

厚生労働科学研究研究費補助金

政策科学総合研究事業

(統計情報総合研究事業)

医療分野 IT 化に対応した統計調査の在り方に関する研究

(H18-統計-プロ-001)

総括研究報告書

平成19年(2007年)3月

主任研究者 木村通男(浜松医科大学)

目 次

I. 総括研究報告書	2
医療分野 IT 化に対応した統計調査の在り方に関する研究	
主任研究者 木村通男	
II. 分担研究報告書	
1. 統計データ収集電子化の可能性	6
木村通男 (浜松医科大学), 藤田伸輔 (千葉大学), 大江和彦 (東京大学)	
2. 統計情報の改善: 電子レセの与える影響	8
大江和彦 (東京大学), 藤田伸輔 (千葉大学), 清谷哲朗 (労働者健康福祉機構)	
木村通男 (浜松医科大学)	
3. 統計情報の改善 オントロジーの導入	10
溝口理一郎 (大阪大学), 宮本正喜 (兵庫医科大学)	
研究協力者: 佐藤幹也 (帝京大学)	
4. 統計継続性についての研究	86
清谷哲朗 (労働者健康福祉機構), 藤田伸輔 (千葉大学)	
5. 地域医療連携に必要な情報の検討	88
小泉俊三 (佐賀大学), 藤田伸輔 (千葉大学), 宮本正喜 (兵庫医科大学)	
6. 国際統計情報に対する期待と改善への動き	92
藤田伸輔 (千葉大学), 大江和彦 (東京大学), 木村通男 (浜松医科大学)	
7. 国際的な統計調査との比較	98
広井良典 (千葉大学)	
研究協力者: 野村眞弓 (千葉大学)	

I . 総括研究報告書

平成18年度 厚生労働科学研究費補助金
政策科学総合研究事業（統計情報総合研究事業）
総括研究報告書

医療分野 IT化に対応した統計調査の在り方に関する研究

主任研究者 木村 通男 浜松医科大学医療情報部・教授

研究要旨

医療分野におけるIT化が急速に進展しつつある状況において、政府が実施する統計調査についても調査票に基づく限定的な情報収集の形式から、膨大な情報の中から真に必要なデータを峻別し、収集するという方向性に向けて、大きな転換が求められている。本研究は、プロジェクト提案型研究として、行政との対話を踏まえながら、時代のニーズに即した医療分野における統計調査の最適化を図るため、総合的かつ効果的な研究計画を策定するものである。

分担研究者・所属機関・職位

藤田伸輔・千葉大学・助教授
宮本正喜・兵庫医科大学・教授
清谷哲朗・労働者健康福祉機構・医療情報管理官
小泉俊三・佐賀大学・教授
廣井良典・千葉大学・教授
溝口理一郎・大阪大学・教授
大江和彦・東京大学・教授

A. 研究目的

（研究の目的及び必要性）

国民の医療行政に対するニーズが高度化し、また複雑多様化する状況において、医療行政の立案・評価の基盤となる統計データについてもより精緻なものが求められるようになってきている。一方、医療分野におけるIT化の急速な進展により、医療機関において膨大な電子データの蓄積が可能となり、政府が実施する統計調査についても調査票（紙ベース）に基づく限定的な情報収集の形式から、膨大な情報の中から真に必要なデータを峻別し、収集するという方針に向けての転換が迫られている。そこで以下のような観点から統計調査のあり方を再度検討する必要がある。（1）既存の調査事項に追加するという視点ではなく、膨大な情報の中で真に必要なデータは何である

かという視点からの検討。（2）データ提供サイド（医療機関）の負担が人的労力から費用負担に大きくシフトするため①データ提供サイドのコスト、②得られる情報の価値（データ収集サイドのみでなくデータ提供サイドにも価値のあるものでなくてはならない）、③両者を比較考量した費用対効果の検証方法等に関する検討（3）大量の情報が集まれば何でも解析できるという幻想を断ち切り、収集された情報によって解析しうる内容の限界を予定しておくという視点からの検討（4）電子化することが困難な他の補完的な調査（受療行動調査など）とのリンケージのあり方という視点からの検討（5）データの標準化及びバリデーションという観点からの検討（6）IT化への移行期における施設間の格差や時系列的な連続性を踏まえた調査設計のあり方という視点からの検討これらの検討課題は互いに連携しており、複雑かつ多面的な要素を含んでいるためプロジェクト提案型研究事業として高度で専門的な意見集約を行うことが必要である。

B. 研究方法

医療分野における統計の現状調査とあるべき形についての検討をグループディスカッションで討論することを中心に研究した。

C. 研究成果

(1) 効果的かつ効率的な統計調査に向けての調査設計の変更 (2) 政策立案・評価における統計データの利活用の促進 (3) 統計データ提供サイド(医療機関等)における便益の向上等が期待される。

各分担研究の概要

1. 統計データ収集電子化の可能性(分担研究1)

これまで医療機関のIT化は計画通りのスピードでは進んでこなかった。しかし平成20年度に開始される特定健診制度において健診結果を電子化して分析することから、国民の健康への意識とともに健康情報の電子化についても急速に関心が高まる可能性があると考えられる。その結果診療情報についても電子化し、自分で管理したいと考えるものも少なからず出現するであろう。患者からの要求は医療機関がIT化するうえでのインセンティブになると考えられる。またレセプトのオンライン化によって医療機関は経済的にもIT化を迫られることになる。このような状況によって今後IT化が急速に進展する可能性がある。

医療に関する統計はほとんどのものが医療機関を情報発信源とするものであり、医療機関のIT化に応じて統計情報収集についてもIT化することが望ましい。情報収集のためのフォーマットはCDAなどの標準的規格に従ったものとするのが望ましい。

2. 統計情報改善についての検討(分担研究2)

