

# 日本における「医療情報DBの活用における連携」のあり方（案）

日本製薬団体連合会・日本製薬工業協会

平成 21 年 10 月

## 要旨

従来より欧米では、種々の医療情報データベース（DB）の二次利用による臨床研究及び薬剤/臨床疫学研究が行われてきた。近年、医薬品の安全監視のため等、より大規模なDBの必要性が高まり、主としてDBの規模拡大の手法の1つとして複数の医療情報DBを用いる「分散型研究ネットワーク」が構築・活用され始めている。

複数のDBを用いる目的・手法は規模の拡大とデータの補完に大別される。前者は、複数の医療機関のDBから同じ項目、形式のデータを収集し合算して扱う研究に代表される。後者は、レセプトDB等大規模で網羅性、追跡性は高いが、情報の質に制約のあるDBを核として、不足する情報を他のDBと連携(リンク/突合)する場合である。

研究に複数のDBを用いるには、各所の医療機関、保険者組織等に散在する複数のデータベース(DB)からオンライン、オフラインを問わず、定型化された手続きによりデータの抽出(データそのもの、もしくはデータのサブセットあるいはサマリー(半集計))および集計が行える状態が重要である。個々のDBにおいて標準化されたデータリポジトリとそこからのデータ検索・集計を可能とするデータウェアハウスが構築されることが理想であるが、メタデータサーバー等により、標準データ交換仕様への変換の仕組みが用意されることで複数のDBに対して、コモン・プロトコール・アプローチにより大規模で網羅的な population の研究成果を得ることが可能となる。

本邦においてはこれまで研究におけるDBの利用体制構築が遅れており、二次利用は限定的であったが、近年、電子カルテ普及に伴う各医療機関内におけるDBの利活用の発展および、現在、厚生労働省による保険診療の全レセプトデータおよび特定健康診査によるデータベース(以下、高確法DB)の構築等により、DBの二次利用への期待が高まっている。

皆保険である本邦の高確法DBには、情報量が限られるが、ほぼ全数の診療基本情報が集められ、網羅性と追跡率が高いという特徴があり、高確法DBを中心として、他のDBにより情報を補完する、または逆に医療機関のDBの結果を高確法DBにより背景の調整を行い、全体を推定するなどの活用様式が考えられる。

単独、複数のDBを問わず、DBを用いた研究は、一例ずつ症例を収集する研究手法に比べて、回顧的研究が容易、大規模な症例を扱える、短期間で実施できる、網羅性が高い、比較的低コストである等の利点がある。DBを研究テーマに合わせて構築するのではなく、すでに存在するDBの特性に合わせた研究を計画する必要がある。また、二次利用であるため、情報の発生(測定)自体が何らかの目的あるいは結果によるものであり、大きなバイアスが存在しうる事を前提とした認識が必要である。DBを用いた研究には種々の制限や限界があることを理解し、研究の目的によっては従来の研究手法と組合せて行うこととなる。

## 「医療情報の二次利用に対する取り組み」の留意事項

- ・ **現存**の電子医療情報の価値を考える
- ・ **標準**を使用する
- ・ **出来ないこと**より、**できること**を考える。

調査・研究をDBに合わせる。

- ・ 一度に全てのことを行わない。
- ・ トライ&エラーを恐れない。
- ・ 完全な域に達するまで待たない。
- ・ 技術ではなく、データを重視する
- ・ 技術面では機械学習も有用である。

## 目次

1 . 前書き .....	4
2 . スコープ .....	5
2.1. データソース.....	5
2.2. 実現可能性 .....	5
3 . 目的 .....	5
4 . データベースの種類.....	5
5 . 複数のデータベース利用の手法.....	6
5.1. 連携データベースモデル例.....	6
5.2. データソース間の併合方法.....	6
5.3. データの標準化要件.....	9
6 . 医療情報DBが有効な場面例 .....	9
6.1. 医薬品、医療機器副作用等、安全性情報の収集、評価 .....	9
6.2. 疾患に関する疫学的情報 .....	9
6.3. 製造販売後調査、副作用自発報告（詳細調査）、調査票を用いた各種疾患研究、行政の調査・申請書類に おける医師・医療機関の負担軽減 .....	9
7 . DB 研究を行なう場合の留意点 .....	10

