

平成27年12月9日	資料2
第26回レセプト情報等の 提供に関する有識者会議	

サンプリングデータセットの内容充実について

平成27年12月9日

厚生労働省保険局医療介護連携政策課
保険システム高度化推進室

今回の検討の背景：規制改革実施計画（平成27年6月30日閣議決定）

No.	事項名	規制改革の内容	実施時期
13	レセプト情報・特定健診等情報データベースの研究利用の法的位置付けの検討	「行政機関の保有する個人情報の保護に関する法律」の見直しの検討状況を踏まえ、NDBデータの公益目的での研究利用の法律上の位置付けや制度的枠組みについて検討し、結論を得る。	「行政機関の保有する個人情報の保護に関する法律」の見直しに合わせて検討・結論
14	レセプト情報・特定健診等情報データベースにおける民間活用の拡大	民間企業でも公益性の高い研究は可能であることから、民間企業に所属する研究者であってもNDBデータの公益目的での利用が可能となるよう、民間企業からの提案に基づき、厚生労働省においてNDBデータを基にした集計表を作成する枠組みを構築する。	平成27年度検討・結論、平成28年度措置
15	レセプト情報・特定健診等情報データベースにおける探索的研究の充実	これまで「サンプリングデータセット」を提供した関係者等の意見も踏まえつつ、探索的研究が可能な「サンプリングデータセット」の内容の充実を図る。	平成27年度措置
16	レセプト情報・特定健診等情報データベースにおける探索的研究の充実	平成27年度に開設されたオンサイトリサーチセンターについて、システム運用に資する検証を行いつつ、利用者の範囲や利用方法などの運用ルールの確立を図る。その上で、精度の高い研究の実施に資する	（オンサイトリサーチセンターの運用ルールの確立） 平成27年度措置 （オンサイトリサーチセンターの特性をいかに
<div style="border: 2px solid red; padding: 10px; display: inline-block;"> <p>サンプリングデータセットの内容の充実を図る必要がある。</p> </div>			
17	レセプト情報・特定健診等情報データベースにおける最少集計単位の検討	研究成果の公表に当たり、集計単位が市区町村の場合に患者数等が100未満になる集計単位が含まれていないことを条件とすることの妥当性について、提供依頼申出者の意見を聴いた上で検討し、結論を得る。	平成27年度検討・結論
18	レセプト情報・特定健診等情報データベースにおける地方公共団体の利用手続簡素化	提供依頼申出者が地方公共団体である場合のNDBデータの提供の枠組みの在り方について、その利用目的等に応じた再整理を行うとともに、特に迅速にデータ提供を行う必要がある場合には、「レセプト情報等の提供に関する有識者会議」の意見聴取を省略することを検討し、結論を得る。	措置済み
19	レセプト情報・特定健診等情報データベースによる分析の効率化	NDBを活用したレセプトデータ分析がより容易になるよう、電子レセプト上で省略されている各診療行為等の点数や回数、診療識別を補完する等、NDBのシステム改修を行う。	措置済み
20	レセプト情報・特定健診等情報データベースによる分析の効率化	研究者等が自らの研究にNDBデータを活用することが可能か事前に判断できるようにするため、項目ごとの出現率などのデータ精度に関する情報等、NDBデータの分析に役立つ情報について精査し、公表する。	平成27年度検討・結論、平成28年度措置

提供されているデータの種類

	特別抽出	サンプリング データセット	基本データセット	集計表情報
基本的な イメージ	申出者の要望に応じ、データベースにある全データのなかから、該当する個票の情報を抽出し、提供する	探索的研究へのニーズに対応し、抽出、匿名化などを施して安全性に十分配慮した、単月分のデータセット	入院、外来、疾患別など目的に合わせて年度ごとに紐付けが可能で、簡易に分析することが可能なデータセット	申出者の要望に応じ、データを加工して作成した集計表を提供する
提供データ	個票	一部匿名化等を行った個票	大幅に加工した個票	集計表
含まれている データ項目例	レセプト情報、特定健診等情報に含まれている、ほぼすべての項目	希少な情報があらかじめ匿名化・削除されたレセプトデータ	患者の基本属性情報以外は、主傷病名、診療識別情報、要望に応じたコードなど	集計表
利用にあたり 具備すべき セキュリティ	データ利用時に、情報セキュリティマネジメントシステムを確実に運用できる利用環境を整える	特別抽出で求められるセキュリティ水準と比較してある程度具備しやすいセキュリティ水準での利用が可能		
想定される 利用者像	レセプト研究に一定の知見があり、申出内容や抽出条件を吟味し、大量のデータを高速に処理することを想定している利用者	レセプト研究に関心はあるが経験がまだ十分でなく、データの特徴や各項目の概要を把握したいと考えている利用者	レセプトの構造を踏まえながら研究するよりも、基本的項目について簡単に分析を試みたいと考えている利用者	集計された結果を必要とし、データ処理を行うことを想定していない利用者
提供実績 (計66件)	42件	17件	—	7件

