

資料

情報バイアスの定義, 考え方と実例

スズキ 鈴木
サダオ 貞夫*

目的 本稿は、分析疫学研究における「情報バイアス」の基本的な考え方を解説することを目的とする。とくに、事例資料のほとんどない共変数に由来する情報バイアスが存在する場合に、どのような研究結果の偏りが、なぜ生じるのかを示す。また、実際の事例について、そのメカニズムを考察し、研究デザインにおける共変数の定義の重要性を提示する。

方法 情報バイアスは「曝露、アウトカム、あるいは共変数の測定に欠陥があることで、そのために比較群間の情報の質（正確さ）が異なってしまう」と定義されている。定義に基づき、結果の歪みについて説明し、名古屋市 HPV ワクチン分析疫学研究のデータを使用した研究で共変数として導入された「Study Period (sp)」の情報バイアスを分析する。sp は比較群間でその始点が異なっており、接種群で12歳時から初回接種日までの期間が短く定義されている。情報バイアスの定義に合致する共変数 sp を用いた調整が、結果に与える影響について、名古屋市のデータを使用して検証する。

結果 分析の結果、本来調整されるべき年齢の代わりに用いられた sp を共変数として使用した場合、結果に偏りが生じることが明らかになった。sp は、ワクチン非接種群と比べた接種群の平均観察期間が短縮するという差別的誤分類を含み、調整により研究結果に歪みが生じていた。この結果は、報告されたオッズ比が、共変数 sp の情報バイアスの影響により歪んでいることを表している。

結論 分析疫学研究において、共変数の定義が比較群間で異なる場合、情報バイアスが生じ、研究結果に偏りをもたらすことをシミュレーションにより明らかにした。情報バイアスが共変数に生じている事例は、曝露やアウトカムに生じる場合に比べて、教育的な事例資料が圧倒的に少なく、メカニズムも複雑であるため、バイアスが生じることが認識されにくい。共変数の定義の均一性を確保することは、妥当性の高い分析疫学研究を行ううえで極めて重要であり、今後の研究や教育活動において、この側面がさらに強調されるべきである。

Key words : 情報バイアス, 共変数, 系統誤差, Study Period, 差別的誤分類

日本公衆衛生雑誌 2026; 73(6): 567-573. doi:10.11236/jph.25-088

I はじめに

分析疫学の妥当性（内的妥当性）を阻害するものを系統誤差と言うが、系統誤差を分類すると、選択バイアス、情報バイアス、交絡に分けられる。妥当性とその裏概念としての系統誤差、あるいはバイアスといった用語は重要であり、公衆衛生学、疫学の基礎概念でもある。しかし、たとえば、「バイアス」をネット検索すると、近年デフォルトで提示される

ようになったAIによる概要【「バイアス」とは、日本語では「偏り」「偏見」「先入観」などを意味する言葉で、思考や判断に特定の偏りをもたらす要因を指します】をはじめ、思考の歪み、経験に基づいた思い込み、歪んだかたちで認識してしまう現象、などの解説が並んでおり、分析疫学におけるバイアスにたどり着くのは困難である。用語を絞って「情報バイアス」を検索しても、AIによる概要は【情報の収集や解釈の際に、何らかの理由で一方向に偏ってしまうことです。たとえば、情報収集の際に自分が興味のある情報だけを集めたり、解釈の際に先入観にとらわれたりすると、情報バイアスが生じます】となっており、これも単に「情報が偏っている」ということを述べているだけで、分析疫学の情

* 名古屋市立大学大学院医学研究科公衆衛生学分野
責任著者連絡先：〒467-8601 名古屋市瑞穂区瑞穂町川澄1
名古屋市立大学大学院医学研究科公衆衛生学分野
鈴木貞夫
E-mail : ssuzuki@med.nagoya-cu.ac.jp

