

食品安全委員会評価技術企画ワーキンググループ

(第12回) 議事録

1. 日時 平成30年12月14日(金) 10:00～11:54

2. 場所 食品安全委員会中会議室(赤坂パークビル22階)

3. 議事

(1) 食品健康影響評価におけるベンチマークドーズ法の更なる活用について

(2) その他

4. 出席者

(専門委員)

川村座長、赤堀専門委員、小坂専門委員、小関専門委員、

西浦専門委員、広瀬専門委員、山田専門委員

(専門参考人)

井上専門参考人、岡田専門参考人

(食品安全委員会)

佐藤委員長、山本委員、川西委員、吉田委員

(事務局)

川島事務局長、小平事務局次長、中山評価第一課長、吉岡評価第二課長、

池田評価情報分析官、橘評価技術企画推進室長、窪崎課長補佐、青木評価専門官、

石神技術参与

5. 配付資料

資料1 食品健康影響評価におけるベンチマークドーズ法の活用に関する指針(骨子案、動物試験で得られた用量反応データへの適用)

参考資料 新たな時代に対応した評価技術の検討～BMD法の更なる活用に向けて～

6. 議事内容

○川村座長 それでは、定刻となりましたので、ただいまから第12回「評価技術企画ワー

キンググループ」を開催いたします。先生方には、この年末の慌ただしい時期に、御出席いただきましてありがとうございます。

机上に配付いたしました資料の、一番上の議事次第をごらんください。

本日は、7名の専門委員の先生方全員が御出席です。また、専門参考人として、井上専門参考人、岡田専門参考人に御出席いただいております。

食品安全委員会からは4名の委員の先生に御出席いただいております。

それでは、議事を進めます。本日の議題は「食品健康影響評価におけるベンチマークドーズ法の更なる活用について」として、ベンチマークドーズ法の指針作成に向けた議論を行います。

開催通知等で御連絡しましたように、本日の会議につきましては、公開で行いますので、よろしく願いいたします。

事務局より資料確認をお願いいたします。

○窪崎課長補佐 お手元に議事次第、座席表、専門委員名簿のほか、資料1の「食品健康影響評価におけるベンチマークドーズ法の活用に関する指針（素案、動物試験で得られた用量反応データへの適用）」のほか、参考資料として、7月に取りまとめていただきましたBMD法の指針作成に向けた議論の報告書をお配りしております。

これらは近日中にホームページに掲載させていただきます。

また、机上配付資料に加えて、過去のワーキンググループでお配りした資料、海外機関等のガイダンス等について、ドッジファイルにとじて御用意しております。

さらに、海外機関等のガイダンスをはじめ、BMD法の報告書において参照した文献、本日の配付資料などがお手元のiPadに保存されています。審議内容に応じて適宜参照ください。

資料の不足等がございましたら、事務局までお申し出いただければと思います。

○川村座長 ありがとうございます。資料はよろしいでしょうか。御確認ください。

続きまして、事務局から「食品安全委員会における調査審議方法等について」に基づき、必要となる専門委員の調査審議等への参加に関する事項について報告してください。

○窪崎課長補佐 それでは、本日の議事に関する専門委員等の調査審議等への参加に関する事項について報告します。

本日の議事について、専門委員の先生方から御提出いただいた確認書を確認したところ、調査審議等に参加しないこととなる事由に該当する専門委員はいらっしゃいません。

○川村座長 先生方、御提出いただいた確認書に相違はございませんでしょうか。

よろしいですね。ありがとうございます。

それでは、議事に入らせていただきます。

議事の「(1) 食品健康影響評価におけるベンチマークドーズ法の更なる活用について」です。

前回のワーキンググループでは、本年度、食品健康影響評価技術研究事業において設置され、BMD法の指針作成に当たっての検討課題について科学的知見の集積を進めていただく

研究班から、研究の計画・コンセプトなどを御発表いただき、意見交換させていただきました。

その後、本ワーキンググループで審議するBMD法の指針について、その趣旨や構成を確認させていただいたほか、一部の手順、基準等は、研究班の研究成果等をもとに改めて検討、更新していくことと、まずは動物試験編の審議を進め、来年夏をめどにその第一版を取りまとめるといった指針作成のスケジュールを確認させていただきました。

さらには、7月に取りまとめた「議論の報告」では統一的な考え方を示していませんでした用量反応モデリングについて、集中的に御議論いただきました。

本日は、事務局において、前回の議論の内容を整理した上で、私も含め先生方と相談しながら、指針の素案を用意していただきましたので、この内容について審議したいと思います。

審議ですけれども、資料1を幾つかのパートに分けて、パートごとに事務局から説明していただいた後、その内容を審議したいと思います。

パートの分け方は、最初に「第1 目的」と「第2 定義」について、次に「第3 食品安全委員会が行う毒性評価におけるBMD法の活用」のうちの1「用量反応データの収集、選択」と2「BMRの設定」を、3つめに、同じく第3の3「用量反応モデリングとBMD等の算出」を、最後に第3の4「結果の報告」と「第4 指針の見直し」の4つに分けたいと思いますので、よろしく願いいたします。

それでは、事務局から御説明をお願いします。

○青木評価専門官 それでは、お手元に資料1「食品健康影響評価におけるベンチマークドーズ法の活用に関する指針（素案） [動物試験で得られた用量反応データへの適用]」をお手元に御用意いただければと思います。

なお、委員の皆様には机上配付資料としまして、前回ワーキンググループで事務局よりお示ししました指針骨子案からの変更点を、見え消しで整理した資料を別途お配りしておりますので、御審議に際して適宜御参照いただければと存じます。

それでは、事務局のほうから説明させていただきます。

最初に冒頭、枠内の記載でございます。本案の記載のうち、先ほども座長のほうからお話がありました、現在進行中の食品健康影響評価技術研究「ベンチマークドーズ手法の健康影響評価における適用条件の検討」の成果をもとに、改めて御審議いただきたい箇所につきましては、斜体の上、アンダーラインを付した記載としております。この点を御留意いただければと存じます。

「第1 目的」でございますが、食品安全委員会がBMD法を活用する場合の基本的な考え方や手順等を整理することにより、食品安全委員会がBMD法を活用する際のさらなる一貫性及び透明性の確保に資することを目的とする、とさせていただきます。

「第2 定義」でございますが、本指針で使用するBMD法関連用語の定義を記載しております。

この内容につきましては、基本的に本年7月におまとめいただきました議論の報告の用語の解説から引用しております。

「(1) ベンチマークドーズ法 (Benchmark Dose approach)」に始まりまして、「(2) ベンチマークドーズ (Benchmark Dose:BMD)」 「(3) BMR (Benchmark Response)」 「(4) BMDL (Benchmark Dose Lower Confidence Limit)」。

2 ページに「AIC (Akaike Information Criterion)」 赤池情報量規準。「(6) 二値データ」 「(7) 順次カテゴリーデータ」 「(8) 連続値データ」 「(9) Restriction」 の9用語を定義いたしております。

事務局からの第1、第2の説明は以上でございます。

○川村座長 ありがとうございます。

この最初の部分につきまして、御質問とか御意見はございませんでしょうか。

1つ変更点として、BMDLのところに補足の説明を加えておりまして「BMDの信頼区間の下限値」という記述だったのですけれども、「BMDの90%信頼区間の下限値」というふうに、より正確にといいますか、ここでの運用上の信頼区間のとり方を書いております。

一般には95%という言葉がよく用いられますけれども、両側に分布するものについては両方のリミットを合わせた数字を書きます。ですから、下限5%、上限5%をとりますと、両側で90%信頼区間ということになりますので、そここのところの誤解を招かないように追記していただいております。

運用の仕方ということで、信頼区間というのは、科学的には別に1%でも5%でも10%でも全然構わないのですけれども、我々がどういう値をとるか、国民の理解を得やすいか、あるいは統計学上の常識といったものに照らして、値を選択することになります。よろしいでしょうか。

それでは、現在の新たな記述を採用します。

そのほかはよろしいでしょうか。

それでは、後で重いものがいっぱいあるので、このあたりは特に異論がなければ、次に進ませていただきたいと思います。

それでは、進みますが、次に「第3 食品安全委員会が行う毒性評価におけるBMDの活用」のうち「1. BMD法を適用する用量反応データの収集、選択」と「2. BMRの設定」を審議します。

事務局から説明をお願いいたします。

○青木評価専門官 御説明させていただきます。

資料1の2ページ目、23行目の「第3 食品安全委員会が行う毒性評価におけるBMD法の活用」になります。

第3では、7月におまとめいただいた議論の報告の内容をもとに、BMD法を活用する際の主な作業ごとに基本的な考え方、手順等を整理いたしております。

最初に24行目以降、BMD法を活用する際の基本事項として3点挙げております。

1つ目としまして、化学物質の毒性評価でBMD法を活用する際は、本指針に基づき、BMD等の算出を行うこと。

2つ目としまして、BMD等の算出に必要な生物学的、統計学的な判断は、各分野の専門家の御意見に従うこと。

3つ目としまして、本指針によらない考え方や手法をとる場合は、BMD等の算出結果の報告時に、その内容及び理由を併記することとさせていただきます。

30行目は「1. BMD法を適用する用量反応データの収集、選択」です。

31行目から3ページの5行目にかけて、評価対象とする化学物質について、利用可能な全ての用量反応データを収集するとした上で、専門家の御判断に基づいて、試験もしくはエンドポイントの選択を取り得るとしております。

