

データマイニング手法の導入に関する 検討結果報告書

平成 21 年 3 月

独立行政法人 医薬品医療機器総合機構

目次

(I) 平成 21 年度調査報告	1
1. 調査概要.....	1
1.1 調査目的.....	1
1.2 調査内容.....	1
1.3 検討委員会の設置	3
2. 新業務フロー試行実験の実施	5
2.1 新業務フロー.....	5
2.2 試行実験.....	7
2.2.1 目的.....	7
2.2.2 実験方法.....	7
2.2.3 実験結果.....	9
2.3 新業務フロー試行実験のまとめ	13
3. Ad Hoc 解析の検討.....	14
3.1 Ad Hoc 解析.....	14
3.2 ケーススタディ	14
3.3 Ad Hoc 解析のまとめ	18
4. 重複報告検出手法の検討	19
4.1 背景と目的	19
4.2 重複報告類似スコアによる評価	19
4.2.1 類似スコアによる重複報告の検出	19
4.2.2 評価実験.....	27
4.2.3 実験結果.....	29
4.3 各属性の類似スコアにもとづく重複報告判別モデル導出の検討	37
4.3.1 背景と目的	37
4.3.2 重複報告判別モデルの導出	37
4.3.3 評価実験.....	39
4.3.4 実験結果.....	40
4.4 今後の検討課題.....	42
5. 調査のまとめ.....	44

(Ⅱ) 中期計画期間のまとめと今後の検討方針	46
1. 中期計画期間のまとめ	46
1.1 年度別調査・検討内容のまとめ	46
1.2 調査・検討内容のまとめ	48
2. 今後の検討方針.....	52
参考文献	54

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
Lareb	Netherland Pharmacovigilance Centre	オランダ薬剤監視センター
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
MHRA	Medicines and Healthcare products Regulatory Agency	英国医薬品庁
FDA	Food and Drug Administration	米国食品医薬品局

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 Shriner	—
MGPS	Multi-item Gamma Poisson Shriner	—
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
HLT	High Level Terms	高位語
SMQ	Standardized MedDRA Queries	MedDRA 標準検索式

その他

略語	正式名称	日本語名
SMO	Sequential Minimal Optimization	—

(I) 平成 20 年度調査報告

1. 調査概要

1.1 調査目的

独立行政法人 医薬品医療機器総合機構(PMDA)の中期計画(平成 16 年度～平成 20 年度)では、安全対策業務に資するデータマイニング手法を導入すると定められている。

この中期計画を実現するため、平成 16 年度に中期計画期間内に導入するデータマイニング手法の概要と中期計画期間内の実施スケジュールの策定を行った。また平成 17 年度は、業務に導入するデータマイニング手法を確定するために、主にデータマイニング手法導入後の業務フローの検討、および基本的シグナル検出手法の詳細検討を行った。さらに、平成 18 年度は基本的シグナル検出手法の精度の向上と高度化の検討を行い、平成 19 年度は、データマイニング手法を導入した業務支援システムの開発を行うのに並行して、データマイニング手法を用いた安全対策業務の業務フローの検討、およびデータマイニング手法の高度化検討を行った。

平成 20 年度は、昨年度に引き続きデータマイニング手法の高度化検討と、データマイニング手法の安全対策業務への試適用、および中期計画期間の成果のとりまとめと今後の方針についての検討を行った。

中期計画期間中のスケジュールを図 1-1 に示す。

1.2 調査内容

本年度、調査および検討を行った項目は以下のとおりである。

(1) データマイニングの高度化検討

PMDA では、平成 20 年度末までに PMDA の安全対策業務で利用を開始するデータマイニングの機能を有する業務支援システムを開発中であり、これまでに検討したデータマイニング手法(基本的シグナル検出手法)を使用する。

今後、さらなる安全対策業務の効率化のためにデータマイニング手法の高度化を進めていくことを目的として、以下の項目について検討した。

① 重複報告検出手法の検討

シグナル検出手法により算出される指標値の信頼性を高めるために、既存手法を参考にした重複報告検出手法について検討した。

② Ad Hoc 解析の検討

薬剤群、副作用群のグルーピング手法やスモールセットによる分析手法について検討

を行い、これらを用いたシグナル検出計算の有効性について評価した。

③ 併用薬との相互作用分析の検討

併用薬との相互作用分析について、既存手法について調査を行い、分析プロセス、分析支援内容、分析支援手法について検討した。

④ 試適用システム用データマイニングソフトウェアの改修

三菱総合研究所が試適用システムのために開発したデータマイニングソフトウェアについて、上記で検討を行った手法等を用いた実験が行えるように、必要な改修を行った。

(2) データマイニングの安全対策業務への試適用検討支援

PMDA では、現在開発中のデータマイニング手法を使用した業務支援システムを導入後の安全対策業務フローを試適用し、業務効率性の観点からの評価と、業務フローの改善点の抽出等の検討を行っている。この検討に関連して、以下の項目を実施した。

① 試適用実施内容の検討

業務効率性の評価、およびシグナル指標値によるスクリーニングの評価を行うための試適用の実施内容について検討した。

② 試適用結果の分析

試適用を実施した結果として収集されたデータについて分析を行い、スクリーニングを含む業務フローの改善方針について検討した。

(3) データマイニング導入検討の成果とりまとめ支援

平成 20 年度に総合機構の中期計画として予定されていたデータマイニングの導入が完了するにあたり、その成果のとりまとめを行った。

① 中期計画期間中の成果評価

中期計画期間中に得られた成果の評価と今後の課題の整理を行った。

② 資料作成

中期計画期間中に得られた成果を資料としてとりまとめを行った。

1.3 検討委員会の設置

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いた。

【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
	鷲尾 隆	大阪大学 産業科学研究所 教授

本報告書は、平成 20 年度「データマイニング手法の検討を行うための支援業務」において株式会社三菱総合研究所が作成したものである。

¹ 所属は 2009 年 3 月時点

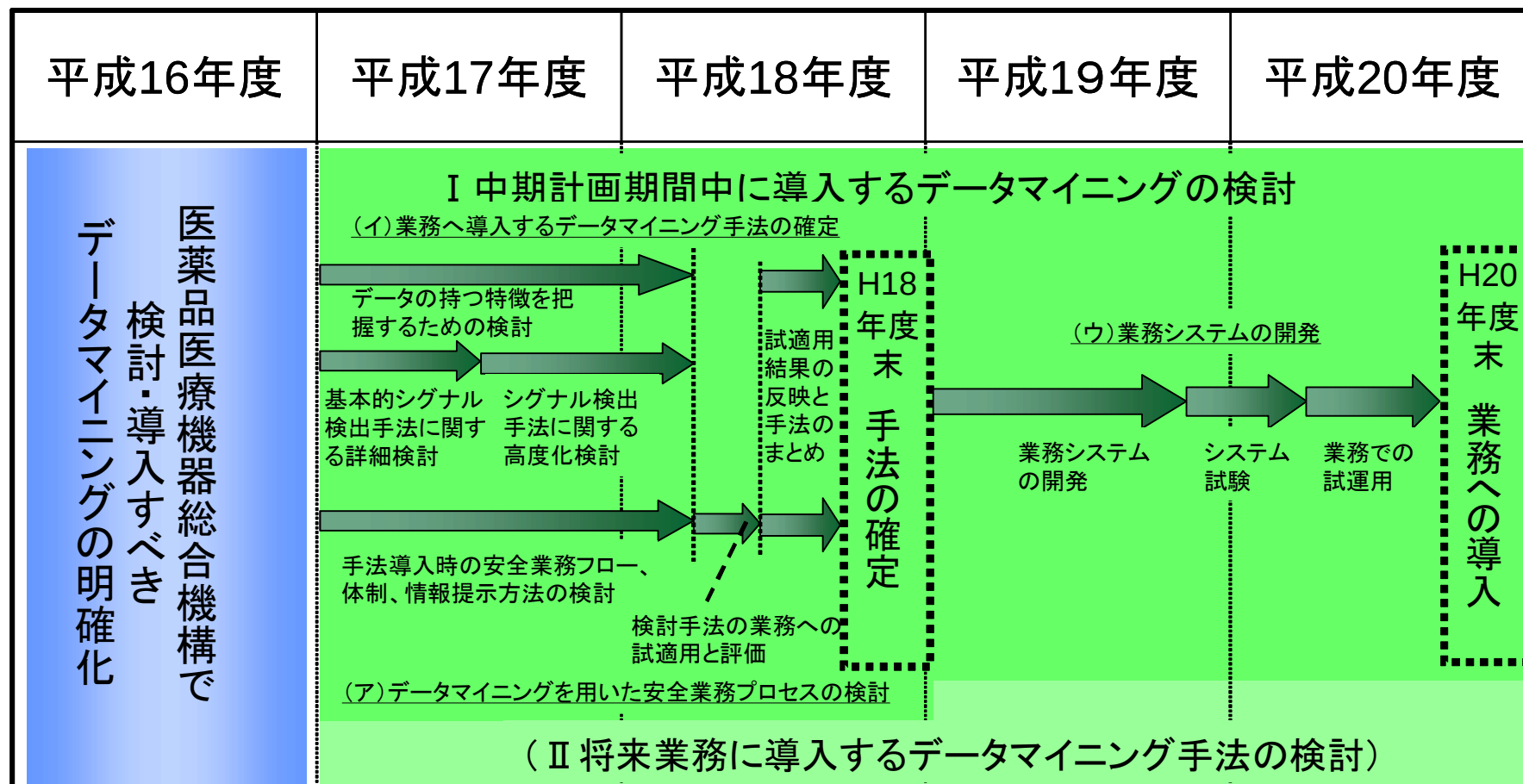


図 1-1 中期計画期間中のスケジュール

2. 新業務フロー試行実験の実施

2.1 新業務フロー

平成 21 年度からのシグナル検出手法の安全分析業務への導入に向け、平成 19 年度から手法を取り込んだ新たな業務フローの検討を進めている^[4]。

業務フローの検討、およびデータマイニングシステムの開発に伴い、安全対策業務における分析単位を整理し、改めて用語を定義した。これを表 2-1 に示す。分析単位はユニット、コンビネーション、案件の 3 種類があり、それぞれ安全対策業務の各段階で使われることになる。シグナル検出指標値の算出はコンビネーション単位で実行される。

表 2-1 安全対策業務における分析単位の定義

単位名	意味
ユニット	特定の症例内の被疑薬(再審査コード上 7 桁)と副作用名 (MedDRA PT) で構成される単位 <特定症例×医薬品 1×副作用 1>
コンビネーション	医薬品の一般名(再審査コード上 7 桁)に副作用名 (MedDRA PT) が 1 対 1 で対応して構成される単位 <医薬品 1×副作用 1>
案件	医薬品の安全対策等における単位であり、複数のコンビネーションをグループ化したもの <医薬品 M×副作用 N>
(参考) 症例	第 1 被疑薬を含む複数医薬品に対して複数副作用が記載されている

未知のユニットに対する新業務フローを図 2-1 に、業務フローとの分析単位の間係を図 2-2 に示す。なお、シグナル検出手法は主に未知のユニットに対して導入することを想定しているため、未知のユニットに対する業務フローのみを示している。

報告された症例には第 1 被疑薬を含む複数医薬品に対して、複数の副作用が記載されている。これをユニットに分解し、それぞれのユニットについて未知・既知を判定し、未知のユニットに対して因果関係評価を行う。因果関係評価結果はコンビネーション単位で蓄積する。蓄積された因果関係評価結果、およびシグナル検出指標値²をスクリーニング条件に用いて 1 次スクリーニングを機械的に実施し、1 次スクリーニングを通過したコンビネーションは案件作成・蓄積以降のプロセスへと進むことになる。

