

令和元年10月29日

食品安全委員会

委員長 佐藤 洋 殿

評価技術企画ワーキンググループ

座長 川村 孝

「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」
について

食品安全基本法第21条第1項に規定する基本的事項（平成24年6月29日閣議決定）において、食品安全委員会は食品健康影響評価に関するガイドラインの作成に努めることとされているため、今般、当ワーキンググループにおいて、別添のとおり「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〔動物試験で得られた用量反応データへの適用〕」をとりまとめましたので報告します。

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
[動物試験で得られた用量反応データへの適用]

2019年10月

食品安全委員会
評価技術企画ワーキンググループ

第1 目的

本指針は、化学物質に係る食品健康影響評価において、ベンチマークドーズ法を活用して適切なPOD¹を求める場合の基本的な考え方、手順等を整理することにより、食品安全委員会が、同法を活用するに当たって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。

第2 定義

本指針における用語の定義は以下のとおりとする。

1. BMR (Benchmark Response)

化学物質のばく露量と、当該物質によりもたらされる影響の発生の頻度又は量との関係（用量反応関係）全体に数学的関数（数理モデル）を当てはめて得られた関数（用量反応曲線）におけるバックグラウンド反応からの反応量の変化。

2. BMD (Benchmark Dose)

一定のBMRをもたらす化学物質のばく露量。

3. ベンチマークドーズ法 (BMD法)

化学物質のばく露量と、当該物質によりもたらされる影響の発生の頻度又は量との関係（用量反応関係）は、特定の数学的関数（数理モデル）に従うという仮定の下、BMD及びその信頼区間を算出する方法。

4. BMDL (Benchmark Dose Lower Confidence Limit)

BMDの信頼区間の下限値（信頼下限値）。通常、BMDの90%信頼区間（片側信頼区間とすれば95%信頼区間）の下限値がBMDLとして用いられる。

5. AIC (Akaike Information Criterion)

以下の式で定義され、異なる数理モデル間で、モデルの複雑さと測定デ

¹ Point of Departure。各種の動物試験や疫学研究から得られた用量反応評価の結果から得られる値で、通常、無毒性量（NOAEL）やBMDLを指す。健康影響に基づく指標値（HBGV）を設定する際や、ばく露マージン（MOE）を算出する際等に用いられる。国際的には、Reference Pointということもある。出典：食品の安全性に関する用語集 http://www.fsc.go.jp/yougoshu/kensaku_hyouka.html（食品安全委員会）

ータとの適合度とのバランスを比較するための指標。

$$AIC = -2 \log(L) + 2k$$

$\log(L)$: モデルの最大対数尤度

k : モデルに含まれる推定パラメータ²数

AICの値が小さくなるモデル、すなわち少ないパラメータ数で適切な当てはめを実現するモデルが好ましいとされている。

6. 二値データ

化学物質へのばく露により各個体で観察された影響を、死亡又は生存、腫瘍発生の有無等、起こりうる帰結が特定の2つのいずれかであるとして分類した非連続データ（分類データ）のこと。

7. 順序カテゴリーデータ

化学物質へのばく露により各個体で観察された影響をカテゴリーに分類した非連続データのうち、尿糖検査値（－、±、＋、++他）、がんのステージ等、影響の程度等の順序はあるが量的概念はないカテゴリーに分類したデータのこと。

8. 連続値データ

化学物質へのばく露による影響として、臓器重量、酵素濃度等の連続量について測定したデータのこと。

9. Restriction

パラメータ最適化のプロセスの中でパラメータが取る値の範囲に制限を設けることであり、BMD法においては、数理モデルから得られる用量反応曲線が、生物学的に説明できない用量反応曲線とならないように、数理モデルに含まれるパラメータが取る値の範囲に制限を設けることとなる。

² $y=f(x, a)$ が、 a の種々の値に対して応答変数 y と説明変数 x の関係を表すと考えるとき、 a をパラメータという（変数の数が多い場合も同様）。