3ページの6行目でございます。

(2)として、毒性試験とエンドポイントの組み合わせからなる各データセットのうち、用量反応関係が成立しているものにBMD法を適用するに当たって、1)としまして、試験設計が適切であること。

2)としまして、エンドポイントが、毒性学的に意義のある影響であること。また、ヒトへの外挿性についても考慮すること。

3)としまして、二値データなど、データの種類別にBMD法を適用する上で必要な情報を含むことを確認するとしております。

そして、二値データ、順序カテゴリーデータ、連続値データ、それぞれにつきまして、BMD法を適用する上で必要な情報を記載させていただいております。

21行目の(3)ですけれども、データセットの統合について記載しております。標本数をふやして、より精緻な検討を行うため、同質性が期待できるデータセットが複数存在する場合、それらのデータセットを統合することが可能であるとしつつ、用量反応関係が不明確になることなどを防ぐために、事前に統計学的、生物学的な観点から、慎重な検討が必要であるとし、そのために把握すべき情報を列記させていただいております。

ここでは各データセットにおける用量の範囲、動物種、標本サイズ、データ測定方法等の違いを把握する必要があるとしております。

本段落冒頭のデータ統合に当たって求められるデータの同質性とリンクする部分ではございますが、この記載内容、求められる同質性に関しまして、ここにはない情報も加味すべきではないか。もしくは、さらにはデータ統合そのものに関する御意見、御審議がございましたらお願いできればと考えております。

さらに、可能であれば個体データまでさかのぼって活用することが望ましいとしております。

31行目の(4)からでございますが、グレーディングがありますような順序カテゴリーデータを二値データとして取り扱うことについては、影響の種類やグレードのつけ方を精査しつつ、慎重に検討すべきであるとして、具体的な注意点などは研究成果をもとに改め

て御検討いただくということにしております。

4 ページの「2. BMRの設定」でございます。

(1) としまして、二値データを用いた場合のBMRの設定でございます。

3 行目以降の1) において、他の化学物質又はエンドポイント等の比較の観点、NOAEL相当の用量で観察し得る過剰リスクも踏まえ、BMRとして過剰リスク10%を用いたBMD及びBMDLは基本的に算出するとしております。

一方、2) で、この過剰リスク10%は、BMRとして、全ての用量反応データに最適な値ではないとした上で、各BMRは、生物学的な意義づけのほか、用いる用量反応データの特性を考慮しつつ、専門家の関与のもとで設定するとしております。

後段ですが、10%とは異なるBMRを設定することが考えられる例として、BMRが最小用量で観察された反応よりもかなり小さく、観察範囲よりも大幅に低用量側の推計を伴う場合ですとか、生殖発生毒性のうち次世代への影響に関する用量反応データを用いる場合等を例として挙げさせていただいております。

13行目の(2) としまして、連続値データを用いた場合のBMRの設定でございます。

14行目の1) としまして、基本的なBMRの設定の考え方としまして、生物学的に意味のある反応量の変化を、科学的根拠に基づいて設定できる場合は、その変化量をBMRとして設定するという基本的な考え方の記載をしております。

16行目の2) としまして、連続値の用量反応データは多様な分布を示すということが想定されるとしまして、1) で言及しました生物学的に意味のある反応量の変化が不明な場合は、用量反応ごとに専門家の関与のもとでBMRを設定するとしております。

19行目の3) としまして、科学的に合理的と思われるカットオフ値を設定できる場合は、反応量の分布ばらつきを仮定した上で、カットオフ値に該当する個体割合の一定の増加分をBMRとして設定することでBMD等を算出する。いわゆるハイブリッド法をとり得るということに言及しております。

23行目の4) としまして、個体データが利用可能な場合に、情報量の減少を伴うという点に留意する必要があるとしつつ、カットオフ値に基づきまして、連続値データを二値データに変換した上でBMRを設定し、二値モデルに基づきまして、BMD等を算出することが有用な場合があるとしております。

第3の1及び2は以上でございます。

○川村座長 ありがとうございます。

ここは重要なポイントを幾つか含んでおりますけれども、今の事務局からの説明があったところにつきまして、御意見はございませんでしょうか。

○小坂専門委員 確認なのですが、2 ページの一番下のデータの収集、選択というところなのですが、利用可能な全ての用量反応データを収集するというのは「利用可能な」と書かれているのでいいと思うのですが、何年分までやるのかとか、英語以外の外国語のデータはどうするのか、我々がやるときには少しは気になると思うのです

けれども、その辺のところは余り具体的なところは記さないという方向で行くということですね。

○川村座長　そうですね。測定技術とかそういうのがもう時々刻々と、絶えず進歩しているので、余り古いものは現実問題として使えなくなってきましたし、そのあたりは、どこまでを採用するかというのは物質ごとに異なると思うのです。

そういうことで、余り細かく書ききれなくて、包括的にですけれども「利用可能な」の線を決めること自体も、またエキスパートオピニオンで決まる領域かとは思っております。

このあたりで何か御意見はございませんでしょうか。

それでは、表現としてはこのままですけれども、含みおくこととして、利用可能なというときに、論文の公表をされた年とか、場合によっては、どういう条件の、動物実験ですので余り地域差はないかとは思っているのですけれども、固有の動物とかいろいろあるかと思うので、そういうことはその都度検討していくということになるかと思えます。

そのほかの問題はいかがでしょうか。

どうぞ。

○井上専門参考人　1の(3)の同質性が期待できるデータセットの25行目から「そのためには」以降、いろいろな条件の違いを把握する必要があるというところなのですけれども、ここに記載されているのは、毒性試験の条件が主になっているかと思えます。

「等」で含んでいるのかもしれないのですけれども、例えば病理所見であれば、病変の所見名だけで同じと判断しがちかと思うのですが、その病変の質についても考慮するというのが現実にあるのかと。なので、その辺の記載も入れてはどうかと思って拝見しました。

○川村座長　ありがとうございます。

ここの同質性のところは、かなり大きな問題を内包しております。同質性と一言に言いますが今、御指摘があった点も含めて、非常に多くの要素がありますので、結論から言うと、同質性の評価、あるいは統合に関しては慎重にならざるを得ないと思えます。

データの統合というのは今、国際的な流れで、特にヒトを対象とした有効性評価の研究は、もうメタアナリシスがついて回っているのですけれども、そのあたりと違うところは、今、世界的に進められている統合の動きの主流は治療法の評価で、効くか効かないかという、イエスかノーかを出すクリニカルクエスチョンに対する答えが主流です。

そういうところは統計パワーを増すことによって、より評価が妥当になる。イエスかノーかという質問については出しやすくなるのですけれども、反面、我々のように、dose response curve、dose response relationship、用量反応関係を出そうとすると、途中の、1か0かではない、0から1までの連続的なところを分析しなければならないと思うのです。そのパターンをむしろ評価するので、多少の異質であっても、まぜるとラインどりがはっきりわからなくなるとか、同じ条件のように見えるけれども、何か1つは違っている。多くはそうで、前と全く同じだったら新規に論文を出す価値が減ってしまうので、何らかの形で新しく追加される研究が出てきたということは、条件が異なる場合が多い。そうな

りますと、どこまで統合が現実に可能なのかというのが非常に問題になります。

私自身も国際的なデータ統合の流れでものを考えてきたのですが、最近改めて、観察研究が持つ用量反応関係の分析がある意味、非常にデリケートであって、非常に重要なポイントであると。そうすると、そこが統合によってうまく表現されるのかという疑問がかなり大きくこみ上げてきたことがありますので、同質性について今、御指摘のあった質ということも含めまして、コンフィデンスといえますか、統合に十分足るとき、例えば、同じ研究者がもう一回確認のために同じ条件でやり直したとか、そういうときは比較的やりやすいかとは思いますが、異なる研究者が異なる実験の施設を用いてとか、動物種が微妙に違ったりとか、いろいろなことがあるともはや難しいでしょうし、そういう病理所見のように、もともと1つの軸で評価しにくいものについては、ましてや統合は可能かどうかということもありますので、問題提起としては出したのですが、現実問題、相当慎重でないといけないというのが最近、座長個人として強く感じているところではあります。

したがって、同質性が十分期待できるという場合で、かつ、個々の要素について、チェックリストのようなものでしっかり評価した上で、それから、異質性について統計学的な I^2 という指標もございますので、そういうところも援用して、十分な吟味をしてから入るべきだと思います。

