

医療情報のデータベース等を用いた薬剤疫学研究の実施
に関するガイドライン（案）

初版：平成 25 年 7 月 9 日

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

目次

1. 序文.....	5
(1) 本ガイドラインの目的と範囲.....	5
(2) 本ガイドラインの対象者.....	5
(3) 背景.....	5
(4) 本ガイドラインの改訂.....	7
2. 研究実施計画書の作成.....	8
3. データソース.....	12
(1) 薬剤疫学研究に使用される主なデータソースと選択.....	12
(2) 標準コード体系.....	15
(3) バリデーション.....	16
4. 研究デザインと研究の実施.....	17
(1) 研究デザイン.....	17
(2) 研究デザインに関する留意事項.....	20
(3) データマネジメント、解析の実施.....	24
(4) 品質の確保.....	24
(5) 個人情報保護と倫理.....	25
5. 研究結果報告書の作成.....	25
6. 研究結果の公表.....	28
7. 参考文献.....	29

1 **用語の定義**

2 各用語について、本ガイドラインでは以下のとおり定義する。

3

4 **HIS/EMR データ**

5 Hospital Information System / Electronic Medical Record データのこと。HIS は病院
6 情報システムと呼ばれ、医療機関が診療業務において使用するオーダー等の情報を管理す
7 るシステムであり、傷病名や各種検査、医薬品等のオーダー及び検査結果情報等が含まれ
8 る。EMR は電子カルテを指し、個々の患者に対する診療の内容や経過、診断等が記録されて
9 いる。両者を併せて医療機関における診療に関するデータを指す。

10

11 **一次利用**

12 当該研究実施を目的として収集されたデータを当該研究に利用すること。

13

14 **ケース（症例）**

15 研究対象集団における特定の疾病や健康状態のこと。またはそれらを有する人。薬剤疫
16 学研究では、一人の患者であっても病気、怪我の種類毎にそれぞれ別の症例として数えら
17 れる。

18

19 **シグナル検出**

20 シグナルとは、世界保健機関の定義では「これまで知られていないか根拠が不十分な有
21 害事象と医薬品との因果関係の可能性について報告された情報」のことである^[1]。その意
22 味をより明確にするために「安全性シグナル」と呼ばれることもある。シグナル検出につ
23 いてはこれまでに国際的に合意された定義はないが、本ガイドラインではシグナルを検出
24 することをシグナル検出という。薬剤疫学研究における仮説生成とほぼ同義である。

25

26 **シグナル精査**

27 シグナル精査についてはこれまでに国際的に合意された定義はないが、本ガイドライン
28 では、シグナル検出よりも精度の高い方法により検出された有害事象と医薬品の関連性を
29 推定することをシグナル精査という。

30

31 **シグナル検証**

32 シグナル検証についてはこれまでに国際的に合意された定義はないが、本ガイドライン
33 では、シグナル検出された有害事象と医薬品の関連性を定量的にかつ精度高く推定するこ
34 とをシグナル検証という。薬剤疫学研究における仮説検証とほぼ同義である。

35

36

1 **新規使用者**

2 ある医薬品を新規に使用開始する人のこと。英語では new user や incident user などと
3 呼ばれる。

4

5 **二次利用**

6 当該研究実施以外の目的で収集されたデータを当該研究に二次的に利用すること。

7

8 **未知・既知**

9 副作用が各医薬品について未知の事象であるか、既知の事象であるかを区別する際に使
10 用する。医薬品の添付文書の使用上の注意に記載されていない、または記載されている内
11 容と事象の進行度、重篤度等の臨床状態が異なり、同じ事象と考えられない事象は未知と
12 いう。逆に使用上の注意に記載されている内容に一致する事象は既知という。

13

14 **薬剤疫学研究**

15 薬剤疫学研究には、コホート研究やケース・コントロール研究だけではなく、処方実態調
16 査等の記述的研究やアクティブサーベイランス等を含む。また、PMDA が安全対策の目的で
17 実施する調査、製薬企業が規制の範囲で若しくは自主的に実施する安全性に関する調査及
18 び研究者が行う研究を含む。

19

20 **臨床イベント**

21 临床上発生する事象（疾患や兆候・症状）のこと。

22

23 **レセ電**

24 レセプト電算処理システムの略名。レセプト電算処理システムとは、「保険医療機関又は
25 保険薬局が、電子レセプトをオンライン又は電子媒体により審査支払機関に提出し、審査
26 支払機関において、受付、審査及び請求支払業務を行い、保険者が受け取る仕組みのこと」
27 （社会保険診療報酬支払基金ホームページより）。

28

29 以上の用語以外に本ガイドラインにおいて用いられている専門用語については、適宜専
30 門書を参考とすること。

31

32

- 1 略語
2
- 3 **ATC**
4 解剖治療科学分類法 (Anatomical Therapeutic Chemical Classification System) の略
5 名。
6
- 7 **CCI**
8 チャールソン併存疾患インデックス (Charlson Comorbidity Index) の略名。
9
- 10 **DRS**
11 疾患リスクスコア (Disease Risk Score) の略名。
12
- 13 **FDA**
14 米国食品医薬品局 (U.S. Food and Drug Administration) の略名。
15
- 16 **ICD**
17 疾病及び関連保健問題の国際統計分類 (International Statistical Classification of
18 Diseases and Related Health Problems) の略名。
19
- 20 **ICH**
21 日米 EU 医薬品規制調和国際会議 (International Conference on Harmonisation of
22 Technical Requirements for Registration of Pharmaceuticals for human use) の略名。
23
- 24 **ISPE**
25 国際薬剤疫学会 (International Society for Pharmacoepidemiology) の略名。
26
- 27 **PMDA**
28 独立行政法人医薬品医療機器総合機構 (Pharmaceuticals and Medical Devices Agency)
29 の略名。
30
- 31 **PS**
32 プロペンシティスコア (Propensity Score) の略名。
33
- 34 **WHO**
35 世界保健機関 (World Health Organization) の略名。

1. 序文

(1) 本ガイドラインの目的と範囲

本ガイドラインは、独立行政法人医薬品医療機器総合機構（以下「PMDA」という。）及び製薬企業が医療情報のデータベースを利用して医薬品の安全性評価を行う際に、適切な薬剤疫学研究が実施されるよう留意事項をまとめたものである。また、本ガイドラインは医療情報のデータベースを用いる研究が主な対象となるが、一部の留意事項は従来型のデータ収集による薬剤疫学研究にも共通するため、それらの研究においても部分的に本ガイドラインが参考となるであろう。

(2) 本ガイドラインの対象者

本ガイドラインの対象者は、主に PMDA と製薬企業である。本ガイドラインは、主に学術的な観点から薬剤疫学研究の実施における留意事項をとりまとめているため、学術機関に属する研究者が薬剤疫学研究を実施する際にも参考になり得る。

本ガイドラインが、PMDA、製薬企業及び学術研究者による適切な薬剤疫学研究の実施の推進及び活性化に寄与することが期待される。

(3) 背景

医薬品の安全性評価方法の現状

これまで我が国の医薬品の安全対策は、医療機関及び製薬企業からの副作用自発報告及び製薬企業による製造販売後調査を主たる情報源として行われてきた。副作用自発報告は、稀な副作用や重篤な副作用の検出等には特に有用であり、医薬品の安全性を評価する上で重要な情報源であるが、実際に臨床現場において副作用と認識された事象の一部しか報告されないこと、またその割合が様々な要因で変動するといった報告バイアスが存在することが知られている。また、発生頻度が求められず、対照群を置いたリスクの定量的かつ適切な評価等も行えないため、安全対策上の判断が困難な場合も存在する。例えば、未知の副作用が 1 件のみ報告された場合、実際の発生頻度が分からなければ公衆衛生上の影響度を推定することは難しい。また、ある疾患分野において背景発現率が高いことが知られている臨床イベントが医薬品の使用後に発生した場合や医薬品の作用が小さい場合、原疾患によるものか医薬品の副作用によるものかを副作用自発報告から評価することは困難である。更に、併用薬が多数存在する場合には、被疑薬として報告された特定の医薬品と副作用との因果関係を適切に評価することも困難である。

近年、高齢患者の増加に伴い併存疾患を多く持つ患者が増加していることや、新規医薬品の開発による医療用医薬品の多様化等により、医薬品の使用環境は変化してきている。市販後の医薬品の安全性評価もそのような変化に対応するために、副作用自発報告のみではなく、より多くの情報源と新たな安全性の評価方法が求められるようになってきた。

