

# 1. 第三者提供：サンプリングデータセットについて

# 1. 基本方針について

- 第8回レセプト情報等の提供に関する有識者会議(平成24年2月10日)において検討された事項  
レセプト情報等の提供における申出に対し、個人情報保護の観点から、「対象者が極めて限定される可能性がある」申出や、「多数の項目を用いた探索的研究」、傷病名・診療行為・医薬品コードを「全て求める」などという申出については不承諾とする審査方針を定めたことを踏まえ、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すること。
  
- 第9回レセプト情報等の提供に関する有識者会議(平成24年3月7日)において検討された事項
  - 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サンプリングデータセットを平成23年度中に整備すること。
  - サンプリングデータセットについては、具体的に以下のとおりとすること。
    - 「医科」「DPC」「調剤」レセプトについて、データセットを整備すること
    - 当面は単月のデータで行うこと
    - 抽出の割合については、「医科入院」「DPC」を10%、「医科入院外」「調剤」を1%とし、統計的な偏りを少なくするため、性別、年齢構成(5歳刻み年齢区分)については、全数データの構成割合と同じものとする
    - 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
    - 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1回審査で不承諾となった申出者に限定すること
    - 通常の「レセプト情報等の提供」と同様の申出内容(ただし通常の「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)、セキュリティ要件を求めること
    - 申出者或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

## 2. データセットの仕様(最終案)について

○ 今回提供するサンプリングデータセットの最終的な仕様の概略については以下のとおり。

### (対象レセプト)

- ・「医科入院レセプト(以下「医科入院」)、DPCレセプト(以下「DPC」)、医科入院外レセプト(以下「医科入院外」)、「調剤」レセプト(以下「調剤」)とする。

### (対象期間)

- ・平成23年10月診療分(単月分)とする。

### (抽出率等)

- ・「医科入院」「DPC」は10%、「医科入院外」「調剤」は1%を抽出する。
- ・「医科入院外」については、平成23年10月診療分および11月診療分のハッシュ値で紐付けた調剤レセプトをあわせて提供する。
- ・「医科入院外」以外は、ハッシュ値を付与せず提供する。「医科入院外」は、紐付けられる調剤レセプトともども、ハッシュ値1を付与する。
- ・「性別」「5歳刻み年齢区分」において、母集団と構成比率を変えないようにして提供する。

### (匿名化処理)

- ・診療報酬請求における基本マスターのうち、「傷病名コード」「医科診療行為コード」「医薬品コード」においては出現頻度の低いコード、「診断群分類」においては出現頻度の低い診断群分類について、出現するコードの総数のうち少ない者から順に0.1%に達するまで、匿名化処理を行う。

### (高額レセプトの扱い等)

- ・「医科入院」においては、合計点数が700,000点以上、「医科入院外」「調剤」については、合計点数が50,000点以上のレセプトを最初に削除したうえで、抽出を行う。
- ・「DPC」については、「医療機関別係数」の数値で医療機関が限定される可能性があるため、「医療機関別係数」「包括評価点数」「今月包括合計点数」は空欄とする。
- ・同様に、「包括小計点数合算」および出来高部分の合計などで合計点数を一定程度推定できるため、「DPC」では合計点数も空欄とする。

### (その他の空欄化処理等)

- ・都道府県情報、医療機関情報、薬局情報、保険者情報、公費情報に関連する項目は空欄化する。
- ・臓器提供者に関する情報は削除する。
- ・「医科入院」「医科入院外」において、病床数が記載されている場合はカテゴリー化した情報に加工して提供する。
- ・「DPC」では、「入院時年齢」「出生児体重」「Burn Index」については、カテゴリー化した情報に加工して提供する。

### (各種マスターにおけるコードの調整)

- ・マスターにないコード番号が含まれている可能性があるが、これらについての調整は行わないこととする。

○ 承諾された申出者からは、サンプリングデータセットの使用に際して、気付いた点についての意見を提供いただき、今後、サンプリングデータセットの質の改善に向けた情報として活用したい。

# 3. 審査方針(案)について

## (研究内容・抽出について)

- サンプルングデータセットについては、個人の特定可能性を下げる処理がほどこされたデータセットであることから、「研究内容・抽出」については必ずしも「レセプト情報等の提供」と同じ水準の審査方針を適用するのではなく、抽出条件の詳細な指定は求めず、研究の概略が把握できるような申出であれば承諾とすることとしてはどうか。