## 1. 前書き

製薬企業は規制当局と共に医薬品の安全性に責任を負っている。医薬品の安全性を確認し、検証するためには、一般的に膨大な症例数が必要であるが迅速性も要求される。また、添付文書を始めとする安全性と適正使用を確保するための措置の有用性についても、バイアスを可能な限り排除した上で適切に評価する必要がある。これらの課題は使用成績調査に代表される一例ずつ症例を収集する調査では解決が困難であり、当該場面における医療情報データベース（DB）の二次利用の必要性は日本を含む世界で広く認識されている。

米国では、HMO（Health Maintenance Organization, 民間医療保険）がレセプトデータと EHR（Electric Health Record）の混合型のデータベースを産官学に提供するビジネスモデルが確立しており、医薬品・医療機器の安全性対策にも広く活用されている。2007年9月に成立した FDA 再生法のもとで、最近複数のデータソースを利用する、分散型研究ネットワーク（distributed research network）の取り組みが開始され、すでにその有用性の可能性を示している報告もある。このような分散型研究ネットワークは、データをデータ保持者の元に置いたまま、大規模な症例数を必要とする臨床研究・疫学研究をより迅速に可能にすると期待されている。欧州でも The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) と称される EMEA によるプロジェクトがあり、薬剤疫学者及びファーマコビジランス専門家が参画して、関連研究施設、医療機関、保険者 DB、及び希少疾病/疾患領域、副作用に関わる電子化された登録制又は現存ネットワークの活用を試みることにより、医薬品市販後の安全性や薬効欠如に関わる多施設共同研究の実施を促進することで、EU における医薬品のモニタリングを強化することを目的として進行中である。

対して、本邦においては、医療機関において DB は構築されつつあるが、各々の規模が小さいこと、システムが標準化されていないことなどにより、DB の二次利用は発達してこなかった。しかしながら、電子カルテにおける情報交換仕様の標準化及び、厚生労働省により進められている保険診療の全レセプトデータおよび特定健康診査によるデータベース（以下、高確法 DB）の構築等により、DB の二次利用への期待が高まっている。

高確法 DB は本邦のほぼ全ての医療行為の基礎情報を有する計画であり、標本調査ではなく、悉皆調査をルーチンに行うことが可能となり、その潜在的価値は高いが、レセプトを主とした情報源であるため、日付情報が月単位であることや、検査結果、治療の評価を含まず、死亡情報、妊娠出産情報等も一部しか含まれないなどの制約がある。そのため、他の DB による情報の補完、連携が必要な場合が多い。国民医療保険、処方箋、診療、住民票データ等、性質の異なる DB を連携させた DB はデンマーク、スウェーデン、フィンランドで構築されている。

本試案においては高確法 DB を核とした種々の DB との連携及び、医療機関個々の DB を並列して扱う分散型研究ネットワークについて示す。

### ➤ 基本的な考え

薬物治療のアウトカム（安全性・有効性）のスタンダードは、治験、製造販売後調査、副作用・感染症自発報告の詳細調査等（以下、総称して"Primary data collection" とする。）患者背景、曝露、アウトカムの関係を明確にして収集された情報である。

診療・臨床情報の様々なデータにおいては、行為（曝露）と結果の対応が明確で無い場合が多く、また、安全性に関するアウトカムの記録のされ方は様々である。

従って、如何なる DB でも単独では情報が不足であり、個々の DB における情報の不足は、複数のデータソースで補完できる場合がある。

また、二次利用であるため、情報の発生（測定）自体が何らかの目的あるいは結果によるものであり、大きなバイアスが存在しうる事を前提とした認識が必要である。

DBの規模と情報の粒度には逆相関の関係があり、特性に合わせた研究を行う必要がある。(研究に相手を合わせることは現実的でない)

以上のように、DBを用いた臨床研究には種々の制限や限界があり、Primary data collection に代わるものではなく、研究の目的および必要な情報の定義をDBに合わせる必要がある。