報バイアスのきちんとした定義ではない。

検索ワードを「分析疫学 情報バイアスとは」にすると、【分析疫学における情報バイアスとは、曝露やアウトカムを測定する際に、情報の取り違えや測定方法の不十分さによって、一方向に偏った結果が得られてしまうことです】と、ようやくそれらしきものが出現するが、そのあと、【より詳しく説明すると、情報バイアスは主に以下のような要因で発生します】と続き、思い出しバイアス、測定バイアス、報告バイアス、盲検バイアスなどが、整理されない形で列挙されている。日本疫学会の疫学用語の基礎知識でも、情報バイアスは【曝露やアウトカムを測定する際、情報の取り違いや測定方法が不十分であるために一方向に偏って測定結果がでてしまうことを情報バイアスという】と説明されており、例として、思い出しバイアスや質問者バイアスがあげられている。初学者が疫学を学ぶときの困難のひとつは、ネットや書籍を含めた「教材」における理論体系の不徹底と、事例の不足があると考えられる。情報バイアスを思い出しバイアスに代表させるのは、医療関係の国家試験レベルまではこなせても、そこから先に、公衆衛生関係の業務に就くには、あまりに心許ない。

「疫学辞典第5版¹⁾」によると、情報バイアスは、【曝露、共変数あるいは帰結の測定に欠陥があることで、そのために比較群間の情報の質（正確さ）が異なってしまう】とある。これは、Wikipediaにある Information bias (epidemiology) の定義【A flaw in measuring exposure, covariate, or outcome variables that results in different quality (accuracy) of information between comparison groups】(2025年11月10日アクセス可能)と一致している。「疫学の事典²⁾」では、情報バイアスの説明として、【研究に必要な情報の真の値と測定値との間で生じる系統誤差】とし、研究に必要な情報として、上記の英文定義と同じ曝露、共変数 (covariate)、アウトカムの3つをあげている。この説明は、「疫学辞典」の定義と一致しているが、発生源別（測定者、対象者、測定手段に関するもの）で事例が列挙されるにとどまり、比較群間の情報の質の差という本質についての明確な記載がない。また、初学者向けの「はじめて学ぶやさしい疫学³⁾」でも、【対象集団から曝露状況やアウトカムの情報を得る時点で起こるバイアスのことである。測定バイアスともいう】と説明があるのみで、初学者に対する説明としては不足しており、単なる測定の誤差と誤解を招く恐れがある。また、共変数についての記載はない。本稿を始めるにあたって、情報バイアスの定義が、「比較群間の情報

の質の差異」であるということを確認しておきたい。なお、因果における原因側は、曝露、説明変数、要因、結果側は帰結、目的変数、アウトカムなどとも呼ばれるが、この稿では、それぞれ曝露、アウトカムと標記する。また、「疫学辞典」で共変数とあるのは、交絡要因、またはその候補として調整目的で解析に組み込まれる変数を指すが、ここでは共変数と標記する。

本稿では、情報バイアスの定義に従って、「比較群間の情報の質（正確さ）が異なる」ことが、曝露、アウトカム、共変数の測定に起きた場合に、どのようなバイアスが生じるのかということ、実例を用いて説明することを目的とする。この中でも、実例が少なく、理解も困難と思われる、共変数由来の情報バイアスについて、詳細に説明する。

II 対象と方法

曝露、アウトカム、共変数の測定のそれぞれについて、情報バイアスがどのような仕組みで起きるかを概説し、情報バイアスを含む論文を紹介する。どのような系統誤差が生じ、結果がどのように歪むか、また、なぜ歪むかについて、実データを用いて分析する。使用するのは、2015年に名古屋市で行われたヒトパピローマウイルス (HPV) ワクチンの分析疫学研究のデータで、ネット上に公開されている⁴⁾ものである。統計ソフトは、SAS (ver. 9.4) を使用した。

III 結 果

1. 曝露、アウトカムに情報バイアスがあるとき

曝露、アウトカムの比較群間での情報の質が異なるということは、曝露における比較群間で、アウトカムの情報の質が異なる場合と、アウトカムにおける比較群間で曝露情報の質が異なる場合とがある。いずれについても、時間的に先に定義された比較群の性質が、あとから収集される情報に影響することがまず考えられる。すなわち前向きコホート研究の枠組みでは、曝露の有無にアウトカムの情報が影響を受けること、症例対照研究の枠組みでは、症例と対照とで曝露の情報の質に差異が生じることが起こりやすい。

