

平成24年2月10日	資料2
第8回レセプト情報等の 提供に関する有識者会議	

基本データセットについて

平成24年2月10日

厚生労働省保険局総務課

目次

第7回有識者会議を踏まえて	3
整理する基本データセットの類別	4
基本データセット整備の考え方	5
「サンプリングデータセット」の具体案: データ内容	6
「サンプリングデータセット」の具体案: 申出等	12
「サンプリングデータセット」の具体案: 提供予定(案)	14
「教育/練習用データセット」の具体案	15
(参考) 関連資料	16

第7回有識者会議を踏まえて

抜粋資料

➤ 『「事前審査の論点を踏まえた、事務局審査方針(研究内容・抽出について)②」』

同じく「個人の識別可能性を下げる」という原則に鑑み、多数の項目を用いた探索的研究や、SYの「傷病名コード」、SIの「診療行為コード」、IYの「医薬品コード」(DPCの場合にはBU, SB, CD)どれかひとつでも「全て求める」という要望の申出は、事務局審査では提供依頼について不承諾とした。

- 多変量解析、propensity score分析、重回帰分析など多くの項目を必要とする申出も探索的研究と考え、今回は不承諾とした。
- 何らかの抽出を経たあとのサンプルに対しても、上記のような申し出に対しては同様の判断とした。
- この方針は特定健診の情報を用いた研究にも適用した。
- 上記のように、研究手法によってはこうした抽出条件とならざるを得ないものもあり、これも今後検討予定の「基本データセット(仮称、後述)」構築によって対応することが考えられる。

➤ 『「今後の予定について」「基本データセット(仮称)の作成について」』

- 「傷病名」「診療行為」「医薬品」コードにおいて「全数希望」の申出が今回複数みられたが、個人が特定化される可能性、というリスクを考えると、データベースから全ての情報を提供することは困難である。
- 一方で、レセプト情報等の活用が最善でありかつ公共性の高い研究については、個人が特定されてしまう可能性のみを根拠として一概に研究の門戸を閉ざすべきではない、という意見もある。
- 次回以降の有識者会議で、ナショナルデータベースから一定の抽出を行い匿名性を高めた「基本データセット(仮称)」の構築について検討していくこととしたい。
- 「基本データセット(仮称)」の活用により、研究デザイン等の理由からやむをえず項目の全数が必要となる研究においても、個人が特定される可能性を可能な限り低めることができうと考えられる。

➤ 『「今後の予定について」「その他」』

- 申出者や研究者が、レセプト情報等でどのような研究ができるのか等の理解の一助となるよう、ダミー値のみで構成された架空のデータセットの作成についても、検討を行っていく。

➤ 探索的研究に対するニーズ

- 現在の審査方針では、「傷病名」「診療行為」「医薬品」コードのいずれかすべてを必要とする申出は、事務局審査方針「研究内容・抽出について②」に照らせば、承諾することができなくなる。
- 一方、**第一回申出で不承諾とした申出の複数**がこうした申出に該当しており、探索的研究に対する研究者のニーズが高いことをうかがわせる。こうした状況を考えると、探索的研究に対しても門戸を開くことができるよう、海外の先進事例も踏まえ、何らかの対応を検討する必要がある。

整備する基本データセットの類型

➤ 台湾でのデータセット提供の事例を参考として

種別	概要	作成方法等
基本資料データ	特定項目の集計表情報	入院レセプト、外来レセプトの特定項目を集計したもの
系統抽出データ	一定の割合で抽出をかけ、匿名性を高めた月単位のデータセット	入院レセプトの5%を抽出 外来レセプトの0.2%を抽出
特定主題データ	疾患に応じ、母集団の構成等を調整し抽出したデータセット	16種類の特定疾患等（悪性新生物等）に沿って抽出したデータファイル
ランダム抽出データ	まとまった数のレセプトを抽出し一定期間紐付けたデータセット	2000年は20万人（1996年～2007年分）、 2005年は100万人分を抽出
教育用データ	研究者向けの練習用ツール	1000人分のランダムサンプリング 教育用に無償提供

前回の有識者会議における議論を踏まえれば、我が国において整備の緊急性および必要性が高いものは、上記のデータセット種別のうち

系統抽出データ

教育用データ

なのではないか。

基本データセット整備の考え方

1. 今回整備する基本データセットとは

- 第7回の有識者会議を経て浮かび上がってきた「探索的研究」への大きな要望や、申出に対する今後の継続的な提供体制について考えたとき、現在のような「全数データ」以外に個人の特定可能性の観点から安全性に優れ、探索的研究などにも使用できる基本データセット：前頁の「系統抽出データ」を参考とした

「サンプリングデータセット」（仮称）

を早急に整備する必要があると考えられる。

- 一方、傷病や診療行為等、さまざまな研究目的に対して最適化されたデータセットの整備は、平成24年度の厚生労働科学研究を活用することにより、学術的な観点からの精度を確保したデータセットの作成を予定している。
- 加えて、レセプト情報等でどのような研究ができるのかの理解の一助となる(いわば教育や練習用の)架空のデータセット（「教育/練習用データセット（仮称）」）の提供もホームページ等を利用して進めていきたい。

