

平成18年度データマイニング手法の導入に関する検討概要について

平成19年6月

独立行政法人医薬品医療機器総合機構

独立行政法人医薬品医療機器総合機構（PMDA）では、中期計画期間中に審査等業務及び安全対策の向上に係る目標を達成するためにとるべき措置の1つとして、データマイニングとよばれる新規手法を平成18年度までに確立し、平成20年度までに業務へ導入するものとしております。

平成16年度に策定したスケジュール（図1参照）に従い、平成18年度は、業務に導入するデータマイニング手法の確定を目的として、主にシグナル検出手法の妥当性評価、シグナル検出手法の高度化、およびデータマイニング手法導入後の業務フローと支援システムのあり方についての検討を行いました。

1. シグナル検出手法の妥当性評価

（1）感度分析・特異度分析

シグナル検出手法が検出するシグナルの妥当性の評価を行うために、感度分析と特異度分析を行いました。感度とは、検出されるべきシグナルを検出する割合、特異度とは検出されるべきでないシグナルを検出しない割合のことを言います。シグナル検出結果を利用する立場からは、感度、特異度ともに高いことが望まれます。

評価対象のシグナル検出手法は、平成17年度までに検討してきた6手法とし、シグナル検出には、平成17年度と同様PMDAの副作用報告データベースに納められているデータを用いました。また、評価を進める中で、シグナル検出に用いるデータベース、度数のカウント方法、シグナル検出基準値等、シグナル検出に関わる設定条件についての検討を行いました。

評価実験を行った結果、6手法の何れについても感度、特異度の観点から、一定の妥当性を有することがわかりました。ただし、感度、特異度については、更なる向上が望まれるものであり、今後も検討を進めて参ります。

また、シグナル検出に関わる設定条件については、評価実験の結果に基づき当面は以下のように設定することとしました。

- シグナル検出を行う副作用報告データ

シグナル検出に用いる副作用報告データ（データベース）は、感度を重視して新データベース（現行の様式で報告されたデータ）と旧データベース（以前の様式で報告されたデータ）の両方を用いることとしました。

- 度数のカウント方法

度数のカウント方法については、医薬品と副作用の組合せ数を基本とする方法と症例数を基本とする方法の2通りで実験を行い比較しましたが、優劣は付け難い結果となりました。そこで諸外国規制当局が採用している方法に倣って、症例数を基本とする方法で度数のカウントを行うこととしました。

- シグナル検出基準値

シグナル検出基準を変更することにより、感度、特異度も変化します。また、感度、特異度は二律背反の関係があり、最適なバランスはシグナル検出結果を利用する分析者が決定する必要があります。そこで、導入当初は標準的なシグナル検出基準を用い、随時、分析者の意見を採りいれながら調整することとしました。

(2) シグナル指標値の相関評価

これまで6つのシグナル検出手法をとりあげ、手法ごとにシグナル検出を行ってきました。一方、各シグナル検出手法で算出されるシグナル指標値は、互いに定義が類似していることから、そのシグナル指標値に相関があると考えられます。

シグナル指標値に相関がある場合には、シグナル検出結果も類似するため、複数のシグナル検出手法間でシグナル検出結果が類似しているのであれば、まずは何れかの結果に注目すればよいことになります。そこでシグナル指標値の相関関係について、Thiessard らの研究[1]を参考に評価を行いました。その結果、6手法のシグナル検出結果の扱いについて、次のような結論を得ました。

