

200/0310

厚生科学研究研究費補助金

21世紀型医療開拓推進研究事業

標準データ項目セットを用いた知的データベースによる診療根拠の

動的生成に関する研究

平成13年度 総括研究報告書

主任研究者 山本 隆一

平成14（2002）年4月

目 次

I. 総括研究報告書

標準データ項目セットを用いた知的データベース

による診療根拠の動的生成に関する研究 ----- 1

山本 隆一

II. 分担研究報告書

1. データマイニングを用いた動的な診療根拠の生成システムの

フレームワーク構築に関する研究 ----- 23

2. 患者プライバシーを保護したデータベース

アーキテクチャの研究 ----- 35

坂本 憲広

3. データ項目セットの整備と利用方法 ----- 41

大江 和彦

III. 研究成果の刊行に関する一覧表 ----- 43

IV. 研究成果の刊行物・別冊 ----- 45

厚生科学研究費補助金（21世紀型医療開拓推進研究事業）

総括研究報告書

標準データ項目セットを用いた知的データベースによる診療根拠の動的生成に関する研究

総括研究者 山本 隆一 大阪医科大学病院医療情報部助教授

研究要旨

EBMが重要であることは論をまたないが、一般的な文献を基礎とする手法では一過性で流行性の疾患や薬剤の副作用の速やかな発見などには適用できない。そこで診療情報の電子化が行われることを前提に、それらの情報をプライバシー保護を確保して上で収集し、データマイニングの手法を用いてダイナミックに知識を抽出し診療根拠とすることを目標とし、そのための基礎研究をおこなった。平成13年度はデータマイニングの手法として相関ルールとそのため前処理に関して検討し、動的に診療根拠を生成するためのフレームワークを研究するとともに、データ項目セットの適応性、およびプライバシー保護と安全性確保のためのデータの無名化の検討と、プライバシー・センシティブな情報の扱いについて研究をおこない一定の成果を得た。

分担研究者：

大江和彦

東京大学医学部附属病院中央医療情報部教授

坂本憲広

九州大学医学部附属病院 講師

増田 剛

大阪医科大学病院医療情報部 助手

本研究の目的は平成11年度の厚生省によるデータ項目セット開発事業の成果を利用し、この項目セットに準拠して収集された診療データから動的に診療根拠を抽出する方法を研究開発することにある。

証拠に基づく診療（EBM）が重要であることは論を待たないが、一般的なEBMのように文献的な証拠に基づく場合、インフルエンザに対する抗ウイルス剤の効果のような流行性で急性の疾患

A. 研究目的

への対処や、薬品の副作用などの迅速な対応を必要とする場合などでは十分な効果が期待できないことがある。例えば今年のインフルエンザにアマンタジンが効果を示すかどうかといった場合、文献的な根拠を待つことができないために、診療現場からの経験が厳密な検証なく、また統計的な処理がほとんど行われずに流布する形でわずかに現場医療に活かされている状況にある。このような急性で流行性の疾患や、集団食中毒のような、時間的に従来のEBMの適応が難しい場合、臨床現場から診療情報を動的に収集し、適切な情報処理と解析を行うことにより、意味のあるルールを導き、早期に診療現場にフィードバックすることで、なんらかのヒントを与えることができれば医療の質の向上に結びつく。また薬品の副作用も相当な例数の蓄積と回顧的な解析が必要であるが、副作用を疑う医療現場からの報告に依存しており、出現頻度が1%以下といった低頻度で、相当な数の症例を扱わない限り疑うことが難しい状況では調査そのものが遅れる可能性がある。仮に新規に発売された医薬品が結果的に1%の患者に肝障害の副作用がある場合、1つの医療機関で50例に投与してもいずれかの患者に肝障害が起こる確率は0.5で、仮に起こったとしても当該医薬品との関連に処方医が気づく可能性は100%で

はない。熟練した医療従事者の印象は実際には複雑な知識背景のもとに下される判断で、高く評価する必要があるが、客観性や確実性は不十分といわざるを得ない。しかし、もし広範囲の医療施設から診療情報を動的に収集し、適切な情報処理と解析を行えば、低頻度の副作用も比較的早期に発見できる可能性がある。

平成11年度に開発されたデータ項目セットは電子化診療情報を共通の標識で整理することで、異なる医療機関の診療情報を統一的に扱うことを可能とするもので、これを活用することにより、広範囲から診療情報をリアルタイムに収集することが可能になる。

一方で情報工学の分野では知的データベースやデータマイニングと呼ばれる手法の研究が活発に行われている。これは網羅的に集められたデータの集合から意味のある関係を自動的に抽出する手法であり、人の印象に頼らないデータ解析を行うことができる。着眼点を指定しなくても動的に特異な関係を抽出できるために人が気づき難い関係や、気づくのに時間がかかる関係を早期に抽出するのに極めて有用な方法と考えられる。もちろん通常のEBMで用いられる回顧的な方法にくらべて精度低下は予想されるし、背景となるべき医学理論までを類推することはできない。したがってデータマイニングや知的

データベースで得られた知識は通常のEBMと同等に論じられるものではなく、あくまでもヒントを与える方法であり、本来のEBMが目指す科学的な精度を得るためには回顧的な研究方法を併用する必要がある場合が多い。しかし少なくとも着眼点を得るまでの時間は大幅に短縮され、医療現場へのフィードバックもそれだけ有効になることが期待できる。