サンプリングデータセットの内容充実によるデータ提供速度の向上


○平成23年11月～平成27年11月の間

サンプリングデータセットが提供された案件

(本来利用・行政機関・自己都合による受け取り遅延は除いた10件が該当)

⇒平均待機期間3.6ヶ月

(データセットは準備された状態でNDBに格納済なので提供は迅速となる)



迅速に提供できるサンプリングデータセットの内容を充実させ、研究者の利用を促すことで、NDBデータの利活用の推進及び研究現場での浸透が期待できるのではないか

サンプリングデータセットの使用経験

「亜急性期入院医療管理料を算定した患者の診療内容に関する研究」報告書より

【提供依頼申請者】吉田愛(公益社団法人全日本病院協会)

<報告書 抜粋>

「年齢が含まれないために、各管理料や入院基本料の平均年齢を算出できず、本研究では75歳以上の患者が占める割合で代替することによって、患者属性を推定せざるをえなかった。」

「また算定されたすべての傷病名コードは含まれているものの、主傷病名とそれ以外の病名を判別できないので、どのような疾患が契機となり入院に至ったかは把握できなかった。」

○サンプリングデータセットに年齢、主傷病名に関する情報が含まれていないため、分析する上での制約となっている。

○これらの点について改善を行うことで、利活用の推進に資する事ができるのではないか。

【改善点1】年齢階級コードを追加し選択可能とする

年齢階級コード1	
コード	年齢
100	0-4歳
101	5-9歳
102	10-14歳
103	15-19歳
104	20-24歳
105	25-29歳
106	30-34歳
107	35-39歳
108	40-44歳
109	45-49歳
110	50-54歳
111	55-59歳
112	60-64歳
113	65-69歳
114	70-74歳
115	75-79歳
116	80歳-

年齢階級コード2			
コード	年齢	コード	年齢
200	0歳	217	45-49歳
201	1歳	218	50-54歳
202	2歳	219	55-59歳
203	3歳	220	60-64歳
204	4歳	221	65-69歳
205	5歳	222	70-74歳
206	6歳	223	75-79歳
207	7歳	224	80-84歳
208	8歳	225	85-89歳
209	9歳	226	90-94歳
210	10-14歳	227	95-99歳
211	15-19歳	228	100歳-
212	20-24歳		
213	25-29歳		
214	30-34歳		
215	35-39歳		
216	40-44歳		

新たに「年齢階級コード2」が選択可能

新システムでは、年齢階級コード1に加え、年齢階級コード2も選択可能である。0歳～9歳は1歳刻み、10歳以上は5歳刻み、最大は、100歳以上となり詳細な分析が可能となる。

【改善点2】疑い病名・主傷病決定フラグを導入する

＜旧システム＞

疑い病名の判定

SYレコードの項目「修飾語コード」に疑い病名コード”8002”という文字が含まれていた場合、該当レコードが疑い病名となる。

主傷病の判定

一連の医療行為の中で主傷病区分=01となるレコードが複数存在する場合があるため、どれが主傷病か判断できない。



＜新システム＞

項目「疑い病名フラグ」を導入し、疑い病名の場合に1を設定する。

項目「主傷病決定フラグ」を導入し、主傷病の場合に1を設定する。

＜研究者の利便性の向上＞

- ・研究者がデータの内部を見て判断する必要があった疑い病名と主傷病を判断するフラグを導入する。

【改善点3】サンプリングデータセット【冬】【春】【夏】を導入する

＜旧システム＞

平成23・24・25年10月診療分
合計3データセットのみ準備・提供



＜新システム＞

平成23・24・25・26年1月・4月・7月・10月診療分
合計16データセットを準備・提供する

＜研究者の利便性の向上＞

・1月（冬）、4月（春）、7月（夏）に特徴的な診療行為・受療行動や季節変動が分析可能となる。

（研究例）季節性の感染症、季節性の気分障害、アレルギー性疾患
皮膚科・小児科領域での受療行動の分析 等

サンプリングデータセットの匿名性について

レセプト原票、レセプト情報等データベースに格納されたレセプト情報等データ、サンプリングデータセットのそれぞれに含まれる主なデータ項目、匿名化処理されている主な項目は以下の通りである。