1) 曝露、アウトカムの差別的誤分類により情報バイアスが生じるメカニズム

分析疫学情報において、比較群間での情報の質が異なるとしたら、後から収集される情報が、比較群の影響を受けやすいということはすでに述べた。たとえば、職域で過去にアスベスト曝露があったとき、悪性中皮腫の早期発見のための検診が濃厚に行

われことがある。この場合、曝露あり群と比較して、曝露なし群では悪性中皮腫の発見の機会が少なく、本来「中皮腫あり」であるケースが、「中皮腫なし」に誤分類されることがある。その結果、相対危険度は過大評価される。これは、職域でのアスベスト曝露の有無により追跡システムが異なり、情報の正確さに差異が出ることによる情報バイアスである。アスベストと悪性中皮腫の関連を症例対照研究の枠組みで研究するとき、たとえば、カルテから見つかった悪性中皮腫のうち、死亡例については、アスベスト曝露を配偶者からの聞き取りで行ったとすると、症例の結婚前の職歴を知らないと、曝露ありが曝露なしに誤分類される可能性がある。この場合、相対危険度は過小評価される。ランダム割り付け比較試験（RCT）であれば、どちらの群に割り付けられても、後からリンクされるアウトカム情報は一律であり、比較群間で違いは生じない。一方で観察研究では、比較群そのものが後からリンクされる情報に影響を与えることがあり、情報バイアスへの注意が必要である。このように比較群で後からつく情報の誤分類の向きや程度が異なることを差別的（differential）誤分類というが、これが曝露とアウトカム関連を歪めることで情報バイアスの原因となっている。

2) 鉄欠乏症と低レベル鉛中毒の実例

Wright 論文⁵⁾は、都市部のプライマリケア外来でのスクリーニングデータを用いて測定された鉄欠乏症が、血中鉛濃度の増加というアウトカムと関連しているかを検討する分析疫学である。論文のなかで、鉄欠乏症と鉛中毒の関連についての先行研究として、Yip 論文⁶⁾に言及している。それによると、Yip 論文では、鉛中毒のスクリーニングにエリスロサイトプロトポルフィリン濃度が用いられており、鉄欠乏自体がエリスロサイトプロトポルフィリン濃度を上昇させることが、バイアスにつながっていると主張している。鉄欠乏がエリスロサイトプロトポルフィリン濃度を上昇させるとすれば、本来「鉄欠乏あり、鉛中毒なし」に分類されるはずの対象の一部が、「鉄欠乏あり、鉛中毒あり」に誤分類される。鉄欠乏がない場合には、このような誤分類はほとんど生じないため、Yip 論文で推計されるオッズ比は、真のオッズ比より過大に推計される。これ

は、定義より情報バイアスと考えるのが妥当であるが、Wright 論文では、これを選択バイアスとしている。

2. 共変数に情報バイアスがあるとき

共変数とは、解析モデルに組み込む曝露、アウトカム以外の変数のことを指し、多くは交絡要因である。曝露やアウトカムに情報バイアスがあることに比べ、共変数に情報バイアスがある事例の紹介は非常に少なく、また、どのように結果が歪むのかについての解説も見当たらない。ここでは、共変数に情報バイアスがあるとはどのようなことか、また、どのような結果の歪みがどういうメカニズムで生じるのかについて、実例を引きながら解説する。接種群と非接種群の比較の際に年齢調整が必要な状況では、実際の年齢（変数名：age）に加えて、b_age と f_age という、接種群だけ短くした変数を実験的に設定する。いずれの変数も、非接種群では age と同じ値とする。この比較群間で定義の異なる変数を使用することにより、情報バイアスによる結果の歪みを観察し、そのメカニズムを考察する。

1) 名古屋市 HPV ワクチン分析疫学研究と年齢交絡

名古屋市 HPV ワクチンの分析疫学研究は、2015年に HPV ワクチンの接種歴と接種後に多いとされる24症状について、1995–2001年度生まれの名古屋市在住女性、約7万人全員を対象に質問票を郵送したものである。このうち、約3万人から得たデータを解析した、後ろ向きコホート研究である。この稿で使用する対象者は、年齢と接種の有無、さらに接種群については初回接種時の情報のある23,887人とする。この稿で問題提起するのは、Yaju 論文⁷⁾で用いられた Study Period という変数であるが、それに先立ち、年齢交絡について説明する。

