

平成30年3月14日	資料4
第1回要介護認定情報・介護レセプト等情報の 提供に関する有識者会議	

サンプリングデータセットについて

1. サンプルングデータセットの整備について

(現状)

- レセプト情報・特定健診等情報データベース（以降、NDB）においては、「探索的研究へのニーズに対応し、安全性（個人の識別可能性）に十分配慮したデータセット」として位置づけられている。（サンプルングデータセットの提供に関する事前説明会資料（2012年3月21日））
- 単月のレセプト情報からの抽出データであるという制約はあるが、特別抽出と比較して簡易な審査によって取得できることから、研究者が探索的研究を試行的に実施することが容易に可能であるという長所を持っている。

(対応案)

- NDBにおいて研究者が探索的研究を試行的に実施できるデータセットとして有効に活用されている実態を考慮し、介護保険総合データベース（介護DB）においてもNDBと同様の位置づけにおいてサンプルングデータセットを整備する。

2. サンプルングデータセットの仕様として定める事項について

(現状)

- サンプルングデータセットを整備するために検討が必要な仕様として、NDBのサンプルングデータセットでは、以下に示す項目が定められている。

レセプトの種類	対象時期	レセプトの抽出方法
匿名化処理方法	その他の処理	

(対応案)

- 介護DBのサンプルングデータセットを整備するため、NDBと同様に、次のような技術的な観点からの仕様について定めることとしてはどうか。

項目	仕様で定める事項
対象とするデータの種類 及び対象時期	<ul style="list-style-type: none">❑ 介護DBで保有する介護レセプト等情報のうち、対象とする範囲❑ 対象とする1ヶ月分の定義（サービス提供月・審査月）
対象データの抽出方法	<ul style="list-style-type: none">❑ 1か月分の全レセプトから除外する高額レセプト等の定義❑ 抽出率及び層化基準
匿名化処理方法	<ul style="list-style-type: none">❑ 秘匿化処理の対象を判別するための出現頻度の基準の定義❑ コードの置き換え・空欄化を行うためのルール
その他の処理	<ul style="list-style-type: none">❑ 空欄化、カテゴリー化、削除等を行う項目の定義

2. a. 対象とするデータの種類及び対象時期

(現状)

- NDBのサンプリングデータセットで対象としているデータは以下の通りである。

項目	(参考) NDBにおける仕様
レセプトの種類	□ 医科入院レセプト、DPCレセプト、医科入院外レセプト、調剤レセプトの4種類

(対応案)

- 介護DBが保有する介護レセプト等情報について、以下の通り対応することとしてはどうか。

データの種類	データの概要	対応案
明細書情報	□ 利用者単位の請求明細書	<ul style="list-style-type: none">□ 明細情報レコード及び居宅サービス計画費情報レコードを対象とする。□ 請求時情報と決定後情報の双方が含まれるが、決定後情報を対象とする。□ 2号被保険者については1グループとして集計する。
給付管理票情報	□ 給付管理票	<ul style="list-style-type: none">□ 収載されている情報は明細書にも含まれる情報であるため対象外とする。
台帳情報	<ul style="list-style-type: none">□ 事業所台帳□ 保険者台帳□ 受給者台帳	<ul style="list-style-type: none">□ 台帳情報はNDBと同様に対象外とする。

2. a. 対象とするデータの種類及び対象時期

(現状)

- NDBでは、レセプトの対象時期について、以下のように定められている。

項目	(参考) NDBにおける仕様
対象時期	<ul style="list-style-type: none">□ 平成 23 年 10 月診療分で、かつ平成 23 年 11 月、平成 23 年 12 月、平成 24 年 1 月に審査されたレセプト□ 以降、各年の 1 月、4 月、7 月、10 月サービス提供分

(対応案)

- 介護DBにおいても同様の時期で提供することとしてはどうか。

項目	対応案
対象時期	<ul style="list-style-type: none">□ 平成 24 年 4 月サービス提供分で、かつ平成 24 年 5 月、平成 24 年 6 月、平成 24 年 7 月に審査された介護レセプト□ 以降、各年の 1 月、4 月、7 月、10 月サービス提供分について作成する

2. b. 対象データの抽出方法の検討

(現状)

- NDBのサンプリングデータセットの抽出方法は、以下のように定められている。

項目	NDBにおける仕様
母集団の定義 (高額レセプトの基準)	<input type="checkbox"/> 「医科入院」では 700,000 点以上、「医科入院外」「調剤」では 50,000 点以上のレセプトを除く集団
層化基準	<input type="checkbox"/> 性別、および5歳刻みの年齢層 <input type="checkbox"/> 層別の構成比率は母集団と同様になるように抽出
抽出率	<input type="checkbox"/> 「医科入院」「DPC」は 10% <input type="checkbox"/> 「医科入院外」「調剤」は1%
抽出単位	<input type="checkbox"/> レセプト単位

