

## データマイニング手法の導入に関する検討内容について

平成17年4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年度までに手法を確立し、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年度は、中期計画期間中に導入するデータマイニング手法と導入に関わるスケジュールについて検討を行いました。検討にあたっては、以下の委員により構成される「データマイニングに関する検討委員会」を設置し、検討内容についてご審議を頂きました。

### 【データマイニングに関する検討委員会】

座長	藤田 利治	国立保健医療科学院 疫学部 疫学情報室 室長
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
	野口 茂	日本製薬団体連合会 安全性委員会 委員
	平山 佳伸	厚生労働省 医薬食品局安全対策課 課長
	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
	鷲尾 隆	大阪大学 産業科学研究所 助教授
	渡邊 伸一	厚生労働省 医薬食品局安全対策課 課長補佐

## 1. 中期計画期間中に導入する新規手法(データマイニング手法)

### (1) 新規手法(データマイニング手法)

近年、データから業務に有用な知識や情報を発掘するデータマイニング手法の研究が進められ、様々な分野で業務への導入が進められています。

市販後医薬品の安全対策業務の分野では、データマイニング手法の1つとして、自発報告データにおいて特定の医薬品と副作用の組に関する報告数が他の医薬品や副作用に比べて有意に多い場合に、その組を安全性について対策が必要である可能性があるものとして抽出するシグナル検出手法の研究開発が行われており、諸外国の規制当局では業務への導入が進められています。この自発報告データ<sup>1</sup>をもとに安全対策業務の観点から注目すべき医薬品と副作用の組をシグナルとして抽出するシグナル検出手法を特に基本的シグナル検出手法とよびます。

自発報告データから、安全対策業務の観点から注目すべき医薬品と副作用を抽出する業務は現在も安全部担当者が実施していますが、シグナル検出手法を導入することにより、従来よりもさらに早期に抽出できるようになり、安全対策の向上に資するものとなります。

安全対策業務に必要な情報という観点からは、基本的シグナル検出手法により抽出される医薬品と副作用のシグナル以外にも、安全対策業務に資するものとして、次の情報の抽出が期待されます。

#### 1. 層別シグナル

「性別」や高齢者、小児といった「年齢」など層別のシグナル情報。安全対策業務では、これら層別に注目した分析を行っています。

#### 2. 併用薬シグナル

併用薬の使用状況を考慮したシグナル情報。安全対策業務では、実際に併用薬を念頭に置いた分析を行っています。

### (2) 中期計画期間中に導入するデータマイニング手法

中期計画期間中に導入するデータマイニング手法は、「基本的シグナル検出手法を中心として安全対策業務の向上に資するように高度化したもの」を対象とします。基本的シグナル検出手法からの高度化の内容については、以下で挙げる高度化の候補内容について技術的検討を通じて、中期計画期間中の導入の可否の判断を行います。

---

<sup>1</sup> ここでの自発報告データとは、主に企業から医薬品医療機器総合機構に報告された個別の副作用症例報告のことをさします。

#### (ア) データの持つ特性把握のための検討

シグナルをはじめとする有益な情報を得るためには、データマイニング手法を適用する前にデータの持つ特性を十分に分析、理解することが必要となります。これに関わる内容として以下の検討を実施します。

- データの特性に関する分析：報告者毎に報告の記述やコーディングに異なる傾向が見られる、あるいは市販後の時期に関係して報告数が異なる等の理由により、自発報告データは一定の傾向、特性を含みます。データマイニングを適用し、結果を理解する上では、この特性を十分に理解することが必須となります。そこでデータの持つ特性とこれが生じする背景を理解するために、データの分析を行いません。
- 背景知識の利用検討：専門家や分析担当者は、上記のようなデータの持つ特性をはじめとしてデータを正しく解釈するのに必要な知識（背景知識）を用いています。データマイニングのプロセスに、これらの背景知識を利用する方法について検討を行います。
- データの前処理に関する検討：データマイニングの結果の質を向上させるために、データの誤り、欠損値、およびデータの持つ特性に対処する方法について検討を行います。
- データベース構成に関する検討：シグナル検出を中心としたデータマイニングに必要なデータのデータベースでの管理方法について検討を行います。

#### (イ) 基本的シグナル検出手法に関する詳細検討

基本的シグナル検出手法とは、自発報告データに基づき、安全対策業務の観点から医薬品・副作用の組を抽出する手法です。基本的シグナル検出手法については、既に研究開発された既存手法<sup>2</sup>がありますので、これらを中心に業務への導入に必要な詳細検討を行います。

詳細検討で実施する内容としては、以下のようなものがあげられます。

##### (a) 手法に関する検討

- 適用手法(アルゴリズム)の選択：海外の規制当局で利用されている PRR、BCPNN、EBM など既存手法<sup>2</sup>について、PMDA が保有する自発報告データを用いてシグナル検出の試行、および評価を行います。
- 検出基準の設定：シグナル検出を行う際の検出基準（閾値）の設定を行います。

##### (b) データに関する検討

- 医薬品、副作用に関する集計単位：医薬品、および副作用についてシグナ

---

<sup>2</sup> 既存手法の詳細については、資料[1]、[2]をご参照下さい。

ル検出のために報告数を集計する際の集計単位を検討します。特に副作用の集計単位については、MedDRA (医薬規制用語集)<sup>3</sup>における用語階層レベルや SMQ (Standardized MedDRA Queries, MedDRA 標準検索式)<sup>4</sup>の利用等について検討します。