しかしデータ項目セットは開発されたばかりで、実装応用例は少なく、このような目的のデータ収集に十分な能力があるか検証する必要がある。またデータマイニングは情報工学の分野で活発に研究されており、いくつかの分野では実用的な応用例も存在するものの、医療分野での定まった適用方法はない。データマイニングにはさまざまな手法があり、またその前提となるデータの前処理にも定まった方法はない。医療で実用的に用いるためには実証的な研究が必要である。また動的に診療情報を収集する以上、プライバシー保護はきわめて重要な課題となる。そこで本研究ではこのような用途に関してのデータ項目セットの検討を分担研究者の大江を中心に行い、プライバシー・センシティブな情報項目の扱いを分担研究者の坂本を中心に行い、データマイニングおよび前処理の手法の研究を増田を中心に、またデータ収集の際のプライバ

シー保護を確実にを行うための匿名化を含めた安全性確保の研究、およびこれらの総括を山本が行うこととした。

データマイニングおよび前処理について、本年度は、既に収集された静的な医療データベースから、診療根拠につながるルールを発見する手法について研究を行うと同時に、このような動的な飲料根拠生成が有用になりうり場面を解析し、その場面ごとに必要な要件を抽出し、フレームワークを作成した。

安全性確保の研究については実際の患者基本情報データベースから実験用に変換したデータを用い、データの無名化の定量化を試み、また安全でプライバシー保護が可能な伝送を実証する目的で廉価に構築できるVirtual Private Network (VPN) 環境を作成し、次年度の研究に備えた。

データマイニングのフレームワーク、データ項目セットおよびプライバシー・センシティブな項目の扱いについては分担研究報告書も参照されたい。

なお、本報告の一部には平成12年度の厚生科学研究費で実施した研究結果が含まれている。これは本研究の研究補助金の分類が変更になったために報告書に一貫性を持たせるためである。

B. 研究方法

B-1. プライバシー保護と匿名化について

a 大阪医科大学附属病院に過去5年間で利用された患者情報32万件を用い、厚生労働省の補助で作成された「電子化された診療情報交換のためのデータ項目セット(以下J-MIXと呼ぶ)」にしたがって項目整理を行った。

b 上記のデータベースを用い、単一項目814項目について最小特定人数を計算した。また郵便番号は上3桁のみ、住所は市や町名レベルでも計算した。また年齢は階層でも計算し、生年月日は月単位、年単位でも計算した。

c 2の単一項目のうち、よく使われると思われる組み合わせ80組について最小特定人数を計算した。

なお、上記a, b, cはデータの無名性を研究するものであり、研究のためのデータは実データを使用した。しかし、特定の程度だけを指標として用いるために、何に特定されるかは問題ではない。そこで、氏名などの特定性の強い項目ははじめから除外して扱った。

d 個人情報保護法案が国会に上程されているが、この法案を解析し、診療情報のプライバシー保護に関するガイドラインを試作した。

B-2. 実験用VPN環境の整備について

当初は専用のVPNコンセントレータを用いて構築する予定であったが、機器の値上がり等で適当な製品がなく、また、Linux等のPC-UNIX上で安定したVPNルーティングが可能になってきたために、既存のLinuxサーバにipchainsおよびFreeS/WANを用いてVPN実験環境を構築した。クライアントとしてはNote PC上で、Windows XPおよびLinux + FreeS/WANでアクセステストを行った。

B-3. データマイニングおよび前処理について

a 相関ルール発見

大量に蓄積された診療データからルールを発見する手法として、データ中の属性間の相関関係を抽出する相関ルール発見手法を適用する。データの属性集合をI、データベースをDとすると相関ルールは $X \Rightarrow Y$ で表現され、 $X, Y \subset I$ 、 $X \cap Y = \text{空集合}$ である。相関ルールはルールが持つ支持度(support)と確信度(confidence)の2つの値を使ってその有意性を示す。相関ルール $X \Rightarrow Y$ の支持度 $\text{support}(X \Rightarrow Y)$ はD全体に対しXとYをともに含む事例の割合 $\text{support}(X \cup Y)$ と定義される。確信度 $\text{confidence}(X \Rightarrow Y)$ はD中でXを含む事例のうち、XとYを共に含む事例の割合、すなわち $\text{support}(X \cup Y) / \text{support}(X)$ によっ

て定義される。相関ルール発見とは、ユーザが指定する最小支持示と最小確信度を満足するすべてのルールを抽出する問題となる。相関ルール発見の効率的なアルゴリズムの1つに、IBMアルマデン研究所のAgrawalらが提案したアプリアルゴリズムがある。これは、最小支持度の性質を利用し探索空間を枝刈りしながら長さ1のルールから順に条件を満たす相関ルールを抽出する。本研究ではこのアプリアルゴリズムを適用し、診療データから相関ルールを発見する。

b データの前処理

対象となる診療データベースとして、大阪医科大学付属病院で実際に蓄積された糖尿病データベースを用いた。本データベースは、60個の属性から構成されており1251の事例を含んでいる。60個の属性のうち、「年齢」や「Body Mass Index」といった属性値として数値をとる連続属性は31個存在する。一方、「網膜症有無」や「喫煙歴有無」といった、属性値としてあらかじめ定義された離散値をとる離散属性は29個である。それぞれの属性は欠損値を含む。これらのすべての属性を使用しアプリアルゴリズムをそのまま適用すると、大量のルールが生成され、ルールの発見にかかる時間も大きくなる。そこで、相関ルール発見手法をより効果的に適用す