1 安全性評価への薬剤疫学的手法の導入

2 副作用自発報告のみでは評価が困難な上記のような事例では、個々の副作用自発報告症
3 例の評価よりも集団を対象とした薬剤疫学的手法が有用となることがある。薬剤疫
4 学とは人の集団における薬物の使用とその効果や影響を研究する学問であり、医薬品の使
5 用実態などに関する記述的研究の他、ある医薬品とある有害事象との関連性を統計解析に
6 よって定量的に評価する分析的研究が含まれ、医薬品の安全性評価に密接に関わっている。
7 従来の薬剤疫学研究では、研究対象とする集団に関する情報の収集は、研究毎に症例報告
8 用の様式を作成し、関係者に記入を依頼し、得られた情報を手入力により電子化すると
9 う手作業が中心であった。これには人的資金的なコストと長期にわたる研究期間を要し、
10 実施可能な薬剤疫学研究の範囲や規模は大きく制約されていた。しかしながら、ここ数年、
11 国内において医療情報のデータベースが少しずつ利用可能となっており、研究のため
12 に収集されたデータを一次利用するという従来型の方法の他に、医療情報のデータを二次
13 利用する薬剤疫学研究の実施が可能となりつつある。

14 以上のような背景から、行政において医療情報のデータベースを利用した薬剤疫学的手法
15 を市販後医薬品の安全性評価に利用する動きが生じている。厚生労働省により設置され
16 た「薬害肝炎事件の検証及び再発防止のための医薬品行政のあり方検討委員会」がとりま
17 とめた提言「薬害再発防止のための医薬品行政等の見直しについて（最終提言）」（平成 22
18 年 4 月 28 日）において、「電子レセプト等のデータベースを活用し、副作用等の発生に関
19 しての医薬品使用者母数の把握や投薬情報と疾病（副作用等）発生情報の双方を含む頻度
20 情報や安全対策措置の効果の評価のための情報基盤の整備を進めるべきである」と述べら
21 れており、医療情報のデータベース化と医薬品の副作用等に関する情報収集・評価の手法
22 や体制の構築は、市販後安全対策の重要な課題の一つと認識されている。更に、同省によ
23 り設置された「医薬品の安全対策における医療関係データベースの活用方策に関する懇談
24 会」がとりまとめた提言「電子化された医療情報データベースの活用による医薬品等の安
25 全・安心に関する提言（日本のセンチネル・プロジェクト）」（平成 22 年 8 月 25 日）にお
26 いては、医薬品等の安全対策に資するための医療情報データベース構築の必要性、デー
27 タベースの規模やナショナルレセプトデータベースとの連携への期待等が示された。

28 海外においては、数百万人、数千万人規模の医療情報のデータベースを用いた薬剤疫学
29 研究が既に盛んに実施され、多くの研究成果が公表されている。米国 Food and Drug
30 Administration（以下「FDA」という。）では Sentinel Initiative として FDA の医薬品等
31 の安全対策を目的に、大規模な医療情報のデータベースに薬剤疫学的手法を適用した評価
32 体制の構築等が進められている。また、欧州においては European Medicines Agency が薬
33 剤疫学と市販後安全性監視を強化するため、ENCePP と呼ばれる、薬剤疫学研究を実施す
34 るための基盤整備が進められている。

35 我が国においてもナショナルレセプトデータベースの利用に関し、平成 22 年度より「レ
36 セプト情報等の提供に関する有識者会議」が開催され、平成 23 年度から平成 24 年度まで

1 の「試行期間」には申出のあった研究者に対してレセプトデータの提供が行われ、平成 25
2 年度からは本運用が開始される。更に、PMDA においては前述の提言等を踏まえ、MIHARI プ
3 ロジェクトとして平成 21 年度より薬剤疫学的手法を医療情報のデータベースに適用し、市
4 販後医薬品の安全性評価に利用するための検討が行われている。また、平成 23 年度より厚
5 生労働省と PMDA が協力して医療情報データベース基盤整備事業として、全国の 10 の拠点
6 医療機関に標準化された医療情報のデータベースと連携のためのネットワークの構築を進
7 めている。平成 24 年 4 月には、厚生労働省医薬食品局安全対策課長および、審査管理課長
8 の二課長通知により医薬品リスク管理計画の導入が公表された。医薬品リスク管理計画で
9 は、製薬企業は新薬等の医薬品の安全対策のため、懸念されるリスクに適した安全性の監
10 視計画を策定することが求められており、その安全性監視の手段の一つとして医療情報の
11 データベースを用いた薬剤疫学的手法が位置づけられている。

12 上述のような流れから、今後、医療情報のデータベースを用いた薬剤疫学研究が実施さ
13 れる機会は増加すると見込まれる。研究実施に際しては、データベースの選択、データの
14 特性の把握、アウトカム定義のバリデーション等、留意すべき事項が多々あるため、本ガ
15 イドラインにそれらの留意事項をまとめた。

16 17 (4) 本ガイドラインの改訂

18 本ガイドラインに示す留意事項は、作成時点における医療情報のデータベースの状況や
19 薬剤疫学の一般的な知見・手法等を反映したものであり、これらの情報は時間経過と共に
20 変わり得るものである。従って、本ガイドラインは今後新たな情報等を踏まえて改訂され
21 ることがある。

22 23 24 25 薬剤疫学研究の実施に関するガイドライン作成検討会委員（五十音順）

26	赤沢 学	明治薬科大学 公衆衛生・疫学教授
27	岡田 美保子	川崎医療福祉大学 医療福祉マネジメント学部医療情報学科教授
28	○久保田 潔	東京大学大学院医学系研究科薬剤疫学講座教授
29	小出 大介	東京大学大学院医学系研究科臨床疫学システム学講座特任准教授
30	杉森 裕樹	大東文化大学大学院予防医学教授
31	中山 健夫	京都大学大学院医学研究科健康情報学分野教授
32	林 邦彦	群馬大学大学院保健学研究科教授
33	山口 拓洋	東北大学大学院医学系研究科医学統計学分野教授

34 (○：座長)

2. 研究実施計画書の作成

薬剤疫学研究の実施に先立ち、研究実施計画書を作成し、それに従って研究を実施する必要がある。以下に、研究実施計画書に記載すべき項目と記載方法を示す。項目の記載順序、計画書の体裁を示すものではないが、これらの項目が全て含まれるよう作成すること。なお、研究実施者が研究を計画及び実施する際に留意すべき事項については、3、4及び5を参照すること。

・計画書標題

標題は、研究の内容が端的に理解できるよう、注目する曝露やアウトカム、研究デザインを含めて記載することが望ましい。

・計画書更新履歴（作成・更新日、変更理由、変更内容）

計画書は、一度確定した後にも変更が必要となることがある。変更過程を記録するため、作成日の他に更新日、変更の理由と内容を記録する。

・用語の定義

計画書内で使用する重要な用語は定義を明確にし、用語は一貫性を持って使用する。

・目的

国際薬剤疫学会（International Society for Pharmacoepidemiology、以下「ISPE」という。）による Good Pharmacoepidemiology Practices ガイドライン⁽²⁾を参考に、次の3点を記載する。

1) 研究目的：研究から得ようとする知見を述べる。リサーチクエスションとしてどのような知見を得たいのかを疑問文として示してもよい。

2) 具体的な目標：研究において注目する曝露とアウトカム、及び評価されるべき仮説を記載する。

3) 論理的根拠：具体的な目標の達成がどのように研究目的の実現につながるのかを論理的に説明する。

・背景・先行研究

研究を実施する背景となった事項を述べ、なぜこの研究を行うのかを明確にする。また、参考とした先行研究があれば、その研究結果や方法に関して問題と考えられる点等、特記すべき事項を記載する。注目する医薬品の市販後の使用状況、同種同効薬の有無等についても記載する。

1 ・研究実施体制

2 研究を行う組織や個人は、その研究に対して全責任を負わなければならない。研究
3 を委託する組織や個人との関係、役割、及び責任を計画書に明記する。研究を委託す
4 る場合は、委託先を含め、研究に関与する全ての人員とその役割を記載する。

5
6 ・研究期間

7 研究計画作成開始から報告書作成終了までのおおよそのスケジュールについて、各
8 マイルストーン（研究計画書完成、研究実施、解析終了、報告書完成等）を含めて記
9 載する。

10
11 ・データソースの選択、データの入手方法

12 研究に使用する全てのデータソースを記載し、研究目的や仮説に照らして、そのデ
13 ータソースの使用がどのような点で適切であるかを説明する。

14 データベースの二次利用においては、そのデータが本来どのような目的で収集され、
15 データの特徴が研究においてどのような限界につながり得るかについても簡潔に述べ
16 ること。また、これまでの疫学研究への利用実績を知り得る範囲で記載する。更にデ
17 ータの入手方法も記載する。

18 既存のデータベースに加えて、研究のためにデータを収集する場合は、その収集方
19 法を記載する。データ収集のために症例報告票を作成する場合にはその様式を計画書
20 に添付する。