第3 食品安全委員会が行う食品健康影響評価におけるBMD法の活用

化学物質の食品健康影響評価においてBMD法を活用する際は、本指針に基づきBMD及びその信頼区間の算出を行う。

また、BMD等の算出に必要な生物学的及び統計学的な判断は、各分野の専門家の意見に従う。

なお、本指針に拠らない考え方又は手法を採る場合は、BMD等の算出結果の報告時にその内容及び採った理由を併記する。

1. BMD法を適用する用量反応データの収集、選択

(1) 評価対象とする化学物質（以下「評価対象物質」という。）に関して、その用量と特定の反応との関係を観察したデータ（以下「用量反応データ」という。）のうち、利用可能なものを全て収集する。

その際、利用可能な用量反応データの種類及び数によっては、専門家の意見に基づき、対象とする毒性試験を先に選択した上でエンドポイントを選択する、又は対象とするエンドポイントを先に選択した上で毒性試験を選択するという手順を取り得る。

(2) 毒性試験とエンドポイントの組合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ついて、以下の条件に該当するか否かを専門家の関与の下で確認する。

- ① 動物の種及び数、投与方法、投与量等の試験設計が適切であること。
- ② 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。なお、ヒトへの外挿性についても考慮する。
- ③ データセットが、BMD法を適用する上で必要となる以下の情報を含むこと。なお、用量反応モデリング³の際に共変量⁴となりうる変数（生殖・発生毒性における同腹児数等）が想定されるデータセットについては、それらの情報を含むことも確認する。
 - a 二値データ：各用量群における試験動物数及び反応を示した動物数。
 - b 順序カテゴリーデータ：各用量群における試験動物数、反応を示した動物数及び各個体の反応の程度（グレード）。

³ 化学物質のばく露量と当該物質によりもたらされる影響との間の用量反応関係に、数理モデルを当てはめることで用量反応曲線を得ること。

⁴ 主要な関心の対象でないが、応答変数に影響を与えられ、モデリング及び解析のために付け加えられる変数。

- c 連続値データ：各個体データ。これが利用可能でない場合は、各用量群における試験動物数、反応量の代表値（平均値又は中央値）及び反応量のばらつきに関する指標（標準偏差又は四分位範囲）

(3) 同質性が期待できるデータセットが複数存在する場合、標本数を増やし、より精緻な検討を行うため、専門家の関与の下、当該データセットを統合し、サンプリングバイアスを改善したデータセットとして解析することが可能である。これにより用量反応関係が不明確になること等を防ぐため、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試験設計の違いを把握し、事前に生物学的及び統計学的な観点からその妥当性を判断する必要がある。

さらに、原データ（個体データ）を活用することで、BMRに対応する用量域における毒性について、より詳細な解析が可能となることから、統合するデータセットの個体データが利用可能な場合は、それぞれ個体データまでさかのぼり、活用することが望ましい。

(4) 観察された影響（病理所見）について、グレード別の発生頻度が記録されている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は、慎重に検討すべきである。

具体的には、対照群における病理所見の発生頻度を考慮しつつ、病理所見の種類や、そのグレードの付け方及び発生頻度に基づき、専門家の関与の下で毒性学的意義があると判断されたグレード以上の病理所見が発生する頻度の合計値を二値データとして取り扱うことを検討する。

2. BMRの設定

(1) 二値データ

- ① 通常の動物試験における試験動物数から得られる統計学的精度に加え、化学物質間又はエンドポイント間における比較及びNOAEL法を用いた毒性評価との連続性⁵も考慮し、BMRとして過剰リスク⁶10%を用いた

⁵ S. Sand *et al.*, A Signal-to-Noise Crossover Dose as the Point of Departure for Health Risk Assessment. *Environmental Health Perspectives*. 2011, 119: 1766-1774.を参照。

⁶ 過剰リスクは、記録された反応の発生率のバックグラウンド時からの増加分である追加リスク： $P(d) - P(0)$ を、バックグラウンド時の反応の非発生率で除した値であり、過剰リスク= $\{P(d) - P(0)\} / \{1 - P(0)\}$ で定義される。

P(d)：用量 d における反応の発生率

P(0)：バックグラウンド時（対照群）における反応の発生率

BMD関連指標は基本的に算出する。

- ② 当該過剰リスク10%は、BMRの値として、あらゆる種類のデータセットに最適な値になるとは限らない。各BMRの値は、生物学的な意義付けのほか、用いる用量反応データの特性を考慮しつつ、専門家の関与の下で設定する。例えば、過剰リスク10%が、最小用量で観察された反応よりもかなり小さく、最小用量よりも大幅に低用量側の推計を伴う場合は、より大きい値が妥当な場合もある。また、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は、統計学的に、過剰リスク10%より小さい値が妥当な場合もある。

