

## 糖尿病に関するトピックス紹介 by 研究推進委員会 No.35

医療情報データベースを利用したコホート研究を計画する上での期間設定及び共変量設定の留意点について

執筆担当：齋藤 智之（順天堂大学大学院 革新的医療技術開発研究センター）

トピックス担当：林 太祐（日本医科大学付属病院 薬剤部）

### 1. はじめに

研究の手法にはさまざまな方法があるが、今回のトピックスでは、タイトルにもある通り医療情報データベースを利用したコホート研究の期間及び共変量の設定について記載した。昨今、云うまでもなく医療情報の電子化が進んでおり、データベースに蓄積された医療情報を利用した薬剤疫学研究が世界的に盛んに行われている。医療情報のデータ利用といえば、匿名レセプト情報・匿名特定健診等情報データベース（NDB）<sup>※1</sup>や、その他の医療ビッグデータ<sup>※2</sup>を想像するが、NDB利用のハードルが高いことや、その他の医療ビッグデータにおいてはコスト面等、臨床の現場で働く薬剤師にとっては現実的に利用することが難しい状況もある。一方で、自施設の電子カルテ等に蓄積されたデータであれば、比較的利用しやすいのではなかろうか。そのため、臨床の現場で働く多くの薬剤師にとっても、データベースを利用した研究を実施するチャンスがあると考えている。

そこで、今回はデータベースを利用したコホート研究を行う上で、研究計画書作成時の留意点について、特に「期間」及び「共変量」の設定について取り上げる。この2つを挙げた理由として、執筆者自身もこれら2つの設定に紛らわしさを感じることや、研究計画書の確認を依頼された際に、設定が十分でないことをしばしば経験するためである。そのため、本文が、特に、これからデータベースで研究を始めるにあたり、研究計画書の作成を検討されている方の参考になれば幸いである。研究計画書を作成するタイミングについては、その時の状況にもよるが、研究計画の実施可能性を明らかにすることが大切であること、倫理審査時に倫理審査の資料に添付して提出が求められる場合があることを踏まえて、早めに準備することをオススメする。

なお、本文中の記載は執筆者の理解に基づく記載であるため、専門書や文献等の記載と異なる場合があることにご留意いただく必要がある。また、本文の執筆にあたっては「製造販売後データベース調査実施計画書の記載事例集」<sup>1</sup>を参考にしている。具体的な記載例については、こちらを参照されたい。なお、本文中で使用する用語については表1に示した。

---

※1 厚生労働省は、「高齢者の医療の確保に関する法律」に基づき、レセプト情報と特定健診等の情報を収集し、NDBに格納してきた。行政機関や研究者等にNDBデータの提供を行っている。

※2 メディカル・データ・ビジョン株式会社（MDV）や株式会社JMDC等の薬剤疫学研究に利用可能なデータベース。

表 1. 用語の定義

用語	定義
データ期間	データベースからデータを抽出する期間
対象集団	データ期間内で研究対象とする集団
組み入れ基準	対象集団に組み入れる基準
除外基準	対象集団に加えない基準
組み入れ期間	組み入れ基準と除外基準を適応してコホートエントリーする期間
共変量	研究結果への影響が予測される因子
Look back period	過去の曝露状況の確認や共変量の情報を取得する期間
観察期間	データ期間内の最初の医療記録日から最後の医療記録日までの期間
曝露期間	Gap period 及び Grace period で考慮して、複数の処方期間を繋いだ一連の期間
処方期間	処方開始日から処方終了日までの期間
処方継続期間	曝露期間と同一であることが多い
Index date	コホート研究ではコホートエントリー日であることが多い
Gap period	前回処方終了日から次回処方日の間の期間であり、処方が継続していると判断する一定の期間
Grace period	最後の処方終了日から薬効が継続していると考えられる一定の期間
追跡期間	疾患や副作用等の発生を評価するために、個人や特定の集団を観察する期間
感度分析	仮定の変化等が、結果にどの程度影響するかを調べることで、結果の頑健性を確認する方法

