

平成20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21年3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います。

平成16年度に策定したスケジュール（図1参照）にしたがい、中期計画期間の最終年度である平成20年度は、データマイニング手法を用いた業務フロー（新業務フロー）の業務量評価と、データマイニング手法の高度化検討として、Ad Hoc 解析手法、重複報告検出手法の検討を行いました。また、業務支援システムの開発を進め、平成21年度よりデータマイニング手法を導入した安全対策業務とする準備が整いました。

1. 新業務フロー試行実験

平成21年度より導入する安全対策業務へデータマイニング手法（シグナル検出手法）を導入した新業務フローの実用性を検証するため、試行実験を実施しました。

未知報告に関する新業務フローでは、詳細な検討を行うコンビネーション（医薬品と副作用の組み合わせ）のスクリーニング基準の1つとして、シグナル検出手法とその指標値を使用します。今回の試行実験では、シグナル検出を含むスクリーニングの各通過基準値と使用するシグナル検出手法を変えながら、スクリーニングを通過するコンビネーション数や内容について確認を行いました。試行実験に用いる評価用データには、評価を実際に即したものとするため、過去のある1週間に到着した報告を使用しました。

シグナル検出手法以外の2つのスクリーニング基準について、基準値を9パターンに変化させて通過するコンビネーション数を算出しましたが、設定した9パターンではスクリーニングを通過する数に大きな差が無いことが確認されました。

また、シグナル検出手法として、ROR、BCPNN、GPSの各手法を適用した場合について通過するコンビネーション数を比較しましたが、RORを適用した場合に比べて、BCPNNを適用した場合は4～5割、GPSを適用した場合は3～4割と、それぞれ半数以下でした。また、いずれの手法でもRORで通過したコンビネーションにほぼ含まれることが確認されました。RORを用いた場合に1次スクリーニングを通過する数はコンビネーション単位で300弱であり、今後予定されている安全対策担当職員の増員を考慮すると、対応可能な業務量であることが確認されました。

以上の試行実験の結果から、新業務フローの1次スクリーニングには、主としてRORを

用いて、BCPNN と GPS は参考値として算出、提示するものとしました。

2. Ad Hoc 解析の検討

非定型業務におけるデータマイニング手法を活用のひとつとして、Ad Hoc 解析について昨年度に引き続き、検討、試行を実施しました。Ad Hoc 解析では、副作用報告データベースのうち、特定の医薬品、もしくは副作用に関連する部分のみのデータ（スモールセット）に範囲を絞り込み、このデータを対象としてシグナル検出指標値を算出するものです。Ad Hoc 解析を行うことにより、特定医薬品による特定副作用の発現リスクの大きさに関する参考指標を得ることが期待できます。

検討結果から、シグナル指標値の順位表を見る際には、シグナル指標値が副作用報告傾向により強い影響を受けることを考慮する必要があることが示唆されました。また、データベース全体での報告件数と併せて、特定の医薬品や副作用に限定したスモールセットのデータでの評価を行うことで、注目した副作用のリスクの大きさをとらえるための参考指標として利用可能な知見が得られることがわかりました。

Ad Hoc 解析のような非定型業務については、経験の積み重ねが重要であることから、今後も引き続き試行を行う予定です。

3. 重複報告検出手法の検討

副作用報告の中には、同じ症例について複数の経路（製薬企業等）から報告されたものが含まれる場合があります。このような報告を重複報告とよびます。WHO-UMC や EMEA など海外の規制当局では、このような重複報告を自動的に検出する数理的な手法の検討が行われています。今回、PMDA においてもデータマイニング高度化の1つとして、この重複報告検出手法の検討を行いました。

まず、WHO-UMC が提案している手法を国内の副作用報告データベースに適用するところから着手しました。WHO-UMC の手法は、ある注目している報告（基準報告）と重複を確認する対象の報告（対象報告）について、それぞれの報告中の複数の項目の内容から重複報告であることが疑われる度合いを総合スコアとして算出するものです。総合スコアは、あらかじめ総合スコアを算出するための数理モデルを導出しておき、このモデルにより計算されます。実際に、PMDA の副作用報告データベースに適用したところ、重複報告には高いスコア値が算出され、国内の副作用報告データベースでも重複報告の検出に有効であることが確認されました。

