対応をしていくことが可能となる。さらに、人的および時間的リソースを削減することができ、それらのリソースをより価値の高い活動に集中させることができる。例えば、機械学習が検出した異常データやパターンおよび洞察に対し、適切な

対応と意思決定をしていくことや、リスクの大きい被検者および施設、リスクの大きいプロセスやオペレーションを早期に特定し、将来起こり得る問題に対応していくこと、などである。臨床開発プロセスの早い段階でデータ品質の問題と将来起こりうるリスクへの対応をしていくことは、早期の承認取得だけでなく、データ信頼性に掛かる失敗コストを最小限に抑えることに繋がる。

参考文献

- 1) Medidata Solutions, "Using Data and Analytics in Clinical Development : An Industry Perspective, An eBook From Medidata Drawn from an online survey of Clinical Research Companies fielded in conjunction with PharmaVoice", 2017
https://www.mdsol.com/sites/default/files/Analytics-Survey_2017_0612_Medidata_eBook.pdf
- 2) Hastie, Tibshirani and Friedman, The Elements of Statistical Learning, New York : Springer, 2001
- 3) Gower, "A general coefficient of similarity and some of its properties", Biometrics, pp. 27, 623-637, 1971
- 4) Tibshirani, Walther and Hastie, "Estimating the number of data clusters via the Gap statistic", Journal of the Royal Statistical Society B, pp. 63, 411-423, 2001
- 5) https://en.wikipedia.org/wiki/Naive_Bayes_classifier
- 6) Adam Ross Miller, "Machine Learning within the Clinical Trial Process", Applied Clinical Trials, Nov 15, 2017