(2) 連続値データ

- ① 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。
- ② 連続値の用量反応データは多様な分布を示すことが想定されるため、生物学的に意味のある反応量の変化が不明な場合は、用量反応データごとに専門家の関与の下でBMRを設定する。
- ③ 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、反応量の分布を仮定した上で、カットオフ値を超える反応量を示す個体割合が、対照群に比べて、あらかじめ定めた増加分（BMR）だけ増える投与量をBMDとして定義した上で、BMD関連指標を算出するハイブリッド法⁷等を取り得る。
- ④ 科学的に合理的と思われるカットオフ値を設定でき、かつ、個体データが利用可能である場合は、情報量の減少を伴う点に留意する必要があるが、カットオフ値に基づい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BMD関連指標を算出することが有用である場合もある。

3. 用量反応モデリングと結果の評価及びPODの決定

(1) 用量反応モデリング

- ① 鋭敏なエンドポイントの排除を防止する等の観点から、1 (2) の条件を満たしたデータセットについては、全て用量反応モデリングを行う。
- ② 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

⁷ ハイブリッド法の手順の詳細については、同法が提案された K. S. Crump, Calculation of Benchmark Doses from Continuous Data. Risk Analysis, 1995. 15: 79-89 や、U.S. EPA, Benchmark Dose Technical Guidance. 2012.を参照。

量反応データを用いて用量反応モデリングを行う（最高用量群等の特定のデータを用量反応モデリングの際に除外することは行わない。）。

- ③ 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こととなる。
そのような「発現機序の本質を捉えた」数理モデルがある場合は、当該モデルからBMD等を算出することが優先される。
- ④ 「発現機序の本質を捉えた」数理モデルが利用できない場合は、BMD等の算出を行うソフトウェアやオンラインツール（以下「ソフトウェア」という。）を使用し、それらに収載されている数理モデルを用いて用量反応モデリングを行う。
- ⑤ ソフトウェアは、使用実績及びBMD等の計算方法に関する論理的背景も踏まえつつ、専門家の関与の下で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。
- ⑥ 使用するソフトウェアにおいて、モデル平均化⁸が可能である場合は、その機能を用いた用量反応モデリングを併せて行う。その際、モデル平均化は、収束⁹した全ての数理モデルを用いて行う。
- ⑦ 個別モデルを用いた用量反応モデリングを行う際にモデルのパラメータの値を制限するRestrictionが可能な場合は、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）は基本的には別個の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用量反応モデリングを行う。

（2）用量反応モデリング結果の評価

データセットごとに、モデル平均化の結果及び各数理モデルを用いた用量反応モデリング結果を、以下の場合分けに基づき評価する。

① モデル平均化を行っている場合

モデル平均化に供した個別数理モデルの用量反応モデリング結果につ

⁸ AIC等の指標を重みとして、収束した数理モデルに基づく推定反応量を用量別に加重平均した結果から新たな数理モデルを導き、BMD関連指標を算出する方法。

⁹ ソフトウェアによるパラメータ推定のための、最尤推定法を使った数値計算が完了すること（解が得られること）。

いて、用量反応データへの適合度等を評価した結果も勘案しつつ、モデル平均化の結果が妥当であることを専門家の関与の下で確認したときは、当該結果を採用する。

なお、その際、モデル平均化に供した数理モデルのうち、視覚的又は適合度検定¹⁰等によって用量反応曲線が用量反応データを明らかに説明できていないと判断されたものを除外した上で、モデル平均化を行い、その結果を評価するという手順を取り得る。

② モデル平均化により妥当性若しくは信頼性が十分に担保された結果は得られ難いと判断された場合、又はモデル平均化を行っていない場合

個別の数理モデルを用いた用量反応モデリング結果を以下の手順により評価する。

a 用量反応モデリングの結果得られた用量反応曲線について、専門家の関与の下、用量反応データへの適合度を視覚的に又は適合度検定等によって評価した結果、用量反応データを明らかに説明できていないと判断されたものは、あらかじめ評価対象から除外する。