21
22 ・研究デザイン

23 研究目的や仮説に照らして、選択した研究デザインの使用が適切である理由を記載
24 する。研究デザインの例として、コホート研究、ケース・コントロール研究、ネステ
25 ッドケース・コントロール研究、ケースクロスオーバー研究、セルフコントロールド
26 ケースシリーズ研究等がある。薬剤疫学研究では、対照群を置いたデザインを用いる
27 ことが多いが、適切な同時比較対照群が無いなどの理由で同時比較対照群を設定しな
28 い場合には、その旨と理由を記載する。同時比較対照群に替えてヒストリカルコン
29 トロールを用いる場合には、そのヒストリカルコントロールが適切である理由を記載す
30 る。ケースクロスオーバー研究やセルフコントロールドケースシリーズ研究のデザイ
31 ンを用いる場合には、そのデザインを用いることがどのような限界につながり得るか
32 を記載する。

33 また、研究デザインに関し、「前向き」、「後向き」という用語が使用されることがあ
34 るが、特に統一された定義ではないため、使用は可能な限り避け、使用する必要がある
35 場合には必ず用語の定義を行った上で使用すること。

1 ・サンプルサイズ、検出力

2 研究実施に必要なサンプルサイズ、または使用するデータソースによって得られる
3 検出力を算出方法と併せて可能な限り記載する。

4
5 ・対象集団の定義

6 対象とする集団の定義を明確に記載する。対象集団は、通常、対象者の属性、場所、
7 時間によって、または適格基準と除外基準によって定義される。観察開始と打ち切り
8 の条件等も明示し、これらの定義を適切とする理由を述べること。また、コホート研
9 究においては観察開始前のアウトカムや共変量を測定する期間、観察開始後のアウト
10 カム発生を測定する期間、ケース・コントロール研究やネステッドケース・コントロ
11 ール研究においては曝露や共変量を測定する期間などを記載し、研究に使用するデー
12 タがこれらの期間に照らして適切であることを説明する。必要に応じて図を示して説
13 明するとよい。

14
15 ・曝露の定義

16 研究において注目する曝露の定義を明確に記載する。データベースから医薬品のデー
17 ータを抽出する場合には、データソース及び医薬品コードのリストを計画書に添付す
18 る。複数の曝露に注目する場合には、全ての曝露の定義を記載する。曝露期間の定義
19 には、処方日または調剤日、処方期間、次回処方（調剤）が前回の処方期間から推測
20 される日より遅れた時にこれを中断と扱うか継続と扱うかの基準、処方期間が重複し
21 た時の扱い、曝露期間の終了日等が含まれる。対象者を医薬品の新規使用開始者（以
22 下「新規使用者」という。）に限定する場合は、新規使用の定義も記載する。各定義に
23 ついてその定義が適切である理由を述べること。必要に応じて図を示して説明すると
24 よい。

25
26 ・アウトカムの定義

27 研究において注目するアウトカムの定義を明確に記載する。データベースからアウ
28 トカムのデータを抽出する場合には、データソース及びアウトカムの指標となる病名
29 コードやその他のコードをリストとして計画書に添付する。病名と治療薬または病名
30 と診療行為等を組み合わせてアウトカムを定義する場合はその組み合わせ方、複数の
31 アウトカムを定義する場合には全てのアウトカムの定義をそれぞれ記載する。初回発
32 生のアウトカムに限定して測定するか複数回の発生を測定するかについても記載する。
33 また、どの日付をアウトカム発生日とするかについても記載する。各定義についてそ
34 の定義が適切である理由を記載すること。

35 使用するアウトカム定義に対するバリデーションスタディについて予定も含めて実
36 施の有無と、実施する場合はそのスタディの計画等を参考資料として添付する。実施

1 しない場合は、その理由を記載する。また、既存のバリデーションスタディの結果を
2 利用してアウトカムを定義する場合には、そのバリデーションスタディの結果を利用
3 することが適切な理由を記載し、参考とした資料を計画書に添付する。

4 5 ・共変量の定義

6 曝露とアウトカム以外に測定する患者の基本情報や、交絡因子または効果修飾因子
7 として働く可能性のある因子を示し、各共変量の測定方法（曝露やアウトカムと同様
8 にコードで定義する場合にはコードリストと情報を抽出するデータソース等）を記載
9 する。また、使用するデータソースにおいて測定が困難であるが重要と考えられる共
10 変量が存在し、研究デザインや解析において対処する場合はその方法を記載する。

11 12 ・データマネジメント、統計解析

13 データクリーニング、データチェック、データセットの作成から最終的な解析に至
14 るまでの各段階で実施するデータの取扱いの内容と方法を記載すること。また、研究
15 目的に対応する結果を求める解析を主要解析として記載する。主要解析を補足するた
16 めに実施する解析は副次解析として記載する。主要解析の前に予備的な解析を行う場
17 合はその内容も記載する。予め計画した解析内容は全て計画書に記載することにより、
18 研究の途中で追加する解析とは明確に区別することが重要である。解析に使用するモ
19 デル、効果指標、共変量の扱い方等詳細を記述すること。解析に使用する統計ソフト
20 はバージョンも含めて記載する。

21 コホート研究における注目する曝露群と対照群とのマッチングや、ケース・コント
22 ロール研究におけるケースとコントロールのマッチング等を行う場合は、マッチング
23 因子とマッチング方法、及びそれらを使用する理由を記載する。マッチング方法は具
24 体的かつ詳細に手順を記載すること。また、横断研究（断面研究）等においてサン
25 プリングを行う場合はサンプリング方法を記載する。

26 解析結果の頑健性を確認するために感度解析を実施する場合は、その内容を記載す
27 る。特にデータベースを用いた薬剤疫学研究では、曝露、アウトカムまたは共変量の
28 定義等研究デザインにより解析結果が大きく変わる場合があるため、感度解析は非常
29 に重要である。予め計画した感度解析は全て記載し、研究の途中で追加する感度解析
30 とは区別できるようにしておくこと。

31 32 ・研究の限界

33 研究目的、仮説に対応する結果を得るために適切に研究を実施するべきであるが、
34 何らかの事情により適切な対処が困難で研究結果の妥当性が十分確保できないことが
35 予測される場合には、これを予想される研究の限界として記載する。特に研究の限界
36 が結論の内的妥当性及び一般化可能性（外的妥当性）にどの程度影響するかについて

1 想定可能な範囲で記載すること。

2
3 ・結果の公表の有無、公表方法

4 医薬品の安全性評価に影響する重要な結果は公表されるべきである。結果の公表を
5 予定する場合はその方法と公表先を記載する。

6
7 ・個人情報保護と倫理

8 個人情報の使用の有無を記載する。個人情報を取り扱う場合にはセキュリティ面の
9 対応方法について概要を示すこと。個人情報に関する項目が取り除かれたデータソー
10 スを使う場合にはその旨、研究の過程でデータソースから個人情報に関する項目を取
11 り除く処理がなされる場合はその主な処理内容を記載する。その他、倫理的に留意す
12 べき事項や対処すべき事項があれば記載する。また、疫学研究の倫理指針の対象とな
13 るか否か、及び倫理審査委員会に付議する場合には、委員会名及び委員会が設置され
14 ている組織を記載する。

15
16 ・利益相反、透明性の確保

17 研究に関与する全ての人員について、研究の妥当性に影響を与える可能性のある全
18 々の要因に関する利益相反状況を記載する。利益相反が存在する場合には、どのよう
19 に研究の透明性を確保するのか、その手段を明記すること。

20
21 ・参考文献

22 計画書の作成に際して参考とした文献を示す。

23
24
25 **3. データソース**

26 薬剤疫学研究の実施に当たっては、研究目的や仮説に対して適切なデータソースを選択
27 し、かつデータの特性を理解し適切に利用することが重要である。以下に主なデータソー
28 スと選択、標準コード体系、及びバリデーションに関する概要と留意事項を述べる。これ
29 らの留意事項等を検討した結果、従来型のデータ収集等による方法の方がより適切と判断
30 される場合には、データベースの二次利用以外の方法により薬剤疫学研究を行う。

31
32 **(1) 薬剤疫学研究に使用される主なデータソースと選択**

33 薬剤疫学研究に使用されるデータには、レセプトデータなど医療費償還のために発行さ
34 れる請求書データや HIS/EMR データ、レジストリ等があり、それぞれの研究において適し
35 たデータソースを用いるべきである。また、どのようなデータソースを使用する場合であ
36 っても、そのデータが生成された背景に関する情報（例えば、データが生成された医療機

1 関の所在地域、規模、診療科、データが得られた期間、データに含まれる集団の年齢や性
2 別の分布等の人口統計学的情報等)を入手し、データの特徴を確認すべきである。