- PRR、ROR、Yule's Q の3つの手法については、シグナル検出結果が類似しており、何れも早期かつ多数のシグナルを検出させるという特徴があることが確認されました。このうち、ROR は、オランダの規制当局(Lareb)が既に導入しており、また、Rothman ら[2]によると、PRR に比べて自発報告データを対象とした指標として適した特性を持つことから、特に注目する指標としました。
- GPS、BCPNN は、報告数が少ない場合にそれぞれ独自のシグナル検出結果を出す可能性があることが確認されました。この特徴を重視し、GPS、BCPNN の何れについても注目することとしました。
- SPRTについては、他手法に比べて特徴的なシグナルを検出する場合もあるが、検出されるシグナルが少なく、各国規制当局で採用された実績もないため、当面は参考として位置づけることにしました。
- 今回特に注目する ROR, GPS, BCPNN 以外の手法についても、当分の間は平行して計算を実施し、実運用の中でそれらの性能の確認をあらためて行うこととしました。

2. シグナル検出手法の高度化検討

シグナル検出手法の高度化を目的として、主に層別分析、相互作用分析、副作用グルーピングの3つについて検討を行いました。

(1) 層別分析

分析の対象とする層として性別と年齢を設定し検討を実施しました。GPS については層別後の調整を含めた計算方法が示されているため、これを用いてシグナル検出を行いました。また、その他の手法については層ごとに抽出されたデータに対してシグナル検出を実施し、シグナルの有無、特徴等を確認しました。

評価実験を行った結果、ある特定の性別、年齢層など層別でのみ検出されるシグナルが確認されました。これらのシグナルは、詳細な分析を行うきっかけとなる可能性があります。

(2) 相互作用分析（併用薬分析）

一般的に処方される薬は複数あり、それらは併用して服用されています。副作用を検討する上では、このような併用薬による相互作用の影響も考慮しなくてはなりません。そこで、併用薬による相互作用の影響を考慮したシグナル検出手法について検討を行いました。

併用薬分析ではMGPSにより2種類の医薬品と1種類の副作用の3アイテムに対するシグナル検出実験を行いました。添付文書の相互作用に追記された事例から評価組合せを抽出し、それらについて検討を試みましたが、現状では該当する組合せの報告件数が少なく、評価組合せをシグナル検出指標値から検出することは困難であることが分かりました。しかし、MGPSは閾値を設けてシグナルを検出することを目的としたものではないことから、他の組合せの指標値との相対的な関係を考慮しながら、その指標値の持つ意味を検討していきたいと考えています。

(3) 副作用グルーピングの影響

シグナル検出計算における副作用名のコーディングにはMedDRA/PTを用いています。多くの諸外国の規制当局においても副作用名のコーディングにMedDRA/PTを用いていますが、シグナル検出の単位としては細分化されすぎていることがあります。

シグナル検出に適したレベルで副作用をコーディングすることを目的として、平成17年度の検討においてMedDRAの体系の1つであるSMQを用いて、MedDRA/PTの複数の副作用を1つの副作用と見なしてシグナル検出を行う副作用グルーピングを試行しました[3]。平成18年度は、添付文書の改訂根拠に用いられた症例に記載されている全PTをシグナル検出上の1つの副作用としてグルーピングする方法を試行しました。

評価用データに対する感度分析を実施しましたが、副作用グルーピングの有無による有効性は確認できませんでした。しかし、実際に人が分析を行う際には複数の PT を同じ副作用と判断しており、これを機械的な計算にも取り入れることができれば、実際の分析業務での判断により近付くことでシグナル検出の精度が向上する可能性があることから、引き続き検討を行っていきたいと考えています。

3. インパクト分析手法の試行

英国 MHRA では、主に検出されたシグナルの対応優先順序づけ（フィルタリング）を目的として、インパクト分析を行っています。平成 17 年度までの検討から、PMDA 保有のデータベースを用いてシグナル検出を行った場合にも、多くのシグナルが検出されることが確認されており、シグナル検出結果を業務で活用することを考えると何らかの優先順序付け（フィルタリング）を行うことが必要です。

平成 18 年度はシグナルのフィルタリング方法導入の初期検討として、英国 MHRA における実施内容を参考にインパクト分析を試行しました。MHRA で用いられている評価基準の一部を適用できなかったものの、本試行において一定の優先順位付けが可能であったことから、今後、インパクト分析手法がシグナルのフィルタリングとして機能するよう、評点付けや評価基準の見直しを行うなど、精度の向上を図っていきます。