電子レセの導入に伴い、これを医療・保健分野の統計調査に活用することはこれまで調査していない新たな指標を得られる可能性がある。またその網羅性からもよりきめの細かいデータを医療機関の負担を増やさずに収集できる可能性がある。しかし電子レセを統計情報源として用いることは目的外使用であるための問題点もある。すなわち医療経済の適正化の観点から診療報酬制度の改正が行われるため、なるべく同じデータを得たい統計の立場からは対応方法を熟考すべきであり、目的が異なるため同じ用語であっても診療報酬制度と統計とは意味が違う場合がある。これらの点に配慮して電子レセを統計に活用す

るための体制と行動指針を作成しなければならない。

3. 統計情報の改善、オントロジーの導入(分担研究3)

わが国の医療・保健分野では数多くの統計調査が実施されている。これらを適切に用いれば非常に有用なデータを得ることができ、海外の医療・保健分野の統計調査に決して引けをとるものではない。しかしこれらの調査の頻度や内容などを一覧できる仕組みは無く、その全貌を知ることが容易ではない。統計調査にIT化を取り入れるなどこれからの改善計画を実施していくうえでも全体像を確認しながら進めていく必要があり、本研究ではオントロジーを導入し、統計調査を整理することの意義を検討した。

4. 統計継続性についての検討(分担研究4)

統計の目的は過去のデータとの比較が重要であることを鑑みて、統計の継続性について十分な配慮が必要である。統計データの採取は実験におけるデータ測定と同様に、真の値に対して測定方法による測定誤差が発生する。測定方法を変更する際には同じデータを二方法で測定したり、同じ群に属すると思われるデータを二方法に無作為割付したりして測定方法による違いを推定する。統計におけるデータ収集方法の電子化では、紙と電子の両方でデータ提出を求めても電子データを出力するだけであろうから、データ提出方法を紙と電子に割り付けるなどを検討すべきである。

5. 地域医療連携に必要な統計情報(分担研究5)

地域医療連携において重要な情報は患者・家族居住地から医療機関及び医療・介護施設への到達性である。目的とする医療サービスによって医療機関及び医療・介護施設を抽出する機能抽出と、対象施設への交通手段によるアクセス時間抽出が必要となる。アクセス時間の推定などは住所情報をGPSデータに変換し、交通手段によってアクセス時間を推定する地図ソフトなどを活用可能であるが、医療機関及び衣料・介護施設の特性データは現在の医療機関調査で知りえない。地域医療連携および地域医療計画の作成といった観点から医療機関における診療行為内容について

の調査が臨まれ、医療機関の位置情報には住所に加えて GPS データを付加することが望まれる。

6. 国際統計情報に対する期待と改善への動き(分担研究6)

国際社会において地球温暖化、人口爆発、新興感染症、災害被害の甚大化、食糧危機などが問題になっている。これらの影響を正しく評価するためには統計調査そのものを大幅に見直す必要がある。このような観点から WHO においては Health Metrics Network 研究が実施されている。また世界規模の健康危機に備えるため International Health Regulation が 36 年ぶりに改訂され、平成 19 年 6 月から施行される。わが国においてもこれらの動きと無縁ではいられないためこれら 2 つについて検討を行った。

7. 国際的な統計調査との比較(分担研究8)

本研究は統計情報の電子化に関するプロジェクト型研究という観点から、保健医療分野の国際標準統計の電子的な利用について検討した。

OECD や EU では、各国の国内統計を基礎的な情報源から編成される国際統計を整備し、各種の指標を開発している。それらの指標は、重要性や科学的合理性とともに、データの入手可能性と比較可能性が評価されている。

今後、わが国の医療統計の電子化においては、国際的な標準統計へのデータの提供可能性や比較可能性を研究する必要があると考えられた。また、Eurostat における加盟国間の統計情報の収集・分析・提供方法をベンチマークとした調査研究が有用となろう。

D. 考察

グローバル化、地球温暖化、高齢化社会の到来などにより医療・保健分野についてより細やかなデータを迅速に収集・分析・公開しなければならないようになってきている。このような目的に沿うためには統計調査の IT 化を検討しなければならない。また社会全体の IT 化に伴い統計調査を IT 化することも医療従事者のリテラシーの向上によりハードルが低くなりつつあり、電子レセ制度や患者本人への診療情報提供 (SS-MIX) や特定健診制度など情報を電子的に収集する環境も整い

つつある。

統計は社会の変化を記載する手段であり、変化を捉えるためには継続性という重要な使命がある。一方社会の変化を捉えるためには新たな指標の作成も随時行わなければならない。継続性という観点からは統計調査の IT 化によってデータの意味が変化したり内容に偏りが出たりする可能性もあり、この点についての検討が必要である。とくに電子レセを情報源として用いる場合には診療報酬制度の改正に対してどのように対処するのかを十分に検討しなければならない。

社会の変化を捉えるために新たな指標を作成するという観点からは作成すべき指標と、その指標をどこからどのように集めたデータによって算出するのかという検討が必要であり、本研究の中では欲しい指標といった観点からの検討を行ったが、実現のためにはデータ収集の具体策についての研究も必要である。

E. 結論

平成 20 年度に開始される特定健診制度およびレセプトオンライン化によって医療情報の電子化が急速に進展すると予想される。このため医療分野における統計調査も電子化に対応してより迅速に、かつ詳細なデータを収集できるようにすることが必要である。その一方で統計の継続性という点にも配慮し、移行に際してデータの偏りがどの程度発生するのか、過去のデータと比較する際に考慮すべき点は何かといった事項を十分検討しておかなければならない。

