

2009J7060A

厚生労働科学研究費補助金
地域医療基盤開発推進研究事業

電子化された医薬品添付情報の利活用に関する研究

平成21年度 総括研究報告書

研究代表者 土屋 文人

平成22年3月

厚生労働科学研究費補助金

地域医療基盤開発推進研究事業

電子化された医薬品添付情報の利活用に関する研究

平成21年度 総括研究報告書

研究代表者 土屋 文人

平成22年3月

目 次

I. 総括研究報告

添付文書情報を利用した医薬品データマスタの検討 1

土屋 文人

II. 分担研究報告

自然言語処理による添付文書情報の利活用

医薬品データマスタ用データベースの開発 9

木村 昌臣

III. 資料

添付文書情報を利用した医薬品データマスタの検討

研究代表者 土屋 文人 東京医科歯科大学

研究要旨

医薬品の適正使用を担保するために、オーダーリングシステム等の病院情報システムで、各種のチェックが実効性をもつようにするための方策を検討した。添付文書の記述内容が標準化されていないことから、チェックが必要な項目をフラッグ化を行った。また、過去に医薬品の名称類似による選択エラーが少なからず発生していることから、これらを防止するための方策についても検討した。また、緊急安全性情報等、厚労省から出される情報を、処方時に医師に伝達するための方策についても検討を行った。

研究分担者 木村昌臣・芝浦工業大学 准教授

A. 研究目的

我が国においてはオーダーリングシステムや電子カルテにおいて、処方データを電子的に作成することが普及している。しかしながら処方に関しては、医薬品の誤選択や用法・用量に関するエラーに起因した医療事故が少なからず発生している。現在、これらのデータチェックについては、病院情報システムに備えられた医薬品マスタが基本となっているが、その構造はオーダーリングシステムが最初に開発された約 30 年前から殆ど変更がなされていない。情報処理技術はこの間に飛躍的に進歩したが、残念ながら、それらが処方のチェックに活かされていないのが現状である。

一方、添付文書情報の電子化は一応行われ、（独）医薬品医療機器総合機構のHPをはじめとして構造化・電子化されたデータが存在しているが、これらの添付文書の情報については基本構造が定められているものの、その内容の記述に関する標準化が行われていないため、データ作成者による解釈に差異が生じることとなり、せっかく構造化・電子化された添付文書情報を医薬品データ

マスタの作成等に利活用しようとするとならずしもうまくいかないのが実情である。

本研究では、添付文書上必要なチェックをオーダーリングシステムで行うことができるように、添付文書情報の解析を行うとともに、データマスタの基本を作成する。

B. 研究方法

第一段階として、医薬品名に関する情報を構成要素として区切ることを行った。次に医薬品の適正使用を担保するために必要な情報項目及び構造について検討を行った。またそれらに該当する添付文書の記載上の問題点を明らかにするためにPMDAから入手できるSGML方式の添付文書情報データのみならず、市販されているデータも可能な限り入手して検討を行った。

また、処方時に問題となる用法については、医療機関等で使用されている用法マスタを参考に、標準用法マスタ作成のための基盤整備を行った。

また、緊急安全情報等、医薬品の適正使用に関する情報が処方を行う医師に確実に伝わるようにするために、これらの情報をオーダーリングシ

テム等に反映できるようにするための方策及び課題についても検討を行った。

C. 研究結果

I. 基本情報

基本情報としては、販売名の構造に関する検討、成分名に関する情報処理上の課題の抽出等を行った。

1-1 販売名の構成要素について

医療用医薬品の販売名は平成12年9月19日に
出された通知（医薬発935号）により以下の
ように定められている。

- (1) 保健衛生上の危害の発生するおそれのないものであり、かつ、医薬品としての品位を保つものであること。
- (2) 原則として、剤型及び有効成分の含量（又は濃度等）に関する情報を付すこと。

例) ブランド名+剤型+含量（又は濃度）

- (3) 販売名の一部の記載が省略された場合に、他に該当する製剤が存在しないこと

また、この通知は当初、新規に販売される医薬品を対象としていたが、その後既発売品に対しても適用され、「医療事故防止のための代替新規申請」として名称変更申請受付が行われ、一応昨年9月の薬価収載で販売名に関するルールの一が一段落したところである。

一方、後発医薬品については、平成17年9月に新たにブランド名を付与することが禁止され、「一般名」+「剤形」+「規格」+「会社名」の4要素から構成されるという、一般名を基本とした命名ルールが定められた（薬食審査発第0922001号）が、平成16年に日本薬局方の名称表記方法が変更になった（例：塩酸トリヘキシフェニジル→トリヘキシフェニジル塩酸塩）ことから、薬価収載されている名称を対象とすると、新旧の販売名が経過措置の関係で併存するため、デ

ータとしては下例のように一般名の表記が販売名において両者が混在することになっている。

例 リン酸コデイン散1%「タケダ」
コデインリン酸塩散1%「タケダ」

また、現時点においては販売名の構成要素である「ブランド」、「剤形」、「規格」の3要素の記載順位は必ずしも決まっていないことから、規格がブランド名よりも前に位置する販売名も存在している（例：2mgセルシン錠）ことから、販売名を構成要素に分ける場合には、各種の処理が前措置として必要であった。

1-2 販売名からのブランド部分の抽出

販売名からブランド名を抽出するためには、ブランド名が基本的にはカタカナであることに着目して、剤形を示す漢字（「錠」「散」「細粒」「軟膏」等）がある場合には、その漢字の前のカタカナ文字列をブランド名候補としてリストアップした。しかしながら、この処理では剤形がカタカナである場合（「カプセル」、「ドライシロップ」、「シロップ」、「クリーム」等）これらとブランド部分とを区分するために、剤形を登録して、これらを含む部分を削除した。また、生薬、漢方製剤は対象外とするとともに、一般名が含まれている販売名についてはブランド名がないものと判断した。

これらの処理を行った結果において、剤形や一般名的な名称を含むものが見られたので、第二次の処理として削除する用語を別途定義し、これらを除くための作業を行った。今回対象とした医薬品（19017品目）の最終処理結果を巻末資料1として掲載することとする。

