

厚生労働科学研究費補助金(医療技術実用化総合研究事業(臨床研究・治験推進研究事業))
分担研究報告書

患者数調査のためのデータベースの構築に関する研究

研究分担者：松村泰志

大阪大学大学院医学系研究科医学専攻情報統合医学講座医療情報学 教授

研究要旨

治験ネットワークを構築する際に、各医療施設について、対象疾患の患者をどれくらい診療しているかを推定できると、バランスの良い配分数を割り当てることができる。レセプトデータの病名データは信頼性が低いとされるが、仕様された薬剤から逆引きすることで、ある程度高い精度で病名が推定できるのではないかと考えた。この目的で、昨年度構築した患者数推定システムを、今年度、更に次の処理を改良した。1) 推定対象を ICD10 の小分類までから細分類まで拡張した。2) 閾値を超えなかった場合に上位分類病名での処理を変更した。3) 病名推定に、性、年齢層による限定を付けた。4) 疑い病名を除いた。5) 規格が異なる同一成分の薬剤を同一薬剤として処理した。6) 外来、入院データを統合して処理した。新システムの処理結果を、ランダムに選択した 51 名の患者についてカルテレビューをして評価した。161 件の病名が推定された(細分類: 38 件、小分類: 94 件、中分類: 139 件、大分類: 161 件) 1 患者につき平均 3.2 病名が推定された。推定された病名のうち、正しい推定が 108 件(67.1%) 妥当な推定が 37 件(23.0%) 不適な推定が 16 件(9.9%) であった。不適な推定が入院患者に多いことから、更に改善策を講じる余地があることが分かった。以上の結果を踏まえ、次年度の改善点が明確となった。

研究協力者

武田理宏(大阪大学大学院医学系研究科)

A. 研究目的

臨床試験を行う場合、被験者を確保することが成否の鍵を握る。日本の場合、ある疾患を特定の医療機関に集中させるようなコントロールはされていないため、個々の医療機関では、十分な数の被験者が得られないことが多い。そこで、治験ネットワークを構築し、複数病院で仮想的に大規模病院を構成し、それぞれの臨床試験に対して十分な数の被験者を確保する体制作りが求められている。

治験ネットワークを構築する上で、各医療機関が診療する各疾患の患者数が分かると、治験の各医療機関への担当配分数を決める際に有効である。近年、多くの病院で電子カルテが普及し、そのデータを Data Warehouse (DWH) に保管することで、様々なデータ検索が可能となっている。こうした機能を利用することで、同一施設内であれば、各疾患の患者数を求めることは難しくない。近年、医療データの標準化が進んでいるものの、多くの病院ではハウスコードを使用するなど、統一したデータ構造となっていない。治験は規模の異なる複数の病院が参加し、場合によっては電子カ

ルテが導入されていない病院が参加することもある。臨床試験のニーズを満たすためには、複数の医療機関で統一的な方法で情報検索できなければ意味をなさない。現状では、レセプトオンライン化が制度化され、レセプトデータ、DPC データはコード、形式が標準化されており、かつ、各病院から既に出力機能が備わっている。

本研究では、レセプトデータのみを使用し、各疾患の患者数を推定することを目的とした。レセプトデータは、病名が含まれているが信頼性に乏しいことが問題である。一方、薬剤は、疾患に対して患者に投与されるものである。投与された薬剤の情報から逆引き的に病名の推定がある程度可能ではないかと考えた。昨年度、レセプトに登録される病名、薬剤名より各疾患の患者数を推定するシステムを構築した。薬剤名から病名への変換は一般財団法人日本医薬情報センター（JAPIC）が用意する「医薬品と対応病名データ」を活用した。病名は ICD-10 コードで与えられるが、ICD-10 コードは疾患の上位下位概念が明確でない。そこで大分類、中分類、小分類をもつ疾患コードを作成し、ICD-10 コード（3 桁）を「疾患コード」に変換した。病名の推定にあたっては、小分類までの推定を目標としたが、推定が出来ない場合は、中分類、大分類までの推定を行うものとした。病名の推定には疾患点数の概念を導入し、レセプトに記載された薬剤に対し疾患点数 10 点を付与し、「医薬品と対応病名データ」で対応づけられる病名に分配した。レセプトに登録される病名に対しては、疾患点数 4 点を付与した。疾患点数が 5 点を超える病名については、疾患名として採択した。病名の推定は小分類まで行うことを目標としたが、小分類では疾患点数が 5 点を越えなかった疾患については、中分類、大分類での推定を行った。