DB研究は、Primary data collection による研究に比べて、回顧的研究が容易、大規模な症例を扱える、短期間で実施できる、網羅性(一般化可能性)が高い、低コストである等の利点がある。

DB研究は、上記の利点と欠点をバランスさせて、行うこととなる。

## 2. スコープ

### 2.1. データソース

現在存在するDB(電子的)及びこれらから派生を見込むDB

現時点で全く電子化されていないデータソースは原則考慮していない。

### 2.2. 実現可能性

解決が必要な課題として、費用、法令、社会的認知(世論、支持)、技術的課題(データの発生者である医療機関等の負担を含む)があるが、ここではこのうち前者3つについては考慮していない。

## 3. 目的

本文書は、国民の健康増進のため、迅速に効率よく網羅的に医療情報を二次利用する情報基盤を整備することを目的とし、来院患者、医療情報を生成する各医療機関の関係者(医療従事者・医事関連業務従事者・医療情報部関係者)及び二次的に活用すると考えられる産官学医の各関係者を対象とする。

## 4. データベースの種類

臨床疫学、薬剤疫学に二次利用可能なDBには、患者登録型と自動集積型がある。前者は、各種疾患登録などで、後述するような標準化が行われ、人的・金銭的に継続的メンテナンスが可能であれば、二次利用の可能性は大きく広がる。後者には大きく分けて2つのタイプがある。臨床情報型と医事会計型である。

表 4.1 自動集積型データベースの種類とその特徴

	利点	限界点
レセプトDB(高確法レセプトDB、商用レセプトDB)	<ul style="list-style-type: none"><li>高確法レセプトDBにおいては、将来的にほぼ全ての保険診療の情報が収集され、抽出によるバイアスを考慮する必要が無い</li><li>保険診療を受けている限り、転院、他施設の受信も完全に追跡されている。</li><li>主要な薬剤の処方及び医療行為は高い精度で特定できる</li></ul>	<ul style="list-style-type: none"><li>データの日付精度が「月」である</li><li>アウトカムの種類により特定の精度が大きく影響を受け、アウトカムの「発現率」が過大評価・過小評価になりやすい</li><li>カルテに戻って確認することができないため、データ(特に診断)の適切性が制限され、何らかのバリデーションが必要。</li><li>検査結果が得られない</li></ul>

診療記録DB（例：各医療機関内HIS）	<ul style="list-style-type: none"> <li>・ 検査結果を含む詳細なデータを保持している</li> </ul>	<ul style="list-style-type: none"> <li>・ 比較的規模が小さい</li> <li>・ 転院などにより追跡が不能になる</li> <li>・ データが複雑で扱いにくい</li> <li>・ コーディング・入カールールが統一されていない。</li> </ul>
---------------------	--	---

このようにデータベースには各々特徴（強み、弱み）がある。例えば、レセプトDBを用いて得られた「発現率」を医療機関の診療情報DBから得られた情報（交絡要因と発現率）を用いて補完し、より精度の高い情報を得るなどの対策も考えられる。

## 5. 複数のデータベース利用の手法

### 5.1. 連携データベースモデル例

ここでは一つの研究目的に対して、複数のDBを用いる場合の利用手法について、いくつかのモデルを示す。複数のDBに個々にアプローチする場合には、一つのDBを利用する手法と同様であるため、複数のデータベースに一貫した方法でアプローチする場合の方法について述べる。この場合、複数のDBの代表窓口となる機関（ここでは仮にポータルサイトと呼ぶ）が必要であると考えられる。

表 3.1 複数のデータベースを一貫した方法で利用するためのモデル例

	中央統合型データベース	分散型データベース	データベースネットワーク
特徴	医療機関や、保険者など、各々が保有するデータベースからデータを収集し、中央にデータベースを構築する。	中央データベースはデータを保有せず（メタデータ、変換マスターは保有する）コンピューターネットワークを介した医療機関や、保険者など、各々が保有するデータベースへの接続、問合せ機能のみを有する。	各データベースがWEBを介してアクセス可能なインタフェースを構築する。
データリポジトリ	データは集められ中央に蓄積される。	データは各サイトに蓄積される。	データは各サイトに蓄積される。
アクセス	利用者は、中央データベースのみにアクセスする。	利用者は、中央集約型データベース同様、中央データベースのみにアクセスする。	研究者は、医療機関や、保険者など、各々が保有するデータベースへWEBを介してアクセスし、必要なデータを抽出する。データベースへのアクセス及びデータ抽出手続きは共通のインタフェースにより行える。