1-3 ブランド名の類似性について

ブランド名の類似性については、過去に医薬品の選択エラーをはじめとして医療事故、ヒヤリ・

ハット事例が少なからず発生している。これらについては報告のあったもの等については厚労省からの通知や日本病院薬剤師会等による通知により警告が出されている。一方、我が国においては医薬品の承認時に土屋らによって開発された類似名称検索システムが活用され、類似性の高い名称は承認時にチェックを受ける仕組みになっている。しかしながら、比較的古くから存在している後発品はブランド名を維持することが許されていることと、これらの後発薬の名称は現状では新薬のようにチェックを受けていない。ここ数年の後発品推進策により、医療機関における後発品の採用は確実に進み、本年4月からの診療報酬体系の変更により、更に後発品の使用が拡大すると思われる。また、従来は先発品同士の類似性が問題となっていたアルマールとアマリールの事故の他に、後発品と先発品の類似名称（アルマトールとアマリール）による医療事故も発生した。そこで、これらの類似性に対して予め警告を出すことが必要と思われることから、1-2で抽出されたブランド名を対象として、類似名称検索システムのロジックを利用して、類似度の高い名称をリストアップすることとした。平成14年の厚労科研特別研究の報告書において指摘されていた1文字違い（edit=1）のブランド名の組み合わせリストと同様のリストを作成した。その結果を巻末資料2として掲載した。これらの情報は処方時に選択エラーを防止するための確認・警告画面を準備することで対応可能である。

1-4 ブランド名が同一の医薬品に関する注意喚起

医薬品が関係した事故及びヒヤリ・ハット事例報告ではその3分の1程度が、同一ブランド下における規格違い、剤形違いである。そこで、同一ブランドで複数規格が存在する場合、複数剤形が存在する場合には、そのことを予め明示して注意

喚起を行うことが可能である。また、平成22年4月からの診療報酬改訂により、保険薬局において類似剤形間の変更が可能になったことから、その面からも同一ブランドで複数規格、複数剤形が存在することを示すフラッグの存在が有用と思われる。そこで、複数規格、複数剤形の存在濃霧を示すフラッグをたてるよう処理を行った。

II 適正使用に必要なチェック項目

医薬品の適正使用を担保するために必要な添付文書情報項目の検討を行った。検討結果を表1に示す。

チェック項目	発生しうるエラー要因等	特記事項
禁忌薬剤	病名との関係	
	併用禁忌(相互作用)	
	小児	
	妊婦	
	授乳婦	
	高齢者	
	高度臓器機能障害患者	
	投与禁止条件(検査値)	
薬剤選択	継続投与中止条件(検査値)	
	名称類似薬品	
	ハイリスク薬警告	
	専門性の有無	
	第2選択薬	
	単独使用禁止(併用必須)	
	定期的な検査の必要性	
	同意書の必要性	
	経過措置期限超過	
	法的規制薬剤関係	
後発薬品関連		
剤形	複数剤形	
	複数規格	
	徐放性	
	腸溶性	
	割線の有無	
	粉碎の可否	
用法	投与経路	
	投与回数	
	投与時期	
	投与速度	
	休薬期間	
	適応外使用時	
用量(投与量)	用量計算違い	体表面積 体重 年齢
	制限量超過	1回制限量 1日制限量 総投与量
	常用量超過	適宜増減
	常用量下限未満	適宜増減
	制限量超過(高齢者)	1回制限量 1日制限量 総投与量
	常用量超過(高齢者)	適宜増減
	常用量下限未満(高齢者)	適宜増減
	分量の単位(対薬価基準)	
薬価基準単位との調整係数		
用量(投与日数等)	投与量の減量条件	
	新医薬品投与日数制限超過	
	保険処方可能日数超過	
	総投与日数制限	
	総投与量制限	
	適応外使用時	

表1 適正使用に必要なチェック項目内容

表 1 に示された項目の多くは基本的には、各医薬品の制限的条件の有無を示すチェックボックス（フラッグ）とその記述（添付文書の記述をそのまま引用）とで構成される。

各項目の内フラッグ処理を行う際に、最適化するために、禁忌患者情報については入手できた添付文書情報や添付文書データを利用してさらに細分化を行った。

1. 禁忌患者情報

小児、高齢者、妊婦、授乳婦を対象に添付文書上、市販データにおける表現について検討を行った。

(1) 小児について

小児が禁忌として使用されている分類としては「乳児」、「幼児」、「小児」を基本として、これに「〇歳未満」というような更に細かい分類を示す用語が付加されていた。しかしながら、「2歳以下の乳児・小児」「2歳未満の乳幼児」「2歳未満の乳・小児」というように、「乳児」、「幼児」、「小児」の用語の使用方法が標準化されていないため、電子化された添付文書情報が存在しても、添付文書の表現のままでは、一定のチェックをかけることは不可能であることが確認された。日本製薬工業協会の「使用上の注意の記載要領とその運用自主基準」によれば、「*参考：小児等に用いている年齢区分（おおよその目安）」として

小児： 15歳未満

幼児： 7歳未満

乳児： 1歳未満

新生児： 出生後4週間未満

低出生体重児： 体重2500g未満

と定めているが、実際の添付文書の記載は必ずしもこのような区分になっていない。従って、処方時のチェックとして小児を対象とする場合には、年齢を基本としてフラッグをたてることが考え

られる。現時点において実際に小児の項目に具体的な数字が出ている年齢区分としては、「3ヶ月」「6ヶ月」「9ヶ月」「2歳」「6歳」「12歳」であることと、前述に自主基準の区分とを考慮すると以下のように区分することが妥当と思われる。

低出生体重児

4週未満

3ヶ月未満

6ヶ月未満

9ヶ月未満

1歳未満

2歳未満

6歳未満

7歳未満

12歳未満

15歳未満

(2) 高齢者について

高齢者については、年齢区分を示す用語としては「高齢者」の表現の他に、「老齢の衰弱した患者」との表現が存在していた。これらから区分としては
高齢者
老齢の衰弱した患者
とすることにした。

(3) 妊婦について

妊婦については「妊婦又は妊娠している可能性のある婦人」「妊娠初期」「妊娠後期」「妊娠末期」「妊娠中」「妊娠3ヶ月以内」「妊娠12週未満」「妊娠16週未満」等の用語が使用されていた。また、これらの妊娠後の期間の用語の他に「ビタミンA欠乏症の婦人は除く」等の条件が記載されている場合もあった。そこで妊婦については、妊娠後の期間と付随する条件というような形で区分を行うこととした。

妊娠している可能性のある婦人

妊婦

初期

12 週未満

16 週未満

中期

後期

末期

その他の条件の有無

(4) 授乳婦について

授乳婦に使用されている用語としては「授乳中」「授乳婦」「授乳期」「授乳を望む母親」等の用語があった。

2. 禁忌病名情報

禁忌病名については添付文書で記載されている禁忌病名のリスト化を行った。禁忌病名の記述には「本剤の成分に対し過敏症の既往歴のある患者」の表現が最も多く見受けられた。また、「悪性症候群、横紋筋融解症又はこれらの既往歴のある患者」というように、病名が列挙された後に「これらの既往歴のある患者」という表現が多く見受けられた。電子カルテシステム等における自動チェックを考慮した場合に、この項目は病名記載欄を参照することから、既往歴の対応は特に対応しなくても病名のみでのチェックで対応可能と考え病名のリスト（資料3）でのチェックを行うこととした。