2. b. 対象データの抽出方法の検討

- 抽出率に関する基礎的分析（請求単位数の分布）
 - 平成26年10月サービス提供分の介護レセプト等情報を用いて、区分支給限度基準額が適用されるサービスについて、利用者一人1月あたりの請求単位数の分布を集計すると以下の通りであった。

項目	要支援 1	要支援 2	要介護 1	要介護 2	要介護 3	要介護 4	要介護 5
区分支給限度基準額 (単位)	5,003	10,473	16,692	19,616	26,931	30,806	36,065
平均 (単位)	2,384	4,392	7,893	10,975	16,593	20,383	25,209
標準偏差(単位)	1,100	2,290	4,792	6,277	8,663	10,072	11,452
中央値 (単位)	2,409	4,546	7,314	10,815	17,197	22,456	28,653

- 例えば、平均 + 2 標準偏差 (SD) 以上の者を除くとした場合、要介護度別の該当者の割合は以下の通りであった。

項目	要支援 1	要支援 2	要介護 1	要介護 2	要介護 3	要介護 4	要介護 5
平均+2SD (単位)	4,585	8,972	17,478	23,528	33,919	40,526	48,112
概要者割合(%)	5.7	3.1	1.5	1.0	0.3	0.3	0.3

2. b. 対象データの抽出方法の検討

- 抽出率に関する基礎的分析（標準誤差率）
 - 平成26年10月サービス提供分の介護レセプト等情報を用いて算出した抽出率を1%、3%、5%、10%とした場合の標準誤差率の推定値は以下の通りであった。
 - 例えば、抽出率を3%とした場合、層別（性・年齢階級別）の標準誤差率は全ての層で2%以下となった。

標本サイズ（人） [抽出率]	利用者1人1月あたりの平均単位数の標準誤差率	層別標準誤差率の 最大値
48,380人[1%]	約0.32%	3.4%
145,140人[3%]	約0.18%	1.9%
241,901人[5%]	約0.14%	1.5%
483,801人[10%]	約0.10%	1.0%

(*1)平成26年10月サービス提供分の利用者4,838,010人を母集団として算出

(*2)請求明細書の利用者1人1月あたりの平均レコード数は約7.95レコード

2. b. 対象データの抽出方法の検討

(参考) 利用者1人1月あたりの平均単位数の層別 (性・年齢階級別) 標準誤差率

[抽出率1%]

年齢階級	性	標準誤差率
64歳以下	男	2.7%
	女	3.4%
65-69歳	男	2.3%
	女	2.8%
70-74歳	男	1.8%
	女	1.9%
75-79歳	男	1.5%
	女	1.3%
80-84歳	男	1.3%
	女	0.9%
85歳以上	男	1.0%
	女	0.5%

[抽出率3%]

年齢階級	性	標準誤差率
64歳以下	男	1.6%
	女	1.9%
65-69歳	男	1.3%
	女	1.6%
70-74歳	男	1.0%
	女	1.1%
75-79歳	男	0.9%
	女	0.7%
80-84歳	男	0.7%
	女	0.5%
85歳以上	男	0.6%
	女	0.3%

2. b. 対象データの抽出方法の検討

(参考) 利用者1人1月あたりの平均単位数の層別 (性・年齢階級別) 標準誤差率

[抽出率5%]

年齢階級	性	標準誤差率
64歳以下	男	1.2%
	女	1.5%
65-69歳	男	1.0%
	女	1.2%
70-74歳	男	0.8%
	女	0.8%
75-79歳	男	0.7%
	女	0.6%
80-84歳	男	0.6%
	女	0.4%
85歳以上	男	0.4%
	女	0.2%

[抽出率10%]

年齢階級	性	標準誤差率
64歳以下	男	0.8%
	女	1.0%
65-69歳	男	0.7%
	女	0.8%
70-74歳	男	0.5%
	女	0.6%
75-79歳	男	0.5%
	女	0.4%
80-84歳	男	0.4%
	女	0.3%
85歳以上	男	0.3%
	女	0.1%

