

サイエンスダイレクトで公開されている [コンテンツ一覧](#)

## Environment International

出版物ホームページ : [www.elsevier.com/locate/envint](http://www.elsevier.com/locate/envint)

### 研究の質を評価するための提案：バイオモニタリング、環境疫学、短寿命化学物質 (BEES-C) の指標

Judy S. LaKind <sup>a,b,c,\*</sup>, Jon R. Sobus <sup>d</sup>, Michael Goodman <sup>e</sup>, Dana Boyd Barr <sup>f</sup>, Peter Fürst <sup>g</sup>, Richard J. Albertini <sup>h</sup>, Tye E. Arbuckle <sup>i</sup>, Greet Schoeters <sup>j,k</sup>, Yu-Mei Tan <sup>d</sup>, Justin Teeguarden <sup>l</sup>, Rogelio Tornero-Velez <sup>d</sup>, Clifford P. Weisel <sup>m</sup>

<sup>a</sup> LaKind Associates, LLC 106 Oakdale Avenue, Catonsville, MD 21228, USA

<sup>b</sup> Department of Epidemiology and Public Health, University of Maryland School of Medicine, USA

<sup>c</sup> Department of Pediatrics, Penn State University College of Medicine, Milton S. Hershey Medical Center, USA

<sup>d</sup> National Exposure Research Laboratory, Human Exposure and Atmospheric Sciences Division, US Environmental Protection Agency, Research Triangle Park, NC 27711, USA

<sup>e</sup> Department of Epidemiology, Rollins School of Public Health, Emory University, 1518 Clifton Rd., Atlanta, GA 30322, USA

<sup>f</sup> Department of Environmental and Occupational Health, Rollins School of Public Health, Emory University, 1518 Clifton Road, NE, Room 272, Atlanta, GA 30322, USA

<sup>g</sup> Chemical and Veterinary Analytical Institute, Münsterland-Emscher-Lippe (CVUA-MEL) Joseph-König-Straße 40, D-48147, Münster D-48151, Germany

<sup>h</sup> University of Vermont College of Medicine, P.O. Box 168, Underhill Center, VT 05490, USA

<sup>i</sup> Population Studies Division, Healthy Environments and Consumer Safety Branch, Health Canada, 50 Colombine Dr., A.L. 0801A, Ottawa, ON K1A 0K9, Canada

<sup>j</sup> Environmental Risk and Health Unit, VITO, Industriezone Vlasmeer 7, 2400 Mol, Belgium

<sup>k</sup> University of Antwerp, Department of Biomedical Sciences, Belgium

<sup>l</sup> Pacific Northwest National Laboratory, 902 Battelle Boulevard, P.O. Box 999, MSIN P7-59, Richland, WA 99352, USA

<sup>m</sup> Environmental and Occupational Health Sciences Institute, Robert Wood Johnson Medical School, UMDNJ, 170 Frelinghuysen Road, Piscataway, NJ 08854, USA

#### 記事情報

##### 記事の履歴

2014年2月26日受領

2014年7月16日に受理

オンラインで利用可能な xxxx

##### キーワード

BEES-C

バイオモニタリング

至る所に存在する (ユビキタス) 化学物質

短い生理的半減期

評価

指標

環境疫学

## 要約

ばく露評価の質は、環境疫学研究の全体的な質を決定する主要な要素である。生理的半減期の短い至る所に存在する（ユビキタス）化学物質へのばく露を評価するためのツールとしてのバイオモニタリングの使用は比較的最近になってから始まった。これらの化学物質には、分析室やサンプリング機器に含まれる化学物質の存在、横断研究における時間的順序の確立の難しさ、ばく露とバイオマーカーの濃度の短期及び長期の変動、適切なばく露分類に必要な測定数に関する情報不足など、いくつかの問題がある。現在までのところ、科学研究分野において、ばく露指標としてバイオモニタリングを使用する短寿命化学物質の研究を計画、実行、解釈するための体系的なガイドラインを開発しておらず、また、この種の研究の質を評価するための WOE 評価、または助成金や出版物のピアレビューのための体系的なガイドラインを開発していない。本研究では、短寿命化学物質のバイオモニタリングデータを用いた疫学研究に影響を与える重要な問題点を説明し、短寿命化学物質のバイオモニタリングデータを取り入れた研究提案や研究の質を評価するための体系的な指標であるバイオモニタリング、環境疫学、短寿命化学物質（BEES-C）を提案する。短寿命化学物質の生物学的測定を含む疫学研究の評価に不可欠と考えられる 3 つの分野の品質基準について説明している。1) バイオマーカーの選択と測定、2) 試験デザインと実行、3) 一般的な疫学研究デザインに関する考察。BEES-C のような評価ツールの開発は単純ではなく、議論の余地のないものでもないことを認識している。

©2014 年 The Authors. 発行 : Elsevier Ltd. これは CC BY-NC-ND ライセンスの下でのオープン・アクセス論文である。

(<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

\* 対応する著者: LaKind Associates, LLC, 106 Oakdale Avenue, Catonsville, MD 21228, USA. Tel.: +1 410 788 8639.

*E-mail addresses:* [lakindassoc@gmail.com](mailto:lakindassoc@gmail.com) (J.S. LaKind),

[Sobus.Jon@epa.gov](mailto:Sobus.Jon@epa.gov) (J.R. Sobus),

[mgoodm2@emory.edu](mailto:mgoodm2@emory.edu) (M. Goodman), [dbbarr@emory.edu](mailto:dbbarr@emory.edu) (D.B. Barr),

[Peter.Fuerst@cvua-mel.de](mailto:Peter.Fuerst@cvua-mel.de) (P. Fürst), [Ralbert315@aol.com](mailto:Ralbert315@aol.com) (R.J. Albertini), [Tye.Arbuckle@hsc.gc.ca](mailto:Tye.Arbuckle@hsc.gc.ca) (T.E. Arbuckle), [greet.schoeters@vito.be](mailto:greet.schoeters@vito.be) (G. Schoeters), [Tan.Cecilia@epa.gov](mailto:Tan.Cecilia@epa.gov) (Y.-M.

Tan), [jt@pnnl.gov](mailto:jt@pnnl.gov) (J. Teeguarden),

[Tornero-Velez.Rogelio@epa.gov](mailto:Tornero-Velez.Rogelio@epa.gov) (R. Tornero-Velez),

[weisel@eohsi.rutgers.edu](mailto:weisel@eohsi.rutgers.edu) (C.P. Weisel).

## 1. 序文

疫学研究は、様々な化学的、物理的、生物学的、放射線学的、行動学的なばく露がヒトの健康に及ぼす影響を評価する上で重要な役割を果たしている。しかし、ヒトにおける因果関係の仮説を検証するために特別に設計され、厳密に実行された疫学研究であっても、しばしば矛盾した結果が報告される。保健政策の提言を行う規制機関やコンセンサス・パネルは、一般的に疫学研究結果を評価するために WOE (weight-of-evidence) アプローチに依存している。

<http://dx.doi.org/10.1016/j.envint.2014.07.011>

0160-4120/©2014 年 The Authors. 発行：Elsevier Ltd. これは CC BY-NC-ND ライセンスの下でのオープン・アクセス論文である。 (<http://creativecommons.org/licenses/by-nc-nd/3.0/>)

以下 196 ページ

WOE 評価は、結論が利用可能な最も強力な証拠に基づいていることを確認するために研究の質を評価しない場合、不完全であったり、誤解を招く可能性がある。さらに、助成金申請書や原稿のピアレビューにおける研究の質を評価することで、ヒトのばく露や健康に関する研究の全体的な質を高めるのに役立つ。

研究の質の決定には常に一定の程度専門的な判断が必要であるが、疫学的証拠の強さの評価は、体系的に適用される合意された基準に依存すべきであるという考えが出てきている (Vandenbroucke ら、2007 年)。これらの考えが、いくつかの研究の質評価ツールの開発と改善の動機となった。これらのツールのいくつか (例えば、STROBE (Vandenbroucke ら、2007 年) ; CONSORT (Moher ら、2001 年) ) は、分野を超えて適用される一般的な問題を扱っている。他のツールは、医学や生命科学の様々な分野に特化して開発された (例えば、遺伝学的研究のための STREGA (Little ら、2009 年)、治療効果の比較研究のための GRADE (Owens ら、2010 年)、診断精度の研究のための STARD (Bossuyt ら、2004 年) ) 。

研究の質を取り入れた WOE 評価の標準化が進んでいる現在の傾向に鑑みると、環境疫学研究を評価するための手段が、研究デザインの開発段階や原稿の審査段階でも、相対的に少なくなっていることは注目に値するが、説明が難しい。エビデンスの重み付け評価のための研究の質の評価に焦点を当てた評価スキーム (Harmonization of Neurodevelopmental Environmental Epidemiology Studies) (Youngstrom ら、2011 年) では、診断精度研究の質の評価 (QUADAS) をコーディングツールの基礎として使用しているが (Whiting ら、2003 年)、その名が示すように、この評価指標は神経発達研究を中心としたものである。National Toxicology Program は最近、研究の質を評価するためのアプローチを開発し (NTP、2013 年)、これを用いて環境化学物質と糖尿病に関する文献を調査した (Kuo ら、2003 年)。この事業計画には、疫学及び毒性学の文献の評価を含み、非難分解性及び難分解性の化学物質は含まれていたが、短寿命化学物質のバイオモニタリング固有の問題は含まれていなかった。

環境疫学研究の最適な実行に関する体系的なガイダンスを提供するツールがないことは、集団ベースの研究に依存する規制上の意思決定にとって重要な制約となっている。他の研究分野とは異なり、ヒトに有害性影響を誘発することを目的とした介入研究は、倫理的に可能であるとしても、ほとんど行われていないため、環境疫学データに基づく WOE 評価は、独特のものである。そのため、環境疫学研究はほとんど常に観察的な研究であり、様々な情報源から生じる避けられない不確実性の影響を受けている。環境疫学における不確実性の重要なソースの一つであると同時に、急速に進展している分野でもあるのが、ばく露科学である。

ばく露評価は、生理学的半減期の短い化学物質を含む環境疫学研究 (Hertz-Picciotto、1998 年) において、全体的なデータの質を決定する主要な要素である。短寿命化学物質とは、化学物質分量の半分を体外に排出するのに必要な時間が、およそ数分から数時間あるいは数日である化学物質のことである。短寿命化学物質のばく露評価の質は、ばく露の様々な側面を調べるためのバイオモニタリングを用いた研究と同様に、健康影響との関連性を評価するためのデータの有用性と密接に関連している。近年、ばく露科学の手法は、バイオモニタリングによる環境化学物質の検出能力の向上により、特に恩恵を受けている。バイオモニタリングとは、血液、尿、呼気、

母乳、毛髪など、ヒトの様々なマトリックス中の化学物質を測定することである。バイオモニタリングデータは、すべての経路（経口、吸入、経皮、経胎盤）からのばく露を統合したものであり、以下のような場合に有用である。(1) 集団参照範囲の設定、(2) 亜集団における異常ばく露の特定、(3) 集団内の時間的変動と傾向の評価、(4) 個人のばく露量を推定するためにデザインされたアンケートの検証、(5) 疫学研究における健康影響との関連性の検討。

難分解性有機汚染物質や金属のばく露量を測定するための基礎としてのバイオモニタリングを用いた疫学研究は、何十年にもわたって実行されてきた。対照的に、至る所に存在する（ユビキタス）生理的半減期の短い化学物質（ベンゼン、フタル酸エステル類、特定の農薬など）のバイオモニタリングは、比較的最近始まったもので、ばく露のばらつきや、分析室やサンプリング機器に含まれる多くの化学物質のユビキタスな性質によって複雑になるため、これらの化学物質に関するデータの解釈にはいくつかの新たな問題がある。また、これらの化学物質は、研究に使用するマトリックスを選択する際にも問題となる。現在までのところ、これらの化学物質のバイオモニタリング研究を実行し、解釈するための体系的なガイドラインは開発されていない。同様に、WOE 評価あるいは助成金や出版物の査読のため、この種の研究の質を評価するための公表された方法はない。

この知識ギャップは、2013 年の国際ワークショップ「Best Practices for Obtaining, Interpreting and Using Human Biomonitoring Data in Epidemiology and Risk Assessment: Chemicals with Short Biological Half-Lives（疫学とリスク評価におけるヒトのバイオモニタリングデータの取得、解釈、利用のための最適な実行：生物学的半減期の短い化学物質）」の明確な目的となった。このワークショップには、分析化学、ばく露とリスクの評価、疫学、医学、生理学的薬物動態（PBPK）モデリング、臨床バイオマーカーを専門とする政府機関、学界、民間機関の専門家が参加した。ワークショップの目的は以下の通りであった。(i) 生理的半減期の短い化学物質の生物学的モニタリングデータを用いた疫学研究に影響を与える問題を説明すること。(ii) 短寿命化学物質のバイオモニタリングデータを取り入れた研究提案や研究の質を評価するための体系的なスキームを構築する。