そういう意味で、今の井上専門参考人の御意見の表現を一部取り込んだほうが良いとおもっておりますけれども、よろしいでしょうか。

どうぞ、お願いいたします。

○吉田委員 反論をするようで恐縮なのですが、病理所見についてはそんなに大きな問題はないように思うのです。というのは、同質性を考えるようなものは、そういう試験の実施年が変わっていないというような、何十年も離れたものをそもそも同じように考えることはございませんので、用語の統一化ということは毒性学方面でも進めます。

この同質性については重要だということは同じなのですが、むしろ毒性試験では余りこの統合というよりも、むしろ例えば、Concurrent Controlが一番、ジェネティカリーに同じ動物を配分して試験をしているので、Concurrent Controlが一番重要であるというのは、多分動物実験をする人間にとって基本だと思うのです。それに対して、異質というか別に行われた実験を統合するというのは、私もかなり慎重であるべきだと思うのですが、私のコメントとしては一つ、最近よくバックグラウンドでヒストリカルコントロールデータといって、同じ試験、同じ試験施設、大体プラスマイナス5 yearぐらいのもので出た所見が、たまたまこのConcurrent Controlの頻度が0であったため、あるいは、10であったために起き得たことについてというのは、非常に重要な議論になるのですが、これはベンチマークドーズ法を適用するに当たり考慮すべきということにはならないのかを先生方にお尋ねしたいという、Concurrent Controlだけではなく、ヒストリカルコントロールデータをどう考えるのかということはどうなのかをお尋ねしたいのです。

○川村座長 今、おっしゃったヒストリカルコントロールというのは、過去にやった実験

のリファレンスという意味ではなくて。

○吉田委員 定義がございまして、厳密な意味では、毒性試験で用いられているのは、同じ試験実施施設、同じ種、同じ系統で、プラスマイナスこのぐらいのyear、同じ投与形態、例えば強制経口ではなくて、同じ混餌投与で集められたデータでのコントロールのデータだけ、対照群のデータだけをヒストリカルコントロールデータと定義しています。

○川村座長 そのコントロールだけ、厳密に言えば別であるという場合ですね。なるほど。

○吉田委員 いえ、別ではなくて、もちろんその試験はコントロールがあるのですけれども、実際に比較してみたときに、コントロールの値が0だった。しかし、通常は5とかいうのが対照群では主だったということが結構あるのです。そういう場合は、このベンチマークドーズ法との共通点というか、関連性はないと考えてよいのか、それとも、そういう点についてもということは、もし考えていただけるとありがたいと思います。

○川村座長 同じような実験だけれども、コントロールだけを持ってくることになるのですか。それとも、実験全体は持ってこないのですか。

○吉田委員 その場合はコントロールだけです。というのは、実験全体はもう投与してしまっていますので、投与による影響が出てくる可能性が否定できないということで、広瀬先生が何かフォローしていただけるとありがたいです。

○広瀬専門委員 まず、データ統合の話は毎回出てきますけれども、現実的に動物実験で統合した例は、本当に繰り返した実験の場合がほとんどです。違う目的で違った実験を合わせるということではなくて、大体過去に、例えばアクリルアミドもそうなのですが、随分昔の80年代にやったけれども、同じような実験で2000年にもう一回やって、用量相関も同じだったので使ったということです。

ですから、先ほど同じ試験は論文にならないと言いましたけれども、それをあえて論文にするケースがまれに化学物質はあるので、だから、割と同質性の範囲内でもともとやるのです。わざと昔の実験を繰り返してやっている。しかも、そのときに用量もふやすので、それも合わせるとうまくいくという場合がほとんどなので、そのときは発がん性も微妙に増殖性と悪性とかという違いはもちろんありますけれども、陰性、陽性で割とクリティカルでやっているのです。多分病理的な問題を含めて、そこまで一緒にしてやるというケースは、経験的にはまずないです。発がん性のところでしかやっていないので、そういう意味では、一般病理のほうもちゃんと井上委員の言われたところも加えたほうがいいと思います。

現実としては本当に再現した場合でしか動物実験でも統合していません。そういうことで、もしやるのだとすれば慎重にやったほうがいいと思います。

今の対照群の話は確かに難しいですけれども、一応そういうコントロールのぶれも含めてフィッティングをやっているはずなので、余りヒストリカルコントロールを持ち出してきて、ベンチマークドーズを一緒にやるということは基本的にはしないと思います。

もしやるのだったら今、言ったように、そのときに対照群の匹数もたくさんふやして、

もう一回実験をやって、そのときにベンチマークドーズを当てるということをやりますけれども、たまたま今回高かったか低かったかというのを加味してしまうと、統計学的なモデルのフィッティングの意味がなくなってくるように私は感じるのですけれども、西浦先生にこの辺をフォローしていただけると。

○西浦専門委員 おっしゃるとおりで、Benchmark Responseをとるために、どこで基礎のデータをとるかということで、たくさんサンプルがあったほうがわかりやすいので、そういうときに種が同じだったら、ヒストリカルコントロールを使った例は私は見たことがないのですけれども、theoreticalには大丈夫なのだと思います。

今のお話で、皆さん、実験生物学の専門の先生方が多いので、統合にすごく懐疑的なのですけれども、私は本当に正反対の考え方をされていて、EFSAの研究だとか、アメリカNIOSHとかのEPA研究とかでは、複数実験データが日常的に同じテーブル上で合併して使用されています。今、広瀬先生がアクリルアミドを言及されましたけれども、ほぼ種に関してと性別に関して同質性を担保する、同じ実験設定であるという場合には、複数の実験データがそのまま一緒にまぜた状態で適合しているという研究が普段から多数あります。

Risk Analysisという雑誌ですと、チタンの影響とかを特に分析しましたというときは、6つか7つの実験データをそのまま、何も説明なくがちゃっと一緒にまとめた上で分析している論文が普通に出ていますので、どこまで異質性に関してケアをしているかというのと、今のこの議論と比べると、海外のスタンダードは相当ルーズなのだと思います。

何というか、ためらいなくどんどん統合しているというのが実情です。同質か異質かというのは、同質ということは「データのばらつきがランダム性で全てが決まっている」と想定することと等価です。つまり、例えばデータが同じドーズで、バリエーションが出たとしても、それはランダムのもので決まっているということが言えるので、統合してしまえばいいということなのですけれども、それが本当は異質性のために生じたのかどうか、動物実験の今のOECDガイドラインに従っている動物実験の動物数では検討できないというのが実情です。3匹、5匹、10匹ぐらいの単位のものだと、実験間の異質性というのも、 I^2 とかで有意にその後検定ができるのですけれども、出せないのです、どちらかということ、政策的な事情でサンプル数がふえると、BMDLが1つの実験に基づくサンプル数に頼って出したときよりも、複数実験のほうが、サンプル数が高くなる影響でBMDLが高いドーズのところになりますから、そういう理由もあってどんどんと統合しているのだろうというのが、私が見て理解しているところです。

○川村座長 どうぞ。お願いします。

○赤堀専門委員 同質性はすごく大事だというのは理解できるのですけれども、あらかじめ同質性を定義するというか、実際にプロットしてみたときに外れているところは、どういう理由で外れているのかという理由を考えた上で、オミットしていくのではないのでしょうか。例えばラットとマウスでやったとしても、ある程度通常の評価では同じに扱っているわけですね。完全に同じではないと思いますけれども、そこに出てくるNOAELの値とかは

近似している。

ですから、大体同じようなところで反応はするということが大前提で今までの評価があるのであれば、どちらかという、先にそれを定義するというよりも、一回全部を乗せてみて、外れているところは、ここはラボ間差がすごく大きかったとか、何かこういう事情があるのかもしれないという感じではないでしょうか。それをばらつきという形で解釈することもできるのかという印象があって、あらかじめ同質性を定義することのディフィカルティーのほうが多分大きいのではないかということをお個人的に思いました。

○川村座長 ありがとうございます。

また頭が一層混乱してきましたけれども、動物実験でかなり統合例があるということで、それはどうでしょうか。用量反応関係を見たものなののでしょうか。それとも、有害性があるかないかという評価をしているものなののでしょうか。

○西浦専門委員 ベンチマークドーズ用の実験が複数同じ物質で、同じアウトカムに関してたくさんあります。

○川村座長 そうなると、現実にはそういう複数の同じものを扱った論文に出会う可能性は十分にあるということになりますね。

それから、このように、いろいろなデータがIT化というウェブ上に保存される時代になると、必ずしも十分パブリッシュとしてきちんとされていないものでも、通常で言うところのハンドサーチレベルでひっかかるものでも出てきてしまう可能性があって、ここはちゃんとパブリッシュされたもの、査読を経て載ったものしか採用しないということを通していいかもしれませんが、インフォーマルなデータも、ひょっとしたらいろいろ利用可能になってしまう恐れという可能性もあって、そうなると、データ統合も必ずしもあり得ない話ではないということになってきます。

そうすると、心構えは用意しておかないといけないですね。

○赤堀専門委員 今、西浦先生が、海外でやられていることがあるということで、そこから学習するというのも一つかと思えます。なので、今、こうしたらいいのではないかということをやっていますけれども、海外の状況を少し踏まえて、それを実際見て、いや、これはないのではないかということなのか、ここのセンスはありますねというところは取り入れるというのが現実的ではないかと思えます。