2. 基本データセット整備のスケジュール案

「サンプリングデータセット（仮称）」：平成23年度中に

「教育/練習用データセット（仮称）」：平成23年度中に

その他の基本データセット：平成24年度～

「サンプリングデータセット」の具体案：データ内容

1. どのレセプトデータで整備するか？

- 「医科」「DPC」「調剤」「歯科」「特定健診」「特定保健指導」すべてで整備するか、それとも整備するレセプトを一定の範囲に絞るかを決める必要がある。
- まず第一回申出で利用希望が多かった「医科」「DPC」「調剤」のデータセット作成から開始することとしてはどうか。
- 「歯科」については、第一回申出において利用の申出が少なく、電子化率も40%前後にとどまっているため、データセット作成は今後の課題としてはどうか。
- 同様に、「特定健診」「特定保健指導」についても、第一回申出において利用の申出が少なかったことなどから今後の課題としてはどうか。

2. どの程度の期間のデータを提供するのか？

- 複数月のデータを提供することとなると、単月分のみとの情報提供とは異なり、紐付け等により個人の特定可能性が高まるなどの課題がある。こうしたデータセットの作成および提供は初めての試みであり、試行期間中でもあることから、当面は単月のデータのみを提供することとしてはどうか。

3. 抽出の割合、手法をどうするか？

- 「医科」では1カ月当たり7000万件以上(平成21年度)にもものぼり、ある程度の抽出を行わなければ、以下の点で支障を来す恐れがある。このため、一定の抽出を行ってはどうか。
 - 一般的な可搬媒体(DVD-ROMなど)に収まりきれない恐れがある。
 - データ量が大きいと研究者の操作が煩雑になる。
- 抽出割合の設定や、母数の異なる入院レセプトと外来レセプトで抽出割合を変えるかどうかも検討課題である。
 - 例：台湾での抽出割合は入院では5%、外来では0.2%
- 抽出に際しては、統計的な偏りを少なくするための最低限の措置として、患者の基本的な属性である性別、年齢構成(5歳刻み年齢区分)については、構成割合をデータベースに保存されている全数データの構成割合と同じものとしてはどうか。

「サンプリングデータセット」の具体案：データ内容

4. 匿名化処理をどう行うか？

- レセプトに出現する回数が少ない情報(たとえば「傷病名」、「診療行為」、「医薬品」コード)が含まれていると、それらの情報から個人が特定されてしまう可能性が高くなる。このため、レセプトに出現する回数が少ないコードについては、**一定の割合で匿名化処理を行う**こととしてはどうか。
- マスターが用意されている各コード(「傷病名」「診療行為」「医薬品」など)において出現回数の低いものを一定数匿名化すると仮定する。その際、レセプトに出現する回数を基準にして匿名化の基準を定めるとなれば、どの程度の数の傷病名コードを匿名化することになるだろうか？

例：循環器内科外来に通院する方の以下AからEの5枚のレセプトにおいて、個人が特定される可能性を下げるため、これら5枚のレセプトに記録されている傷病名を、出現回数を基準として少ないものから**10%**匿名化するとしたら？

※ この事例は架空の設定にもとづいたものであり、必ずしも実態を反映したものではない。

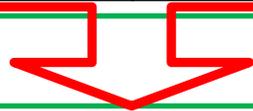
The image displays five sample prescriptions, labeled A through E, each with a red circle highlighting a specific injury name. Below each prescription is a callout box listing the injury names for that prescription.

- A**: 傷病名
 - 高血圧
 - 高脂血症
 - 糖尿病
 - うつ病
- B**: 傷病名
 - 高血圧
 - 糖尿病
 - 狭心症
 - 痛風
 - 触覚鈍麻
- C**: 傷病名
 - 高血圧
 - 糖尿病
 - 狭心症
 - 痛風
 - 硝子体炎
- D**: 傷病名
 - 高血圧
 - 高脂血症
 - 狭心症
- E**: 傷病名
 - 高血圧
 - 高脂血症
 - うつ病

「サンプリングデータセット」の具体案: データ内容

集計結果

	1	2	3	4	5	6	7	8	
傷病名	触覚鈍麻	硝子体炎	うつ病	痛風	糖尿病	高脂血症	狭心症	高血圧	合計
出現回数	1	1	2	2	3	3	3	5	20
レセプト	B	C	A, E	B, C	A, B, C	A, D, E	B, C, D	A, B, C, D, E	
全出現回数に占める割合	5%	5%	10%	10%	15%	15%	15%	25%	100%