4. データマイニング手法導入後の業務フロー検討

平成 17 年度に引き続き、データマイニング手法導入後の安全対策業務フローについて検討しました。平成 18 年度は特にシグナル検出結果など分析に資する情報の整理と分析者への提示方法について整理しました。

また、データマイニング手法導入後、シグナル検出システムの運用中に生じる課題とその改善策を整理しました。副作用報告データやシステムを利用する分析者の経験は常に蓄積、変化していくことから、今後も定期的にシグナル検出結果の評価や分析者に対するヒアリングを実施し、システムに反映するものとししました。

5. 手法の確定及び今後の予定

今回の検討結果から、PMDA が行うデータマイニング手法には、主として ROR, GPS, BCPNN の 3 つの手法を組み合わせで行うこととし、さらにインパクト分析を導入することとしました。

今後は、平成 16 年度に策定したスケジュール（図 1）に従い、平成 18 年度までの検討結果に基づき、平成 19 年度から業務システムの開発を行います。

また、今回確定したシグナル検出手法に関しても、業務での試適用等を行いながら、更なる高度化のための検討を進めていきます。

7. 検討委員会

本調査検討の遂行にあたっては、以下の委員からなる「データマイニングに関する検討委員会」を設置し、調査内容、およびシグナル検出実験の結果について検討を頂きました。

【データマイニングに関する検討委員会】

座 長：藤田 利治 統計数理研究所 データ科学研究系 教授
委 員：岩崎 学 成蹊大学 工学部 経営・情報工学科 教授
岡田 孝 関西学院大学 理工学部 教授
岡田 美保子 川崎医療福祉大学 医療情報学科 教授
酒井 弘憲 日本製薬工業協会 医薬品評価委員会 統計・DM 部会 副部長
櫻井 靖郎 財団法人 日本公定書協会 JMO 事業部 事業部長
藤原 昭雄 日本製薬団体連合会 安全性委員会 委員
望月 眞弓 北里大学 薬学部 臨床薬学研究センター 医薬品情報部門 教授
鷺尾 隆 大阪大学 産業科学研究所 教授

注) 平成18年度における調査結果の詳細については、別途報告書として医薬品医療機器情報提供ホームページ等により公開致します。

略語一覧

規制機関

略語	正式名称	日本語名
PMDA	Pharmaceuticals and Medical Devices Agency	独立行政法人 医薬品医療機器総合機構
MHRA	Medicines and Healthcare products Regulatory Agency	英国医薬品庁

シグナル検出手法

略語	正式名称	日本語名
BCPNN	Bayesian Confidence Propagation Neural Network	—
PRR	Proportional Reporting Ratios	—
GPS	Gamma-Poisson Shrinker	—
MGPS	Multi-item Gamma Poisson Shrinker	—
ROR	Reporting Odds Ratio	—
SPRT	Sequential Probability Ratio Tests	—

MedDRA用語

略語	正式名称	日本語名
MedDRA	Medical Dictionary for Regulatory Activities	ICH 国際医薬用語集
PT	Preferred Terms	基本語
SMQ	Standardized MedDRA Queries	MedDRA 標準検索式

参考資料

- [1] F. Thiessard et al., Comparaison de méthodes de génération automatique de signal pour la détection des effets indésirables des médicaments, Journées Francophones d'Informatique Médicale, Lille 12-13 mai 2005, (2005).
- [2] K. J. Rothman et al., The reporting odds ratio and its advantages over the proportional reporting ratio, *Pharmacoepidemiology and Drug Safety*, Vol. 13, pp.519-523, (2004).
- [3] データマイニング手法の検討を行うための支援業務報告書, 三菱総合研究所, 平成 18 年.