3 以下に、各データソースが持つ主な特徴の観点からの、データソースの選択に関する留
4 意事項を示す。

5 ①請求書データ

6 レセプトデータは医療機関や薬局が診療報酬または調剤報酬請求のために作成するもの
7 であり、主に内科、歯科、調剤及びDPCの4種類が存在する。国内には研究に利用可能な
8 請求書データがいくつか存在し、解析等へ利用可能なデータセットとして提供する、また
9 はそのデータを用いて解析業務を請け負うなどのサービスがある。また、レセプトデータ
10 と特定健康診査・特定保健指導に関する情報を連結して解析を行うサービスも存在する。

11 これらの請求書データは複数の医療機関や薬局から収集されたデータの場合の他、健康保
12 険組合から提供されたデータの場合もあり、後者では組合員が受診した全ての医療保険診
13 療の請求情報が網羅される。いずれもデータの規模は小さいのに対し、厚生労働省が構築
14 しているナショナルレセプトデータベース(特定健康診査・特定保健指導に関する情報
15 を含む)は、ごく一部を除く国民の大半の医療保険診療の請求情報が集約されており大規模
16 なデータベースである。発生が稀な副作用や疾患を注目するアウトカムとする研究を行う
17 場合には、このような大規模データベースが必要となるが、現時点ではナショナルレセプ
18 トデータベースの利用は学術研究者等に限定されており、利用には申出や審査等に時間を
19 要し、研究実施者はセキュリティ体制を構築する必要がある等多くの制約がある。

20 請求書データが一箇所に集約されると、データが存在する期間内は全ての医療機関の診
21 療情報が得られることになるため、曝露からアウトカム発生までの網羅的な追跡が可能で
22 ある。このように追跡期間が確保される点は薬剤疫学研究に適している。海外、特に米国
23 においては民間の各保険会社が数千万人単位の大規模な請求書データを保有し、それら
24 を使った薬剤疫学研究が数多く実施されている。

25 ②HIS/EMRデータ

26 HIS/EMRデータは医療機関が保有するデータであり、レセプトデータよりも診療情報に関
27 するデータ項目数が多く情報は豊富であるが、医療機関毎にデータ項目やデータ形式が異
28 なることがあるため、複数施設のデータを統合して研究に使用する場合には、データ形式
29 の標準化が大きな課題となることが多い。データ項目やそこに入力されるデータ形式の違
30 い、傷病名、医薬品名及び臨床検査項目に使われるコードの違い等を明らかにし、違いが
31 あれば対処方法を検討すべきである。また、HIS/EMRデータは医療機関単位のデータであ
32 るため、外来患者については並行して受診している医療機関があっても、注目する曝露、
33 交絡因子、またはアウトカム発生等の診療情報は通常得られない。そこで、妥当な研究結
34 果を得るために、急性イベントの場合には追跡期間を入院期間に限定する等の対処方法が
35 必要となることもある。外来患者を対象に含める場合は、少なくとも受診期間、受診回数
36 等の情報から継続して受診している患者を特定し、研究対象集団を限定したり、その割合

1 を推定して研究結果に与える影響を十分検討したりするべきである。このように、HIS/EMR
2 データが適切なデータソースとなる疫学研究の種類や範囲には制限があるが、将来的に地
3 域医療連携の発展に伴い診療情報が医療機関間で共有されるようになれば、レセプトデー
4 タと同様に外来診療の期間を含めて対象者の追跡の正確性が向上することが期待され、薬
5 剤疫学研究における利用価値は格段に上昇する。なお、ナショナルレセプトデータベース
6 とは別に、厚生労働省とPMDAにより、複数の医療機関のHIS/EMRデータを標準化し統合し
7 て利用することの可能なネットワークの構築が進められている。

8 ③レジストリ

9 レジストリはある特定の情報の収集を目的に作成されるものであり、収集する情報の種
10 類で分けると、アウトカムレジストリ（がん登録等）、曝露レジストリ（製造販売後調査等）、
11 及び診療行為レジストリ（特定の手術に関連する情報の集積等）等が挙げられる。レジス
12 トリでは個々の目的に沿って特定のデータ項目が収集されるため、研究のために作成した
13 レジストリの一次利用においては、研究に必要なデータ項目のうち収集可能な情報は概ね
14 確保されている。既に作成されたレジストリを二次利用する場合には、レジストリの作成
15 目的と二次利用の研究目的が異なれば必要なデータ項目の欠如等により情報が不足する可
16 能性があり、追加の情報収集や他のデータとのリンケージ（後述）の検討が必要となるこ
17 とがある。

18 その他、研究目的や仮説によっては海外のデータベースを利用するという選択肢も存在
19 する。国内における医薬品の安全性に関する懸念について海外のデータベースを使用する
20 場合は、研究結果が我が国においても当てはまるかを慎重に検討する必要がある。国内と
21 海外の医療保険制度、医療行為（診断基準、医薬品の選択等）の違いや人種差、用いられ
22 るデータの収集方法やコーディング等を比較し、海外のデータベースを選択する明確な理
23 由がある場合は、研究結果に影響し得るこれら要因の検討結果と選択理由を計画書や報告
24 書に示すべきである。

25
26 上述の各データソースの基本的特徴の違い以外にもデータソース選択時には、研究内容
27 に関連するデータの1) 追跡期間、2) 時期、3) 対象者、及び4) データ項目などの検討が
28 必要である。1) 追跡期間は、その期間内に研究において注目する共変量やアウトカム発生
29 の測定が可能であることが明示される必要があり、目的に応じてこれらの条件を十分に満
30 たすデータを選択するべきである。例えば、注目するアウトカムが遅発性で曝露からアウ
31 トカム発生までの期間が長い場合は長期の追跡期間が必要である。2) 時期については、あ
32 る特定の時期を対象とした研究を行う場合は当然その期間を含むデータが必要である。3)
33 対象者は、使用するデータが目的とする集団を含むことを確認するべきである。例えば、
34 小児を対象とする場合は健康保険組合が有する請求書データが利用可能である。これに対
35 し、高齢者を対象とする場合、多くの労働者は定年退職に伴い健康保険組合を脱退するた
36 め、高齢者のデータは健康保険組合の請求書データにはごく僅かしか含まれておらず、研

1 究への利用は適当とは言えない。従って、HIS/EMR データまたはレジストリなど、健康保険
2 組合の請求書データ以外の利用を検討すべきである。4) データ項目として何を保有してい
3 るのか情報を入手し、対象集団、曝露、アウトカム等の定義に使用する情報がデータ項目
4 として存在しているかを確認する。例えば、調剤レセプトデータには傷病名が含まれてお
5 らず、また、DPC レセプトデータの傷病名には日付情報が含まれていないため、これらの項
6 目を必要とする場合には、他のデータソースの使用について検討が必要となるかもしれな
7 い。

8 我が国において、薬剤疫学研究に利用可能な医療情報のデータベースは徐々に増えてき
9 てはいるものの、選択肢はさほど多くはないのが現状である。従って、研究目的や仮説に
10 適した理想的なデータベースが常に存在する訳ではない。限られた選択肢の中で最も適切
11 なデータベースを選択し、そのデータベースの利用における研究の限界を計画書と報告書
12 に明記すべきである。なお、3の冒頭に示したように、データベースの選択に関して検
13 討を行った結果、一次データの収集等の方がデータベースの二次利用よりも適切と判断さ
14 れることもある。

15 単独のデータベースでは情報量が不足している場合、リンケージと呼ばれる複数のデー
16 タベースを連結して使用方法がある。例えば、請求書データの処方情報から曝露を特
17 定し、アウトカムを HIS/EMR データの臨床検査値、またはレジストリに登録された病名か
18 ら特定することにより、単独のデータソースを使用する場合よりも質の高い研究が実施可
19 能となることが期待される。ただし、リンケージされたデータはリンケージする前の個別
20 のデータベースに比べて情報量が増えることにより個人特定の可能性が高まることある
21 ため、リンケージの実施に際しては十分な配慮が必要である。リンケージを行う場合は、
22 対象データの方法及びその妥当性を計画書及び報告書に明記すること。

23 24 (2) 標準コード体系

25 データベースを用いた研究において曝露やアウトカムを特定する際には、通常これらの
26 情報がコード化されたデータを用いる。使用されているコード体系が、研究において特定
27 しようとする曝露やアウトカムを定義する上で適切であるかをデータソース選択の際に確
28 認すべきである。傷病名、医薬品、臨床検査について我が国で用いられている標準的な
29 コード体系は次のとおりである。

30 傷病名では、疾病及び関連保健問題の国際統計分類第 10 版（以下「ICD-10」という。）
31 コードとそれに対応付けが可能なより詳細な傷病名である標準病名（コードとして一般的
32 に交換用コードが用いられる）が存在し、HIS/EMR データではこれらのコードを有すること
33 が多い。また、レセプトデータはレセプト電算処理システム（以下「レセ電」という。）傷
34 病名マスタが用いられ、これは標準病名と対応付けされている。これらのコードとは別に、
35 主に副作用、有害事象を表す MedDRA コードがあり、製薬企業等が行政に副作用、有害事象
36 の症例を報告する場合などに用いられている。