F. 健康被害情報

現在のところ報告すべき情報はない。

G. 本研究の成果(予定を含む)

医療・保健分野の統計を改善することはわが国にとって必要不可欠であり、グローバル化や地球温暖化問題などから国際社会から見ての不可欠となっている。本研究はプロジェクト型研究という性格上実際の改善のための作業には着手できなかったが、今後本研究の成果を生かして研究を発展させられる環境を切に要望する。

II. 分担研究報告書

平成18年度 厚生労働科学研究費補助金
政策科学総合研究事業（統計情報総合研究事業）
分担研究報告書

1. 統計データ収集電子化の可能性

分担研究者 浜松医科大学 木村通男
千葉大学 藤田伸輔
東京大学 大江和彦

研究要旨

これまで医療機関のIT化は計画通りのスピードでは進んでこなかった。しかし平成20年度に開始される特定健診制度において健診結果を電子化して分析することから、国民の健康への意識とともに健康情報の電子化についても急速に関心が高まる可能性があると考えられる。その結果診療情報についても電子化し、自分で管理したいと考えるものも少なからず出現するであろう。患者からの要求は医療機関がIT化するうえでのインセンティブになると考えられる。またレセプトのオンライン化によって医療機関は経済的にもIT化を迫られることになる。このような状況によって今後IT化が急速に進展する可能性がある。

A. 研究目的

これまで医療機関のIT化は計画通りのスピードでは進んでこなかった。しかし平成20年度に開始される特定健診制度において健診結果を電子化して分析することから、国民の健康への意識とともに健康情報の電子化についても急速に関心が高まる可能性があると考えられる。その結果診療情報についても電子化し、自分で管理したいと考えるものも少なからず出現するであろう。患者からの要求は医療機関がIT化するうえでのインセンティブになると考えられる。またレセプトのオンライン化によって医療機関は経済的にもIT化を迫られることになる。このような状況によって今後IT化が急速に進展する可能性がある。

医療分野の統計調査のデータは多くが医療機関で採取される。患者調査など本来はもっと頻繁に調査し、詳細なデータを収集したほうがきめ細やかな政策に反映できる。しかし紙ベースでの個票収集では、記入、収集、整理、登録といった統計を使えるようにする準備に多大な労力を要しコストパフォーマンスの点からもこれ以上の頻度での調査は不可能である。

統計調査の電子化は医療機関での個票記入から統計データベースへのデータ格納までのプロセスに要する労力を大幅に削減し、現状のまま継続するのであればコスト削減を、同じコストをかけるのであればよりサンプル数を増加させることを可能とする。

本研究ではこのような観点から医療機関を情報源とする統計情報について電子化収拾の可能性について検討した。

B. 研究方法

電子化情報収集の基盤として、データのフォーマット、伝送方法、セキュリティについて検討した。

C. 研究結果

① 電子的収集の可能性

現在用いられている医療分野の統計情報は多くが医療機関から収集されるものである。個票に記入されるデータのほとんどが医事コンピュータに入力されている項目でもあり、医事のデータを利用することも可能であろう。しかし手書きの個票のデータと医事コンピュータ（レセコン）から出力されるデータとは必ずしも一致しないことが問題となり、この点については清谷の分担報告書を参照されたい。

② データのフォーマット、

電子的に抽出する場合そのデータフォーマットが問題となるが、Eurostatでも、あるいはHealth Metrics Networkでも、データフォーマットについては規定されていない。できるだけ国際的な規格に準じたほうが良いと考え、またコンピュータでも人間でも理解可能なフォーマットということで、HL7のCDAに準じるのが最適であるとの結論となった。ただしCDAではカバーされない項目が多数存在し、詳細な検討を加えて日本からHL7に提案することが最適と判断した。

レセコンから出力すると仮定して、個票の項目をコンピュータから抽出し、CDA 規格に沿って記録し、暗号化してから抽出することになるが、個票の種類は多数ある。多種存在する個票の中で同一の調査項目も存在するため、それぞれの個票に対応したプログラムを作成するのではなく、同じプログラムで全ての調査に対応できることが理想である。すなわちレセコンへのデータ出力要求は各ベンダー独自仕様になるが、同じフォーマットで出力されれば、そこからの CDA 化は [A11:胸痛]プログラムとする事が可能である。

③ データ暗号化と通信

個票には匿名化できないものも存在し、データ暗号化が重要となる。暗号化の強度等の検討は専門家に譲るが、HL7CDA に準じて個票を作成すると、個票一枚一枚に対して暗号化を行うことになり、それぞれに違う複合キーを設定すればセキュリティは向上するが、非常に面倒なものとなる。一施設分をひとつの CDA で扱う Document とみなすと暗号化・複合化の手間はほぼ妥当なものとなる。

一施設分の個票データをひとつの CDA Document として CD-R 等に収録し、その搬送中に紛失すると全てのデータを読まれてしまう可能性が高くなる。個票は匿名化されたデータであり、それを暗号化しているので万一情報を読み取られても直接的被害はほとんど発生しないとも考えられるが、社会的には大きなダメージとなることも考えられる。紙に記載された個票ではその体積と重さが搬送中も注意を惹起するが、電子媒体は小型軽量であるがゆえに危険性を増すという矛盾をはらんでいる。

また新興感染症やテロリズムなど甚大な健康被害の発生が懸念されてまとめられた IHR2005 に対応するためには情報収集の即時性という問題も浮上する。IHR2005 では 24 時間体制で WHO との間で情報交換することが規定されており、医療情報の IT 化が強く求められている。このため CD-R などの電子媒体を用いた情報伝達に対して決して安全とはいえないが、Network を用いた情報伝達についても積極的に対応すべきであろう。

④ データの保管と公開

個票データを整理匿名化し、データベース

に保管するプログラム、現在公開されている帳票の形にデータベースからデータを抽出するプログラムが必要である。

国民が統計資料を利用する際にはいくつかの帳票をあわせて分析しているので、Eurostat のようにデータベースから必要な項目を抽出できれば便利である。データの効率的利用を推進するためには各データが持つ意味を明確にすることも重要である。