また、薬物-薬物間相互作用により投与が禁忌の場合には、添付文書では「〇〇の投与中の患者」という表現がなされている（資料4）。そこで、これらについても禁忌病名のリストに含まれる者として処理を行った。

3. 用法の標準化について

平成21年5月に厚労省医政局に設置された「内

服薬処方せんの記載方法の在り方に関する検討会」において、用法の標準化が必要であることが議論されたことから、本研究において、標準用法についても検討を行うこととした。

用法の標準化を行うために、オーダリングシステムが導入されている大学病院や一般病院8施設において使用されている用法マスタを収集し、用法の整理を行った。これに現在レセプトにおいて使用されている用法マスタ内容を併せて、標準用法マスタ(案)を作成した（資料5）。

4 用法・用量について

用法・用量は基本的には承認事項であることから、記載内容の統一が図られていてもおかしくないが、現実には、他の項目同様、様々な表現が混在していた。

大まかな基本構造でいえば、適応症があつて用量が定められていることになる。用量の記載は、「成人」といったように、年齢区分の記載が多いが、小児用医薬品においては体重あたり、抗悪性腫瘍剤等においては体表面積当たりの用量が定められている例が多かった。一般的に最も多い年齢区分による表記であっても、判断のぶれを最小限にするためには、1回当たりの用量が記載され、その回数が記載されるべきと思われる。最近承認された医薬品には殆どのものがこの形態をとっていることは、国際治験が実施されていることと無縁ではないと思われる。

しかしながら承認が古い時期であればあるほど、その表現の違いが目につくこととなる。「1日1~3錠を1~3回で・・・」といった表現はその典型的な例であり、このような文学的表現をいかに分解するかが問題となる。従来は1日量を基本とした記載が多いが、用法・用量をよりわかりやすくするためには、処方せん記載の在り方検討会で議論されたように、今後は1回量をベースとして考慮すべきである。市販のデータも多くが1日量

をベースとしているが、最近承認されている医薬品のように1回量ベースに変換すべきと思われる。

5 その他の項目

表1に示された各項目については、基本的に条件等を示すフラッグをたてることができるように、データ処理を行った。

6 厚労省から出された安全性情報等

厚労省から出された「緊急安全性情報」、医薬品医療機器総合機構から出された「安全性情報」、日本医療機能評価機構から出された「安全性情報」をベースに、これらの情報の発出の有無、発出された年月日、情報の内容等をデータあるいは画像(PDF)として活用できるように処理を行った。

これらの研究結果をふまえ、表1に示されるような処方時に必要なチェックが処方オーダーリングシステムでかかるようにするために、添付文書の記載内容をフラッグ化したテンプレートを試作した。

D. 考察

医療用医薬品添付文書は薬務局長通知により、記載項目及び記載順序が定められており、かつ記載要領が定められ、かつ記載上の一般的留意事項、各項目に関する留意事項が定められているものである。また、同じく使用上の注意の記載要領も通知により定められているものである。

しかしながら、記載されている内容そのものは、標準化が行われていないこと、また、承認された時代により、表現内容は異なっているのが現状である。また、過去ソリブジン事故に問題となったように、例えば「併用しないこと」という記載は「併用を絶対的に禁止」しているのか、「原則と

しては併用しないが、絶対的禁止ではない」のかが読み手によって解釈が分かれてしまうようなことが少なからずある。これらが医師の裁量との兼ね合いで難しい問題を含むことは理解できるが、この曖昧さが、添付文書情報の電子化において最大の障壁となっていることも事実である。

また、添付文書の著作権は基本的には製薬企業に属しているが、そのため、同一成分の薬剤であっても、使用されている用語にさまざまな点の違いが生じている。例えば「ジアゼパム錠2mg」のいわゆる標榜薬効と呼ばれる記載についても、「マイナートランキライザー」「精神安定剤」「鎮静・抗不安剤」「抗不安剤」「精神神経安定剤」というような表現がされており、意味としては同じと判断できても、電子的にはこれらが同一内容であることを担保するためには、これらの表現を全て同義語辞書として登録することが必要となる。

添付文書の記載についての法的判断としては、平成8年1月23日の最高裁で、「医薬品の添付文書に記載された使用上の注意事項に従わず、それによって医療事故が発生した場合には、これに従わなかったことにつき特段の合理的理由がない限り、当該医師の過失が推定される」とする判決がなされている。この判決に示されているように、「特段の合理的理由」がある場合には、使用上の注意に従わなかったとしても過失とは認定されないことになっているわけで、その意味で、医師の裁量権を認めつつ、一定のルールが存在の必要性を示していることになる。

今回の研究課程においても、随所に表現の違いが現れており、いわば同義語辞書造りが必要不可欠であることが確認された。

しかしながら、情報処理技術の発達により、依然殆ど不可能であった解析が、今やかなりの面で可能となったことから、添付文書のように、その解釈の違いが直接生命に危険を及ぼす可能性があるものについては、一義的な解釈が要求され

るといっても過言ではない。一義的に判断できるような構造にすることが、解釈の余地を全く無くしてしまうことには結びつかないことはいうまでもない。先の最高裁判決に示されるように、「特段の合理的理由」が存在すれば、過失は認定されないからである。

とするならば、今回いくつかの項目で行ったように、添付文書の重要部分について、記載の有無を示すフラッグをたてることや、注意の程度を示す分類区分を定め、これに従って判断可能とするような手法が有効と思われる。

添付文書に使用されている「文学的表現」をいくらか電子化（フォント化）しても、実効性は保証できないことから、従来行われてきた添付文書の電子化については、再考する必要があることを今回の研究は示している。

元来「物」と「情報」は1対1であるべきであるが、添付文書の場合は「1対N」の関係となっていることも問題といえる。製薬企業の合理化のために、添付文書そのものが「1対N」になることはやむを得ないかもしれないが、少なくとも情報は「1対1」を死守すべきである。そのためには、添付文書の著作権を有する製薬企業が、今回試作されたテンプレートに、1品目毎に添付文書の内容と同値になるような形でデータを埋めることが必要と思われる。

これらのデータを製薬企業に任せることは、データのぶれを生じる恐れがあることはいうまでもない。しかしながら、電子的なデータであれば、これを情報処理の手法を利用して統一していくことは、時間は要するかもしれないが、それほど難しいことではない。既発売の医薬品に対して、

添付文書と内容的に同値のデータシートを作成することが重要であり、その場合に全ての医薬品を対象とすることは理想であるが、まずハイリスク薬からはじめ、次いで注射剤というように段階を追って有効ではないかと考える。

もちろん、新薬に関しては承認時（あるいは薬価収載時）には、添付文書とデータシートの提出を義務づけることが必要と考える。このような形でメリハリを付けてデータシートを作成することにより、医薬品の適正使用が進展するものと考えられる。