### 5.2. データソース間の併合方法

データの併合の方法には link と stack がある。

同一個人異なる情報の統合を行う link は、データソース間で共通の個人を識別する情報が必須である。

stack は例えば複数の医療機関の診療情報の統合で用いるが、症例の重複への対応や時間軸の処理のため、link も必要な場合がある。

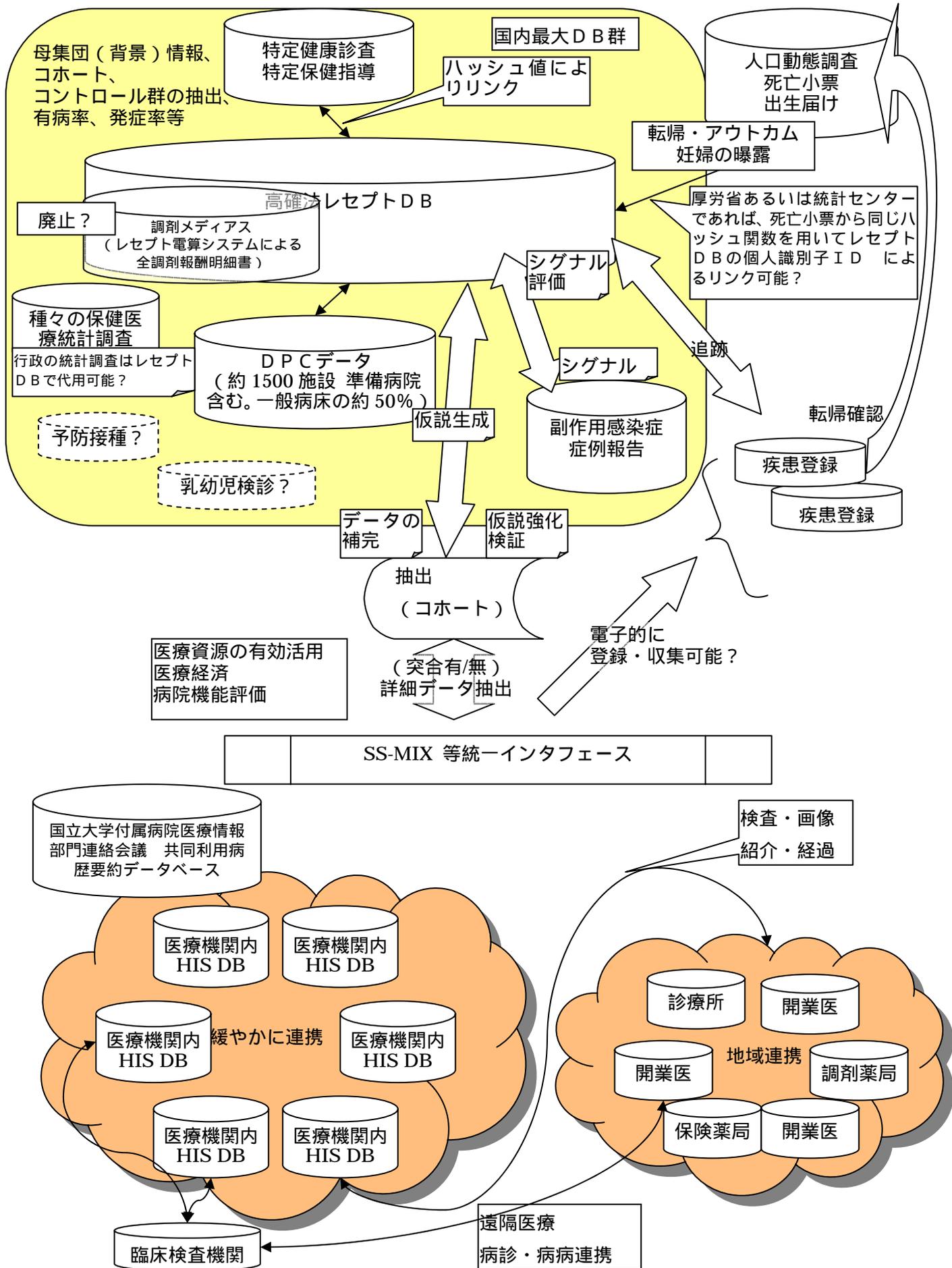
現在、個人情報の保護等のため、データソースの二次利用においては、個人を容易に特定可能な情報を使用することは困難であり、個人識別情報から作成されたハッシュ値（暗号化個人識別子）または、種々の情報から確率的に突合せることとなる。