本論文では、短寿命化学物質の生物学的測定を含む疫学研究の評価の基礎となると考えられる 3 つの分野の品質基準について説明する。1) バイオマーカーの選択と測定、2) 試験デザインと実行、3) 一般的な疫学研究デザインの考慮事項である。これらのトピック領域の主要な側面について議論し、次に、階層化されたマトリックスとして構成された評価指標（バイオモニタリング（Biomonitoring）、環境疫学（Environmental Epidemiology）、短寿命化学物質（Short-Lived Chemicals（BEES-C））の指標）の提案に組み入れた（表 1）。提案されている評価指標のいくつかの側面には、難分解性化学物質と短寿命化学物質の両方の疫学研究に関連する研究デザインの要素が含まれている。実際、STROBE のような広く受け入れられている評価指標の側面は、ここで提案する評価指標に意図的に織り込まれている（Gallo ら、2011 年；Little ら、2009 年；Vandenbroucke ら、2007 年）（STROBE は、観察研究の報告を改善するための方法や、これらの研究を批判的に評価するためのガイダンスを提供している。STROBE は査読者、ジャーナル編集者、読者が利用できるようにデザインされている「（Vandenbroucke ら、2007）」）。STROBE の既定及び新規の側面は、ばく露評価のアプローチとして短寿命化学物質のバイオモニタリングを用いた研究の質を評価する上で批判的な要素であるが、本コミュニケーションの第一の目的は、短寿命化学物質の研究の批判的な側面をカバーすることである。

ある研究を評価するために使用できる品質問題のリストは長い。すべてを網羅しているが扱いにくい手段を開発することと、最も重要な問題のみを含む（短寿命化学物質に特有の、あるいは特に重要な研究の側面に焦点を当てている）より差別化された実用的な手段を開発することとの間とは対立する状態にある。提案された BEES-C 指標を開発するにあたり、後者を選択した。

この指標は、ばく露と健康影響との関係を調べる研究だけでなく、ばく露の様々な側面（例えば、時間的・空間的傾向）を調べるバイオモニタリングデータを用いた研究にも適用可能である。ここで提起され、BEES-C 指標によって対処された問題は、職業に関係のある研究と栄養疫学を含めて短寿命化学物質の生物学的指標に関わる複数の分野に影響を及ぼす。

環境疫学研究における短命化学物質の特徴として、特に注意が必要なのは、対象となる健康影響に関連するばく露範囲を表すサンプルの数と時期、パーソナルケア製品、実験器具、塵埃、食品など、現在製造されている製品にはこれらの化学物質の多くが普遍的に存在しているため、サ

ンプルの汚染を避けるために特別な配慮が生じること、適切な生物学的マトリックスの選択により、1つのサンプルで多数の化学物質を測定することができるため、完全な報告と多重比較に関する問題に注意を払う必要性が高まることである。

以下 197 ページ

表 1

バイオモニタリング、環境疫学、短寿命化学物質 (BEES-C) の評価指標。生理的半減期の短い化学物質のバイオモニタリングを含む疫学研究の質を評価するための評価基準。評価基準は、バイオモニタリングをばく露指標とした環境疫学研究のいくつかの側面をカバーしている (表の下部に頭字語が定義されている)。Justification 欄は、意思決定のプロセスの透明性を高めるために使用される。

研究評価の構成要素	TIER 1	TIER 2	TIER 3	Justification
バイオマーカーの選択と測定				
生物学的関連性 (親化合物/代替の関係)				
ばく露バイオマーカー	指定されたマトリックス中のバイオマーカーは、外部ばく露、内部ばく露、目標ばく露量と精密・正確な定量的関係を持つ。	指定されたマトリックス内のバイオマーカーと外部ばく露、内部ばく露、または目標ばく露量との間に関係があることを示す証拠が存在する。	指定されたマトリックスに含まれるバイオマーカーは、ばく露/用量に対する代替性が低い (正確さ・精度が低い)。	
効果バイオマーカー	AOP における重要なイベントのバイオインジケーター。	健康影響との関連性が示された効果のバイオマーカーであるが、その作用機序は不明である。	バイオマーカーは未確定の結果をもたらす (例えば、バイオマーカーは健康影響に特異的ではない)。	
特異度	バイオマーカーは、1つの親化合物へのばく露から得られる。	バイオマーカーは、類似の有害性エンドポイントを持つ複数の親化合物に由来する。	バイオマーカーは、有害性エンドポイントの種類が異なる複数の親化合物に由来している。	
手法の感度 (検出限界)	検出限界は、研究課題に取り組むのに十分な割合のサンプルから化学物質を検出するのに十分な低さである。	NA	検出頻度が低すぎて研究仮説に取り組めない。	
バイオマーカーの安定性	既知の履歴と記録された安定性データを持つサンプル、またはリアルタイム測定を使用しているサンプル。	サンプルは保存中の損失が知られているが、低量ばく露と高量ばく露の違いを定性的に評価することができる。	履歴が不明なサンプル、または目的の分析物の安定性データがないサンプル。	
サンプルの汚染	サンプルは、収集時から測定時まで無汚染である (例えば、	これらの手順を使用しない/記録しないことを検討する。	汚染の問題が知られており、問題が解決されたという	

	分析物・無含有保証の収集用品や標準物質の使用、現場と実験室の両方でのブランクの適切な使用など)。研究には、研究データの信頼性を保証するために必要な手順を記録したものが含まれる。		記録はない。	
手法の要件	バイオマーカーの明確な同定と定量を必要とする感度で提供する機器 (例: GC-HRMS、GC-MS/MS、LC-MS/MS)。	高い信頼性と必要とする感度でバイオマーカーの同定を可能にする機器 (例: GC-MS、GC-ECD)。	バイオマーカーの定量が可能で、その手法が既知の干渉物質を持っている機器 (例: GC-FID、分光法)。	
マトリックス調整	調整が必要な場合には、調整済み濃度と非調整済み濃度の結果が含まれている。	研究では、1つの手法(マトリックス調整済みか否か)を用いた結果のみを提供している。	調整手法が確立されていない(例: 毛髪調整)。	
<b>研究デザインと実行</b>				
時系列性	ばく露と発現事象の間に確立された時間順序; ばく露と発現事象/再構成されたばく露の間の関連する間隔、及び関連するばく露範囲の適切な考慮。	ばく露と発現事象の間に時間的順序が確立されているが、関連するばく露範囲は考慮されていない。	ばく露と発現事象の間に時間的順序が確立されていない研究。	
ばく露変動と誤分類	十分なサンプル数。精度(例えば、感度と特異度)と信頼性(例えば、ICC)の尺度を計算することによって考慮される誤差。 1つのサンプルが使用されている場合、1つの測定値からの誤差が無視できる程度であるという証拠がある。	1つ以上のサンプルを収集したが、エラーを明示的に評価していない。	誤差を考慮せずに単一のサンプルに基づいたばく露。	
<b>一般的な疫学研究デザインの考察</b>				
研究の論拠	先験的に定式化された仮説を評価するために特別にデザインされた研究。	既存のサンプルやデータを用いて、先験的に定式化された仮説を評価する研究。	あらかじめ指定された仮説を持たないデータマイニング研究; 複数の同時仮説検証。	
研究参加者	母集団ベースの不偏的選択プロトコル; 高い応答率または脱	集団ベースの不偏的選択プロトコル; 応答率が低い、または	サンプル選択の方法、応答率/脱落率は報告されてい	

	落例が少ない。	脱落例が多い。	ない。	
データ分析	因果モデルと予測モデルの明確な区別；効果修飾の評価と交絡因子の調整を伴う外的因子への十分な配慮；感度分析。	外的因子を十分に考慮しているが、感度分析は伴っていない。	外的因子の制御が不十分。	
報告	研究は、その目的を明確に述べ、読者が検証された仮説の数（結果が与えられた仮説の数だけではなく）を評価できるようにする。複数の同時仮説検証が行われる場合は、その影響をなるべく PFP または FP : FN 比を推定することで評価する。発現事象報告のバイアスの証拠がなく、結論は観察された結果を超えたものにはならない。	結論は保証されたように見えるが、検証された仮説の数不明瞭である（明示されていないか、判別が困難である）、または複数の検証を考慮していない。	データ要約を選択的に報告し、提示された結果の方法や選択に関して透明性を欠いている研究。	

AOP = 有害性転帰経路；FP = 偽陽性；FN = 偽陰性；GC-HRMS = ガスクロマトグラフィー/高分解能質量分析計；GC-MS = ガスクロマトグラフィー/質量分析；GC-ECD = ガスクロマトグラフィー電子捕足検出器；GC-FID = ガスクロマトグラフィー-炎イオン化検出器、ICC = クラス内相関係数；NA = 非該当；PFP = 偽陽性確率

以下 198 ページ

これらについては、以下のセクションで詳しく説明し、それぞれの問題について例を示す。機器のトピックのほとんどはばく露のバイオマーカーに関連するものであるが、効果のバイオマーカーについても関連性があれば記載している。

## 2. BEES-C 指標の使用

BEES-C 指標は、研究デザインの開発、助成金提案書のレビュー、論文原稿のピアレビュー、WOE 評価の実行など、複数の目的に使用することができる。

### 2.1. BEES-C の使用目的

BEES-C ツールの最終的な目標は、ヒトにおける短寿命化学物質の研究に関する文献全体の改善で研究者を支援することである。BEES-C は、以下のような目的で使用されることを意図したものではない：(i) 研究者が仮説作成のための研究を行わないようにすること、(ii) 低レベルの研究を WOE 評価に含めないようにすること。

どのようなタイプの検査機器でもそうであるように、評価プロセスには、階層化の観点からも、また、検査機器のどの点が特定の研究に関連しているかを判断するためにも専門家の判断が必要である。

以下の項では、BEES-C の主な特徴を事例を交えて説明する。ここでは、BEES-C を利用するための推奨事項について述べる。BEES-C は、ばく露のバイオマーカーと何らかの効果指標（例：効果のバイオマーカー、医師による診断疾患）との関連性に関する疫学研究の一部である

ヒトのバイオモニタリング研究に適用されるが、他の目的（例：時間的または空間的傾向分析のためのばく露評価）のために計画されたヒトのバイオモニタリング研究には、BEES-C 指標の一部のみが適用される。

## 2.2. BEES-C の利用方法

表 1 は、研究デザインの側面（行）と評価段階（列）に沿って整理されている。審査対象の各研究について、重要な側面が行ごとに評価され、該当するセルは色分けされている（図 1）。これにより、研究者／査読者は研究の質の全体像を把握することができる。この指標の使用者は、各決定に対して正当な根拠を示す必要がある（表 1）；これはプロセスの透明性を高めることになる。BEES-C 指標は以下のように使用される：(i) 提案された研究デザインを評価する研究者が、研究の質が最大化されていることを確認するための手段として使用、(ii) 研究の質を体系的に評価し、質を向上させることができる分野を特定するために、論文や出版物の査読者が使用、(iii) 意思決定（例：規制基準の策定に使用するには十分に質の高い研究か？研究はメタ解析に含まれるか？）を通知する手段として研究の質を評価するためにシステムティックレビューを実行している者が使用、(iv) BEES-C を現在の既存のレビュースキームに組み入れたいと望む者が使用。例えば、短寿命化学物質に特異的に適用可能な我々の提案するアプローチの問題点の多くは、OHAT（Office of Health Assessment and Translation Approach）の草案（NTP、2013 年）にはまだ含まれていないが、「環境化学物質、天然物質、または混合物（総称して「物質」と呼ばれる）が有害性健康影響を引き起こすという証拠を評価するための文献ベースの評価」を実行するためのアプローチに組み込まれる可能性がある。

本研究の品質評価指標には、以下の各項目について、原稿または企画書で明示的に報告することが暗黙の了解となっている。言い換えれば、その研究が特定のレベルの基準を満たしているかどうかを評価するためには、その問題に関する情報が明確に記述されていなければならないということである。以前に発表されたバイオモニタリングデータ（例：米国国民健康栄養調査（NHANES））に依存している研究については、同じ報告要件を満たさなければならない。著者は、研究の検出限界、相対標準偏差、関連する品質管理パラメータなどの関係がある項目を含め、方法の記述を明確にすべきである。

このプロセスに数値スコアリングがないのは意図的なものである。ほとんどの構成要素については質の高い研究であるにもかかわらず、研究結果の信憑性を大幅に低下させるような重要な問題に対処していない場合があることは間違いない。全体的に高い「スコア」は、この問題を覆い隠す。その代わりに、柔軟性を高める質的アプローチを提案する。

補足：評価のすべての側面で Tier 1 に分類される研究は知られていない。すべての側面で Tier 1 に分類される研究は確かに目標であり、強固なデータを提供することになるが、ほとんどの研究は Tier 2 または Tier 3 と考えられる側面を含むことになる。研究データに対する利用者の意図にもよるが、評価上の問題点によっては問題にならない場合もある。一方、Tier 3 の指定により、その研究の有用性が低いと判断される問題もある（例：サンプルに汚染がないことを証明できない）。