E. 結論

添付文書の記述内容は統一がとれておらず、情報処理上でさまざまな問題があることが確認できた。医薬品の適正使用を推進するためには、今回試作したテンプレートを利用した、データシート等を利用することで、記述内容のチェックがわかり易くなると思われる

F. 健康危険情報

なし

G. 研究発表

1. 論文発表

なし

2. 学会発表

H. 知的財産権の出願・登録状況

なし

自然言語処理による添付文書情報の利活用・医薬品データマスタ用データベースの開発

分担研究者 木村 昌臣 芝浦工業大学

研究要旨

添付文書情報やこれをもとに作成されたデータベースをもとに、医薬品に関わる医療事故防止のための処方オーダーリングシステムの仕組みに資するデータの解析・調査を行った。

医薬品の販売名は、ブランド名・剤型・規格単位・用途・括弧内に含まれる文字列により構成されていることがわかり高齢者・妊婦・小児への投与情報にも表現に揺らぎがあることがわかった。これは添付文書情報をもとにしているため、元になっている添付文書自身にもそのような困難があると考えられる。また、緊急安全性情報・DSUに含まれる情報をもつ構造を同定も行った。

A. 研究目的

処方オーダーリングシステムが原因となる医療事故を防止するためには、医薬品に関わる事故防止に資する情報を処方医師に伝え、また正しい情報の入力を促す仕組みが必要である。本研究では、まず処方オーダーリングシステムにて必ず入力される医薬品名称の構造を解析し、そこに含まれる情報抽出する方法についての検討を行った。また、医薬品に関して処方時に提供すべき情報として高齢者・妊婦・小児への投与に関して添付文書を元にして作成されたデータベースにおける問題点についての検討を行い、さらに医薬品に関して医師に提供されるべき情報が含まれている「緊急安全性情報（ドクターレター）」および「DSU（医薬品安全対策情報）」内のテキストを解析し、抽出可能なデータ項目を求めた。以下、各事項についてのより詳細な説明を行う。

(ア) 医薬品販売名に含まれる情報の抽出

医薬品を処方オーダーリングシステムに入力する際には、処方として意図した医薬品名称を正しく入力できなければならない。一方で、935号通知で求められているように、医薬品販売名にはブランド名や一般名に関わる部分（以降、まとめて

語幹部分と呼ぶ）や剤型、規格単位などの情報が含まれる。先行研究ではこれらの情報に位置情報を併せて23種類の情報が含まれていることが示されている。意図した医薬品を指定するためには語幹部分の正しい同定が必要であるのは言うまでもないが、正しい投与方法と処方の整合性を保証するには剤型情報が必要であり、正しい投与量を処方情報として伝えるためには、投与すべき医薬品の規格単位を正しく指定する必要がある。これらのことから、製品名の語幹部分だけでなく他の情報も正しく使用されるために必要な情報であると考えることが出来る。処方する医師が、入力のたびに医薬品名称を支援なしにすべて手入力することは現実的でないため、これらの情報をもとに意図した医薬品を正しく入力する仕組みが必要であり、さらに、そのためには医薬品の販売名に含まれている情報を抽出する必要がある。

目視により情報を同定することも可能であるが、医薬品に含まれる文字列をまとめる観点が作業員によって異なる可能性があること、また医薬品数が2万件程度あり医薬品の名称が変更になるたびに人手で対応する必要性が生じることから、本研究では情報処理技術を利用した抽出方法の検

討を行い、その結果を利用した解析を行った。

情報を抽出するために解決する必要がある問題は以下の点にある。

- 製品名にこれらの情報が欠落しているものがあること。例えば、有効成分が複数含まれる配合剤については規格単位を製品名に付加することが困難であるが、少なくとも単味剤では規格単位を名称に付加することが出来ると考えられる。しかしながら、規格単位が付加可能な販売名の必ずしも全てに規格単位が含まれているわけではない。
- 製品名にこれらの情報が含まれていても必ずしもそれらの販売名における位置が定まっていないこと。例えば「トリアラム錠 0.25mg」と「カムリトン 0.25mg 錠」はともに規格単位 0.25mg の錠剤であるが、製品名に現れる剤型と規格単位の位置が逆転している。そのため、存在すべき位置を手がかりに情報を抽出することは困難である。

以上のために、医薬品の販売名に含まれる情報を抽出することは単純には行えない。前者については必要な情報が販売名に含まれていないことから、販売名から必要な情報を抽出するのではなく、むしろ情報を追加するための仕組みが必要となる。後者については、必要な情報が販売名に含まれているため、本研究では販売名から情報を抽出することに焦点を置くこととした。

(イ) 添付文書における有効成分情報の表現の調査

有効成分の記載方法において、化合物の薬効をもつ主たる部分のみの場合とそれ以外の塩基や酸の部分を含めた場合の場合、および無水物の場合と水和物の場合でそれぞれ含量としての表示が変わる。添付文書ではどの粒度で有効成分を表記するのかについても特定の決まりがないようにみられる。そこで、配合剤の有効成分情報の括

弧をつけて付記している部分の特に規格単位に関する部分に注目し、括弧内に「として」という言葉が含まれているものを集計してその表記ゆれについて調査した。

(ウ) 添付文書をもとに作成されたデータベースにおける高齢者・妊婦・小児への投与に関する情報抽出

処方オーダーリングシステムに提示すべき高齢者への投与に関する情報に関する知見を得るため、添付文書にもとづいて作成された医薬品使用安全性データベースの「高齢者への投与情報」の分析を行った。

(エ) 緊急安全性情報ファイルおよび DSU ファイルからの情報抽出

緊急性を要する医薬品に関する安全性情報の情報源となる「緊急安全性情報（ドクターレター）」および「DSU（医薬品安全対策情報）」は、医薬品医療機器総合機構から PDF ファイル形式で公開されている。この各ファイルに共通に含まれている情報を同定し、抽出可能なデータ項目として提案する。これにより、処方オーダーリング利用時に緊急安全性情報にもとづく注意を促すための情報を表示することが可能になると考えられる。

B. 研究方法

(ア) 医薬品販売名に含まれる情報の抽出

医薬品商品名から必要な方法をその位置にもとづいて抽出することは先述の通りであるので、各情報をもつ文字列パターンから抽出する必要がある。方法の候補となる技術として挙げられるものは次のいずれかであると考えられる。

- 形態素解析による単語抽出
- マスターデータにもとづくパターンマッチング

前者は自然言語処理によるもの、後者は情報抽出分野で扱われる手法である。前者の手法を利用するには、出現する単語があらかじめ辞書として登録されておく必要があり、かつ名称等をはじめとする複合語を分解するのは得意としない方法である。後者は、やはり出現する単語を辞書として用意する必要があるが、辞書に含まれる単語を文字列パターンとして合致するか否かをみる方法であれば複合語などであってもそこに含まれる情報が何かを特定することが可能である。