○川村座長 ありがとうございます。

まだ我々の学習が続くという。

○小坂専門委員 我々もデータを集めるときに、地方衛研のデータとかあったので、必ずしもPubMedだけではなくハンドサーチも大事だなと思いました。もう一つ、試験動物がgeneticallyに一緒ではない場合も、実際に多様性があり得るわけですね。それを統合しても良いのではないかと思いました。

○川村座長 吉田先生、お願いします。

○吉田委員 practicallyにそれはもう、0%に近いぐらい実現性はないかと思えます。そ

う言いますのは、実験動物、例えばラット、マウス、ウサギぐらいまでですけれども、そのように交配、育種を繰り返して、実験に適した遺伝的なコントロールをしてきておりますので、そういったものをあえてそこでしていますというのがバックグラウンドになっています。

実際、出てくる病変とかいうものが、ほとんど別の種と言っていいぐらい違うものが、同じラットでも多くございます。同じように出てくるような病変もありますけれども、特に加齢性の病変です。かなり違います。それは病変だけではなくて、繁殖性なり繁殖能、あるいは子供の数、いろいろなものが全て系統によってかなり違うので、私は今の段階では系統を統合するというのは、かなりチャレンジングではないか。それを説明できるだけの十分な科学的エビデンスはまだ我々は蓄積していないのかというのが、私のコメントでございます。

○小坂専門委員 現実的にはいろいろなラットの多様性があるわけですね。ですから、それをまとめて統合してヒトの評価に使うということを経験してはいけないうのですかと言われたときに、何と答えられるでしょうか。

○川村座長 突然吉田先生の組に入ってしまったりするのですけれども、種が違っていると、多分バイオアベイラビリティとかディストリビューションとかが変わってきて、いわゆるばく露量とは違う臓器レベルでの、あるいは細胞レベルでのばく露量が変わってきたり、ゼロレベルでの発生の様子が変わってきたりするのではないかと。それから、それは仮に同じ動物種であっても、例えば週齢が違っているとそれも変わってくる恐れがないのかと。例えば、非常に幼若時の動物でやった場合と、成獣でやった場合は、それも同じ結果が得られないのではないかと思うので、動物種だけでもないのかと思ったりするのですが、このあたりは経験のある人の御意見を聞かないといけないと思います。

○吉田委員 今、小坂先生の御質問で思い出したのですけれども、何年前か、今まで私たちは毒性なりの影響をよりセンシティブに見つけようとして、そういった系統の育種を重ねてきた動物を使ったのですけれども、それは違うのではないかと。たしかいろいろなものをミクスチャーしたようなマウスだったかラットかの研究が、しばらく前に論文で出ていたように思います。

ただ、それ以来、それが進展したということは聞いていないので、もし論文を見つけましたらお知らせしたいと思います。

○川村座長 ありがとうございます。何かすぐに結論が出なくなってしまったのですけれども。

佐藤先生。

○佐藤委員長 それはデータを統合して何を求めているのかによるのだらうと思うのです。

先ほど座長がおっしゃっていた疫学なんかだとみんな一緒にしてしまうメタアナリシスが結構多いというのは、多分ヒトだっているいろいろなポピュレーションは違うのだらうけれども、それを超えて何か起きていのかどうかを見ようというのが目的なわけですね。

動物実験の場合にはピュアにしてきたわけだけれども、吉田先生が最後のところで言われたように、ホモジニアスでない集団の中でも何が起こるのかという考え方もあり得るということだろうと思うのだけれども、結局このデータを統合するときに、どういう目的でやろうかということを考えて、それがいいのか悪いのかという考え方にしていってほしいのだと思うのです。

よりピュアな形で、集団で見たほうがいい場合と、そうではなくて、もっとホモジニアスでなくても起こり得るようなことを見るという場合で、考え方は違うのだと思うのです。

この文章は割合よくできていると思うのですがけれども、標本数をふやし、より精緻な検討を行うために、その精緻というのが今、言ったようなことになるのか。しかし、専門家の関与のもとに書いてあるので、そういう目的を考えた上で考えましようというふうにとれるのではないかと思うのです。

文章としてはこういうことで、多分よろしかろうと。今、議論したことを頭の中に置いておいて、赤堀先生がおっしゃったように、先行事例みたいなものも勉強しつつ、また必要があればこれを書き直していただくということではいかがでしょうか。

○川村座長 ありがとうございます。

大岡裁きみたいでありがたいのですけれども、確かにここで細かいことを決めきれないこともありますし、まだこれ自体が一つの食品安全委員会として知りたい知見でもあるので、ここにはあえて細かくは書かず、基本的に現在の記述を残すということで、今後の検討材料として認識しておくことにしたいと思います。

動物実験のデータ統合に関してというのは、それ自体で一つの解決されるべき大きな課題だと思いますので、このところは現在の基本的な流れで、一部字句の修正はあるとしても、この表現を、こういう基準の仕方を残すということにしたいと思います。それでよろしいでしょうか。

ありがとうございます。

同質性の前に、エンドポイントの（２）の中の２）ですけれども「エンドポイントが、毒性学的に意義のある影響であること」という言葉を使ってもらっております。

ここはあえてこういう書き方をしてもらったところですが、生物学的な変化が全て病的な意味があるとは限らないので、ヒトにおいて病気につながるものという意味で、こういう表現にしているところがあります。

いわゆるパソロジカルな変化以外に、リアクションとしての変化があり得るかと思うのです。異物の侵入に対して何かの反応を起こす。これは生物として当たり前のことで、起きなかったら逆にまずいでしょうという、あって当たり前の反応の変化には違いがないので、そのあたりを論文の中で、病的な意味がどうあるのかということを確認して、エンドポイントを設定するという意味で、この文言を追加しております。

そういうことで、この毒性試験の選択の仕方、あるいは取り扱いについて、試験動物の

種類や数、あるいは試験設計に加えて、エンドポイントを特記しております。

それから、データの種類によって、求める必要な情報がいろいろあるということは従前どおりであります。

特に今度はBMRについてですけれども、資料の4ページですが、こちらの連続値データにつきましてはなかなか取り扱いが難しく、どこからが病的な意味があるのかというのをなかなか評価しがたいところがあります。

もちろん、科学的根拠が出てきて、これ以上の反応が異常であるということが明快に説明できればよろしいのですが、変化は捉えたけれども、それが果たして問題のあるレベルなのか、それほどではないのかというところがなかなか難しく、一応その連続値データについては1、2、3、4という列記するやり方をしております。

これについてはいかがでしょうか。

お願いします。

○西浦専門委員 1つだけ、細かいところなのですが、連続値データの3)なのですが、ちょうど今、疫学データに対するBMD法の適用とかで、連続値データの cohorts とかを詳しく勉強している最中だったのですが、ハイブリッド法を取り得るということと言及していただいているのですが、今の国外の状況を見ていると、ハイブリッド法が新しい方法に置き換えられそうで、2017年ぐらいをきっかけに、quantile impairment threshold method というのですが、直訳すると分位変動閾値法とでもいえるのでしょうか。NIOSHの統計学者らによってそういうのがつくられてきているのですが、ハイブリッド法は今まで平均値 (mean) を使っていたのが、ベイズ統計の手法を利用して分位を使って連続データの閾値を決めてくださいというものです。技術面での日進月歩な部分もあるのですが、ハイブリッド法が完全にリプレースされそうな気配なのです。

しかし、新しい変化にずっとガイダンスで最初から対応するのは厳しいことです。少なくとも、例えばハイブリッド法「等」とかにしていただくとか、そうやっているとうまくいくのかと思います。

○川村座長 ありがとうございます。

ここの連続値データをいかに評価するかということで、ハイブリッド法は一つの注目すべき方法として見ていたのですが、激変が起こる可能性があるということで、これもまだ現時点で決定しきれないところかと思えます。というか、研究班が一生懸命活動できればされるほど、未知の部分もふえていくという、一般に真実探求というのはそうなのですが、やればやるほどわからなくなる。わからない部分もふえてきます。

それはいいことでありまして、科学は永遠にチャレンジということになりますが、これも今、御指摘がありましたように、もしハイブリッド法や連続量の評価が変わってくるということであれば、私どもはまだそこを十分承知していないところですので、ハイブリッド法の後に「等」を入れて、とりあえずしのいでおいて、我々は逆に言うと、次から次へと課題がふえますけれども、このワーキングでまた新しいものについて取り上げるという

ことになるかと思えます。

○佐藤委員長 質問をいいのですか。

西浦先生に教えていただきたいのですけれども、前のほうの記述は変えることはないのですか。大丈夫ですか。

○西浦専門委員 前というのはどこですか。

○佐藤委員長 「ハイブリッド法を取り得る」の前に3行分ぐらい書いてあるけれども、その記述を変える必要はないですか。

○西浦専門委員 ハイブリッド法自体の考える連続データを扱うコンセプト自体は維持されているので、特に影響はないのですけれども、ハイブリッド法は平均（mean）を必ず使ってしまったという、ちょっとした問題が指摘されていて、それを分位にかえると、すごくリスク評価がうまくいくというのが今、皆の間で、そういえばそうだなという認識が広がっているみたいで、そこだけの問題です。