1 医薬品に関しては、HIS/EMR データでは薬価基準収載医薬品コードまたは個別医薬品コー
2 ド（YJ コード）が付与されていることが多く、医療機関によっては H0T コードも付与され
3 ている。レセプトデータはレセ電医薬品マスタが用いられ、このレセ電医薬品コード、薬
4 価基準収載医薬品コード、および個別医薬品コードは H0T コードにより対応づけられてい
5 るため、いずれかのコードが付与されていれば、他のコード体系への変換が可能である。
6 また、世界保健機関（World Health Organization、（以下「WHO」という。）が管理する解
7 剖治療化学分類法（Anatomical Therapeutic Chemical Classification System、以下「ATC」
8 という。）コードについては個別医薬品コードとの対応付けが進められている。ATC コード、
9 個別医薬品コードとも医薬品の薬効別分類を定義するのに有用なコードである。

10 臨床検査項目のコードでは、臨床検査学会が管理する JLAC10 があるが、あまり普及は進
11 んでおらず、HIS/EMR データに JLAC10 を使用している医療機関は現時点では少数に留まっ
12 ている。また、レセプトデータでは診療行為及び特定器材にもレセ電マスタが整備されて
13 おり、これらのコードも利用可能である。

14 これらのコードはバージョンの更新により、コード自体やそれに対応する名称が変更さ
15 れることがあるため、研究に使用するデータで用いられているコードマスタのバージョン
16 を確認する必要がある。曝露やアウトカム等の定義として用意するコードリストは、使用
17 するデータソースと同じバージョンのマスタを用いて作成すること。

20 (3) バリデーション

21 バリデーションとは、ある方法が正しいことを確認する過程を指す⁽³⁾。医療情報のデー
22 タを二次利用する場合に、研究に用いる曝露やアウトカム等の定義の正確性を評価するた
23 めに、バリデーションスタディを実施することを推奨する。特に請求書データの傷病名コ
24 ードのみでアウトカムを定義する場合には、バリデーションスタディが必要となることが
25 ある。

26 バリデーションスタディは、データベースに適用される曝露やアウトカムなどの予め設
27 定した定義について、至適標準（gold standard）とする情報源（信頼性が確立されている
28 レジストリのデータや診療録等を適切な手順で評価した結果）と照らし合わせて、その定
29 義がどの程度正確であるのかを評価するものである。至適標準に診療録を用いる場合の手
30 順として、例えば、あるデータベースより曝露またはアウトカム定義に該当する対象者等
31 必要な対象者を全て特定し、その対象者の診療録等の調査によりデータベース上の定義が
32 曝露またはアウトカムを正しく特定しているか等を判定する。アウトカムの判定基準は、
33 関連する臨床ガイドライン等に定められた標準的な診断基準があればそれを参考に作成す
34 る。関連する臨床ガイドライン等がない場合であっても、可能な限り一般的に確立された
35 診断基準を参考にした判定基準を設定することが望ましい。アウトカムの判定は判定基準
36 に従って行われるが、臨床的知識が必要とされるため臨床医が判定者となることが多い。

1 また、判定は、結果の客観性を担保するために複数の判定者によって実施され、判定結果
2 のバラツキを抑えるために少数の判定者によって実施されることが望ましい。なお、複数
3 の医療機関に渡って診療録を確認する必要がある場合に、海外では特別にトレーニングを
4 受けた者が、各医療機関の診療録から予め様式を定めた用紙に、該当症例の判定に必要な
5 情報のみを抽出する方法をとることがある。国内では実績が乏しいが、このような方法に
6 より少数の判定者により複数の医療機関の判定を実施することが可能となる。

7 判定により曝露またはアウトカムが判定基準に該当する場合を真とする。データベース
8 から定義に該当するとして抽出された対象者に対する真と判定された対象者の割合は陽性
9 的中度¹と呼ばれ、妥当性の指標値の一つである。その他の指標値には陰性的中度、感度、
10 特異度があり、感度、特異度はバリデーションスタディの対象とした集団の特性の構成に
11 影響されない普遍的な指標値であるため推定することが望ましい。しかし、推定のためには
12 信頼性の高い疾患レジストリなど対象集団の一部または全部に関する至適標準による判
13 定がなされたデータソースが必要であり、このようなデータソースを用意することは困難
14 であることが多い。従って、実施可能性の観点から多くの場合陽性的中度しか求められな
15 い。

16 バリデーションスタディは、本来、研究毎に特定のデータベースと定義に対して実施さ
17 れるものであり、陽性的中度もこれらに特異的な指標値である。しかし、研究に使用した
18 データの対象者について診療録等の情報を入手し妥当性の判定を行うことができない場合、
19 他のデータベースで同様の定義に関して実施したバリデーションスタディの結果を参考に
20 することもある。

21 また、バリデーションスタディとは異なる方法であるが、データベースに含まれる情報
22 を用いて適切に曝露またはアウトカムを定義することができない場合、または妥当性の高
23 い研究結果を得ようとする場合には、診療録等のレビューにより正確に曝露またはアウト
24 カムを特定するべきである。

27 4. 研究デザインと研究の実施

28 薬剤疫学研究に用いられる主な研究デザインの概要、研究デザイン方法に関する留意事
29 項、その他薬剤疫学研究実施時に留意すべき事項を記載する。

31 (1) 研究デザイン

32 研究デザインは数多くあり、研究の目的や仮説毎に最も適切なものを選択するべきであ
33 る。以下に薬剤疫学研究で用いられる主な研究デザインを示すが、これらの中でも対照群
34 を置くデザインが比較的良好に使用される。

¹ 陽性反応的中率、陽性適中度等様々な呼び方があるが、本ガイドラインでは「疫学辞典 第5版」に倣った

ここに挙げたデザインは参考として示すものである。また、各デザインの詳細については専門書を参照すること。

横断研究（断面研究）

ある一時点の対象集団についてデータを収集する研究デザインで、疾患の有病割合や医薬品の使用/処方実態の調査などに用いられる。また、複数回に分けて横断研究を行うことで経時的な傾向を分析することができ、安全対策措置（例えば、添付文書への警告欄の新設）前後における処方動向の調査等による安全対策措置の影響の評価に用いられる。横断研究は、対照群を設定可能であるが、曝露とアウトカムの時間的前後関係が不明、有病割合の変化が疾患の発生と罹病期間のいずれの変化によるかが不明などの理由から曝露とアウトカムの因果関係を推定することは困難である。

アクティブサーベイランス

日米 EU 医薬品規制調和国際会議（International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for human use、以下「ICH」という。）E2E ガイドライン⁽⁴⁾によれば、アクティブサーベイランスは「予め計画された継続的なプロセスを介して有害事象の発現数を完全に把握しようとする手法である」と定義される。例えば、特定の薬剤投与を受けた患者のリスク管理プログラムによる追跡がある。これは、この薬剤の処方を受ける患者に、簡単な記入様式への記入及びその後調査者より患者へ連絡をとる許可を依頼するものである。

FDA は、市販後の医薬品についてほぼリアルタイムで安全性を監視するシステムとして Sentinel System の構築を進めており、これは、医療情報のデータを二次利用した新たな形式のアクティブサーベイランスである。例えば、承認後間もない新薬について懸念される有害事象等に関して、可能な限り速やかに且つある程度妥当性を持つような安全性シグナルを特定することなどを目指している。

ケースシリーズ（症例集積）

ある共通のアウトカムを持つ症例を収集するデザインであり、曝露と無関係に特定のアウトカムを持つ症例の集積、特定の曝露を受けてかつ特定のアウトカムを持つ症例の集積のいずれも含まれる⁽⁵⁾。後者の場合は曝露とアウトカムとの関連性に関する仮説の生成につながり得るが、通常、絶対リスクを推定することはできない。

セルフコントロールドケースシリーズ（SCCS）、ケースクロスオーバー（CCO）

両者とも対照群を設定せず自己を対照とするデザインであり、時間経過により変動しない個人内の交絡因子の影響を排除できること、基本的にケースのみを対象とするため必要な対象者数が少なく効率性が高いなどの利点がある。ただし、時間経過とともに変動する交絡因子の影響は適切に調整しない限り排除できない、曝露後アウトカ

1 ムが発生するリスクのある期間として設定するリスク期間が解析結果に大きな影響を
2 与える、比較的急性のイベントに用いられることが多く、因果の逆転（reverse
3 causality）が起りやすい等の問題もあり、使用に当たってはこれらのデザインを使用
4 することの妥当性やこれらのデザインを適切に使用するために必要条件などを十分
5 検討する必要がある。