D. 考察

本研究はプロジェクト型研究であるため、医療分野の統計を電子化する際の方向性を探るにとどめた。

医療分野の統計データ収集を標準化する方法としては HL7・CDA 規格を拡張することが妥当であると考えるが、わが国単独で拡張して使用するのではなく、International Health Regulation 2005 や Health Metrics Network の精神を尊重し、WHO, Eurostat, OECD などと協調していくことが重要である。

一方で国民が利用しやすいデータベースという視点も重要である。現在公表されているいくつかの帳票から目的とする地域のデータを抜き出して利用するといったことが日常的に行われている。このような作業を毎年繰り返していることを考慮して、目的とするデータを効率的に収集する仕組みについても、統計情報の電子化の際に対応することが望ましい。

E. 結論

医療機関のほとんどがレセコンを導入している現在、統計調査の多くの部分を電子化する事が可能である。しかしその導入に当たっては個票データの収集に HL7・CDA 規格の拡張、および電子化された統計情報データベース利用の促進のため、Eurostat を参考にデータ利用のための環境改善が必要である。

F. 健康被害情報

現在のところ報告すべき情報はない。

G. 参考文献

1. HL7・CDA
2. Eurostat
3. ISO/TC215

2. 統計情報の改善：電子レセの与える影響

分担研究者	東京大学	大江和彦
	千葉大学	藤田伸輔
	労働者健康福祉機構	清谷哲朗
	浜松医科大学	木村通男

研究要旨

医療・保健分野の統計調査はその多くの部分が医療機関を情報発信源とするものであり、さらにその大部分がレセプトコンピュータに収められた情報である。レセプトが電子化され、オンライン請求となろうとしている中、統計調査を電子レセプト情報から収集すればよいという考え方も成り立つ。電子レセを情報源とした場合、個票作成のための労力を削減できること、診療費の分析が可能になることなど大きなメリットが存在する。一方診療科の分類や診療報酬算定におけるルールの影響などレセプトデータを解釈する上で配慮しなければならない点もある。そこで電子レセを情報源とする場合のメリットと注意点についての検討が必要である。

A. 研究目的

医療・保健分野の統計情報収集のIT化はその医療・保健分野の統計調査はその多くの部分が医療機関を情報発信源とするものであり、さらにその大部分がレセプトコンピュータに収められた情報である。レセプトが電子化され、オンライン請求となろうとしている中、統計調査を電子レセプト情報から収集すればよいという考え方も成り立つが、収集方法を変更すれば得られるデータも変化する。本研究では電子レセを情報収集源に変更する場合に考慮すべきことは何かについて検討した。

B. 研究方法

医療・保健分野の統計調査の個票を検討し、電子レセプトから情報収集する場合の問題点について討論した。

C. 研究結果

1) 診療科

現在の調査は内科・外科分類を基盤とする診療科分類（診療科Ⅰ類、Ⅱ類、Ⅲ類）が採用されている。しかし「現実に行われている診療＝医療機関で外来に使用されている診療科名」は老人科、成人病科といった内科分類の中での複合化、循環器科から冠動脈疾患外来と高血圧外来の分化や外科から肝胆膵外科・食道胃腸外科・乳腺外科への分化などより細分化にむかうもの、消化器内科と消化器外科を癒合せた消

化器科といった診療科分類をまたぐ融合、女性科といった診療科分類との位置づけが困難な分類の出現などが存在する。電子レセプトではこれらの診療科で行われた行為も診療科分類上のどこかで行われた行為として請求されている。レセプト上では診療行為の内容とレセプト上での診療科名との整合性を厳密には要求されず、日々の請求の中で診療行為によって分類しなおして請求することも困難なため、医療機関における診療科名を一義的に診療科分類に当てはめて請求していることが多い。一方患者調査ではどの診療科が主たる診療を行ったのかを人手により調査し個票に記入しているため、電子レセプトから診療科に関する調査を行うと大きなバイアスが発生する可能性がある。また電子レセプトから主たる診療科を自動決定することも困難であり、この点でも問題が発生する。一方電子レセプトから全ての受診診療科を調査できれば診療科ごとの診療費を分析し、さらに病名ごとの診療費を計算することも可能になり、新たな指標として活用できる可能性もある。

2) 診療費

電子レセから統計情報を得ることのメリットとして診療費の分析が可能になることがある。高齢社会における医療費の高騰が非常に大きな問題であり、医療費の分析を行えるようになることは非常に重要である。

現在の診療報酬の枠組みでは一医療機関内で複数の診療科を受診した場合は単一診療科とみなして（複数診療科の個別初診料を認めない・同一日の複数科再診料を認めないなど）いるが、複数医療機関を受診すればそれぞれ独立して扱っている。このため診療費の正確な解析のためには医療機関内に存在する診療科と電子レセの情報を付き合わせ、さらに診療報酬制度とつき合わせた解析が必要になる。

診療報酬制度はおおむね2年に一度の改正が行われ、必要に応じて年に2回行われたこともあった。統計調査は継続性が問題となるため、できるだけ同じ基準に基づくデータを数年以上にわたり収集すべきであるため、電子レセを情報収集源とする場合は統計調査でデータを振り分けるロジックについても診療報酬制度の改正にあわせて作り直す必要がある。

3) 北欧での診療報酬データ活用

北欧各国では一般医からの診療報酬制度をICDではなくICPCを用いたオンライン請求としている。ICPCはICD10に比較すると非常に荒い分類であり、一種の外来版DPC制度である。このためコンピュータ通信の時代からオンライン請求制度が発達し、政府はリアルタイムで診療状況の概略を把握している。一般医＝かかりつけ医＝家庭医は疾患にかかわらず住民一人一人を総合的に診察しているため一般医が下す診断の割合はどの一般医もある程度同じになるはずである。そこで診断分布を調査して偏りの大きい医師や病院への紹介率の高い医師を抽出し、調査を行ったり指導したりしている。本邦においても外来診療へのDPC導入を実現すれば同じような調査を行い、医療の均質化に貢献できる可能性がある。