2. b. 対象データの抽出方法の検討

(参考) 利用者1人1月あたりの平均単位数の標準誤差率の算定式

$$\sqrt{V} = \frac{1}{N} \sqrt{\sum_{h=1}^L N_h (N_h - n_h) \frac{\text{Var}(X_h)}{n_h}}$$

$$e = \frac{\sqrt{V}}{\bar{X}}$$

e : 利用者1人1月あたり平均単位数の推定値の標準誤差率

\sqrt{V} : 利用者1人1月あたり平均単位数の推定値の標準誤差

\bar{X} : 利用者1人1月あたり平均単位数の推定値

L : 層の数

N_h : 第 h 層における母集団利用者数

N : 母集団利用者数

n_h : 第 h 層における標本利用者数

$\text{Var}(X_h)$: 第 h 層における利用者1人1月あたり単位数の分散

2. b. 対象データの抽出方法の検討

(対応案)

- 以下の通りサンプリングデータセット整備に用いる対象データを抽出することとしてはどうか。

項目	論点	対応案
高額レセプトの基準	<ul style="list-style-type: none">「高額レセプト」の定義は請求実績の分布を分析した上で検討してはどうか。	<ul style="list-style-type: none">区分支給限度基準額対象サービスについて、平均 + 2SD を超える利用者を抽出から除外する。
層化基準	<ul style="list-style-type: none">無作為抽出であれば確率的には母集団の構成が維持されるが、特定の属性の集団が確実に含まれるように層化抽出を行うべきではないか。	<ul style="list-style-type: none">性・年齢階級別（5歳刻み）に層化抽出を行う。
抽出率	<ul style="list-style-type: none">抽出率は、「サンプリングデータがどの程度母集団を代表しているか」「研究者が容易に扱えるデータ量であるか」の両面から検討してはどうか。	<ul style="list-style-type: none">全ての層で標準誤差率が2%以下となる、3%の抽出率で抽出する。
抽出単位	<ul style="list-style-type: none">「レセプト単位」と「利用者単位」のどちらで抽出すべきか。	<ul style="list-style-type: none">利用者単位で抽出する。

2. c. 匿名化処理の検討

(現状)

- NDBでは以下の匿名化処理が実施されている。

項目	(参考) NDBにおける仕様
匿名化処理	<ul style="list-style-type: none"> 出現頻度の低い「医科診療行為コード」等について匿名化処理を実施。 「傷病名コード」「医科診療行為コード」「医薬品コード」「診断群分類」においては0.1%、ただし、「医科入院」「医科入院外」における「医科診療行為コード」においては0.01%に達するまでのコードを対象。 該当するコードについては「9999999」等の特定のコードに置き換える、点数を空欄化する 等を実施。

- 介護レセプトのサービスコードについて

– サービスコードは2桁のサービス種類コードと4桁のサービス項目コードで構成される

例1) 1 1 4 4 8 5 (165単位)

例2) 1 1 4 4 8 6 (206単位)

訪問介護 身体介護が中心 20分未満

訪問介護 身体介護が中心 20分未満
夜間早朝の場合 25%加算

- サービスコード出現頻度について

– 平成26年10月サービス提供分の介護レセプト等情報を用いて、サービスコードの出現頻度を確認すると、以下の通りであった。

出現件数	10件未満	10件以上100件未満	100件以上1,000件未満	1,000件以上10,000件未満
サービスコード割合	32%	27%	19%	12%

2. c. 匿名化処理の検討

(対応案)

- 以下の通り、匿名化処理を行う。

項目	論点	対応案
匿名化処理の対象	<ul style="list-style-type: none">□ NDBを参考に匿名化処理の対象を決定してはどうか。	<ul style="list-style-type: none">□ 10件未満のサービスコードを匿名化する。
匿名化処理の方法	<ul style="list-style-type: none">□ NDBでの取組を参考に、匿名化処理を行ってはどうか。	<ul style="list-style-type: none">□ 「サービス項目コード」を「9999」等の特定のコードに置き換える。□ 単位数は抽出された利用者が利用している合計単位数が把握できるように合計して表示する。□ 日数・回数については合計ができないため、空欄化処理を実施する。

2. d. その他の処理の検討

(現状)

- NDBでは以下の処理が実施されている。

項目	(参考) NDBにおける仕様
その他の処理	□ 「医療機関コード」「レセプト番号」等の空欄化、「病床数」等のカテゴリー化、「臓器提供医療機関情報レコード」等の削除

(対応案)

- 介護DBにおいても、NDBの取組を参考に、以下の処理を行ってはどうか。

項目	対応案
その他の処理	□ 「事業所番号」「居宅サービス計画 事業所番号（居宅介護支援事業所）」等を空欄化する。