なお、次節以降に示す業務量の評価は、図 2-1 に示すように 1 次スクリーニング部分を対象とする。

² シグナル検出指標値の算出の際には、第 1 被疑薬を含むすべての被疑薬と副作用のコンビネーションを対象とする。

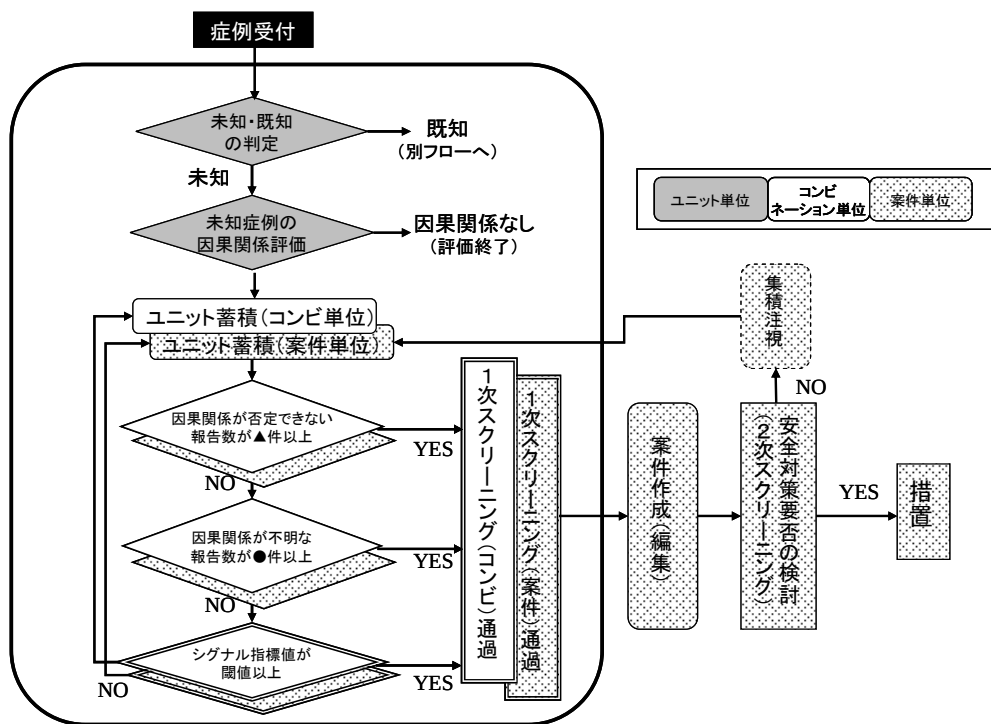


図 2-1 未知のユニットに対する新業務フロー

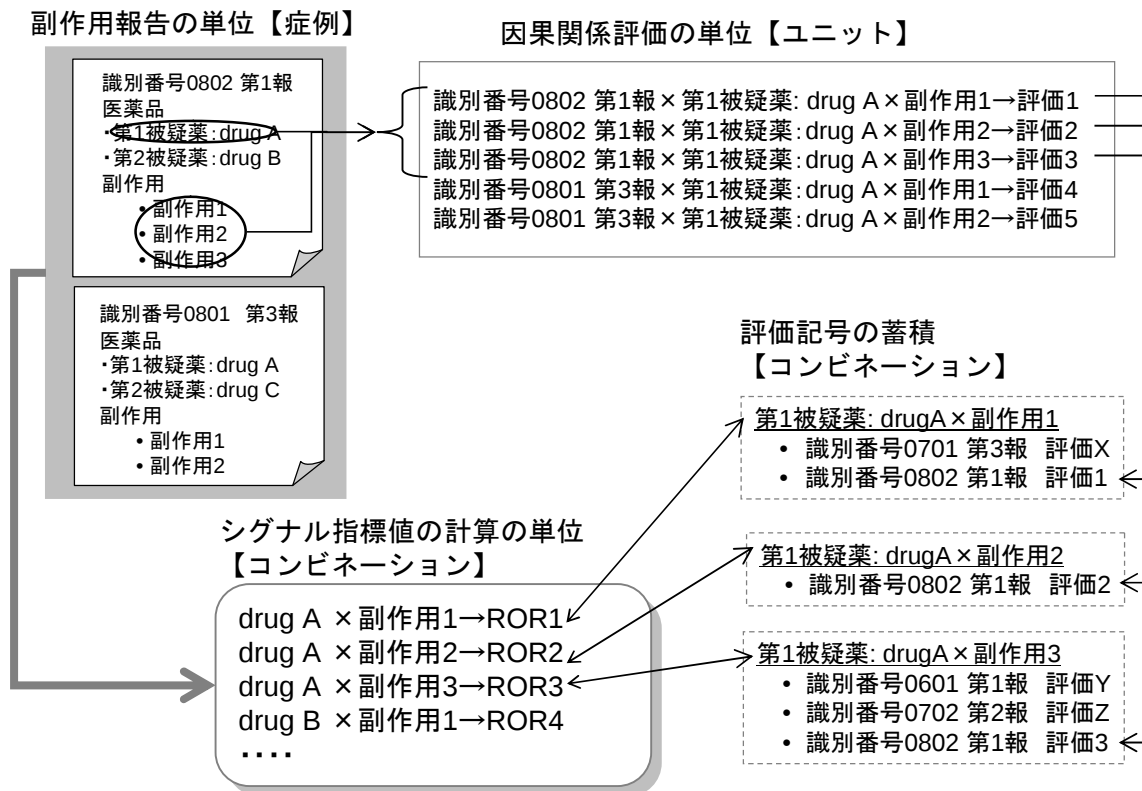


図 2-2 業務フローと分析単位の関係

2.2 試行実験

2.2.1 目的

新業務フローの試行実験は平成 19 年度の検討においても実施している^[4]。しかしながら、平成 19 年度の評価は簡易的なものであり、また新業務フローの検討も進行中の段階で行ったものであった。そこで、今年度はさらに検討を重ねた業務フローに対して、実際に平成 21 年度以降に導入される業務プロセスを再現した形でより詳細な評価を実施する。

新業務フローの試行実験は 1 次スクリーニングに対して実施する。試行実験により、1 次スクリーニングに関して以下の点を明らかにする。

- 1 次スクリーニングにおける機械的基準（件数、シグナル検出指標値）
1 次スクリーニングを通過したコンビネーション数について分析者の業務量を考慮した上で、現実的なスクリーニングの基準値を定める。
- シグナル検出計算に用いる手法（ROR、BCPNN、GPS の 3 手法から絞込み）
シグナル検出手法ごとに、1 次スクリーニングを通過したコンビネーションの種類の傾向を確認し、1 次スクリーニングに適した手法を選択する。

2.2.2 実験方法

試行実験はある 1 週間に実際に報告されたデータに対して実施する。表 2-2 に試行実験対象データの概要を示す。試行実験対象期間は 2008 年 7 月 28 日から 8 月 1 日の 5 日間であり、期間中に報告された症例数は 1,219 例（追加報も含めると計 1,230 件となる）、ユニット数で 1,846、コンビネーションとしては 1,547 であった。

表 2-2 試行実験対象データの概要

試行実験対象期間		2008 年 7 月 28 日（月）～8 月 1 日（金）（5 日間）
報告総数	症例数	1,219 例（1,230 件）
	ユニット数	1,846
	コンビネーション数	1,547

試行実験は以下の流れで実施した。

① 未知・既知判定

図 2-1 に示した通り、報告された症例に記載されているユニットごとに未知・既知の判定が必要となる。未知・既知データが蓄積されていれば、未知・既知の判定は機械的に行うことができるが、現段階では未知・既知データは整備されていないため、試行実験期間の全ユニットに対して未知・既知の判定を行った。

なお、平成 21 年度以降は、一度未知・既知判定を行ったコンビネーションについてはデータが蓄積されるため、徐々に未知・既知データが整備されることとなる。

② 因果関係評価データのクリーニング

1 次スクリーニングを行うためには、実験対象期間に報告された症例のうち、未知と判断されたコンビネーションに関しては、これまでの因果関係評価結果の情報が必要となる。しかし、現段階では因果関係評価結果はコンビネーション単位でカウントできるように整備されていない。したがって、試行実験の実施に関連するコンビネーションについては、過去の因果関係評価データのクリーニングを行い、スクリーニングが実施できる形に整備した。

なお、試行実験に関連するコンビネーション以外の過去の因果関係評価データについてもクリーニングを進めており、新業務フロー導入時には利用可能となる予定である。

③ 試行実験項目の設定

図 2-3 に示すとおり 1 次スクリーニングを実施するためには、項目 1～3 について基準値を定めなくてはならない。そこで、過去の経験等にもとづき表 2-3 に示す基準値パターンを設定した。項目 1 について 3 パターン、項目 2 について 3 パターン、項目 3 について 3 種の手法を適用し、全部で 27 通りのスクリーニングを実施する。

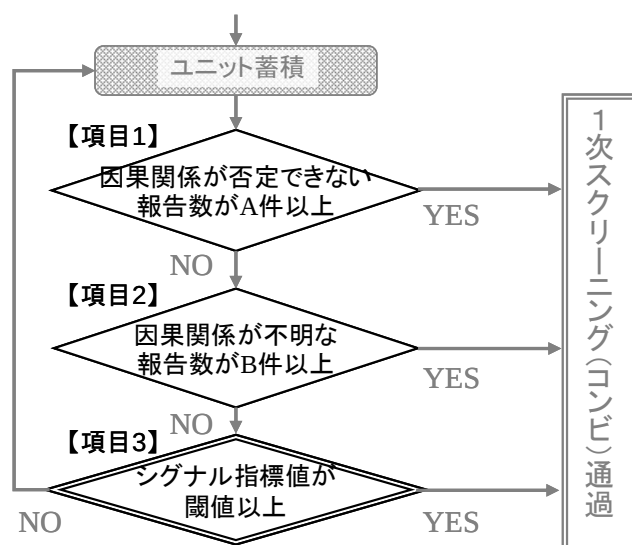


図 2-3 1 次スクリーニングの試行実験項目

表 2-3 試行実験項目と基準値設定パターン

試行実験項目		基準値設定パターン
【項目 1】A (因果関係が否定できない報告数)		A=1、2、3
【項目 2】B (因果関係が不明な報告数)		B=3、4、5
【項目 3】シグナル検出指標値 ³	ROR	ROR (95%信頼下限) ≥ 1
	BCPNN	IC025(95%確率区間の下限) > 0
	GPS	EB05 (5%点) ≥ 2

④ 1 次スクリーニングの実施

①～③の準備を整え、実験対象期間における 1 次スクリーニングを実施した。シグナル検出指標値の算出は、実験対象期間最終日である 2008 年 8 月 1 日時点で旧データベースおよび新データベース⁴に蓄積された症例を対象として実行した。

2.2.3 実験結果

(1) 量の評価

表 2-3 に示した基準値設定パターンごとのスクリーニング実験結果を表 2-4 から表 2-9 に示す。順に項目 3 において用いたシグナル検出手法が ROR、BCPNN、GPS の結果であり、さらにカウント単位がコンビネーションの場合、ユニットの場合でそれぞれの結果を示している。

スクリーニング抽出件数が多いのは ROR の場合であり、300 弱のコンビネーションが抽出された。これに対し、BCPNN では 130 前後、GPS では 100 前後のコンビネーション数となった。

³ 各手法の指標、閾値については参考文献[1]2.5.2 参照。

⁴ 新データベースとは 2003 年 10 月 27 日以降の副作用報告が蓄積されたデータベースであり、旧データベースとは 1994 年 9 月 1 日から 2003 年 10 月 26 日までの副作用報告が蓄積されたデータベース（累積報告件数約 17 万件）である。

表 2-4 シグナル検出手法に ROR を用いた場合の試行実験結果
(コンビネーション単位)

試行実験条件 (基準値)	A	1	1	1	2	2	2	3	3	3
	B	3	4	5	3	4	5	3	4	5
未知報告数		499	499	499	499	499	499	499	499	499
項目 1 通過報告数		23	23	23	1	1	1	0	0	0
項目 2 通過報告数		19	13	9	23	16	11	23	16	11
項目 3 通過報告数		252	258	262	258	265	270	258	265	270
1 次スクリーニング 通過総件数		294	294	294	282	282	282	281	281	281

表 2-5 シグナル検出手法に BCPNN を用いた場合の試行実験結果
(コンビネーション単位)

試行実験条件 (基準値)	A	1	1	1	2	2	2	3	3	3
	B	3	4	5	3	4	5	3	4	5
未知報告数		499	499	499	499	499	499	499	499	499
項目 1 通過報告数		23	23	23	1	1	1	0	0	0
項目 2 通過報告数		19	13	9	23	16	11	23	16	11
項目 3 通過報告数		100	105	108	102	108	112	102	108	112
1 次スクリーニング 通過総件数		142	141	140	126	125	124	125	124	123

表 2-6 シグナル検出手法に GPS を用いた場合の試行実験結果
(コンビネーション単位)

試行実験条件 (基準値)	A	1	1	1	2	2	2	3	3	3
	B	3	4	5	3	4	5	3	4	5
未知報告数		499	499	499	499	499	499	499	499	499
項目 1 通過報告数		23	23	23	1	1	1	0	0	0
項目 2 通過報告数		19	13	9	23	16	11	23	16	11
項目 3 通過報告数		70	73	74	70	74	76	70	74	76
1 次スクリーニング 通過総件数		112	109	106	94	91	88	93	90	87

表 2-7 シグナル検出手法に ROR を用いた場合の試行実験結果
(ユニット単位)

試行実験条件 (基準値)	A	1	1	1	2	2	2	3	3	3
	B	3	4	5	3	4	5	3	4	5
未知報告数		530	530	530	530	530	530	530	530	530
項目 1 通過報告数		23	23	23	1	1	1	0	0	0
項目 2 通過報告数		22	16	12	26	19	14	26	19	14
項目 3 通過報告数		277	283	287	283	290	295	283	290	295
1 次スクリーニング 通過総件数		322	322	322	310	310	310	309	309	309

表 2-8 シグナル検出手法に BCPNN を用いた場合の試行実験結果
(ユニット単位)

試行実験条件 (基準値)	A	1	1	1	2	2	2	3	3	3
	B	3	4	5	3	4	5	3	4	5
未知報告数		530	530	530	530	530	530	530	530	530
項目 1 通過報告数		23	23	23	1	1	1	0	0	0
項目 2 通過報告数		22	16	12	26	19	14	26	19	14
項目 3 通過報告数		121	126	129	123	129	133	123	129	133
1 次スクリーニング 通過総件数		166	165	164	150	149	148	149	148	147

表 2-9 シグナル検出手法に GPS を用いた場合の試行実験結果
(ユニット単位)

試行実験条件 (基準値)	A	1	1	1	2	2	2	3	3	3
	B	3	4	5	3	4	5	3	4	5
未知報告数		530	530	530	530	530	530	530	530	530
項目 1 通過報告数		23	23	23	1	1	1	0	0	0
項目 2 通過報告数		22	16	12	26	19	14	26	19	14
項目 3 通過報告数		90	93	94	90	94	96	90	94	96
1 次スクリーニング 通過総件数		135	132	129	117	114	111	116	113	110

(2) 内容の評価

ROR、BCPNN、GPS の 3 つの手法を用いてスクリーニングの試行実験を実施し、手法によるシグナル検出傾向の違いを確認する。

表 2-10 に、試行実験期間中に報告のあったコンビネーションに対する ROR、BCPNN、GPS でのシグナル検出数を示す。なお、本結果にはスクリーニング条件の項目 1、項目 2 によるスクリーニングは含まれていない。シグナル検出数が最も多かったのは ROR であり、総コンビネーションに対して約 65%でシグナルを検出した。次いで、BCPNN、GPS の順となった。また、検出されたコンビネーションの傾向に関しては、ROR で検出されず BCPNN と GPS で検出されたものが 1 組、GPS のみで検出されたものが 1 組、という結果であり、ROR で検出されたコンビネーションは他の 2 つの手法で検出されたコンビネーションの大半を含むことが確認された。