今年度は、昨年度に構築した患者数推定システムの病名推定ロジックを見直し、推定精度の向上を行った。

B．研究方法

1．ICD-10 小分類までの推定を細分類までの推定に変更

昨年度、ICD-10 コードの 3 桁を使用し、小分類までを疾患コードに変換した。しかし、表 1 に示すように（09, 07, 08）その他の脳血管疾患などのように、「その他」に分類される疾患を多数認めた。その他の脳血管疾患には、脳動脈瘤や脳動脈のアテローム硬化、もやもや病など明らかに病態の異なる疾患名が包括される。そこで、疾患コードを（大分類、中分類、小分類）に新たに細分類を追加し、ICD-10 コード 4 桁との対応付けを行った。細分類は 1 桁で表し、ICD-10 コード 4 桁の下 1 桁と一致する。推定アルゴリズムを以下のように変更した。

- ・ 大分類（2 桁） 中分類（2 桁） 小分類（2 桁） 細分類（1 桁）とした（表 1）。
- ・ ICD-10 コードが 3 桁の病名については、細分類を ^ としてコードを付与する（例：（01,01,01,^））。
- ・ 病名推定の際は、小分類→中分類→大分類の順の推定から、細分類→小分類→中分類→大分類の順に行く。

2．疾患点数加算方法の変更

昨年度では、病名に中分類まで共通する病名（例：（01,01,01）（01,01,02））が登録されていた場合、小分類の推定ではそれぞれに対し疾患点数 4 点が割り振られ治療薬との組み合わせで病名が採択されるが、中分類の推定では合計点数が割り振られ（上記の例では（01,01,^）に対し 8 点）レセプト病名のみで病名が採択されてしまった。そこで、今年度は病名が重複する場

合は、1 病名として疾患点数 4 点を割り振ることとし、以下のようにアルゴリズムを変更した。

- ・ 小分類が同一の（細分類のみ異なる）病名を、小分類の病名推測に用いる場合、1 病名として取り扱い、疾患点数 4 点を割り振る方針とした。
- ・ 中分類が同一の（小分類から異なる）病名を、中分類の病名推測に用いる場合、1 病名として取り扱い、疾患点数 4 点を割り振る。
- ・ 大分類が同一の（中分類から異なる）病名を、大分類の病名推測に用いる場合、1 病名として取り扱い、疾患点数 4 点を割り振る。

3．病名から対象が限定される疾患の取り扱い

病名には男性器の疾患や女性器の疾患、妊娠に関わる疾患、出生に関わる疾患など、性別や年齢が限定される疾患がある。男性患者に対し女性のみに発生する疾患に疾患点数を振ることは意味がなく、その疾患が採択された場合はデータの精度が低下する。そこで疾患コードに対して、対象患者を限定するフラグを立てた。

- ・ 男性のフラグ
(02：新生物, 09：男性性器, ^^, ^)
(04：内分泌, 栄養および代謝疾患, 04：その他の内分泌腺障害, 10：精巣＜睾丸＞機能障害, ^)、
(14：尿路性器系の疾患, 07：男性性器の疾患, ^^, ^) など
- ・ 女性のフラグ
(02：新生物, 08：女性性器, ^^, ^)
(04：内分泌, 栄養および代謝疾患, 04：その他の内分泌腺障害, 09：卵巣機能障害, ^)
(14：尿路性器系の疾患, 09：女性骨盤臓器の炎症性障害, ^^, ^)

(14：尿路性器系の疾患, 10：女性性器の非炎症性障害, ^^, ^)

(15：妊娠, 分娩および産じょく, ^^, ^^, ^) など

- ・ 年齢<1 歳のフラグ

(16：周産期に発生した病態, ^^, ^^, ^)