(詳細は、「オンライン又は光ディスク等による請求に係る記録条件仕様(医科用) (http://www.iryohoken.go.jp/shinryohoshu/file/spec/24bt1_1_kiroku.pdf)等を参照)

	項目	匿名化や削除などの処理が行われている主な項目等	
	レセプト原票	レセプト情報等データ(特別抽出)	サンプリングデータセット
医療機関情報	<ul style="list-style-type: none"> 都道府県 医療機関コード 医療機関名、電話番号等 	<ul style="list-style-type: none"> 医療機関名、電話番号は削除 	<p>【左記に加え】</p> <ul style="list-style-type: none"> 都道府県情報は削除 医療機関コードは削除
レセプトの各種属性情報	<ul style="list-style-type: none"> レセプト番号やカルテ番号 生年月日、氏名および性別 診療年月や入院年月日等 	<ul style="list-style-type: none"> カルテ番号は削除 生年月日の「日」は削除 氏名は匿名化(ハッシュ値1) 	<p>【左記に加え】</p> <ul style="list-style-type: none"> 単月分(10月診療)データ 性別及び5歳刻み年齢階層区分で元データの構成比率を保ちながら、1%もしくは10%の割合で抽出
保険者情報	<ul style="list-style-type: none"> 保険者番号 被保険者証等の記号および番号 合計点数等 	<ul style="list-style-type: none"> 被保険者証等記号・番号は匿名化(ハッシュ値2) 	<p>【左記に加え】</p> <ul style="list-style-type: none"> 保険者番号は削除
傷病名 診療行為 医薬品	<ul style="list-style-type: none"> 含まれる 	<ul style="list-style-type: none"> レセプトに記載されている情報は全て含まれる 	<ul style="list-style-type: none"> 希少なものは匿名化
その他	<ul style="list-style-type: none"> 保険外の診療は記録されない 	<ul style="list-style-type: none"> コメントや未コード化傷病名など、テキスト情報は削除される 	<ul style="list-style-type: none"> 高額レセプトは削除 DPCレセプトでは医療機関別係数及びその影響がある項目は削除 患者や医療機関を特定できる可能性のあるその他記載項目(病床数等)はカテゴリー化

○ 特別抽出で提供されるレセプト情報等データと比較して、サンプリングデータセットに含まれている個票データは、十分に個人の特定可能性を下げる処理をした、匿名性の高いデータとして位置づけられるのではないかと。

2. 2 サンプルングデータセットにおける公表基準の適用について

- サンプルングデータセットの利用者から、「公表にあたっての基準のひとつである『最小集計単位10以上』という原則を適用する必要があるのか」という意見が提出され、第16回有識者会議においても「その是非について議論すべきである。」とされたところである。
- サンプルングデータセットは、1ヶ月分のレセプトデータを一定の比率で抽出し、都道府県、保険者、医療機関等の情報は削除し、また希少な傷病名、診療行為、医薬品等の情報に対しては匿名化処理をほどこしたデータである。
- このデータにおいては、医療機関情報は削除されているので医療機関に関する公表基準が適用されることはない。
- 患者数に関する集計表上の1つのセルの人数については何人程度の患者が母集団にいるのか、抽出率から以下のように推定される。

	抽出率	仮に、従来の基準である10以上という原則を適用した場合、母集団にはどの程度の患者がいると推定されるか	仮に、患者数が1人であった場合、母集団にはどの程度の患者がいると推定されるか	一般の公表にあたっての基準
医科入院	10%	100人以上	10人	≧ 10人
医科入院外	1%	1000人以上	100人	> 10人
DPC	10%	100人以上	10人	≧ 10人
調剤	1%	1000人以上	100人	> 10人

仮に「1人」で公表しても母集団には10人以上の該当者がいると推定される

- サンプルングデータセットを用いた成果物の公表にあたっては、公表基準の適用を求めないこととしてはどうか。
- サンプルングデータセットにおいては探索的研究を認めていることから、公表に際しての事前の確認は原則として事務局のみで行うこととし、特に必要な場合を除き有識者における事前確認は求めないこととしてはどうか。