D. 考 察

医療・保健分野の統計調査に電子レセを活用することは現在未調査の情報を得られる可能性がある。しかし診療報酬という目的から、医療経済の指標を得る場合、あるいはさらに目的を離れて患者調査といった疾病構造の調査に利用する場合、目的外使用によるデータのバイアスをどのようにコントロールするのかという観点からも検討が必要である。

診療報酬制度の改正は社会情勢の変化に応じて医療経済の観点から随時行われる。これはなるべく継続性を維持しなければならない統計調査とは相容れないものであるが、統計調査上の都合から診療報酬制度の改定を制限することは本末転倒になる。このため電子レセを活

用する場合、診療報酬制度改正があった場合にどのように対処するのか、対応のための体制作り、その体制における指針作成などを検討しなければならない。

電子レセと医療・保健分野の統計調査では診療科名のように同じ名称であっても意味が異なる場合が存在する。DPC制度における主たる診療科名と患者調査における主たる診療科名は同一でない場合も存在する。そこで矛盾が発生する可能性があるかどうか一語一語について検討することが必要である。

E. 結 論

電子レセを統計調査の情報源として用いることは新たな指標を得たり、よりきめの細かいデータを得たりすることができる可能性がある。しかし同一に見えて違うデータとなる可能性も存在しこれを整理すること、診療報酬制度の改正にあわせて統計調査を再検討する必要があることなど十分な検討が必要である。

F. 健康被害情報

現在のところ報告すべき情報はない。

G. 参考文献

特に無い

平成18年度 厚生労働科学研究費補助金
政策科学総合研究事業（統計情報総合研究事業）
分担研究報告書

3. 統計情報の改善 オントロジーの導入

分担研究者 大阪大学 溝口理一郎
兵庫医科大学 宮本正喜
研究協力者 帝京大学 佐藤幹也

研究要旨

医療分野の統計は数多くの帳票が公開されている。これらを有効利用するためには各帳票がどのような意味を持つのかを理解しやすく整理することが重要である。本研究ではオントロジーエディター法造を用いて医療分野の統計についてのオントロジー作成を行った。医療分野の統計を IT 化する際には統計全般を概観し、データ発生から正規化を経てデータベース格納までを IT の視点から整理しなおす必要がある。オントロジーはこのような目的に非常に有用であることを確認した。

A. 研究目的

統計情報収集の IT 化に際して、収集項目や頻度の見直しを行うことが望ましい。統計改善のためにその使用目的を明確にすること、情報提供者に過度の負担とならないように留意すること、Eurostat などとの比較を考慮し、国際標準化を目指している Health Metrics Network の動きに注目すること、統計の継続性について配慮することが必要である。

統計使用目的の明確化は統計情報を国民が広く活用する上で不可欠である。調査項目と目的を明確化する手段としてオントロジーを用いることを検討し、統計調査についてのオントロジーを構築し、その有用性を検討することとした。

B. 研究方法

現在公開されている医療分野の統計を元にオントロジーを構築した。オントロジー構築ツールとして法造^{1,2}を用いた。

C. 研究結果

医療分野の統計についてのオントロジーを作成した。対象は厚生労働省が実施主体となっている医療分野の統計である。調査主体からみたオントロジーの一部を Fig.1 に示す。どの部局がどの統計調査を行っているかを表現した。ロールの観点からは厚生労働省内の部局の表現は isa 関係ではなく part-of 関係で示すべきであるが、ここでは統計調査に絞って表現したため isa 関係で表現した。

調査の種類から見た調査項目のオントロジーの一部を Fig.2 に示す。このオントロジーでは実際には存在しないが、概念の包括関係を示す上で追加したものには「p-」のラベルを付して表現した。また今後情報中に含めたほうが資料活用に有用だと思われる情報、たとえば医療機関の住所から一義的に決まるが併記されているほうが便利な東経・北緯といった情報には「r-」のラベルを付して表現した。

D. 考察

オントロジーとは概念を整理するためのツールであり、統計分野の複雑な情報を整理するものとして最適である。本研究ではオントロジーエディターとして法造を用いた。世界的には Protégé がもっとも流通しているが、本研究では概念関係の表現がより厳格な法造を採用した。法造は古崎助手を中心として溝口研究室で開発してきたオントロジー構築・利用環境である。その特徴をあげると以下のようになる。

- (1) is-a と part-of / attribute-of 関係を中心としたフレーム（ノードとスロット）構造を基本としている。
- (2) フレンドリーな GUI を備えており、使いやすい。
- (3) 先端的ロール理論に準拠しており、ロールを簡単に表現できる世界唯一のツールである。
- (4) オントロジー記述言語 OWL への出力機能を持つ。

(5)オントロジーを複数のコンポーネントに分割して複数の人間の協調による分散開発を支援する機能がある。

(6)オントロジーの変化を明示する GUI とバージョン管理機能がある。

(7)オントロジー利用環境として必要な API が整備されている。

(8)多言語対応も可能であり、国産ではあるが成果物を海外に公表する際に障害はない。

本研究で法造をベースに統計オントロジーを開発する理由は上記の特徴の特に1と3と5にある。ロール理論に基づき厳格な isa 関係と part-of 関係の整理はオントロジー構築に際し不自由と指摘する意見もある。しかし不自由とを感じる場面ではひとつと考えていた概念が複数の意味を持っているがための混乱による場合がほとんどであると考えられる。このような概念の曖昧さは大きな概念集合を扱う上で致命的な誤解を生む可能性を秘めており、本研究では法造を選択した。