図1左は実際の年齢の関与を示したもので、①年長者ほどワクチン接種が多い、②年長者ほど症状の発生が多いという現象のため、年齢が交絡因子となり、ワクチン接種と症状発生の間に見かけ上の関連が生じる。解析で年齢（2015年4月1日時点、変数名：age）を考慮しない場合、実際は年長であるために多発している症状が、年長者で多いワクチン接種由来と誤認識される（図1左の曲線矢印）。その結果、ワクチンのオッズ比が過大評価されることに

図1 ワクチン接種と症状発生に年齢関連変数が及ぼす影響

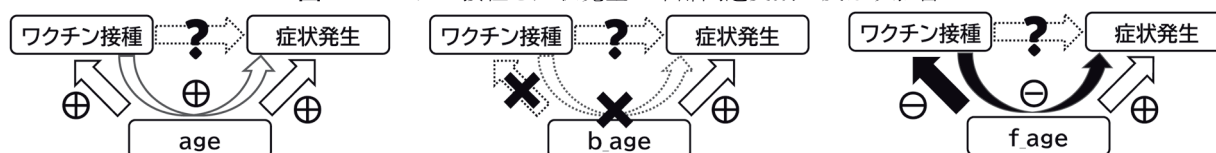
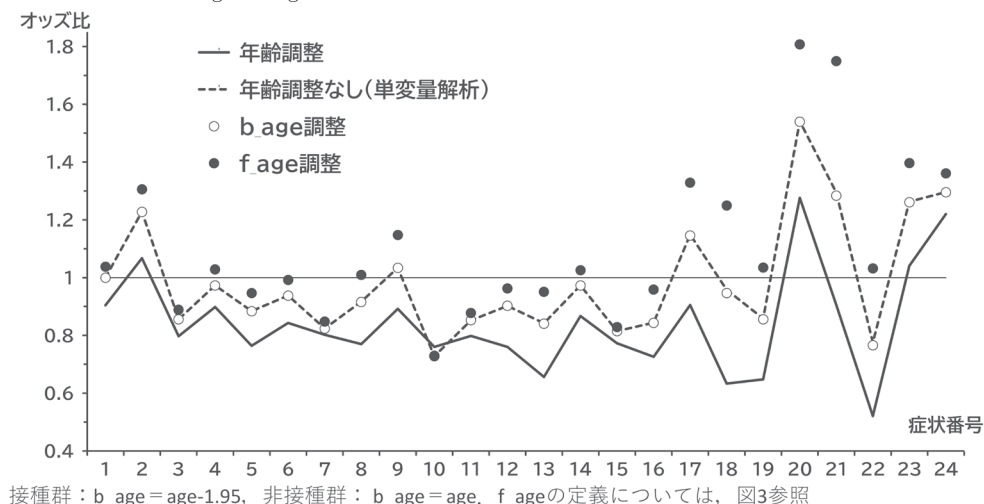


図2 年齢, b_age , f_age で調整した HPV ワクチンのオッズ比：単変量解析との比較



なる。実験的に年齢変数として、接種群のみ群間の平均年齢差（1.95歳）を減じた変数、 b_age を使用して解析すると、ワクチン接種群と非接種群の平均年齢は同じになる。このように、ワクチン接種と年齢の関連を非常に小さくすると、交絡要件を満たさなくなり、 b_age 調整解析は、単変量解析と極めて近いオッズ比が算出されると予測できる（図1中央参照）。実際のデータを再解析したワクチンのオッズ比を図2に示すが、予測通り、 b_age 調整解析は、本来されるべき調整がされず、単変量解析と極めて近いオッズ比を示している。これは、「実験的に比較群間の年齢差をなくした変数 b_age を実年齢の代わりに使用しても、本来交絡要因として作用する年齢が交絡と認識されず、調整が機能しない」現象と理解される。ここで行ったことは、年齢という共変数について、比較群間で異なる定義を用いたということであり、これは情報バイアスの定義と一致する。