A

傷病名

- 高血圧
- 高脂血症
- 糖尿病
- うつ病

B

傷病名

- 高血圧
- 糖尿病
- 狭心症
- 痛風
- 触覚鈍麻

C

傷病名

- 高血圧
- 糖尿病
- 狭心症
- 痛風
- 硝子体炎

D

傷病名

- 高血圧
- 高脂血症
- 狭心症

E

傷病名

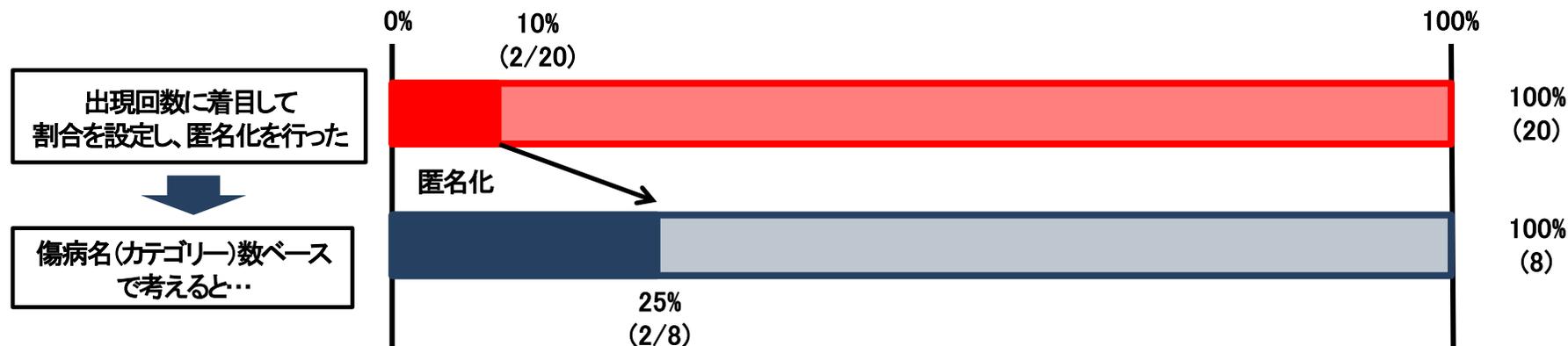
- 高血圧
- 高脂血症
- うつ病

BとCの区別がつかなくなった

「サンプリングデータセット」の具体案: データ内容

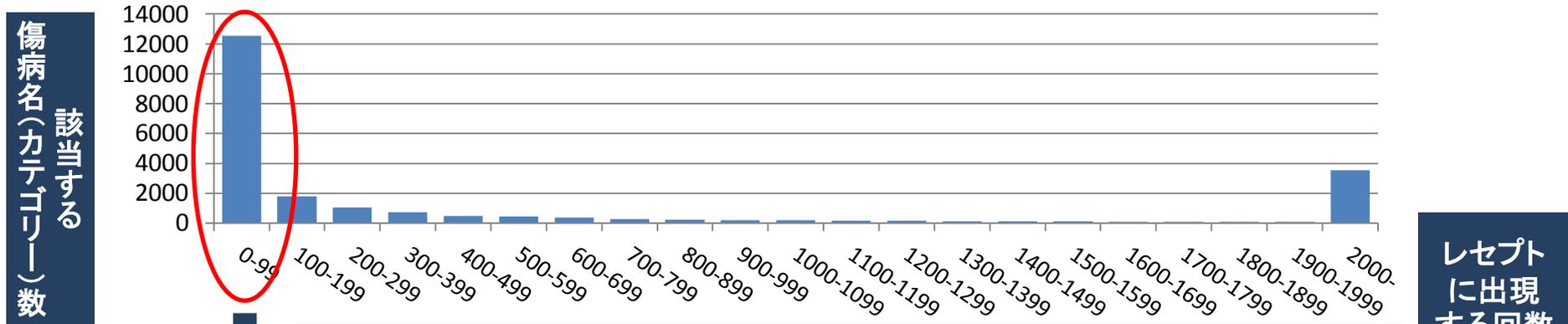
傷病名(カテゴリー)数とレセプト出現回数の関係

- この事例では5枚のレセプトの匿名性を高めるため、5枚のレセプトに出現する傷病名の出現回数の少ないものから「10%」を匿名化することを考えた。
- 集計結果から、1度しか出現しなかった「触覚鈍麻」「硝子体炎」を合計すると10%に達したためこれらを匿名化した。その結果、傷病名からは[B]と[C]の区別がつけられなくなるなど、5枚のレセプトの匿名性を高めることができた。
- しかし、「出現回数」を「10%」に設定することで匿名化した傷病名は「触覚鈍麻」と「硝子体炎」の2傷病名(カテゴリー)であり、これはこの5枚のレセプトに出現する全ての傷病名(8傷病名(カテゴリー)):「触覚鈍麻」「硝子体炎」のほか、「うつ病」「痛風」「狭心症」「高脂血症」「糖尿病」「高血圧」のうち、「25%」に相当する。
- つまり、出現回数の少ない傷病名や出現回数の多い傷病名があるため、傷病名(カテゴリー)数からみた匿名化の割合は、「出現回数」を基準にして設定した匿名化の割合よりも高い割合をとることとなる。これを帯グラフで表すと、以下のようになる。