- ・ 年齢<20 歳のフラグ

(5：精神および行動の障害, 10：小児期および青年期に通常発症する行動および情緒の障害, ^^, ^)

4．疑い病名の取り扱い

昨年度は、レセプトに登録された病名は全て疾患点数を付与していたが、今年度はレセプト病名で疑い病名であった場合に、疾患点数を付与しないこととした。

5．同一薬剤の取り扱い

昨年度は「ワーファリン 1mg」と「ワーファリン 0.5mg」など、同一成分の薬剤であっても異なる製剤が併用されている場合、疾患点数が 2 重に加点された。今年度は薬剤コードを薬価基準収載医薬品コードの上 7 桁に対応させ、同一コードの薬剤を同一薬剤として処理した。

6．入院、外来の評価の統合

昨年度は入院、外来を分けて評価を行っていた。しかし、同一患者に入院と外来のデータがあった場合は一つのデータとして処理を行うことが望ましい。そこで、今年度は、入院、外来をまとめた評価とする処理を行った。この際、入院、外来で同一疾患名が登録された場合は、一方を除外する処理を追加した。

7．検証対象データ

検証は2015年1月に大阪大学医学部附属病院を受診した入院患者、外来患者のレセプト情報を使用し、病名の推定を行った。

入院患者1,238人、外来患者27,375人が対象となり、うち19,119人が病名推定の対象となった（薬剤を投与されていた）患者であった。投薬データはEFファイルより取得し、入院患者が74,820レコード、外来患者が86,010レコードであった。病名データはレセプト電算から取得し、入院が9,647レコード、外来が221,568レコードであり、2014年度システムに利用した。この中から疑い病名を除いた入院患者8,421レコード、外来患者200,058レコードを2015年度システムで利用した。

8．システム構成

使用したコンピュータは、CPUはIntel Core2 DUO 3.33GHz、メモリーは4.00GB、OSはWindows Vista Business (SP2)とした。この上で、今年度構築したシステム（2014年度システム）と昨年度構築したシステム（2013年度システム）を動かした。

C．研究結果

1．処理速度

疾患評価ファイルの作成に2013年度システムで13時間39分8秒、2014年度システムで28時間12分57秒、集計表ファイル作成に2013年度システムで15秒、2014年度システムで22秒を費やした。

2013年度システムでは患者17,206名に対し、83,515件の疾患名（1患者あたり 4.85 ± 5.88 件、小分類：567病名、76,640件、中分類：176病名、79,564件、大分類：20病名、83,515件）が、2014年度システムでは患者18,393名に対し、58,526件の疾患名（1患者あたり 3.18 ± 2.36 件、細分類：392病名、16,057件、小分類：338病名、41,610件、中分類：142病名、52,788件、大分類：20病名、58,526件）が推定された。疾患コードの細分類処理、病名から対象が限定される疾患の取り扱い、同一薬剤対応で処理時間は大幅に増加したと考えられる。病名推

定件数の減少は、同一薬剤対応や入院、外来の評価の統合などの影響と考えられる。

2．推定された病名

推定された大分類の病名を2014年度システムで推定件数の多かった順に表2に示す。11：消化器系の疾患、09：循環器系の疾患は病名件数に大きな変化はないものの、04：内分泌、栄養および代謝疾患は推定病名件数が9,894件から6,809件に減少していた。また、19：損傷、中毒およびその他の外因の影響、21：健康状態に影響をおよぼす要因および保健サービスの利用、16：周産期に発生した病態、15：妊娠、分娩および産じょくは、大きく推定病名数を減らした。04：内分泌、栄養および代謝疾患については、疑い病名の除外（2 - 4）と疾患点数加算方法の変更（2 - 2）の効果と考えられる。16：周産期に発生した病態、15：妊娠、分娩および産じょくは、病名から対象が限定される疾患の取り扱い（2 - 3）による効果と考えられる。さらに、21：健康状態に影響をおよぼす要因および保健サービスの利用は通常の病院受診で付きにくい疾患が記載されており、2014年度システムにより疾患推定の精度が上昇していることが推定された。