さらに、検出精度を高めるために、WHO-UMC 手法で算出される報告中の各項目のスコアから重複か否かを判定する重複報告判別モデルを導出し、これを用いて重複報告の判定を行う手法について検討を行いました。WHO-UMC 手法では、各項目のスコア値の総和を総合スコアとして用いるのに対して、検討する判別モデルでは最適化された重み和にもとづいて判別するため、検出精度が高まることが期待できます。評価実験を行った結果、

WHO-UMC の提案手法よりも高い精度で判別できることが確認されました。

今後は、重複報告判別モデルの導出に用いる事例の構成を変化させることにより、さらに精度の改善を図ったり、実際の業務への導入方法や手法の利用方法について検討を行う予定です。

4. 業務支援システムの開発

昨年度（平成19年度）に引き続き、業務支援システムの開発を行い、完成しました。この業務支援システムは、データマイニングが導入された新しい業務フロー（新業務フロー）に対応しており、シグナル値を含む分析に必要な情報が参照しやすいものとなっています。

5. 中期計画期間のまとめ

平成16年度から平成20年度までの中期計画期間に、基本的シグナル検出手法を中心とした業務への導入検討を行い、シグナル検出手法を導入した新しい安全対策業務フローを構築し、これを支援する業務支援システムの開発、導入を行いました。

平成21年度より、データマイニング手法（シグナル検出手法）が導入された新しい業務フローに沿って安全対策業務が実施される予定となっています。

6. 今後の予定

平成21年度以後は、導入した安全対策業務の適切な運用と業務の高度化に向けて、以下のような検討を実施していく予定です。

- スクリーニング基準等の見直し

今後予想される分析要員の変更等の体制の変化、副作用報告数の変化などの安全対策業務の外部環境の変化に対応して、定期的にシグナル検出基準を含むスクリーニング基準について継続的に評価、見直しを行います。

- 新たなデータマイニング手法の業務への導入検討

安全対策業務をさらに効率化、高度化するため、新たなデータマイニング手法の導入について検討を行います。検討内容については、以下のようなものを予定しています。

- － 層別・併用薬分析

評価実験を試行した結果、参考指標として有効であることが確認されています。今後は、業務での利用方法を想定し、利用する場合の留意点等について、試行を重ねる中で検討していく予定です。

- － Ad Hoc 解析

これまでの試行から、有効な情報が得られるという結果が得られています。

試行を重ねることにより、業務で使用する際の留意点の抽出、および業務での具体的な利用方法、さらには業務支援システムへの組み込み等について検討していく予定です。

－ 重複報告検出

業務への導入のための検討として、さらに検出手法の高精度化のための検討と、業務での利用方法を考慮した試行実験を実施していきます。

- データ整備・拡充

今回整備に着手した未知・既知データ、案件（グルーピング）情報等について、業務を通じてさらに蓄積を進めていきます。

- 検討委員会の設置

上記の検討を推進するため、次年度以後も検討委員会を設置し、開催します。

7. 調査報告書の公開

平成20年度に、外部シンクタンクに委託した調査検討の結果は、報告書として別途公開致します。

8. 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、調査結果の報告内容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】¹

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 理工学部 情報科学科 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 副部会長
	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 副委員長
	望月 眞弓	慶應義塾大学 薬学部 教授
	鷺尾 隆	大阪大学 産業科学研究所 教授

¹ 所属は2009年3月時点

略語一覧

規制機関名称

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
EMA	European Medicines Agency	欧州医薬品庁
WHO	World Health Organization	世界保健機関
UMC	Uppsala Monitoring Center	WHO 医薬品副作用情報収集センター

シグナル検出手法

略語	正式名称	日本語名
BCPNN	Bayesian Confidence Propagation Neural Network	—
GPS	Gamma-Poisson Shrinker	—
ROR	Reporting Odds Ratio	—