今回の論点

●新システムの導入を受け、以下の点につき内容の充実を行ってはどうか。

【改善点1】年齢階級コードを追加し選択可能とする

【改善点2】転帰区分・疑い病名・主傷病・主傷病決定フラグを導入する

【改善点3】サンプリングデータセット【春】【夏】【冬】を導入する

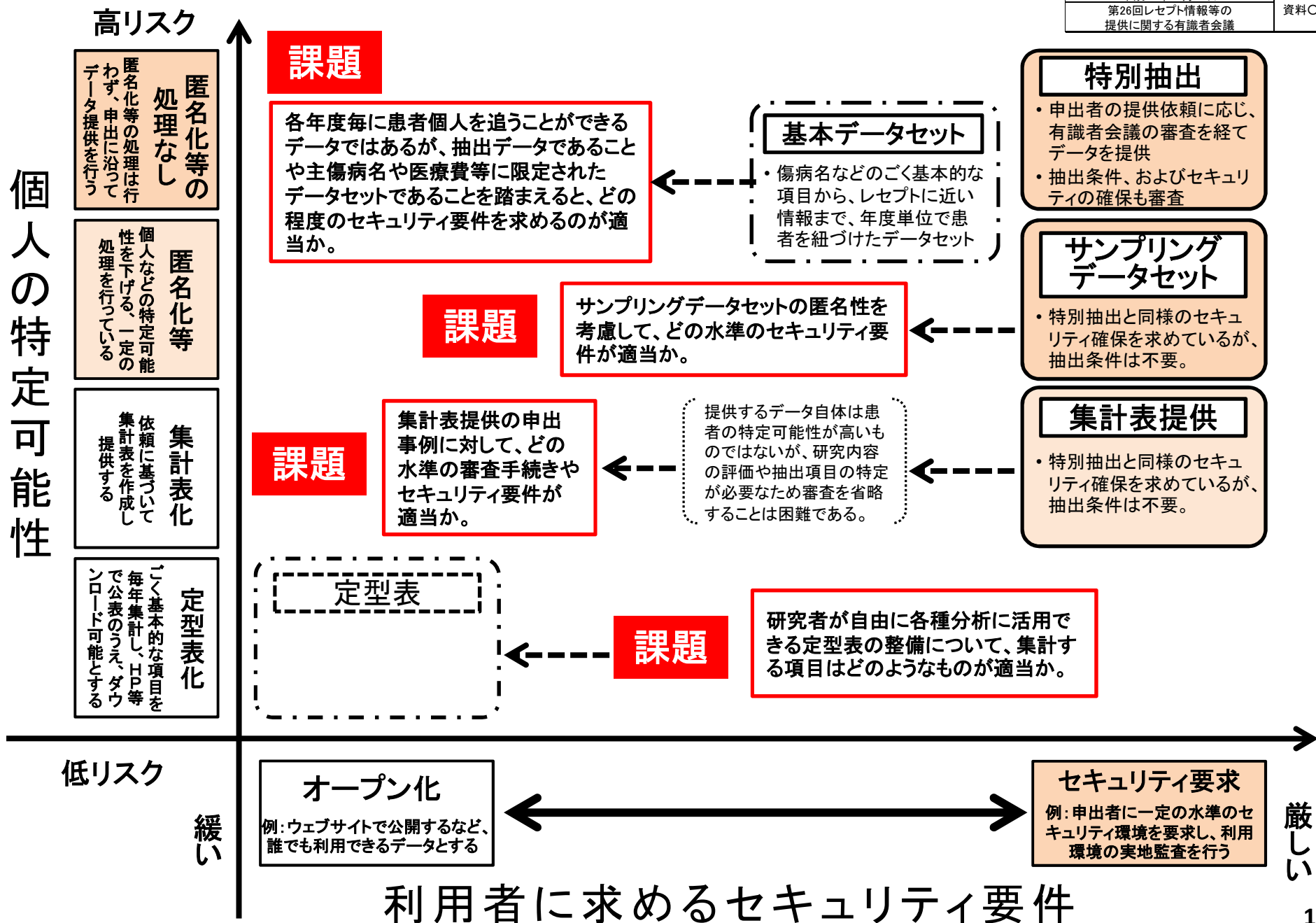
●サンプリングデータセットの特徴から、上記改善を行っても個人の特定性への影響は考えにくいいため、引き続き成果物の公表にあたっては現行の公表基準を継続することとしてはどうか。