中分類まで推定された病名を表3に示す。2014年度システムで推定数の多かった疾患については、2013年度システムでも多く推定され、推定数も大きな変化は認めなかった。一方、2013年度システムでは多く（500件以上）推定されたが、2014年度システムでは少ない（50件以下）推定数であった疾患は、1110：消化器系のその他の疾患（2013年度：1731件、2014年度：7件）、1921：損傷、中毒およびその他の外因による影響の続発・後遺症（2013年度：1323件、2014年度：2件）、1502：妊娠、分娩および産じょくにおける浮腫、たんぱく尿および高血圧性障害（2013年度：1221件、2014年度：25

件) 2101: 検査および診査のための保健サービスの利用者 (2013年度: 1134件、2014年度: 0件) 0506: 生理的障害および身体的要因に関連した行動症候群 (2013年度: 647件、2014年度: 10件) 1508: その他の産科的病態,他に分類されないもの (2013年度: 606件、2014年度: 4件)であった。その他に分類される病名、女性のみが発生する病名、病院で診療を行う疾患としては適切でない病名が大きく数を減らす結果となった。

小分類まで推定された病名を表4に示す。中分類と同様に、2014年度システムで推定数の多かった疾患については、2013年度システムでも多く推定され、推定数も大きな変化は認めなかった。一方、2013年度システムでは多く(500件以上)推定されたが、2014年度システムでは少ない(20件以下)推定数であった疾患は、192002: 処置の合併症,他に分類されないもの (2013年度: 1754件、2014年度: 11件) 110609: 腸のその他の疾患 (2013年度: 761件、2014年度: 1件) 192009: 外科的および内科的ケアのその他の合併症,他に分類されないもの (2013年度: 661件、2014年度: 2件) 150805: 他に分類されるが、妊娠,分娩および産じょくに合併するその他の母体疾患 (2013年度: 599件、2014年度: 4件) 180807: けいれん,他に分類されないもの (2013年度: 548件、2014年度: 13件) 050605: 産じょくに関連した精神および行動の障害,他に分類されないもの (2013年度: 503件、2014年度: 1件)であった。いずれもその他に分類される病名であり、2014年度システム改造により推定の精度が上がっていると予想される。

最後に細分類まで推定された病名を表5に示す。細分類までの推定は2014年度システムのみであるが、日常臨床で多くみられる疾患が推定されている。

3 .2014 年度システムで推定された病名のカルテレレビューによる評価

次に初診が 2010 年 1 月 1 日以降 (ペーパレス電子カルテ運用開始後) の患者のうち、患者 ID の下 2 桁が 00 の患者に対して、カルテレレビューを行い 2014 年度システムで病名推定が正しく行われているか検討を行った。カルテレレビューの対象となった患者 51 名に対し、161 件の病名 (細分類: 38 件、小分類: 94 件、中分類: 139 件、大分類: 161 件) 患者当たり平均 3.2 病名が推定された。大分類の推定では、「11: 消化器系の疾患」が 38 件、「09: 循環器系の疾患」が 31 件、「13: 筋骨格系および結合組織の疾患」が 22 件、「06: 神経系の疾患」が 13 件と上位を占めた。カルテレレビューの結果、推定された病名のうち、正しい推定が 108 件 (67.1%)、妥当な推定が 37 件 (23.0%)、不適な推定が 16 件 (9.9%)であった。妥当な推定としては、抗凝固療法や非ステロイド性消炎鎮痛剤投与に伴うプロトンポンプインヒビターや H₂ ブロッカー投与による、胃潰瘍、胃炎の推定などであった。不適な推定として、本態性高血圧を二次性高血圧と推定したり、二次性肺高血圧を原発性高血圧と推定するなど、中分類、大分類では正しく推定されたと考えられる推定を 2 件認めた。他の不適な推定は、循環器疾患に対して施行される抗凝固療法を脳血管疾患と判定するなどであったが、「疑い」を付与されていない保険病名が登録されている症例が多かった。カルテレレビュー対象患者 51 名のうち 27 患者 (52.9%) は主病名 (病院受診の契機となっている病名) が正しく推定され、15 患者 (29.4%) は期間中に投与された薬剤では主病名の推測が困難な症例であった。9 患者 (17.6%) は主病名に対して薬剤が投与されていたが病名の推定漏れと考えられた症例であったが、うち 4 患者は眼科の点眼薬投与で正し

