

厚生科学研究研究費補助金

厚生科学特別研究事業

医薬品等副作用症例報告における重要シグナル自動認識
及び抽出のための論理構造とIT化に関する研究

平成 13 年度 総括・分担研究報告書

主任研究者 久保田 潔

平成 14 年 4 月

目 次

I. 総括研究報告	
医薬品等副作用症例報告における重要シグナル自動認識及び抽出 のための論理構造とIT化に関する研究	----- 1
久保田 潔	
(資料) 予備的プログラム CD-ROM	
II. 分担研究報告	
1. データベースの構造と解析手法との関係の検討に関する研究	----- 24
小出 大介	
III. 研究成果の刊行に関する一覧表	----- 36
IV. 研究成果の刊行物・別刷	----- 別添

厚生科学研究費補助金（厚生科学特別研究事業）
総括研究報告書

医薬品等副作用症例報告における重要シグナル自動認識及び抽出のための論理構造とIT化に関する研究

主任研究者 久保田 潔 東京大学

研究要旨 国が収集する副作用情報など副作用に関する症例報告が集積されたデータから、特に重要な情報を選び出す「シグナル検出」をわが国の環境下で効率的に実施するために、これに必要なシステムに組みこみ可能な予備的プログラムを作成した。また、関連する問題についての研究を実施した。本研究では、特に、近年英米規制当局やWHO Uppsala Monitoring Centre (UMC) で実利用の段階にまで研究が進められてきた「データマイニング」の方法に着目し、これらの新しい方法によるシグナル検出を可能とする予備的プログラムを厚生労働省の副作用情報処理システムに組み込み可能な形で作成した。ただし、厚生労働科学研究として実施される研究である点に鑑み一般パソコンユーザーが十分利用できるものを作成した。また、関連テーマ1として、「E2B/M2仕様の電子的な個別症例安全性報告からのシグナル検出のための情報抽出」をとりあげたが、この研究は分担研究として実施された。ほかに、関連テーマ2として「ICH国際医薬用語集 (Medical Dictionary for regulatory activities terminology、MedDRA) のシグナル検出への利用」をとりあげた。作成された予備的プログラムは、個別症例安全性情報から抽出される「DRUG-EVENT発生回数データ」の存在を前提にしており、このデータから、米国FDAのGPSの方法で計算して得られたSignal Scoreによって薬-副作用をランクづけし、さらに、特定の薬-イベントについて英国MCAのPRRの方法で当該イベントに関して報告された全ての薬に関するPRR対カイ二乗値のデータをプロットしたグラフで検索し、また、UMCの方法であるBCPNNのInformation Componentとその95%信頼区間の経時的推移をグラフ的に検討することを可能とするものである。機能検証のために厚生労働省がWebを通じて公開している平成10-12年度の報告副作用一覧データ合計109207件を用いた予備的検討を行い、結果を相互比較し、BCPNNの方法が最も厳しい基準であり、BCPNNでシグナルと認識される薬-副作用の96%以上が他の方法でもシグナルとして認識されることなどを明らかにした。「DRUG-EVENT発生回数データ」は、個別症例安全性情報から必要項目を抽出して作成されるものであるが、関連テーマ1に関する分担研究では、「DRUG-EVENT発生回数データ」作成のために抽出するべき項目のほか、3例目に達した未知の重篤な反応を選び出すなど、旧来のシグナル検出のために抽出しておくべき項目についても検討した。厚生労働省のシステムに本予備的プログラムを組みこむ際には、英米規制当局やUMCの新しい方法だけでなく、旧来の方法によるシグナル検出の方法も実施できることが想定されており、その際には分担研究で明らかにされた結果が参考になると考えられる。また、関連テーマ2ではシグナル検出をMedDRAを有効に利用するためには、基本語 (PT) 以上の階層に現在とは別のシグナル検出のための階層構造を作成される必要性が示唆されるとともに、シグナル検出への利用の問題が未知反応と既知反応の識別の問題と密接に関連すること、MedDRAに比べると安定性などシグナル検出という意味ではより適していると考えられるWHO-ARTを適切に利用することを検討すべきことなどを明らかにした。

分担研究者 小出 大介
国際医療福祉大学
助教授

A. 研究目的
国が収集する副作用情報は年々増加の一端をたどっており、年間55,000症例を上回

る状況となっている。これらの報告はすべて重篤または未知の副作用情報であり、重要な情報の見落としがあつては医薬品の安全対策の根幹が揺らいでしまう。万全を期するためには、情報の処理について最新の電子的処理を導入し、自動化を図ってミスを防止する必要がある。これが、欧米規制当局で研究が進められているシソーラス、論理的構造の研究や実証実験など、高度な情報科学技術による集学的な研究にもとづいて、特に重要な情報を選び出すシグナル検出="Prioritization"である。この手法を導入することにより、副作用情報の処理及び解析を迅速かつ効率的に行うことが可能となり、医薬品等の安全対策を速やかに講じることへの寄与が期待される。

ただし、シグナル検出の研究は、現在急速に新たな発展を遂げつつある分野であり、本研究は、わが国の副作用情報処理のあり方が今後の先進国で進むであろう急速な展開に取り残されないために最低限必要と考えられる、英米規制当局とWHOのUppsala Monitoring Centre (UMC)で実利用の段階にまで進んできた「データマイニング」の手法を副作用のシグナル検出に活用する方法論を、我が国の副作用情報解析に利用するためにも利用可能とすることを中心的課題として据えた。

本研究では、これらの新しい方法によるシグナルを検出を可能とする予備的プログラムを厚生労働省が副作用情報処理のために構築してきたシステムに組み込み可能な形で作成する。したがって、本研究は厚生労働省の業務強化への寄与を一義的な目的とするものであるが、本研究が厚生科学研究費で実施される研究である点に鑑み、単に厚生労働省におけるシステムへの組み込み可能性だけでなく、民間の企業や研究者が自ら有するデータを解析したり、既に行行政によって公開された自発報告から検出されたシグナルを検討する上でも役立つよう、一般のパソコンユーザーが十分利用できるものとするにも留意する。ただし、商用プログラムの開発を目指すものではなく、ユ

ーザーには使用にあたって、本プログラムの計算プロセスを十分理解できる程度にシグナル検出について学習した上で、自らの責任の下に本プログラムを適宜利用することを求めるものとする。

さらに、本研究で作成される予備的プログラムを十分に機能させるためには、今後解決しなければならない問題がいくつか存在するとの認識に立って、関連する問題についても検討を加える。

今後解決すべき本研究で作成される予備的プログラムに関連する問題は多岐にわたるが、主要な問題には以下の二つが含まれる。

関連テーマ1：今後大勢を占めると考えられるICHで合意されたE2B/M2仕様の電子的な個別症例安全性報告からのシグナル検出のための情報抽出に関する問題を明らかにする。

関連テーマ2：E2B/M2でもその使用が前提になっているICH国際医薬用語集

(Medical Dictionary for regulatory activities terminology, MedDRA)を用いて報告されたイベントからシグナル検出を効率よく行うときに必要な用語の集計方法を明らかにする。

本研究は、これら関連する問題も扱い、最終的結論を出すに至らなくても、今後の研究の発展に資する成果を出すことを目標とした。このうち第一点目は、「国際的動向を視野に入れた医薬品安全性情報の電子的伝達システムに関する研究」班(平成13-14年度厚生科学研究医薬安全総合事業)の班員でもある分担研究者(小出大介)によって実施された。

B. 研究方法

シグナル検出のための予備的プログラム作成：

以下の3つの方法にもとづき、厚生労働省の副作用情報処理システムに組み込み可能な予備的プログラムを作成した。

1) 米国Food and Drug Administration (FDA)が用いるGamma Poisson Shrinker (GPS) Program(American Statistician 1999;53:177-190)

2) 英国Medicines Control Agency (MCA) が用いるProportional reporting Ratio (PRR) (*Pharmacoepidemiol. Drug Safty* 2001;10:483-496.)