## 3. BEES-C の構成要素

最初に、短寿命バイオマーカーに関連した BEES-C の構成要素について説明する。これに続いて、より一般的な疫学研究デザインの問題に関連する BEES-C の側面について述べる。

### 3.1. バイオマーカーの選択と測定

バイオマーカー/生物学的マーカーは、「生物学的システムにおける変化や事象の指標」と定義されている。「ばく露の生物学的マーカーとは、組織、細胞、または体内液などの生物学的媒体から得られる細胞状、生化学的、分析的、または分子的な測定値を指し、薬剤へのばく露を示すものである」（Zartarian ら、2005 年）。このように、バイオマーカーは、体内の化学物質やその代謝物の量を測定することで、化学物質へのばく露を評価するために使用することができる。さらに、バイオマーカーは健康影響の指標として使用することができる。ばく露と効果のバイオ



マーカーの多くは短寿命であり、環境化学物質へのばく露と環境化学物質による健康影響に関するヒトの研究では、両方のタイプのバイオマーカーが一般的に使用されている。この評価ツールは主にばく露のバイオマーカーに焦点を当てているが、ここで明らかにした原則の多くは効果のバイオマーカーにも適用される。

一般的なルールとして、ばく露と健康影響との関連を観察するように設計された研究は、適切に確立されたバイオマーカーをばく露または健康のエンドポイントの代替として使用される場合において、より正当性が高まる。バイオマーカーが高品質とみなされるために満たすべき基準については、一般的なコンセンサスが得られている（NRC、2006年；Zelenkaら、2011年）。これらの基準の中には、バイオマーカーに固有の特性（例えば、化学物質ばく露との関連性や生物学的関連性）に基づくものもある。その他の基準は、バイオマーカーの測定に関するもので、すなわちバイオマーカーを定量するために使用される方法の精度と正確さ、保存中のバイオマーカーの安定性、バイオマーカーの定量に誤差をもたらすサンプル汚染の可能性、測定誤差を引き起こす可能性のある生物学的マトリックスの効果を調整する必要性などがある。バイオマーカーの選択と測定の重要な側面については以下のサブセクションで説明し、BEES-Cのために提案している階層化スキームを表1に示す。

### 3.1.1. 関連性

化学物質の生成からヒトへの接触、目標の投与量、ならびにその後の分子、細胞、器官、個体、集団での反応までの経路を示すために発生源から結果までの連続性が頻繁に使用される。

以下 199 ページ

仮想研究 1				仮想研究 2			
研究評価の構成要素	TIER 1	TIER 2	TIER 3	研究評価の構成要素	TIER 1	TIER 2	TIER 3
<b>バイオマーカーの選択と測定</b>				<b>バイオマーカーの選択と測定</b>			
生物学的関連性				生物学的関連性			
ばく露バイオマーカー				ばく露バイオマーカー			
効果バイオマーカー				効果バイオマーカー			
特異度				特異度			
手法の感度				手法の感度			
バイオマーカーの安定性				バイオマーカーの安定性			
サンプル汚染				サンプル汚染			
手法の要件				手法の要件			
マトリックス調整				マトリックス調整			
<b>研究デザインと実行</b>				<b>研究デザインと実行</b>			
時系列性				時系列性			
ばく露変動と誤分類				ばく露変動と誤分類			
<b>一般的な疫学研究デザインの考察</b>				<b>一般的な疫学研究デザインの考察</b>			
研究の論拠				研究の論拠			
研究参加者				研究参加者			
報告				報告			
データ分析				データ分析			

図 1. BEES-C 指標を用いてバイオモニターされた短寿命化学物質の 2 つの仮想研究についての品質比較の例。各仮想研究のレビューでは、重要な側面が行ごとに評価され、適切なセルが色分けされているため、研究者/査読者は研究の質の全体像を把握することができる。セル内のテキストは読みやすくするために削除されている。

またバイオマーカーは、ばく露、用量、及び生物学的反応を経験的に特徴付ける手段として使用されることがある。本セクションでは、ばく露のバイオマーカー（すなわち、標的における親化合物、代謝物、相互作用生成物（WHO、2001年））と効果のバイオマーカー（すなわち、測定可能な生化学的または生理学的反応で、健康影響に関連するもの（WHO、2001年））の両方を、ばく露と健康影響との関連性に関する疫学研究の重要な構成要素として検討する。

3.1.1.1. ばく露のバイオマーカー 疫学研究は、仮説駆動型の研究であってもよいし、より仮説

生成に向けた研究であってもよい。後者の場合、ばく露の最も適切なバイオマーカーは、外部ばく露または内部ばく露の正確で精密な代替となるものである。強固な生物学的根拠が存在し、生物学的な「標的」が分かっている場合、最も適したバイオマーカーは、標的（分子、細胞、または臓器レベル）で直接測定されるもの、または目標の用量を正確かつ精密に代替するものである。

理想的には、疫学研究で使用されるバイオマーカーについて、ばく露、用量、バイオマーカーのレベル間の定量的な関連性を明確に理解していることが必要である。標的組織のサンプリングの侵襲性を考慮すると、ほとんどのバイオマーカーベースの疫学研究では、血液、尿、毛髪、またはその他の容易にアクセス可能なマトリックスのサンプルを利用している。これらのマトリックスからのバイオマーカー測定値とばく露/用量のレベルとの間の定量的関係を解明するには、化学物質の吸収、分布、代謝、排泄（ADME）の理解が必要である；これらのプロセスは、しばしば薬物動態（PK）モデルまたは生理学的薬物動態（PBPK）モデルを用いて記述される。疫学研究でバイオマーカーを使用する前に、ADME に影響を与える可能性のある内的要因（例えば、遺伝、ライフステージ、妊娠、性別）と外的要因（例えば、食事、薬、病状）と同様に、化学物質の ADME についてもしっかりと理解しておく必要がある。さらに、短寿命のバイオマーカーについては、サンプル採取に関連して、特定のタイミングの詳細（例えば、時間帯、食事ばく露に関連する化学物質については最後の食事をしてからの時間、最後の尿を排泄してからの時間）を知ることが重要である。理想的には、バイオマーカー濃度とばく露/用量のレベルとの関係、及び疫学研究でバイオマーカーを使用する前に、これらの関係に対する内的、外的、タイミングの要因の影響を徹底的に評価する。（ばく露/用量に関する）バイオマーカーを適切に解釈するために必要な重要情報は、その後、研究の一部として収集され、慎重に評価されるべきである。ばく露の各バイオマーカーの経費と利益は、疫学評価の一部として慎重に検討され、解釈されるべきである。

マトリックスの選択はばく露研究や疫学研究に不可欠な要素であり、測定能力、汚染の問題、ばく露や健康影響と標的分析物との関連性など、複数の要因を考慮しなければならないことに注意することが重要である。BEES-C では、これらの問題を個別に扱う。

**3.1.1.1.1. 短寿命化学物質の例** ビスフェノール A（BPA）は、遊離形（親化合物）、硫酸抱合体またはグルクロン酸抱合体、または遊離形と抱合体の組み合わせ（総 BPA）として尿中に存在する（Harthéら、2012年；LaKindら、2012年a；Völkelら、2008年；Yeら、2005年）。最近のいくつかの研究では、BPA ばく露に関連した内分泌関連の健康影響が調査されている。最も生物学的に有用性の高いバイオマーカーは遊離（親化合物）BPA であるが、これは親化合物 BPA のみがエストロゲン活性を有すると考えられているためである（EPA、2013年；WHO、2011年）。尿中の遊離 BPA の定量は、非抱合体の BPA がごく一部しか存在しないため、分析が困難である（Yeら、2005年）。この限界を考えると、抱合体または総 BPA の測定が遊離 BPA の有用な代替となる可能性がある。特に、個人内と個々人の間での遊離 BPA と抱合体 BPA の比率にばらつきが少ない場合（ばく露レベルのばらつきに対して）、抱合体 BPA または総 BPA は、遊離 BPA 及び一般的な BPA ばく露での正確で精密な代替となりうる。この事例は、ばく露とバイオマーカー、異なるタイプのバイオマーカー（親化合物と代謝物のマトリックス）、バイオマーカーと生物学的標的の関係を理解することの重要性を強調している。さらに、個々のばく露のバイオマーカーを選択する際には、トレードオフの可能性のあることを強調している（BPA の場合、生物学的関連性を最適化するために、化学物質の検出能力を犠牲にすることがある）。

以下 200 ページ

**3.1.1.1.2. 研究の評価 (表 1)** 標的マトリックス中のばく露の Tier 1 バイオマーカーは、目標用量（既知の標的がある仮説駆動型の研究の場合）または外部ばく露（既知の標的がない研究の場合）の正確で精密な代替である。Tier 2 のバイオマーカーについては、特定マトリックス中のバイオマーカーと外部ばく露、内部ばく露、または目標用量との間に関係があることを示す証拠が存在する。標的マトリックス中の Tier 3 バイオマーカーは、ばく露/用量の低い代替（精度と正確さが低い）である。

3.1.1.2. 効果のバイオマーカー 疫学研究では、短期的なバイオマーカー測定値と長期的な健康影響の意味のある比較することが困難な場合がある。特に横断的な研究では、現在のバイオマーカーレベルが疾患発症に関連する時間帯の過去のばく露を再現していることが重要な前提となる。効果のバイオマーカーのレベルは、疾患発症前の重要な時間帯に、対象となる集団におけるばく露と反応の関係を評価する手段となる。調査結果は、ばく露のバイオマーカーとばく露効果、及びばく露効果のバイオマーカーと有害性健康影響の関連性の強さに基づいて解釈される。

ばく露事象から健康への有害性影響への進行は、有害性転帰経路（AOP）を用いて定義することができる（Ankley ら、2010 年）。特定の健康影響の AOP は、体内の標的における分子的な開始イベントから始まる。分子レベルでの効果は、ばく露イベントによって開始され、細胞レベル、体内液レベル、臓器レベルでの効果へと進行し、最終的には生体全体へと効果を及ぼすことになる。「キーイベント」とは、AOP に沿った中間的なステップであり、AOP に沿った進行を評価するために実験的にモニターすることができる。これらの重要な事象を、生体から採取可能な媒体を用いて測定することを、バイオマーカーと呼ぶ。バイオマーカーは、特定の有害性発現事象に関連する生物学的機能を反映しているため、理想的な効果をもつバイオマーカーと考えられ、「個人または集団における潜在的有害性影響の予測で信頼性が高い」（[www.epa.gov/pesticides/science/biomarker.html](http://www.epa.gov/pesticides/science/biomarker.html)）とされている。一方、「未確定の結果」に分類される効果のバイオマーカーは、変化を特定疾患の発現事象への結びつける経路があまり確実ではないことを反映している（[www.epa.gov/pesticides/science/biomarker.html](http://www.epa.gov/pesticides/science/biomarker.html)）。したがって、バイオ指標の代わりにこれらのバイオマーカーを使用した場合、個人または集団の発現事象の予測は、確実性が低くなる。

3.1.1.2.1 研究の評価（表 1） Tier 1 効果のバイオマーカーは、AOP における重要なイベントのバイオ指標である。Tier 2 の効果のバイオマーカーは、健康影響との関係があることが示されているが、作用機序は解明されていない。結果が不確定な効果のバイオマーカーは、Tier 3 と考えられる。

### 3.1.2. 特異度

単一のばく露のバイオマーカーが複数の親化合物に由来する場合があります。親化合物へのばく露の評価は困難である（Barr and Needham、2002 年; Barr ら、1999 年、2006 年）。ばく露評価や疫学研究の解釈において、親化合物の毒性や作用機序が異なる場合に特に問題となる。さらに、親化合物へのばく露評価を阻害する事例として、代謝物の一つが環境中にも存在する場合（外的ソース）がある。

3.1.2.1. 短寿命化学物質の例 3-フェノキシ安息香酸（3PBA）は、研究の質を評価する際に特異度の評価の重要性を強調する短寿命化学物質の一例である。3PBA は、少なくとも 18 種類の合成ピレスロイドの代謝物であり（Barr ら、2010 年 ; Leng ら、1997 年）、3PBA 環境分解物である 3-フェノキシベンジルアルコールの潜在的な代謝物でもある。尿中 3PBA 測定は、神経毒性が知られていない環境分解物へのばく露に加えて、神経毒性の程度が異なる複数の殺虫剤へのばく露を表している（Barr ら、2010 年）。尿中 3PBA 測定は、ピレスロイドへのばく露の保守的な推定値を提供することができるが、追加のばく露データがない場合、ピレスロイド系殺虫剤へのばく露に関連した神経毒性作用を正確に推定することはできない。神経毒性とばく露の関係を明らかにすることは、真のばく露量が不明であるため、より困難であると考えられる。