○川村座長 得られた数値の切り方の問題でしょうか。

○西浦専門委員 はい。代表値のどこを使うかということだと思います。

○川村座長 わかりました。

それでは、基本的な考え方自体は変化がなくて、数字の取り扱い、切り方のところが変わるということなので、本文自体は生きるということですね。

○西浦専門委員 はい。

○川村座長 それでは、お願いいたします。

ほかはよろしいでしょうか。

それでは、ここの部分はいろいろ意見が出て、課題はたくさん残りましたがけれども、原案を一部修正して、指針をつくってまいりたいと思います。

それでは、第3の「3. 用量反応モデリングとBMD等の算出」を審議します。

事務局から御説明ください。

○青木評価専門官 それでは、御説明いたします。

4 ページの26行目、第3の「3. 用量反応モデリングとBMD等の算出」でございます。

前回集中的に御審議いただいた内容をもとに、この項を新たに整理させていただきました。

「(1) 用量反応モデリング」としまして、5 ページの1行目にかけて1) としまして、ここでは、先ほどの用量反応データの選定の部分で3つ挙げました条件を満たしたデータにつきましては、鋭敏なエンドポイントの排除を防止する観点から、全てモデリングを行う、とさせていただきます。

5 ページの2行目の2) としまして「生物学的な根拠がある場合を除き、データセットに含まれる全ての用量反応データを用いてモデリングを行う（最高用量群のデータのみをモデリングの際に除外するといったことは行わない）」ということに記載しております。

5行目の3) ですけども「特定の化学物質とエンドポイントの組合せに関する用量反

応関係は、理想的には、普遍的なトキシコキネティクス及びトキシコダイナミクスを説明する、生物学的見地に基づく単一の数理モデルで説明される」としまして、そのような「発現機序の本質を捉えた数理モデル」があれば、当該モデルからBMD等を算出することが優先されるとしております。

10行目の4) としまして「『発現機序の本質を捉えた』数理モデルが存在しない場合は、BMD等の算出を行うソフトウェアやオンラインツールを使用し、それらに収載されている数理モデルを用いてモデリングを行う」こととしております。そして、使用するソフトウェアにつきましては「過去の使用実績や論理的背景も踏まえつつ、専門家の関与のもとでBMD等の算出結果が妥当であると判断したソフトウェアを用いる」とさせていただいております。

16行目の5) としまして「使用するソフトウェアにおいて、モデルのパラメータの値を制限するRestriction機能が利用可能な場合は、同一モデルであっても、Restrictionを行った場合(ON)と行わなかった場合(OFF)は基本的には違うモデルであるとの前提に立ち、ONとOFF両方のモデルを用いたモデリングを行う」としております。

また、研究成果も踏まえて、その適否も改めて御検討いただくことになると考えておりますが、モデル平均化が可能な場合は、同機能を用いたモデリングをあわせて実施するという記事を記載しております。

22行目の「(2) 用量反応モデリング結果の評価」になります。

これ以降、point of departure (POD) の決定までの内容につきましては、前回ワーキンググループで集中的に御審議いただいた内容をもとに整理しております。

23行目の1) としまして「データセット（データセットを統合した場合は、統合後のデータセット）毎に、各モデルを用いたモデリング結果を評価する」としております。

その際、脚注をごらんください。「同一動物種を対象とした同一エンドポイントのデータセットが複数存在する場合は、専門家の判断に基づき、普遍的に各データセットへの適合性が良いモデルを選択可能であれば、当該モデルから算出されるBMD又はBMDLを優先して採用する」という記載にしております。

モデリングの結果が27行目以降の①から、ページをおめくりいただいた③に該当するモデルについては、用量反応データを明らかに説明できていないものとして、あらかじめBMD等の算出、選択の対象から除外するとしております。

その除外の要件の具体でございます。27行目の①として「適合度検定の結果、用量反応データが当該モデルから説明できないと判断されたモデル」を挙げております。適合度検定の臨界値については、研究の成果も踏まえて改めて、通常0.1が海外のガイダンス等では記載されていますが、そこについても改めて御検討していただくとしております。

31行目から6ページにかけまして、②として「プロットしたモデリング結果を目視した結果、用量反応データ全体、特にBMRに対応する反応量の変化を引き起こす用量領域を明らかに説明できていないモデル」を除外するとしております。

ここで本案では、7月にまとめていただいた議論の報告ですとか、海外のガイダンスも参考に、プロットの目視におきましては、特に低用量領域を重視する旨を記載させていただいておりますが、この重みづけにつきまして、原案のとおり低用量領域を特に重視すべきなのか。それとも、BMD法の特徴であります全体の用量反応データを利用するという観点から、全体のフィッティングを重視するような記載にしたほうがよいのかということについて、御意見をいただければと考えております。

6ページの3行目の③としまして、BMDの関連指標。例えば、BMDL/BMDといった値をモデル除外の基準として設定することに妥当性があるのかといったところについては今、研究のほうでシミュレーションをさせていただいておりますので、その結果をもとに改めて御検討いただくということにしております。

6ページの4行目、2)としまして、先ほどの除外基準に該当しなかった各モデルにつきまして、BMRに対応する反応量の変化を引き起こす用量域におけるそれぞれの適合度合いを考慮しつつということで、改めて低用量領域を重視するという記載をしております。

その先のモデルの選択基準につきましては、3つ挙げておりますけれども、そのほかの選択肢も含めまして、研究成果も踏まえて、改めて検討するとさせていただいておりますが、一定の基準に従ってモデルを選択し、そこから算出されるBMDなどを採用していくという手順を整理しております。

この中で指標の案の一つとしまして、案1として、AICの値が小さいことを指標の案として挙げております。この指標を採用した場合、例えばAICが最小の値であった数理モデルの結果を採用すべきなのか、それとも、AICの値が十分に小さい一定の範囲にあるモデルを選択すべきかについて、御意見をいただければと考えております。

その際、後者、一定の幅にあるモデルを採用するという考え方に拠る場合ですけれども、十分に小さい範囲としてどの程度の値を考えればよいのかというところです。過去の御審議では2～3程度の範囲内にある数理モデルは、同等にフィットしているという御指摘もございました。海外の機関では2でしたり4という数字を挙げている場合もございますので、この点についてもあわせて御意見をいただければと考えております。

10行目の3)としまして、単一のモデルを上記のような手順で選択しきれない場合であって、かつモデル平均化を行っている場合は、モデル平均化により得られた結果を採用することも検討するというふうにしております。

この場合のモデル平均化の手法の詳細も含めまして、その是非については研究結果も踏まえて、改めて御検討いただくということで、斜体、アンダーラインを付しております。

7ページの枠囲みの中に、先ほど御説明しました用量反応モデリング結果の評価手順について、代替案を整理しております。

先ほどの案と同様に、データセット、データセットを統合した場合は統合後のデータセット毎に評価を行うとした上で、1)にございますとおり、モデル平均化で得られたBMD等が妥当であるということを確認した場合は、平均化により得られ

る結果を優先して検討してはどうかという手順を整理しております。

2) としまして、モデル平均化を行っていない場合ですとか、平均化の結果の妥当性や信頼性が十分に担保されがたいと判断された場合は、先ほど御説明したような案と同様に、個別の数理モデルを用いたモデリング結果を評価する手順としております。

モデル平均化を採用するだけでなく、その結果を従来の単一モデルのフィッティング結果に基づく検討に先んじて、優先して検討する案として、代替案をお示ししております。

本代替案につきましては、そもそものモデル平均化の是非も始めまして、案全体について、研究結果を基にして改めて御審議いただきたいと思っておりますが、現時点で本案に対する御意見、御質問等がございましたら、ぜひいただければと考えております。

7 ページ下段の 1 行目になりますけれども、「(3) HBGV算出に用いるPODの決定」になります。

先ほどの手順を通じまして、データセットごとに独立したモデリング結果、すなわち、代表選手となりますBMD等が得られることとなりますので、それらの結果について用いた用量反応データの特徴ですとか、BMDやその信頼下限値、BMDLの値などをもとに、あるデータセットを選択し、当該データセットにおいて算出されたBMDの信頼下限値であるBMDLを、リスク判定のためのHBGV算出に用いるPODとするとしております。

第3の3.の御説明は以上でございます。

○川村座長 ありがとうございます。

この部分では、フィッティングについて全体あるいは低用量領域を特に重視すべきかという問題と、モデルの選択の指標、あるいは平均化の課題というところになります。

最初のフィッティングを確認するということで、何か御意見はございませんでしょうか。

お願いします。

○広瀬専門委員 確かにフィッティングを目で確認するというところで、BMR付近を重要視するというのは、ある意味当たり前ではあるのですが、実際できることは、よほど盲目的に数値だけ見ていくとすごくフィットしているけれども、波打ったような曲線が出たりするので、そういうのを一応目で確認するという意味合いのほうがここは強いはずなので、余り強調するほどのことはないのかという感じがします。