そこで本研究では、規格単位や剤型など各情報として含まれるべきものをマスターデータとしてあらかじめ用意し、これと文字列として合致するものを探す手法により、医薬品商品名に含まれる情報を同定・整理する方法をとる。

ただし、単にマスターデータに含まれるデータ文字列ごとにひとつひとつ医薬品名称と照合を行うと処理時間がかかるために工夫が必要となる。そこで、マスターデータの文字列は医薬品商品名の部分文字列となるものであるため、その文字列長は比較的短いことを考慮にいて、まず自然言語処理で用いられる N-gram という手法を用いて医薬品商品名を部分文字列に分解する(図 1)。N-gram とは、対象の文字列を先頭から長さ N の文字列とりだし、さらに先頭を一文字ずつずらしながら同様に長さ N の文字列を取り出していく手法である。ここでは、N を 1 から商品名の文字列の長さまで変化させて得られた部分文字列を対象とし、それぞれとマスターデータの照合を行う。ただし、マスターデータには、例えば「注射液」とその略語である「注」のようにマスターデータの部分文字列自身もマスターデータに含まれることがある。もし商品名に「注射液」という文字列が含まれていれば、上述の部分文字列の取得方法から必ず「注」にも合致してしまう。このことから、マスターデータと合致する文字列が他の合致する文字列の部分文字列となる場合はこ

れを結果から外す処理も行った。

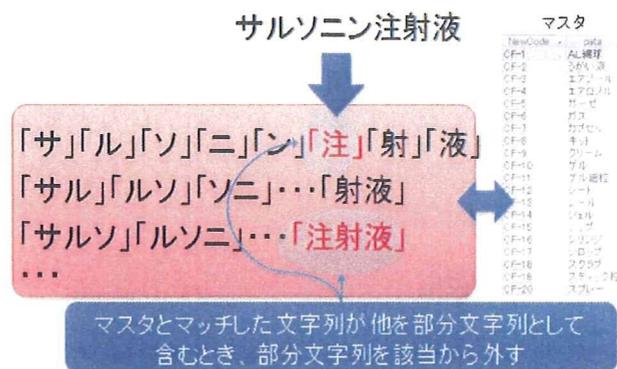


図 1 N-gram による合致文字列の抽出方法

合致した文字列のうち、他の部分文字列になっているものを対象外とする。

(イ) 添付文書における有効成分情報の表現の調査

配合剤の添付文書をもとにつくられたデータベースに対し、括弧部分を取り出すが「日局」等の有効成分に直接関係ない語句を省くため、「として」という文言が入っているもののみを対象とし、これを抽出し、集計を行った。

(ウ) 添付文書をもとに作成されたデータベースにおける高齢者・妊婦・小児への投与に関する情報抽出

高齢者への投与についての情報には 52 種類あり、「安全未確立」「十分考慮」などおおよそ二つの単語から構成されている傾向がみてとれる。そのため、これを構成するふたつの単語に分割し、それぞれの関係についての情報を、Web グラフを利用して整理した。

(エ) 緊急安全性情報ファイルおよび DSU ファイルからの情報抽出

緊急性を要する医薬品に関する安全性情報の情報源となる「緊急安全性情報 (ドクターレター)」および「DSU (医薬品安全対策情報)」は、医薬品医療機器総合機構から PDF ファイル形式

で公開されている。この各ファイルに共通に含まれている情報を同定し、抽出可能なデータ項目として提案する。これにより、処方オーダーリング利用時に緊急安全性情報にもとづく注意を促すための情報を表示することが可能になると考えられる。

C. 研究結果

(ア) 医薬品販売名に含まれる情報の抽出

本研究ではマスターデータとして規格単位マスター、剤型マスター、用途情報マスター、括弧内文字列マスターおよび記号マスターを作成した。規格単位マスターには、MEDIS 標準医薬品マスターの「規格単位」データを括弧等の記号により分割し、さらに「数値+数値以外」のパターンを持つ文字列を抽出したデータを用いた。ただし、医薬品名称に現れる規格単位には数値のみで単位が見つからないものがあるため、該当パターンの文字列から数値のみを抽出したのも規格単位マスターに含めた(図 2)。

剤型マスターは、平成 20 年度厚生労働科学研究「医薬品辞書データ項目と基準(ICHM5)に関する研究」の成果である医薬品辞書データにおける「基本剤形」データをもとに作成した(図 3)。

935 通知の趣旨から、本来であれば「語幹+剤形+規格単位」が望ましい組み合わせではあるが、後発品ではこれ以外に企業名が入る場合があり、また剤形として直接「注射液」と書かずに「注射用」と書く場合がある。これらの情報も拾うために、規格単位マスター・剤形マスター以外に、用途マスター、括弧内文字列マスターと記号マスターを用意した(図 4、図 5、図 6)。

記号マスターには MEDIS 標準医薬品マスターの「公示名称」データから漢字・片仮名・平仮名・アルファベット、数字以外の文字を抽出した文字群を、括弧内文字列マスターは記号マスターの括弧に相当するもので囲まれている文字列群(後発

品の製薬企業名など)に含まれる)を、用途マスターは MEDIS 標準医薬品マスターの「公示名称」をデータ漢字・片仮名・平仮名・アルファベット・数字・その他の文字の 5 つの文字種単位で切り分け、「用」で終わる漢字文字列群(ただし「用」の直前の文字が別の文字種であればそのあとに「用」を付けた文字列とした)を収集して用いた。また、それぞれのマスターにはデータに対応したコード番号を付加し、どの情報が医薬品名称に含まれているのかをコード番号に基づき把握できるように工夫している。

id	data
U1008	6.6mg
U1009	60
U101	0.25mL
U1010	60μg
U1011	60ピスター
U1012	60倍
U1013	60%
U1014	60,000
U1015	60,000単位
U1016	600
U1017	600μg溶解液局注射
U1018	600万
U1019	600万国際単位
U102	0.25mg
U1020	600国際単位
U1021	6000
U1022	6000単位入
U1023	6000国際単位
U1024	60000
U1025	60000単位
U1026	600MBq

図 2 規格単位マスター (抜粋)

id	data
F0	うがい液
F1	エアゾール
F10	シップ
F100	静注
F101	静注液
F102	顆粒
F103	髄注
F104	AL綿球
F11	シリンジ
F12	シロップ
F13	シロップ用
F14	シート
F15	シール
F16	ジェル
F17	スクラブ
F18	スティック粒
F19	スプレー
F2	エアロゾル
F20	ゼリー
F21	ゾル
F22	チュアブル錠
F23	チンキ