3.1.2.2. 研究の評価（表 1） Tier 1 研究には、1つの親化合物へのばく露に由来するばく露のバイオマーカーが含まれる。Tier 2 研究には、同様の種類の有害性エンドポイントを持つ複数の親化合物に由来するバイオマーカーが含まれる。Tier 3 研究には、有害性エンドポイントの種類が異なる複数の親化合物に由来するバイオマーカーが含まれる。

### 3.1.3. 手法の感度

バイオマーカーは、分析対象のマトリックス中に評価できるほどに存在しなければならない

(Calafat and Needham, 2008 年)。ばく露に関係なく、マトリックス中で頻繁に検出されないバイオマーカーは、結果の有用性が限定的となる可能性があるため、環境疫学研究では望ましくない。

**3.1.3.1. 短寿命化学物質の例** 4 つ以上の環を持ついくつかの多環芳香族炭化水素 (PAH) は、ヒトでの発がん性が疑われているか、または既知の発がん性物質である (例: benzo[a]pyrene)。標準的な分析法 (例えば、GC-MS [ガスクロマトグラフィー/質量分析] または LC-MS/MS [液体クロマトグラフィー/タンデム質量分析]) は、採取可能な媒体 (例えば、尿) 中でのこれらの PAH の代謝物を定量するのに十分な感度を持っていないことが多く (Bouchard and Viau, 1997 年)、疫学調査の妨げとなっている。ナフタレン、フェナントレン及びピレンを含む低分子量の PAH のバイオマーカーは、より高分子量の発がん性化学種の代替として評価されている (Bouchard ら、1998 年; Sobus ら、2009 年; Viau ら、1999 年; Withey ら、1991 年)。これらの代替物質は分析上の限界を克服する手段を提供するが、標的化学種のばく露を反映する能力、PAH 間の共起を評価する能力、ばく露ソースの相関関係に関する情報を評価する能力について十分に評価されなければならない。

**3.1.3.2. 研究の評価 (表 1)** Tier 1 研究手法は、研究の問題に対処するために十分な割合のサンプルから化学物質を検出するのに十分な検出限界を持っている (例えば、研究仮説が母集団濃度の代表値と上端の両方の推定を必要とする場合、50~60% の検出値を有する) (Barr ら、2010 年; Zota ら、2014 年)。この要素は Tier 2 にはない。Tier 3 の研究では、検出頻度が低すぎて研究仮説に対応できない。

#### 3.1.4. バイオマーカーの安定性

バイオマーカーは、保存及び使用の時間にわたって、所定のマトリックス中で安定でなければならない (Barr ら、2005 年 a)。サンプルの安定性は記録されるべきである。凍結/融解サイクルを経たサンプルを用いた研究では、それらのサンプルの安定性を実証すべきである。試料の採取から測定までの時間を記録すべきである。

**3.1.4.1. 短寿命化学物質の例** 難分解性有機汚染物質は、通常、 $-20^{\circ}\text{C}$  以下で凍結した場合、血液製剤中では永久に安定であるが、難分解性でない化学物質は血液中での安定性が低い場合がある。例えば、現在使用されている殺虫剤は非常に反応性が高く、血液中の酵素で容易に分解される (Barr ら、1999 年)。EDTA を事前に添加した血液は、エステラーゼ活性を最小限に抑えるが、測定は採取後数ヶ月以内に行う必要がある。解凍/再凍結の過程や熱湯での試料の解凍も分解の原因となる。長期間保存された尿または血液サンプルを使用することで、背景的に収集されたサンプル (NHANES III サンプルなど) のデータが得られるが、多くの場合、解凍/再凍結の過程が発生し、敏感な化学物質の劣化やサンプル自体の汚染が生じる可能性がある。背景的なサンプルの安定性を確認するために、1 つのサンプルを小量で複数に等分して保存する必要がある。バイオマーカーの損失は、容器の壁との結合や揮発によっても発生する可能性がある。

以下 201 ページ

プラスチック容器はガラスに比べて安価で取り扱いや凍結が容易であるが、一部の化学物質の汚染源になることがある。また、金属や有機化合物を吸収してしまうため、化学物質の濃度を過小評価してしまうこともある。保存中や解凍-凍結の過程で損失がないことを検証するために、実際のサンプルに含まれると予想されるレベルと一致するレベルのスパイクされたマトリックスを使用した保存研究や、保存前のサンプルに安定同位体標識された化合物を添加することが必要である。

**3.1.4.2. 研究の評価 (表 1)** Tier 1 研究には、既知の履歴と記録された安定性データを持つサンプルが含まれる。Tier 2 の研究では、保存中の損失は判明しているが、低量ばく露と高量ばく露の違いを定性的に評価することができる (すなわち、研究の目的上、試験参加者を低量ばく露

か高量ばく露かのどちらかに分類することは十分に可能である)。Tier 3の研究では、目的とする分析物の履歴が不明なサンプル、または安定性データがないサンプルを使用している。

### 3.1.5. 試料の汚染

このBEES-C評価基準は、至る所に存在する(ユビキタス)短寿命化学物質を測定する研究を評価する上で最も重要な基準の一つである。これは、汚染を避けるために多大な時間をかけているにもかかわらず、これらの化学物質の多くについて、採取時から測定時までの間に試料が汚染される可能性があることが実証されているからである。生理的半減期の短い化学物質は、環境的に至る所に存在するもの(ユビキタス)であるだけでなく、疫学研究で使用されるサンプリングや分析機器にも含まれている可能性がある。したがって、試料採取から試料測定までの研究のすべての段階において試料汚染を回避/防止するためには、細心の注意が必要である(Barrら、1999年; Calafat and Needham、2008年、2009年; Needhamら、2007年)。試料採取中に、標的の化学物質を含む物や、採取材料やマトリックスを環境媒体(空気や水など)にさらすと、測定された濃度が誤って上昇する可能性がある。予防措置を講じたとしても、研究では分析物の汚染が問題であることが報告されており、研究結果の解釈に不確実性をもたらしている。

3.1.5.1. 短寿命化学物質の例 Yeら(2013年)は、最善の努力をしたにもかかわらず、米国疾病予防管理センター(Centers for Disease Control Prevention)の研究室のサンプルがトリクロサンで汚染されていたことを指摘している。同様に、いくつかの研究グループは、このような汚染を避けるために細心の注意を払っているにもかかわらず、(溶剤や試薬を含む)汚染が原因で、血液サンプル中のBPAを測定することが困難であったことを指摘している(Calafatら、2013年、Markhamら、2010年、Teegardenら、2011年、Yeら、2013年)。

3.1.5.2. 研究の評価(表1) Tier 1研究では、サンプルが収集時から測定時まで汚染されていないことを保証する(例えば、分析物を含まない保証された収集用品や標準物質を使用したり、現場と実験室の両方でブランクを適切に使用したりする)。研究では、研究データが信頼性が高く正確であることを保証するために必要な手順を記録したものを含む。これらの手順を使用していない、または記録していない研究はTier 2に分類される。Tier 3の研究では、既知の汚染問題があり、その問題に対処したという記録がない。

### 3.1.6. 手法の要件

ばく露を評価するためのバイオマーカーの品質は、測定に使用される手法の質に大きく依存する。これは、バイオマーカー測定の評価が難しい側面となる。例えば、習熟度テスト練習への研究室参加と成果は、Tier 1研究のための合理的なテストのように見えるかもしれないが、多くの習熟度テスト研究では、200%変動する可能性のある許容範囲がある(すなわち、「許容できる」分析物濃度値は、真の値の±200%になる)。一般的に、研究手法は適切な機器を備え、付随する手順(例えば、QC、手法の強固性、確認イオンの存在、同位体希釈の使用)を記述しておく必要がある。

3.1.6.1. 研究の評価(表1) Tier 1研究では、バイオマーカーの明確な同定と定量を必要な感度で提供する機器が含まれている(例: GC-HRMS [ガスクロマトグラフィー/高分解能質量分析]、GC-MS/MS、LC-MS/MS)。Tier 2研究では、高度な信頼性と必要な感度でバイオマーカーの同定を可能にする機器を使用する(例: GC-MS、GC-ECD [ガスクロマトグラフィー-電子捕足検出器])。Tier 3の研究では、バイオマーカーの定量化のみを可能にする機器を使用するが、その手法には既知の干渉物質が含まれている(例えば、GC-FID [ガスクロマトグラフィー-炎イオン化検出器]、分光器)。

### 3.1.7. マトリックスの調整

バイオマーカーは濃度の単位で最も一般的に測定され、報告される。尿のバイオマーカー濃度は、尿量変動(食事、運動、水分補給、年齢、病状などによって引き起こされる)によって強く影響を受け、血液中のバイオマーカー濃度は、血液量と脂肪含量によって強く影響を受ける。尿

のバイオマーカー濃度は、クレアチニン濃度（筋肉内でのクレアチン・リン酸分解に由来する）、比重、尿量、及びその他の手法を用いて、被験者間及び被験者内で尿希釈の変動を補正するために標準化されている。しかし、補助情報のないスポット試料の未補正の尿中濃度は、一般に報告され、ばく露や健康影響との関係の評価に利用される（Barr ら、2005 年 b; LaKind and Naiman、2008 年、2011 年; Lorber ら、2011 年; Meeker ら、2005 年）。尿希釈の変動に対する尿のバイオマーカー測定値を「補正」するための最良の手法については、現在のところコンセンサスはない。最低限、研究間での適切な比較を可能にするために、容量ベースの濃度と補正された（クレアチニンまたは他の手法）濃度の両方が提供されるべきである。また、全排尿量と排尿間隔時間を取得することも有益である。

血液ベースのバイオマーカーレベルは、全血、血清、血漿、及び脂質調整値として報告されている。脂質調整を決定するために使用した手法、または血液中の脂質の異なる成分を分離するために使用した手法を提供し、入手可能な場合には、すべての濃度を報告すべきである（例えば、全体量及び脂質調整値）。同様に、親油性化学物質の測定における絶食サンプル及び血清脂質調整に関連する問題も考慮しなければならない（Schisterman ら、2005 年）。脂質やその他の組織成分の調整の妥当性は、現在使用されている農薬のような特定の短寿命化学物質については確立されていない。このような場合には、調整の妥当性が確立されていないことを明記して、全量濃度と調整濃度を報告すべきである。さらに、血漿量は妊娠中に増加し（また、いくつかの既往症や健康状態でも増加する可能性がある）、妊娠前や集団間で血漿濃度を比較する際にも考慮する必要があるかもしれない（Hyttén、1985 年）。

サンプル収集要件とマトリックス処理に関する情報は、研究間のデータを比較したり、基準範囲と比較したりする際に重要である。異なる政府機関（例えば、欧州連合、特定の欧州諸国、米国 NHANES、カナダ健康測定調査、欧州規模でヒト・バイオモニタリングを実行する共同体、州ベースの HANES）及び他の大規模なバイオモニタリングデータ収集による研究では、マトリックス及び報告されたバイオマーカー濃度を変化させる可能性のあるサンプルの収集及び処理のための異なるプロトコルを持っている可能性がある。例えば、サンプル採取前の絶食についての参加者への指示は、血液中の脂質含有量を最小化することができるため、サンプル中の親油性バイオマーカー濃度を最小化することができる（Barr ら、2005 年 a）が、これらの指示は国によって必ずしも同じではない（LaKind ら、2012 年 a）。同様に、朝の起床直後の尿では、単純なスポットサンプルよりもマトリックス成分の濃度が高い可能性があり、分析対象物を検出または識別する能力が変化する可能性がある（Kissel ら、2005 年; Scher ら、2007 年）。

以下 202 ページ

さらに、朝の起床直後の尿収集は、以前のばく露とサンプルの収集・測定の間関係によってデータにバイアス（システムティックエラー）をもたらす。抗凝固剤として EDTA とヘパリンを用いて採取した血漿は、マトリックスの特性を変化させる可能性がある（Barr ら、2005 年 a）。採取要件及びサンプル処理の違い（及びバイオマーカー濃度に影響を与える可能性のある腎臓病などの研究参加者の健康状態）は、研究間で結果を比較する際に報告され、考慮され、それに応じて重み付けされる必要がある。

**3.1.7.1. 研究の評価 (表 1)** マトリックス調整のための最良の手法は、調査する仮説と目的の特定化学物質が密接に関連していることであるが、この分野でのコンセンサスはまだ得られていないことを認識している。ただし、調整は研究結果に大きな影響を与える可能性がある。したがって、Tier 1 研究では、調整濃度と非調整濃度（調整が必要な場合）の結果を提供し、それによってマトリックス調整の影響について読者が独自の結論に達することができるようにすることを提案する。Tier 2 研究は、1 つの手法（マトリックス調整の有無にかかわらず）を用いた結果のみを提示する研究である。Tier 3 の研究には、有効な調整手法がまだ確立されていないマトリックス中の化学物質の測定が含まれている。