○川村座長 ありがとうございます。よろしいでしょうか。

確かにそのとおりで、まず、ここにいる方々の認識を共通化しておかないといけないと思うのですが、NOAEL法と違って、BMD法というのはそもそも生物学的なモデルに従って反応を起こすはずであるという前提を是として、それで実験のデータを見るということなので、各実験のデータにモデルを当てはめるのではなくて、モデルを仮定して、そこに実験データを流し込むという発想になります。検定も一緒に、差がないという仮説を立てて、そこに観察したデータを当てはめたらどうなるのかという考え方なので、基本的にはモデル先にありきなのです。

モデルはどうやって選んだかという、長年の経験に基づいて、こういうモデルは今までの生命現象をよく説明してきたというふうに、有形、無形の経験から選び抜かれたものが今、装備されている。ロジスティックとかプロフィットとかワイブルとかマルチステージとか、そういったもののはずなのです。あるいは、理論的に構築されたものもあるかもしれませんが、それが生物学的な現象をよく説明するであろうという前提で、それを是とした上で、実際に動物実験で観察したデータを流し込んで、逸脱していないかどうかを確認するという、そもそもの統計学の根本の考え方を共通化しておく必要があるかと思います。

そうすると、広瀬先生がおっしゃっているように、基本的には全体で見るものであって、念のために、極端に逸脱したものは心証的にもとりづらいので目視、それから、幾つかの指標で逸脱をあらわす指標で確認だけをするということ、余り特定のところにこだわらず、全体を見るのが筋であると考えておりますけれども、そういう考えでよろしいかどうかということ、皆さん、御意見がありましたらお願いいたします。

よろしければ、念のための確認で、特にこれが重要というよりは、むしろ決めるときの心証を固めるときの補助的なものであるという感じで、これを厳密に当てはめて、イエスかノーか決めるといふほどのものではないと認識しておりますけれども、そういう雰囲気を書き方でよろしいでしょうか。

ありがとうございます。

次に、モデルの選択あるいは平均化ということなのですが、このところは先に西浦先生から見解をお願いしたいと思うのです。

○西浦専門委員 段々、研究班ではまだ完全に結果は出ていないのですが、平均化をいろいろなパターンで検討しています。

平均化といってもいろいろな平均化がありますから、収束したモデルを全部平均化するのか、トップ3のいいものだけ平均化するのかとか、AICでどれくらいの範囲にあるものだけを拾い上げて平均化するのかとか、フィルターを何らかの形でかけてから平均化するのかとか、考え得るのですけれども、考え得る全てのパターンを計算して、どれがいいのかというのを検討していて、その結果はまた御報告できる機会を見つめます。

平均化ということのフィロソフィーに関しては、できればもうちょっと、恐らくAICが最も小さいものだったり、BMDLが最も小さいという基準で選択したものと比べて、よりBMDLがもともとのNOAELあるいはシミュレーション上のBMDLの真の値というのを再現できるものが来る。つまり、モデル平均化のほうが、ほかの指標を使ったときよりもパフォーマンスのいいものがくるということは、恐らく大丈夫だろうと思うのですけれども、まだそれは結果が完全に出さることが必要です。また、このシミュレーションに関しては海外の査読をちゃんと経ようと思いますので、自分たちがやったら多分大丈夫ですけれども、外で査読されると意外に間違っていたりとか、覆されたりなど、そういうことが全くないわけではないので、慎重にこの先、その経過を見て、また報告させていただきたいと思って

おります。

AICだけで決めていったら簡単なのですけれども、モデルの選択の基準がたくさんあるので、このBMD法の最後のリファレンスドーズの決定というのが混沌とするということなので、一つこれの議題に上がっていたことではあるのですけれども、AICの小さいものだけで、ベストモデルを何らかの形でフィックスするという事は、少し難しいのかという印象は持っています。

とりあえずそれぐらいにさせていただきます。

○川村座長 ありがとうございます。

研究班の展望というか結論の見通しが、少しペンディングと聞かせていただいておりますが、確定したものではありませんので、書き込むには少し限界があるかもしれませんけれども、AICというのは重要な指標ではありますが、それがオールマイティーとは言えないということで、AICは使ってもいいけれども、押しきれぬかどうか分からないということです。

そのAICを使った場合も、ナンバー1を選ぶのか、トップ3を選ぶのか、一定の範囲を決めてそこに入るものは全部使うのかとか、いろいろあるかと思えますし、それだけ迷うのだったら最初から、平均化した場合にどこまで入れるかということと、AICだけで単一のものを選ぶかというのが大事にされるのかもしれないけれども、生物学的に十分説明しきれたモデルが存在しない場合の行政上の選択、あるいはサイエンスとしても、リスクヘッジみたいな形で平均化をとるという考え方もあるかもしれません。

このあたり、施策の面も一つありますし、AICのパフォーマンス自体の問題もあるかと思えます。

今の段階で指針案をどこまで詳しく書き込めるかというのが難しいのですけれども、仮の案ということで、今の見通しで書いておくことができるのかどうかです。こちら辺をお願いします。

○広瀬専門委員 そういう意味で西浦先生からも補足していただきましたけれども、AICを重要視しつつ、除外基準も同時にやりながら、BMDLも低いものをとるという、どれかが必ず順番にシーケンスに優先するのではなくて、ほぼ同時に、どういう組み合わせがいいかというのは研究班で今、検討しているので、申しわけないですけれども、なかなか書き込めません。

こういうことは優先しているというところまでは書けますけれども、どういう基準で、どうやるというところまでは書けないというのは、研究班の最終的な結果はまだ確定できていないので、そこは御勘弁いただければと思っていますところでは。

○小坂専門委員 6ページ一番上の、反応量の変化を引き起こすことを明らかに説明できていないモデルのことに絡むのですけれども、既存のモデルがある場合はそれを使ってもいいわけですね。既存のモデルがないところで、例えば限られたデータでいろいろ精緻なことをやろうとすると、かえってオーバーフィッティングになってしまうので、平均化

をしましようというのが趣旨だと思うのです。その点のところは少し書いてもいいのかと。

ですから、BMRとか関係しないようなフィッティングモデルも平均化のときに入れるかどうか、本当に外してしまっているのかはいかがでしょうか。

○西浦専門委員 今、適合するのに使っているモデルは、例えば、この食安委でよく使われているBMDSというアメリカのソフトウェアを使っていると、プロトタイプにしているモデルは8つだけなのです。Restrictionとかをつけると、それにプラス4つぐらいとかになりますかね。それの中の、どれも用量反応関係で、生体の中で、一定の理由でメカニズムが数式になっているものの理屈が完全に成立して固定されているようなモデルはないのです。

ですから、何かのモデルがうまく当てはまっても、それで今後も行けるという保証がないので、いつもがららぼんと全て適合してしまうというのが実践になっていると理解しています。

そのときに、オーバーフィッティングするから平均化をしないといけないというわけではなくて、どのselection criteriaで最終のモデルを選ぶと、真のBMDLというものを返せるのだろうかという命題に対して、今はまだ最も良いアプローチに関する解がないのです。

その答えの一つが、平均化を何らかの形でやるとうまくいくかもしれないからということでやられているということなのですけれども、とりあえず今のコンベンショナルなプラクティスというのは、その2パラメータと3パラメータぐらいのモデルというのは、もう生物学的には現象を記述するには十分ではないので、それぞれの意味論というのは、基礎は知っておくことは必要なのですけれども、とりあえずはその動物実験の与えられたサンプル数の少ないデータが目前にあったら、既存のデフォルトになっているモデルをどんどん適合していったということで、その中から盲目的に今、やっているのが現状だと理解しています。何かを恣意的に外すことができそうなモデルはないのではないかと思います。

○川村座長 ありがとうございます。

先生方の御意見を総合しつつ思うところですが、やはり生物学的なアプローチで単一のモデルを選ぶということはまだ難しいので、AICを参考指標として一定の許容範囲のものを平均化する。それによって、モデルが本質を必ずしも言い得ていない場合のリスクをヘッジするといいますか、緩和するといいますか、そういう方法で平均化をとれる場合はとる。

それを数値の範囲で決めるのか、順位で決めるのかはまだ検討の余地がありますが、先ほど言いましたように、基本的にはモデルに数値を当てはめる。なぜかという、モデルのほうがより確かなもので、実際の個々の実験データはランダムエラーをいっぱい含むものなのです。ですから、個々の実験データでモデルを引き出すということはもともと本質的に無理だと思いますので、過去の有形、無形の経験から経験知として得られた提案されるモデルを用意しておいて、その中で性能のよさそうなものについて幾つか選んで、それを是として得られた実験データを流し込んでみて、基準となるPODを決めていくということではないかと思っております。

そういう意味で、モデルを選択して、一定の手順で、主にAICを中心として選んで、念のためにその外れぐあいとかを見ることは見るのですけれども、それも先ほどと同じように、より納得して進めるための一つの参考指標にとどまると思います。