## 2. 各期間の設定について

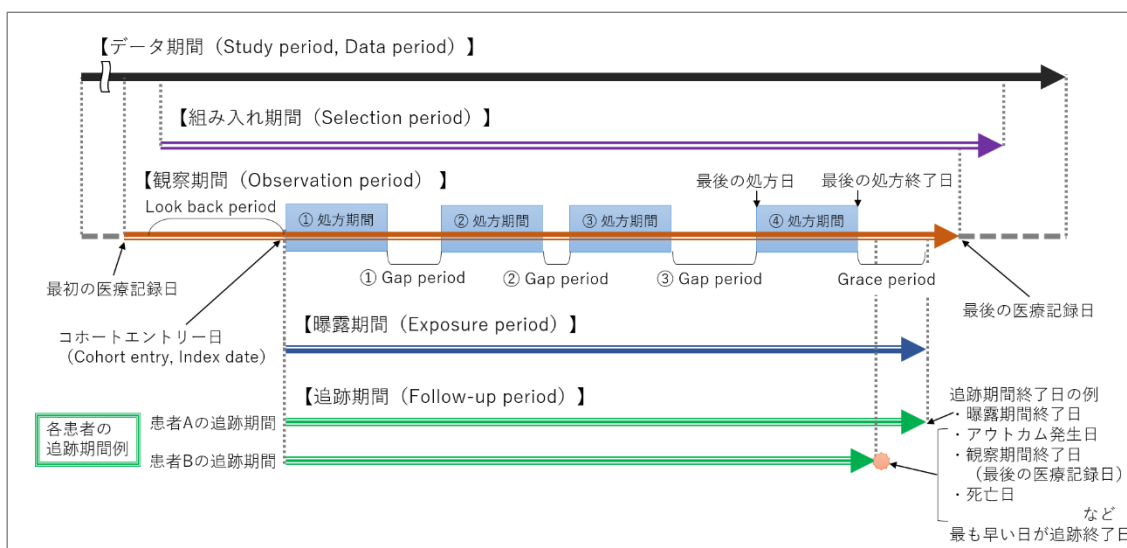


図 1. データベース研究の各期間設定について

各期間について図 1 に示した。ここで、一部の設定に関して補足を加える。まず、追跡期間について、特に、追跡期間の終了日は研究によって設定が異なり、図 1 の中で例を挙げているが、研究毎に十分な検討が必要である。追跡期間の終了日が適切に設定されていない場

合、リスク評価できない期間を追跡期間として含めてしまうことがあり、リスクを過小評価する可能性がある。その理由として、コホート研究での発生率(incidence rate)の計算方法(単位人・時間あたりの発生数)は以下の式<sup>2</sup>で表され、分母に時間を含むためである。

$$I = \frac{A}{T}$$

$I$ : 発生率

$A$ : 発生数

$T$ : リスク人・時間

データベース研究では、最後の医療記録日以降はアウトカムが発現したか否かの判断ができないため、最後の医療記録日を追跡期間の終了日として設定することをオススメする。(曝露期間の終了日に追跡期間の終了日を設定することでも差し支えない)

次に、Look back period の設定について、Look back period は過去の曝露状況の確認や共変量の情報を取得する期間として一定期間設定するため、除外基準の一つに「Look back period に設定した期間より短いデータ期間の患者」を設定することがあり、その場合はむやみに長いLook back period を設定すると対象者が減ってしまうことに注意が必要である。そのため、Look back period の期間の設定に迷う場合は、例えば、主解析ではLook back period を6か月と設定し、主解析の結果の頑健性を確認することを目的に、Look back period を9か月又は12か月に変更した感度分析を実施することもある。

3つ目にGap period 及び Grace period の設定について、Gap period は受診のサイクルや残薬を考慮して設定することが多く、Grace period は処方終了日から薬効が継続していると考ええる一定の期間と考えて設定することが多い。どちらの期間設定においても明確な根拠を示せることが多くないこともあり、それぞれの期間設定は様々な角度から検討することになる。なお、例えば、図1に示す③Gap period が Gap period として設定した期間より長い場合は、最後の処方期間を③処方期間と判断し、③処方期間の後ろに Grace period を加えた期間を追跡期間として判断することになる。

図1に示すいずれの設定においても、研究毎に考慮すべき事項が異なるため、先行研究を参考にしつつも、利用するデータベースの特性や取得可能な対象者の数等を考慮して設定することになるだろう。なお、図1に示した記載について、必ずしも全てが定義されていないこともあるが、共同研究者と共通認識を持って研究を進めることが大切であると考え、研究途中で共同研究者の引継ぎがあり得ることを踏まえると、より正確に設定し、かつ当該設定理由については、いずれも詳細に検討することをオススメする。

### 3. 共変量の設定について

表2. 共変量設定の例

因子	情報取得期間	区分	元データ	変数名	コード
性別	コホートエントリー日	2値(男/女)	Pt_Data	Sex	
年齢	コホートエントリー日	連続変数	Pt_Data	Age	

診療科	コホートエントリー日 (曝露薬剤を処方した診療科)	分類変数	Pt_Data	Department	
高血圧	Look back period	2 値(有/無)	Disease	ICD10code*	I10
脂質異常症	Look back period	2 値(有/無)	Disease	ICD10code*	E78.5
妊娠糖尿病	コホートエントリー日以前の 観察期間	2 値(有/無)	Disease	ICD10code*	O24
ステロイド*	Look back period	2 値(有/無)	Medication	ATCcode†	H02A2, H02B
β遮断薬	Look back period	2 値(有/無)	Medication	ATCcode†	C07A
免疫チェック ポイント阻害薬	コホートエントリー日以前の 観察期間	2 値(有/無)	Medication	ATCcode†	L1G5

\* International Classification of Diseases (ICD) 10 code

† The Anatomical Classification of Pharmaceutical Products has been developed and maintained by the European Pharmaceutical Market Research Association (EphMRA).