図 3 剤型マスター (抜粋)

id	data
P0	ろ過用
P1	カントップ用
P10	人用
P100	顆粒含嗽用
P101	顆粒懸濁用
P102	鼻用
P103	鼻腔用
P104	mL用
P11	今津消毒用
P12	内服用
P13	内用
P14	内痔核内注射
P15	冠動注用
P16	動注用
P17	医療用
P18	印消毒用
P19	口腔用
P2	ケフレックスシ
P20	含嗽用
P21	吸入用
P22	噴霧用
P23	坐剤小児用
P24	塩酸塩内用
P25	塩酸塩注射用

図4 用途情報マスター (抜粋)

id	data
E0	あすか
E1	わかもと
E10	イナバタ
E100	バイエル
E101	バクスター
E102	ヒカリ
E103	ヒサミツ
E104	ヒシヤマ
E105	ビケン
E106	ビケンCAM
E107	ビケンF
E108	ビケンHA
E109	ファイザー
E11	イヌイ
E110	ファルマー
E111	フェリング
E112	フジ
E113	フジモト
E114	フソー
E115	フゼミ

図5 括弧内文字列マスター (抜粋)

このマスターデータに基づき、商品名に含まれる情報を抽出した結果の一部をエラー! 参照元が見つかりません。に示す。なお、同一マスターデータの複数のデータと合致した部分文字列が複数ある場合、コロンで結合して表示している。この図では「抱水クロラール」の「水」に剤型の「水」が該当していることを除けば、それぞれのマスターに該当する情報が抽出出来ていることがわか

る。ランダムに1% (203件)を抽出し、マスター該当部分を商品名から取り除きカタカナ部分を抽出したところ、85.7%は正しく抽出されていた。精度はコードをもつテーブル等を今後改善していくと考えられる。

id	data
S0	
S1	X
S10	II
S11	III
S12	V
S13	VII
S14	VIII
S15	IX
S16	\
S17	{
S18	}
S19	[
S2	α
S20]
S21	[
S22]
S23	.
S24	#
S25	(
S26)

図6 記号マスター (抜粋)

プロバリン原末		F45	原末
プロモバレリル原末「ヨシダ」			
プロムワレリル原末「JG」			
プロモバレリル原末「マルイシ」		F45	原末
抱水クロラール		F58	水
抱水クロラール「ホエイ」		F58	水
エスクレ坐剤「250J」	U666	250	F49 坐剤
エスクレ坐剤「500J」	U654	500	F49 坐剤
エスクレ注射用キット「500J」	U654	500	F6 キット
ユーロジン散1%	U218	1%	F56 散
ユーロジン1mg錠	U532	1mg	F97 錠
エスタゾラム錠1mg「アメル」	U532	1mg	F97 錠
ユーロジン2mg錠	U705	2mg	F97 錠
エスタゾラム錠2mg「アメル」	U705	2mg	F97 錠
ベシジュールカプセル10	U290	10	F3 カプセル
ダルメールカプセル15	U431	15	F3 カプセル
ベシジュールカプセル15	U431	15	F3 カプセル
ネルボン散1%	U218	1%	F56 散
ベンザリン細粒1%	U218	1%	F83 細粒
ネルロレン細粒1%	U218	1%	F83 細粒
ベンザリン錠2	U535	2	F97 錠

図7 照合結果 (抜粋)

以上で得られた対応関係をもとに、規格単位や剤型などの頻度分布を求めた。

①規格単位の分布

規格単位を含まないものを除き、規格単位情報を集計したものが図8である。100mg、10mg、50mg、5mg等が比較的多い傾向がみられる。②で述べるように錠剤が多いことから、さらにその

中でこれらの規格単位が比較的多く用いられると考えることができる。

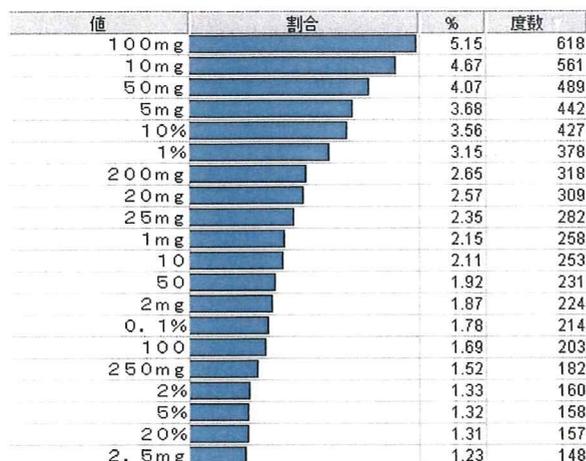


図8 規格単位分布 (上位)

②剤型の分布

剤型情報を含まないものを除き、剤型情報を集計したものが図9である。錠剤が4877品目でもっとも多く、次いで注射液(注を含む)が1794品目等と多いことがわかる。



図9 剤型分布 (上位)

③単味剤の規格単位

配合剤のリスト(1732品目)および漢方薬(YJコードの先頭が5であるもの。2734品目)を除いた単味剤14603品目のうち、約3200品目に規格単位がついていない(\$null\$となっているもの)ことがわかる。

④単味剤の剤型

③で規格単位が商品名にない品目の剤型を赤

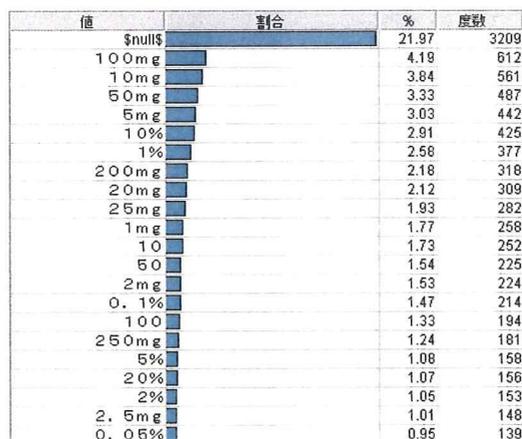


図10 単味剤の規格単位分布 (上位)

色に、あるものを青色にして単味剤の剤型の集計を行った(図11)。その結果、剤型が液体であるもの(注射液、水、油など)比較的規格単位がついていないものが多いことが伺える(ただし、点眼液や静注には規格単位がないものが少ない)。また、剤型の情報がないものは規格単位もついていないものが多い傾向もみられる。

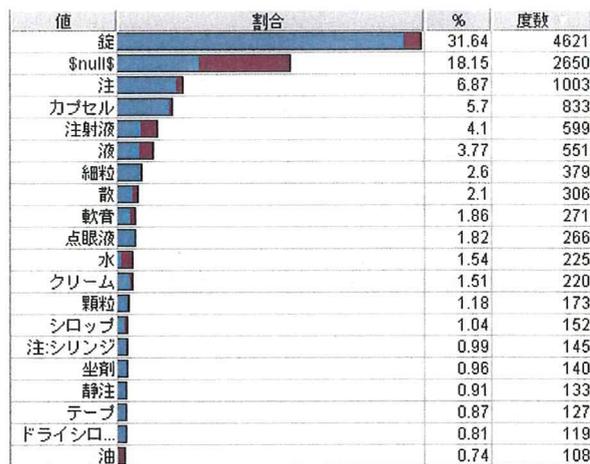


図11 単味剤の剤型分布 (上位)