表 2-11 には、スクリーニング条件の項目 3 に ROR、BCPNN、GPS のそれぞれを適用した際の、スクリーニング通過件数の比較を示す。このとき、スクリーニング条件の項目 1 と項目 2 については、当該項目での通過件数が最も少なくなる条件である A=3、B=5 を採用している。表 2-11 より、ROR で通過したコンビネーションは BCPNN で通過したものをすべて含んでおり、GPS に関しても GPS のみで通過したものは 1 組のみであることがわかる。本結果から、表 2-10 に示した結果と同様に、ROR で通過したコンビネーションは、他の 2 つの手法で通過したコンビネーションの大半を含んでいることが確認された。

表 2-10 ROR、BCPNN、GPS における実験期間中のシグナル検出数

(コンビネーション数)

総数			1,547
	ROR で検出		1,013
		ROR のみで検出	274
		ROR かつ BCPNN で検出	187
		ROR かつ GPS で検出	7
		ROR かつ BCPNN かつ GPS で検出	545
	BCPNN で検出		733
		BCPNN のみで検出	0
		BCPNN かつ GPS で検出	1
	GPS で検出		554
		GPS のみで検出	1

表 2-11 項目 1 : A=3、項目 2 : B=5 における項目 3 (シグナル検出指標値) での
3 手法間の通過件数の比較

(コンビネーション数)

ROR で通過		270
	ROR のみで通過	154
	ROR かつ BCPNN で通過	41
	ROR かつ GPS で通過	4
	ROR かつ BCPNN かつ GPS で通過	71
BCPNN で通過		112
	BCPNN のみで通過	0
	BCPNN かつ GPS で通過	0
GPS で通過		76
	GPS のみで通過	1

2.3 新業務フロー試行実験のまとめ

本章では、平成 21 年度の安全対策業務へのシグナル検出手法の導入に向けた新業務フローの実用性を検証するための試行実験を実施した。

未知報告の新業務フローでは 1 次スクリーニングとして、過去の評価データを用いて設定する 2 つの項目と、シグナル検出手法を用いて設定される 1 つの項目の合計 3 項目からなる条件を設定することとしている。そこで、本試行実験において、過去の評価データを用いて設定される項目 1、2 については、その通過基準値を変化させた場合の通過シグナル数を確認し、シグナル検出手法を用いる項目 3 については、ROR、BCPNN、GPS の 3 手法を適用した場合に、それぞれの手法で検出されるコンビネーションの違いを確認した。

項目 1、2 については基準値を 9 パターンに変化させて通過報告数を算出したが、設定した 9 パターンでは通過報告数に大きな差が無いことが確認された。項目 3 については、ROR と比べ、BCPNN、GPS での通過報告数はそれぞれ 4～5 割、3～4 割と、半数以下であった。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た。ROR を用いた際の 1 次スクリーニング通過報告数はコンビネーション単位で 300 弱であり、今後の安全対策担当職員の増員を考慮すると、対応可能な業務量と言える。以上の結果から、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、BCPNN と GPS は参考値として算出、提示することとする。

3. Ad Hoc 解析の検討

3.1 Ad Hoc 解析

PMDA では定型業務とは別に、非定型業務でのデータマイニング手法の活用のひとつとして、昨年度から Ad Hoc 解析の試行を実施している。

通常のシグナル検出指標値は、評価者が注目する医薬品における、注目する副作用の発現リスクを副作用報告データベースの全データを対象として算出する。これに対し、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指標値を算出する。これにより、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される。

3.2 ケーススタディ

Ad Hoc 解析により得られる特定医薬品での特定副作用の発現リスクに関する指標値の、安全分析業務における参考情報としての利用可能性の探索を目的とし、ケーススタディとして医薬品 A と間質性肺炎を採り上げ、Ad Hoc 解析を試行する。

(1) 背景

間質性肺炎の副作用報告が寄せられている医薬品の中で、特定の医薬品が他剤と比較してリスクが高いのかどうかを知りたい。

(2) 試行の目的

特定の副作用が報告されている医薬品についての順位表の利用可能性を探索する。

(3) 方法

表 3-1 に本事例での実験実施方法をまとめる。

表 3-1 Ad Hoc 解析の方法

注目した医薬品と 副作用	医薬品	医薬品 A を含む全医薬品
	副作用	間質性肺疾患
シグナル指標値 計算手法		ROR (シグナル検出閾値：ROR 95%CI 下限値 $\geq$ 1)
注目した副作用が報告されている医薬品について以下 2 つの順位表を作成し、その傾向を比較		
・ 報告件数順		
・ 全データセットを用いて算出したシグナル指標値（95%CI 下限値）順		

(CI：信頼区間)

(4) 結果

1) 副作用を間質性肺炎に限定した場合の報告件数順

報告件数順に並べたときに上位を占めた薬効群を表 3-2 に示す。

上位を占めた薬効群の特徴としては、抗がん剤（422 代謝拮抗剤、424 抗腫瘍性植物成分製剤、429 その他の腫瘍用薬）、抗リウマチ剤（399 他に分類されない代謝性医薬品、442 刺激療法剤）、抗不整脈薬（212 不整脈用剤）等の間質性肺炎が副作用として発現することが一般的によく知られている薬剤ばかりが頻出している点が挙げられる。

なお、報告件数が100件を超えていても、シグナル指標値の閾値を超えないものも含まれていた。

表 3-2 報告件数順に並べたときに上位を占めた薬効群

No (報告 件数順)	薬効分類		報告 件数	ROR		
				シグ ナル	計算値	95%CI 下限値
1	429	その他の腫瘍用薬	1,715	1	36.10	33.52
2	399	他に分類されない代謝性医薬品	648	1	11.08	10.08
3	212	不整脈用剤	399	1	20.99	18.35
4	422	代謝拮抗剤	390	1	5.41	4.84
5	520	漢方製剤	354	1	51.50	42.88
6	424	抗腫瘍性植物成分製剤	350	1	1.65	1.48
7	424	抗腫瘍性植物成分製剤	232	1	2.81	2.45
8	442	刺激療法剤	210	1	5.96	5.12
9	424	抗腫瘍性植物成分製剤	177	0	0.55	0.47
10	422	代謝拮抗剤	173	1	1.68	1.44
11	422	代謝拮抗剤	163	1	9.70	8.09
12	424	抗腫瘍性植物成分製剤	162	1	4.74	4.00
13	249	その他のホルモン剤	159	1	2.89	2.45
14	639	その他の生物学的製剤	158	1	1.35	1.15
15	429	その他の腫瘍用薬B	143	1	2.63	2.21
16	339	他に分類されない代謝性医薬品	132	1	8.19	6.72
17	639	その他の生物学的製剤	129	1	3.00	2.50
18	399	他に分類されない代謝性医薬品	117	1	3.28	2.70
19	429	その他の腫瘍用薬	114	1	2.26	1.86
20	639	その他の生物学的製剤	112	1	3.32	2.72

2) 副作用を間質性肺炎に限定した場合の 95%CI 下限値順

ROR 値の 95%CI 下限値順で並べたときに、上位を占めた薬効群を表 3-3 に示す。

上位を占めた薬効群の特徴としては、漢方製剤は報告件数が比較的少なくても上位を占める一方、抗がん剤は報告件数が比較的多くても上位ではない事が挙げられる。

表 3-3 報告件数順に並べたときに上位を占めた薬効群

No (95% CI 下限 値順)	No (報告 件数 順)	薬効分類		報告 件数	ROR		
					シグ ナル	計算値	95%CI 下限値
1	5	520	漢方製剤	354	1	51.50	42.88
2	1	429	その他の腫瘍用薬	1,715	1	36.10	33.52
3	89	520	漢方製剤	22	1	38.21	19.55
4	3	212	不整脈用剤	399	1	20.99	18.35
5	93	520	漢方製剤	21	1	31.91	16.65
6	24	423	抗腫瘍性抗生物質製剤	77	1	20.23	14.94
7	21	520	漢方製剤	101	1	19.31	14.87
8	23	429	その他の腫瘍用薬	90	1	18.81	14.28
9	55	423	抗腫瘍性抗生物質製剤	42	1	19.30	12.87
10	2	399	他に分類されない代謝性医薬品	648	1	11.08	10.08
11	176	520	漢方製剤	8	1	27.76	10.06
12	76	520	漢方製剤	26	1	15.81	9.65
13	154	520	漢方製剤	11	1	17.82	8.18
14	48	442	他に分類されない代謝性医薬品	47	1	11.57	8.17
15	11	422	代謝拮抗剤	163	1	9.70	8.09
16	50	429	その他の腫瘍用薬	46	1	11.32	7.97
17	123	520	漢方製剤	15	1	15.19	7.97
18	138	520	漢方製剤	13	1	15.79	7.85
19	188	520	漢方製剤	7	1	21.25	7.70
20	130	520	漢方製剤	14	1	14.18	7.33

3) 医薬品を漢方製剤と抗がん剤に限定した場合の結果の比較

漢方製剤と抗がん剤に限定した結果の比較を表 3-4 に示す。

副作用が「間質性肺炎」での報告件数は、抗がん剤のほうが漢方製剤より多い傾向があったにも関わらず、シグナル指標値が閾値を越えないものがあった。一方、漢方製剤は報告件数が比較的少なくてもシグナル指標値が閾値を越えていることが確認された。

ここで、N11/N12 の値を考えてみると、漢方製剤では「間質性肺炎」が「他の副作用」よりも報告件数が多いこと、抗がん剤では「他の副作用」が「間質性肺炎」よりも報告件数が多いことから、漢方製剤では「間質性肺炎」が「他の副作用」よりも報告されやすい傾向にあることが推測された。

表 3-4 漢方製剤と抗がん剤に限定した結果の比較

No	No	薬効分類		N11*	N12*	N21*	N22*	ROR 値	95%CI 下限値
1	5	520	漢方製剤	354	172	11,235	281,133	51.5	42.88
3	89	520	漢方製剤	22	14	11,567	281,291	38.21	19.55
5	93	520	漢方製剤	21	16	11,568	281,289	31.91	16.65
7	21	520	漢方製剤	101	128	11,488	281,177	19.31	14.87
11	176	520	漢方製剤	8	7	11,581	281,298	27.76	10.06
12	76	520	漢方製剤	26	40	11,563	281,265	15.81	9.65
13	154	520	漢方製剤	11	15	11,578	281,290	17.82	8.18
17	123	520	漢方製剤	15	24	11,574	281,281	15.19	7.97
18	138	520	漢方製剤	13	20	11,576	281,285	15.79	7.85
19	188	520	漢方製剤	7	8	11,582	281,297	21.25	7.7
20	130	520	漢方製剤	14	24	11,575	281,281	14.18	7.33
2	1	429	その他の 腫瘍用薬	1,715	1,347	9,874	279,958	36.1	33.52
6	24	423	抗腫瘍性抗 生物質製剤	77	93	11,512	281,212	20.23	14.94
8	23	429	その他の 腫瘍用薬	90	117	11,499	281,188	18.81	14.28
9	55	423	抗腫瘍性抗 生物質製剤	42	53	11,547	281,252	19.3	12.87
15	11	422	代謝拮抗剤	163	413	11,426	280,892	9.7	8.09
16	50	429	その他の 腫瘍用薬	46	99	11,543	281,206	11.32	7.97

※ N11, N12, N21, N22 は以下のセル度数の数字を表す

	注目する副作用	その他の副作用
注目する医薬品	N11	N12
その他の医薬品	N21	N22

3.3 Ad Hoc 解析のまとめ

本章における検討結果より、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と考えられる。

データマイニング手法の利用可能性に関する検討の中で、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、今後も引続き試行を行う予定である。

4. 重複報告検出手法の検討

4.1 背景と目的

副作用報告データベースの中に重複報告⁵が存在すると、1つの症例報告（報告）を複数としてカウントすることになる。そのため実際よりも多くの報告をカウントしていることになり、評価、分析に影響を与える。EMEA や Lareb は、データクリーニングの1つとして、データベースから重複報告を検出し、評価の際に取り除くプロセスを導入している。

シグナル検出手法により算出される指標値の信頼性を高めることを目的とし、副作用報告データベースを対象に、既存手法を参考にした重複報告検出手法について検討を行った。

4.2 重複報告類似スコアによる評価

WHO-UMC は、参考文献[5]で重複報告検出手法を提案している。今回の重複報告検出手法の検討では、この手法を PMDA の副作用報告データベースに適用することから着手した。

文献に示された既存手法の概要をまとめる。

タイトル	Duplicate detection in adverse drug reaction surveillance
著者名	G. Niklas Norén, Roland Orre, Andrew Bate, I.R. Edwards
掲載誌	Data Mining and Knowledge Discovery, 2007, 14(3) , 305-328

4.2.1 類似スコアによる重複報告の検出

(1) 概要

WHO-UMC が提案している手法は、Hit-miss モデル（混合分布モデル）にもとづく重複報告の評価モデルを作成し、そのモデルにもとづいて、報告対の類似度（重複報告の疑わしさ）を類似スコアとして計算するものである。