6 SCCS では、効果指標（作用指標）として発生率比が求められる。SCCS は本来ワクチン
7 と急性イベントの関連性の評価に使用されてきたデザインであるが、最近は他の医
8 薬品への応用が検討されている。

9 CC0 では、効果指標としてオッズ比が求められる。CC0 は処方傾向の時間的変動の影響
10 を受けやすいため、コントロール期間をケース期間の前だけに設定する通常の CC0
11 とは異なり、ケース期間の後にも設定するデザインや case-time-control デザインと
12 いったデザインが使用されることもある。Case-time-control デザインは、処方傾向の
13 時間的変動を調整する目的で、ケースが発生した集団と同じ集団から時間でマッチン
14 グした非ケース群（コントロール）を特定し、調整オッズ比を求めるデザインである⁽⁶⁾。

15 16 **コホート研究**

17 コホートとは、「ある期間追跡される、特定の目的のために選ばれた個人の集まり」
18 と幅広く定義され、コホート研究とは、一つないしそれ以上のコホート内で起こる疾
19 病の発生を測定するもので、全ての疫学研究の原型である⁽⁷⁾。コホート研究では単一
20 のコホートで複数のアウトカムを測定することが可能であるが、アウトカムが稀な場
21 合は非常に大きなコホートサイズが必要となる。対照群を置く場合は、効果指標とし
22 てリスク差²、リスク比³、または発生率差、発生率比等を用いる。対照群を設定しない
23 場合は発生割合や発生率を測定することはできるが、効果指標を求めることはできな
24 い。

25 シグナル検証を目的として注目する医薬品と注目するアウトカムの関連性を研究す
26 る場合には適切に評価するために対照群を設定すべきである。また、背景発現率の高
27 いアウトカムと曝露、または医薬品の作用が小さな曝露とアウトカムの関連性を推定
28 する場合には基本的に対照群を設定する必要性が生じる。注目する曝露群と対照群は、
29 可能な限り年代、性別、疾病、併用治療等が似たような集団とする。適切な同時比較
30 対照群がない場合であっても、過去の当該曝露を受けていない集団に関する比較可能

² リスク差の代わりに寄与リスク（寄与危険度）（attributable risk）という語が用いられることもあるが、寄与リスクという語は、比較対照となる2群間でアウトカム発生に影響する曝露以外の因子が同等であることが含意されているほか、リスク差を曝露群における発生割合で除した attributable fraction やその他の異なる指標を意味する語として用いられることもある。

³ リスク比と発生率比の両方を包含する「相対リスク（危険度）」という用語が用いられることがある。また、症例対照研究におけるオッズ比がリスク比または発生率比と解釈できる場合に、オッズ比に「相対リスク」という用語が使用されることもある

1 かつ信頼に足るデータが存在する場合は、そのデータ（ヒストリカルコントロール）
2 と比較することも可能である。

3 コホート研究は、従来型のデータ収集方式ではデータ収集に労力を要する研究デザ
4 インであるが、大規模なデータベースを利用する場合はデータ収集効率はさほど問題
5 にならないため、研究デザインの基本形であるコホート研究が選択されることが多い。

7 **ケース・コントロール（症例対照）研究、ネステッドケース・コントロール（コホー 8 ト内症例対照）研究**

9 コントロールを標本抽出することで、より効率的にコホート研究と同じ目標へ到達
10 することを目指す研究デザインである⁽⁷⁾。単一のアウトカムと複数の曝露の関係を推
11 定することが可能であり、効果指標としてオッズ比が求められる。

12 一般にコントロールは、ケースが発生したソース集団から選択されるべきである。
13 ソース集団を明確に定義できない場合には、適切なコントロールを選択するための方
14 法を十分に検討するべきである。最初にコホートを設定し、そのコホート内でケース・
15 コントロール研究を行うネステッドケース・コントロールデザインは、ソース集団か
16 ら確実にコントロールを選択できる優れた方法である。

17 従来型のデータ収集方式では、本デザインはコホート研究より収集するデータ量が
18 少なく済むことから研究効率が良く比較的良好に用いられてきた方法である。大規模
19 なデータベースが利用可能な場合は、コホートを設定しネステッドケース・コント
20 ロール研究を実施することもある。

22 **（2）研究デザインに関する留意事項**

23 **対象集団の定義**

24 対象集団を定義する場合、対象集団に含まれる各個人の追跡期間を使用するデー
25 タの中で確認する必要がある。追跡期間の定義は、例えば請求書データでは、データ収
26 集開始日からデータ収集終了日までなどであり、HIS/EMR では、入院日から退院日まで
27 等である。請求書データ等において、追跡期間の定義に用いるこれらの情報が得られ
28 ない場合には、追跡期間をどのように定義するのが適切かを検討する必要がある。

29 対象集団は、研究目的・仮説に適切に対応する集団である必要があり、そのような
30 集団が構成されるように適格基準と除外基準を設定する。研究実施時にはデータにこ
31 れらの基準を適用して対象集団を特定した際、特に除外基準については除外した人の
32 情報を確認し、研究目的、仮説に照らして不適切な除外をしていないか検討するべき
33 である。また、観察の打ち切りが特定の曝露やアウトカムの発生と関連して生じてい
34 ないことの確認も必要である。

35 観察単位は人時間、人数、または処方件数など適切な単位を設定する。

1 曝露の定義

2 医薬品の処方に関する情報は、データソースにより情報の粒度や含まれる医薬品の
3 種類が異なるため、これらの点を十分に考慮して曝露を定義する。例えば、使用され
4 ている医薬品のコード体系、処方について日付までと月までのいずれが含まれている
5 か、院外処方または院内処方の医薬品が含まれているか等を考慮して定義の検討を行
6 う。注目する曝露及び対照群の曝露とする医薬品コードをリスト化し、複数の剤形を
7 持つ医薬品の場合は曝露に含める剤形も明確にする。

8 通常、医療情報のデータベースを二次利用する場合には、処方及び調剤情報は得ら
9 れるが、医薬品の使用に関する情報は得られないことが多い。従って、実際の医薬品
10 の使用状況を処方及び調剤情報から類推して曝露期間を定義する必要がある。曝露期
11 間に関して、曝露開始日は処方日か調剤日か、処方と処方の間隔 (gap) を曝露が継続
12 した期間とみなすか、処方終了後の猶予期間 (grace period) を設けるか、曝露期間
13 終了の条件は何か等を設定する必要がある。また、抗がん剤の化学療法のように一定
14 周期を単位とする処方や頓服処方の場合、曝露期間をどのように設定するのが最も適
15 切であるのか検討する必要がある。処方用量に関する情報が必要な場合は、測定方法、
16 一日用量や累積用量などの算出方法を検討する必要がある。また、医薬品の新規使用
17 者に限定する必要性について検討し、必要な場合は新規使用者と定義するための条件
18 (例えば、処方開始前6ヶ月間に当該医薬品の処方がないこと)を設定する。

19 アウトカムの定義

20 データの二次利用の場合、データに含まれる情報の中から工夫して妥当性の高いア
21 ウトカム定義を設定することが必須である。アウトカムに関する傷病名のみではなく、
22 そのアウトカムの診断や治療に使用される医薬品や臨床検査、処置などと組み合わせ
23 て定義することにより、妥当性の高いアウトカム定義とすることが可能か検討する。
24 アウトカムの定義に使用する傷病名、医薬品名、臨床検査名または診療行為名等のコ
25 ードをリスト化する。また、アウトカム発生日の定義も必要であり、特に複数の情報
26 を組み合わせてアウトカムを定義する場合には、どの情報に関する日付を発生日とす
27 るか検討する。HIS/EMR データのように臨床検査値が含まれる場合は、アウトカム定義
28 に臨床検査値の利用を検討することが推奨される。例えば、アウトカムが白血球減少
29 や血小板減少の場合は、臨床検査値にアウトカムとみなす基準値を設定することによ
30 り妥当性の高い定義となることが期待される。ただし、曝露開始前の検査値や曝露か
31 ら検査までの時間間隔についても併せて検討が必要である。また、複数の医療機関の
32 データを用いて研究を行う場合には、医療機関毎に検査値の基準範囲が異なる点に留
33 意してアウトカムを定義する。

34
35 アウトカムの定義に傷病名コードを用いる場合、データソースにより傷病名に関す
36 る日付が付与されるタイミングが異なる場合があることも留意が必要である。例えば、

1 同一症例の請求書データの診療開始日と、HIS/EMR データの診断日が一致しないことが
2 ある。また、アウトカム発生までの潜伏期間 (latent period) や初回に発生したアウ
3 トカムに限定するか複数回発生したアウトカムを全て測定するか等についても検討す
4 る。

5 アウトカム定義の妥当性を確認するためには、バリデーションスタディを実施する、
6 または既存のバリデーションスタディの結果を参考とする方法がある。バリデーショ
7 ンスタディの詳細は、3の(3)を参照のこと。また、シグナル検証を目的とする場合
8 等、妥当性の高い研究結果が必要な場合や適切なアウトカム定義が存在しない場合は、
9 データベースから診療録等に戻り、診療録等を確認して妥当なケースを特定すること
10 が適切と判断される場合もある。