## 3.2. 研究デザインと実行

研究デザイン、ばく露のばらつきと誤分類の両方を考慮することは、短寿命化学物質では特に重要である。

### 3.2.1. 疫学研究のデザイン

短寿命化学物質に関するバイオモニタリングデータと疾患との関連性を探る研究は、バイオマーカーの血中または尿中のレベルが、通常数時間前または数日前に発生した最近のばく露を再現しており、バイオマーカーのサンプル採取に関連したばく露のタイミングは通常知られていないため、独特の問題を抱えている。しかし、目的的健康影響の多くは、発症に数年から数十年を要する慢性疾患（例：肥満、高血圧、生殖機能の測定値）である。このため、短寿命化学物質を測定する研究における因果関係仮説の評価は複雑であり、状況によっては実行不可能な場合もある。因果関係の決定的で異論のない特性は時系列性であり、発現事象に先行する因果ばく露の観察によって裏付けられなければならないということである（Potischman and Weed、1999年；Rothman and Greenland、2005年；Weed、1997年；Weed and Gorelic、1996年）。

時系列性を確立することができるのは、発症時の新たな疾患症例や、ベースラインと比較した健康関連指標の変化など、健康関連のイベントを特定する「発生率」研究のみである（Pearce、2012年）。発生率研究には、介入研究（例：臨床試験）と観察研究（コホート研究や症例を確認する症例対照研究）がある。しかし、どのようなデザインであっても、発生率研究の主な特徴は、疾病の発症時期（または少なくとも診断時期）を特定できることであり、それによってばく露と発現事象の順序を評価できる可能性がある。ばく露レベルが時間の経過とともに急速に変化する場合、有用なアプローチは、ばく露の反復測定値と健康バイオマーカーの反復測定値との関係性を評価する縦断研究である。

因果関係を評価するためには時間的関係を確立する能力が重要であるが、環境疫学研究ではばく露と研究の発現事象との間の間隔が別の研究デザインの問題となっている。ヒトのバイオモニタリングデータを病因研究に利用するためには、ばく露は疾患発症に関連する時期に測定されなければならない。これは簡単な作業ではないが、関連する時間帯に難分解性化学物質のばく露を調査し、それらのばく露と特定の有害性発現事象の発生とを相関させたバイオモニタリング研究の成功例がある。

例えば、鉛の血中濃度は、過去5～6週間のばく露を反映しており、適切に実行された疫学研究では、子供の血中濃度と認知能力への有害性影響を関連付けることができた（Lanphearら、2000年）。しかし、半減期の短い化学物質の場合、関連するばく露と疾患発症の間隔を評価することは困難である。研究デザインは、後述するばく露の誤分類とともに、短寿命化学物質のバイオモニタリング研究において最も重要であり、まだ十分に調査されていない側面である。

「有病率」研究では、「罹患」研究に比べて時系列性の立証が困難であり、因果関係についての結論を出すことが困難である。典型的な有病率調査は、ばく露と疾病情報を同時に把握する横断的デザインに依存している（Rothman and Greenland、1998年）。短寿命化学物質に焦点を当てた研究の場合、多くの症例対照研究は、たとえ症例を用いていたとしても、バイオマーカーレベルが病気の発症に先行するのではなく、最近のばく露を反映しているため、解釈は困難である。特筆すべき例外は、コホート内症例対照研究やケース・コホート研究で行われているように、将来の使用のために採取したサンプルを使用する研究である（Gordis、2008年）。

**3.2.1.1. 短寿命化学物質の例** フタル酸代謝物と肥満、糖尿病、心血管疾患との関連性に関する最近の疫学文献のレビュー（Goodmanら、2014年）において、研究の大部分は横断的デザインであった。研究結果は発現事象の間で一貫性がなく、フタル酸代謝物への事前ばく露とそれに伴う健康影響との関係を識別する能力における重要な制限要因として、時系列性の欠如が考えられた。

**3.2.1.2. 研究の評価（表1）** Tier 1研究は、追跡期間または反復測定値の縦断分析を含む発生率研究であり、ばく露と発現事象の間の時間順序と関連する間隔の両方が確立される（表1）。Tier 2研究には、ばく露が発現事象に先行する発生率研究が含まれるが、ばく露の特定の関連範

囲は考慮されない。最も情報量の少ない (Tier 3) 研究は、現在のばく露 (例: 化学物質の血中レベル) と、急性ばく露ではなく慢性ばく露に関連している可能性が高くて頻繁に測定される発現事象 (例: BMI) との関連を検討する研究である (この評価基準は、集団内または集団間の時間的または空間的な関係を調査する研究といったばく露のみに焦点を当てた研究には適用されないことに注意する。)

### 3.2.2. ばく露のばらつきと誤分類

多くの短寿命化学物質については、個人内での時間的変動が大きい場合があり、そのような化学物質と疾患との間の関連を特定しようとすることは支持できない。個人の食生活、健康状態、製品の使用、活動または場所の変化による短寿命化学物質のバイオモニターされたレベルの違いが予想される (Pleil and Sobus, 2013 年)。Meeker ら (2013 年) が指摘しているように「ばく露測定基準における時間的変動を特徴づけることは、特に非難分解性化合物のバイオマーカーにおいて、ばく露測定誤差の可能性に関連する疫学研究をデザインし解釈する上で重要なステップである。」

慢性ばく露または平均的ばく露を推定しようとする短寿命化学物質の多くの発表された研究は、尿または血液の一回限りのサンプルを使用したばく露の一つの尺度に依存しているため、誤差が生じる可能性がある (Goodman ら、2014 年; LaKind ら、2012 年 b、2014 年; Preau ら、2010 年; Wielgomas、2013 年)。バイオマーカーレベルの時間的変動を平均化するために、同じ個人から異なる時間に複数のサンプルを採取することで、ばく露を推定する能力を向上させることができる (NRC、2006 年)。信頼性は通常、クラス内相関係数 (ICC) を計算することによって測定される。ICC は、数時間、数日または数週間に渡って反復収集されたサンプル中の化学物質を測定し、個々人の間の分散を総分散で割った値を計算することで推定できる。

以下 203 ページ

ICC の値は 0 から 1 の範囲であり、1 に等しいまたは 1 に近い ICC 値は、単一のサンプルから集団の長期ばく露を推定する際の信頼性が高いことを示唆している。Symanski ら (1996 年) は、職業ばく露における安定しない行動を考慮するために混合効果モデルを使用し、ばく露における系統的な変化が適切にモデル化されていない場合、分散成分 (ICC の計算に使用される) の推定値が大幅に偏る可能性があることを発見した。「もし ICC が数週間または数ヶ月間の反復サンプルを採取して作成された場合、その値は目的の多くの慢性疾患の発症の時間枠である数年間のばく露に関連するか?」という疑問について提起されなければならない。この問題については、標的の短寿命化学物質の多くについて現在のところ研究が進んでいない。

短寿命化学物質の単一の測定値を使用する場合のもう一つの問題は、ばく露の誤分類を引き起こす可能性のあるエラーである。ばく露の誤分類は、割り当てられたばく露が実際のばく露レベルやカテゴリーを正しく再現していない場合に起こる。ばく露の誤分類は、方向性と大きさの両方の観点から予測することが困難であることが示されている (Cantor ら、1992 年; Copeland ら、1977 年; Copeland ら、1977 年; Dosemeci ら、1990 年; Sorahan and Gilthorpe、1994 年; Wacholder ら、1995 年)。ばく露誤差とばく露誤分類の用量反応関係への効果は問題である (Rhomberg ら、2011 年)。ばく露の誤分類は、重大なばく露が発生する際に関連するサンプル収集のタイミングを含め、測定誤差の多くのソースから発生する。例えば、多くの揮発性有機化合物は半減期が数分のオーダーである。すなわちばく露は毎日発生しても、短い時間間隔で発生する。このように、ばく露のバイオマーカーの濃度は、ばく露が発生した際に関連してサンプルが採取された時期に大きく依存しており、体内の長期的なレベルを適切に反映しない場合がある。

複数のサンプルの使用または長時間 (例: 24 時間) のサンプル収集は、時間的変動、研究の亜集団の特性、及びサンプルに関連する問題の効果を軽減することにより、誤差を減少させるのに役立つかもしれない (Scher ら、2007 年)。誤差が回避できない場合 (例: 利用可能なすべてのサンプルが絶食後に得られた場合) は、感度と特異度を計算することにより、ばく露特性の精度を評価することが重要である (Jurek ら、2006 年)。感度とは、もしも個人が本当に高いばく露カテゴリーに属しているならば、高レベルのばく露を受けた個人と正しく分類できる確率で



ある。特異度とは、本当に低レベルのばく露を受けた参加者を正しく低レベルばく露と割り当てる確率である。感度と特異度の推定値は、多くの測定値に対する真の感度と特異度は不明であるので、被験者ごとに複数のサンプルを絶対的基準として使用して1つの尿サンプルについて計算される。これは、研究期間中に収集された各個人の反復サンプルの中から無作為に1つのサンプルを選択することによって実行される（2008年に Adibi らによってフタル酸エステル類について実証されたように）。

**3.2.2.1. 短寿命化学物質の例** Goodman ら（2014年）は、フタル酸エステル類と肥満、糖尿病、心血管疾患との関連性に関する疫学文献の最近のシステムティックレビューにおいて、利用可能な26件の研究のうち、3件を除くすべての研究がフタル酸エステル類の単一の測定値に依存していることを明らかにした。同様に、BPAと肥満、糖尿病、心血管疾患に関するシステムティックレビューでは、LaKind ら（2014年）は、利用可能な45件の研究のうち、4件を除くすべての研究がBPAの単一測定値に依存していたことを明らかにした。しかし、BPAの個人内変動は大きく（ICCは0.10～0.35の範囲）（Lassen ら、2013年；Teitelbaum ら、2008年）、1人の長期ばく露を記述するには複数のばく露尺度が必要である。フタル酸エステル類のICCはBPAよりも高いことが報告されているが（例えば、フタル酸モノエチルは0.18～0.61、フタル酸モノイソブチルは0.21～0.51、フタル酸モノ（2-エチルヘキシル）は0.08～0.27（Goodman ら、2014年でレビュー）、個人間のばらつきは依然として大きい。最近、Attfield ら（2014年）は、小児における尿中農薬測定値の変動性に関する研究において、各研究参加者からのサンプル数が少ない研究では、「...不正確な四分位分類によるばく露の誤分類の可能性が高く、ばく露を正確に特定のカテゴリーに分類するための保証はほとんどない」と述べた。

**3.2.2.2. 研究の評価（表1）** 上記の考察により、利用可能な文献を以下の段階に分けることができる（表1）。Tier 1には、ばく露評価が、適切な期間にわたるばく露量を推定するために、一人当たりのサンプル数が十分な数であるか、または適切な長期サンプリング（例えば、複数の24時間尿採取）を使用している研究が含まれている。Tier 1に含まれるためには、研究は、精度（例：感度及び特異度）及び信頼性（例：ICC）の測定値を計算することによって誤差を評価すべきである。化学物質によっては、ばく露を完全に特徴づけるには1つのサンプルで十分である可能性がある。このような場合、Tier 1研究では、単一の測定の誤差が十分に小さいと考えられるという証拠を提供する必要がある。これは常に実現可能なことではないが、研究者がバリデーション研究を実行する必要があると考える状況があることは理解している（Teegarden ら、2011年）。Tier 2には、2つ以上のサンプルを使用する研究が含まれるが、測定数の選択に根拠を示さず、誤差の明示的な評価も含まれていない。Tier 3は、誤差を考慮せずに単一のサンプルに基づいてばく露評価を行う研究のためにある。

### 3.3. 一般的な疫学研究デザインの考察

このセクションでは、必ずしも短寿命化学物質に特化したものではないが、研究全体の質を評価する上で重要な研究デザインの側面について議論する。これらの問題のいくつかは、ばく露と健康影響との関連性を検討する研究に適用できるものもあれば、ばく露のみに焦点を当てた研究に適用できるものもある。

#### 3.3.1. 研究の論拠

このセクションでは、バイオモニタリングデータと健康影響データの間に関連性を検証する仮説検証研究を述べる。臨床観察または基礎科学実験から生じる十分に練られた仮説は、研究の種類にかかわらず、疫学調査の基礎となるものである（Boet ら、2012年；Fisher and Wood、2007年；Moher and Tricco、2008年）。現在、様々な分野で推奨されているのは、目的の集団、調査対象のばく露（またはそれに対応するマーカー）及び懸念される結果に関する情報を伝えるように構成された調査課題を提起することの重要性を強調している（Sampson ら、2009年；Walker ら、2012年）。