重複報告の判定閾値を定め、判定閾値以上の類似スコアを有する報告、あるいは類似スコアの順位が上位にくる報告を、重複報告として検出する。

重複報告検出の処理フローを図 4-1 に示す。処理は以下の 3 つのフローで構成される。

⁵ 同一症例に対して、複数の経路（企業）から報告され、データベースに登録されている報告組

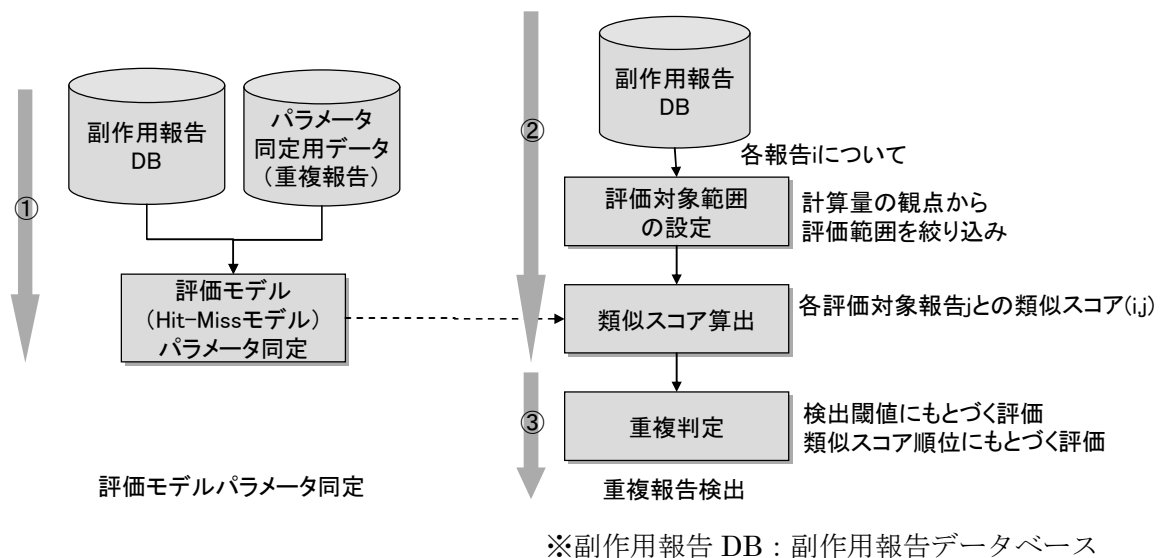


図 4-1 重複報告検出の処理フロー

① 評価モデルパラメータ同定

報告対の類似スコア算出に用いる評価モデルのパラメータ同定を行う。評価モデルは報告中の評価に用いる属性（症例報告に含まれるデータ項目）ごとに1つ作成され、属性値（属性に入力されている値）の種類（離散値、ベクトル値、連続値）に応じて異なる評価モデルが用いられる。

パラメータの同定には以下の2種類のデータを用いる。

- ・ 副作用報告データベース
- ・ パラメータ同定用データ（あらかじめ重複報告であるとわかっている報告組（正解データ）のリスト）

② 類似スコア算出

副作用報告データベース内の報告のうち、重複判定を行う報告の評価対象範囲を設定する。副作用報告データベース内にあるすべての報告との判定を行うと、計算量が大きく、また明らかに重複の可能性がない報告についても判定を行うことになる。例えば、性別が一致する、一致する医薬品（一般名）が1項目以上ある、などの条件を設定し、重複報告である可能性がある報告のみを評価対象として設定する。

設定した評価対象範囲内の報告について、重複報告を評価する基準となる報告との類似スコアを算出する。類似スコアの算出は、①で得た評価モデルの計算式

を用いて行う。

③ 重複報告検出

算出された類似スコアにもとづき、重複報告の判定を行う。判定方法には、以下の2通りがある。

- ・ 判定閾値による判定：あらかじめ設定した判定閾値と類似スコアを比較し、類似スコアが判定閾値を上回る報告を重複報告と判定する。
- ・ 順位による判定：報告 i に対する評価対象範囲内の各報告の類似スコア順位をみて、順位が上位にくる報告を重複報告と判定する。

(2) 重複報告の評価モデル

評価モデルは属性値ごとに異なり、離散値、ベクトル値、連続値の3種類がある。各評価モデルについて以下に示す。

① 離散値属性の評価モデル

報告 X、報告 Y について、属性 A の属性値がそれぞれ j、k である場合を考える。

表 4-1 離散値モデルにおける報告と属性のイメージ

報告	属性 A
X	X=j
Y	Y=k

報告対の属性 A の値 j、k にもとづき、離散値の評価モデルでは、属性ごとのスコア値 W_{jk} を以下のように求める。

a) $j \neq k$ (報告 X と報告 Y の属性値が異なる場合) :

$$W_{jk} = \log_2 c - 2\log_2(1-b)$$

b) $j=k$ (報告 X と報告 Y の属性値が一致する場合) :

$$W_{jk} = \log_2 \{1 - c(1 - \beta_j)(1-b)^{-2}\} - \log_2 \beta_j$$

c) j 又は k が欠損値（報告 X または報告 Y のいずれかが欠損値の場合）：

$$W_{jk} = 0$$

モデル中のパラメータは b 、 β_j 、 c 、の 3 つであり、それぞれの意味は以下のとおりである。

- ・ b ：データベースの全報告中の欠損値の割合。
- ・ β_j ：データベースの欠損値を除いた報告のうち属性値が j である報告の割合。
- ・ c ：重複報告である場合に、属性値が一致しない確率を評価したパラメータ。

② ベクトル値属性の評価モデル

表 4-2 に示すように、報告中に当該医薬品（副作用）の報告があるときを 1、ないときを 0 とし、医薬品（副作用）は医薬品（副作用）の種類数の次元をもつベクトルで値を表現する。

表 4-2 ベクトル値モデルにおける報告と属性のイメージ

報告	医薬品#1	医薬品#2	...	医薬品#t	...	医薬品#m
X	X[1]	X[2]	...	X[t]	...	X[m]
Y	Y[1]	Y[2]	...	Y[t]	...	Y[m]

※ $\{X[t] \mid 0,1\}$ 、 $\{Y[t] \mid 0,1\}$

ベクトル値の評価モデルでは、医薬品（副作用）ごとにスコア値 W_t を以下のよう求める。医薬品（副作用）全体のスコア値は、各医薬品（副作用）について求められたスコア値の合計となる。

a) $X[t]=1$ かつ $Y[t]=0$ 、又は $X[t]=0$ かつ $Y[t]=1$ （医薬品（副作用） t が片方の報告にしかない場合）：

$$W_t = \log_2 c - 2\log_2(1-b)$$

b) $X[t]=1$ かつ $Y[t]=1$ （両方の報告が医薬品（副作用） t を含む場合）：

$$W_t = \log_2 \{1 - c(1 - \beta_j)(1 - b)^{-2}\} - \log_2 \beta_j$$

c) $X[t]=0$ かつ $Y[t]=0$ （両方の報告が医薬品（副作用） t を含まない場合）：

$$W_t = 0$$

モデル中のパラメータは b 、 β_j 、 c 、の 3 つであり、それぞれの意味は以下のとおり。

- b : データベースの全報告中の欠損値（医薬品（副作用）が 1 つも報告されていない）の割合。
- β_j : 医薬品（副作用）の報告が 1 つ以上ある報告のうち、当該医薬品（副作用） t の報告がある（ $X[t]=1$ である）報告の割合。
- c : 重複報告である場合に、属性値が一致しない確率を評価したパラメータ。

③ 連続値属性の評価モデル

報告 X 、報告 Y について、属性 A の属性値がそれぞれ j 、 k である場合を考える。

表 4-3 連続値モデルにおける報告と属性のイメージ

報告	属性 A
X	$X=j$
Y	$Y=k$

スコア値 $W_{(d)}$ は、評価する報告対の属性値の差 $d=k-j$ にもとづき、次のように求める。 $P_d(d)$ は混合分布（4 つの異なる確率分布の重み付き和）を示す式となっている。

$$W_{(d)} = \log_2 \left(\frac{\int p_d(d) \delta d}{\int p_u(d) \delta d} \right)$$

ここで、

$$p_d(d) = (1 - a_1 - a_2 - b)^2 \cdot \delta(d) + a_2(2 - a_2 - 2b) \cdot f(d) \\ + 2a_1(1 - a_1 - a_2 - b) \cdot \phi(d; 0, \sigma_1^2) + a_1^2 \cdot \phi(d; 0, 2\sigma_1^2)$$

$$p_u(d) = (1 - b)^2 \cdot f(d)$$

モデル中には $\delta(d)$ 、 $f(d)=\phi(d;0,\sigma_2^2)$ 、 $\phi(d;0,\sigma_1^2)$ 、 $\phi(d;0,2\sigma_1^2)$ の 4 とおりの関数が用いられ、 b 、 a_1 、 a_2 、 σ_1^2 、 σ_2^2 が同定の必要なパラメータである。それぞれの意味は以下のとおりである。

- ・ $\delta(d)$ 、 $f(d)=\phi(d;0,\sigma_2^2)$ 、 $\phi(d;0,\sigma_1^2)$ 、 $\phi(d;0,2\sigma_1^2)$ ：混合分布を構成する各確率分布の確率密度関数。

※ $\delta(d)$ ：ディラックのデルタ関数。

$$\delta(d)=\begin{cases} 1 & (d=0) \\ 0 & (d \neq 0) \end{cases}$$

- ・ b ：データベースの全報告中の欠損値の割合。
- ・ a_1 、 a_2 ：確率分布モデルの事前確率に関するパラメータ。
- ・ σ_1^2 、 σ_2^2 ：確率密度関数の分散。

WHO-UMC 論文では d の積分範囲を、年齢、受理日のそれぞれについて以下のように設定している。

- ・ 年齢： $-1 \leq d [\text{year}] \leq 1$
- ・ 受理日： $-12 \leq d [\text{day}] \leq 20$

(3) 類似スコアの算出

全ての評価属性のスコア値の総和を、その報告対の類似スコアとする。WHO-UMC の手法では、類似スコア算出に用いる属性として以下の 7 属性を用いている。

- ・ 性別（離散値属性）
- ・ 転帰（離散値属性）
- ・ 報告国（離散値属性）
- ・ 医薬品（一般名）（ベクトル値属性）
- ・ 副作用（MedDRA・HLT⁶）（ベクトル値属性）
- ・ 年齢（連続値属性）
- ・ 受理日（連続値属性）

⁶ WHO-UMC 論文^[6]中には、用語集として MedDRA、WHO-ART のいずれを、またいずれの階層を採用したのかという点について明記されていない。ただ、実験のパラメータとして副作用の種類数が 1950 と示されている。PMDA で主に使用している MedDRA では、この種類数は HLT 階層の語数に近いたため、今回の実験では MedDRA の HLT 階層を使用した。

類似スコア算出のイメージを図 4-2 に示す。報告対の類似度が高いほどスコアは大きくなり、類似度が低いほどスコアは小さくなる。

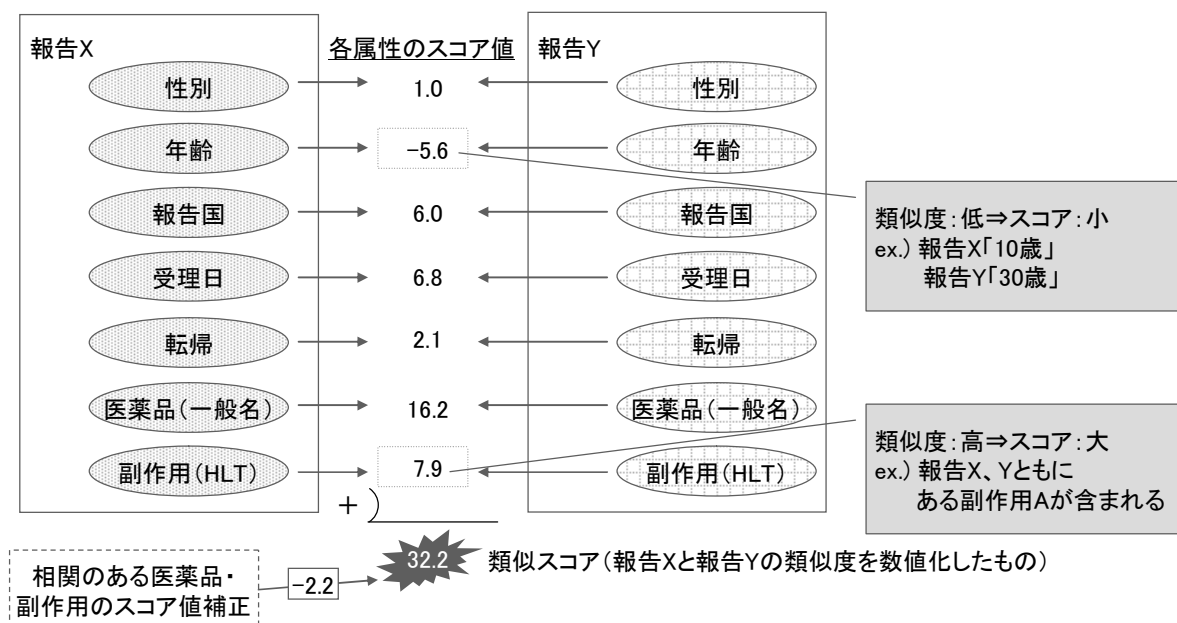


図 4-2 類似スコア算出と類似度評価

(4) 相関のある項目間の類似スコア補正

医薬品と副作用については、ある医薬品 A が処方されると特定の副作用 B が発生しやすい（あるいは特定の医薬品 C が同時に処方されやすい）、という相関関係がある。相関性が類似度スコアに与える影響を考えると、例えば報告 X と報告 Y において、相関関係のある医薬品 A と副作用 B がともに一致する場合と、相関関係のない医薬品 A と副作用 D がともに一致する場合とを比較した場合、報告 X と報告 Y が重複報告である可能性は後者の方が前者よりも高いと考えられる。

そこで類似スコアの算出においては、医薬品のスコア値と副作用のスコア値の和を取る際に、相関性の強さに応じてスコア値を減じる補正を行っている。

● 医薬品（副作用）と医薬品（副作用）の相関性

全医薬品、副作用 (ma=m1+m2、m1: 医薬品の種類数、m2: 副作用の種類数) について考える。ある医薬品（または副作用）と別の医薬品（または副作用）の相関性（共起性）は以下の指標 IC で表される。

$$IC_{s,t} = \log_2 \frac{P(X[s]=1|X[t]=1)}{P(X[t]=1)}$$

$$\times 0 \leq s, t \leq ma$$

IC の値が大きいほど、その医薬品と副作用（医薬品）は相関性が高いことになる。

- 相関補正を考慮した医薬品と副作用のスコア値

相関補正分のスコア値 $IC_t^* - W_t$ は、医薬品、副作用のそれぞれについて以下のよう求める。

$$IC_t^* - W_t = \arg \max_{s; s < t} IC_{s,t}$$

報告対のいずれかに含まれていない（ $X[s]=0$ 又は $Y[s]=0$ となる）医薬品、副作用との相関は $IC_{s,t} = 0$ として扱い、 $s < t$ の範囲で最大となる $IC_{s,t}$ を $IC_t^* - W_t$ とする。各医薬品、副作用のスコアの総和から相関補正分を引いた値が、医薬品と副作用のスコア値の和 $\sum_{t=1}^{ma} W_t$ となる。

$$\sum_{t=1}^{ma} W_t = \sum_{t=1}^{m1} W_t + \sum_{t=1}^{m2} W_t - \sum_{t=1}^{ma} IC_t^* - W_t$$