赤色は規格単位がついていないもの、青色は規格単位がついているものを示す。

⑤配合剤の剤型

配合剤については935号通知では「原則として、

配合剤にあたっては含量の記載はしないこと」としている。単味剤と同様、配合剤（漢方薬を除く）について規格単位情報の有無と剤型の関係性を求めた。規格単位情報が含まれている割合は多いものの、腹膜透析液・透析剤には規格単位が含まれている割合が他剤と比べて多いこともみられる。

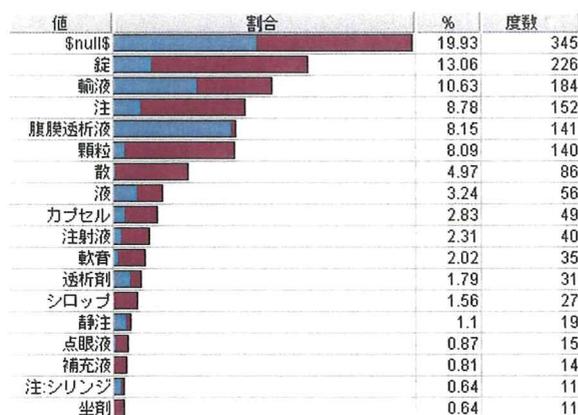


図 12 配合剤の剤型分布（上位）

赤色は規格単位がっていないもの、青色は規格単位がついているものを示す。

⑥用途情報

用途情報を含まないものを除いた、各用途情報を含む品目数分布が図 13 である。

注射薬や外用薬等の用法が多いが、それ以外にも医療用・小児用・内用等も比較的多く用いられていることがわかる。



図 1 用途情報分布（上位）

⑦括弧内文字列情報

括弧内文字列マスターに合致しないものを除き、括弧内文字列マスターのいずれかのデータに合致する品目数の分布を求めたものが図 13 である。図 13 に現れているものの多くは、後発品を発売していると考えられる企業名であるが、M、F、R、NP など英字が多く現れることも特徴的である。ただし、本手法ではこれらの英字は括弧内外によらず品目数に算入するため、括弧内に記述されているこれらの英字の実数はこれより少ないと考えられる。特に、M や F 等は語幹に相当する部分に現れる英字記号と合致している場合がある。これは、英字が略語を含め様々な意味を指していることを示しており、手法からくる問題のみならず、医薬品の商品名において意味が明確に定義されていない（または各社が独自に意味を決めてしまっている）記号が用いられているということにも起因する問題である。これを解消するためには商品名の語幹部分に現れる英字の意味についての標準を決める必要があると考えられる。



図 13 括弧内文字列情報分布（上位）

(イ) 添付文書における有効成分情報の表現の調査

得られた結果の上位のものを図 14 に示す。明らかにわかる表現の揺れは「乳酸ナトリウムとして」と「乳酸ナトリウムとして 1.120mg」のような記載方法にみられる。前者は乳酸ナトリウムとしての含量のみの記述であるが、後者は乳酸ナト

リウムとしての含量に加えて乳酸ナトリウム液としての含量について記述されている。

この点では「乾燥重量として」「無水物として」なども同様であり、何を基準に含量を表現するのかが決まっていなかったためにこのような表現をとらざるを得ないとも考えられる。

乾燥重量として	37
乳酸ナトリウムとして	36
酸化アルミニウムとして224mg	24
グリチルリチン酸として40mg	17
L-システイン塩酸塩として20mg	12
サルメテロールとして50μg	12
無水物として	10
酸化アルミニウムとして2.24g	9
グリチルリチン酸として	9
シラスタチンとして	7
パントテン酸として7.5mg	7
乳酸ナトリウムとして1.120g	6
トロンビンとして	6
乳酸ナトリウムとして0.448g	6
L-乳酸ナトリウムとして(C3H5NaO3	6
パントテン酸として15mg	6

図 14 括弧内に含まれる「として」を含む配合剤の有効成分（6件以上のもの。右の数字は該当する表現をもつYJコードにもとづく）

(ウ) 添付文書をもとに作成されたデータベースにおける高齢者・妊婦・小児への投与に関する情報抽出

図 15 より、「特記なし」・「特になし」、「低用量開始」・「少量開始」、「～注意」・「～留意」・「～考慮」、「使用」・「投与」、「用法」・「用法等」、「用量」・「用量等」は、類似の内容をもち、表記の揺れが存在することがわかる。これはもともとの添付文書に同様の表記揺れが存在することによるため、添付文書から情報を取得することができるためには少なくとも記述する語彙についての標準を決める必要がある。

また、調整等の用法・用量に関する注意、腎・肝の機能低下についての注意、観察を行うなどの慎重な投与の必要性などの情報が含まれていることがわかる。

これらについては、そこに含まれる概念は必ずしも独立したものではないため、まずはこれらを分類したうえで、独立した概念としてどのようなものがあるかを整理する必要がある。またさらにその上で、それらの概念がただ一つ該当する保証もないため、複数の概念と各医薬品の情報を結びつける必要がある。そのためには、例えば各概念に対応するフラグを用意し、そのオン・オフをもとに高齢者への投与に関する注意点を表現するなどの工夫が必要になると考えられる。

なお、妊婦・小児についても(図 16,図 17)同様の傾向が見られたため、以上のことは高齢者だけではなく、これらの対象者への投与方法についても同様の仕組みが必要であると考えられる。

(エ) 緊急安全性情報ファイルおよび DSU ファイルからの情報抽出

緊急安全性情報については、手順 6. で得られる、名詞が含まれる文節から名詞部分を取り除いて得られる助詞相当部分については、図 18 に示すように「を」「が」「に」が最も多く、次いで「は」「と」等が続くことがわかる。図 19 は、「を」「が」「に」「は」を対象にこの順にそれぞれの助詞が現れるように出現順序を正規化してから頻度を集計したものである。その結果、「～ことがある」「処置を行う」「投与を中止する」「～ことを確認する」「観察を十分に行う」などの頻度が比較的多いことがわかる。(頻度は、対象全文書で現れる回数を表す)

値	割合	%	度数
を		23.85	192
が		17.39	140
に		15.03	121
、		12.3	99
は		3.48	28
と		2.36	19
には		2.36	19
で		1.61	13
について		1.61	13
」		1.49	12
の		1.24	10
）		1.24	10
から		1.12	9
後		1.12	9
により		0.87	7
）を		0.75	6
において		0.62	5
も		0.62	5