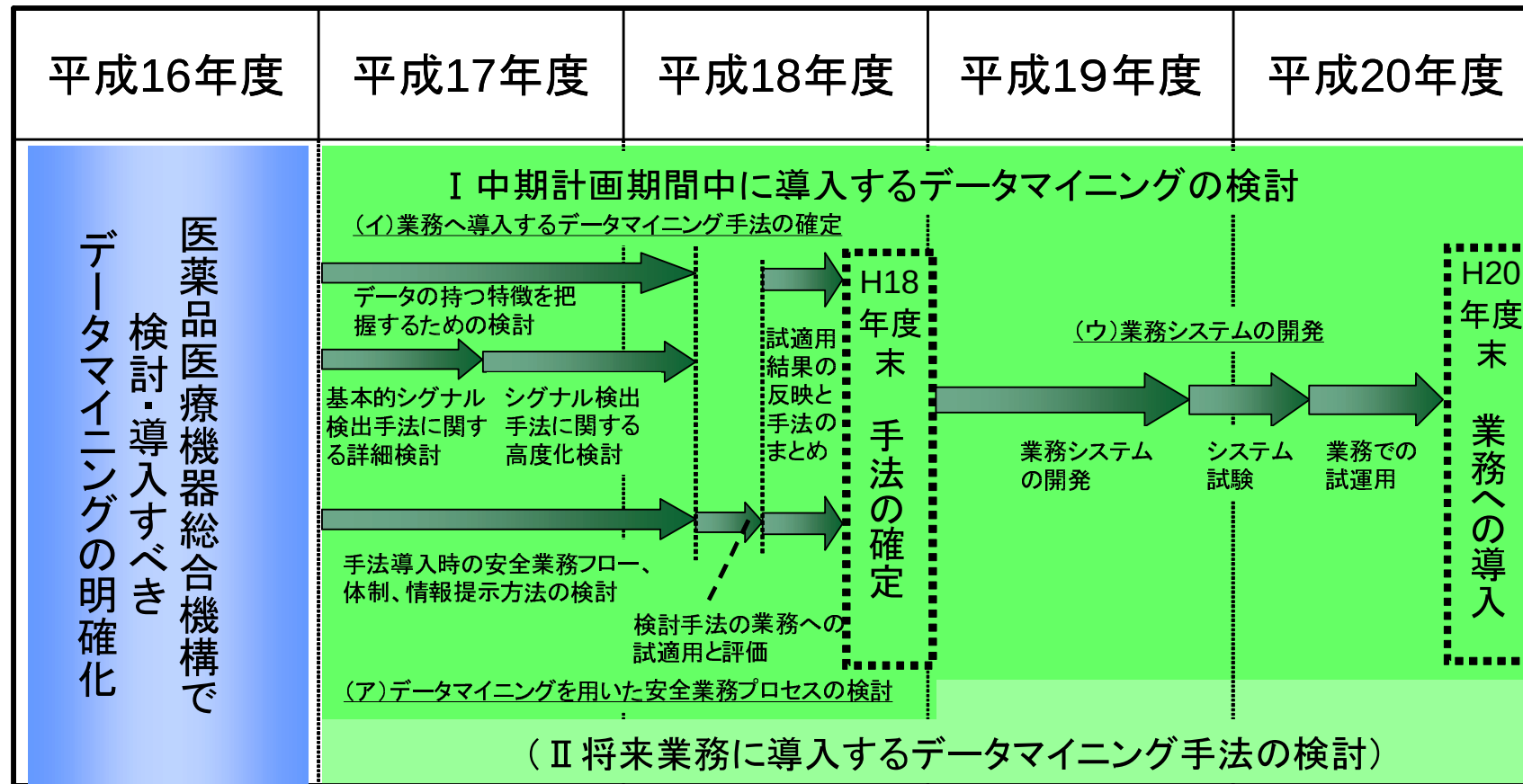


図1 中期計画期間中のスケジュール