本研究が目指す統計オントロジーの第一の目的は医療分野の統計を IT ベースで改善する際に利用することである。各統計の調査実施者、調査対象、調査頻度、調査方法、調査項目などを整理することが不可欠であり、本研究はその基盤とするべく実施した。

現在厚生労働省統計表データベース³からは目的とする統計表を容易に入手可能であり、調査項目を示す個票の用紙も入手可能である。しかしこれらの資料は膨大であり、統計資料の全体像を把握するにはかなりの努力を要する。本研究ではこれらのデータを用いてオントロジーを作成した。調査主体から見たオントロジーは厚生労働省の各部局がどのような調査を行っているかを示すものであり、調査名から検索

すればどの部局が調査しているのかが明瞭になる。また IT 化に際して必要な調整もどの範囲で行えばよいか明瞭である。一方統計項目のオントロジーは、各調査間で重複する項目が明瞭であり、データ収集から正規化をへてデータベースへ格納するまでのデータフローを作成する際に有用である。もちろんこの二つで IT 化やユーザに便利なオントロジーを全て構築できたわけでもなく、またこれらのオントロジーはプロトタイプであるが、その有用性は評価できた。

厚生労働省統計表データベースには非常に多くの情報が公開されており、Eurostat にも引けをとらない規模であるが、どのような統計があるのかを知らないとい検索も困難である。本研究の成果は医療分野の統計の全体像を俯瞰しやすいことにあり、国民が統計諸表を利用する際にも有用な道しるべになる。本研究はこのような観点からの使用にも耐えうるものへと発展させたい。

E. 結論

医療分野の統計は非常に数が多く、また調査項目も多い。調査項目は複数の調査で共通するものもあり、これらはコード化されている。従って統計調査を IT 化するには調査内容をオントロジーで整理し、全体の整合性を保ちながら IT 化の準備を進めることが必須である。またこの準備は資料として公開すれば、国民が統計調査の全貌を知り、データを有効活用する際の補助となると考えられる。

F. 健康被害情報

現在のところ報告すべき情報はない。

Fig. 1. 医療分野の統計のオートロジ

法連 オートロジエタ
 フォル 編集 表示 ウィンドウ Role ヘルプ
 *編集パネル - 【厚生労働統計総括表to.xml】

概念ツリー 検索
 概念ツリー表示 関係概念表示

全体概念表示

75%

part-of

NE

概念表示 RDF表示

ラベル 血液製剤使用状況調査
 上位概念 (血液製剤使用状況調査)
 内容説明 公理
 スロウト 継承スロウト

値

概念表示 RDF表示

厚生労働統計総括表to.xml activated ...

3 Window... 2 Internet... JP
 法連... CHANDLO... 2 Internet... JP
 12:49 日曜日

G. 参考文献

1. 溝口：オントロジー工学，オーム社，2005
2. 溝口編著：オントロジー構築入門，オーム社，2006
3. 厚生労働省統計表データベース
<http://www.dbtk.mhlw.go.jp/toukei/index.html>