- データの利用期間：シグナル検出に利用する報告データの期間の検討を行います。
- 既知・未知データの利用：添付文書の副作用に関する項目への反映済（既知）未反映（未知）のデータを利用するシグナル検出手法の検討を行います

#### (ウ)シグナル検出手法に関する高度化検討

基本的シグナル検出手法を安全対策業務の向上に資するものとして高度化する方向としては、以下の2つに分類されます。

- 抽出する知識・情報（シグナル）の内容を高度化する
- 抽出する知識・情報（シグナル）の精度を高度化する

内容の高度化については、層別シグナル抽出、併用薬シグナルの抽出を候補とします。

抽出する知識・情報の精度向上については、分析に利用する手法（アルゴリズム）分析対象データについて以下のような検討を行います。

##### (a) 手法（アルゴリズム）

- 複数手法・基準の利用：医薬品群等ごとに異なるシグナル検出基準、手法を利用することについて検討します。
- 複数手法の併用：複数手法によりシグナル検出を行い、さらに各手法によるシグナル検出結果に基づいて、最終的なシグナル検出を行う手法を検討します。機械学習手法の分野で注目されている Boosting 手法、Bagging 手法などアンサンブル学習法<sup>5</sup>の適用の検討を含みます。

##### (b) 分析対象データ

- 重み付けデータの利用：症例報告の情報量に基づいて付与された重み付けデータを利用するシグナル検出手法について検討を行います。

---

<sup>3</sup> MedDRA の詳細については、資料[3]をご参照下さい。

<sup>4</sup> SMQ の詳細については、資料[4]をご参照下さい。

<sup>5</sup> Boosting 手法、Bagging 手法などアンサンブル学習法の詳細については、資料[5]をご参照ください。

- 医薬品の流通量に関するデータの利用：処方情報など各医薬品の流通量に関するデータの利用についても実現可能性の検討を行います。

### (3) 今後の導入を目指して検討を行うデータマイニング手法

中期計画期間中の安全対策業務への導入は困難だと判断されますが、今後の導入を目指して検討を行うべきであるデータマイニング手法として以下のようなものが挙げられます。

- 症例報告票記述データの分析：症例報告票に記述されている情報のうち、コード化されていない記述（テキスト）データをシグナル検出手法に利用するための方法について検討を行います。
- 化学構造・薬理作用・薬物動態データ（知識）ベースの適用：分析担当者が分析時に用いている当該知識について知識ベース化し、これを分析に利用するシグナル検出手法について検討を行います。

### (4) 自発報告データの質および量の向上のための提言

シグナル検出をはじめとするデータマイニング手法を適用することに限らず、安全対策業務の高度化のためには、自発報告データの質および量の更なる向上が期待されます。これを実現するため、データマイニング手法の検討を踏まえて製薬企業・医療機関に対して必要な働きかけを行っていきます。

## 2. 業務導入までのスケジュール

### (1) 実施項目

平成18年度までの手法確立、平成20年度までの業務への導入に向けて、検討会を設け専門家の意見を頂きつつ、以下の項目について実施します（付表参照）。

【平成17年度～平成18年度】

平成18年度末までの手法の確立を目的として以下の検討を行います。

#### (ア) データマイニングを用いた安全業務プロセスの検討

今回導入するデータマイニング手法（シグナル検出手法）が提供する情報に基づき安全業務プロセスの信頼性向上を実現するための具体的な方法について検討を行います。検討する内容には以下の点を含みます。

- 検出されたシグナルをはじめとする情報に基づく安全対策業務上の判断基準、業務フローの検討
- 次項（イ）で検討されたデータマイニング手法の業務への試適用と評価

- 検出されたシグナル、およびシグナル検出のために算出された指標などの情報の分析担当者への提示インタフェース、提供リストに含むべき情報項目の検討
- データマイニング手法の運用を含む人員体制

#### (イ) 業務に導入するデータマイニング手法の確定

業務へ導入する「基本的シグナル検出手法を中心として安全対策業務の向上に資するように高度化した」データマイニング手法を確定するために、以下の検討を行います。

- データの持つ特徴を把握するための検討
- 基本的シグナル検出手法に関する詳細検討
- シグナル検出手法に関する高度化検討

これらの検討を行うことにより、平成18年度末には、手法の確立を完了し、業務への仮導入が可能な状況とします。

【平成19年度～平成20年度】

#### (ウ)業務システムの開発

平成18年度末までに確立されたデータマイニング手法を実際に業務に導入するための業務システムの開発と業務の試運用を平成20年度末までに実施し、業務への導入を完了します。

#### (2)実施状況の公表

実施状況については適宜公表します。

#### 参考資料

- [1] 藤田利治他，医薬品の副作用自発報告によるシグナル検出の実用化に向けての検討，厚生労働科学研究補助金（医薬品等医療技術リスク評価研究事業）分担研究報告書，(2004).
- [2] 渡邊裕之他，重要な安全性情報を早期に検出する仕組み - シグナル検出の最近の手法について - ，計量生物学，Vol.25，No.1，37-60，(2004).
- [3] （財）日本公定書協会 JMO 事業部，医薬規制用語集(MedDRA)バージョン 8.0 手引書，(2005).
- [4] CIOMS，JMO 事業部（翻訳），MedDRA 標準検索式の開発と適正使用について (CIOMS-WG 報告書)，(2004).
- [5] 上田修功，アンサンブル学習，計測と制御，Vol.41，No.3，248，(2002).

付表: 中期計画期間中のスケジュール