その際においても、使用できる情報(項目、コード体系)により、突合の効率は大きく変わるため、各データベースにおいて可能な手立てを講じられることが望まれる。

例えば

- ・ 暗号化個人識別子の仕様とその利用に関するルール策定
- ・ 施設毎の各種マスター（コード）の統一あるいはクロスインデックスの作成
- ・ データ項目仕様とガイダンス作成
- ・ 副作用・感染症症例報告への報告者情報（ICH E2B では A.2.1 に規定されているが、現在は PMDA の指導で報告されていない）の運用として、施設コードの送信
- ・ 副作用・感染症症例報告への生年月日の追加（ICH E2B では B.1.2.1 に規定されているが、現在は PMDA の指導で報告されていない）

図 各種データベースの連携可能性



### 5.3. データの標準化要件

複数(数件～数百件)の医療機関内の診療情報DBに保存されているデータを用いて臨床研究を行う場合、収集するデータは症例単位である必要は無く、研究に必要なデータ項目でサマライズされた集計表(項目データの組み合わせ毎の症例数)で可能な場合がある。個々の施設から必要な項目と内容を得るためには、データが共通のフォーマット、コード体系で出力されることが望ましい。また、データの取り出し条件はデータの内容に影響するため、問合せ(クエリー、例えば、「3カ月以内に糖尿病を発症した症例」という条件)に対して同じ解釈がされる共通のインタフェース(コモン・プロトコール・アプローチ)が望ましい。医療機関内では、日々の診療から生じるデータを標準仕様のDBに蓄積(データリポジトリ、データウェアハウス)されていることが理想であるが、施設固有の仕様であっても、メタデータサーバー等、標準データ交換仕様への変換の仕組みが用意されていれば利用可能である。

## 6. 医療情報DBが有効な場面例

より良い医療の提供には、現状把握が不可欠である。しかし、一つ一つの疾患に着目し、それぞれの患者登録を継続してデータを収集する方法は、多大なる時間と労力を要する。例えば保険診療の大半を含むレセプトDBデータベースを経時的に分析することにより、現状をすばやく把握できる疾患もあるだろう。薬剤の使用実態においても、継続・新規別の使用患者数が得られ、安全性を検討する上で重要なリスクサイズを得られる。医療情報に関する種々の活動が連携されれば、より大きな効果が得られる。

### 6.1. 医薬品、医療機器副作用等、安全性情報の収集、評価

個別の症例報告等から得られるシグナルの評価、安全性仮説の検証。

高確法DBでは、曝露・アウトカムの間関係の精度が月単位であること、アウトカムに関する情報が欠損する場合があるなどのため、仮説強化は可能であるが、検証は困難な場合が多い。診療情報DB等と連携あるいは、併用が必要となる。

### 6.2. 疾患に関する疫学的情報

有病率、発症率等の基礎情報、進展・二次病変の発現と治療の関係等。

疾患登録のフォロー強化

現在、多くの組織(官、学)が各々データの収集、登録及び追跡等に多大な労力と費用を費やしている。医療機関のDBから電子的にデータを取り出す、症例の追跡に高確法DB、死亡小票と連携する等により、労力並びに費用の削減のみならず、質の向上(追跡率向上、データ収集の迅速化など)も可能となる。