<医療分野の統計のオントロジー>

<?xml version="1.0"?>

<rdf:RDF

xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">

xmlns:owl="http://www.w3.org/2002/07/owl#">

xml:base="http://www.hozo.jp/owl/厚生労働統計総括_sato.xml/">

>

<owl:Ontology rdf:about="">

<rdfs:comment>

HOZO:OWL Export TEST

</rdfs:comment>

</owl:Ontology>

<owl:Class rdf:ID="RelationalConcept">

<rdfs:label>RelationalConcept</rdfs:label>

</owl:Class>

<owl:ObjectProperty rdf:ID="hasPart">

<rdfs:label>hasPart</rdfs:label>

</owl:ObjectProperty>

<owl:ObjectProperty

rdf:ID="hasAttribute">

<rdfs:label>hasAttribute</rdfs:label>

</owl:ObjectProperty>

<owl:Class rdf:ID="保健統計室">

<rdfs:label>保健統計室</rdfs:label>

<rdfs:subClassOf rdf:resource="#人口動態_保健統計課" />

</owl:Class>

<owl:Class rdf:ID="指定統計">

<rdfs:label>指定統計</rdfs:label>

<rdfs:subClassOf rdf:resource="#統計調査" />

<rdfs:subClassOf>

<owl:Restriction>

<owl:cardinality

rdf:datatype="http://www.w3.org/2001/XMLSchema#int">

1

</owl:cardinality>

<owl:onProperty

rdf:resource="#has_実施周期"/>

</owl:Restriction>

</rdfs:subClassOf>

<rdfs:subClassOf>

<owl:Restriction>

<owl:onProperty rdf:resource="#has_実施周期"/>

<owl:allValuesFrom rdf:resource="#周期" />

</owl:Restriction>

</rdfs:subClassOf>

<rdfs:subClassOf>

<owl:Restriction>

<owl:cardinality

rdf:datatype="http://www.w3.org/2001/XMLSchema#int">

1

</owl:cardinality>

<owl:onProperty

rdf:resource="#has_最終調査年度"/>

</owl:Restriction>

</rdfs:subClassOf>

<rdfs:subClassOf>

<owl:Restriction>

<owl:onProperty rdf:resource="#has_最終調査年度"/>

<owl:allValuesFrom rdf:resource="#年度" />

</owl:Restriction>

</rdfs:subClassOf>

<rdfs:subClassOf>

<owl:Restriction>


```

        <owl:cardinality
rdf:datatype="http://www.w3.org/2001/X
MLSchema#int">
            1
        </owl:cardinality>
        <owl:onProperty
rdf:resource="#has_指定統計番号"/>
        </owl:Restriction>
        </rdfs:subClassOf>
        <rdfs:subClassOf>
        <owl:Restriction>
        <owl:onProperty rdf:resource="#has_
指定統計番号"/>
        <owl:allValuesFrom rdf:resource="#
実数" />
        </owl:Restriction>
        </rdfs:subClassOf>
        <rdfs:subClassOf>
        <owl:Restriction>
        <owl:cardinality
rdf:datatype="http://www.w3.org/2001/X
MLSchema#int">
            1
        </owl:cardinality>
        <owl:onProperty
rdf:resource="#has_指定年月日"/>
        </owl:Restriction>
        </rdfs:subClassOf>
        <rdfs:subClassOf>
        <owl:Restriction>
        <owl:onProperty rdf:resource="#has_
指定年月日"/>
        <owl:allValuesFrom rdf:resource="#
年月日" />
        </owl:Restriction>
        </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="加工統計">
    <rdfs:label>加工統計</rdfs:label>
    <rdfs:subClassOf rdf:resource="# 統計
調査" />
    <rdfs:subClassOf>
    <owl:Restriction>
        <owl:cardinality

```

```

MLSchema#int">
            1
        </owl:cardinality>
        <owl:onProperty
rdf:resource="#has_実施周期"/>
        </owl:Restriction>
        </rdfs:subClassOf>
        <rdfs:subClassOf>
        <owl:Restriction>
        <owl:onProperty rdf:resource="#has_
実施周期"/>
        <owl:allValuesFrom rdf:resource="#
周期" />
        </owl:Restriction>
        </rdfs:subClassOf>
        <rdfs:subClassOf>
        <owl:Restriction>
        <owl:cardinality
rdf:datatype="http://www.w3.org/2001/X
MLSchema#int">
            1
        </owl:cardinality>
        <owl:onProperty
rdf:resource="#has_最終調査年度"/>
        </owl:Restriction>
        </rdfs:subClassOf>
        <rdfs:subClassOf>
        <owl:Restriction>
        <owl:onProperty rdf:resource="#has_
最終調査年度"/>
        <owl:allValuesFrom rdf:resource="#
年度" />
        </owl:Restriction>
        </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="医療施設静態調査">
    <rdfs:label>医療施設静態調査
</rdfs:label>
    <rdfs:subClassOf rdf:resource="# 医療
施設調査" />
</owl:Class>
<owl:Class rdf:ID="歯科診療所票（医療施設
統計）">
    <rdfs:label>歯科診療所票（医療施設統
計）</rdfs:label>
    <rdfs:subClassOf rdf:resource="# 医療
施設静態調査" />

```

```

</owl:Class>
<owl:Class rdf:ID="一般診療所票（医療施設統計）">
  <rdfs:label>一般診療所票（医療施設統計）</rdfs:label>
  <rdfs:subClassOf rdf:resource="#医療施設静態調査"/>
</owl:Class>
<owl:Class rdf:ID="衛生行政報告例調査票">
  <rdfs:label>衛生行政報告例調査票</rdfs:label>
  <rdfs:subClassOf rdf:resource="#衛生行政報告例"/>
</owl:Class>
<owl:Class rdf:ID="医療施設動態調査">
  <rdfs:label>医療施設動態調査</rdfs:label>
  <rdfs:subClassOf rdf:resource="#医療施設調査"/>
</owl:Class>
<owl:Class rdf:ID="患者調査">
  <rdfs:label>患者調査</rdfs:label>
  <rdfs:subClassOf rdf:resource="#保健統計室"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:cardinality
rdf:datatype="http://www.w3.org/2001/XMLSchema#int">
        1
      </owl:cardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty
rdf:resource="#has_統計の種類"/>
      </owl:Restriction>
    </rdfs:subClassOf>
  </rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty rdf:resource="#has_統計の種類"/>
    <owl:allValuesFrom rdf:resource="#指定統計"/>
  </owl:Restriction>
</rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="一般診療所票（患者調

```

```

査）">
  <rdfs:label>一般診療所票（患者調査）</rdfs:label>
  <rdfs:subClassOf rdf:resource="#一般診療所調査"/>
</owl:Class>
<owl:Class rdf:ID="一般診療所退院票">
  <rdfs:label>一般診療所退院票</rdfs:label>
  <rdfs:subClassOf rdf:resource="#一般診療所調査"/>
</owl:Class>
<owl:Class rdf:ID="病院（偶数）票">
  <rdfs:label>病院（偶数）票</rdfs:label>
  <rdfs:subClassOf rdf:resource="#病院調査"/>
</owl:Class>
<owl:Class rdf:ID="病院入院（奇数）票">
  <rdfs:label>病院入院（奇数）票</rdfs:label>
  <rdfs:subClassOf rdf:resource="#病院調査"/>
</owl:Class>
<owl:Class rdf:ID="社会福祉施設等調査">
  <rdfs:label>社会福祉施設等調査</rdfs:label>
  <rdfs:subClassOf rdf:resource="#社会統計課"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:cardinality
rdf:datatype="http://www.w3.org/2001/XMLSchema#int">
        1
      </owl:cardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty
rdf:resource="#has_統計の種類"/>
      </owl:Restriction>
    </rdfs:subClassOf>
  </rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty rdf:resource="#has_統計の種類"/>
    <owl:allValuesFrom rdf:resource="#承認統計"/>
  </owl:Restriction>
</owl:Class>