## (セキュリティ要件について)

- 「セキュリティ要件」については、有識者会議における検討においても、従来の「レセプト情報等」の提供と同様のセキュリティを確保するよう求めていることから、従来の「レセプト情報等の提供」と同じ水準の審査方針を踏襲することとしてはどうか。
- 通常のレセプト情報等データベースを用いた研究と比較して、サンプルングデータセットを用いて行った研究成果が、個人情報保護の観点から問題がないかを確認するため、研究成果を公表する前には通常のレセプト情報等データベースを用いた研究と同様に、事務局に公表物を提出し、事務局での確認のうえ、必要に応じて有識者会議での承認を得る手続きをとることとしてはどうか。

## (その他)

- 個別の審査は、非公開の場で審査を行うこととしてはどうか。
- 「1人の申出者・利用者に対して同時に提供することが出来るレセプト情報等は1件まで」としていることから、申出者があらかじめ希望している優先順位を参考に、内容も確認したうえで一つの申出に対してのみデータ提供を行うこととしてはどうか。
- 今後、通常のレセプト情報等データベースについてもあわせて申出する場合、仮に両方の申出が承諾された場合、まずサンプルングデータセットを用いた研究を行ってもらい、その研究を終了してからレセプト情報等の提供の資格を得られるよう調整等行うこととしてはどうか。

# 4. 進捗状況及び今後の予定

## 有識者会議

## 申出者への対応等

平成23年11月

- データセット作成の提案  
(11月10日)

平成23年12月

平成24年1月

平成24年2月

- データセット仕様の検討  
(2月10日、3月7日)

平成24年3月

平成24年4月

平成24年5月

- サンプリングデータセット作成のための各種予備調査
  - 各種レセプトにおけるデータ量の評価
  - 希少なコードの同定
  - 母集団と構成比率を変えずに抽出する手法の確認
  - 等々

平成24年6月

- 申出審査  
(6月4日)

平成24年7月

- (予定)サンプリングデータセット作成

平成24年8月

- (予定)サンプリングデータセット提供

- 事前説明会の開催  
(3月21日)
- 事前相談の受付  
(4月16日～5月11日)
- 申出受付  
(5月14日～5月18日)

- (予定)誓約書等提出

## 5. 今後の方針について

### (1: 今回の申出審査にあたって)

今回のサンプリングデータセットについて、申出者は6名、申出研究は7件であり、先に確認した審査方針のもと、非公開の場で審査を行うこととしてはどうか。

### (2: 今後のデータセットのあり方)

#### ① サンプリングデータセットの改善

今回承諾となった申出者からの指摘・意見の集約を行うこととしてはどうか。

#### ② その他データセットの整備

第8回、第9回有識者会議における検討を踏まえ、平成24年度の厚生労働科学研究を活用し、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について、学術的な観点からの精度を確保しながら検討を進めることとした。

# その他データセット作成：平成24年度厚生労働科学研究について

○ 第8回有識者会議において「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」について検討がなされ、「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」をすすめるべきとの議論を踏まえ、平成23年11月24日に発出された厚生労働科学研究費補助金政策科学研究事業において研究課題「汎用性の高いレセプト基本データセット作成に関する研究」として公募を行った。

○ 今年度は、政策科学研究事業「汎用性の高いレセプト基本データセット作成に関する研究」(副部長：満武巨裕氏(医療経済研究機構))が採択となった。

○ この研究班では、以下の目的に沿ってデータセットの設計と作成を行うこととしている。

本研究の目的は、レセプト情報等データベース(以下、NDB)の利用促進のために、匿名化技術等のセキュリティ対策を施した、研究用途における汎用性の高いレセプト基本データセットの設計と作成、...を整備することである。データセットの抽出率、アウトライヤー(裾切)処理、母集団の性質を有しているか等の統計的検証も行う...。詳細な診療情報を生かした、研究利用しやすい

- (1)一定期間のレセプトを個人ID毎に統合する等の処理、且つ診療行為も含むデータセット
- (2)特定健診・保健指導データとレセプトとリンクさせたデータセット

を設計・施行する。

- (3)上記以外のデータセットも検討・作成する。

○ 研究にあたっては実際のデータセットの形態を確認したうえですすめる必要があることから、レセプト情報・特定健診等情報データベースに格納されているデータを分析する必要がある。

○ 本研究班には、提供にあたっては「本来目的以外での利用」の枠組みにおいて申出者に求めているものと同等の申出書類、利用環境を準備することを研究班に義務づけ、セキュリティ要件の確保に細心の注意を払うことを前提に、レセプト情報等データベースのデータを提供することとしてはどうか。