●今後の検討課題

歯科レセプトの電子化も相当進んだことを踏まえ、平成28年度以降に歯科レセプトのサンプリングデータセットの作成を検討してはどうか。

參考資料

2.1 提供するデータの特徴とセキュリティの確保について



サンプリングデータセットの利用環境のあり方について

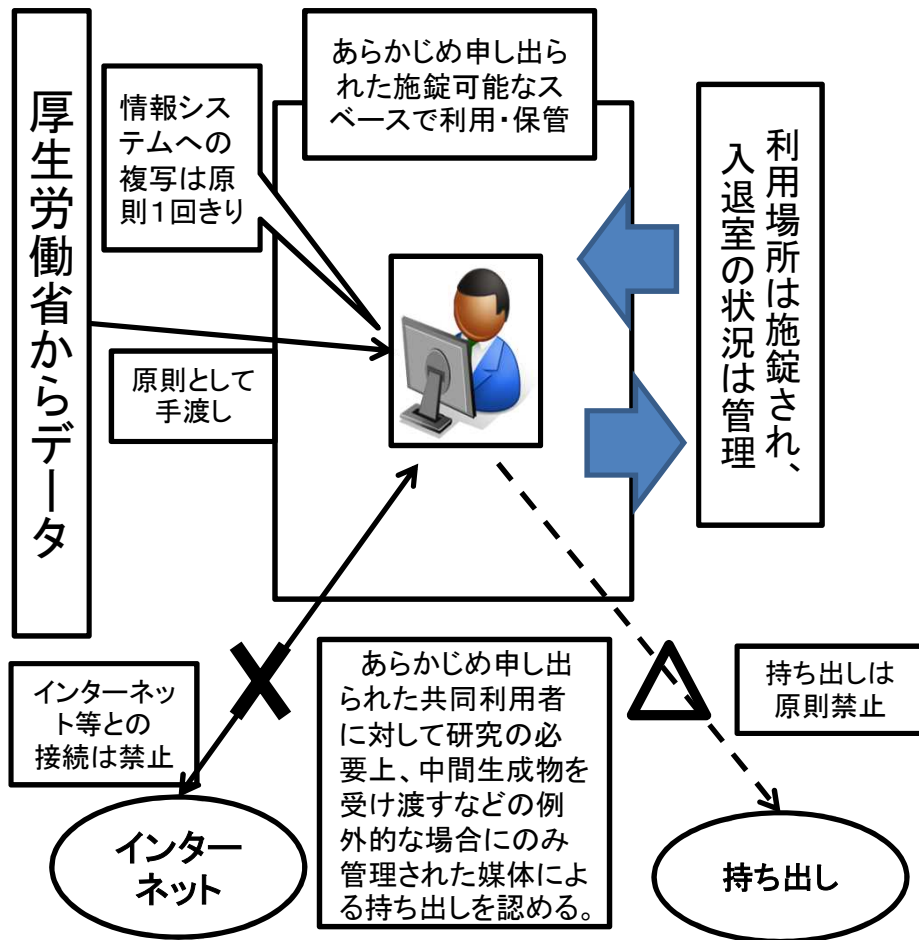
平成25年12月19日
第26回レセプト情報等の
提供に関する有識者会議

資料〇

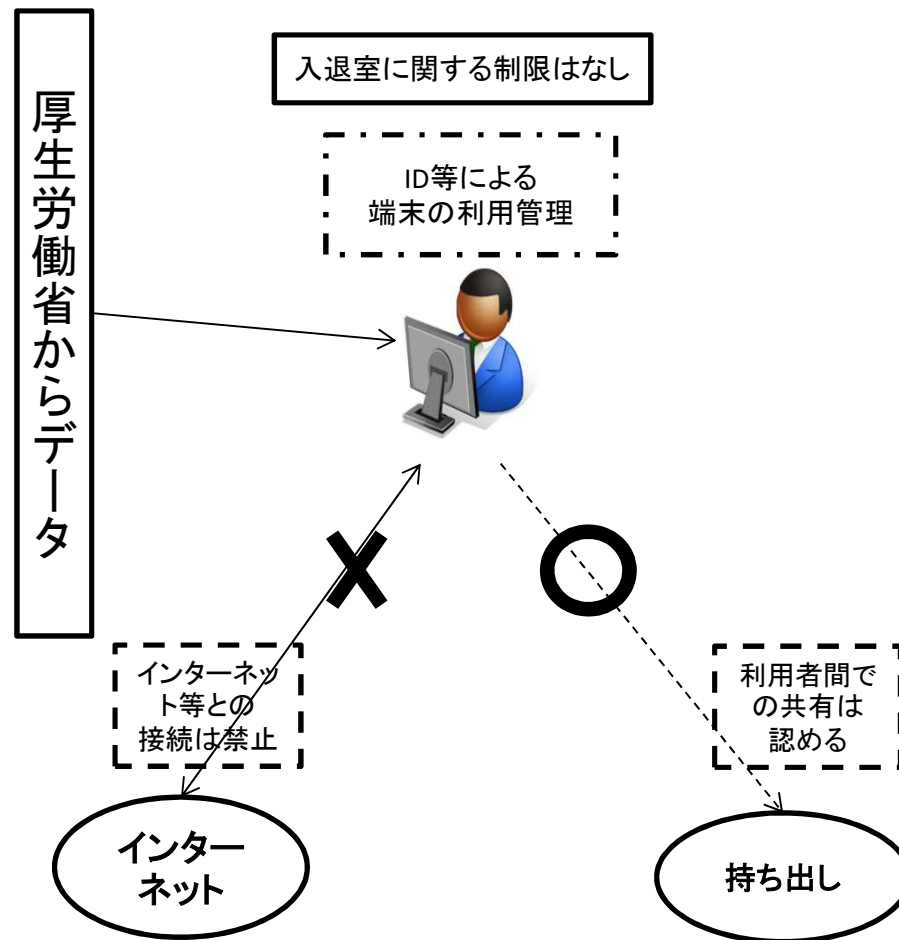
サンプリングデータセットの利用環境の現状および課題

- 「サンプリングデータセット」とは、単月分のデータから一定の割合で抽出し、傷病名や診療行為等の患者関連情報について出現回数の少ないものに一定の匿名化処理を行った「探索的研究へのニーズに対応し、安全性に十分配慮したデータセット」(第9回有識者会議資料より)である。
 - ※ サンプリングデータセットでは、「診断群分類」を除き、傷病名、医科診療行為、医薬品の各コードについて、30%~60%程度の匿名化処理がほどこされている。
- しかし、サンプリングデータセットが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等に鑑み、これまで提供依頼申出者に対して、特別抽出におけるデータ提供と同様のセキュリティ要件を求めてきた。
- 一方、データ利用者からは「探索的研究へのニーズに対応し、安全性に十分配慮したデータセット」であるというサンプリングデータセットの特性に鑑みて、セキュリティ要件の緩和を検討できないか、といった意見が寄せられている(第16回有識者会議資料より)。