2) Study Period と情報バイアス

ここでさらに、接種可能最年少の12歳時から HPV ワクチン初回接種までの猶予期間（変数名： gp ）を減じた年齢（変数名： f_age ）を、接種群にのみ使用する。すなわち、接種群では $f_age = age - gp$ 、非接種群では $f_age = age$ と定義する。共変数として f_age で調整した解析結果も図2に示した。ほとんどの症状で、 f_age 調整オッズ比が b_age 調整オッズ比よりさらに上がっていることが観察される。これは、接種群における b_age と実際の年齢との差が1.95歳だったのに対し、 f_age との差は平均2.55歳とさらに大きかったことによる。メカニズムについては、図1の右に示した通り、 f_age を年齢として使用すると、非接種者の平均年齢が接種者を

上回るため、 f_age とワクチン接種は負の関連を持つようになることで、交絡の方向が正から負へ逆転するためである。このように、情報バイアスによる結果の歪みは、効果の方向と差の大きさの影響を受けている。たとえば、症状10のように年齢交絡の方向が他の症状と逆向き（過小評価方向）なものは、情報バイアスの作用も逆向きになっている。ただし有意差はない。

Yaju 論文⁷⁾では、Study Period（変数名： sp ）というオリジナル変数を導入している。これは、「質問紙調査期間中における症状発症（経験）を確認した期間を次のように、接種群と非接種群であえて異なる扱いとした」というもので、具体的には、「接種群での期間は、症例ごとに接種時から2015年9月の調査時点までの期間とし、非接種群については質問紙調査期間全体（調査で規定していた12歳から2015年9月の調査時点までの期間）」とした。この概念に対して、「症状を経験することが可能であった期間、すなわち“リスク・ピリオド”という意味において、接種群、非接種群に共通する概念として扱うことができるものである」と説明している。

図3に比較群ごとの変数の定義を示す。接種群、非接種群とも $f_age = sp + 12$ であることが分かる。 age が一致する2人で、 sp は接種群が gp 分（平均2.55年）短くなるので、 sp 調整により固定される年齢の対応に「ずれ」が生じることも確認しておきたい。

対象者全員について、 sp は f_age より一律に定数12小さい ($sp = f_age - 12$)。ロジスティック解析で算出される sp の回帰係数は、「曝露（説明変数）が1単位増加することに伴うアウトカム（目的変数）の変化量」なので、定数の加減の影響を受けず、

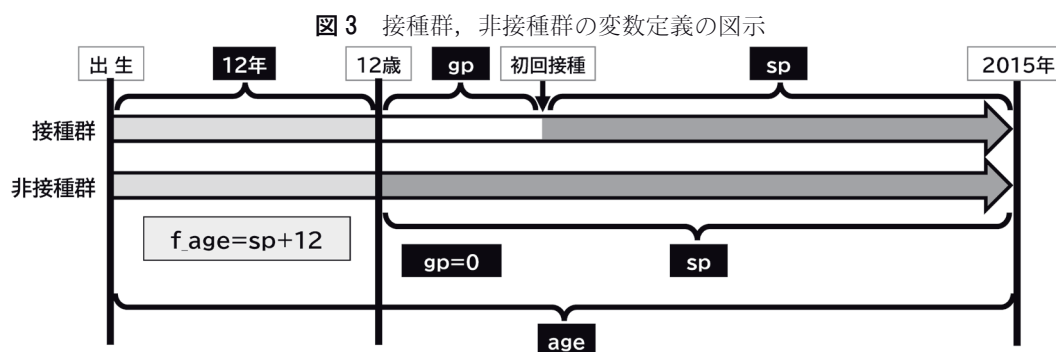


表1 ロジスティック解析のアウトプットの一部：f_age 調整（左）と sp 調整

Parameter	Estimate	Standard Error	Parameter	Estimate	Standard Error
Intercept	-8.7817	1.0244	Intercept	-6.5335	0.4390
vac	0.5920	0.1681	vac	0.5920	0.1681
f_age	0.1873	0.0538	sp	0.1873	0.0538