(5) 重複判定の閾値設定

算出された類似スコアが大きい報告対ほど、重複報告である可能性が高いと評価できる。判定閾値にもとづいて重複報告を検出する場合に用いる判定閾値は、類似スコアが s の報告対が重複報告（dup）である確率 $P(\text{dup} | s)$ に関する以下の式から設定する。

$$P(\text{dup} | s) = \frac{\frac{0.05}{3.00 \times 10^6} \cdot \phi(s; \hat{\mu}_r, \hat{\sigma}_r^2)}{\frac{0.05}{3.00 \times 10^6} \cdot \phi(s; \hat{\mu}_r, \hat{\sigma}_r^2) + \left(1 - \frac{0.05}{3.00 \times 10^6}\right) \cdot \phi(s; \hat{\mu}_u, \hat{\sigma}_u^2)}$$

- ・ s : 類似スコア
- ・ $\phi(s; \hat{\mu}_u, \hat{\sigma}_u^2)$: データベース全体の類似スコア分布。パラメータはデータベース内からランダムに選んだ 10,000 組の症例報告から求めた類

似スコア s の分布から求める。

- ・ $\phi(s, \hat{\mu}_r, \hat{\sigma}_r^2)$: 正解データリストの類似スコア分布。パラメータは正解データ中の全ての重複報告対から求めた類似スコア s の分布から求める。

判定閾値を決める前提条件として、重複報告含有率 $Z(\%)$ を設定する。 Z は副作用報告データベースの全報告に占める重複報告の割合であり、上記の式は、 $Z=5\%$ という WHO-UMC のおいた仮定にもとづく式である。重複報告の類似スコア、データベースに含まれる全報告の類似スコアがそれぞれ別々の正規分布で表されると仮定したとき、データベースから取り出したある報告対が重複報告である確率を示している。

WHO-UMC では、取り出したある報告対が 95% 以上の確率で重複報告である ($P(\text{dup}|s)=0.95$) ときの類似スコア s を判定閾値として用いた。本検討においても、この基準を用いて判定閾値を設定する。

4.2.2 評価実験

(1) 利用データ

評価実験には以下の 2 種類のデータを用いた。

- ・ 副作用報告データベース (新データベース)
- ・ 正解データリスト

副作用報告データベース (新データベース) には約 13 万件の報告が含まれる⁷。また、正解データリストは「過去の伝送で記載されたその他の症例識別子はあるか (A1.11)」で重複報告の症例識別子が登録されている報告を抽出し、作成した。正解データは新データベースの中から 307 グループ (ペア、または 3 つ以上の報告が互いに重複である報告のグループ) が確認された。このうち 150 グループを評価モデルのパラメータ同定に利用し、残りの 157 グループを評価用データとした。

(2) 基準データとテストデータ

評価用データの 1 グループのうち、第一被疑薬の投与開始日が最も新しい報告を基準データ、それ以外の報告をテストデータと定義する。

(3) 利用属性

評価実験において類似スコア算出に用いた属性は以下のとおり。PMDA に報告される症

⁷ 2008 年度時点

例報告は、同一症例であっても企業によって受理日が大きく異なる場合がある。このため WHO-UMC の提案する手法に対し、受理日を（基準データの）第一被疑薬の投与開始日に変更した。また、報告国は利用属性から除外した。

- ・ 性別
- ・ 転帰
- ・ 医薬品（一般名）
- ・ 副作用（HLT）
- ・ 年齢
- ・ （基準データの）第一被疑薬の投与開始日

(4) 評価対象範囲

重複報告の検出を行う評価の対象範囲について、表 4-4 に示す 4 通りを設定した。Case 0 は参考とした WHO-UMC の論文と同様の設定である。

表 4-4 評価対象範囲の設定

	Case 0	Case 1	Case 2	Case 3
医薬品（一般名）1項目以上一致	○	○	○	○
副作用(SOC)1項目以上一致	○	○	○	○
性別一致		○	○	○
転帰3段階以内			○	○
受理日の差700日以内				○
年齢の差15歳以内				○
評価対象範囲の報告数	795,029	402,568	318,837	124,669

(5) 評価方法

評価用データの基準データについて、重複報告検出手法を用いた重複報告の判定・検出を行い、テストデータの検出の有無を評価した。評価は以下の 2 通りの方法で行った。

- ① 判定閾値による評価
- ② 類似スコア順位による評価

①では、判定閾値の値によって、テストデータの検出率の変化を確認した。

②では、類似スコア順位に着目した。評価用データは重複報告であるので類似スコアは大きくなると考えられ、テストデータの順位は 1 位など上位となることが望ましい。テス

トデータが位置する順位を確認し、1 位にならなかった場合について考察を行った。

図 4-3 に、評価実験の流れを示す。

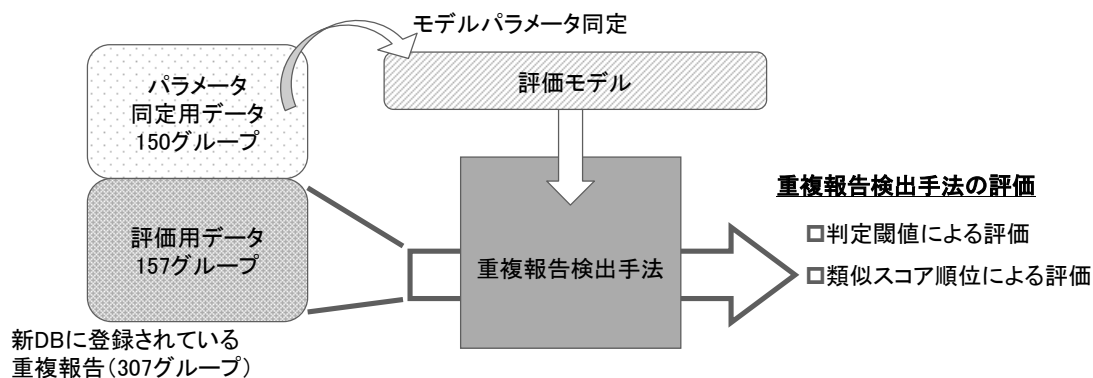


図 4-3 評価実験の流れ

4.2.3 実験結果

(1) 判定閾値による評価結果

類似スコアが判定閾値以上の報告を重複報告として検出する場合について、判定閾値と重複報告として検出される評価用データの関係を図 4-4 に示す。

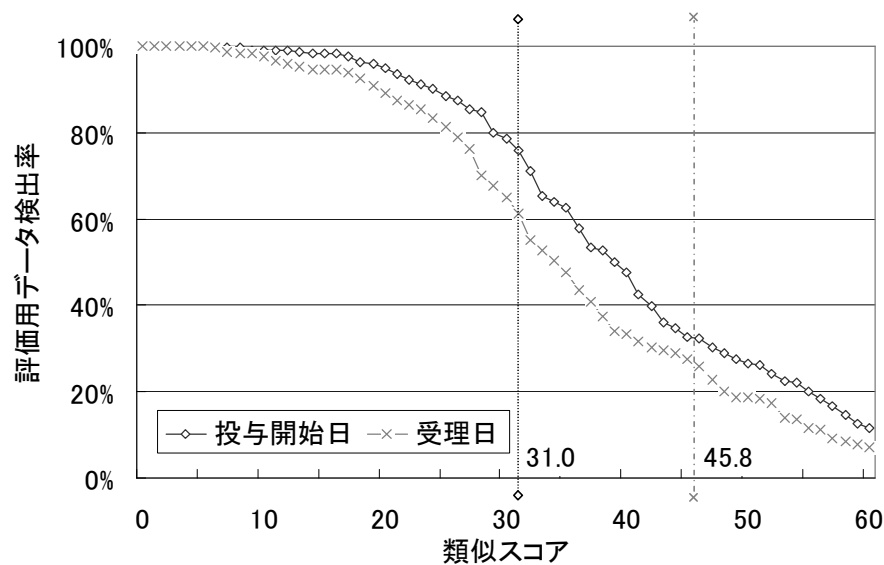


図 4-4 判定閾値と評価用データ検出率の関係

比較のため、投与開始日ではなく受理日を用いた場合の結果を合わせて示した。類似スコア $s=31.0$ (点線)、 45.8 (一点鎖線) は、それぞれ投与開始日を用いた場合、受理日を用いた場合の判定閾値である。類似スコア s の報告が重複報告である確率 (投与開始日を

用いた場合) について、図 4-5 に重複報告含有率 Z 別に示す。

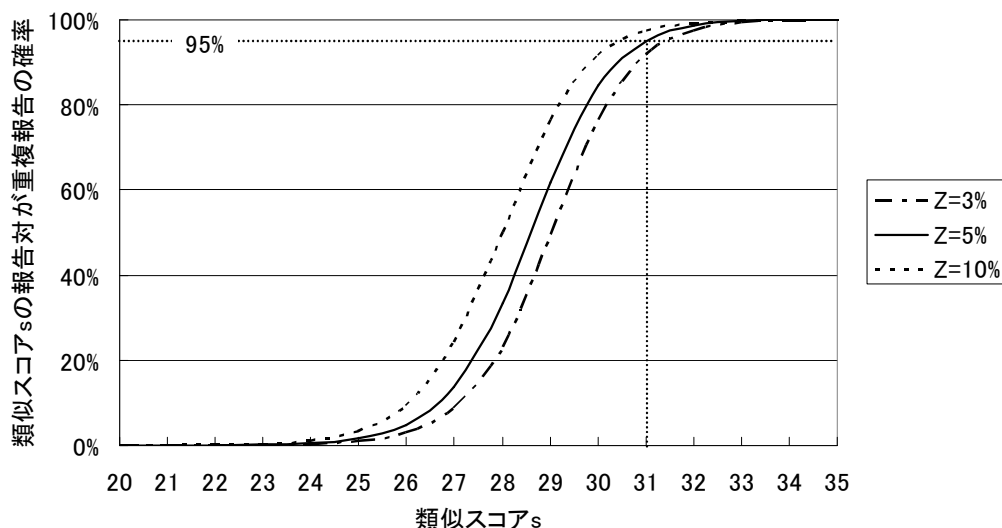


図 4-5 類似スコア s の報告が重複報告の確率 (重複報告含有率 Z 別)
(投与開始日を用いた場合)

Z を大きく取ると、同じ類似スコアでも重複報告の確率が上昇することがわかる。ここでは WHO-UMC の論文と同様に重複報告含有率 $Z=5\%$ 、類似スコア s の報告対が重複報告である確率 95% を基準に判定閾値を設定し、投与開始日を用いた場合の判定閾値 31.0 を得た。同様にして、受理日を用いた場合の判定閾値は 45.8 となった。

図 4-4 に戻ると、判定閾値が小さくなるほど評価用データ検出率は上昇する。しかし、同時に類似度の低い報告対を重複報告として検出する割合も多くなる。逆に判定閾値を大きくとると類似度のより高い報告対が重複報告として検出されるようになるが、評価用データ検出率は低下する。

また、投与開始日を利用した場合と受理日を利用した場合を比べると、前者の方が同じ判定閾値に対する評価用データ検出率が高くなった。したがって投与開始日を利用した場合の方が重複報告検出の精度が高いといえる。

一方、判定閾値の類似スコアは投与開始日の方が小さくなっているが、これは利用する属性が変わったことで副作用報告データベースの類似スコア分布が変化したことが主な原因と思われる。副作用報告データベースに含まれる報告対の類似スコア、および正解データリストに含まれる重複報告対の類似スコアの確率分布を図 4-6 に示す。

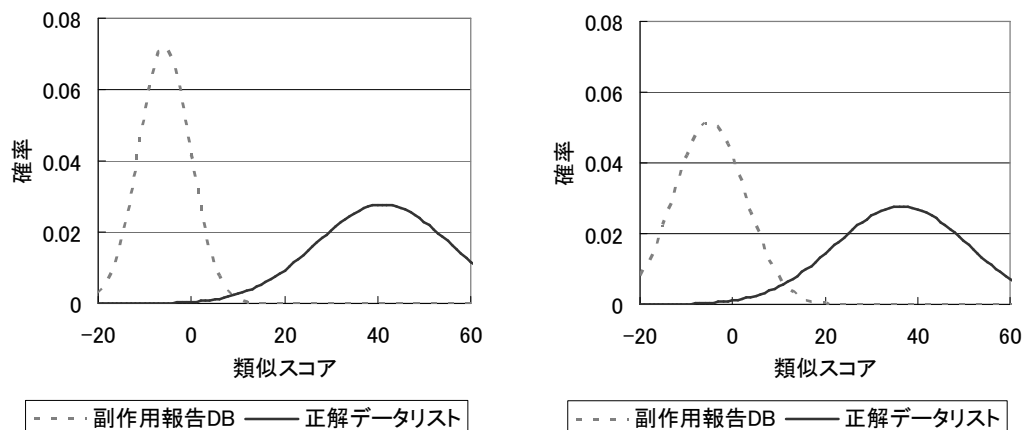


図 4-6 副作用報告データベースと正解データリストの類似スコア確率分布
(左：投与開始日を用いた場合、右：受理日を用いた場合)