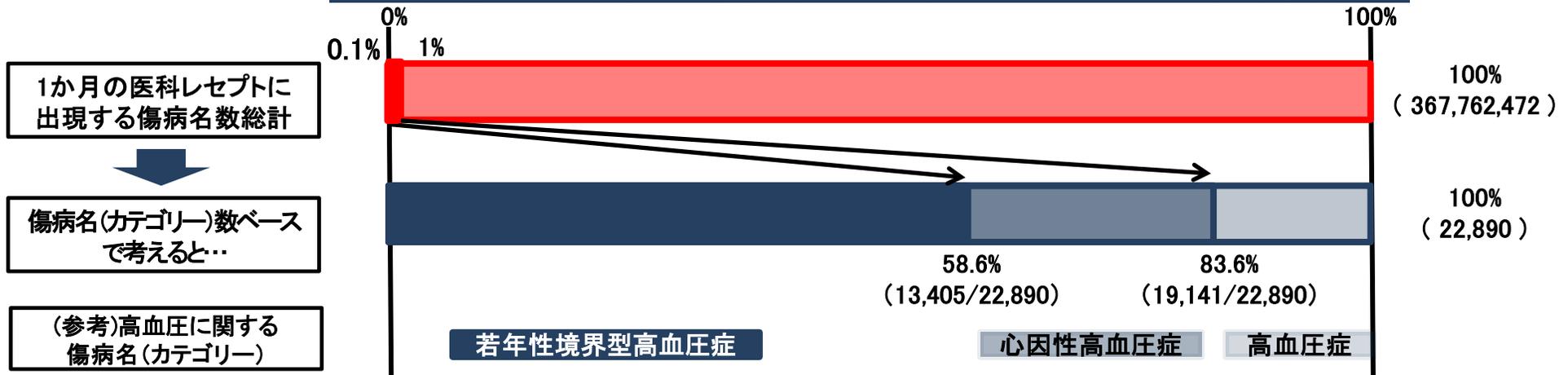
「サンプリングデータセット」の具体案：データ内容

(参考) ある月の医療レセプトにおける各傷病名(カテゴリー)の出現回数から



ひと月に100回未満しかレセプトに出現しない傷病名(カテゴリー)が12,000以上と、傷病名(カテゴリー)全体(22,890として計算。傷病名(カテゴリー)数はマスターの更新時期によって変動する)の半分超を占めている。したがって、レセプトに出現してくる傷病名(カテゴリー)のほとんどは、出現回数の高い数10パーセント程度の傷病名(カテゴリー)でカバーされているのが実態である。

例：この月の場合、レセプトに記録される傷病名の出現回数のうち99%は、傷病名(カテゴリー)全体の16.4%、3,749の傷病名(カテゴリー)のみでカバーされている。下図参照。



「サンプリングデータセット」の具体案：データ内容

5. 「サンプリングデータセット」高額のリセプトをどう扱うか？

- 請求点数が高額となるリセプトも、その個人が特定される可能性が高くなるため、一定の点数以上のリセプトを高額群として同一のコードに再整理して扱うのはどうか。(例：保険局調査課で使用している高額群の基準…「入院：700,000点」「入院外、調剤：50,000点」)
→ただし、医療費に関する調査への利便性も考慮し、上記高額群として扱われたリセプトの点数については、実際の総点数を追加事項として記載し、研究者が適宜活用できるようにすることでどうか。

6. 「サンプリングデータセット」に含めるデータの範囲をどう設定するか？

- 匿名性をより高め、また抽出により構成割合がナショナルデータベースに保存されている全数データと同一にならなくなるという観点から、次のようなデータ項目等については提供しないこととしてはどうか。
 - 「公費医療(生活保護を除く)」対象リセプト
 - 「医療機関コード」
 - 「保険者番号」
 - 「都道府県」

「サンプリングデータセット」の具体案：申出等

1. 申出資格等について

- 第一回申出で複数の申出者が該当した審査方針「研究内容・抽出について②」は、基本データセットの審査においては許容されるため、申出の多数が承諾となる可能性が高い。申出資格を制限しない場合、現行のデータ提供等の体制を考慮すると、データ提供が大幅に遅れるなどの懸念がある。
- 一方、第一回審査で不承諾となった申出者においては、審査の過程で事務局とのコミュニケーションが図られており、審査方針やセキュリティ要件を具備する意義についても既に理解していただいていると考えられる。
 - こうした背景を踏まえ、例えば初回のサンプリングデータセットの申出資格としては、**第一回審査で不承諾となった申出者に限定すること**としてはどうか。
 - なお、提供するデータの期間を単月に限定した場合、複数月、複数年のデータ利用を前提とした第一回申出の研究内容についてはそのまま実施することは不可能となる。この場合、**申出内容（研究方法）を必要に応じて変更したうえで**申し出ていただくこととなる。