く主病名を推測できないケースであった。推定が難しかった主病名については、入院の手術症例が多く認められた。これは、投与される薬剤が主病名に対するものでなく、手術の副作用（抗生剤など）によるが多かったことによると考えられた。これらの病名は DPC 病名から取得することが可能であり、病名推定に DPC 病名の主病名と入院の契機となった病名については、疾患点数を 5 点付与するなど病名として確定させる、医療資源病名については疾患点数 4 点を付与して治療薬と合わせて病名の推定を行うことで、病名推定の精度を上げられると考えられた。

D. 考察

今年度までの取り組みにより、レセプト情報からある程度の精度をもって疾患名を推定できることが明らかとなった。しかし、更に改善させるべき点があることも判明した。次年度には、以下を実行する予定である。

まず、入院患者の推定精度を上げるために、DPC 病名を用いた疾患の推定を、ロジックを加えることとする。また、本プログラムは 1 ヶ月のレセプトデータを対象としていたが、1 ヶ月では患者が検査のみで来院するなど、治療薬と病名を正確に把握できない問題があった。このため、複数月（3 カ月を想定）のレセプトデータを取りまとめて、疾患推定プログラムに適用することが可能となるようにプログラム改造を行う。

さらに複数の病院の疾患推定プログラムを走らせることを考える。疾患推定プログラムの解析にはかなりの時間を要することが分かった。一方、患者 ID 等の個人識別情報は不要である。そこで、各病院のレセプトデータから疾患推定プログラムに必要なデータだけを抽出し匿名化処理をするプログラムを構築する。このデータをセキュ

アなネットワーク回線または CD 等にて解析施設（本研究では大阪大学医学部附属病院）に送り、解析施設で疾患推定プログラムを走らせる運用を想定する。

実際の臨床試験での活用の際には、DWH を持つ電子カルテを運用している複数の病院をモデル病院、DWH で抽出可能な複数の疾患をモデル疾患として、モデル病院ごとのモデル疾患の患者数を登録する。モデル病院を含む本プロジェクト参加病院は、レセプトデータから疾患数を推定し、モデル病院のモデル疾患の患者数からデータを補正することで、より正しい患者数の推定を行うことを想定する。

E. 結論

昨年度構築したレセプトデータ、DPC の EF データから患者数を推定するシステムについて以下のように処理を改良した。1) 昨年度システムでは ICD10 の小分類までを推定対象としたが、今年度は細分類までの推定対象とした。これによりその他分類の疾患を減らすことができた。2) 詳細分類病名で閾値を超えなかった場合に上位分類病名で再処理を行う際に、同一分類病名の点数を加算しないよう変更した。3) 病名推定に性、年齢層による限定を加えた。4) 疑い病名を除いた。5) 規格が異なる同一成分の薬剤を同一薬剤として処理した。6) 外来、入院データを統合して処理した。新しいプログラムで処理した結果を、ランダムに選択した 51 名の患者についてカルテレビューをして評価した。161 件の病名（細分類：38 件、小分類：94 件、中分類：139 件、大分類：161 件）患者当たり平均 3.2 病名が推定された。推定された病名のうち、正しい推定が 108 件（67.1%）、妥当な推定が 37 件（23.0%）、不適な推定が 16 件（9.9%）であった。不適な推定が入院患者に多いことから、更に改善策を講じる余地があることが分かった。以上を踏まえ、