1) WHO Uppsala Monitoring Centre (UMC)が用いるBayesian Confidence Propagation Neural Network (BCPNN) (*Eur J Clin Pharmacol* 1998; 54:315-351) 及び *Computational Statistics and Data Analysis* 2000;34:473-493)

以上の方法はいずれも「データマイニング」の手法にもとづく新しい方法である。これに対し、旧来のシグナル検出のための論理（たとえば、3例目に達した未知・重篤な副作用症例を特に重要なものとしてとりあげる）については、その論理構造は比較的単純であるため、予備的プログラムを厚生労働省のシステムに実際に組み込む時に機能を付加することを想定し、予備的プログラムでは特段の考慮をしなかった。

関連テーマ1：E2B/M2仕様の電子的な個別症例安全性報告からのシグナル検出のための情報抽出

分担研究報告書（「データベースの構造と解析手法との関係の検討」）参照
関連テーマ2：ICH国際医薬用語集

(Medical Dictionary for regulatory activities terminology, MedDRA) のシグナル検出への利用 MedDRA/Jの内容検討、特に各バージョン間の異同の比較検討と、本予備的プログラムを「医薬品情報提供システム」

(<http://www.pharnasys.gr.jp/>) に公表されている「副作用が疑われる症例報告に関する情報」の平成10-12年度の「報告副作用一覧」からえた集積データに応用してえられた計算結果を検討した。

(倫理面への配慮)

本研究は、厚生労働省が収集した自発報告データを倫理面に十分配慮した上で国が公表した連結不可能な匿名化がされたデータを二次的に利用して実施されており、倫理面での問題は存在しないと考えられる。

C. 研究結果

シグナル検出のための予備的プログラム作成：

システムの概要：

本プログラムは下記の特長を有する。

(1)解析のためのデータは、Drug、Event、Drug-Eventでの発生回数、サブクラスの4項目から成るデータを用意する。

(2)世界的に有用と考えられているシグナル検出の3手法を組み合わせて解析でき、解析結果は一覧表やグラフの形で作成されるので、結果の分析が容易である。シグナルの検出では3手法を組み合わせて解析できる。

図1に、予備的プログラムの概要を示す。シグナル検出に使用すべきE2B/M2仕様の電子的個別安全性情報の項目についての詳細は分担研究報告のC1.1「PRR, GPS, BCPNNおよび国内(旧来)に共通して必要な項目」、C1.2「GPSに必要な項目」に示されているが、予備的プログラムでは、これらの項目に関して情報が抽出されたデータ（「DRUG-EVENT発生回数データ」、図1）の存在が前提にされている。プログラム言語にはVisual Basic (VB)およびExcelのマクロ言語であるVisual Basic for Application (VBA)を使用した。

「DRUG-EVENT発生回数データ」は、各薬とイベントの組み合わせについての報告件数を、さらに報告年度や患者の性・年齢などでサブクラスにわけた（層別化した）データである。

本プログラムは、厚生労働省のシステムに組み込まれることを想定しており、本プログラムが組み込まれた場合に想定される将来の副作用報告データの処理の流れを示したのが図2である。この場合、本予備的プログラムで前提にされている「DRUG-EVENT発生回数データ」は、E2B仕様の電子的個別症例安全性報告を集積されて作られるデータベースである「副作用発生DB」から作成される。また、予備的プログラムを図示した図1と異なり、厚生労働省のシステムへの組込みを想定した図2では「DRUG-EVENT発生回数デー

タ」から旧来のシグナル検出の基準（たとえば新たに3例目に達した未知・重篤症例）に沿ってシグナルを検出することも想定されている。また、将来のシステムではGPSの結果またはPRRの結果を示す画面から特定の薬-イベントの組み合わせを指定することにより、該当する対象症例にもどり、その一覧が表示されることが想定されている。

図3に本予備的プログラムの計算の流れについての概要を示す。「DRUG-EVENT発生回数データ」（図3A①、予備的プログラムではマイクロソフトアクセスのテーブルの形式を想定）から、指定された条件にそってイニシャルテーブル（②）が作成される。このイニシャルテーブルと薬剤およびイベントマスタを用いて、DRUG辞書、EVENT辞書、STARTA辞書とイニシャルデータファイルが全てテキストファイルとして作成される。これらのデータを用いて、まずFDAの方法であるWeb (<ftp://ftp.research.att.com/dist/gps>) から無償で提供されているGPSプログラム (GPS.exe) を起動する。また、今回、一連の計算の流れを”Seamless”に行うために試みに作成した、SASの Interactive Matrix Language (IML) を用いて、GPS.exeと同様の結果を自動で得ることも可能とする。いずれかの方法で計算されたGPSアウトプットファイル（⑦テキストファイル）をマイクロソフトエクセルのシート (MS-Excel) に変換する。図3Bに示した通り、同時にイニシャルデータからPRRに関する計算とBCPNNに関する計算を実施し、計算結果をワークシート（⑩PRR、⑪BCPNN）として示すとともに、このワークシートのデータからグラフ（⑭PRR、⑮BCPNN）を作成して表示する。

図3で①から⑦の記号で示したテーブルやファイルの一覧を示したのが図4である。

また、図1、図2にも示したメインページの詳細を図5Aと図5Bに示す。なお、図1に示したように、予備的プログラムが前提にする「DRUG-EVENT発生回数データ」はメインページ図5Aにある②の層化項目が指定された上で抽出されて作成される

データであり、予備的プログラムでは動作イメージが示されるにとどまる。

図5Aと図5Bに示したメインページでは、抽出にあたっての層化項目を指定し、解析にあたっての環境設定を行い、GPS.exeまたはSAS/IMLを実行してGPSの方法に従って計算を行い、その結果を表示する。図6にGPS.exeを実行した時に作成されるテキストファイルをエクセルファイルに変換する指定ボックスを示す。

図7に環境設定画面を示す。この画面で「DRUG-EVENT発生回数データ」から副作用や層が異なる薬剤を1つにまとめた1薬剤が1レコードに示された薬剤情報 (Drug辞書)、同様に1副作用が1レコードに示された副作用情報 (EVENT辞書)、1層が1レコードに示された層別項目情報 (STRATA辞書)、各薬剤-副作用ごとに層別項目情報を考慮して計算された期待値と実測値が1レコードに示された基本情報ファイル (Initial Dataファイル) およびGPSの計算結果のテキスト名およびエクセルファイル名を指定する。