2. 申出書内容・セキュリティ要件について

- 通常の「レセプト情報等の提供」と同様の申出書内容、セキュリティ要件を求めるかどうか
 - 匿名性が高められたデータとはいえ個票データであることから、基本的には同様のセキュリティ要件としてはどうか。
 - 申出書の様式は基本的に通常の「レセプト情報等の提供」を使うこととしてはどうか。ただし、探索的研究を見据えた匿名性の高いデータセットであることから、申出書に記載される公表形式には、**通常の「レセプト情報等の提供」ほどの具体性および網羅性を求めない**こととしてはどうか。

「サンプリングデータセット」の具体案：申出等

3. 提供・審査体制について

- 今回の「サンプリングデータセット」提供については、今年度末～来年度に予定している第二回申出と、受付時期をほぼ同一とする予定である。ただし、基本データセットの申出審査と第二回申出審査の時期については、「サンプリングデータセット」の審査および提供を先に行うこととして、両者の審査期間が重ならないように調整してはどうか。
- 第一回申出で不承諾となった申出者においては「第二回申出」と「サンプリングデータセット」の両方を申出することができる。但しデータ提供においてはサンプリングデータセットも個票として扱われるものである。このため、
 - ① 原則としてこれら情報は他情報との照合を認めていないこと、また
 - ② できるだけ多くの方々に使用されることによりその課題を明らかにする必要があること

から、申出者あるいは利用者がレセプト情報等を用いた複数の研究に同時に関与することは認めないこととしてはどうか。

- この考え方をあてはめれば、「サンプリングデータセット」の提供を承諾された申出者・利用者は、このデータを用いた研究を終了するまで、レセプト情報等の第三者提供を活用した他の研究が不可能となる。
- この考え方は、第一回申出で承諾を受けた研究者にも適用してはどうか。
- 模擬申出についてはガイドライン作成以前に行ったものであり、この考え方をあてはめないこととしてはどうか。

4. 「サンプリングデータセット」の今後について

- 今後は、この「サンプリングデータセット」を利用した研究者の方々からも様々な意見を集めるとともに、平成24年度厚生労働科学研究での検討も踏まえ、データセットの匿名性に配慮しつつ、さらなる利便性の向上を図っていくこととしたい。

「サンプリングデータセット」の具体案：提供予定（案）

第一回申出で不承諾となった者が次回申出を行う場合

サンプリングデータセット申出

第二回申出

事前説明会
(両方同時申請可能)

受付期間

審査

提供

受付期間

事前相談

審査

提供

「サンプリングデータセット」の提供を受けた場合、利用期間中は通常の全数データの提供を受けることはできない。

- 提供までの所要が短期間
- 全数調査ができない

- 提供までに要する日数が長期にわたる
- 承諾された場合、全数調査が可能となる

3月

4月

5月

6月

7月

8月

「教育/練習用データセット」の具体案

➤ 作成方法について

- 台湾で提供している教育用データセットは、レセプト研究を行う大学院生等に向け、多数のデータを申請に応じてCD-ROM形式の媒体で提供されるものである。一方、今回整備を予定している「教育/練習用データセット」は、データベースに保存されているレセプト情報等の実形式について研究者が簡単に把握でき、イメージを深めてもらうためのサンプルとしての位置づけにとどめることとし、以下のような形式としてはどうか。
 - データセットはホームページから閲覧、およびダウンロードできる程度の容量に限り、安全性を十分に保ったものとする。おおよそ数十例程度を各レセプト種別ごとに用意する。
 - 研究者がイメージしやすいよう、このデータセットはデータベースの提供形式と同一のものとする。
 - 「医科」「DPC」「調剤」「歯科」「特定健診」「特定保健指導」すべてに対し、データセットを準備する。

➤ 提供方法について

- 「教育/練習用データセット」の提供は、整備ができ次第、順次厚生労働省ホームページ内にある

「レセプト情報・特定健診等情報提供に関するホームページ」

(http://www.mhlw.go.jp/seisakunitsuite/bunya/kenkou_iryuu/iryuuhoken/reseputo/)