```

```

    </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="介護利用型軽費老人ホーム（ケアハウス）入所者票">
  <rdfs:label>介護利用型軽費老人ホーム（ケアハウス）入所者票</rdfs:label>
  <rdfs:subClassOf rdf:resource="#社会福祉施設等調査" />
</owl:Class>
<owl:Class rdf:ID="医師届出票">
  <rdfs:label>医師届出票</rdfs:label>
  <rdfs:subClassOf rdf:resource="#医師_歯科医師_薬剤師調査" />
</owl:Class>
<owl:Class rdf:ID="病院票">
  <rdfs:label>病院票</rdfs:label>
  <rdfs:subClassOf rdf:resource="#医療施設静態調査" />
</owl:Class>
<owl:Class rdf:ID="歯科診療所票（患者調査）">
  <rdfs:label>歯科診療所票（患者調査）</rdfs:label>
  <rdfs:subClassOf rdf:resource="#歯科診療所調査" />
</owl:Class>
<owl:Class rdf:ID="病院外来（奇数）票">
  <rdfs:label>病院外来（奇数）票</rdfs:label>
  <rdfs:subClassOf rdf:resource="#病院調査" />
</owl:Class>
<owl:Class rdf:ID="病院退院票">
  <rdfs:label>病院退院票</rdfs:label>
  <rdfs:subClassOf rdf:resource="#病院調査" />
</owl:Class>
<owl:Class rdf:ID="歯科医師届出票">
  <rdfs:label>歯科医師届出票</rdfs:label>
  <rdfs:subClassOf rdf:resource="#医師_歯科医師_薬剤師調査" />
</owl:Class>
<owl:Class rdf:ID="厚生省報告例">
  <rdfs:label>厚生省報告例</rdfs:label>
  <rdfs:subClassOf rdf:resource="#衛生

```

```

行政報告例" />
</owl:Class>
<owl:Class rdf:ID="児童福祉施設等調査票">
  <rdfs:label>児童福祉施設等調査票</rdfs:label>
  <rdfs:subClassOf rdf:resource="#社会福祉施設等調査" />
</owl:Class>
<owl:Class rdf:ID="保育所調査票">
  <rdfs:label>保育所調査票</rdfs:label>
  <rdfs:subClassOf rdf:resource="#社会福祉施設等調査" />
</owl:Class>
<owl:Class rdf:ID="有料老人ホーム入居者票">
  <rdfs:label>有料老人ホーム入居者票</rdfs:label>
  <rdfs:subClassOf rdf:resource="#社会福祉施設等調査" />
</owl:Class>
<owl:Class rdf:ID="特別養護老人ホーム入所者票">
  <rdfs:label>特別養護老人ホーム入所者票</rdfs:label>
  <rdfs:subClassOf rdf:resource="#社会福祉施設等調査" />
</owl:Class>
<owl:Class rdf:ID="老人_障害者福祉施設等調査票">
  <rdfs:label>老人・障害者福祉施設等調査票</rdfs:label>
  <rdfs:subClassOf rdf:resource="#社会福祉施設等調査" />
</owl:Class>
<owl:Class rdf:ID="受療行動調査">
  <rdfs:label>受療行動調査</rdfs:label>
  <rdfs:subClassOf rdf:resource="#保健統計室" />
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:cardinality
rdf:datatype="http://www.w3.org/2001/XMLSchema#int">
        1
      </owl:cardinality>

```

```

    <owl:onProperty
rdf:resource="#has_統計の種類"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#has_
統計の種類"/>
      <owl:allValuesFrom rdf:resource="#
承認統計" />
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="外来患者票">
  <rdfs:label>外来患者票</rdfs:label>
  <rdfs:subClassOf rdf:resource="# 受療
行動調査" />
</owl:Class>
<owl:Class rdf:ID="入院患者票">
  <rdfs:label>入院患者票</rdfs:label>
  <rdfs:subClassOf rdf:resource="# 受療
行動調査" />
</owl:Class>
<owl:Class rdf:ID="地域保健事業報告調査
票">
  <rdfs:label>地域保健事業報告調査票
</rdfs:label>
  <rdfs:subClassOf rdf:resource="# 地域
保健事業報告 (保健所運営報告) " />
</owl:Class>
<owl:Class rdf:ID="地域保健_老人保健事
業報告調査票">
  <rdfs:label>地域保健・老人保健事業報告
調査票</rdfs:label>
  <rdfs:subClassOf rdf:resource="# 地域
保健_老人保健事業報告" />
</owl:Class>
<owl:Class rdf:ID="伝染病月報">
  <rdfs:label>伝染病月報</rdfs:label>
  <rdfs:subClassOf rdf:resource="# 伝染
病統計" />
</owl:Class>
<owl:Class rdf:ID="擬似症_インフルエン
ザ_ましん患者票">
  <rdfs:label>擬似症・インフルエンザ・ま
しん患者票</rdfs:label>
  <rdfs:subClassOf rdf:resource="# 伝染

```

```

病月報" />
</owl:Class>
<owl:Class rdf:ID="性病者票">
  <rdfs:label>性病者票</rdfs:label>
  <rdfs:subClassOf rdf:resource="# 伝染
病月報" />
</owl:Class>
<owl:Class rdf:ID="伝染病患者票">
  <rdfs:label>伝染病患者票</rdfs:label>
  <rdfs:subClassOf rdf:resource="# 伝染
病月報" />
</owl:Class>
<owl:Class rdf:ID="病院報告">
  <rdfs:label>病院報告</rdfs:label>
  <rdfs:subClassOf rdf:resource="# 保健
統計室" />
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:cardinality
rdf:datatype="http://www.w3.org/2001/X
MLSchema#int">
        1
      </owl:cardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <owl:onProperty
rdf:resource="#has_統計の種類"/>
  </owl:Restriction>
</rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#has_
統計の種類"/>
      <owl:allValuesFrom rdf:resource="#
承認統計" />
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="患者票">
  <rdfs:label>患者票</rdfs:label>
  <rdfs:subClassOf rdf:resource="# 病院
報告" />
</owl:Class>
<owl:Class rdf:ID="従事者票">
  <rdfs:label>従事者票</rdfs:label>
  <rdfs:subClassOf rdf:resource="# 病院
報告" />
</owl:Class>

```