図8にGPSの計算結果の画面を示す。PRR (図9) とBCPNN (図10) については、この画面のボタンを押すことによって、その都度図11に示す計算式にしたがって求めた結果を閲覧することができる。

「医薬品情報提供システム」 (<http://www.pharmasys.gr.jp/>) に公表されているデータの解析:

「医薬品情報提供システム」 (<http://www.pharmasys.gr.jp/>) に公表されている「副作用が疑われる症例報告に関する情報」の平成10-12年度の「報告副作用一覧」には、平成10年度9636件、平成11年度58565件、平成12年度41006件、合計109207件の副作用報告が公開されている。副作用名はMedDRA/JのV4.0の基本語 (Preferred Term, PT) を用いて表示されているが、薬-イベントの組み合わせ数は51232組に上る。MedDRAのV4.0は14287のPTを含むが、実際に使われているPTは1668語にすぎない。この理由は主に、平成10-12年度の副作用については、その多くが当初J-ARTでコードされており、それがいずれかの時点でMedDRAコードに機械的に変換されたためと考えられる。

本研究は単年度の研究であり、計算のための予備的プログラム作成を中心課題としたため、計算結果自体の検討は十分実施しえていないが、実施した範囲内で計算された値について以下に概括する。

薬-PTの組合せでみた場合、Signal Score 51232組には0.00から606.06までの値が計算され、FDAにならってSignal Score >2をシグナルの基準とすると16869件がこの条件をみたす。これに対し、PRRは報告数が1-2件のものおよび、あるイベントが報告されている薬が1種類だけである場合にはその薬-イベントについてのPRRの値を計算しないため、値が与えられるのは7285件であり、この7285件に0.02-23400の値のPRRが計算され、PRR >2、カイ二乗値 >4のシグナルの基準を満たすものは4367件である。WHO UMCのBCPNN基準であるIC-2SDはSignal Scoreと同様に51232組全てに最小-5.3、最大5.2の値が計算されるが、IC-2SD >0の基準をみたすものは、3942件である。

Signal Score >2の基準をみたす16869件のうち、PRRの基準を満たすものは25%の4244件、BCPNNの基準をみたすものは22%の3780件、3つの基準全てを満たすのは22%の3777件であった。

PRR >2、対応するカイ二乗値 >4、報告数 >2の英国MCAで使用されている「シグナル」の基準を満たす4367件のうち、Signal Score >2の基準を満たすものは97%の4244件、BCPNNの基準を満たすのは88%の3860件、3法全ての基準を満たすのは86%の3777件であった。

BCPNNの基準をみたす3942件のうちSignal Score >2の基準をみたすのは98%の3860件、PRRの基準を満たすのは96%の3780件、3法全ての基準を満たすのは96%の3777件であった。

以上をまとめると、上記3つの方法のうち、BCPNNが最も厳しい基準であり、BCPNNでシグナルとされるもののほとんどが、他の2法でもシグナルとして検出される。ついで、厳しいのがシグナル検出の基準に報告数 >2が含まれるPRRであり、PRRでシグナルの基準を満たすもののほとんどはSignal Score >2の基準も満たす。

最も厳しい基準であるBCPNNで指標として使われるIC-2SDは、報告数が相当多くないと高い値はとらない。たとえば、報告数が1,2,3,4の薬-PTの組み合わせの最大のIC-2SDはそれぞれ-1.25、-0.21、0.42、0.88であり、BCPNNのシグナルの基準はIC-2SD >0なので、本研究で用いたデータを用いる限り、報告数が2をこえないとシグナルは検出されず、これはPRRのシグナルの基準の一つである報告数 >2と結果的に一致する。これに対し、報告数が1,2,3,4の薬-イベントの組み合わせに関するSignal Scoreの最大値はそれぞれ6.0、83.2、151.8、185.0であり、報告数1でもSignalと認識されることがありうる。PRRの方法では報告数が1または2のものについては値を計算しないが、シグナルの基準をみたす報告数が3または4で最大のPRRの値はそれぞれ23400（全PRRのうちの最大値）、4549であり、報告数が小さくとも、PRRは大きな値をとることがありうる。

関連テーマ1：E2B/M2仕様の電子的な個別症例安全性報告からのシグナル検出のための情報抽出

分担研究報告書（「データベースの構造と解析手法との関係の検討」）参照

関連テーマ2：ICH国際医薬用語集

(Medical Dictionary for regulatory activities terminology, MedDRA)のシグナル検出への利用

表1にシグナル検出に関する有用性という観点から考えた現在のMedDRAのもつ問題点と将来解決されるべき問題をまとめた。表1に示した通り、MedDRAはシグナル検出に使用できるように設計されておらず、必要な語の迅速な追加や問題点の迅速な修正などのMedDRAの長所と評価されている点の多くが、シグナル検出に関しては短所として浮かび上がった。

MedDRAは5層構造をなしているが、高位語 (High Level Term, HLT) や高位グループ用語 (High Level Group Term, HLGT) などのPT以上の階層に位置する語はシグナル検出のために設計された語ではない。たとえば1998-1999年度のデータを使用した場合、薬-PT、薬-HLT、薬-HLGTの組み合わせ数は33732、24536、

18373であり、1998-2000年度のデータを使用した場合、それぞれ、40817、29008、24429であり、HLGTを使用してもその数は半分にもならない。したがって、今回検討したデータを用いる限りにおいては、HLT、HLGTの使用は十分に「しほりこみ」の機能を果たしていると評価することはできない。

今後の課題としては、PT以上の階層に現在とは別のシグナル検出のための階層構造を作成する必要性が示唆された。また、シグナル検出と密接な関連をもつバージョンコントロールについては、下層語 (Lowest Level Term、LLT) に関する情報の重要性が示唆された。一般に、下層語はオリジナルの報告をコードする作業で有用ではあるが、規制当局などへの報告ではPTで十分であるとの意見が強いが、MedDRAでは既存のPTが廃止されたり、廃止後新しいPTに変更されることがしばしば起こり、この結果、異なるバージョンのMedDRAを使って報告された自発報告の中には、見かけ上異なるが内容的には等価の複数のPTが混在することになる。これに対し、全てのLLTは、その時点では使用が勧められないことを示すllt_currencyの値が”N”に変更されても、LLTとしては永久保存され、かつ、最新のバージョンではllt_currencyの値が”N”のLLTを含めて全LLTに最新のPTとの関係が示される。したがってLLTとバージョンが指定されれば、全ての報告を最新のバージョンのPTで抽出しなおすことが可能である。

以上の関係を例示するために、表2にLLT「会話障害」とLLT「言語障害NEC」のバージョンによる時間的推移をとりあげた。後者の「言語障害NEC」の英語表記は”Speech disorder NEC”であるが、この日本語訳は平成12年9月のV3.2において、「会話障害NEC」に変更された（英語表記は不変）。平成13年6月のV4.0では、LLTとPTの関係に変更が加えられ、「会話障害NEC」はPTから削除され、LLTに「格下げ」になり、llt_currencyの値は”N”に変更された。代わって「会話障害」がPTに「格上げ」された。このため、LLT「会話障害」でコードされた報告をPTに