るために、前処理として以下の処理を糖尿病データベースに対して施した。

○知識発見に無関係と思われる属性を排除

「患者ID」や「カルテNo」といった知識発見に本質的に関係無い属性を排除した。

○欠損値を多く含む属性を排除

属性の中には、1251事例のうち1000個以上もの事例について、その属性値が欠損しているものが存在した。そのように、欠損値を多く含む属性は、その属性値の生起確率が小さくなるために発見されるルール中には出現しない。そこでそのような属性はあらかじめ排除した。

「Retinopathy Score」や「Nephropathy Score」といった属性がこれに該当する。

○属性の統合

属性の中には、属性間で何らかの相関を持つ時系列属性が含まれる。例えば属性「網膜症有無」は、年度ごとに値がそれぞれ異なる属性として定義されている。これらすべての属性を使用した場合、これらの属性間の陽な関係がルールとして導出される可能性が高いが、それらが有用な知識となる可能性は低い。そこで、これらの属性を、「1994年から2000年までの網膜症の有無」という1つの属性に統合した。

○離散属性値の集約

離散属性「網膜症有無」と「腎症有無」は多

値の属性値をとる。これらは、相関ルール発見アルゴリズムを適用する際それぞれ別個の属性に分解される。これは属性数を増加させるため、ルール発見時の計算時間の増大につながる。そこで、このような多値属性値を二値に集約した。

前処理の結果として、実際にルール発見に使用する属性は22個(連続属性4個、離散属性18個)となった。

さらに、今回適用したアプリアリアルゴリズムは、2値の属性値を持つ離散属性しか扱うことができない。そこで、連続属性については、幾つかの「教師なし大域離散化」を用いて離散化を行う。「教師なし大域離散化」は、各事例が持つ目標概念に関する情報を用いずに、事例集合全体を使って属性間の依存性を考慮せずに各属性を離散化する手法である。今回は、属性値を等範囲で分割するEqual Width Intervalを用い、分割数は2とした。また離散属性は、各属性値についてYesとNoの二値をとる個別の属性に分割した。

c 導出ルール数の抑制

単純に相関ルールを適用すると、特に最小指示度を小さくした場合に非常に大量の相関ルールが導出され、しかもそのほとんどは、人間にとってあたりまえの規則を表現するルールであった。そこで、そのようなあたりまえのルール

を機械的に排除し、導出ルール数を抑制するために、以下に述べる内部確信度を導入する。これは直感的には、相関ルールの条件部を構成する個々の属性は相関ルールの結論節と相関が見られないが、条件部全体が結論節と強い相関を持つルールのみを導出する操作に相当する。たとえば最小確信度を0.8でルール導出を試みて、

$$(1) \quad A \Rightarrow X \quad (\text{確信度 } 0.82)$$

$$(2) \quad B \Rightarrow X \quad (\text{確信度 } 0.81)$$

$$(3) \quad A, B \Rightarrow X \quad (\text{確信度 } 0.80)$$

という結果が得られた場合、(3)は新しいルールとしての意味は乏しい。(1)、(2)の確信度より(3)の確信度が相当程度高い場合に意味がある。組み合わせで確信度が上昇することを表現する場合、割合で示すこともできるが、本研究では目標最小確信度より低い内部確信度を設定し、この例で言えば(1)、(2)は内部確信度より低く、(3)が目標最小確信度より高い場合だけに新しいルールとして導出することとした。

すなわち、相関ルール $R: Y_1, Y_2, \dots, Y_m \Rightarrow X$ に対して以下の相関ルール $R_k: Y_k \Rightarrow X$ ($1 \leq k \leq m$)を考える。相関ルール R_k は、ルール R の条件部に現れる属性を1つだけ条件部に含むルールである。このとき、 $\text{confidence}(R)$ をルール R の確信度とすると、

$$\forall k [\text{confidence}(R_k) < \theta_1 \wedge \text{confidence}(R)$$

> $\theta 2$]

であるようなすべてのルールを導出する。 $\theta 1$ を内部確信度と呼ぶ。 $\theta 2$ は最小確信度であり、 $\theta 1 \leq \theta 2$ である。

なお、本研究では倫理面の問題は存在しない。実診療情報を用いているが、完全に無名化されており、結果には個人を特定できる情報はまったく含まれていない。

B-4. 診療現場でデータマイニングを用いた動的な診療根拠を生成するフレームワーク

EBMは直感や経験にたよる診療から、ピア・レビューを受けた科学的な根拠を持つ証拠に基づいて医療を遂行するという考え方であり、医学に基づく医療としてはいわば当然の方向である。しかし医療の現場にはピア・レビューを受けた根拠となるべき知識だけでは根拠に基づく医療を遂行できない場面が存在する。

その1つは急性で流行性の疾患であり、例えば今年のインフルエンザに特定の抗ウイルス薬が効果があるかどうかは、その知識が集積され、解析され、発表の上ピア・レビューを受けて診療根拠として利用できるころには流行が終了している。また1996年に堺市で0-157大腸菌による集団食中毒が発生したが、このような治療経験の乏しい感染性集団食中毒に対して抗生物質の