投与開始日を用いた場合と受理日を用いた場合を比較すると、正解データリストに含まれる重複報告対の類似スコア分布に大きな変化はないが、データベース全体の類似スコア分布は分散が小さくなり、尖度が高くなった。これにより類似度の高い報告対と類似度の低い報告対の類似スコア分布の重なる領域が狭まり、判定閾値が小さくなったと考えられる。

(2) 類似スコア順位による評価結果

● 評価実験の概要と類似スコア順位表

評価用データの基準データについて、図 4-7 のような評価対象範囲の報告（テストデータ含む）の組合せに対して類似スコアを算出し、順位表を作成した。評価用データには 157 件の基準データと、それに関連した 214 件のテストデータが含まれる。

表 4-4 に示した Case0～3 の 4 種類の評価対象範囲について、類似スコアの計算を行い、基準データごとに、表 4-5 に示すような順位表を作成した。

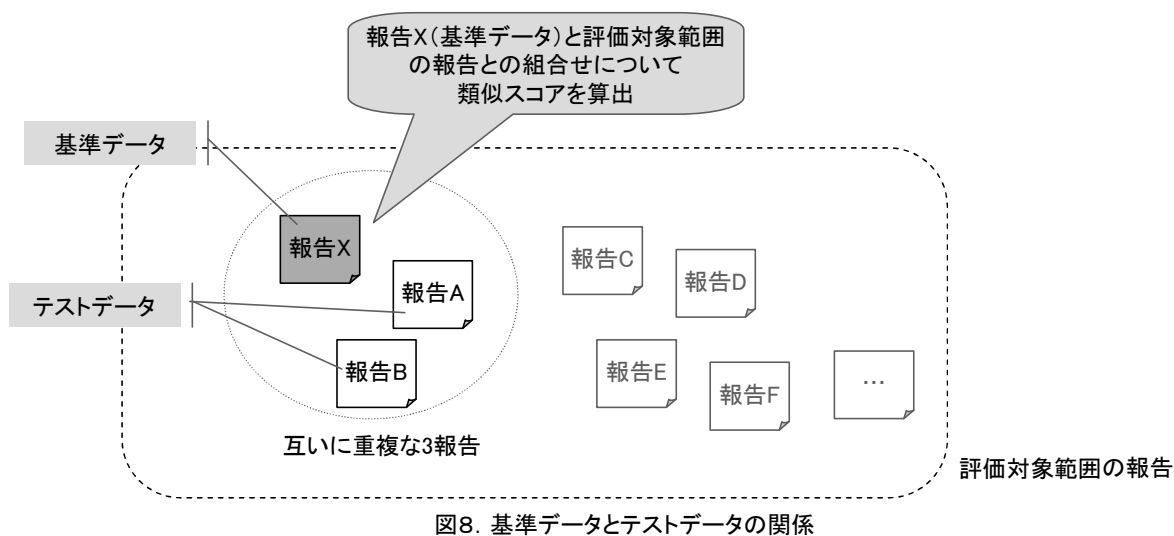


図 4-7 基準データ、テストデータと評価対象範囲の報告との関係

表 4-5 ある基準データに関する類似スコア順位表（イメージ）

順位	性別	転帰	医薬品	副作用	第一被疑薬 投与開始日(※)	年齢	類似スコア	テストデータ
基準データ	女	回復	3 in total	2 in total	2008/2/27	35	—	
1	女	回復	3 of 3 +2	2 of 2	2008/2/27	35	37.20	*
2	女	回復	3 of 3	2 of 2	2008/2/27	34	32.50	
3	女	軽快	1 of 3 +2	1 of 2 +1	2008/2/20	38	28.00	
...	...	...	...	...	...		...	...

※基準データの第一被疑薬の投与開始日

類似スコア順位表の見方は以下のとおりである。

- 順位がついていない、一行目の報告は基準データ。
- 基準データとの組合せで計算した類似スコアで、各報告の順位を表示。
- 各欄の内容は報告の各属性値。
- 医薬品・副作用の欄は、基準データに含まれる医薬品・副作用について、類似スコアの計算対象となる報告に含まれる数を示す。基準データの「3 in total」は基準データに含まれる医薬品が3種類であることを示している。各報告の「1 of 3 +1」のような記載内容は、計算対象の報告が基準データに含まれる3種類の医薬品のうち1種類を含み、また一致しない医薬品も2種類含むことを示す。
- 報告がテストデータの場合は最右欄に「*」を付与。

参考として、表 4-6、表 4-7 に、投与開始日を利用した場合について、各属性のスコアの最大値と最小値を示す。

表 4-6 投与開始日を利用した場合の離散値属性、ベクトル値属性のスコア

属性	属性値が不一致の場合	属性値が一致する場合	
		最大値	最小値
性別	−4.27	1.01	0.92
転帰	−2.04	5.47	0.93
医薬品（一般名）	0.00	16.97	2.41
副作用（HLT）	0.00	16.97	4.29

表 4-7 投与開始日を利用した場合の連続値属性のスコア

属性	いずれかの属性値が欠損値の場合	いずれかの属性値が欠損値でない場合	
		最大値	最小値
年齢	0.00	8.34	−5.64
投与開始日	0.00	8.43	−4.12

- 類似スコア順位表におけるテストデータの順位

属性に投与日を用いた場合と受理日を用いた場合のそれぞれについて、Case ごとの各順位のテストデータ件数および割合を、表 4-8、表 4-9 に示す。

表 4-8 テストデータの順位（投与開始日を用いた場合）

順位	Case 0	Case 1	Case 2	Case 3
1 位	139 (65%)	139 (65%)	139 (65%)	140 (65%)
2 位	63 (29%)	64 (30%)	64 (30%)	64 (30%)
3 位	3 (2%)	2 (1%)	2 (1%)	2 (1%)
4 位以下	9 (4%)	9 (4%)	9 (4%)	8 (4%)

表 4-9 テストデータの順位（受理日を用いた場合）

順位	Case 0	Case 1	Case 2	Case 3
1 位	125 (58%)	126 (59%)	128 (60%)	130 (60%)
2 位	62 (29%)	62 (29%)	60 (28%)	62 (29%)
3 位	8 (4%)	8 (4%)	8 (4%)	6 (3%)
4 位以下	19 (9%)	18 (8%)	18 (8%)	16 (8%)

WHO-UMC の提案手法では受理日を用いていたのに対し、ここでは PMDA の副作用報告データベースの特性を考慮し、重複報告検出の精度をあげるために投与開始日を用いた。表 4-8、表 4-9 をみると、投与開始日を用いた場合にテストデータが 1 位となる割合は約 65%、受理日を用いた場合には約 60%と、投与開始日を用いたことによる重複報告検出の精度向上が認められた。また、投与開始日を用いた場合では約 95%のテストデータが上位 3 位以内の順位となった。

各 Case は計算量の観点から、重複報告の可能性のある報告のみを評価対象とするよう設定したものであるが、評価対象範囲を絞り込んだことによる Case 間での結果の差は見られなかった。したがって、Case3 の評価対象範囲で類似スコアの算出を行った場合でも、Case0 の場合とほぼ同等の重複報告検出率が得られるものと考えられる。

テストデータの類似スコア順位が 1 位とならなかった場合について報告の内容を精査したところ、1 位と 2 位がともにテストデータであった場合を除くと、これらは主に、以下に示す 3 分類のいずれかに該当した。

① 新たな重複報告の可能性あり：

テストデータより上位の報告は、類似スコアが大きく（より基準データに似ている）、新たに重複の可能性のある報告が発見された可能性がある。

表 4-10 に例を示す。1 位の報告は重複報告として副作用報告データベースに登録されていない報告だが、テストデータよりも投与開始日が基準データに近く、より類似度が高かった。報告内容を確認した結果、1 位の報告は新たな重複報告であることが判明した報告であった。

表 4-10 新たな重複報告と見られる場合の類似スコア順位表（例）

順位	性別	転帰	医薬品	副作用	第一被疑薬 投与開始日(※)	年齢	類似スコア	テストデータ
基準データ	女	回復	10 in total	2 in total	2006/1/23	44	-	
1	女	回復	9 of 10 +5	2 of 2	2006/1/23	44	52.14	
2	女	回復	9 of 10 +1	2 of 2	2006/2/6	44	42.03	*

※基準データの第一被疑薬の投与開始日

② 全体的に類似スコアが小さい：

テストデータより上位の報告であっても、判定閾値に比べると類似スコアが小さい。表 4-11 に例を示す。1 位の報告であっても医薬品の一致が 2 件のみで投与開始日は欠損値であるなど、全体的に一致している項目が少なくなっている。判定閾値は 31.0 であるから、類似スコアは約 10 小さい。確認したところ、この報告は重複報告ではなかった。

3 位や 5 位の報告は医薬品が 4 つ一致しているが、一方で投与開始日や年齢が大きく異なるためにこれらの属性に関するスコア値が負になっており、相殺されて、結果的に合計の類似スコアが大きにならない。

なお投与開始日と年齢について、いずれか片方あるいは両方が欠損値の場合、当該属性のスコア値は 0 としている。

表 4-11 全体的に類似スコアが小さいと見られる場合の類似スコア順位表（例）

順位	性別	転帰	医薬品	副作用	第一被疑薬 投与開始日(※)	年齢	類似スコア	テストデータ
基準データ	女	回復	5 in total	1 in total	2006/4/13	37	-	
1	女	回復	2 of 5 +1	1 of 1	-	37	20.93	
2	女	回復	1 of 5 +2	1 of 1	-	37	19.84	
3	女	回復	4 of 5 +7	1 of 1 +2	2006/2/22	37	18.68	
4	女	回復	1 of 5 +6	1 of 1 +1	-	37	18.16	
5	女	回復	4of5+4	1of1	2006/4/13	60	17.21	*
6	女	回復	3of5+3	1of1+1	2005/10/27	37	16.94	
7	女	回復	3of5+5	1of1	2006/4/13	60	16.20	*

※基準データの第一被疑薬の投与開始日

③ 医薬品が多い：

基準データに含まれる医薬品の数が多いため、医薬品の属性に関するスコアが支配的になりすぎている。例えば表 4-12 では性別が異なっていたり、年齢が 10 以上離れていたりする報告であってもテストデータより上位になっていた。

医薬品の属性に関するスコアは一致する医薬品の数に比例する形で増加する。そのため類似スコアの内訳において医薬品のスコアが支配的になり、医薬品以外の属性値が大きく異なっている場合、類似スコアが大きい報告が存在していると考えられる。

報告内容からも、これら上位の報告の中には重複ではない報告が複数含まれていることが確認されており、今後の課題となっている。

表 4-12 医薬品が多いと見られる場合の類似スコア順位表（例）

順位	性別	転帰	医薬品	副作用	第一被疑薬 投与開始日(※)	年齢	類似スコア	テストデータ
基準データ	男	回復	33 in total	1 in total	2005/4/12	35	—	
1	男	回復	19 of 33 +26	1 of 1	2006/1/12	35	57.76	
2	男	回復	20 of 33 +3	1 of 1 +1	2005/4/12	35	48.21	
3	男	回復	20 of 33 +31	1 of 1	2005/5/21	48	44.49	
4	男	回復	16 of 33 +21	1 of 1 +2	2000/1/1	56	42.24	
5	男	回復	13 of 33 +21	1 of 1 +3	2006/1/12	35	40.90	
6	男	軽快	15 of 33 +12	1 of 1 +3	2005/10/25	35	40.84	
7	男	回復	19 of 33 +10	1 of 1	2005/3/7	26	40.64	
8	男	回復	11 of 33 +15	1 of 1 +2	—	34	35.94	
9	男	回復	12 of 33 +13	0 of 1 +4	2006/1/12	35	35.78	
10	男	回復	12 of 33 +4	1 of 1	2005/4/21	33	33.95	
11	男	回復	14 of 33 +16	1 of 1	2005/3/25	60	33.44	
12	男	回復	9 of 33	1 of 1	2005/10/25	35	33.36	
13	女	回復	15 of 33 +10	1 of 1	2005/9/13	33	30.11	
14	男	回復	2 of 33	1 of 1	2005/4/12	35	29.68	
15	男	回復	3 of 33 +1	0 of 1 +3	2005/4/12	35	29.49	*

※基準データの第一被疑薬の投与開始日

4.3 各属性の類似スコアにもとづく重複報告判別モデル導出の検討

4.3.1 背景と目的

前節までに説明した WHO-UMC が提案している類似スコアにもとづく重複報告検出手法では、類似スコアの算出に用いた各属性のスコア値の単純和を類似スコア（総合スコア）とし、この類似スコアにもとづいて重複報告を検出する。

前節の評価実験の結果から、類似スコアの順に評価すると、多くの重複報告が上位ランクに評価されていることがわかった。ただし、上位にランクされた報告の中にも、実際には重複報告でないものが含まれていた。また、判定閾値による評価を用いると多くの報告が重複報告の候補として検出される。この中には、誤検出されたもの、すなわち本当は重複報告ではないが、重複報告候補として検出されているものが含まれる。確認作業の負荷を考えると、誤検出を低減することが望まれる。

そこで、WHO-UMC 手法が算出する各属性のスコア値を入力とし、重複報告であるか否かを判別する重複報告判別モデルを導出することについて検討した。WHO-UMC 手法が各属性のスコア値の単純和にもとづいて評価するのに対して、各スコア値を総合的に判定する判別モデルを構築することにより、誤検出の問題を改善するというのが検討のねらいである。

4.3.2 重複報告判別モデルの導出

(1) 概要

WHO-UMC 手法による重複報告検出と、重複報告判別モデルによる重複報告検出の流れについて、それぞれ図 4-8、図 4-9 に示す。評価モデルにもとづいて各属性のスコア値を算出するところまでは同じであるが、その後の検出までの処理が異なる。

WHO-UMC 手法では、各属性のスコア値の総和である類似スコアにもとづいて、判定閾値を超えたもの、あるいは類似スコア値が上位となるものを重複報告として検出する。これに対して、重複報告判別モデルを用いた検出では、各属性のスコア値を入力とし、重複報告であるか、否かを判別する重複報告判別モデルを準備し、この判別モデルの判別結果をもって重複報告の検出を行う。この重複報告判別モデルの判別の質を高めることにより、精度が高い重複報告の検出が可能となる。