その上で実験データを流し込んで、具体的な値を決めていくということではないかと思っております。

そこで、一応現時点では平均化を優先できるものについてはする。そこに使う主たる指標がAICである。具体的にどこまでをとるかは研究班の御報告を待って確定するということかと思っております。

○西浦専門委員 平均化を必ず最初にするかどうかは、まだペンディングにさせていただきたい次第です。ちゃんと後日にエビデンスをつくります。それができれば、外で査読を受けたものということで見ただけだと思うのです。

あとは何ですかね。1個だけなのですけれども、AICで平均化をしていく話で使うことでもあるのですけれども、例えばもともとの、2)の指標案1、指標案2、指標案3とありますね。その中のlowest BMDLといって、BMDLが最も低い値を与えるようなものも、実を言うと、シミュレーションではそんなに悪い成果ではないのですということも開陳させていただきたいのです。

AICはフィットがいいだけで、フィットがいいというのは、BMDLがシミュレーション前の真のBMDLを返すかどうかというのとは独立の話なので、実はAICだけに頼っていると、BMDLをリカバーするもののパフォーマンスは悪いのですけれども、lowest BMDLという選択手段は、ひどいモデルを省ける限りはそんなに悪くなくて、P値で悪いモデルを除外した上であれば、あるいは、BMD割るBMDLとか、何らかのモデル選択の別の条件です。除外条件を事前に適用した上で、lowest BMDLというものを使うと、実は平均化に対比するぐらい、パフォーマンスはそんなに悪くないという成果が得られていたりしています。この辺は微妙なラインの計算をしているので、平均化できる場合はするということで着地するかどうかも、ちょっとだけペンディングにさせていただきたいとは思っています。

○川村座長 ありがとうございます。

ただ、ものの決定のフィロソフィーとしては、いろいろやって、しかも信頼区間下限値で一番低いというのは、サイエンスを放棄しているみたいに感じるころがあって、モデルはより合理的と信じられるものにして、そこに流し込むということだと思うのですけれども、AIC自体も実際にデータを使ってやっているのです、自己撞着に陥るのですけれども、値が一番低いようなものを選択的にとるとするのは、行政上の目的のものを除くと、サイエンスとしてはあり得ない考えだと思うのです。

ただ、サイエンスとして決めたものを、実際の行政の現場でどう使うかというときは、もちろん行政的な配慮は必要で、安全サイドに立つということですが、今はまだ行政的な選択以前に、サイエンスとしてどういうモデルを採用するかであって、BMDやBMDLを決定するというサイエンスレベルの段階で結果を見てから、結果というか、数値の一番

低いものを選ぶというのは、論理として納得しがたいものがどうしても残るのです。

○広瀬専門委員 それはそうなのですけれども、今、西浦先生のところでシミュレーションしているのは、BMDLが一番低いということを強調しているからあれなのですけれども、真のBMDを出すときのクライテリアとしてBMDLというのを幅でとってみると、真のBMDをたくさん補足できることをクライテリアの一つに、サイエンスベースでなるということ、今やろうとしているので、そこをもし説明できたら、余地はまだ残っているということを説明しています。必ずしも低いものをとらなければいけないというポリシーではないです。なので、その辺はここで固定しきれない。要するに、BMDLをとるのを完全に最初から外すのは、今の段階で、研究班の結果が出る前に何かできないかということで、ちょっとお持ちいただければというのが本心です。

○川村座長 そうしますと、このところは今、決定的には書けないので、AIC等を参照して、単一のモデルもしくは平均化したモデルを用いるというところまでしか書けないということで、また同時並行。要するに、この指針の案づくりと研究班の研究が同時並行なのでそうなるのですけれども、究極は融合するとして、現時点ではAICだけで選ぶわけではないのですけれども、AICを主たる目安として1つのモデルを選ぶ。もしくは、AICを参照して、こちらにもかかるのですけれども、モデルの平均化を行うという記述で、研究班の御成果を待って、そこの部分は修正するというにしたいと思います。

ですから、余り進展がないというと、ちょっと苦しいのですけれども、サイエンスの議論としてはそうなります。

事務局、コメントがあったら出していただいてもいいです。

○青木評価専門官 ありがとうございます。研究班の成果を待って、改めて御審議いただければと思います。

1点確認で恐縮ですけれども、用量反応、プロットしたモデリング結果の目視の部分でございますけれども、先生方の御意見を総合しますと、低用量域は当然大事なわけけれども、見なければいけないのは全体ということなので、今、特に低用量域、BMRに対応する用量領域をことさら強調した記載が2点ほどございますが、そこは修正したほうがよろしいというお考えでしょうか。

○西浦専門委員 削除されたほうがいいと思います。

低用量領域でフィットが悪いものだけを恣意的にどけるとエラーが起こってしまうので、その部分だけは要らないです。目視で明らかにフィットしていないものを除外するという内容の部分だけでいいので、特にここは要らないと思います。

○川村座長 今、西浦先生がおっしゃったとおり、全体を見て決める。全体にフィットが悪ければ、やむを得ず外さざるを得ないものも存在し得るかもしれないのですけれども、基本的には低用量領域だけを特別視はしないということによろしいかと思います。

そういうことで、あとは現在までのところは、基本的にはサイエンスベース等で論理的に正しいということと、統計学的な本来の推測統計学の考え方にのっかって、最初に仮定

を置き、それで現実のデータを入れ込んで、妥当性を評価しながらPODに相当するところの数値を出していく。

その後は、(実験は)サンプルですので、サンプルに伴うランダムエラーが発生するから、信頼区間というものを考慮する。

さらにその後には、動物実験の結果だから、動物種の問題とか、評価の期間、長期慢性影響を見たいけれども比較的短期間の影響で見たなどの要素から、安全係数などを掛ける。これは行政上の配慮の部分といったものが、後から足されていくと思います。

まずは現在、フィッティングについては全体については見る。それから、AICが主たる指標にはなるけれども、必ずしもそれのみになるわけではなく、モデルが1本容易に選択できれば、妥当と判断できればそうなりますし、そうでなければ平均化が無難ということで、あるいは、真の値を推定するのにシミュレーションで一番よかったということであれば、それを優先するというので、ここは1本には書ききらず、含みを残して、今のところの指針案とすることになるかと思いますが、ここまでのところはよろしいでしょうか。

ありがとうございます。

第3の「4. 結果の報告」と「第4 指針の見直し」に進ませていただきます。

事務局から御説明ください。

○青木評価専門官 資料1の8ページをごらんください。第3のうちの「4. 結果の報告」でございます。

BMD法を活用した結果については文書化の上、食品健康影響評価の審議に当たって、適宜報告するとしております。

文書化に当たって含める情報につきましては、4行目以降、①から⑥にその具体を整理しております。

① としまして、使用したソフトウェアの名称とそのバージョン。

② としまして、BMD法を適用した各データセットに関しまして、その試験設計の概要や出典、また、データ別に、データの試験の概要を示すような情報を整理するとしております。

③ としまして、BMD等の算出に用いたBMRの値と、そのBMRを設定した理由。

④ としまして、モデルのプロットなどを含みます各データセットにおけるAICですとかBMD等の計算結果を記載するというのを整理しております。

⑤ としまして、各データセットにおいて、モデリング結果の評価に用いた数理モデルの除外の基準ですとか、選択の基準。

⑥ としまして、最終的に決定したPODの値及び当該PODを導いたエンドポイント及びデータセットを採用した理由。

以上の情報を含める形で文書化し、報告するという形で整理しております。

最後に「第4 指針の見直し」として、BMD法に関する国際的な動向や、食品健康影響評価における活用実績等を踏まえ、必要に応じて指針を改訂する。本指針の見直し規定を最後に設けております。

説明は以上でございます。

○川村座長 ありがとうございます。

ただいまのところはいかがでしょう。

どこまで報告書に書くかということで、ちょっと気になったのは、実際に使ったモデルで諸指標を出すのは当然としましても、下準備の段階でいろいろやったのをどこまで表現するかということがあると思うのです。

特に採用されなかったモデルでの計算結果などを出すと、余計なと言うと変ですけども、採用しないものについては計算しなくていいはずのものなので、そこまで書いてしまったりしないほうがいいのかと思ったりもするのですが、AICはもちろん出すのでしょけれども、BMDやBMDLの値については、選んだデータセット、選んだモデルで出すだけでいいのかとも思ったりしております。

このあたりは御意見がありましたらお願いします。

○西浦専門委員 BMD法の思想的（フィロソフィカル）な部分というのは、用量反応関係のS字曲線が背景で、何らかの真のモデルがあるのですけれども、その真の今、目の前にあるデータを統御している真の反応関係がわからないわけですね。