使用が有効か有害かといった判断も通常のEBMでは間に合わない。この2つの例のような、時間的に早い経過をたどる流行性疾患ないし集団発生疾患が、データマイニングを用いた診療根拠の動的生成のターゲットの1つになると考えられる。

2つ目は低頻度の疾患または病的状態で、低頻度の薬剤副作用がその代表である。例えば結果的に肝機能障害が1%以下の確率で起こる薬品で、その副作用が知られていない状態を想定する。この場合、ある医療機関で50例にこの薬品を処方していたとして、いずれかの患者に当該薬剤の副作用としての肝機能障害が見られる確率は0.5である。またかりに一例の肝機能障害が発見されても、処方医が副作用であると気づかない可能性がある。このような場合でも多施設から診療情報を収集し、データマイニングを適切に行えば早期に副作用の可能性を示すことができる可能性がある。

以上の2つの場合に注目をし、それぞれについてデータマイニングおよびその結果の解析のフレームワークを検討した。

C. 研究結果

C-1. データの無名化について

表1は生年月日の粒度別や特定の年齢、およ

表1 最小特定人数の計算例

患者.生年月日(年、月、日)	:30.6人
患者.生年月日(年、月、日) + 患者.性別	:15.3人
患者.生年月日(年、月)	:368人
患者.生年月日(年、月) + 患者.性別(女性)	:152人
患者.年齢(60歳)	:2万5千人
患者.住所(高槻市)	:12万8千人
患者.住所(高槻市安岡寺)	:4332人
患者.年齢(60歳) + 患者.性別(女性)	:1万1千人
患者.年齢(60歳) + 患者.性別(女性) + 患者.住所(安岡寺)	:1170人
患者.年齢(60歳) + 患者.性別(女性) + 保険適用.傷病.名称(胃がん)	:89人

び特定の住所の最小特定人数および他の情報項目との組み合わせの最小特定人数を示す。

C-2 個人情報保護法とプライバシーガイドラインについて。

個人情報保護法案を分析したところ、総則、各論、実施上の事務的な項目などにわかれており、法律が成立した場合、医療従事者から見た場合に複数の章から1つの対策を読み取る必要があった。したがって効率よくガイドラインを作成するために主に医療従事者から見て、対応すべき項目ごとに条文を並べ替えポイントリストとした。以下にその一部を示す。

Section 1-1利用目的の特定と目的外利用の制限

第四条 個人情報とは、その利用の目的が明確にされるときに、当該目的の達成に必要な範囲内で取り扱われなければならない。

第二十条 個人情報取扱事業者は、個人情報を取り扱うに当たっては、その利用の目的（以下「利用目的」という。）をできる限り特定しなければならない。

2 個人情報取扱事業者は、利用目的を変更する場合には、変更前の利用目的と相当の関連性を有すると合理的に認められる範囲を超えて行ってはならない。

第二十一条 個人情報取扱事業者は、あらかじめ本人の同意を得ないで、二十条の規定により特定された利用目的の達成に必要な範囲を超えて、個人情報を取り扱ってはならない。

2 個人情報取扱事業者は、合併その他の事由により他の個人情報取扱事業者から事業を承継することに伴って個人情報を取得した場合は、あらかじめ本人の同意を得ないで、承継前における当該個人情報の利用目的の達成に必要な範囲を超えて、当該個人情報を取り扱ってはならない。

3 前二項の規定は、次に掲げる場合については、適用しない。

一 法令に基づく場合

二 人の生命、身体又は財産の保護のために必要がある場合であって、本人の同意を得ることが困難であるとき。

三 公衆衛生の向上又は児童の健全な育成の推進のために特に必要がある場合であって、本人の同意を得ることが困難であるとき。

四 国の機関若しくは地方公共団体又はその委託を受けた者が法令の定める事務を遂行することに対して協力する必要がある場合であって、本人の同意を得ることにより当該事務の遂行に支障を及ぼすおそれがあるとき。

Section 1-2取得に際しての利用目的の通知等

第二十三条 個人情報取扱事業者は、個人情報を取得した場合は、あらかじめその利用目的を公表している場合を除き、速やかに、その利用目的を、本人に通知し、又は公表しなければならない。

2 個人情報取扱事業者は、前項の規定にかかわらず、

本人との間で契約を締結することに伴って契約書その他の書面（電子的方法、磁気的方法その他の知覚によっては認識することができない方法で作られる記録を含む。以下この項において同じ。）に記載された当該本人の個人情報を取得する場合その他本人から直接書面に記載された当該本人の個人情報を取得する場合は、あらかじめ、本人に対し、その利用目的を明示しなければならない。ただし、人の生命、身体又は財産の保護のために緊急に必要がある場合は、この限りでない。

3 個人情報取扱事業者は、利用目的を変更した場合は、変更された利用目的について、本人に通知し、又は公表しなければならない。

4 前三項の規定は、次に掲げる場合については、適用しない。

一 利用目的を本人に通知し、又は公表することにより本人又は第三者の生命、身体、財産その他の権利利益を害するおそれがある場合

二 利用目的を本人に通知し、又は公表することにより当該個人情報取扱事業者の権利又は正当な利益を害するおそれがある場合

三 国の機関又は地方公共団体が法令の定める事務を遂行することに対して協力する必要がある場合であって、利用目的を本人に通知し、又は公表することにより当該事務の遂行に支障を及ぼすおそれがあるとき。