変換したときのpt_kanjiは、V3.1までは「言語障害NEC」、V3.2とV3.3では「会話障害NEC」、V4.0以後は「会話障害」になる。即ち、異なる時期にPTに変換されて報告された語を集計すると、LLTは同一であるにもかかわらず、日本語PTは3種類の異なる語が混在することになる。PTの英語表記または、そのコード番号を使用した場合でもPTの集計データにはV3.3までの”Speech disorder NEC” (10041468)とV4.0からの”Speech disorder” (10041466)が混在することになる。これに対し、LLTが「会話障害」であることと、PTに変換する際に使用するMedDRAのバージョンが指定されれば、全てのLLT「会話障害」は指定されたバージョンの単一のPTに変更されることが可能である。

仮に、今後“Speech disorder”が廃止 (llt_currencyの値がNへ変更) され、これにかわる新しいLLTが採用されても、その2つの語が同一のPTに属するとの指定がなされる限り、この方法（オリジナルのレポートのLLTと最新のMedDRAにおけるLLTとPTの関係の指定を用いて最新版のMedDRAのPTとしての抽出する）を用いれば、ある用語が廃止されても、廃止された古い語と廃止された語に相当する新設された語を同一のPTとして抽出することは可能である。

D. 考察

シグナル検出のための予備的プログラム：本研究班の最大の目標は、副作用情報から、シグナルを検出するための予備的プログラムの作成であり、その目標は達し得た。

本システムの特徴は、世界的に有用であることが確認されているPRR、GPS、BCPNNの3つの方法を同一プログラム内で計算・比較することが可能である点である。これまで一部でBCPNNとPRRの比較、PRRとGPSの比較を行った研究は存在するが、3者を計算するシステムは存在していなかった。

今後本システムを有効に使うためには多くの課題が残っている。たとえば3つの方法をどのように組み合わせると最も有効なシグナル検出が可能であるかなど、

その運用の在り方の研究をさらに行い、その結果にしたがって、システムを改良する必要がある。また、新しいシグナルを収集する立場にある企業や規制当局などでは、新しくえられた報告について、その都度一つ一つ検討するプロセスも重要と考えられ、その場合、当該の薬または同効のグループの薬について、あるいは当該のイベントまたは、そのイベントと類似するイベント群についてしばってデータを表示することなども今後本システムの「使い勝手」を改良するために重要な事項と思われる。

関連テーマ1：E2B/M2仕様の電子的な個別症例安全性報告からのシグナル検出のための情報抽出

分担研究報告書（「データベースの構造と解析手法との関係の検討」）にもあるが、本研究で作成されたシステムはPRR、GPS、BCPNNという3つの新しい方法による解析が中心であり、予備的プログラムではこの3つの方法に必要な項目について抽出したデータの存在が前提にされている。その他の項目については、旧来のシグナル検出の方法に必要な項目である。旧来の方法で重要な情報としては、因果性を判断する際に必要な併用薬や原疾患に関する情報のほかに、重篤であるか否かも重要である。また、関連テーマ2でとりあげられるべき問題でもあるが、未知か既知かの別（添付文書に載っているか、外国で同じ症例があるか、PDRに書いてあるか）に関する情報、類似の薬で同じ副作用が出ているかの区別などの項目も重要と考えられる。分担研究報告書に必要なと考えられる項目は列挙してあるが、今後、それぞれの項目がどのように使用されるべきかの詳細を明らかにする必要がある。

関連テーマ2：ICH国際医薬用語集

(Medical Dictionary for regulatory activities terminology, MedDRA)のシグナル検出への利用

表1に示した問題点を解決するために、MedDRAのシグナル検出における有効利用に関する研究を続けることが必要である。シグナル検出の問題は未知か既知かの判断の問題と密接な関係を有する。

未知か既知かの判断で重要なのは、MedDRAのPTとしては異なり、みかけ上表記の違う語であっても、同一の副作用と考えるよい場合があり、これをどのように識別するかという問題である。ここで、同一の副作用と考えることが通常許される複数のMedDRAの用語は、シグナル検出においても単一の語としてまとめられるべきであろう。

本研究では、結果の項に記述した通り、異なるバージョンの語を同一のPTとして抽出するために、蓄積される症例に付せられたLLTが重要であることが示唆されたが、シグナル検出のために、PT以上の新たな階層構造を作成する問題は依然として今後の課題として残されている。

シグナル検出に使用する用語については、今後の研究によるべきところが大きいであるが、長期間に蓄積されたデータを使用するので、あまり変化が激しくないものであることが望ましい。MedDRAがリリースされてからも、WHO Uppsala Monitoring CentreではWHO-ARTのメンテナンスを続けており、UMCではMedDRAで報告された語をWHO-ARTに変換して、UMCのデータベースに蓄積してシグナル検出などに用いている。この作業に伴い、WHO-ARTの語数は徐々に増加しているが、その速度はより緩慢であり、シグナル検出に使う用語がもつべき安定性の要件を満たす。さらに、WHO-ARTに一度採用された語が通常廃止されない点はMedDRAのバージョン管理のあり方と根本的に異なり、頻繁にPTが廃止されたり、廃止後これに代わる新たな語が新設されるMedDRAに比べて一度採用されたPTが原則として廃止されないWHO-ARTはシグナル検出により適していると考えられる。今後MedDRAのシグナル検出への利用にあたっては、WHO-ARTの利用が大きな課題になると考えられる。

E. 結論

我が国の副作用情報解析に使用する英米規制当局やWHO Uppsala Monitoring Centreで実利用が開始されている「データマイニング」の手法によりシグナルを検出するシステムを作成した。本システムは厚生労働省における副作用情報解析

のためのシステムに組み込みを想定した予備的プログラムであると同時に、企業や研究者が自ら保有するデータを解析したり、厚生労働省から公開されているデータの解析結果を閲覧することに使用することも可能である。

同時に、関連テーマ1として、現在安全性情報伝達の主要手段となりつつあるE2B/M2仕様の電子的個別症例安全性報告からシグナル検出するために抽出すべき項目を検討した（分担研究参照）

また、関連テーマ2として今後のシグナル検出へのMedDRAの利用に関する問題点を抽出した。シグナル検出へのMedDRAの利用の問題は、未知反応と既知反応の区別の問題と密接な関係をもち、解決方法を今後研究する上でWHO-ARTの応用が重要となる可能性が高いことが結論づけられた。

F. 健康危険情報

今後、本研究で成果として作成されたシステムで副作用情報の解析を実施し、新たな健康危険情報を得る可能性は否定できないが、現時点では特記すべき事項はない。

G. 研究発表

本研究で作成された予備的プログラムは、平成13年度厚生労働科学研究「患者による副作用早期発見のための適切な情報の収集及び提供の在り方に関する研究」の分担研究「自発報告からのシグナル検出の方法論に関する研究」によって明らかになった成果にもとづいて作成された。上記分担研究は本研究の主任研究者によって実施されており、以下の研究発表が行われた。

1. 論文発表

久保田潔 自発報告からのシグナル検出：英国MCA、米国FDA、WHOの新しい方法。薬剤疫学6(2)：101-108, 2001.