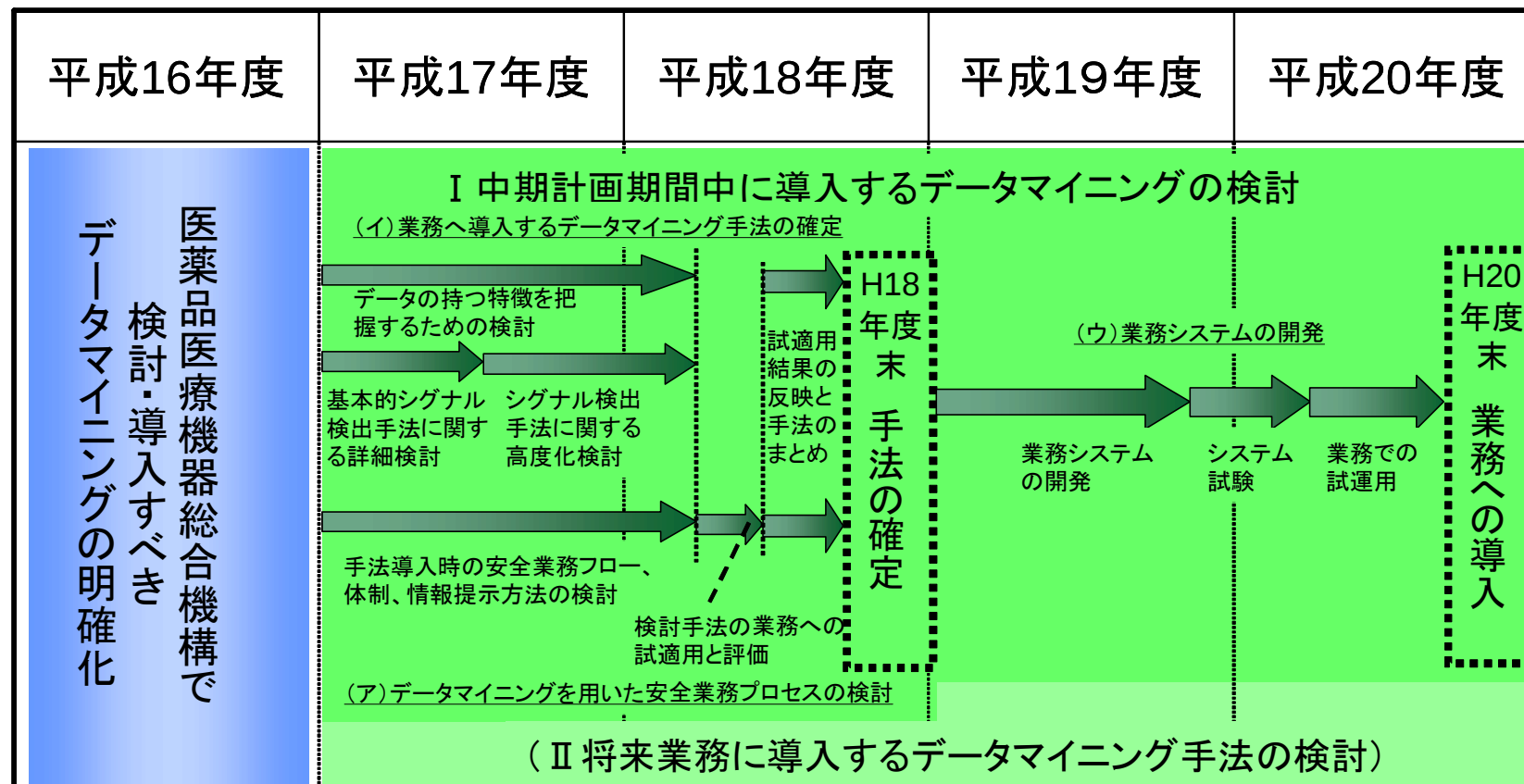


図1 中期計画期間中のスケジュール