11 12 **バイアス・交絡への対処**

13 疫学研究における結果や推定の誤差には大きく分けて偶然誤差と系統誤差の二種類
14 がある。前者は偶然発生する誤差のことで系統誤差では説明できないデータのばらつ
15 きであり、理論上は、対象集団のサイズを無限大にすると偶然誤差による平均値のバ
16 ラツキはゼロにまで減らすことができる。後者はバイアスとも呼ばれ真実からの系統
17 的なずれであり、対象集団のサイズが無限大になっても平均値のずれの度合いは変わ
18 らない⁽⁷⁾。バイアスは多くの種類に細分化されることもあるが、大別すると選択バ
19 イアス、情報バイアス、交絡の三種類がある。各バイアスの詳細については、専門書
20 を参考のこと。

21 交絡のうち、薬剤疫学研究で特に問題となるのが「適応による交絡」であり、「処方
22 理由による交絡」とも呼ばれる。ある医薬品使用者と非使用者を比較する場合には、
23 使用者群における該当医薬品の適応症を持つ割合は、当然非使用者群より高い。同一
24 の適応症を持つ二つの医薬品を比較する場合でも、それぞれの医薬品が異なる重症度
25 や併存疾患をもつ患者に対して処方されるのであれば、二群間で重症度や併存疾患の
26 分布の違いが生じ得る。適応症の有無またはその重症度や併存疾患の有無が、研究で
27 注目するアウトカムの発生に影響を与える場合、アウトカムの発生は曝露の有無また
28 は曝露の違いとともに、適応症とその重症度及び併存疾患の違いのいずれからも影響
29 を受ける。適応による交絡とは、医薬品そのものによる効果とそれ以外の因子による
30 効果、つまり適応による効果の混合、または混合した結果を指し、適応による交絡を
31 適切に調整しないと効果指標を正しく推定することができない⁽⁸⁾。

32 一方、ある曝露とアウトカム以外の因子のレベルによって曝露のアウトカムに対す
33 る効果が異なるのであれば、その因子は効果修飾因子と呼ばれる。例えば、男性と女
34 性である医薬品による副作用の発現の程度が異なるのであれば性は効果修飾因子であ
35 る。なお、効果修飾は用いる効果指標によって異なることから、効果修飾ではなく、
36 効果指標の修飾と呼ばれることもある。

1 研究計画段階から考え得る全ての交絡について、どのように対処するのかを十分検
2 討する必要がある。また、効果修飾因子の有無を調査し、重要な効果修飾因子がある
3 場合はその影響の推定を含めた解析計画を作成する。

4 交絡因子等の共変量についても曝露やアウトカム定義と同様にコードリストを作成
5 し、特定すべき因子を設定する。通常、共変量は、注目する曝露開始以前に設定した
6 期間において測定する。

7 8 **交絡因子の調整**

9 解析において多くの交絡因子を同時に調整するには、通常各因子を共変量として重
10 回帰モデルに含めるが、アウトカムの発生数に対して交絡因子が多過ぎると解析に時
11 間を要し、また効果指標が正しく推定できないことが知られている⁽⁹⁾。対応策として、
12 複数の共変量を要約したスコアを作成し、そのスコアを重回帰モデルに含めることで
13 共変量の数を減らすことが可能である。複数の共変量から曝露を予測する要約スコア
14 がプロペンシティスコア（以下「PS」という。）であり、アウトカムの発生を予測する
15 要約スコアが Disease Risk Score（以下「DRS」という。）である。

16 17 **① プロペンシティスコア**

18 PS は曝露を受ける確率を予測する共変量を要約したスコアであり、各個人について
19 アウトカムとは無関係に複数の共変量から推定するものである⁽¹⁰⁾。注目する曝露（医
20 薬品 A とする。）と比較対照の曝露（医薬品 B とする。）とがある場合には、実際に処
21 方された医薬品が A であるか B であるかは独立に推定された各個人の医薬品 A の処
22 方を受ける確率が PS である。ある個人の共変量から求められる PS が 1 に近ければ、
23 実際には医薬品 B が処方されていても医薬品 A が処方される確率が高いことになり、
24 逆に PS が 0 に近ければ実際には医薬品 A が処方されていても医薬品 A が処方される確
25 率は低いと評価される。個人毎に求められた PS は、マッチング、層別化または回帰モ
26 デルに用いられる。

27 PS は 1983 年に初めて提唱され⁽¹¹⁾、近年では多くの研究に採用されており、特にデ
28 ータベースを用いたコホートデザインによる研究では、二群間の共変量のバランスを
29 取る標準的方法として確立されている。しかし、観測される共変量しか調整ができな
30 いこと、データソースの規模が小さい場合には正確な PS が得られない可能性があるこ
31 となどに留意すべきである。PS を推定するための変数選択に関しては様々な意見があ
32 るが、最低限、治療法の選択に関係しかつアウトカムに影響を及ぼす全ての因子を可
33 可能な限り測定して PS を推定すべきである。

34 35 **② Disease Risk Score**

36 DRS は、広くは次項に記載する Charlson Comorbidity Index（以下「CCI」という。）

1 など、疾患のアウトカムや予後を予測する、複数の因子から推定されるスコアである。
2 薬剤疫学において DRS は通常、単にアウトカム発生を予測するだけでなく、交絡調
3 整を目的として当該研究における非曝露群のデータから推定される要約スコアを意味
4 する。⁽¹²⁾これが交絡因子の要約スコアとして層別化や回帰モデルに使用されており、
5 心血管疾患をアウトカムとする研究などで利用例が見られる⁽¹³⁾。薬剤疫学研究におい
6 ては、PS と比較して DRS が使用される機会は圧倒的に少ない。しかし、新薬の販売開
7 始直後など医薬品の処方動向が急速に変化する場合は、PS よりもこの DRS で交絡因子
8 の調整を実施することが適切な場合があるとの指摘や、DRS を効果修飾因子として利用
9 する方法など、近年、その有用性が見直されてきている⁽¹⁴⁾⁽¹⁵⁾。

11 ③ Charlson Comorbidity Index

12 疾患のアウトカム発生の予測を直接の目的とする要約スコアも存在する。例えば、
13 CCI 推定の疾患リストに含まれるがん、心血管疾患、糖尿病など特定の疾患に対応
14 する点数を合計し、死亡リスクを推定するスコアとして CCI が知られている⁽¹⁶⁾。も
15 ともとは診療録の情報を使って作成されたものであるが、それをデータベース研究
16 で利用するため、CCI 推定の疾患リストを ICD-9-CM や ICD-10 に変換した方法が考
17 案されている⁽¹⁷⁾⁽¹⁸⁾。このスコアを共変量に用いた研究は海外で散見されるが、我
18 が国とは診断基準等医療の背景が異なる可能性があるため、我が国の医療情報のデ
19 ータを用いて CCI を推定する場合は、CCI 推定方法の妥当性を確認する必要がある。

21 (3) データマネジメント、解析の実施

22 解析前に実施したデータのチェックやクリーニング等のデータマネジメントについて文
23 書に記録する。

24 データベース研究では曝露やアウトカムの定義等の研究デザインの設定が解析結果に大
25 きく影響することがある。従って、解析結果の頑健性を担保するため、感度解析の実施が
26 重要である。

27 また、計画書に研究開始前に定めた解析と計画書確定後に追加した解析は、明確に区別
28 して行うことが重要である。解析の再現性を担保するため、解析過程が分かるように文書
29 化し、解析プログラムは、第三者が理解しやすいようにコメント等を挿入して保存するべ
30 きである。

32 (4) 品質の確保

33 研究に使用するデータに関し、情報セキュリティ対策やデータへのアクセス記録の作成
34 等によりデータの正確性を保証する。また、研究過程で作成した文書やプログラムについ
35 て品質の確認を行う。

1 (5) 個人情報保護と倫理

2 実施する研究毎に、疫学研究に関する倫理指針や関連する法規に照らし合わせ、必要に応
3 じて指針に従って研究を実施する。また、研究に使用するデータの取扱いについて、情報
4 漏洩防止等のセキュリティ面への対応策について研究組織外の者に対して説明可能な体制
5 を整えておくべきである。

6 7 8 5. 研究結果報告書の作成

9 報告書の提出が必要な場合には、以下に示す項目が全て含まれるよう報告書を作成する
10 こと。ただし、項目の記載順等報告書の体裁を示すものではない。また、報告書が必要と
11 されない場合には、体裁は自由であるが、実施した研究の再現性が担保されるように、以
12 下に示す項目について記録を残しておくべきである。

13 報告書には、計画書に記載した目的、方法等を全て記載する。研究途中で計画書を更新
14 した場合は最終版の内容を報告書に反映し、加えて結果と考察等を記載する。