四 取得の状況からみて利用目的が明らかであると認められる場合

このように整理して得られた項目は以下のよう
になった。

Part1 個人情報取り扱い業者

Section 1-1利用目的の特定と目的外利用の制限

Section 1-2取得に際しての利用目的の通知等

Section 1-3 適正な取得

Section 1-4 正確性の確保

Section 1-5 安全性の確保

Section 1-5-2安全管理措置

Section 1-5-3 従業者の監督

Section 1-5-4 委託先の監督

Section 1-6 第三者提供の制限

Section 1-7 透明性の確保

Section 1-7-2 保有個人データに関する事項の公表等

Section 1-7-3 開示

Section 1-7-4 開示等の求めに応じる手続

Section 1-7-5 手数料

Section 1-7-6 訂正等

Section 1-7-7 利用停止等

Section 1-7-8 理由の説明

Section 1-7-9 個人情報取扱事業者による苦情の処理

Section 1-8 報告の徴収

Section 1-8-2 助言

Section 1-8-3 勧告及び命令

Section 1-8-4 配慮義務

Section 1-8-5 主務大臣

Part 2 認定個人情報保護団体

Section 2-1 認定・廃止・取り消し

Section 2-1-2 欠格条項

Section 2-1-3 認定の基準

Section 2-1-4 認定の取消し

Section 2-1-5 廃止の届出

Section 2-2 業務

Section 2-2-1 対象事業者

Section 2-2-2 苦情の処理

Section 2-2-3 個人情報保護指針

Section 2-3 目的外利用の禁止

Section 2-4 名称の使用制限

Section 2-5 報告の徴収

Section 2-6 命令

Section 2-7 主務大臣

Part 3 雑則

Section 3-1 適用除外

Section 3-2 施行の状況の公表

Part 4 罰則

Part 5 附 則

Section 5-1 施行期日

Section 5-2 本人の同意に関する経過措置

Section 5-3 通知に関する経過措置

Section 5-4 名称の使用制限に関する経過措置

このうち医療機関および医療従事者に対するガイドラインとして重要なのはPart1でありPart1を主体に条文に対応した医療分野用のガイドライン案を作成した。なお作成にあたっては米国のHIPAA Privacy Regulationおよびそれに対応してAcademic Medical Colleges HIPAA Working Groupが作成した“Guidelines for Academic Medical Centers on Security and Privacy”を参考にした。以下にガイドライン試案を示す。

医療における個人情報保護ガイドライン試案

1) 用語

個人情報とは個人が容易に識別できる一体の情報であり、他の情報と照合することで容易に個人が識別できる場合を含んでいる。例えば患者IDと検査項目名、検査値だけが記録された情報は、患者IDと患者本人を結びつける情報がない場合は個人情報ではないが、IDと患者本人を結びつける情報が入手可能な場合は個人情報として扱わなければならない。診療現場では患者から発生したか、患者に関連する診療情報はすべて個人情報として扱う必要がある。

個人情報データベースとは個人情報の集積で検索可能なものを指す。個人情報保護に関する法律ではコンピュータの使用を前提とした一般的なデータベース以外に手作業で検索可能なものも含むとしており、診療現場では紙のカルテや各種伝票、帳票、報告書なども含まれ

ると考えられる。

個人情報取扱い業者とは個人情報データベースを扱う業者であり、すべての医療機関はこれに相当すると考えられる。

提携事業者とは医療機関ではなくて、医療機関が個人情報を含む診療情報を業務上で提供または交換する事業者。例えば検査会社や診療報酬算定業者など。

2) 体制

医療機関は個人情報保護を達成するための体制を整備すべきである。大規模な医療機関では院内に個人情報保護責任者と苦情処理担当者を設置し、次章のポリシーおよび運用規程に明記すべきである。小規模な医療機関では個人情報保護を目的とする民間団体の対象事業者となり、その作成する指針に従うべきである。

3) ポリシーと運用規程

医療機関は個人情報保護に対する理念と後述する利用目的を含む情報の扱いの概要や苦情処理の方法などの基本方針をさだめて個人情報保護に関するポリシーとして文書化し公開すべきである。この場合公開とは受診患者ならびに受診しようとする患者にとって容易にアクセスできることを意味し、例えば院内のよく見えるところに掲示し、パンフレットを用意するなどを指す。またポリシーに従って診療情報の運用規程を定め、要求に応じて開示すべきである。

4) 医療機関における個人情報保護の細目

4-1. 利用目的の明確化と通知

医療機関は診療情報の利用目的を明確にしなければならない。また当該本人にとって自明でない利用目的は容易に当該本人が認識できる方法で通知しなければならない。誰でも見ることができ、なおかつ注意を引き易い場所に掲示するか、個々に説明文書を渡すなどの方法をとる必要がある。

一般に医療機関では下記のような利用目的が考えられる。A、B、Cは医療機関で情報を取得する場合自明の利用目的であり、通知しなくてもかまわない。またD、の法令で定められた利用に関しては通知しなくてもかまわない。