バイオモニタリング研究、特に1つのサンプルが多数の化学物質に関するデータを提供できる

ような短寿命化学物質を対象とした研究では、多くの場合、複数の変数を含むデータを生成し、複数の仮説検証を同時に行う機会を得ることができる。バイオモニタリング研究のこの特徴は、いくつかの関連する発現事象に対して有意な関連性が観察された場合（Lord ら、2004 年）のような状況では、強みとみなすことができる。例えば、仮説を立てられた肥満因子が体重指数、ウエスト周囲長、または体脂肪率に同様の影響を及ぼす場合である。一方で、複数のばく露-発現事象の関連を評価する能力は、特に過去に収集されたデータを扱う場合、結果の解釈を複雑にする（Clarke ら、2003 年; Lee and Huang、2005 年; Marco and Larkin、2000 年）。過去に収集されたデータを使用した研究の中で、優先的に定式化された仮説に導かれた研究と、強い生物学的根拠を持たずに実行された研究を区別することが重要であるが、後者のカテゴリーは新しい仮説の策定に役立つことが証明されている（Liekens ら、2011 年; Oquendo ら、2012 年）。仮説が十分に定式化されている研究は、その研究がこれまでの知識の上に築かれていることを示しており、これは WOE 評価にとって重要な考慮事項である。既存の知識ベースに追加するために特別にデザインされた研究は、より容易に WOE に組み込むことができる。

**3.3.1.1 研究の評価 (表 1)** 優先的に定式化された仮説を、この仮説に対処するために特別に設計されたバイオモニタリング戦略で評価する研究は、最高品質 (Tier 1) と考えるべきである。Tier 2 の研究とは、優先的に定式化された仮説を評価するために既存のサンプルやデータを使用した研究であり、この目的のために特別にデザインされたバイオモニタリング戦略ではない。

以下 204 ページ

Tier 3 の研究では、事前に定められた仮説を持たずに既存のサンプルやデータに依存する研究または複数の同時仮説検証を含む。現在のところ、短寿命化学物質を対象としたほとんどのバイオモニタリング研究の研究理論拠は、Tier 3 研究として記述されることになることと認識している。

### 3.3.2. 研究参加者

研究参加者の選択に関する評価方法は、難分解性化学物質と短寿命化学物質の両方の研究に適用される。疫学研究における参加者選択の目的は、サンプルから得られる情報と対象となる集団について求められる情報との間を「仲立」することである（Kalsbeek and Heiss、2000 年）。偏りのない集団サンプルを選択する実際のプロセスは、症例対照研究、縦断（コホート）研究、横断研究において継続的な問題である（Vandenbroucke ら、2007 年）。

参加者選択の問題は、短寿命化学物質の疫学研究に特有のものではない。しかし、バイオモニタリング研究はこの問題に十分な注意を払っていない。これまでのバイオモニタリング研究のレビューでは、選択バイアスが内的妥当性に対する重要な脅威となる可能性があるという証拠が示されている（Bull ら、2006 年; Faust ら、2004 年）。同じ問題は、フタル酸エステル類のような短寿命化学物質のバイオモニタリング研究にも当てはまる（Durmaz ら、2010 年; Wang ら、2013 年; Wirth ら、2008 年）。

**3.3.2.1. 研究の評価 (表 1)** Tier 1 研究には、偏りのない選択またはフォローアッププロトコルが含まれており、横断研究または症例対照研究では高い（例：80%以上）応答率、コホート研究では低い（例：20%未満）脱落例が認められる。Tier 2 の研究は、偏りのない選択/フォローアッププロトコルを有し、横断研究または症例対照研究では低い（例：50%~80%）、応答率、コホート研究では高い（例：20%~50%）脱落例が認められる。Tier 3 の研究とは、望ましい参加者の数が 50%未満であるか、サンプルの選択方法、無応答率または脱落例率を報告していない研究である。この情報が報告されていない研究は、Tier 3 研究とみなされるべきである。

低い応答率や高頻度の脱落例を選択バイアスと同一視すべきではないことを念頭に置いておくことが重要である。選択バイアスは、最終のデータセットに含まれる個人の割合（別名：選択確率）がばく露と発現事象の両方で異なる場合に起こる（例えば、ばく露群の発症例、非ばく露群の発症例、ばく露群の非発症例、非ばく露群の非発症例の間で）。実際の選択確率は通常不明であるが、望ましい参加者の 10%しか脱落していない研究では、バイアスの大きさは、被験者の 50%以上が脱落している研究でのバイアスの大きさよりもはるかに小さいと予想できる。

### 3.3.3. データ分析

疫学研究におけるデータ分析に必須な側面は他で検討されており、生理学的半減期が短い化学物質には特化したものではない。しかし、提案されている段階的評価システムの完全性のために、これらの考慮事項を簡単に説明する。観察研究における全体的な分析戦略は、研究の主な目的によって決まる。一般的に、統計モデルは予測分析と説明分析の 2 つのカテゴリーに分類される (Shmueli, 2010 年)。予測分析では、モデルへの変数の選択はデータに基づいて行われ、データセットごとに異なる場合がある。このアプローチの目標は、モデル適合を最大化することであり、目的の特定の共変量を保持するかどうかの決定は、目的の特定のばく露なしで、統計的検定と適合良好に基づいて行われる (Bellazzi and Zupan, 2008 年)。説明的 (仮説検証) 分析では、発現事象とリスク因子の関係が交絡している場合に、潜在的に重要な変数を誤って排除したり、交絡因子として作用しない変数を間違えて保持したりする可能性があるため、このアプローチは不適切であるかもしれない (Kleinbaum and Klein, 2002 年)。

事前に定義されたばく露と発現事象の関係に焦点を当てた説明モデルでより重要なことは、コントロール変数 (交絡因子、媒介因子、または効果修飾因子) の包含と排除は、少なくとも部分的には、先験的な理由に基づいて行われるべきである (Beran and Violato, 2010 年; Concato ら, 1993 年; Hernan ら, 2002 年)。

観察研究の結果は必然的に不確実性の影響を受けることを念頭に置くことが重要である。この不確実性は、原因不明のバイアスの様々なソースや、様々なデータ処理の決定や仮定に起因している可能性がある。不確実性の大きさは、定量的な感度分析によって正式に評価できる。感度分析を通じた残留バイアスへの対処方法は、基礎的な理論 (Greenland, 1996 年) と実践的な応用の両面で十分に発達している (Goodman ら, 2007 年; Lash and Fink, 2003 年; Maldonado ら, 2003 年)。代替的な決定や仮定の感度分析に関しては、経済学、ばく露評価、定量的リスク分析の過去の経験から多くを学ぶことができる (Koornneef ら, 2010 年; Leamer, 1985 年; Spiegelman, 2010 年)。

**3.3.3.1. 研究の評価 (表 1)** Tier 1 の研究には、因果関係モデルと予測モデルを明確に区別し、効果修飾と交絡因子の調整を評価して外的要因を十分に考慮していることを論証することが含まれる。Tier 1 の資格を得るためには、研究は正式な感度分析も行うべきである。外部要因の考慮が十分であり、モデル選択が適切であると考えられる場合でも、感度分析のない研究は不完全とみなされる可能性がある。このような研究は Tier 2 に分類される。Tier 3 の研究は、共変量の選択方法が不適切であったり、重要な交絡因子を考慮していなかったり、効果修飾を考慮に入れていなかったりするために、外部要因を適切にコントロールできなかった研究である。

「外部要因」という用語は、対照となるばく露や発現事象以外の参加者の特性を示しており、それらは交絡因子または効果修飾因子あるいはその両方として作用するので、研究のデザインまたは解析の段階で考慮する必要がある (Kleinbaum ら, 2007 年)。

### 3.3.4. 結果報告

報告の 3 つの側面 (透明性、複数のテスト、報告の偏り) を考慮する。

**3.3.4.1. 報告の透明性** STROBE 声明で述べられているように、結果の報告は「観察研究において何が計画され、何が行われ、何が発見されたかを明確に提示」 (Vandenbroucke ら, 2007 年) すべきである。これらの考慮事項はすべての研究に当てはまるが、短寿命化学物質のバイオモニタリング研究には特に関連性のある研究報告の側面がある。

生物学的サンプル分析は、1 回の分析で複数の分析物を迅速に分析できるように、ますます最適化されてきている。このような技術の発展により、ばく露 (及び該当する場合は発現事象) バイオマーカーの完全なリストを含むデータの完全な報告の重要性が増してきており、代表値や分散などの要約統計量の提示も重要である。その他の重要な情報要素には、欠落データのパターンとその取り扱い、及び LOD 以下の測定値の説明が含まれるべきであり、これらはすべて研究結果の解釈に影響を及ぼす (Albert ら, 2010 年; Barnes ら, 2008 年; LaKind ら, 2012 年 b)。さらに、研究参加者の数を決定する際に使用された検出力計算や、有意な関連性を特定する能力

に影響を与えるべく露勾配に関する情報も提供されるべきである。これらの情報の一部は、スペースの制約のために論文には含まれないかもしれないが、補足資料に組み入れたり、要望に応じて提供したりすることが可能となる。

**3.3.4.2. 多重検定の問題点** 複数の仮説検定を行う場合の主な懸念は、偽陽性 (FP) の結果となる可能性が高まることである (Boffetta ら、2008 年; Ioannidis、2014 年; Jager and Leek、2014 年; Rothman、1990 年; Sabatti、2007 年)。その他、FP の結果の問題は、対応する偽陰性 (FN) の問題よりも重要ではないとの主張がある (Blair ら、2009 年)。どのタイプのエラー (FP か、FN か) がより大きな懸念をもたらすかの決定は、化学的特性及び発現事象特性であり、ケース・バイ・ケースで行われるべきである。

以下 205 ページ

遺伝疫学及び分子疫学における最近の進歩は、FN 結果のリスクを増加させることなく FP (PFP) の確率を減少させることを目的とした新しいアプローチの開発された (Datta and Datta、2005 年; Wacholder ら、2004 年)。さらに最近になって、これらのアプローチは、さらに FP : FN 比を計算できるまでに及んだ (Ioannidis ら、2011 年)。

**3.3.4.3. バイアス報告** メタ分析や WOE 評価のために一連の研究を評価する際には、利用可能なデータの分析と統合の両方に影響を及ぼす可能性のある 2 つの特定のバイアスのソース、すなわち、出版バイアスと論文報告バイアスを考慮しなければならない。出版バイアスとは、「研究者や編集者が研究結果の方向性や強さに基づいて研究結果を発表しない傾向」と定義されている (Dickersin and Min、1993 年)。これと密接に関連する概念として、選択的研究内報告 (別名：結果報告バイアス) があり、これは「論文に含めるために記録された元の変数の一部の結果に基づいて選択すること」と定義されている (Dwan ら、2008 年)。

出版バイアスは短寿命化学物質を含む研究には特有のものではない。しかし、上記の理由から、短寿命化学物質の研究では、発現事象報告バイアスがより問題になる可能性がある。特に、高性能分析設備が利用しやすくなることで、より多くのサンプル数でより多くの分析物を測定することが可能になっている。

**3.3.4.4. 研究の評価** Tier1 研究では、その目的が明確に述べられており、読者は検証された仮説の数 (結果が与えられた仮説の数だけではない) を評価することができる。複数の同時仮説検証が行われている場合は、その影響が評価し、可能なら PFP または FP : FN 比を推定することで評価を行う。結果報告バイアスの証拠はなく、結論は観察された結果を超えたものではない。Tier 2 の研究では、結論は正当であると思われるが、検証された仮説の数が不明確 (明示されていないか、識別が困難である) であり、及び/または多重性の検討がなされていない。データ要約を選択的に報告し、提示された結果の方法や選択に関して透明性を欠いている研究は Tier 3 に含まれる。

#### 4. 考察／結論

環境疫学研究の質を評価するための体系的なアプローチの必要性は明らかである。評価スキームを開発するための 2 つの先行研究は、環境化学物質ばく露と神経発達に関する疫学研究に焦点を当てたものであった (Amler ら、2006 年; Youngstrom ら、2011 年)。これらの提案されたスキームで提示された概念の多くは、生理学的半減期の短い化学物質のバイオモニタリングを検討する際に、研究の質を評価し、研究結果を伝える上で貴重なものである。例えば、Amler ら、(2006 年) が提示した基本的な最適な実行/基準は以下とおりである。明確に定義された生物学的に妥当な仮説、前向きで縦断的なコホートデザインの使用、研究間での研究デザインプロトコルの一貫性、研究結果がどの程度決定的で一般化可能であるかについての率直で規律正しく知的に正直な扱い、実際の研究課題、検証手順及び研究での発見についての報告の制限、否定的な結果も肯定的な結果も報告する倫理的義務の研究者の認識ならびに結果を最小化したり誇張したり

しないことの重要性。

生理的半減期が短い化学物質は、分析室やサンプリング機器に存在すること、横断研究における時間的順序の確立が困難であること、ばく露とバイオマーカー濃度の短期及び長期の変動性、正確なばく露分類に必要な測定数に関する情報不足など、いくつかの重要な問題を抱えている。BEES-C は、研究や提案の中でこれらの問題を評価するためにデザインされている。