<https://www.ephmra.org/media/4973/atc-guidelines-2021-final.pdf>

共変量についてもしばしば設定が十分でないことが見受けられ、執筆者自身も共変量設定の甘さを経験することがある。そこで、設定漏れなく、かつ共同研究者との情報共有も踏まえて判りやすく記載する方法として、表 2 のように記載することをオススメする。共変量の設定においては、情報をどこから取得するのか（「元データ」はデータを取得するデータセット名を指し、「変数名」はそのデータセットのどの列であるか）、取得した情報はどのように取り扱うのか（連続変数又は分類変数）等、情報の取得可能性も踏まえて明確にすることが重要であり、研究を計画する時点で、どこにどのようにデータが格納されているかを知ることで、適切に共変量を設定することができると思う。

コホート研究であれば、共変量の情報の多くは Look back period で取得することが考えられるが、一方で、データベース研究の場合、設定した Look back period で情報を取得しきれない可能性もあるため、具体例を交えて共変量設定の留意点について記載する。例えば、過去の治療歴を共変量として設定したい場合、治療が完了している傷病名を Look back period のみで情報を取得できない可能性がある。具体例としては妊娠糖尿病の既往を共変量とする場合等が考えられ、この場合はコホートエントリー以前の観察期間で情報を取得する方法も一案である。同様に、過去に使用した薬剤を共変量に設定したい場合においても、look back period より過去の曝露状況を考慮した設定とすることも考えられる。

その他、共変量設定において留意する事項として、例えば、診療科については、複数の診療科の情報が記録されていることが予想されるため、どの診療科の情報を取得するのか、又は複数の診療科の情報を纏めて分類変数（例えば、内科、外科等）として扱うのか等を明確に定義する必要がある。また、年齢を共変量として設定する場合は、格納されているデータが、○歳○か月なのか、年齢区分（5 歳毎）でしか情報を取得できないのか等の確認が大切である。

上述に加えて、共変量を設定する際に、傷病名や薬剤の特定に用いるコード等を事前に明らかにすることも、研究を正確かつ円滑に進める上で大切である。なお、共変量の設定にお

いても、期間の設定で記載した理由と同様に、設定根拠まで検討し、研究計画書に記載することをオススメする。

#### 4. 最後に

今回は、執筆者の経験を踏まて、医療情報データを利用したコホート研究の期間及び共変量の設定における留意点について記載した。これからデータベースを利用した研究計画を立てようとしている方の参考になれば幸いである。また、データベースを利用した研究論文を読む際の補助としても利用できるであろう。なお、詳細な研究計画を検討する上では、本文に記載した事項以外に多くを検討する必要があるため、「製造販売後データベース調査実施計画書の記載要領<sup>3)</sup>」及び専門書並びに文献等を参照されたい。

なお、研究を実施する際には、データベース研究においても倫理審査を要することは云うまでもないが、執筆者は近年増えているオープンデータを利用した研究であっても倫理審査を受ける必要があると考えている。理由としては、研究倫理は患者の権利保護に留まらず、利益相反や資金の流れ等、多岐に渡るためである。倫理審査で研究の倫理性及び透明性を明らかにし、研究計画書で計画性を示すことで、質の高い研究を目指していきたい。

#### 利益相反

執筆者は本文に関連し、現在、開示すべき利益相反関係にある企業等はない。

#### 参考文献

1. 製薬協医薬品評価委員会. 製造販売後データベース調査実施計画書の記載事例集. 2018年8月.  
[https://www.jpma.or.jp/information/evaluation/results/allotment/lofurc000000b3cb-att/db\\_inspect.pdf](https://www.jpma.or.jp/information/evaluation/results/allotment/lofurc000000b3cb-att/db_inspect.pdf)
2. 小森 哲志, 鍵村 達夫. コホート研究とケース・コントロール研究 研究デザインの最近の動向. 薬剤疫学 (1342-0445)18巻2号 Page95-111(2014.03).
3. 独立行政法人 医薬品医療機器総合機構. 製造販売後データベース調査実施計画書の記載要領. 平成30年1月23日. <https://www.pmda.go.jp/files/000222302.pdf>