Effect	Odds Ratio Estimates			Effect	Odds Ratio Estimates		
	Point Estimate	95% Confidence	Wald Limits		Point Estimate	95% Confidence	Wald Limits
vac	1.808	1.300	2.513	vac	1.808	1.300	2.513
f_age	1.206	1.085	1.340	sp	1.206	1.085	1.340

f_age の回帰係数と一致する。また、HPV ワクチン接種（変数名：vac）の回帰係数にも影響しない。したがって、vac の回帰係数の指数である sp 調整オッズ比は、f_age 調整で算出されるのと同じになる。表1に、例として症状20における実際のロジスティック解析のアウトプットの一部を示すが、f_age と sp との12年の差は、切片（Intercept）に反映される（-8.7817対-6.5335）のみで、vac に対する回帰係数（0.5920）も標準偏差（0.1681）も等しく、ワクチンのオッズ比はともに1.206で、95% 信頼区間（1.085-1.340）を含めて一致している。sp 調整のオッズ比は、24症状とも図2の f_age のもの（●）と完全一致する。

IV 考 察

実際の年齢で調整する代わりに、接種群のみ12歳時から初回接種までの猶予期間 gp を減じた f_age で調整すると結果が歪むこと、また、f_age を sp と入れ替えてもまったく同じオッズ比が得られることを示した。これは、年齢の代替として sp を使用すると、情報バイアスのために結果が歪むことを意味している。年齢の代わりに b_age を使用して比較群で平均年齢を同じにした実験的解析（b_age 調整、図2）では、本来調整されるべき age が情報バイア

スを含んだ b_age で解析に組み込まれたため、調整しても単変量解析と結果が変わらなくなることを確認した。sp を用いることで、比較群間の期間の差が b_age 使用時よりさらに広がる（差の平均：1.95⇒2.55）ことにより、sp 調整では単変量解析や b_age 調整解析よりさらに大きく結果が歪むことも明らかにした。

もともと、sp は、ワクチン初回接種以前からある症状は、ワクチンの副反応ではないという、因果関係の「時間性」を厳守することを目的に導入されている。市販前ワクチンの安全性評価に際して行われる RCT であれば、対照群にもプラセボなどの対照薬が接種され、接種前に存在する症状の除外は可能である。しかし、この研究のような観察研究では、接種群に関しては、ワクチン初回接種と症状との前後関係が分かっているため、ワクチン接種以前からある症状は除外できるが、この正確で厳密な運用は、接種群にしか適応できない。非接種群はそもそも接種をしていないので、初回接種時間のデータを持っておらず、sp の始点が定義できないからだ。理論的には、反実仮想的（counterfactual）な時間設定をすべき状況であり、その議論は可能であるが、12歳時に固定してよいものではない。この方法で sp の値を算出することは、非接種群のみに誤分類

が生じることを意味し、差別的なものである。一方、spを用いない解析では、接種群に「接種前からある症状」の紛れ込みがあるが、非接種群にも現実仮想的な意味での紛れ込みは存在する。この症状の有無に対する誤分類は、比較群間で大きく異なるとは考えにくく、接種の影響を受けないのであれば、非差別的 (non-differential) 誤分類である。2値における誤分類が非差別的であると、曝露の推定値が希釈され実際の効果よりもゼロに近づくなど、差別的誤分類と比べてバイアスの方向性の予測が立てやすい。一方、sp導入のアイデアは一考の価値はあるものの、差別的誤分類により、分析疫学の最重要ポイントである比較妥当性を保つことができなかった。オリジナルの変数導入に際しては、綿密な検討が重要である。

情報バイアスが共変数で起きることを、実例を用いて解説した。spのように比較群間で「始点」の定義が異なるものとして、検診における「リードタイムバイアス」がある。これは、生存時間解析におけるアウトカムの情報バイアスであり、検診群での生存時間の過大評価の過程が容易に理解できる。一方で、「共変数」の情報バイアスの定義については、実例が圧倒的に少なく、また、結果が歪む過程についてのメカニズムも複雑で気が付きにくい。注意喚起のため、教科書などで広く紹介するなど、今後の研究や教育活動において、この側面が強調されるべきである。