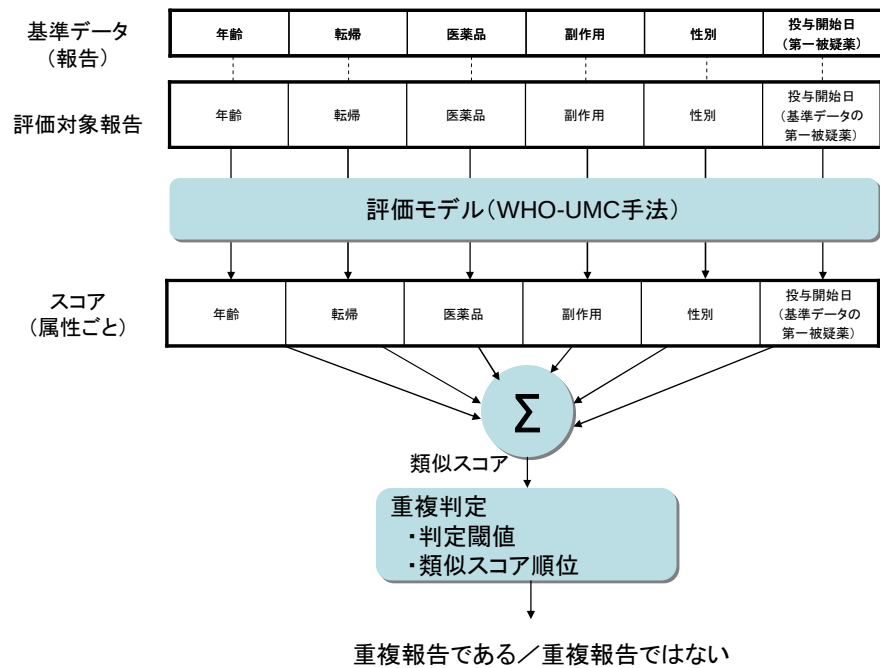


図 4-8 WHO-UMC 手法による重複報告検出

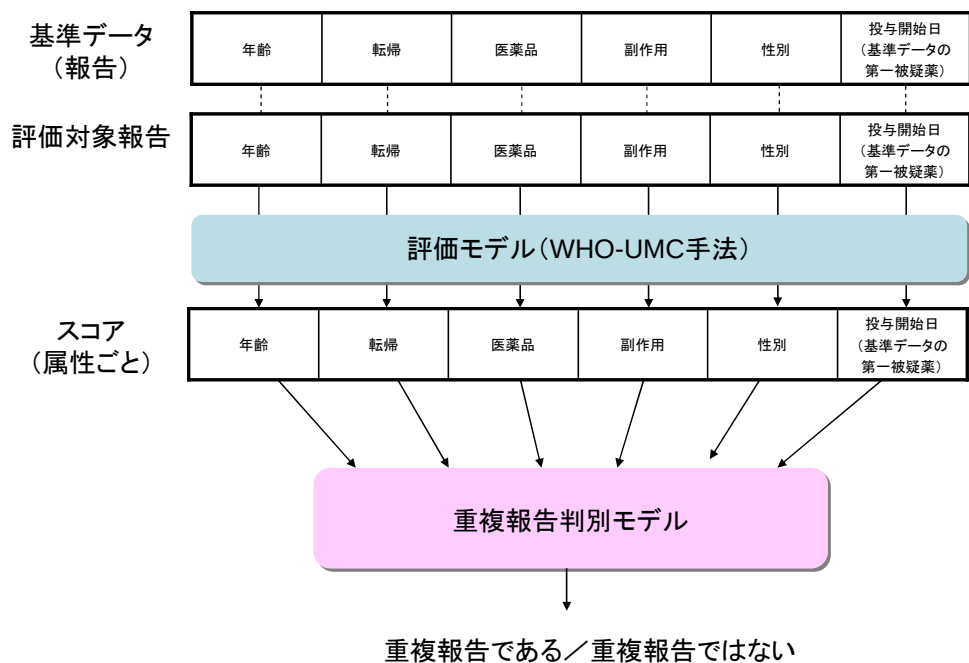


図 4-9 重複報告判別モデルによる重複報告検出

(2) SMOによる重複報告判別モデルの導出

重複報告判別モデル導出の流れを図 4-10 に示す。重複報告判別モデルの導出は、重複報告である正例と重複報告ではない負例からなる評価用データを準備し、カーネル法の 1 つである SMO(Sequential Minimal Optimization)手法^[6]を用いて学習、導出した。

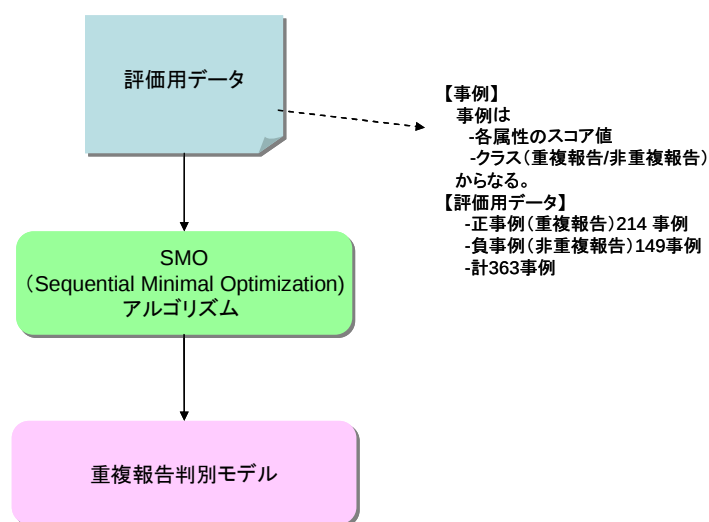


図 4-10 重複報告判別モデル導出の流れ

4.3.3 評価実験

(1) 評価用データ

評価用データとして、重複報告である正例と重複報告ではない負例を準備した。正例については、前節の評価実験で用いたテストデータ 214 事例をそのまま用いた。これらテストデータは、副作用報告データベースの情報から重複報告であることがあらかじめわかっているものである。

一方、負例については、評価実験で類似スコア順位が上位となった報告のうち、分析者が重複報告ではないと推定したものを採用した。

なお、評価用データは、正事例 214 事例、負事例 149 事例の 363 事例からなる。

また、各事例の属性、属性値として、WHO-UMC 手法により算出された以下の属性とその属性値を用いた。

- (1) 年齢
- (2) 性別
- (3) 転帰
- (4) 医薬品 (一般名)

- (5) 副作用（HLT）
- (6) 医薬品・副作用（補正後）
- (7) 基準データの第一被疑薬の投与開始日

(2) 学習アルゴリズム

評価用データから重複報告検出モデルを学習アルゴリズムにより導出する。学習する内容は、7つのスコア（数値属性値）からその事例が重複報告であるか否かというクラスを判別する2クラス分類問題である。

学習アルゴリズムには、カーネル法の中でも高速な学習が可能であることが知られているSMO(Sequential Minimal Optimization)を用いた。なお、今回の実験では、WEKA^[7]に導入されているSMOアルゴリズムを使用した。パラメータは、全てデフォルトのものを使用している。

(3) 実験方法

用意した評価用データを用いて10分割の交差検定を実施し、評価用データの重複報告の判別精度の評価を行った。また、WHO-UMC手法による検出結果との精度比較を実施した。

4.3.4 実験結果

(1) 判別精度

評価用データの分類精度を表4-13に示す。全体で92.3%と比較的高い判別精度が得られた。また、特に重複でない報告については、96.0%と高い判別精度が得られており、誤検出を低減することができる。

表 4-13 重複報告判別モデルの判別精度

		モデルによる判別結果		判別精度 (%)
		重複	非重複	
評価用 データ	重複	192	22	89.7
	非重複	6	143	96.0
合計				92.3

(2) 類似スコア順位にもとづく検出方法との比較

表 4-4 に示した Case 0 を評価対象範囲とした場合について、各類似スコア順位の評価用データに対する判別精度を表 4-14 に示す。ここでの判別精度とは各類似順位スコアの評価用データのうち重複報告であるもの（正事例）について、モデルが重複報告であると正しく判別した割合である。

表 4-14 類似スコア順位にもとづく検出との比較

類似スコア順位にもとづく検出		モデルにもとづく検出
順位	件数	判別精度(%)
1 位	139	94.2
2 位	63	88.9
3 位	3	33.3
4 位以下	9	66.6
合計	214	89.7

類似スコア順位が 4 位以下と低い 9 件のうち、6 件について正しく判別することができた。このように、類似スコア順位が低いものについても、ある程度正しく判別することができる。

(3) 判定閾値にもとづく検出方法との比較

次に判定閾値にもとづく検出方法との比較を行った結果を、表 4-15 に示す。ここで、先に述べたように判定閾値は 31.0 とし、判定閾値にもとづく検出では、閾値以上の報告は全てを重複報告、閾値未満の報告は全てを非重複報告と判別するものとしている。

表 4-15 判定閾値にもとづく検出結果との比較

		モデルにもとづく検出				判定閾値にもとづく検出		
		重複	非重複	判別精度		重複	非重複	判別精度
閾値以上	重複	152	10	93.8	94.7	162	0	86.6
	非重複	0	25	100.0		25	0	
閾値未満	重複	40	12	76.9	90.0	0	52	70.5
	非重複	6	118	96.3		0	124	

モデルにもとづく検出手法は、閾値未満の評価用データについても、約 77%の精度で正しく検出することができた。また誤検出も 4%弱と小さい。このように、類似スコアが閾値未満と小さい報告についても、正しく判別することができる。

以上のように、判別モデルを使用することにより、類似スコア順位が低い、あるいは類似スコアが小さい重複報告についても、効率よく検出することができる。

4.4 今後の検討課題

WHO-UMC の提案する重複報告検出手法を参考とし、PMDA の副作用報告データベースに対して適用した。

- WHO-UMC の手法をそのまま適用するのではなく、PMDA の副作用報告データベースに合わせて利用する属性を選択することで、重複報告検出の精度を向上できることが確認された。
- 重複報告類似スコアにもとづく評価により、データ上は重複とされていない重複報告を検出することができた。

シグナル検出手法により算出される指標値の信頼性を高めるために、データベースから重複報告を検出し評価対象から除外する手法として、重複報告検出手法のさらなる向上が期待される。

重複報告検出手法を今後実用化していくためには、以下のような項目について引き続き検討を行い、評価実験を行っていくことが望まれる。

- PMDA の副作用報告データベースに適した評価に用いる属性について、さらなる検討を行う。
- 医薬品数が多い場合について、類似スコアで医薬品が支配的になってしまう課題を解決し重複報告検出の精度を向上させる。
- 各属性のスコアの単純和を取って算出している類似スコアについて、属性ごとの重み付けを行う。
- 外部の副作用報告データベースなどの評価用データベースを整備し、それを利用した評価実験で手法の有効性を確認する。

さらに、WHO-UMC 手法により算出される各属性のスコア値にもとづき、重複報告であるか否かを判別する重複報告判別モデルの導出について検討を行った。評価用データを用いた実験を行った結果、90%以上の分類精度を得ることができることを確認した。

今後の課題として、類似スコアが小さいケースの分類性能の向上があげられる。分類精度を改善するための方法として、判別モデルの学習に用いる事例の構成等を変更することが考えられる。

5. 調査のまとめ

本年度の調査では、(1) データマイニングの安全対策業務への試適用検討として、新業務フローにおける試行実験の実施と、(2) データマイニングの高度化検討として、Ad Hoc 解析の検討と重複報告検出手法の検討を行った。

新しい安全対策業務フロー（新業務フロー）では、シグナル検出手法を詳細検討を要する報告のスクリーニングに用いる。このスクリーニングの基準を変化させながら、要詳細検討としてスクリーニングを通過する件数（量）と内容（質）について評価を行った。

今回の評価実験では、スクリーニング基準を変化させても通過する件数は大きくは変化しなかった。また、スクリーニングを通過する件数は、現在の体制でも対応可能な範囲であると見込まれた。スクリーニング基準については、今後も体制の変更等に合わせ、適宜見直していく予定である。

評価実験の実施と併せ、因果関係評価データのデータベース化を行った。データベース化された因果関係評価データは、業務支援システムに格納され、スクリーニングをはじめ、安全対策業務で活用される。

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。これにより、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される。

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と考えられる。

データマイニング手法の利用可能性に関する検討の中で、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ある。

データ前処理へのデータマイニング手法活用の1つとして、重複報告の検出を支援する手法について検討を行った。まず、既存手法である WHO-UMC が提案している手法を PMDA の副作用データベースに適用することから検討に着手した。評価実験を通じた検討から、PMDA の副作用データベースにおいても、重複報告の検出に有用であることが確認

された。ただし、重複報告の候補として誤検出される報告も多いため、業務への有効性という観点からは、この誤検出を削減するための改善が望まれる。

誤検出を改善するための方法として、WHO-UMC の提案手法が出力するスコア値から、当該報告が重複であるか、否かを判別する重複報告判別モデルを導出し、その判別結果により、重複報告を検出する方法を検討した。試行実験の結果から、重複報告判別モデルを用いることにより、誤検出の削減が期待できることが確認された。

今後は、重複報告判別モデルに用いる事例の構成を変化させることにより、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ある。

なお、ここで挙げた調査と並行して業務支援システムの開発と、これに必要なデータの整備を行った。データマイニング手法が導入された新しい安全対策業務フロー（新業務フロー）と開発した業務支援システムは、平成 21 年度より運用される予定である。

(Ⅱ) 中期計画期間のまとめと今後の検討方針

1. 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、データマイニング手法の導入を目的として検討した内容と結果の概要を以下にまとめる。