BEES-C のような評価ツールの開発は、単純でもなく、議論の余地のないものでもないことを認識しており、CONSORT やその他の既存方法の一部、あるいは臨床データの質評価であるデータ品質スキームと同様に、これが反復的なプロセスになることを期待している。この種の評価スキームは探索的研究には有用ではなく、むしろ、半減期の短い化学物質へのばく露と有害性健康影響の関連性の理解を深めるために最も有用な研究をデザインし、特定することに焦点が当てられていることにも留意したい。本ワークショップで開発された評価指標が、このテーマについてのさらなる議論のきっかけとなることを期待する。

## 利益相反

本ワークショップは、米国化学協議会 (ACC) の Polycarbonate/BPA Global Group が主催した。ACC は、本ワークショップの設計、運営、開発、原稿の作成、承認には関与していない。ワークショップ参加者、または、その関係機関は、謝礼 (JSL, ES, GS, JS, JT, Y-MT, RT-V, TA を除く) と旅費 (TA, Y-MT, DB, ES を除く) を受け取った。JSL はワークショップの開発とファシリテーションの支援を受けた。MG は定期的に政府や民間セクターのコンサルタントを務めている。その他の利害関係は表明していない。

## 免責事項

ここで表明された見解は著者のものであり、ACC、米国環境保護庁、カナダ保健省、または国立成育医療研究センターの見解を必ずしも代表するものではない。米国環境保護庁は研究開発局を通じて、ここに記載されている研究に協力した。この研究は、米国環境保護庁の審査を受け、出版の承認を受けている。

## 謝辞

本書に記載されている見解は、2013 年 4 月にメリーランド州ボルチモアで開催されたワークショップで発表されたものである。運営委員会には以下が含まれている。運営委員会には、米国環境保護庁国立計算毒性学センターの Elaine Cohen Hubal 博士、Judy S. LaKind 博士、メリーランド大学メリーランド校 LaKind Associates LLC の Judy S. LaKind 博士が参加している。LaKind、博士、LaKind Associates LLC、メリーランド大学医学部及びペンシルバニア州立大学医学部、Enrique F. Schisterman、博士、ユニース・ケネディ・シュライバー国立小児保健・人間発達研究所疫学統計・予防研究部門、国立衛生研究所、及び Justin Teeguarden、博士、DABT、Pacific ノースウエスト国立研究所から構成されている。米国 EPA とカナダ保健省から 3 人の匿名査読者の思慮深いコメントに感謝する。

## 参考文献

- Adibi JJ, Whyatt RM, Williams PL, Calafat AM, Camann D, Herrick R, et al. Characterization of phthalate exposure among pregnant women assessed by repeat air and urine samples. *Environ Health Perspect* 2008;116:467–73.
- Albert PS, Harel O, Perkins N, Browne R. Use of multiple assays subject to detection limits with regression modeling in assessing the relationship between exposure and outcome. *Epidemiology* 2010;21(Suppl. 4):S35–43.
- Amler RW, Barone Jr S, Belger A, Berlin Jr CM, Cox C, Frank H, et al. Hershey Medical Center Technical Workshop Report: optimizing the design and interpretation of epidemiologic studies for assessing neurodevelopmental effects from in utero chemical exposure.

- Neurotoxicology 2006;27:861–74.
- Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, et al. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. *Environ Toxicol Chem* 2010;29:730–41.
- Attfield KR, Hughes MD, Spengler JD, Lu C. Within- and between-child variation in repeated urinary pesticide metabolite measurements over a 1-year period. *Environ Health Perspect* 2014;122:201–6.
- Barnes SA, Mallinckrodt CH, Lindborg SR, Carter MK. The impact of missing data and how it is handled on the rate of false-positive results in drug development. *Pharm Stat* 2008;7:215–25.
- Barr DB, Needham LL. Analytical methods for biological monitoring of exposure to pesticides: a review. *J Chromatogr B Analyt Technol Biomed Life Sci* 2002;778:5–29.

以下 206 ページ

- Barr DB, Barr JR, Driskell WJ, Hill Jr RH, Ashley DL, Needham LL, et al. Strategies for biological monitoring of exposure for contemporary-use pesticides. *Toxicol Ind Health* 1999;15:168–79.
- Barr DB, Wang RY, Needham LL. Biologic monitoring of exposure to environmental chemicals throughout the life stages: requirements and issues for consideration for the National Children's Study. *Environ Health Perspect* 2005a;113:1083–91.
- Barr DB, Wilder LC, Caudill SP, Gonzalez AJ, Needham LL, Pirkle JL. Urinary creatinine concentrations in the U.S. population: implications for urinary biologic monitoring measurements. *Environ Health Perspect* 2005b;113:192–200.
- Barr DB, Landsittel D, Nishioka M, Thomas K, Curwin B, Raymer J, et al. A survey of laboratory and statistical issues related to farmworker exposure studies. *Environ Health Perspect* 2006;114:961–8.
- Barr DB, Olsson AO, Wong LY, Udunka S, Baker SE, Whitehead RD, et al. Urinary concentrations of metabolites of pyrethroid insecticides in the general U.S. population: National Health and Nutrition Examination Survey 1999–2002. *Environ Health Perspect* 2010;118:742–8.
- Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. *Int J Med Inform* 2008;77:81–97.
- Beran TN, Violato C. Structural equation modeling in medical research: a primer. *BMC Res Notes* 2010;3:267.
- Blair A, Saracci R, Vineis P, Cocco P, Forastiere F, Grandjean P, et al. Epidemiology, public health, and the rhetoric of false positives. *Environ Health Perspect* 2009;117: 1809–13.
- Boet S, Sharma S, Goldman J, Reeves S. Review article: medical education research: an overview of methods. *Can J Anaesth* 2012;59:159–70.
- Boffetta P, McLaughlin JK, La Vecchia C, Tarone RE, Lipworth L, Blot WJ. False-positive results in cancer epidemiology: a plea for epistemological modesty. *J Natl Cancer Inst* 2008;100:988–95.
- Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Fam Pract* 2004;21:4–10.
- Bouchard M, Viau C. Urinary excretion of benzo[a]pyrene metabolites following intravenous, oral, and cutaneous benzo[a]pyrene administration. *Can J Physiol Pharmacol* 1997;75:185–92.
- Bouchard M, Krishnan K, Viau C. Kinetics of tissue distribution and elimination of pyrene and 1-hydroxypyrene following intravenous administration of [<sup>14</sup>C]pyrene in rats. *Toxicol Sci* 1998;46:11–20.
- Bull S, Fletcher K, Boobis AR, Battershill JM. Evidence for genotoxicity of pesticides in pesticide applicators: a review. *Mutagenesis* 2006;21:93–103.
- Calafat AM, Needham LL. Factors affecting the evaluation of biomonitoring data for human exposure assessment. *Int J Androl* 2008;31:139–43.

- Calafat AM, Needham LL. What additional factors beyond state-of-the-art analytical methods are needed for optimal generation and interpretation of biomonitoring data? *Environ Health Perspect* 2009;117:1481–5.
- Calafat AM, Koch HM, Swan SH, Hauser R, Goldman LR, Lanphear BP, et al. Misuse of blood serum to assess exposure to bisphenol A and phthalates. *Breast Cancer Res* 2013;15: 403.
- Cantor KP, Blair A, Everett G, Gibson R, Burmeister LF, Brown LM, et al. Pesticides and other agricultural risk factors for non-Hodgkin's lymphoma among men in Iowa and Minnesota. *Cancer Res* 1992;52:2447–55.
- Clarke P, Sproston K, Thomas R. An investigation into expectation-led interviewer effects in health surveys. *Soc Sci Med* 2003;56:2221–8.
- Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. *Ann Intern Med* 1993;118:201–10.
- Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol* 1977;105:488–95.
- Datta S, Datta S. Empirical Bayes screening of many p-values with applications to microarray studies. *Bioinformatics* 2005;21:1987–94.
- Dickersin K, Min YI. Publication bias: the problem that won't go away. *Ann N Y Acad Sci* 1993;703:135–46. [discussion 146–138].
- Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of exposure always bias a true effect toward the null value? *Am J Epidemiol* 1990;132:746–8.
- Durmaz E, Ozmert EN, Erkekoglu P, Giray B, Derman O, Hincal F, et al. Plasma phthalate levels in pubertal gynecomastia. *Pediatrics* 2010;125:e122–9.
- Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One* 2008;3:e3081.
- EPA (US Environmental Protection Agency). *America's children and the environment* 3rd ed. ; 2013 [Available: <http://www.epa.gov/ace/> Accessed November 25, 2013].
- Faust F, Kassie F, Knasmuller S, Boedecker RH, Mann M, Mersch-Sundermann V. The use of the alkaline comet assay with lymphocytes in human biomonitoring studies. *Mutat Res* 2004;566:209–29.
- Fisher CG, Wood KB. Introduction to and techniques of evidence-based medicine. *Spine (Phila Pa 1976)* 2007;32:S66–72.
- Gallo V, Egger M, McCormack V, Farmer PB, Ioannidis JPA, Kirsch-Volders M, et al. Strengthening the Reporting of Observational studies in Epidemiology Molecular Epidemiology STROBE-ME: an extension of the STROBE statement. *J Clin Epidemiol* 2011;64:1350–63.
- Goodman M, Barraj LM, Mink PJ, Britton NL, Yager JW, Flanders WD, et al. Estimating uncertainty in observational studies of associations between continuous variables: example of methylmercury and neuropsychological testing in children. *Epidemiol Perspect Innov* 2007;4:9.
- Goodman M, LaKind JS, Mattison DR. Do phthalates act as obesogens in humans? A systematic review of the epidemiology literature. *Crit Rev Toxicol* 2014;44(2): 151–75.
- Gordis L. *Epidemiology*. Philadelphia, PA: Saunders Elsevier; 2008.
- Greenland S. Basic methods for sensitivity analysis of biases. *Int J Epidemiol* 1996;25: 1107–16.
- Harthé C, Rinaldi S, Achaintre D, de Ravel MR, Mappus E, Pugeat M, et al. Bisphenol A-glucuronide measurement in urine samples. *Talanta* 2012;100:410–3.
- Hernan MA, Hernandez-Diaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol* 2002;155:176–84.
- Hertz-Picciotto I. Environmental epidemiology. In: Rothman KJ, Greenland S, editors. *Modern epidemiology*. Lippincott Williams and Wilkins; 1998.
- Hytten F. Blood volume changes in normal pregnancy. *Clin Haematol* 1985;14:601–12.
- Ioannidis JP. Discussion: why “an estimate of the science-wise false discovery rate and application to the top medical literature” is false. *Biostatistics* 2014;15:28–36. [discussion

- 39–45].
- Ioannidis JP, Tarone R, McLaughlin JK. The false-positive to false-negative ratio in epidemiologic studies. *Epidemiology* 2011;22:450–6.
- Jager LR, Leek JT. An estimate of the science-wise false discovery rate and application to the top medical literature. *Biostatistics* 2014;15:1–12.
- Jurek AM, Maldonado G, Greenland S, Church TR. Exposure-measurement error is frequently ignored when interpreting epidemiologic study results. *Eur J Epidemiol* 2006;21:871–6.
- Kalsbeek W, Heiss G. Building bridges between populations and samples in epidemiological studies. *Annu Rev Public Health* 2000;21:147–69.
- Kissel JC, Curl CL, Kedan G, Lu C, Griffith W, Barr DB, et al. Comparison of organophosphorus pesticide metabolite levels in single and multiple daily urine samples collected from preschool children in Washington State. *J Expo Anal Environ Epidemiol* 2005; 15:164–71.
- Kleinbaum DG, Klein M. *Logistic regression: a self-learning text*. NY: Springer-Verlag New York; 2002.
- Kleinbaum DG, Sullivan KM, Barker ND. *A pocket guide to epidemiology*. New York: Springer Science + Business Media; 2007. p. 228–9.
- Koornneef J, Spruijt M, Molag M, Ramirez A, Turkenburg W, Faaij A. Quantitative risk assessment of CO<sub>2</sub> transport by pipelines — a review of uncertainties and their impacts. *J Hazard Mater* 2010;177:12–27.
- Kuo CC, Moon K, Thayer KA, Navas-Acien A. Environmental chemicals and type 2 diabetes: an updated systematic review of the epidemiologic evidence. *Curr Diab Rep* 2013;13:831–49.
- LaKind JS, Naiman DQ. Bisphenol A (BPA) daily intakes in the United States: estimates from the 2003–2004 NHANES urinary BPA data. *J Expo Sci Environ Epidemiol* 2008; 18:608–15.
- LaKind JS, Naiman DQ. Daily intake of bisphenol A (BPA) and potential sources of exposure — 2005–2006 NHANES. *J Expo Sci Environ Epidemiol* 2011;21:272–9.
- LaKind JS, Levesque J, Dumas P, Bryan S, Clarke J, Naiman DQ. Comparing United States and Canadian population exposures from national biomonitoring surveys: bisphenol A intake as a case study. *J Expo Sci Environ Epidemiol* 2012a;22:219–26.
- LaKind JS, Goodman M, Naiman DQ. Use of NHANES data to link chemical exposures to chronic diseases: a cautionary tale. *PLoS One* 2012b;7(12):e51086. <http://dx.doi.org/10.1371/journal.pone.0051086>.
- LaKind JS, Goodman M, Mattison DR. Bisphenol A and indicators of obesity, glucose metabolism/type 2 diabetes and cardiovascular disease: a systematic review of epidemiologic research. *Crit Rev Toxicol* 2014;44(2):121–50.
- Lanphear BP, Dietrich K, Auinger P, Cox C. Cognitive deficits associated with blood lead concentrations  $\geq 10$   $\mu\text{g}/\text{dL}$  in US children and adolescents. *Public Health Rep* 2000; 115:521–9.
- Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology* 2003;14:451–8.
- Lassen TH, Frederiksen H, Jensen TK, Petersen JH, Main KM, Skakkebaek NE, et al. Temporal variability in urinary excretion of bisphenol A and seven other phenols in spot, morning, and 24-h urine samples. *Environ Res* 2013;126:164–70.
- Leamer EE. Sensitivity analyses would help. *Am Econ Rev* 1985;75:308–13.
- Lee WC, Huang HY. Data-dredging gene-dose analyses in association studies: biases and their corrections. *Cancer Epidemiol Biomarkers Prev* 2005;14:3004–6.
- Leng G, Kuhn KH, Idel H. Biological monitoring of pyrethroids in blood and pyrethroid metabolites in urine: applications and limitations. *Sci Total Environ* 1997;199:173–81.
- Liekens AM, De Knijf J, Daelemans W, Goethals B, De Rijk P, Del-Favero J. Biograph: unsupervised biomedical knowledge discovery via automated hypothesis generation. *Genome Biol* 2011;12:R57.
- Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, von Elm E, et al. Strengthening the reporting of genetic association studies (STREGA): an extension of the strengthening the