有病率、発症率等の基礎情報の収集には、高い追跡率と診断の網羅性、正確性が必要である。高確法DBは前者の条件は満たされるが、いわゆる「レセプト病名」のため、後者については医療機関の診療情報、検査結果などによる補完が有用である。

### 6.3. 製造販売後調査、副作用自発報告(詳細調査)、調査票を用いた各種疾患研究、行政の調査・申請書類における医師・医療機関の負担軽減

専用の調査票を用いた調査において、多くの項目がカルテ等からの転記により行われている。これをSS-MIXなどにより医療機関のDBから電子的に取り出すことにより、医師等は転記が不要となり、機械的に取り出せない項目(多くの場合、評価等、A4 1ページ程度)のみを調査票に記入(DB二次利用とprimary data collectのハイブリット)することでデータの質を高め且つ労力が軽減可能となる。

## 7. DB 研究を行なう場合の留意点

- 副作用（有害事象）症例の探索を例に

副作用症例報告は、極めて一部の症例であると考えられ、未報告の症例の存在が推定されるため、レセプトデータ、臨床データ共（以下DB）によるサーベイランスが期待される。しかしながら、レセプトのデータ、特に傷病名はいわゆる「レセプト病名」、治療の対象とならない未記載病名の存在並びに記載がバリエートされていない等もあり、レセプトデータ単独で任意（特定）の副作用（有害事象）症例を定義、抽出することは困難な場合が多く、抽出条件（クエリー）の設定に試行錯誤が伴い、容易でない。

バリエートされた情報である副作用症例報告症例とDBとの突合による任意（特定）の副作用（有害事象）のDBにおける副作用の表現形をクエリーに用いてDB内での類似症例を検索することにより効率的（半自動的）な抽出が可能となると考えられる。

例えば、ある薬剤Aにおける「横紋筋融解」を調査する場合、PMDAに報告（2004年以降）された全薬剤の「横紋筋融解」（約800例）をDBと突合する。この時、DB内で必ずしも全例に「横紋筋融解」と判断できる記録はなく、何割かは不明（と都合できない症例も）となるであろうが、このことでDBから得られる結果の特異度の推定が可能となる。

そして、判別できた症例に記録されている情報（診断名、検査、処置、使用薬剤等）を用いて、「横紋筋融解」と判別できる症例をDBから抽出する。これにより、薬剤Aのみでなく、類薬や、背景発現率に関する情報も得られ、自発報告では叶わない科学的な評価が可能となる。

各種医療情報DBの特性

項目	レセプト +特定健康診査	D P C	医療機関内 H I S D B	地域診療情報ネ ットワーク	各種の疾患登 録	副作用感染症 症例報告	製造販売後調 査	死亡小票
規模（年間）	4千万例	100万例単位	1万～数十万	数十～数百万例	数百～数万例	25000例	100～1万例	100万例
網羅性（対全 数）	++++	++	-～+	+++	-～+	-	-～+++	+++++
診療行為デー タ	+++ 検査項目の明 細が無い場合 有	++++ Fファイル？ レセプト+	+++++	+++	+～++++	++++	++++	-
日付単位	月	日	日～時分	日	月～日	日	日～時分	日
追跡性	++++ 医療行為を伴 わない死亡は 記録されない。	+	+	+++	+～++++	-	-	-
検査結果	-	-	+++++	-～+++	-～+	-～+++	-～+++	-
アウトカム、 イベント	-～++ 死亡小票の補 完により++++	-～++	++～++++	-～++	+～++++	+++++	++～+++++	+++
アウトカム、 イベントの原 因	- 死亡小票の補 完により++	-	-～++	-～++	-～++	+++++	++～+++++	++++
有病率（患者 数）	++++	-	-	-～++++	-～++	-	-	-
発生（症）率	+～++++	-～+	-～+	-～++++	-～+++	-	-～+	-～++
曝露患者数	+++++	-～+	-～++	+～+++	-～++	-	-～++	-
疾患疫学	+～++++	-～+++	+～++	+～+++	++～++++	-	-～++	-～++