を通じて提供していくこととしたい。

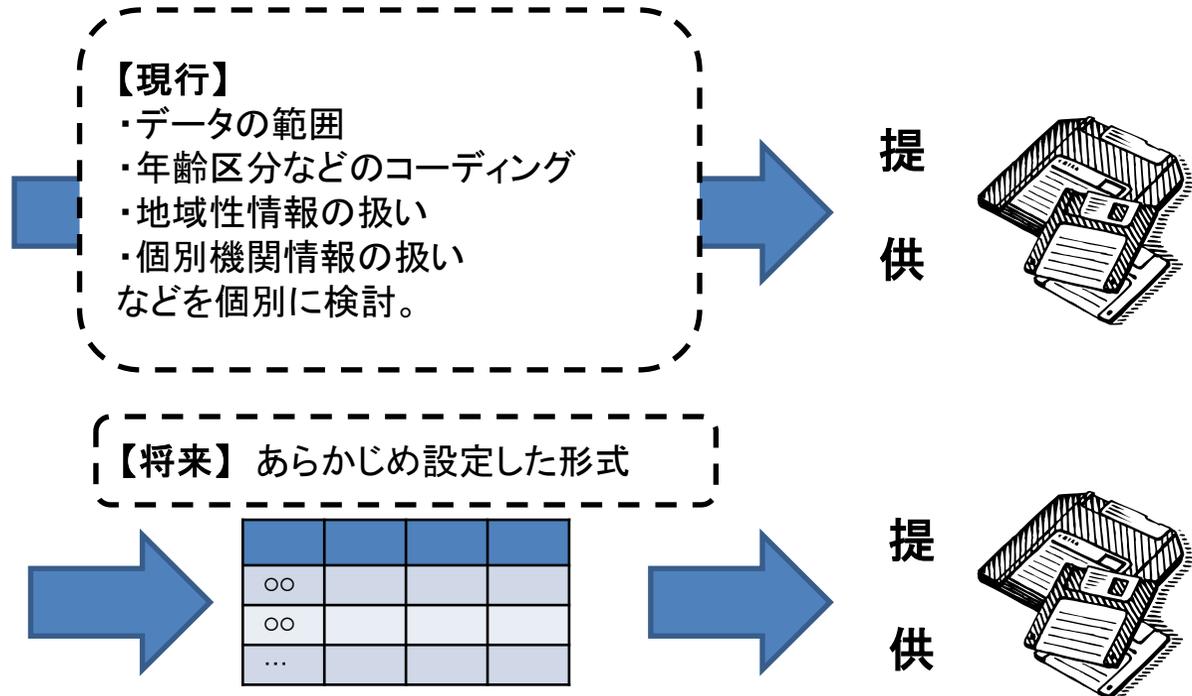
<基本データセットの考え方>

- 平成23年度から24年度までの試行期間においては、それぞれの利用申出毎に個別に提供するデータ形式を検討した上で、有識者会議での議論を踏まえレセプト情報等の提供を行うこととしている。
 その中で、提供するレセプト情報等については、①**個人の特定可能性について、十分に配慮がなされており、かつ、②学術研究の用にも供することができる形式である、**という要件を満たすことが求められる。
- しかしながら、上記のような**検討を個別に行うことは提供までの審査に要する期間や事務コストが膨大なもの**となることが考えられる。そこで、将来的には利用目的や公表形式、データの管理方法等のガイドライン上のルールが遵守されることを前提に、**あらかじめ上記の2点の要件を満たしたデータ形式を設定しておくことが、**利用者の利便性や審査の効率化に資すると考えられる。

<イメージ>

電子レセプト

IY, 21, 5, 612350038, 1, 1, 4
 IY, 21, 5, 611170373, 1, ,
 IY, , 5, 611410121, 1, ,
 IY, , 5, 611120119, 1, 4, 4
 IY, 21, 5, 610453117, 2, ,
 IY, , 5, 610453103, 8, ,
 IY, , 5, 610453099, 4, 10, 4



<1. 個人の特定可能性への配慮>

- レセプト情報等には、患者の氏名等の直接的に個人を識別できる情報はそもそも保管されていない。
- ガイドラインにおいては、有識者会議が特に認める場合を除き、原則として他の情報とレセプト情報等の照合を行ってはならないこととしているが、その上でさらに、個人の特定可能性へ配慮する観点から、以下の点について検討が必要ではないか。

①データの範囲

希少な傷病名や事例の少ない診療行為を含めないなど、データセットに含まれる情報の範囲を検討する必要。

②年齢区分のコーディング

年齢情報については、模擬申出・審査において、概ね5歳刻みでコーディングし、85歳以上トップコーディングを行っている。ただし、0歳児などの取扱いをどうするかなどの検討も必要。

③地域性情報の取扱い

レセプト情報等には、診療・健診を受けた医療機関等のコードが記録されている。これらのコードは原則提供しないこととなっているが、このコードの情報に基づいて、事務局で都道府県、市区町村などの一定の地域区分にデータを整理することは可能。

ただし、地域が限定されれば、それだけ特定個人の識別可能性も高まることが考えられることから、基本データセットにおいて、そもそも地域性の区分を設けるか、設けるとした場合に、その単位は都道府県や市町村といったどの程度の範囲までとするか、検討が必要。

④個別機関情報の取扱い

診療を受けた個別の医療機関や所属する保険者が特定される場合、特定個人の識別可能性も高まることから、どの程度まで個別の医療機関等の属性情報(所在地、病床数、病棟区分など)に配慮するか検討する必要。

<2. 学術研究への有用性の観点>

- 前ページの①から④までの項目の検討にあたっては、学術研究における有用性も考慮する必要があると考えられる。特定個人の識別可能性への配慮により、地域区分が大括りとなりすぎたり、個別の傷病名や診療行為の情報が削除されすぎると学術研究への有用性が低下することも考慮する必要がある。
- また、「社会医療診療行為別調査」や「医療費の動向(メディアス)」などの既存の政府統計との関係についてもそれぞれのデータの性質を考慮して整理することが、データ利用者の利便性に資すると考えられる。
すなわち、レセプト情報等の基本データセットについては、こうした既存の政府統計にはない、レセプト情報等が有する特質を生かしたものとすることが望ましいのではないかと。