2. 学会発表

Kiyoshi Kubota. A tentative Overview of Signal Detection Methods Currently Developed in Western Monitoring Centers: Are the Epidemiological Methods Obsolete? Drug Information Association, The 1st Annual Workshop in Japan for Global Pharmacovigilance. November 1-2, 2001, Tokyo, Japan.

H. 知的財産権の出願・登録状況
該当例なし

図1 予備的プログラムの概要

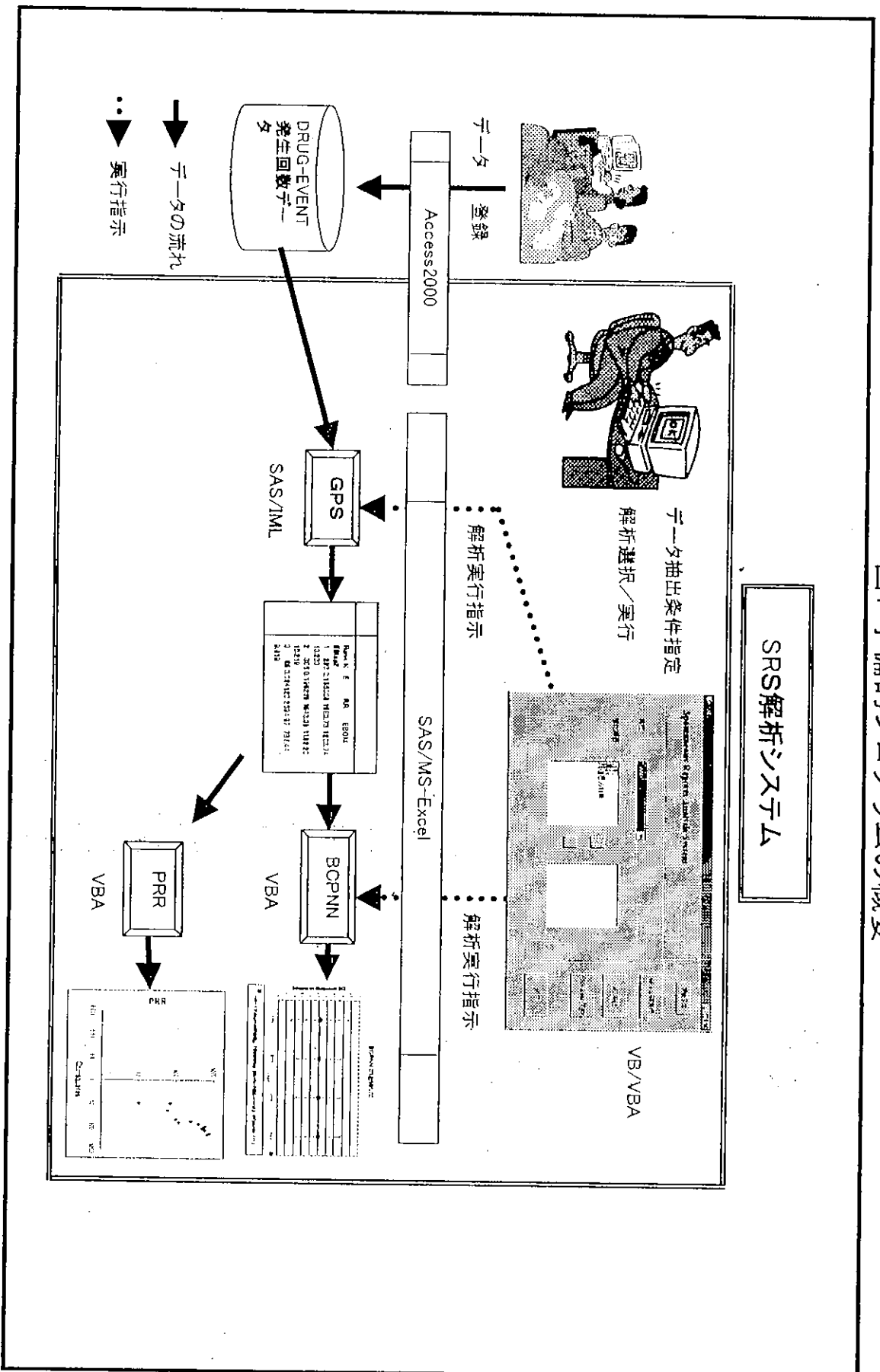


図2 SRS解析システムの将来像

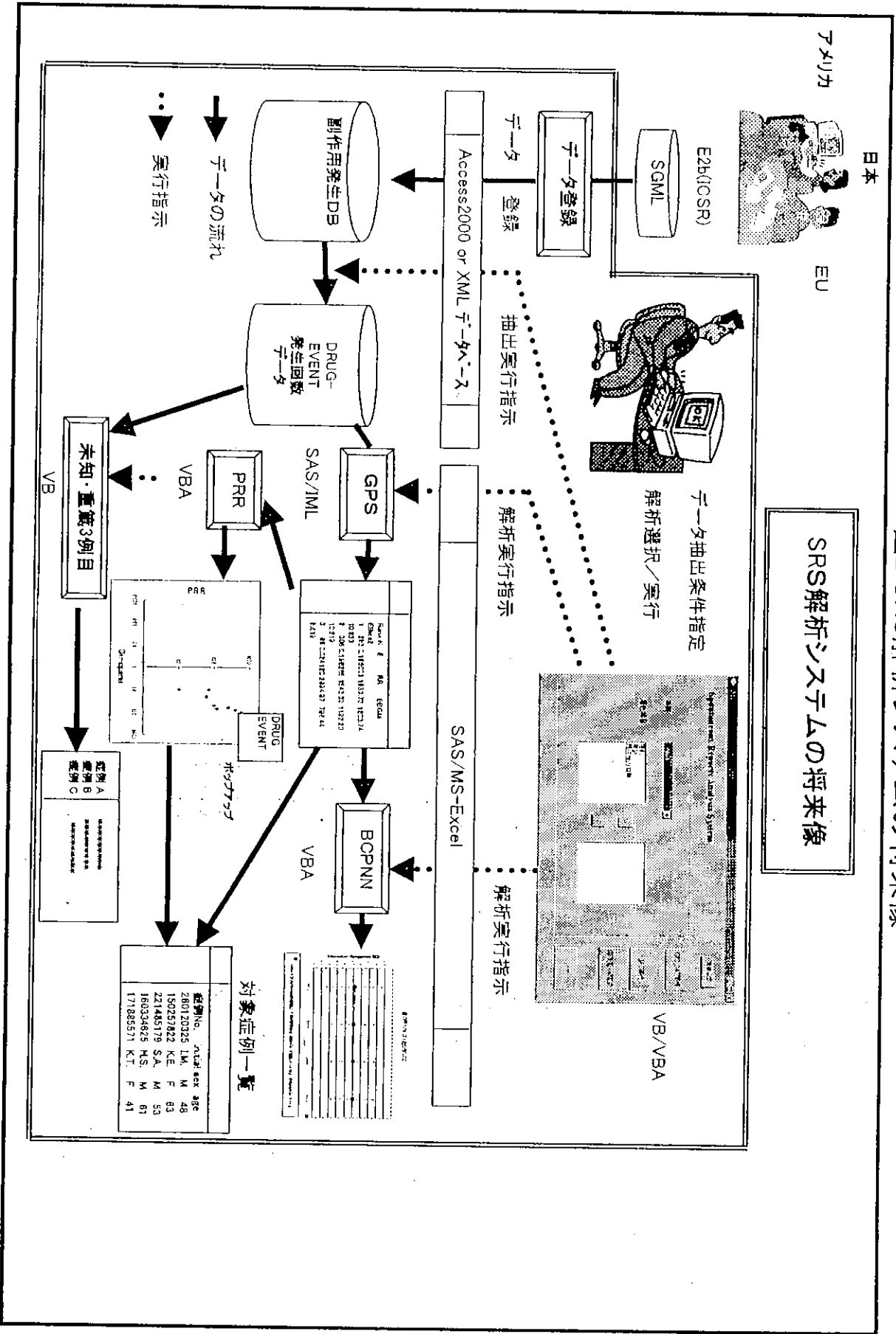


図3A 計算フロー概要(1)