- 有識者会議においても、「データ利用については、サンプリングデータセットが匿名性を向上させた安全性の高いものであることに鑑み、匿名性の程度に応じたセキュリティ要件やその提供のあり方について検討してはどうか。」と提言されている。

現在の利用形態(抽出データと同じ)



米国FTCの考え方等を踏まえた利用形態の一例(案)



サンプリングデータセットのデータの特性に鑑み、どのような利用形態が適切か。

サンプリングデータセットの申出手続きについて

別添番号	準備書類題目	現在の申出手続き	米国FTCの考え方等を踏まえた申出手続きの一例
	レセプト情報等の提供に関する申出書(様式1)	必須	必須
別添1	本人確認	必須	必須
	レセプト情報等を利用した研究に関する承認書(様式1-1)	必須	必須
別添2	別添2-1 運用フロー図	必須	(不要)
	別添2-2 リスク分析・対応表	必須	(不要)
	別添2-3 運用管理規程	必須	(不要)
	別添2-4 自己点検規程	必須	(不要)
別添3	所属組織の個人情報保護に関する規程(プライバシーポリシー、情報セキュリティポリシーなど。複数ある場合は、3-1, 3-2…と付番して提出)	任意	任意
別添4	厚労科研交付通知など、公共性の高い研究であることを示唆する書類	任意	任意
別添5	提供依頼申出者における過去の研究実績を証明するもの(論文の写し等)	任意	任意
別添6	外部委託があれば、守秘義務契約の写し	該当時必須	該当時必須
別添7	倫理委員会承諾書	必須	(不要)
別添8	詳細な抽出条件	不要	不要
別添9	詳細な公表形式	必須	(不要)
	その他適宜必要な書類 (添付の際は「別添10」「別添11」…と番号を振っていただき、どういった書類かを明記して下さい)	その他	(不要)

別添2の運用関連書類、別添7の倫理委員会承諾書、別添9の詳細な公表形式(本欄では暫定的に(不要)と記載)等についての書類は、提出を求める必要があるか。