【レセプト情報等の特質】

- ・ レセプト情報等のデータベースには基本的に電子化された全てのレセプトが保存。全数を基本としている。
- ・ 他の抽出調査と異なり、年間の一時点ではなく経時的に収集されている。
- ・ あくまでレセプトのデータであることから、レセプトの持つ性格(実質的な主傷病の特定が難しい、個別の診療行為の回数の集計が難しい、など)をそのまま有しており、必ずしも研究に適した体裁となっていない。

こうした論点について、有識者会議での議論に加え、必要に応じて、統計の専門家や疫学研究等の研究者の意見も参考にし、検討する必要。

その際、統計的な見地からは、統計法の制度における統計委員会匿名データ部会での議論を、学術研究の有用性の観点からは、学会等の意見を参考にしつつ検討してはどうか。

<概要>

台湾では、1995年に国民健康保険プログラムが導入され、2007年時点で総人口2,296万人中、2,260万人がプログラムに加入。このプログラムに基づく、診療報酬支払いのためのレセプトデータ等を、国民医療保険局(Bureau of National Health Insurance: BNHI)が収集しデータベースを構築しており、その管理を国家健康調査機構(National Health Research Institutes :NHRI)が行っている。

データの利用を希望する研究者は、「一般申請」と「特別申請」の2つの方法で申請ができる。

<全民健康保険研究資料庫 (National Health Insurance Research Database)に含まれるデータ>

レセプトファイル(原始資料檔)	備考
1. 住院費用申請總表主檔 (DT)	入院レセプトファイル(総表)
2. 門診費用申請總表主檔 (CT)	外来レセプトファイル(総表)
3. 住院醫療費用清單明細檔 (DD)	入院レセプトファイル
4. 住院醫療費用醫令清單明細檔 (DO)	入院明細(オーダーリング・処方)ファイル
5. 門診処方及治療明細檔 (CD)	外来レセプトファイル
6. 門診処方醫令明細檔 (OO)	外来明細(オーダーリング・処方)ファイル
7. 特約藥局処方及調劑明細檔 (GD)	調劑明細ファイル
8. 特約藥局処方醫令檔 (GO)	調劑レセプトファイル
9. 承保資料檔 (ID)	被保険者ファイル

基本ファイル(基本資料檔)	備考
1. 醫事機構病床主檔 (BED)	各医療機関の病床種別ファイル
2. 醫事機構診療科別明細檔 (DETA)	各医療機関の開設診療科ファイル
3. 醫事機構基本資料 (HOSB)	医療機関情報 I
4. 醫事機構副檔資料 (HOSX)	医療機関情報 II (開院日情報)
5. 専科醫師證書主 (DOC)	医師情報 I
6. 醫事人員基本資料檔 (PER)	医師情報 II
7. 重大傷病證明明細檔 (HV)	重大傷病証明明細ファイル
8. 醫事機構服務項目檔 (HOX)	医療機関情報 III (開院日情報)
9. 藥品主檔 (DRUG)	医薬品コード
10. 承保資料檔 (ID)	被保険者ファイル

<一般申請>

あらかじめ設定された5種類のデータセットからデータの提供を受けるもの。

(Ⅰ)基本資料データ(基本資料檔):集計データ

基本ファイル(基本資料檔)の10ファイルと、DT(入院レセプトデータ(総表))とCT(外来のレセプトデータ(総表))の2ファイル。

ただし、DTとCTは医療機関単位で集計(月次)されている。

(Ⅱ)系統抽出データ(系統抽様檔):個票データ

入院レセプトファイル(DD)の5%抽出

(入院明細(オーダーリング・処方)ファイル(DO)を含む)

外来レセプトファイル(CD)の0.2%抽出(外来明細

(オーダーリング・処方)ファイル(CO)を含む)

(Ⅲ)特定主題データ(特定主題分檔):個票データ

特定疾患や調査目的に沿って抽出したデータファイル。悪性新生物、糖尿病、精神疾患、交通事故、リハビリ、漢方薬(中醫薬)ファイル等、現在16種類。

(Ⅳ)ランダム抽出データ(抽様歸人檔):個票データ

被保険者ファイル(ID)から、ランダムサンプリングしたもの。

2000年は20万人を抽出し、(1996年～2007年の)入院、外来、調剤レセプトデータを提供している(※)。

(Ⅴ)教育用データ(教學用資料檔):個票データ

2000年のレセプトデータから1,000人をランダムサンプリングしたもの。教育用に無償提供。

※ランダム抽出データの2000年版ファイル構成

ID	ID2000
	入院レセプトファイル(DD)
	入院明細(オーダーリング・処方)ファイル(DO)
	外来レセプトファイル(CD)
	外来明細(オーダーリング・処方)ファイル(OO)
	調剤明細ファイル(GD)
	調剤レセプトファイル(GO)

2000年は20万人をランダムサンプリング。2005年は100万人

<特別申請>

申請人が提出した研究計画に必要なデータを、國家衛生研究院による審査終了後に、4種類のデータ(CD, OO, DD, DO)から抽出して、データセットを作成して提供する(含、調剤データ)。