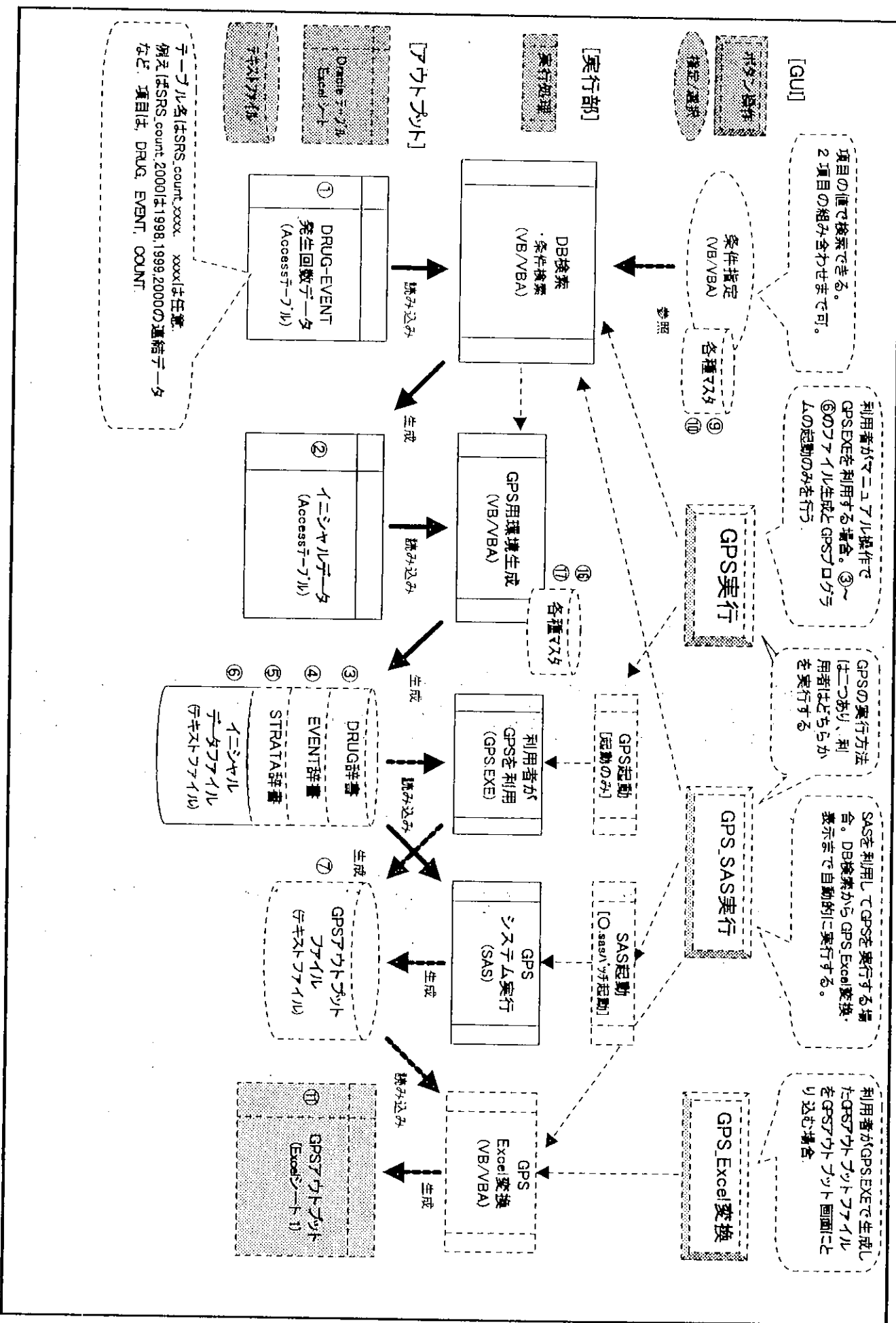


図3B 計算フロー概要(2)