次年度の改善点が明確となった。

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

武田 理宏、三原 直樹、真鍋 史朗、松村 泰志．レセプトデータを活用した患者病名の推定（医療情報学 vol34, 312-315）11月

H．知的財産権の出願・登録状況

なし。

ICD10	大分類		中分類		小分類		細分類	
I670	09	循環器系の疾患	07	脳血管疾患	08	その他の脳血管疾患	0	脳動脈壁の解離, 非破裂性
I671							1	脳動脈瘤, 非破裂性
I672							2	脳動脈のアテローム硬化(症)
I673							3	進行性血管性白質脳症
I674							4	高血圧性脳症
I675							5	もやもや病
I676							6	頭蓋内静脈系の非化膿性血栓症
I677							7	脳動脈炎, 他に分類されないもの
I678							8	その他の明示された脳血管疾患
I679							9	脳血管疾患, 詳細不明

表 1. ICD-10 コード（4 桁）から疾患コードへの変換例

大分類番号	病名(大分類)	件数	
		2014年度	2013年度
11	消化器系の疾患	11,851	12,557
09	循環器系の疾患	9,005	9,478
04	内分泌,栄養および代謝疾患	6,809	9,894
13	筋骨格系および結合組織の疾患	6,302	6,946
06	神経系の疾患	5,373	5,901
10	呼吸器系の疾患	2,941	3,297
12	皮膚および皮下組織の疾患	2,743	3,481
07	眼および付属器の疾患	2,502	2,405
05	精神および行動の障害	2,457	4,117
18	症状,徴候および異常臨床所見・ 異常検査所見で他に分類されないもの	1,727	5,573
01	感染症および寄生虫症	1,639	3,387
02	新生物	1,428	1,805
03	血液および造血器の疾患ならびに 免疫機構の障害	1,222	923
14	尿路性器系の疾患	1,148	2,133
19	損傷,中毒およびその他の外因の影響	949	6,399
08	耳および乳様突起の疾患	258	225
21	健康状態に影響をおよぼす要因および 保健サービスの利用	71	1,829
16	周産期に発生した病態	52	1,109
15	妊娠,分娩および産じょく	33	1,966
17	先天奇形,変形および染色体異常	16	90

表 2 . 推定された病名 (大分類)

中分類番号	病名(中分類)	件数	
		2014年度	2013年度
0408	代謝障害	4,193	5,971
0903	高血圧性疾患	3,861	4,786
0606	挿間性および発作性障害	3,430	2,870
1106	腸のその他の疾患	2,740	3,308
1312	骨障害および軟骨障害	1,967	3,333
0402	糖尿病	1,876	2,382
0906	その他の型の心疾患	1,318	4,383
0904	虚血性心疾患	1,245	1,789
1203	皮膚炎および湿疹	1,136	1,651
0907	脳血管疾患	1,116	1,629

表 3 . 推定された病名 (中分類: 2015 年度システム上位 10 病名)

小分類番号	病名(小分類)	件数	
		2014年度	2013年度
110605	その他の腸の機能障害	2,676	2,453
060607	睡眠障害	2,623	2,602
040809	リポたんぱく代謝障害 およびその他の脂(質)血症	2,596	2,538
110205	胃潰瘍	2,587	1,308
131202	骨粗しょう症, 病的骨折を伴わないもの	1,710	1,686
110209	胃炎および十二指腸炎	1,484	1,700
040202	インスリン非依存性糖尿病< NIDDM >	1,478	1,325
090301	その他のリウマチ性心疾患	1,337	2,196
070201	結膜炎	874	573

表４．推定された病名（小分類：2015年度システム上位10病名）

細分類番号	病名(細分類)	件数	
		2014年度	2013年度
1106050	便秘	2,031	
0606070	睡眠の導入および維持の障害[不眠症]	1,865	
0408100	炎症性関節炎および痛風結節性疾患の 徴候を伴わない高尿酸血症	654	
1302029	慢性関節リウマチ, 詳細不明	553	
0408095	高脂(質)血症, 詳細不明	539	
0607139	多発(性)ニューロパチー, 詳細不明	490	
0408090	純型高コレステロール血症	484	
1312025	特発性骨粗しょう症	322	
0706019	緑内障, 詳細不明	302	
0604019	アルツハイマー病, 詳細不明	292	

表５．推定された病名（細分類：2015年度システム上位10病名）