- A. 当該本人の健康回復および健康維持。
- B. 診療報酬請求の作成のための利用。
- C. 病院管理などの医療機関の健全な運用のための利用
- D. 法令で定められた届け出などのための利用
- E. 医学生、研修医などの教育のための利用
- F. 臨床医学研究および疫学的研究のための利用
- G. 各種監査における利用

利用目的を変更した場合は本人に通知するか公表しなければならない。ただし、国の機関又は地方公共団体が法令の定める事務を遂行することに対して協力する必要がある場合であって、利用目的を本人に通知し、又

は公表することにより当該事務の遂行に支障を及ぼすおそれがあるときを除く。

4-2. 目的外利用の制限

個別に本人に同意を得ることなく、診療情報の目的外利用はしてはならない。ただし以下の場合を除く。

- A. 法令に基づく場合
- B. 人の生命、身体または財産の保護のために必要がある場合で、本人の同意を得ることが困難なとき
- C. 公衆衛生の向上又は児童の健全な育成のために必要がある場合で、本人の同意を得ることが困難なとき
- D. 国の機関若しくは地方公共団体又はその委託を受けた者が法令の定める事務を遂行することに対して協力する必要がある場合であって、本人の同意を得ることにより当該事務の遂行に支障を及ぼすおそれがあるとき。

4-3. 適正な取得

ポリシーの掲示だけでなく、可能な限り説明し同意を得て情報を収集するべきである。意識障害などで同意を得られない場合は、その事情を客観的に診療録に記述する。また未成年者で本人が説明を理解する能力がない場合は親権者または法的な後見人に説明し同意を得ることに勤めなければならない。

4-4. 正確性の確保

診療データの多くはある時点の情報であり、経時的に変化する場合、その時点でのあらたな情報が発生する。このような情報の正確性を確保するために、一般的な精

度管理に勤めなければならない。また例えば病名のような持続する状態を示す情報は適切に転帰を記載するなどの状態を正確に保つことにつとめなければならない。

4-5. 安全管理と従業者の監督

安全管理は診療情報データベースサーバやカルテ倉庫のような物理的な場所の安全管理と、搬送機やネットワーク上で稼動する連携アプリケーションなどの情報の移転の安全管理、および情報にアクセスする従業員の安全管理からなる。

物理的な場所の安全管理は漏水、浸水、地震などによる情報の破損に備える必要がある。可能であれば地下や1階は避け、2階以上で保管するべきである。やむを得ない場合は十分な防水対策等を施す必要がある。また入退出管理は必須であり、操作記録も備える必要がある。また必要に応じて記録やデータを管理場所から持ち出す場合は運用上自明の場合を除き、所在を明らかにする必要がある。

情報の移転の安全管理は紙の媒体の場合、落下を防止し、また容易に第三者が診療情報を観察できるような形態で移転してはならない。診察室などでも現に診察中の患者以外の診療情報が容易に観察できるような形で放置してはならない。ローカルエリアネットワークを用いている場合は、ネットワーク接続機器を正しく管理するとともに、不正なアプリケーションやライブラリを動作させてはならない。

従業員の安全管理は正確に利用者登録を行い、利用者認証を徹底し、アクセス管理をたたく行うことからな

る。またシステムとしてのアクセス管理だけで効率的な運用と安全確保の両立は難しく、運用規程をさだめ遵守することと監査が重要である。

4-6. 提携業者の監督

提携業者に診療情報を提供する必要がある場合は、あらかじめ当該本人にその旨を掲示したポリシー内などで告知しておく必要がある。さもなければ提携業者とはみなされない。

提携業者が当該医療機関と同等以上の個人情報保護対策を講じていることを確認し、その旨を明記した契約を行う必要がある。契約後も適切な監督を行う必要があり、定期的に報告を求めると、監査することが望ましい。

4-7. 第三者提供の制限

第三者への診療情報の提供はポリシーでの掲示などで周知をはかった公益的目的以外は原則として、してはならない。ただし以下の場合を除く。

- A. 法令に基づく場合
- B. 人の生命、身体または財産の保護のために必要がある場合で、本人の同意を得ることが困難なとき
- C. 公衆衛生の向上又は児童の健全な育成のために必要がある場合で、本人の同意を得ることが困難なとき
- D. 国の機関若しくは地方公共団体又はその委託を受けた者が法令の定める事務を遂行することに対して協力する必要がある場合であって、本人の同意を得ることにより当該事務の遂行に支障を及ぼすおそれがあるとき。

個人識別不可能な情報は対象外であるが、識別不可能であることを厳格に確認する必要がある。医療機関や保健機関であつかう情報は住所や年齢など偏りが大きいものもあり、一見個人識別につながらない情報でも相当程度に限定できる場合がある。また項目の組み合わせで強く限定できる場合があることに留意しなくてはならない。

本人の同意を得ており、本人の求めによって中止でき、以下の事項を本人が承知している場合は第三者に提供可能である。

- A. 第三者への提供を利用目的とすること。
- B. 第三者に提供される個人データの項目
- C. 第三者への提供の手段または方法
- D. 本人の求めに応じて当該本人が識別される個人データの第三者への提供を停止すること。

各種保険の診断書などが上記の第三者へ提供することができる場合に相当する。なお、死亡診断書のように本人の同意が不可能な場合は、遺族または家族を代表するものが求める場合、および遺族または家族を代表するものの同意を得て提供すべきである。

