

匿名医療保険等関連情報データベース（NDB）の第三者提供

よくあるご質問（FAQ）

2023年10月作成

④『サンプリングデータセット』に関する質問

1	Q	サンプリングデータセットとはどのようなものか。
	A	NDB から予め一定の割合（※）で抽出したデータを、さらに安全性に配慮した工夫を施した上で提供するものです。特定の患者個人又は医療機関・薬局等の識別性の問題に配慮した上で、1ヶ月分のNDB から一部を抽出しています。 ※詳細な仕様は以下マニュアルの P24-25 をご参照ください。 https://www.mhlw.go.jp/content/12400000/001076618.pdf
2	Q	どのようなデータを提供可能か。
	A	・「医科入院レセプト」、「医科入院外レセプト」、「DPC レセプト」、「調剤レセプト」における、2011年以降の各年1月、4月、7月、10月分が提供可能です。 ・調剤レセプトに限り、①調剤単体分(PHA)／②医科入院外に紐付く調剤分(PHA 外)の2種を提供しております。（※②は、医科入院外を抽出後、ハッシュ値1で紐付けた調剤レセプトを抽出）
3	Q	1月、4月、7月、10月それぞれの診療月、審査月はいつか。
	A	1月：当該年1月診療分、かつ2月、3月、4月に審査されたレセプトが対象です。 4月：当該年4月診療分、かつ5月、6月、7月に審査されたレセプトが対象です。 7月：当該年7月診療分、かつ8月、9月、10月に審査されたレセプトが対象です。 10月：当該年10月診療分、かつ11月、12月、1月に審査されたレセプトが対象です。
4	Q	サンプリングデータセットを利用して、患者単位に診療行為等分析することは可能か。
	A	「医科入院外」のデータセットは患者 ID1n が記録されているため、患者単位の分析が可能です。また、医科入院外に紐付く調剤分(PHA 外)についても可能です。 「医科入院」、「DPC」、「調剤」については、患者 ID1n が記録されていないため、患者単位の分析はできません。
5	Q	サンプリングデータセットでは、1レセプトに含まれる傷病名数に上限はあるのか。
	A	傷病名の数を削除するような処理は行っておりません。希少な傷病名の匿名化については、レコードの削除ではなく、コードの匿名化によって対応しております。従って、サンプリングデータセットに含まれる患者が「何種類の併存病を有するか」については、実データに準ずる情報が把握可能です。

6	Q	サンプリングデータセットでは高額レセプトが抽出の対象外となると思うが、同じ人物が複数の病院にかかり、A 病院から高額レセプトが提出され、B 病院からの高額ではないレセプトが提出された場合は、B 病院からのレセプトは抽出対象となるか。
	A	高額レセプトではないものは抽出の対象となります。サンプリングデータセットは、抽出対象となったデータを個人単位ではなくレセプト単位で絞り込みますので、この例ですと B 病院からのレセプトは抽出対象になりえます。

7	Q	サンプリングデータセットの医科入院レセプトと医科入院外レセプトに記載されている診療行為を患者単位の紐付けて分析することは可能か。
	A	サンプリングデータセットの医科入院レセプトには患者 ID に該当する情報がありませんので、患者単位の紐付けを行うことは不可能です。(サンプリングデータセットの医科入院外レセプトと調剤レセプトは患者単位で紐付けを行うことが可能です。) 異なる種類のレセプトを紐付けて分析したい場合には特別抽出をご検討ください。

8	Q	出現回数の少ない傷病名コード等に対する匿名化処理とはどのようなものか。
	A	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番号について匿名化処理を行います。 「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、抽出したレセプトに出現するコードの総数のうち、少ないものから順に 0.1% に達するまでのコードを、特定のコードにおきかえます。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」については、抽出したレセプトに出現するコードの総数のうち、少ないものから順に 0.01% に達するまで、特定のコードにおきかえます。詳細な仕様は以下マニュアルの P24-25 をご参照ください。 https://www.mhlw.go.jp/content/12400000/001076618.pdf