なお、適合度検定を行う場合の有意水準は、基本的に0.1を用いることとし、用量反応モデリングに当たって特定のモデルを使用する理由がある場合は、状況に応じて異なる有意水準を用いる。

b 専門家の関与の下、各用量反応モデリング結果について、以下の項目全てについて評価する。

(a) 用量反応曲線の形状に生物学的な矛盾がないこと

(b) AICの値が、同値の最も小さい用量反応モデリング結果と比べて、+3以下の範囲にあること

(c) 得られたBMDの信頼区間の幅が、他の用量反応モデリング結果と比べて明らかに大きくないこと

(d) 得られたBMD及びその信頼区間の下限值が、最小用量と比べて著しく低用量側でないこと

c 評価の結果、全ての評価項目を満たす用量反応モデリング結果を選択する。なお、該当する用量反応モデリング結果が存在しない場合は、

¹⁰ 用量反応モデリングの結果得られた用量反応曲線が母集団における用量反応関係を示すと仮定した際に、実際に観察された用量反応データがサンプリングされる確率 p 値を求め、統計学的に通常起こりうるとしてあらかじめ設定する有意水準と比較する検定。 p 値 < 有意水準である場合は、前提とした仮定が統計学的に正しくないと判断して、同仮定を棄却する。

BMD法を適用する是非について専門家の関与の下で判断する。

(3) PODの決定

(2)の結果、データセットごとに単一又は複数の独立した用量反応モデリング結果が得られることとなる。

それらの中から、専門家の関与の下、ヒトへの外挿性等を踏まえた各エンドポイントの意義、観察用量範囲等の用量反応データの特性、BMD及びその信頼下限値であるBMDLの値等を基に用量反応モデリング結果を選択し、当該モデリング結果において算出されたBMDの90%信頼区間の下限値BMDLを、PODとする。

4. 結果の文書化

解析の結果は、以下の情報を含む形で文書化し、各専門調査会等における評価対象物質の食品健康影響評価の審議で利用する。

(1) 使用したソフトウェアの名称及びそのバージョン

(2) BMD法を適用した各データセットの情報

- ① 試験設計概要（試験動物種、エンドポイント、投与方法、投与期間等）
- ② 公開データの場合は、出典（著者名、雑誌名、年号等）
- ③ 試験結果概要
 - a 二値データの場合は、各用量群における試験動物数及び反応を示した動物数
 - b 順序カテゴリーデータの場合は、各用量群における試験動物数、反応を示した動物数及び各個体の反応のグレード並びに当該データを二値データ化した場合はその手順及び根拠
 - c 連続値データの場合は、各用量群における試験動物数並びに反応量の代表値（平均値、中央値）及びばらつき（標準偏差、四分位等）
- ④ 共変量として用いた変数がある場合は、その変数
- ⑤ データセットを統合した場合は、統合後のデータセットに関する③及び④の内容、測定方法等統合前のデータセット間で異なる試験設計及びデータ統合の妥当性に関する考察

(3) BMRの値及びその値を用いた根拠

- (4) BMD法を適用した各データセットにおける用量反応モデリング結果
- ① 各用量反応モデリング結果のプロット（モデル平均化並びに Restriction ON及びOFFの両方により得られた用量反応曲線を含む。）
 - ② 各用量反応モデリング結果における指標（AIC等）並びにBMD及びその90%信頼区間
 - ③ BMD等の信頼区間においては、その推定に用いた統計学的手法（プロファイル尤度法、ブートストラップ法等）
 - ④ モデル平均化を行った場合は、重み付けに用いた指標及び各モデルの重み
- (5) 各データセットにおける用量反応モデリング結果の評価の手順等
- (6) PODの値並びに当該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

第4 指針の見直し

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、必要に応じて本指針を改訂する。

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（動物試験で得られた用量反応データへの適用）に関する審議結果（案）についての意見・情報の募集結果について

1. 実施期間 令和元年8月21日～令和元年9月19日
2. 提出方法 インターネット、ファックス、郵送
3. 提出状況 2件
4. 意見・情報の概要及び食品安全委員会評価技術企画ワーキンググループの回答