15 16 ・ 報告書標題

17 計画書標題と対応を付けやすいよう一貫性のある標題とすることが望ましい。

18 19 ・ 報告書更新履歴

20 計画書と同様、変更過程を記録する。

21 22 ・ 用語の定義

23 計画書に加えて報告書で使用する用語を追加し、同様に記載する。

24 25 ・ 目的

26 計画書と同様に記載する。追記すべき点があれば記載する。

27 28 ・ 背景、先行研究

29 計画書と同様に記載する。追記すべき点があれば記載する。

30 31 ・ 研究実施体制

32 計画書と同様に記載する。追記すべき点があれば記載する。

33 34 ・ 研究期間

35 実際に研究に要した期間を記載する。

36

- 1 ・データソースの選択、データの入手方法
2 計画書の内容に加えて、実際のデータ入手方法等追記すべき点があれば記載する。
3
- 4 ・研究デザイン
5 計画書と同様に記載する。追記すべき点があれば記載する。
6
- 7 ・サンプルサイズ、検出力
8 計画書と同様に記載する。追記すべき点があれば記載する。
9
- 10 ・対象集団の定義
11 計画書と同様に記載する。追記すべき点があれば記載する。
12
- 13 ・曝露の定義
14 計画書と同様に記載する。追記すべき点があれば記載する。
15
- 16 ・アウトカムの定義
17 計画書と同様に記載する。追記すべき点があれば記載する。
18
- 19 ・共変量の定義
20 計画書と同様に記載する。追記すべき点があれば記載する。
21
- 22 ・データマネジメント、統計解析
23 計画書と同様に記載する。データマネジメント過程で欠測値の補完やデータ変換（コ
24 ード化、カテゴリ化等）を実施した場合は、その処理内容を記載する。また測定した
25 変数のうち欠測値の割合が大きいもの等、特記すべき事項があれば記載する。計画書
26 に記載した解析以外に、交絡因子の調整、交互作用の検討、または感度解析の追加等
27 により計画書確定後に追加した解析は追加解析として明確に区別して記載する。また、
28 追加した目的も記載すること。
29
- 30 ・結果の公表の有無、公表方法
31 報告書作成時点における公表予定について計画書と同様に記載する。
32
- 33 ・個人情報保護と倫理
34 個人情報取扱いの方法とその他倫理的に対処した事項があればその点を記載する。
35 倫理審査委員会に付議した場合は、委員会名、委員会が設置されている組織名、審査
36 日、審査結果、及び承認番号を記載する。

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36

- ・利益相反、透明性の確保

計画書と同様に記載する。追記すべき点があれば記載する。

- ・研究結果

対象集団の人数、適格基準及び除外基準の該当人数、注目する曝露群と対照群のサンプルサイズ、またはケースとコントロールの人数等を示す。ソース集団から解析対象集団が特定されるまでの各段階についてフロー図を用いて示すと分かりやすい。

注目する曝露群と対照群、ケースとコントロール等解析の対象とした集団の特徴が分かるよう、測定した基本的な因子（人口統計学的、臨床的、社会学的因子等）や共変量等の分布を示す。また、コホート研究においてマッチングを行った場合には、マッチング前後の分布を提示し、マッチングにより両群の共変量のバランスがどの程度揃ったのかを示すこと。

コホート研究では期間（追跡期間、曝露期間等）の分布やアウトカム発生数、ケース・コントロール研究では曝露の頻度等を示す。モデルによる解析結果を示す前に、粗解析の結果や主要な因子により層化した解析結果を表や図で示す。また、推定した効果指標について 95%信頼区間も示すこと。

主要解析、副次解析、感度解析の結果は明確に区別して示す。計画書に示した解析以外に追加の解析を実施した場合は、追加解析であることを明確化して示すこと。

解析に使用した最終モデルの結果のみを示すのではなく、変数選択の過程も詳細に示す。候補とした全ての変数、変数選択の方法、最終的に選択された変数及び選択されなかった変数を記載する。PS や DRS 等の要約スコアを用いた場合は、これらのスコアの推定に用いた変数のリストとスコアの推定方法も記載する。

- ・考察

研究目的に照らして、得られた結果はどのように解釈されるのかを記載する。文章の構成は自由であり、特に指定するものではないが、例として、1) 主要な研究結果の概要、2) 先行研究の結果との比較、3) 研究結果についての解釈（バイアスや未測定の影響に関する考察を含む）、4) 研究結果の一般化可能性を含む研究の限界、5) 結論の順に記載することが挙げられる。なお、ここに示した内容に限らず必要な点は全て記載する。

- ・参考文献

研究において参考とした文献等を記載する。

1 **6. 研究結果の公表**

2 基本的には研究結果は公益に資するべきであり、また、薬剤疫学研究に対する理解を得
3 るためにも、研究実施者は一般国民に対して研究結果を説明することが望ましい。また、
4 研究内容を学術雑誌に公表する場合は、報告書の内容を適切に要約し報告書と雑誌の内容
5 に齟齬の無いようにするべきである。また、研究目的、仮説に対して否定的な研究結果が
6 得られた場合であっても、公表バイアスを防ぐため可能な限り公表することが推奨される。

7
8
9

7. 参考文献

1. WHO. The UPPSALA MONITERING CENTRE. (オンライン) (引用日: 2013 年 3 月 11 日)
<http://www.who-umc.org/DynPage.aspx?id=115092&mn1=7347&mn2=7252&mn3=7613&mn4=7614>.
2. *Guidelines for Good Pharmacoepidemiology Practices*. ISPE. 2007.
3. Miquel Porta 編 日本疫学会訳. *疫学辞典 第5版*.
4. *E2E Pharmacovibilance Planning*. ICH. 2004.
5. *Distinguishing case series from cohort studies*. O. M. Dekkers, M. Egger, D. G. Altman, J. P. Vandenbroucke. s.l. : *Annals of International Medicine*, 2012, Vol. 156, pp. 37-40.
6. *When should case-only designs be used for safety monitoring of medical products?* M. Maclure, B. Fireman, J. C. Nelson, W. Hua, A. Shoaibi, A. Paredes, D. Madigan. S1, s.l.: *Pharmacoepidemiology and drug safety*, 2012, Vol. 21, pp. 50-61.
7. Rothman 著 矢野栄二・橋本英樹監訳, K J. *ロスマンの疫学*. s.l. : 株式会社篠原出版新社, 2009.
8. K. J. Rothman, S. Greenland, T. L. Lash. *Modern Epidemiology Third edition*.
9. *A simulation study of the number of events per variable in logistic regression analysis*. P. Peduzzi, J. Concato, E. Kemper, et al. s.l. : *Journal of Clinical Epidemiology*, 1996, Vol. 49, pp. 1373-1379.
10. *Estimating Causal Effects from Large Data Sets Using Propensity Scores*. Rubin, D. B. s.l.: *Annual International Medicine*, 1997, Vol. 127, pp. 757-763.
11. *The central role of the propensity score in observational studies for causal effects*. P R Rosenbaum, D B Rubin. 8S, s.l. : *Annals of International Medicine*, 1997, Vol. 127, pp. 757-763.
12. *Stratification by a Multivariate Confounder Score*. Miettinen, O. S. 6, s.l. : *American Journal of Epidemiology*, 1976, Vol. 104, pp. 609-620.
13. *Azithromycin and the Risk of Cardiovascular Death*. W. A. Ray, K. T. Murray, K. Hall. 2012, Vol. 366, pp. 1881-90.
14. *Use of disease risk scores in Pharmacoepidemiologic studies*. P. G. Arbogast, W. A. Ray. s.l.: *Statistical Methods in Medical Research*, 2009, Vol. 18, pp. 67-80.
15. *Role of disease risk scores in comparative effectiveness research with*. R. J. Glynn, J. J. Gagne, S. Schneeweiss. S2, 2012, Vol. 21, pp. 138-147.
16. *A new method of classifying prognostic comorbidity in longitudinal studies: development and validation*. M. E. Charlson, P. Pompei, K. L. Ales, C. R. MacKenzie.

- 1 5, s.l.: Journal of Chronic Diseases, 1987, Vol. 40, pp. 373-383.
- 2 17. *Adapting a clinical comorbidity index for use with ICD-9-CM administrative*
3 *database.* R. A. Deyo, D. C. Cherkin, M. A. Ciol. 6, s.l. : Journal of Clinical
4 Epidemiology, 1992, Vol. 45, pp. 613-619.
- 5 18. *New ICD-10 version of the Charlson Comorbidity Index predicted in-hospital*
6 *mortality.* V. Sundararajan, T. Henderson, C. Perry, A. Muggivan, H. Quan, W. Ghali.
7 s.l.: Journal of Clinical Epidemiology, 2004, Vol. 57, pp. 1288-1294.
- 8