- reporting of observational studies in epidemiology (STROBE) statement. *J Clin Epidemiol* 2009;62(597–608):e594.
- Lorber M, Koch HM, Angerer J. A critical evaluation of the creatinine correction approach: can it underestimate intakes of phthalates? A case study with di-2-ethylhexyl phthalate. *J Expo Sci Environ Epidemiol* 2011;21:576–86.
- Lord SJ, GebSKI VJ, Keech AC. Multiple analyses in clinical trials: sound science or data dredging? *Med J Aust* 2004;181:452–4.
- Maldonado G, Delzell E, Tyl RW, Sever LE. Occupational exposure to glycol ethers and human congenital malformations. *Int Arch Occup Environ Health* 2003;76:405–23.
- Marco CA, Larkin GL. Research ethics: ethical issues of data reporting and the quest for authenticity. *Acad Emerg Med* 2000;7:691–4.
- Markham DA, Waechter Jr JM, Wimber M, Rao N, Connolly P, Chuang JC, et al. Development of a method for the determination of bisphenol A at trace concentrations in human blood and urine and elucidation of factors influencing method accuracy and sensitivity. *J Anal Toxicol* 2010;34:293–303.
- Meeker JD, Barr DB, Ryan L, Herrick RF, Bennett DH, Bravo R, et al. Temporal variability of urinary levels of nonpersistent insecticides in adult men. *J Expo Anal Environ Epidemiol* 2005;15:271–81.

以下 207 ページ

- Meeker JD, Cantonwine DE, Rivera-González LO, Ferguson KK, Mukherjee B, Calafat AM, et al. Distribution, variability, and predictors of urinary concentrations of phenols and parabens among pregnant women in Puerto Rico. *Environ Sci Technol* 2013; 47:3439–47.
- Moher D, Tricco AC. Issues related to the conduct of systematic reviews: a focus on the nutrition field. *Am J Clin Nutr* 2008;88:1191–9.
- Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001; 357:1191–4.
- National Research Council (NRC). Human biomonitoring for environmental chemicals. Washington, DC: The National Academies Press; 2006.
- National Toxicology Program (NTP). Draft OHAT approach for systematic review and evidence integration for literature-based health assessments — February. Division of the National Toxicology Program, National Institute of Environmental Health Sciences, National Institutes of Health; 2013 [Available: <http://ntp.niehs.nih.gov/?objectid=960B6F03-A712-90CB-8856221E90EDA46E> [accessed 25 October 2013]].
- Needham LL, Calafat AM, Barr DB. Uses and issues of biomonitoring. *Int J Hyg Environ Health* 2007;210:229–38.
- Oquendo MA, Baca-García E, Artes-Rodríguez A, Pérez-Cruz F, Galfalvy HC, Blasco-Fontecilla H, et al. Machine learning and data mining: strategies for hypothesis generation. *Mol Psychiatry* 2012;17:956–9.
- Owens DK, Lohr KN, Atkins D, Treadwell JR, Reston JT, Bass EB, et al. AHRQ series paper 5: grading the strength of a body of evidence when comparing medical interventions — Agency for Healthcare Research and Quality and the effective health-care program. *J Clin Epidemiol* 2010;63:513–23.
- Pearce N. Classification of epidemiological study designs. *Int J Epidemiol* 2012;41:393–7.
- Pleil JD, Sobus JR. Estimating lifetime risk from spot biomarker data and intraclass correlation coefficients (ICC). *J Toxicol Environ Health Part A* 2013;76:747–66.
- Potischman N, Weed DL. Causal criteria in nutritional epidemiology. *Am J Clin Nutr* 1999; 69:1309S–14S.
- Preau Jr JL, Wong LY, Silva MJ, Needham LL, Calafat AM. Variability over 1 week in the urinary concentrations of metabolites of diethyl phthalate and di(2-ethylhexyl) phthalate among eight adults: an observational study. *Environ Health Perspect* 2010;118: 1748–54.
- Rhomberg LR, Chandalia JK, Long CM, Goodman JE. Measurement error in environmental

- epidemiology and the shape of exposure-response curves. *Crit Rev Toxicol* 2011;41: 651–71.
- Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology* 1990;1: 43–6.
- Rothman KJ, Greenland S. *Modern epidemiology*. Philadelphia, PA: Lippincott Williams and Wilkins; 1998.
- Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health* 2005;95(Suppl. 1):S144–50.
- Sabatti C. Avoiding false discoveries in association studies. *Methods Mol Biol* 2007;376: 195–211.
- Sampson M, McGowan J, Cogo E, Grimshaw J, Moher D, Lefebvre C. An evidence-based practice guideline for the peer review of electronic search strategies. *J Clin Epidemiol* 2009;62:944–52.
- Scher DP, Alexander BH, Adgate JL, Eberly LE, Mandel JS, Acquavella JF, et al. Agreement of pesticide biomarkers between morning void and 24-h urine samples from farmers and their children. *J Expo Sci Environ Epidemiol* 2007;17:350–7.
- Schisterman EF, Whitcomb BW, Buck Louis GM, Louis TA. Lipid adjustment in the analysis of environmental contaminants and human health risks. *Environ Health Perspect* 2005;113:853–7.
- Shmueli G. To explain or to predict? *Stat Sci* 2010;25:289–310.
- Sobus JR, McClean MD, Herrick RF, Waidyanatha S, Nylander-French LA, Kupper LL, et al. Comparing urinary biomarkers of airborne and dermal exposure to polycyclic aromatic compounds in asphalt-exposed workers. *Ann Occup Hyg* 2009;53:561–71.
- Sorahan T, Gilthorpe MS. Non-differential misclassification of exposure always leads to an underestimate of risk: an incorrect conclusion. *Occup Environ Med* 1994;51:839–40.
- Spiegelman D. Approaches to uncertainty in exposure assessment in environmental epidemiology. *Annu Rev Public Health* 2010;31:149–63.
- Symanski E, Kupper LL, Kromhout H, Rappaport SM. An investigation of systematic changes in occupational exposure. 1996;57:724–35.
- Teeguarden JG, Calafat AM, Ye X, Doerge DR, Churchwell MI, Gunawan R, et al. Twenty-four hour human urine and serum profiles of bisphenol a during high-dietary exposure. *Toxicol Sci* 2011;123:48–57.
- Teitelbaum SL, Britton JA, Calafat AM, Ye X, Silva MJ, Reidy JA, et al. Temporal variability in urinary concentrations of phthalate metabolites, phytoestrogens and phenols among minority children in the United States. *Environ Res* 2008;106:257–69.
- Vandenbroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the reporting of observational studies in epidemiology (strobe): explanation and elaboration. *Epidemiology* 2007;18:805–35.
- Viau C, Bouchard M, Carrier G, Brunet R, Krishnan K. The toxicokinetics of pyrene and its metabolites in rats. *Toxicol Lett* 1999;108:201–7.
- Völkel W, Kiranoglu M, Fromme H. Determination of free and total bisphenol A in human urine to assess daily uptake as a basis for a valid risk assessment. *Toxicol Lett* 2008; 179:155–62.
- Wacholder S, Hartge P, Lubin JH, Dosemeci M. Non-differential misclassification and bias towards the null: a clarification. *Occup Environ Med* 1995;52:557–8.
- Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst* 2004;96:434–42.
- Walker DG, Wilson RF, Sharma R, Bridges J, Niessen L, Bass EB, et al. Best practices for conducting economic evaluations in health care: a systematic review of quality assessment tools. *AHRQ Methods for Effective Health Care*. Rockville (MD): Agency for Healthcare Research and Quality; 2012.
- Wang H, Zhou Y, Tang C, He Y, Wu J, Chen Y, et al. Urinary phthalate metabolites are associated with body mass index and waist circumference in Chinese school children. *PLoS One* 2013;8:e56800.

- Weed DL. On the use of causal criteria. *Int J Epidemiol* 1997;26:1137–41.
- Weed DL, Gorelic LS. The practice of causal inference in cancer epidemiology. *Cancer Epidemiol Biomarkers Prev* 1996;5:303–11.
- Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.
- World Health Organization. Biomarkers & Human Biomonitoring. Children's Health and the Environment WHO Training Package for the Health Sector; 2011. October. [www.who.int/ceh/capacity/biomarkers.pdf](http://www.who.int/ceh/capacity/biomarkers.pdf).
- WHO (World Health Organization). Toxicological and health aspects of bisphenol A. Report of Joint FAO/WHO Expert Meeting. 2–5 November 2010 and Report of Stakeholder Meeting on Bisphenol A; 2011. [Available: [whqlibdoc.who.int/publications/2011/97892141564274\\_eng.pdf](http://whqlibdoc.who.int/publications/2011/97892141564274_eng.pdf) [accessed 25 November 2013]].
- Wielgomas B. Variability of urinary excretion of pyrethroid metabolites in seven persons over seven consecutive days — implications for observational studies. *Toxicol Lett* 2013;221:15–22.
- Wirth JJ, Rossano MG, Potter R, Puscheck E, Daly DC, Paneth N, et al. A pilot study associating urinary concentrations of phthalate metabolites and semen quality. *Syst Biol Reprod Med* 2008;54:143–54.
- Withey JR, Law FC, Endrenyi L. Pharmacokinetics and bioavailability of pyrene in the rat. *J Toxicol Environ Health* 1991;32:429–47.
- Ye X, Kuklennyik Z, Needham LL, Calafat AM. Quantification of urinary conjugates of bisphenol A, 2,5-dichlorophenol, and 2-hydroxy-4-methoxybenzophenone in humans by online solid phase extraction-high performance liquid chromatography–tandem mass spectrometry. *Anal Bioanal Chem* 2005;383:638–44.
- Ye X, Zhou X, Hennings R, Kramer J, Calafat AM. Potential external contamination with bisphenol A and other ubiquitous organic environmental chemicals during biomonitoring analysis: an elusive laboratory challenge. *Environ Health Perspect* 2013;121: 283–6.
- Youngstrom E, Kenworthy L, Lipkin PH, Goodman M, Squibb K, Mattison DR, et al. A proposal to facilitate weight-of-evidence assessments: Harmonization of Neurodevelopmental Environmental Epidemiology Studies (HONEES). *Neurotoxicol Teratol* 2011;33:354–9.
- Zartarian V, Bahadori T, McKone T. Adoption of an official ISEA glossary. *J Expo Anal Environ Epidemiol* 2005;15:1–5.
- Zelenka MP, Barr DB, Nicolich MJ, Lewis RJ, Bird MG, Letinski DJ, et al. A weight of evidence approach for selecting exposure biomarkers for biomonitoring. *Biomarkers* 2011;16:65–73.
- Zota AR, Calafat AM, Woodruff TJ. Temporal trends in phthalate exposures: findings from the National Health and Nutrition Examination Survey, 2001–2010. *Environ Health Perspect* 2014;122:235–41.