1.1 年度別調査・検討内容のまとめ

データマイニング手法を安全対策業務に導入することを目的として、平成 16 年度から 20 年度までに以下のような調査、検討を実施した。

(1) 平成 16 年度

平成 16 年度は、中期計画期間の初年度として、安全対策業務に導入するデータマイニング手法の概念検討と中期計画期間に実施する検討内容の整理を行った。

中期計画期間中に安全対策業務に導入するデータマイニング手法については、諸外国規制当局の検討、導入状況を踏まえ、PRR や ROR などに代表される統計的な指標にもとづくシグナル検出手法（基本的シグナル検出手法）を主な検討対象とするものとした。

また、データマイニング手法の検討とあわせて、データマイニング手法導入後の安全対策業務フローの検討が必要であることを確認した。

(2) 平成 17 年度

平成 17 年度は、平成 16 年度の検討結果にもとづき、既存の代表的な基本的シグナル検出手法をとりあげ、PMDA が保有する副作用報告データベースに適用することにより、シグナル検出量とシグナル検出内容の評価を行った。未知のものに限っても多くのシグナルが検出されるため、安全対策業務におけるシグナルの扱いについて検討することが必要であることを確認した。

海外規制当局におけるシグナル検出手法の導入状況、安全対策業務での利用状況を把握するため、WHO-UMC、MHRA、FDA を訪問した。WHO-UMC ではトリアージ手法、MHRA ではインパクト分析手法とよばれる安全対策業務上の検出されたシグナルへの対応優先度づけ手法（スクリーニング手法）をあわせて導入することにより、シグナル検出される量の課題に対応していることがわかった。

また、その他に SMQ (Standardized MedDRA Queries) を利用した副作用グルーピングの検討を行った。

(3) 平成 18 年度

平成 18 年度は、シグナル検出手法、基準と感度、特異度の関係の評価するために、実験・評価用データを準備し、評価実験を行った。各手法で検出されるシグナルの内容と特徴、およびシグナル指標の理解しやすさ等の観点から、使用するシグナル検出手法を、ROR、

BCPNN、GPS の 3 つとすることにした。

また、シグナル検出に使用する副作用報告データベースについて、新データベースのみと新旧両データベースを使用する場合、報告数のカウント方法として組み合わせ数ベースと症例数ベースの場合で比較実験を行った。実験結果を踏まえて、副作用報告データベースは、新旧両データベースをあわせたもの、カウント方法については、症例数ベースを採用することにした。

シグナル検出手法の高度化検討として、層別分析、MGPS を用いた相互作用分析、副作用グルーピングについて実験を行った。それぞれ特徴的なシグナルが検出されたが、業務での具体的な活用方法等については検討課題として残った。

(4) 平成 19 年度

平成 19 年度は、データマイニング手法導入後の新業務フローについて検討を行った。新業務フローでは、シグナル検出手法を、詳細検討を行うべき対象のスクリーニングに用いるものとした。これを受けて業務量を評価するために、スクリーニングの試行実験を行った。

また、オランダの Lareb、EU の EMEA を訪問し、シグナル検出手法の導入、活用状況と、安全対策業務フロー、および業務支援システム等についてヒアリングを行った。業務支援システムの機能やインタフェース、安全対策業務におけるシグナル指標値の活用に対する考え方等について、参考になる情報を得ることができ、業務支援システムの開発等に反映した。

(5) 平成 20 年度

平成 20 年度は、平成 19 年度に引き続き、新業務フローにおけるスクリーニング部分の試行を行った。試行実験では、症例評価データなど新業務フローで使用するデータを準備し、実際に即した形で実施した。試行結果から、分析者が対応することが必要となる件数は、現在の体制でも対応可能な範囲であることを確認した。

また、データマイニング手法の高度化検討の一環として、Ad Hoc 解析、重複報告検出手法の検討を行った。実験結果はそれぞれ良好なものであったが、業務導入に向けては、業務での具体的な活用方法の検討と、さらに試行を重ねることにより、利用上の留意点等についての知見の蓄積が必要であることを確認した。

なお、上記の検討と並行して業務支援システムの開発と業務支援システムで利用するデータの整備も進め、安全対策業務に導入される見込みである。

1.2 調査・検討内容のまとめ

(1) 業務へ導入するデータマイニング手法の確定

業務へ導入するデータマイニング手法確定のために、以下の検討を行った。カッコ内は、主に検討を行った年度である。

- 業務へ導入するデータマイニング手法（平成 16 年度）

中期計画期間内に導入するデータマイニングについては、諸外国規制当局が基本的シグナル検出手法の導入と活用を進めている現状を鑑みて、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た。

- データの持つ特徴を把握するための検討（平成 16 年度～18 年度）

- － シグナル検出手法の選択

基本的シグナル検出手法として、これまでの提案されているシグナル検出手法である PRR、ROR、GPS、Yule's Q、SPRT、BCPNN の 6 手法をとりあげ、実際に PMDA の副作用報告データベースを用いた評価実験を行った。シグナルの未知・既知を考慮しなかった場合ではあるが、ROR、Yule's Q 等では既存のシグナル検出基準を用いた場合、1 週間あたり 300 から 400 ほどの新規シグナルが検出されることがわかった。また、PRR と BCPNN はシグナル検出される内容が非常に類似することが確認された。

採用するシグナル検出手法については、指標の意味合いのわかりやすさの観点から ROR を、検出されるシグナルが類似する BCPNN と PRR から BCPNN を、他のシグナル検出手法とは検出するシグナルの傾向が少し異なる GPS の 3 手法を採用した。

- － カウント方法

報告数のカウント方法には、組み合わせ数ベースの方法と症例数ベースの方法という 2 つの方法がある。各方法を採用した場合のシグナル検出結果の比較を行ったが、大きな差が見られなかったため、諸外国の規制当局で採用されている症例数ベースの方法を採用することにした。

- － シグナル検出に用いる副作用報告データベース

シグナル検出に使用可能な副作用報告データベースには、新データベース（2003 年 10 月 27 日～）と旧データベース（1994 年 9 月 1 日～2003 年 10 月 26 日）の 2 種類がある。新データベースのみ、および新、旧データベースを併用した場合について、シグナル検出結果を比較した。実験の結果から、若干なが

ら感度が優れていた新データベース、旧データベースの併用を採用するものとした。

(2) シグナル検出手法の応用に関する試行

(1)に示した業務に導入する基本的シグナル検出手法の検討に加え、シグナル検出手法の応用に関して以下について試行した。業務への導入に向けては、さらなる検討と試行が必要である。

- 層別・併用薬シグナル（平成 18 年度）

シグナル検出手法を用いて、性別等の層別を対象としたシグナル検出と、複数の併用されている医薬品と副作用に関するシグナル検出について、実験を行った。層別の、あるいは併用薬の相互作用に関するシグナル検出計算を行うことにより、新たにシグナルとして検出されるものがあることが確認された。

これらのシグナル検出指標の業務での利用方法については、今後の検討課題である。

- グルーピング（平成 17、18 年度）

類似する副作用をグルーピングし、評価するための方法として SMQ の利用について検討を行った。SMQ は、将来的にはグルーピングに利用できる可能性があるが、現状では十分な数が準備されていないことを踏まえ、分析者が通常の業務で行っている医薬品、副作用のグルーピングを案件情報として蓄積し、これを利用することにした。

(3) データマイニングを用いた安全業務プロセスの検討

従来の安全対策業務のうち、通常業務として行われている部分（定常業務）に、データマイニング手法（基本的シグナル検出手法）を導入するものとし、以下のような検討を実施した。

- 安全対策業務におけるシグナル検出手法の利用（平成 17～19 年度）

上述のとおり、シグナル検出手法を業務に導入することを考えると、大量のシグナルが検出されるという点が課題となる。この課題に対応するため、諸外国の規制当局では、例えば WHO-UMC がトリアージ手法、MHRA がインパクト分析手法を採用し、検出されているシグナルのスクリーニングを実施している。

これらのスクリーニングの考え方を参考に、PMDA においても、シグナル検出手法をスクリーニング手法として活用するものとし、安全対策業務フロー中の未

知のユニットの評価プロセスのうち、1次スクリーニングに組み入れた。

- 検討手法の業務への試適用と評価結果（平成 19、20 年度）

業務フロー中の 1 次スクリーニングについて試行実験を行った。実験の結果、スクリーニングにより抽出されるコンビネーション数は、1 週間あたり 300 件弱であり、今後、業務体制を充実させることにより対応可能な量であることが確認された。また、スクリーニングにより抽出された内容のいくつかは、詳細な検討が必要なものであることが確認された。
- 業務支援システムにおける知見・データの活用（平成 20 年度）
 - － 分析者の知見の利用

分析者の知見に関するものとして、未知・既知データ、因果関係評価データ、案件（グルーピング）情報について電子データ化の準備と整備を行った。電子データ化を行うことにより、分析者間の知見の共有や安全対策業務での利用が容易となった。
 - － 医薬品流通量データの導入

医薬品流通量データを導入し、業務支援システムにおいて、分析者が参照できるようにした。

(4) 業務支援システムの開発

データマイニング手法を導入した安全対策業務の効率的な実施を支援する業務支援システムの開発を行った。業務支援システムの特徴等は以下のとおりである。

- 業務支援システムの特徴

諸外国の規制当局で使用されている業務支援システムも参考に、シグナル検出手法を導入した業務支援システムの開発を行った。

この業務支援システムでは、1 次スクリーニング部分にシグナル検出手法が導入されているほか、上述の新しい業務フローに即しており、シグナル指標値など、分析に必要な情報が参照しやすいインタフェースとなっている。
- 各種データの管理

因果関係評価を含む症例の評価データ、未知・既知データなどのデータも、業務支援システム上で入力、管理され、必要に応じて共有、利用することができる。

(5) 海外調査の実施（平成 17 年、19 年度）

平成 17 年度に、WHO-UMC、MHRA、FDA、平成 19 年度に Lareb、EMA を訪問し、海外規制当局におけるデータマイニング手法、特にシグナル検出手法の利用状況、シグナル検出手法を利用した業務フロー、業務支援システム等について調査を実施した。

シグナル検出手法を導入した業務フロー、特に多数検出されるシグナルへのスクリーニングによる対応について参考となる情報が得られた。また、業務支援システムの機能、インタフェース設計についても有益な情報が得られた。得られた情報は新しい安全対策業務フローや業務支援システムに反映されている。

調査結果のまとめについては、平成 19 年度報告書を参照のこと。

(6) 将来業務に導入するためのデータマイニング

将来的に安全対策業務に導入するデータマイニング手法として、以下の手法についての検討を行い、試行した。

● Ad Hoc 解析（平成 19、20 年度）

特定医薬品による特定副作用の発現リスクに関する参考指標として、副作用報告データベースのうち、その医薬品、副作用に関連する部分（スモールセット）に範囲を絞り込み、これに対してシグナル検出手法を適用した指標値を用いる Ad Hoc 解析の検討を行った。

いくつかの事例についてケーススタディを試行した結果、参考指標として有用であるという結果が得られている。

● 重複報告検出（平成 20 年度）

重複報告の自動検出手法について検討を行った。まず、WHO-UMC が提案している重複報告の特徴を複数の項目から数理モデル化し、このモデルが項目ごとに算出するスコアにより重複報告である可能性を評価する手法について、PMDA のデータベースを対象として検討を行った。さらに、項目ごとに計算されるスコア値から真に重複報告である報告を判別するモデルを導出し、これにもとづいて重複報告を精度よく検出する手法について検討を行った。

実用化に向けては、本格的な評価実験等が必要である。

以上のように、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、導入を行った。

平成 21 年度より、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た安全対策業務が実施される。

2. 今後の検討方針

上記のように、平成 16 年度から平成 20 年度までに、基本的シグナル検出手法とよばれるデータマイニング手法を業務への導入が実現された。今後、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、以下の検討を実施する。

- 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を業務量とスクリーニング結果（質）の観点から、再評価し、継続的に見直しを行う。

- 新たなデータマイニング手法の業務への導入検討

本中期計画期間中に業務に導入した基本的シグナル検出手法に加えて、安全対策業務を効率化、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う。検討手法の例として、以下のものを想定している。

- － 層別・併用薬分析

層別、併用薬分析については、本中期計画期間中に評価実験を試行しており、参考指標として有効であることが確認されている。また、併用薬分析のための指標として FDA が使用しており、既に試行実験を行った MGPS のほかに、WHO-UMC で Ω という新しい指標も提案されている。

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する必要がある。

- － Ad Hoc 解析

本中期計画期間に行った試行から、有効な情報が得られるという結果が得られている。

試行を重ねていくことにより、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必要がある。

- － 重複報告検出

実用化のための検討として、さらに十分な評価用、モデル学習用データを準備し、検出モデルの最適化と評価を行う必要がある。

- 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る。
- 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、開催する。

参考文献

- [1] 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 17 年 3 月
- [2] 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 3 月
- [3] データマイニング手法の導入に関する検討結果報告書, 独立行政法人 医薬品医療機器総合機構, 平成 19 年 3 月
- [4] データマイニング手法の導入に関する検討結果報告書, 独立行政法人 医薬品医療機器総合機構, 平成 20 年 3 月
※ [1] ~ [4] の報告書は、PMDA 医薬品医療機器情報提供ホームページ (http://www.info.pmda.go.jp/kyoten_iyaku/dm.html) から閲覧可能。
- [5] G. Niklas Norén, Roland Orre, Andrew Bate, I.R. Edwards (2007) “Duplicate detection in adverse drug reaction surveillance”, Data Mining and Knowledge Discovery, 14 (3), pp.305-328.
- [6] J. Platt (1998) “Fast Training of Support Vector Machines using Sequential Minimal Optimization”, Advances in Kernel Methods - Support Vector Learning, B. Schoelkopf, C. Burges, and A. Smola (eds)., MIT Press.
- [7] I. H. Witten and E. Frank (2005) “Data Mining: Practical machine learning tools and techniques”, 2nd Edition, Morgan Kaufmann, San Francisco.