図 18 緊急安全性情報ファイルに含まれる名詞を含む文節から名詞を削除し抽出された文字列

値	割合	%	度数
を		22.09	1589
が		21.34	1535
、		15.1	1086
に		14.26	1026
は		2.45	176
には		2.25	162
の		2.02	145
と		1.75	126
で		1.74	125
により		1.45	104
は、		0.93	67
」		0.9	65
や		0.79	57
として		0.76	55
において、		0.72	52
的に		0.7	50
には、		0.47	34
において		0.46	33
）が		0.46	33
では		0.42	30

図 20 DSUファイルに含まれる名詞を含む文節から名詞を削除して抽出された文字列

値	割合	%	度数
こと[が] ある		4.88	21
処置[を] 行う		3.49	15
投与[を] 中止する		3.26	14
こと[を] 確認する		2.33	10
観察[を] 十分に 行う		2.33	10
慎重[に] 投与する		1.63	7
状況[を] 確認する		1.63	7
異常[が] 認めらる		1.63	7
〔用法・用量〕に 関連する		0.93	4
慎重[に] 検討する		0.93	4
投与[を] 場合[に] 中断する		0.93	4
必要[に] 応じる		0.93	4
十分[に] 観察する		0.93	4
排便[が] ある		0.7	3
時間[を] かける		0.7	3
誤開塞[を] 起こす		0.7	3
後[に] 投与する		0.7	3
検査等[を] 行う		0.7	3
腸管穿孔[を] 起こす		0.7	3
使用[に] 際す		0.7	3

図 19 頻度の上位に現れる「を」「が」「に」「は」を含む文節とそれが係る動詞の組み合わせ

また、DSUについても図 20 に示すように、名詞が含まれる文節から名詞部分を取り除いて得られる助詞相当部分については緊急安全性情報(図 1)と同様の傾向が見られた。この上位に現れる「を」「が」「に」をこの順番で現れるように正規化して集計したものが図 21 であり、「～ことがある」「処置を行う」「投与を中止する」「観察を十分に 行う」「異常が認め(られ)る」という緊急安全性情報と同じものが上位に現れる一方、「可能性がある」「おそれがある」など、症状発生等の可能性について示唆する文言も比較的上位に現れていることがわかる。

値	%	度数
こと[が] ある	7.39	205
処置[を] 行う	4.02	155
投与[を] 中止する	3.32	120
報告[が] ある	3.27	126
異常[が] 認めらる	2.59	100
観察[を] 十分[に] 行う	2.49	96
可能性[が] ある	1.35	52
おそれ[が] ある	1.19	46
症状[が] あらわれる	1.11	43
慎重[に] 投与する	1.06	41
効能[を] 有する	0.98	38
必要[に] 応じる	0.88	34
〔用法・用量〕に 関連	0.65	25
本剤[を] 投与する	0.65	25
安全性[が] 確立する	0.54	21
血中濃度[が] 上昇する	0.52	20
副作用[他]に 分類する	0.41	16
併用[に] する	0.39	15
こと[が] 報告する	0.36	14
作用[を] 減弱する	0.34	13

図 21 DSUファイルに含まれる文中の「を」「が」「に」を含む文節とそれが係る動詞の組み合わせ

D. 考察

(ア) 医薬品販売名に含まれる情報の抽出

医薬品の販売名は5つのマスターデータに含まれる語句および語幹部分で構成されていると考えられることが出来る。ただし、記号については、括弧等情報を区切るために用いられることが多いため、語幹部分・剤型・規格単位・用途情報・括弧内文字列情報が(それぞれの位置情報を除いて)医薬品を一意に特定するために販売名に含まれるべき情報として必要であることがわかる。即ち

- 語幹部分[必須]
- 剤型[必須]

- 規格単位（含量）[単味剤は必須]
- 用途情報[用法等を付記する必要がある場合]
- 括弧内文字列情報

ただし、括弧内文字列情報自体は「括弧の中にある文字列」という分類であるため、多くは（後発品の製造もしくは販売を行っている製薬会社と考えられる）企業名であるが、今後はこれに含まれる情報の整理および英字記号についての標準化を行う必要がある。

(イ) 添付文書における有効成分情報の表現の調査

複数の粒度の記述方法が併存しているため、このような情報をコンピューターシステム内に持たためにはまずどの粒度で情報を持つのかを決める必要があり、そのためには添付文書においても記述の統一が望まれる。

(ウ) 添付文書をもとに作成されたデータベースにおける高齢者・妊婦・小児への投与に関する情報抽出

添付文書をもとに作成されたデータベースに含まれているデータも、その拠り所は添付文書の記述であるため、表現の揺れなどはもとの添付文書の表記揺れに端を発していると考えられる。実際、添付文書の記述のしかたに対して特定のルールがあるわけではないため、記述の自由度は大きく、同じ意味を表現するために異なる文言が使用されることがあるため、このことを結果は反映していると考えられる。

高齢者・妊婦・小児等の投与に関して配慮すべき点をもとに医薬品を検索する場合、このような表記の揺れがあることにより検索の方法を変える必要がある。すなわち表現に揺れがなければ、検索項目の値の一致・不一致のみを調べればよいが、表現に揺れがある場合はそのように単純にはいかず、予め同義語辞書を作成し、置き換えを行

うなどの処理を行う必要がある。同義語辞書の作成は表現の自由度が大きければそれだけ困難になるため、対応は難しくなる。

添付文書自体は、医薬品の第一義的な情報源であるから、そこに含まれる情報は処方オーダーリングシステムを含む様々なコンピューターシステムで活用されるべきであるが、このような表現の揺れを含む、表現の大きな自由度が却ってそこに含まれる情報の活用を阻害していると考えられ、コンピューターシステムでのそのような情報の活用のためには、情報の構造を調査する必要があると考えられる。

(エ) 緊急安全性情報ファイルおよび DSU ファイルからの情報抽出

緊急安全性情報・DSU のそれぞれについて得られた結果をまとめると以下のような情報が含まれていると考えられる。

- 考えられる異常・症状
 - 異常が認め（られ）る / 症状が現れる
 - 可能性がある / おそれがある / ことがある
- 報告の有無・内容
 - 報告がある
- 処置の要/不要・方法
 - 処置を行う / 慎重に投与する / 投与を中止する
- 確認・観察の要・不要
 - ことを確認する / 観察を十分に行う

よって、処方オーダーリングシステム上で緊急安全性情報等を提示する場合、多く言及されているこれらの情報についての項目を少なくとも持つ必要があると考えられる。さらに、本手法では対象外としたが、読点（、）が含まれる文節を目視で確認したところ、症状等の列挙および「ため」「ところ」「場合」など、条件を表す文言と結びつい