(出典):

國家衛生研究院、National Health Insurance Research Database in Taiwanホームページ

http://w3.nhri.org.tw/nhird/date_01.htm

http://w3.nhri.org.tw/nhird/en/Data_Files.html

http://w3.nhri.org.tw/nhird//date_03_02.php?year=97&list_n=0

http://w3.nhri.org.tw/nhird//file_talk/workshop2008-1.pdf

厚生労働科学研究費補助金、医療ナショナルデータベースに関する諸外国の整備状況および日本におけるデータベースのあり方研究、(財)医療経済研究機構・社会保険福祉協会

(資料作成協力): (財)医療経済研究機構・社会保険福祉協会 満武巨裕 研究部副部長

米国の公的医療保険制度としては、主に高齢者の方を対象とするメディケア (Medicare) と主に低所得者や障害者の方を対象とするメディケイド (Medicaid) が存在。

メディケアの加入者: 約4520万人 メディケイドの加入者: 約4714万人

(※2008年6月末時点)

制度	メディケア	メディケイド
保険者	連邦政府	連邦政府と州の共同資金拠出
対象者・給付内容	<ul style="list-style-type: none"> ・65歳以上の高齢者対象 ・入院医療、高度看護施設、ホスピスケア ・病院外来、臨床検査、在宅ケア 	<ul style="list-style-type: none"> ・義務的困窮者(給付義務付)と選択的困窮者(各州の裁量で給付)等が対象。 ※ 子供のいる低収入家庭、補足的所得受給者、世帯収入が一定水準以下、など。 ・基礎的なサービス(各州が給付義務) 入院・外来、妊婦健診、小児ワクチン 等 ・選択的なサービス(各州が選択的に給付可) 診療所サービス、診断サービス、等。

メディケア・メディケイドのデータ提供は、米国政府のDepartment of Health and Human Services のエージェンシーであるCenter for Medicare and Medicaid Services (CMS)が行っている。

メディケア、メディケイドに関するデータは、①レセプトベースのSAFs(Standard Analytical Files:標準分析ファイル)と、②医療施設に入院したメディケア受給者について、在院日数ベースのMedpar(Medicare Provider Analysis and Review :医療供給者分析ファイル)の2種類が収集されている。

これらのファイルには、外来の場合の治療日、入退院日、誕生日、人種、性別、診断名、治療内容、治療内容に対する支出、治療を受けた医療施設ID、治療を受けた医師のUPIN(Unique Physician Identifier)等が含まれる。

データの名称	概要	提供についての基準
Identifiable Data Files	実際の受給者・医療サービスの提供者について個人を識別できる情報を含む。	<p>ResDAC(The Research Data Assistance Center)(注)による事前審査を経た上で、CMSにおいて、主に以下の内容について審査。</p> <ul style="list-style-type: none"> ○具体的な研究内容・手法の詳細。目的とする研究に申請するデータを使用する科学的根拠。 ○データの管理方法・責任者など。 ○Data Use Agreement の締結。 ○研究資金を受けていることの証明。
Limited Data Set	個票データではあるが、受益者・医療サービス提供者について個人を識別できるデータ含まない。	<ul style="list-style-type: none"> ○Identifiable Data よりは簡易な様式により、研究内容、手法等について申請。 ○Data Use Agreementの締結。
Public Use Files (Non-Identifiable Data Files)	個票データではなくあらかじめ定式化された集計データ	<ul style="list-style-type: none"> ○提供するデータの提供や送付先、支払方法等に係わる簡易なUser Agreementのみ締結。

(注)ResDAC は、手数料と引き替えに研究者等に対してメディケア・メディケイドのデータの利用方法等について技術的な補助を行うCMSと契約を締結した機関であり、ミネソタ大学の疫学、公衆衛生等の専門家によって構成されている。

○ 統計法における第三者提供の枠組みにおいては、まず政府において収集したデータを用いた基本資料を作成した上で、第三者利用に供するための基本データセットを作成している。

レセプト情報・特定健診等情報データベースについては、今般、社会医療診療行為別調査及びメディアスの利用することにより政府統計として、データを公表する。

類型	概要	他制度での例	レセプト情報・特定健診等情報データベース
①政府統計資料として公表するもの	政府において収集したデータに基づいて統計資料等を作成しHP等で公表	政府基幹統計など。	<ul style="list-style-type: none"> ・社会医療診療行為別調査 ・調剤メディアス ・特定健診等基本統計
②基本データセットを作成し提供するもの	個票ベースの基本的なデータセットをあらかじめ作成し、申出に基づいて提供。	統計法の調査票情報、匿名データの提供	<p>今後の検討</p>
③求めに応じた集計等を行うもの(オーダーメイド)	基本データセットから申出に基づき集計した情報を提供。	統計法のオーダーメイド集計	<p>有識者会議における議論により提供。</p>
	個票ベースのデータを個別の申出毎に匿名化処理等を行い個票又は集計表で提供。	—	