4-8. 開示

本人から、診療情報の開示を求められたときは、別に政令で定める方法で遅滞なく診療情報を開示しなければならない。ただし以下の場合を除く。

- A. 本人又は第三者の生命、身体、財産その他の権利利益を害するおそれがある場合

B. 当該個人情報取扱事業者の業務の適正な実施に著しい支障を及ぼすおそれがある場合

C. 他の法令に違反することとなる場合

以上の理由により開示しないことを決定した場合は、遅滞なく本人にその趣旨を伝えなければならない。また開示しない理由を明らかにするようにつとめなければならない。

医療機関は開示を請求するための手続きを定めることができる。手続きを定めた場合、開示を請求するものはこの手続きに従って請求しなければならない。開示を請求する手続きは請求者に過度の負担をかけるものであってはならない。また開示に必要な手数料を定めることができるが、実費を勘案して合理的であると認められる範囲内において、その手数料の額を定めなければならない。

開示請求は政令で定める代理人によってもおこなうことができる。

4-9. 訂正等

本人から当該本人の診療情報が事実でないという理由で理由によって訂正（削除および追加を含む）があった場合は利用目的の達成に必要な範囲内において、遅滞なく必要な調査を行い、その結果に基づき、当該保有診療情報の内容の訂正等を行わなければならない。ただし診療情報の多くは客観的な情報であり、このような情報は本人の求めであっても根拠なく変更してはならない。

訂正を行った場合は訂正の履歴が確認できるように訂正を行ったこと自体を記録するとともに、その理由も記録することがのぞましい。また訂正の事実と内容、訂正を行わなかった場合はその趣旨を本人に通知しなければならない。訂正をおこなわなかった場合にはその理由を明らかにするようにつとめなければならない。

4-10. 利用の停止

診療情報が個人情報保護に関する法律の理念に違反して取り扱われているという理由で本人から診療情報の利用の停止を要求された場合で、その求めに理由があることが判明し、本人および第三者の生命、身体、財産を害する恐れのない場合は、違反を是正するために必要な限度で、遅滞なく、当該診療情報の利用停止等を行わなければならない。ただし、当該診療情報の利用停止等に多額の費用を要する場合その他の利用停止等を行うことが困難な場合であって、本人の権利利益を保護するため必要なこれに代わるべき措置をとるときは、この限りでない。

利用の停止を行った場合、および利用の停止を行わなかった場合は本人に遅滞なく通知し、利用の停止をおこなわなかった場合には理由を明らかにすることにつとめなければならない。

4-11. 苦情の処理

医療機関は個人情報保護に関しての苦情の適切かつ迅速な処理に努めなければならない。またそのために担当者を明確にし、窓口を設けるなどの体制を整備しなけ

ればならない。

5) 認定個人情報保護団体

個人情報保護に関する法律では個人情報保護団体が主務大臣の認可を受けて、苦情の処理、対象事業者への個人情報保護に関する情報の提供、指針の作成などを行うことができる。小規模な医療機関で苦情の処理などを行うことが困難な場合は個人情報保護団体の対象業者となることで、負担を軽減することができる。

C-3. 実験用VPN環境について

PC-UnixであるLinux (Red hat 7.2)をインストールし、ネットワークカードを2枚持つPCをVPNサーバとし、ipchainsおよびFreeS/WANをインストールし、RSA暗号を用いたPKI-VPNと事前鍵交換を前提にする共有鍵VPN環境を実験的に構築した。接続実験はクライアントとして、2台のNote PC Sony VAIO 505S/PDを用い、OSはNote PCに付属のMicrosoft Windows XP professionalとLinux (Red hat 7.2)で行った。クライアントにもLinuxを用い、FreeS/WANを使った場合はPKI-VPN、共有鍵VPNともに円滑に接続でき、処理速度、暗号強度ともに実用的な問題はなかった。しかしクライアントにWindows XPを用いた場合、VPNサーバには問題なく接続できたが、サーバが接続されたLANへのアクセスはできな

表2

最小確信度(%)	最小支持度(%)	内部確信度 < 90%		内部確信度 < 80%	
		ルール数	時間[sec]	ルール数	時間[sec]
90	50	115(87)	15.158	28(0)	14.949
	30	603(574)	109.843	29(0)	109.235
	20	1164(1134)	395.166	35(5)	395.141
	10	2193(2161)	2611.314	42(10)	2619.225
	4	3262(3229)	13297.554	54(21)	16203.318
最小確信度(%)	最小支持度(%)	内部確信度 < 70%		内部確信度 < 50%	
		ルール数	時間[sec]	ルール数	時間[sec]
90	50	28(0)	15.088	28(0)	14.912
	30	29(0)	110.584	29(0)	108.577
	20	30(0)	395.653	30(0)	394.558
	10	32(0)	2619.407	32(0)	3868.889
	4	34(1)	13272.743	33(0)	13334.230

表3

Coronary Disease (Yes), Hyperlipidemia (Yes), Lentinotomy (No) ⇒ Hypertension (Yes)
 (支持度:4.96% / 確信度:91.18%)

Photocoagulation Therapy (Yes) ⇒ Retinopathy (not NDR)
 (支持度:12.63% / 確信度:96.93%)