それは直接観察できないので、不確実性の一種になりますから、複数のモデルを適合するというのは、それに対する不確実性に対応していることになります。なので、不確実性に対応して、直接観察できなくて、どれが正しいか実証（ベリフィケーション）ができない中で、こういう結果を提示するときの一番誠実な対応は、不確実性に全部こうやって幾つかのモデルをフィットして対応しましたという表をちゃんと出して、その代表的なメトリックを出すということだとは理解していますので、その一環で全部の項目に関してまとめて表ができて、それを出して検討するというのは、多分誠実な対応ということの一つなのだと理解しています。

○川村座長 検討したことというか、モデルの選択までに使った指標は出すべきだと思うのですけれども、BMDやBMDLは、BMDLが一番低いモデルを使うという手段をとる場合は出さないといけないけれども、AICを主として使って、BMDは選択には使わないということであれば、モデルを選んでから計算するべきことなので、出すべきではないと思うのです。

○西浦専門委員 モデルを適合すると、BMDとかBMDLというのは、もう自動で計算されているものですね。

○川村座長 出てきてしまうのだけれども、報告書に書くのはモデルを選んで、本来出してほしくないのに勝手に出てきてしまうだけだから、それは報告書には書かない。モデルを選択するのに必要なところだけ出せばいい。

選んだモデルについては全部出す。

○西浦専門委員 選んだものも不確実なので、本当は正解かどうか分からないわけなのです。

正解で選んだBMDとかBMDLだけではなくて、もしかすると正解かもしれないリストが順に、

このモデルのBMDとかBMDLのリストになると思うのです。

○佐藤委員長 ちょっとよろしいですか。

多分これはフィロソフィカルな問題とプラクティカルな問題があって、川村座長の御質問というか御懸念をプラクティカルに言えば、この報告は誰がどの場でするのか。要するに、誰がBMDを計算するのか今の時点でわかりませんが、プラクティカルには調査会にどうやって出すのかという話とつながると思うのです。

もう一つはリスク評価書にどこまで書くのか、どのデータまで載せるのかという話になって、そういうレベルを考えた上で今の話をしていただいたほうが多分いいのだろうと思うのです。

西浦先生がおっしゃるようなことは、多分調査会のところ、あるいは検討するところへ全部出したほうがいいのかと思うし、リスク評価書に書くのだとすると、川村先生がおっしゃったように、ある程度選択がかかった上での話で出したほうがいいのかと私は今の時点で思っていますけれども、そのレベルを決めて議論しないと、うまく結論を導けないのではないかと思います。

もう一つは、多分、食安委というのは多くの場合、調査会を公開でやっているわけなのです。公開の資料は全部ホームページに載せてしまっていて、そのやり方を続けていくと、今のような話は結構難しいのです。

○川村座長 ありがとうございます。

運用の問題と非常に密接に関連することで、検討して、専門調査委員の先生方には使わなかった資料も見せることには全然異存はありませんけれども、報告をする際に、一つの流れとして、モデルを選びました、そのモデルを使って計算するとこの値になりましたという、一つのストーリーのつくり方として、使わなかった数値はモデルに関するBMDやBMDLは出さなくてよいのではないかと思います。

これは運用の仕方なので、高次の話になるかもしれませんが。親委員会の先生方に、高い見地から御判断いただいたほうがよいかもしれないので、お願いしたいと思います。

○佐藤委員長 3行目に「食品健康影響評価の審議に当たり」というところがあって、それは多分、調査会の資料として出すという意味ですね。

違うのですか。

○青木評価専門官 これは評価書に掲載するというよりは、調査会の御審議の中で、最終的にPODですとかHBGVを決めていく際の御検討に資する資料として報告するというのを考えております。

○佐藤委員長 そこをそういう理解にいただければ。

○川村座長 わかりました。

そうしたら、いろいろ専門調査会での議論をメインに考える。個々の資料の公表の有無については別途お考えいただくとして、専門委員の中ではつまびらかにしたほうがいいとは思っていますので、お願いいたします。

○吉田委員 ありがとうございます。

親委員会であることを座長から言っていただきまして、ありがとうございます。

というのは、申請資料もございまして、統計学的指標は全て申請者にさせておりますので、このBMDということについても今後検討していかなくてはなりませんので、そのように言っていただけると大変、今後の我々のほうでもしっかり受けとめて、しっかり話し合いをしていきたいと思えます。

○川村座長 ありがとうございます。

それでは、報告は専門調査会への報告を念頭に今、列記していただいたような項目について御報告するというところでよろしいでしょうか。

それから、指針の見直しに関しても、たった2行ですけれども、随時見直しを図る。佐藤委員長のほうからも、最初のほうにお話がありましたように、やりながら経験を積むところもございまして、方法論自体も進歩するところもありますので、タイムリーに、必要が生じたときに見直しをするということは、これに限らず、方法論全てに言えることかと思えます。

そういうことで、このところは特に修正の必要はなく、このままで結構かと思えます。

それでは、全体を通しまして、この指針案の考え方、あるいは記述の仕方について、もう一度何か、お気づきの点はございますでしょうか。

委員長、どうぞ。

○佐藤委員長 中身にかかわる話ではないのですけれども、HBGVとかPODとかはいきなり出てくるような気がするのだけれども、定義にありますか。

これはもし使うのだったら、定義にあったほうがいいのではないかと。これはいきなり出てきていいですかね。

○川村座長 我々の間では当たり前かもしれませんが、ここには多分、BMDに比較的固有のことが書いてあるのだと思うのです。

○佐藤委員長 定義ではなくて脚注でもいいかもしれないけれども、何か手当てしておいたほうがいいと思うのです。

○青木評価専門官 検討いたします。ありがとうございます。

○佐藤委員長 それから、もう一点。

1 ページの定義の中で、Benchmark Responseのところ、私だけが気になるのだったらいいのですけれども、22行目にバックグラウンド反応に比した一定の反応量の変化のことで書いてあるのですけれども、比したというのが気になるのです。

それは何でかということ、過剰リスク、追加リスクの話もあるかと思うのですけれども、バックグラウンドに比したと言うと、バックグラウンドに対して何パーセントみたいなふうに見えてしまうこともあるし、連続量の場合には、二値のデータの場合と連続量の場合でまた違うのだらうと思えますけれども、バックグラウンドからどれくらい変化したかという、単にシンプルにそれでいいような気がするのですけれども、比したというのはどう

も割合とか比をとるみたいな話になってしまって、余りいい表現ではないかと思うのですが、いかがでしょうか。

○川村座長 そうですね。バックグラウンドのレベルからの増分を見ていることになりま
すので。変化の度合いというか。

○佐藤委員長 例えば、鉛のIQなどだと低下にはなるかと思うのですが、変化でい
いと思うのです。

○川村座長 バックグラウンドレベルからの変化の程度ですね。

実はまだBMRにも多少の謎が残っているところがあるのですが、バックグラウンド
がゼロだったらどうするかとか、その辺もありますので。

○佐藤委員長 それもあるのです。比しただと、0は0だから。

○川村座長 バックグラウンドはかなり条件によって変わりますので、どういう環境で実
験しているかによって、もともとリファレンスレベルが変わっていってしまうので、多分、
西浦先生の御説明にちゃんとあったのだけれども、まだしっかりと頭の中に入っていない
ので、上手な説明の仕方も含めて、番外編で申しわけないのですが、簡単な言葉で
お教えいただければと思います。

よろしいでしょうか。

そのほかに。

事務局、どうぞ。

○青木評価専門官 先ほどの定義の部分でございますけれども、ベンチマークドーズ法の
ところにも比してというのが、バックグラウンドに比してというのがありますので、ここ
をバックグラウンドからの一定の変化とさせていただきます。

○佐藤委員長 ですから、BMRに関係したところはみんな同じように書いてもらわないと。

○青木評価専門官 はい。わかりました。

○川村座長 ありがとうございます。

そのほか、お気づきの点はございませんでしょうか。

それでは、きょうは報告書がものすごく進展したという感じではないかもしれませんが
れども、論点はかなり議論できたと思いますし、研究班の成果もちらほら見えてきて、何
となく心が軽くなったような気がします。

○佐藤委員長 よい年をお迎えできるのではないですか。

○川村座長 そうですね。

先生方、御議論を進めていただきまして本当にありがとうございます。

本日の議論を踏まえまして、事務局には指針案の改定といいますか、さらなる、よりよ
い表現の、より良い内容の指針案の作成を進めていただけますようお願いいたします。

○窪崎課長補佐 承知いたしました。

○川村座長 そのほか、事務局から議題とか連絡はありますか。

○窪崎課長補佐 来年の2月15日の午後を次回ワーキンググループの開催予定日として、

先生方の御予定をいただいておりますが、議事等は改めてお知らせさせていただきます。
○川村座長 ありがとうございました。

それでは、これにて本日の議事は全て終了しました。本年のこのワーキンググループの開催は、今回が最後になります。委員の先生方、お忙しいところ熱心に参加していただきまして、本当にありがとうございます。

ことしの漢字が「災」という字になるように、大変いろいろなことがあって、被災された方もいらっしゃるかとは思いますが、来年は落ち着いて審議ができるように、いい年になってほしいと思っております。

それでは、来年またお会いします。それまで、よいお年をお迎えください。

ありがとうございました。