V 結 語

この稿では、情報バイアス、とくに「共変数」由来のものについて、現象とメカニズム、誤分類との関連を中心に実例をあげながら解説した。疫学研究において、共変数の定義が比較群間で異なる場合、情報バイアスが生じ、研究結果に偏りをもたらすことを明らかにした。情報バイアスが共変数に生じていると、曝露やアウトカムに生じるときと比べて、教育的な事例資料が圧倒的に少なく、メカニズムも

複雑であるため、バイアスが生じることが認識されにくい。部分的な正確さを重視することで、比較妥当性を犠牲にすることのないよう、共変数の定義の均一性を確保することは、妥当性の高い疫学研究を行う上で極めて重要であり、今後の研究や教育活動において、この側面が強調されるべきである。

本研究に関して開示すべき利益相反状態 (COI) 状態はありません。

(受付 2025. 7.15)
(採用 2025.12.12)
(J-STAGE 早期公開 2026. 3. 3)

文 献

- 1) 日本疫学会 (訳), Information bias. Miquel Porta (編). 疫学辞典第5版, 東京: 財団法人 日本公衆衛生協会. 2010; 167.
- 2) 若井建志. 情報バイアス. 日本疫学会 (監修). 疫学の事典. 東京: 朝倉書店. 2022; 414–415.
- 3) 藤吉 朗, 情報バイアス, 日本疫学会 (監修), はじめて学ぶやさしい疫学 (改定第4版). 東京: 南江堂. 2024; 76–77.
- 4) 名古屋市. 子宮頸がん予防接種調査の結果を報告します, 調査回答データ. 2024. <https://www.city.nagoya.jp/kenkofukushi/page/0000088972.html> (2025年11月10日アクセス可能).
- 5) Wright RO, Shannon MW, Wright RJ, et al. Association between iron deficiency and low-level lead poisoning in an urban primary care clinic. *Am J Public Health* 1999; 89: 1049–1053.
- 6) Yip R. Iron deficiency: contemporary scientific issues and international programmatic approaches. *J Nutr* 1994; 124(suppl): 1479S–1490S.
- 7) Yaju Y, Tsubaki H. Safety concerns with human papilloma virus immunization in Japan: Analysis and evaluation of Nagoya City's surveillance data for adverse events. *Jap J Nurs Sci* 2019; 16: 433–449.

Definition, concept, and practical example of information bias

Sadao SUZUKI*

Key words : information bias, covariate, systematic error, study period, differential misclassification

Objectives This paper aims to elucidate the fundamental concept of “information bias” in analytical epidemiological studies. Specifically, it focuses on bias arising from covariates, an area for which educational case examples are scarce, and explains how and why such bias distorts research results. Using a real-world example, the paper examines the underlying mechanism and emphasizes the importance of precise covariate definition in study design.

Methods Information bias is defined as “a flaw in measuring exposure, covariate, or outcome variables that results in different quality (accuracy) of information between comparison groups.” Based on this definition, we explain how distorted results arise and analyze information bias associated with the covariate “Study Period (SP)” in an epidemiological study using human papillomavirus vaccination data from Nagoya City. The starting point of SP differed between comparison groups, with the vaccinated group having a shorter period defined from age 12 years to the first vaccination date. Using the Nagoya data, we evaluated the impact of adjusting for SP, a covariate that meets the definition of information bias, on the study results.

Results Simulation analyses showed that bias emerged when SP was used as a covariate instead of age, which should have been adjusted for. Adjustment for SP introduced differential misclassification because the average observation period was shorter in the vaccinated group than in the non-vaccinated group, leading to distortion in the study results. Consequently, the reported odds ratios were distorted due to information bias originating from the covariate SP.

Conclusion This study demonstrates through simulation that information bias arises when covariate definitions differ between comparison groups in analytical epidemiological studies, leading to distorted results. Compared with bias related to exposure or outcome variables, information bias originating from covariates is rarely addressed in educational materials, and its mechanisms are more complex, making it less likely to be recognized. Ensuring uniform covariate definitions is crucial for valid analytical epidemiological research and warrants greater emphasis in future research and education.

* Department of Public Health, Nagoya City University Graduate School of Medical Sciences