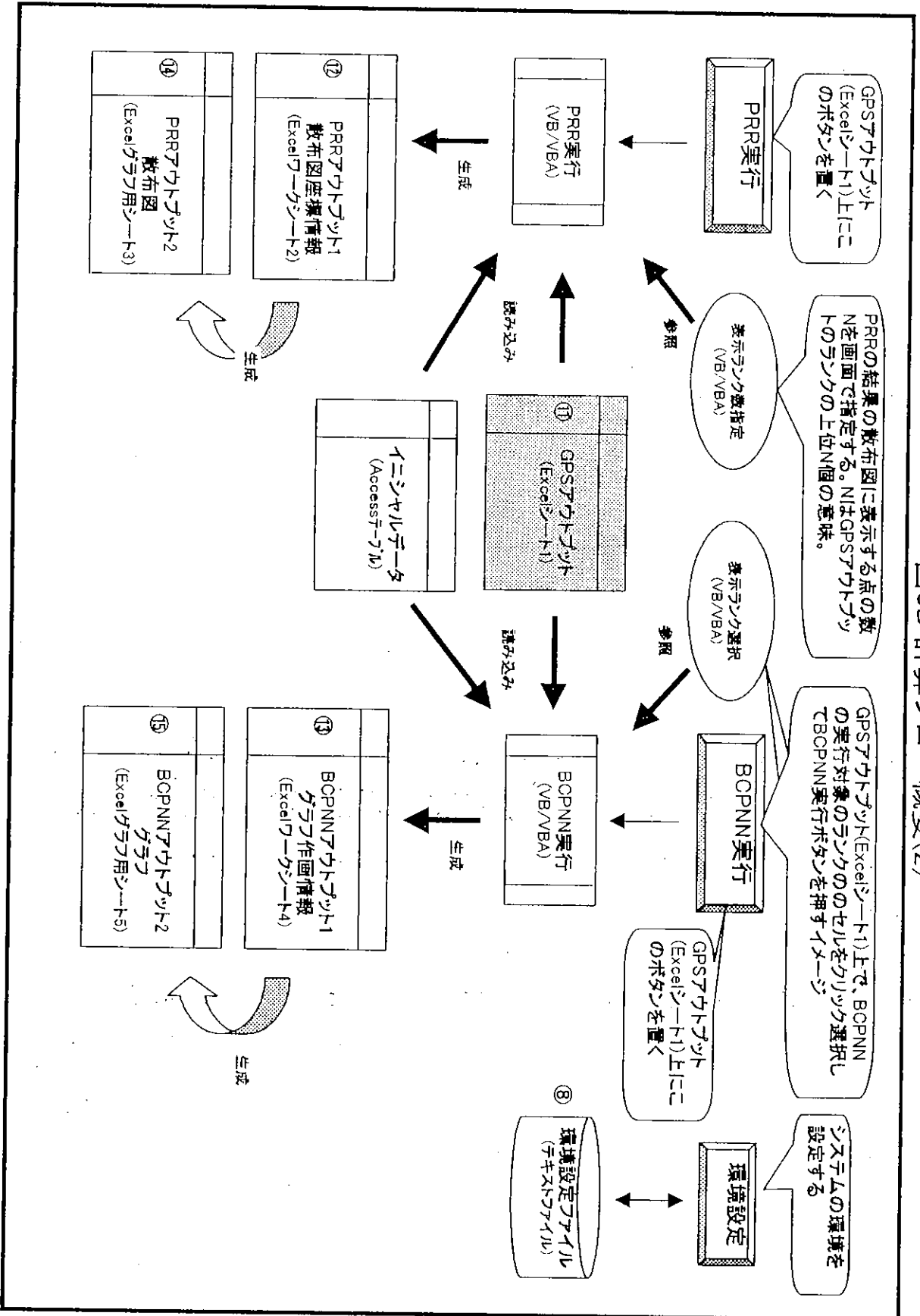


図4 テーブル・ファイル一覧

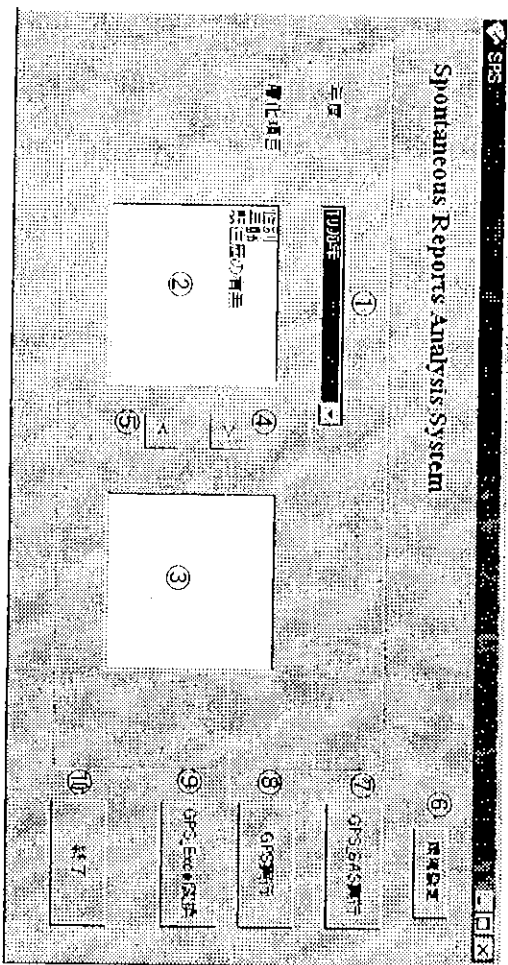
テーブル一覧 (Access2000)			
No.	テーブル	テーブル名	説明
①	DRUG-EVENT発生回数データ	SRS_count.XXXX	薬剤および副作用のカウントデータ。テーブル名のXXXXは任意。
②	イニシャルデータ	SRS_wrk_initdata	GPS計算に必要な各種辞書およびデータファイルを生成するために使用されるワークテーブル
⑬	薬剤マスタ	SRS_Mst_drug	薬剤マスタ
⑭	イベントマスタ	SRS_Mst_event	イベントマスタ

ファイル一覧 (テキスト)				
No.	ファイル	形式	ファイル名	説明
③	DRUG辞書	タグ	SRS_drug.dic	GPS計算用の薬剤情報ファイル
④	EVENT辞書	タグ	SRS_event.dic	GPS計算用の副作用情報ファイル
⑤	STRATA辞書	タグ	SRS_strata.dic	GPS計算用の層別項目情報ファイル
⑥	InitialDataファイル	タグ	SRS_Initial.dat	GPS計算用の基本情報ファイル
⑦	GPSアウトブックファイル	タグ	SRS_output.dat	GPS計算結果ファイル
⑧	環境設定ファイル	csv	SRS_envirom.ini	環境設定画面で設定された情報の保存ファイル
⑨	年度マスタ	csv	SRS_year.mst	メイン画面において年度を表示するためのマスタ
⑩	項目マスタ	csv	SRS_variable.mst	メイン画面において条件設定のための項目を表示するマスタ

シート一覧 (Excel)			
No.	ファイル	ファイル名	説明
①	GPSアウトブック (Excel)	SRS_output.xls	GPS計算結果のExcelブック
②	PRRアウトブック 散布図座標情報	SRS_output.xls	PRRグラフ用座標情報 (PRR値およびカイニ乗値) の格納ワークシート
③	BCPNNアウトブック グラフ制作画情報	SRS_output.xls	BCPNNグラフ制作画情報の格納ワークシート
④	PRRアウトブック 散布図	SRS_output.xls	PRRグラフ用シート
⑤	BCPNNアウトブック	SRS_output.xls	BCPNNグラフ用シート

※ No.欄の番号は「システム化業務フロー」の番号に対応

図 5A メインページ(1)



No	機能概要	タッチ	初期値
1	抽出年度を選択する。表示対象は「年度マスタ」の日本語ラベル	選択	なし
2	抽出項目の候補が表示される。 抽出項目をクリックすることで反転表示される ※ 手続的プログラムでは抽出項目指定は動作イメージのみ	選択	項目
3	抽出項目が表示される。5項目まで選択可。	選択	なし
4	抽出項目を選択する。 ボタンを押下することで選別された項目が③に表示される。	押下	-
5	抽出項目の選択を取り消す。 ボタンを押下することで③で選択された項目が②に表示される。	押下	-
6	参照・出力ファイルのフォルダを指定する。指定できるファイルは以下のとおり ・DRUG辞書 (txt) ・EVENT辞書 (txt) ・STRATA辞書 (txt) ・InitialDataファイル (txt) ・GPSソフトウェアファイル (Excel)	押下	-
7	SASを起動し、GPS計算を実行する。 Excelを起動し、GPS結果画面を表示する。 「年度」が選択されれば押下可能となる。	押下	非活性
8	各種辞書およびインシナルデータファイルを作成し、GPS.exeを起動する。 「年度」が選択されれば押下可能となる。	押下	非活性

図5B メインページ(2)

No.	機能概要	7/23/23	初期値
9	GPSソフトウェアファイルをExcelに読み込み、GPS結果画面を表示する。 GPS結果ファイルが作成されれば押下可能となる。	押下	非応性
10	本画面を終了する。	押下	-