った。これはFreeS/WANがWindows XPのL2TPに対応していないため、L2TPをサポートするソフトウェアモジュールを追加することで解決可能と考えられるが、今回は時間的に実験できなかった。

C-4. データマイニングおよび前処理について

前処理を施した糖尿病データベースに対し相関ルール発見手法を適用した。最小支持度を4、10、30、50%と変化させた。また最小確信度は90%とした。2.3節で述べた内部確信度としてそれぞれ、50、70、80、90%を用いた。実験結果を表2に示す。前述した内部確信度を用いたルール数の抑制は、ルールのサイズ(条件部と帰結

部に含まれる属性数)が2以上のルールに対してしか適用できない。表2中の括弧内の数字は、サイズが2以上のルール数、つまり実際に内部確信度による抑制が適用されたルールの数を示す。また得られたルールの例を表3に示す。

C-5. 動的診療根拠生成のフレームワーク

一つ目のユースケースである、流行性・一過性の疾患に対する診療根拠の動的な生成に関しては、データマイニング手法の一つである相関ルールを用いたルール発見、導出されたルールのルールベースによる管理、意思決定を支援するための問合せ機構によるルールベースの後解析、ユーザへの効果的なルール提示のためのルール削減・要約手法を用いた枠組みを提案する。二つ目のユースケースである、薬品の低頻度の副作用に対する診療根拠の導出に関しては、相関ルール発見の枠組みにおいて薬品の低頻度の副作用を表すルールの性質を考え、ルールの確信度の差に基づくパターンの発見を用いる。本研究により、動的な診療根拠の生成システムに必要な要件を定義することができ、この枠組みに対して適用可能なデータマイニング手法とその問題点を明らかにすることができた。この点については分担研究者の増田の報告書に詳しく

述べる。なお本研究項目についてはスロベニアのマリボル大学工学部システムデザイン研究科のPeter Kokol教授およびどう研究室のMilan Zorman, Matej Sproger, Mitja Lenicと日本学術振興会の二国間共同研究でも実施しており、本研究補助金でも研究遂行に極めて有効であるために、共同研究を行った。この共同研究の成果については2001年9月のMedinfo2001で発表した。

D. 考察

データの無名性に関して、表1は母集団として過去5年間に大阪医科大学付属病院で利用された患者基本情報32万件とその関連データを用いたものであり、たとえば母集団が1万件程度では生年月日を日まで特定すれば最小特定人数はほぼ1になることが予想される。また男女比はほぼ1対1と仮定した場合と実計算された最小特定人数はほぼ同じ値をしめしたが、年齢は対象地域(大阪府北摂地区で人口は約110万人)の年齢分布とは異なり、表には示していないが、60歳代にピークを持つ。また住所も全国的に見れば大きな偏りを示すことは自明であり、計算例でも高槻市が半数近くを占めている。町名レベルでも大きな公的病院のある地域では最小特定人数が著しく低くなる傾向があ

った。

最小特定人数をデータの二次利用の匿名性の根拠として用いる場合、説明の対象は調査の対象となる人であり、たとえば疾患分布などの専門的知識を持たないと考えざるを得ない。したがって疾患分布で母集団に特異性があるかどうかは検討していない。このように母集団の偏りが存在し、説明の対象者に医学知識の存在を仮定できないために、今回おこなった検討では男女比を除いて、予測値を用いることはできないと考えられる。一方で最小特定人数の計算アルゴリズムは単純であり、正しく項目整理されたデータベースがあれば、簡単であり、実用に用いる点で大きな問題はない。

問題は項目整理であり、多施設間研究などでは項目の定義が異なれば、最小特定人数の計算はできない。もちろん多施設間研究そのものために項目の同一定義は必要になり、その都度定義をそろえても理論的には計算可能であるが、データベースの再設計をする必要があり、また単純とはいえ、最小特定人数の計算アルゴリズムも毎回実装しなければならない。本研究ではJ-MIXを用いたが、これはきわめて有効であった。もちろんJ-MIXでは調査のための項目としては不足があることも考えられるが、かなり網羅的であり、少なくともJ-MIXを基本にすることで、

調査のためのデータベースの設計や最小特定人数の計算アルゴリズムの実装は大幅に簡略化されることが考えられる。

個人情報保護法が国会で審議中であり、おそらく平成14年度には成立するであろう。これまで診療情報の守秘義務に関しては各種の倫理綱領と刑法などに規定された医療専門職の守秘義務により守られてきた。しかし包括的な個人情報保護法が成立することで、状況は変化せざるを得ない。特に「守秘」から「プライバシー保護」への概念の変化は重要である。審議中の個人情報保護法はOECDプライバシーガイドラインに基本的に準拠している。OECDのガイドラインが現在のプライバシーに関する学説のすべてに一致しているわけではないが、おおむね先進国の共通の理解と一致している。日本の医療では少なくとも制度的には「守秘」が中心であったが、新しいプライバシーの概念は「守秘」だけではなく、自己情報のコントロール権が中心概念と言ってよい。したがって従来の秘密を守ればよいという考え方から、説明と同意を基本とするプライバシー保護に考え方をシフトする必要がある。本研究が目指す診療情報の収集とデータマイニングによる診療根拠の生成もプライバシー保護なしには実用化はできない。匿名化の研究で一定の成果は得たが、どの程度の匿名