	意見・情報の概要*	評価技術企画ワーキンググループの回答
1	<p>P.6「最高用量群等の特定のデータを用量反応モデリングの際に除外することは行わない。」について。</p> <p>（コメント）一概に「最高用量群等の特定のデータを除外しない」とするよりは、「最高用量群等の特定のデータを除去する場合は、妥当な根拠および除外前後のデータを提示すること」などとしてはいかがでしょうか。この前に記載されている「生物学的な根拠がある場合」に含まれるのかもしれませんが、特定のデータを絶対に除外してはいけないと誤解されることを懸念いたしました。ご検討済みと思いますが、全用量でモデリングして、フィッティングが思わしくない場合、米国EPAのガイダンスでは、専門家判断で最高用量データを除外してもよいとされており、この場合は十分な根拠説明および除外前後の解析結果を明示することとされています。</p>	<p>本指針（案）では、BMD法が、用量反応データ（第3の1（1）を参照。）の一部ではなく、用量反応データ全体に数理モデルを当てはめて得られた用量反応曲線からBMD及びその信頼区間を算出する方法であること、特定のデータの除外は用量反応モデリングに利用可能な情報量の損失につながることを考慮し、特定のデータの除外自体は否定しないものの、その実施を正当化する生物学的な根拠が必要であると判断し「生物学的な根拠がある場合を除き、最高用量群等の特定のデータを用量反応モデリングの際に除外することは行わない。」としています。</p> <p>“Benchmark Dose Technical Guidance”（U.S. EPA, 2012）においても、最高用量データの除外を含めた特定のデータの除外について、その企画、実施は慎重に行わなければならないとした上で、それを正当化する根拠を示さなければならないとしています。</p> <p>なお、同ガイダンスでは、毒性試験データで特定のデータを除外することは、用量群数の少なさから非常に制限されることに言及しています。このため、本指針（案）が対象とする動物試験で得られた用量反応データにBMD法を適用する場合において、特定のデータを除外した上で用量反応モデリングを行うことは、統計学的な観点からも一般的な手順ではないと考えます。</p>

<p>2</p>	<p>P.8 「ヒトへの外挿性等を踏まえた各エンドポイントの意義、観察用量範囲等の用量反応データの特性、BMD及びその信頼下限値であるBMDLの値等を基に用量反応モデリング結果を選択し、当該モデリング結果において算出されたBMDの90%信頼区間の下限値BMDLを、PODとする。」について。</p> <p>(コメント) 上記PODの基準にAICを含めて「ヒトへの外挿性等を踏まえた各エンドポイントの意義、観察用量範囲等の用量反応データの特性、AIC、BMD及びその信頼下限値であるBMDLの値等を基に用量反応モデリング結果を選択し、当該モデリング結果において算出されたBMDの90%信頼区間の下限値BMDLを、PODとする。」としてはいかがでしょうか。ご検討済みと思いますが、米国EPA、WHO/IPCSおよびEFSAのガイダンスでは、AICがPOD決定の基準に採用されております。</p>	<p>本指針(案)では、BMD法を活用して評価対象物質におけるPODを求めるに当たって、</p> <ol style="list-style-type: none"> ① 毒性試験とエンドポイントの組合せからなるデータセット毎に、複数の数理モデルを用いた用量反応モデリングを実施(指針第3の3(1))、 ② データセット毎に①の結果を評価(モデル平均化結果又は単一若しくは少数のモデリング結果の採用、指針第3の3(2))、 ③ データセット毎に行った②の評価結果を総覧し、PODとするBMDLを決定(指針第3の3(3)) <p>という手順を採用しています。</p> <p>御指摘の記述箇所は、上記③の手順に該当する記載です。</p> <p>AICについては、異なる数理モデル間で、同一の用量反応データへの適合度の良さを評価する際に用いる指標であり、異なる用量反応データに当てはめた数理モデル間の評価には用いることができません。このため、御指摘の上記③の手順ではAICを評価指標として用いていませんが、上記②の手順において評価指標の一つとして用いています。</p> <p>なお、”Benchmark Dose Technical Guidance”(U.S. EPA, 2012)、”Principles for modelling dose-response for the risk assessment of chemicals”(IPCS, 2009)及び”Update: use of the benchmark dose approach in risk assessment”(EFSA, 2017)のいずれにおいても、AICを評価指標として用いているのは、各ガイダンスで上記②に該当する手順になります。</p>
----------	--	--

* 頂いた意見・情報はそのまま掲載しています。