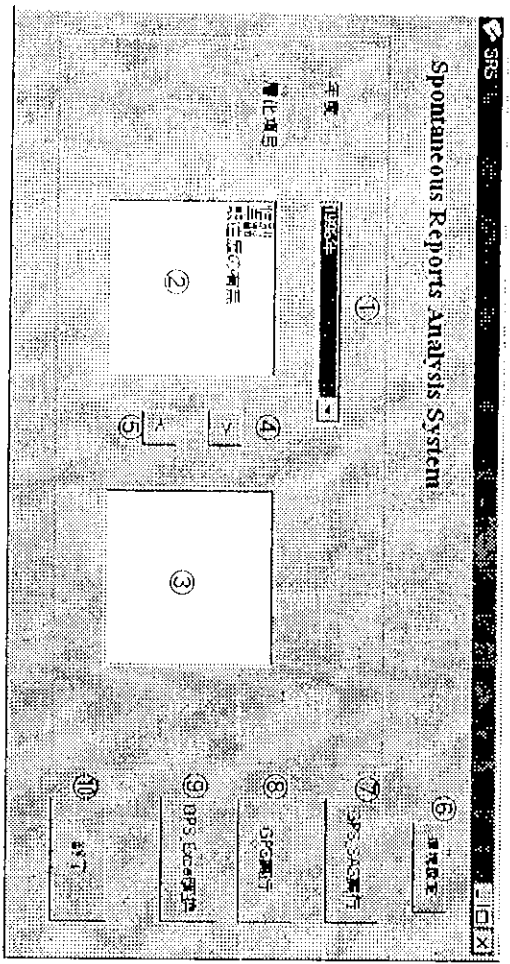


図6 GPSテキストファイルのエクセルファイルへの変換（GPS起動の場合）

No.	機能概要	アクション	初期値
1	GPS Excel 変換の関連ファイルの保存先を指定する。		
2	GPSアウトプットファイル（テキスト）の参照先を指定する。 デフォルトは環境設定に従う。	入力 参照	環境設定
3	GPSアウトプットファイル（Excelブック）の保存先を指定する。 デフォルトは環境設定に従う。	入力 参照	環境設定
4	各参照・保存先を参照・選択する画面を開く	押下	-
5	設定された情報を保存して画面をクローズする	押下	-
5	設定された情報を取り消して画面をクローズする	押下	-

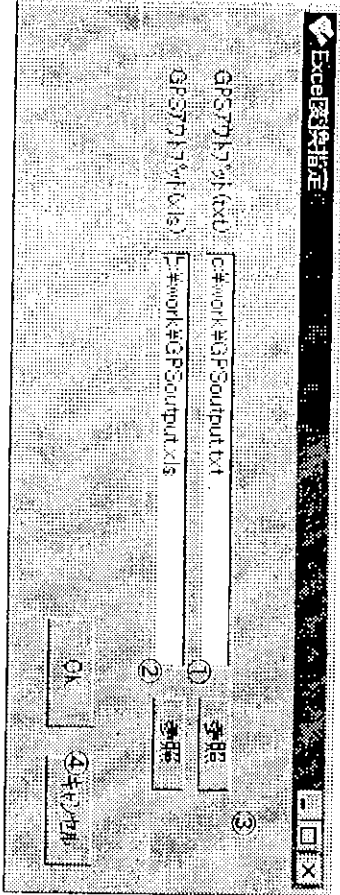


図7 環境設定画面

No.	機能概要	アクション	初期値
1	生成ファイルの保存フォルダを指定する	入力参照	前回の設定値
2	副作用情報ファイルの保存先を指定する	入力参照	前回の設定値
3	個別項目情報ファイルの保存先を指定する	入力参照	前回の設定値
4	基本情報ファイルの保存先を指定する	入力参照	前回の設定値
5	GPS結果ファイル(テキスト)の保存先を指定する	入力参照	前回の設定値
6	GPS結果ファイル(Excelブック)の保存先を指定する	入力参照	前回の設定値
7	各保存先を参照・選択する画面を開く	押下	—
8	PRR実行時に選択されるランク上位のデフォルト数	入力	前回の設定値
9	設定された情報を保存して画面をクローズする	押下	—
10	設定された情報を取り消して画面をクローズする	押下	—

図8 GPS 結果

⑨ BCPNN実行								⑩ PRR実行		
① Rank	② N	③ E	④ RR	⑤ EBGM	⑥ EBlog2	⑦ Dru	⑧ Even	No.	機能概要	初期値
1	293	0.185006	1583.73	1203.74	10.233	ニトラゼパム	肝機能異常NOS	1	Excelを起動し、GPS実行結果をExcelシートで表示する。	-
2	306	0.198265	1543.39	1192.2	10.219	ニトラゼパム	発熱	2	薬剤×副作用の組み合わせをEBGM(Signal Score)の大きい順に並び替えたときの順位。	-
3	65	0.024120	2689.34	786.23	9.619	トリゾラム	咽喉頭疼痛	3	薬剤×副作用の組み合わせでの副作用発生回数(例数)。	-
4	387	0.541982	724.33	644.44	9.332	ゾロチゾラム	口腔内不快感	4	副作用発生回数の期待値(E)	-
5	289	0.424531	680.74	598.34	9.224	ニトラゼパム	嘔気	5	相対リスク(RR)	-
6	150	0.223421	656.32	500.34	9.003	トリゾラム	舌障害NOS	6	副作用発生回数の期待値に対する副作用発生回数の比	-
7	82	0.104533	780.23	480.41	8.988	ゾロチゾラム	流涎過多	7	実証的なベイズの事後確率分布の幾何平均(EBGM)	-
8	71	0.043345	797.45	452.34	8.902	ロルメタゼパム	振戦	8	2のべき乗(べき数はEBlog2)で計算される。	-
9	351	0.343234	645.56	395.65	8.654	ニトラゼパム	急性腎不全	9	相対リスクのベイズ流期待値(EBlog2)	-
10	60	0.058934	594.34	369.32	8.556	トマイソパム	睡眠障害NOS	10	Rankにおける薬剤名(日本語)	-
11	58	0.459822	626.21	356.44	8.532	ロルメタゼパム	暗点		異なるRankに同じ薬剤名が存在することもある。	-
12	37	0.568923	346.45	320.45	8.435	トリゾラム	易興奮性	8	Rankにおける副作用名(日本語)	-
13	46	0.982343	707.22	304.21	8.235	ロルメタゼパム	譫妄	9	異なるRankに同じ副作用が存在することもある。	-
14	41	0.029484	879.34	297.45	8.034	トマイソパム	血圧低下		BCPNN実行ボタン、グラフ表示するデータを1つ選択(Excelの行選択でデータを選択)して押下する。	押下
15	34	0.345433	452.45	264.56	7.893	ロルメタゼパム	耳鳴		ボタンの押下によって同じグラフにグラフ用シートが追加される。	押下
16	257	0.683204	452.45	235.89	7.756	ゾロチゾラム	口腔乾燥		PRR実行ボタン、グラフ表示するデータを1つ選択(Excelの行選択でデータを選択)して押下する。	押下
17	218	0.234492	785.43	223.65	7.734	トマイソパム	鼻咽喉炎		ボタンの押下によって同じグラフにグラフ用シートが追加される。	押下