

別記様式第3号(第4の2関係)

研 究 成 果 報 告 書

令和 2年 6月30日

内閣府食品安全委員会事務局長 殿

(代表受託者)

住 所 神奈川県川崎市川崎区殿町3-25-26

機 関 名 国立医薬品食品衛生研究所

氏 名 広瀬明彦

印

平成30～令和元年度食品健康影響評価技術研究「課題名：ベンチマークドース手法の健康影響評価における適用条件の検討（課題番号：1804）」

委託契約総額 13,550,000円

上記委託研究について、食品安全委員会食品健康影響評価技術研究実施要領（平成17年5月18日食品安全委員会事務局長決定）第4の2の規定に基づき、別紙のとおり研究成果報告書を提出します。

平成30～令和元年度 食品健康影響評価技術研究 研究成果報告書（終了時）

（※研究全体の期間（初年度及び2年目以降）について記載する。）

研究課題名 （研究項目名）	ベンチマークドース手法の健康影響評価における適用条件の検討 （課題番号：1804） （3 新たなリスク評価方法等の確立（2）確率論的評価方法に関する研究）
主任研究者	研究者名： 広瀬明彦 所属機関： 国立医薬品食品衛生研究所

## I 研究期間及び研究目的等

### 1 研究期間

平成30年度～令和元年度（2年間）

### 2 研究目的

近年、化学物質の健康影響に係るリスクアセスメントの用量反応評価の客観的手法としてベンチマークドース（BMD）法が取り入れられるようになってきているが、実際に適用する場合には、適用できるエンドポイントの反応率の設定、適合性の良い数理モデルの選択手法などにおいて、国際的に標準化された手法がなく、恣意的な選択基準も使用されているのが実情である。本研究では、数理モデルやエンドポイントの選択におけるBMD手法の信頼性のより高い適用基準を確立するために、実データを基に拡張したシミュレーション技術を用いた検証研究を行い、BMD法適用ガイダンス案を作成する。

### 3 研究体制

研究項目名	個別課題名	研究担当者名（所属機関名）
（1） 動物実験データを 基にしたシミュレ ーション技術の活 用によるBMD法適用 における留意点特 定のための事例検 討とBMD法適用ガイ ダンス案作成	1) BMD法における妥当なカットオフとNOAELとの関連の検討	広瀬明彦（国立医薬品食品衛生研究所） 西浦博（北海道大学）
	2) BMD法のモデル選択基準に関する検討	広瀬明彦（国立医薬品食品衛生研究所） 西浦博（北海道大学）
	3) BMD法におけるモデル除外基準の考案	広瀬明彦（国立医薬品食品衛生研究所） 西浦博（北海道大学）
	4) BMD法におけるパラメータ制約の有用性の検討	広瀬明彦（国立医薬品食品衛生研究所） 西浦博（北海道大学）
（2） 病理学的データに BMD法を適用する 際の留意点	1) 個別課題名：病理学的データにBMD法を適用する際の留意点	梅村隆志（ヤマザキ動物看護大学） 広瀬明彦（国立医薬品食品衛生研究所）
（3） 疫学データに基づ くBMD法の適用	1) 個別課題名：既存の疫学データに対するBMD法適用の事例検討	西浦博（北海道大学）
	2) 疫学研究デザイン別のBMD法の使用実績に関する文献研究と技術的ピットフォールの特定	西浦博（北海道大学）

#### 4 倫理面への配慮について

本研究の実施については、研究の特性上、研究対象者に対する人権擁護上の配慮及び実験動物に対する動物愛護上の配慮を要することはなく、研究方法による研究対象者に対する不利益及び危険性の排除や説明と理解（インフォームド・コンセント）に関わる状況が生じることはない。

## II 研究内容及び成果等

(1) 研究項目名：動物実験データを基にしたシミュレーション技術の活用によるBMD法適用における留意点特定のための事例検討とBMD法適用ガイダンス案作成

1) 個別課題名：BMD法における妥当なカットオフとNOAELとの関連の検討

(研究担当者：広瀬明彦（国立医薬品食品衛生研究所）、西浦博（北海道大学）)

### 【研究内容・方法】

用量反応関係に関する動物実験データは、統計学的な観点から考えるとサンプル数が不足しがちであり、実験データのNOAELも極めて限られた数の動物で反応が見られなかった用量を指し示しているに過ぎず、評価対象となる母集団の本質的なNOAELを捉えている保証があるわけではない。その点を考慮してBMD法ではBMDL（BMRを与えるBMDの片側95%信頼区間下限値）を利用するが、その使用の妥当性と従来から使用されてきたNOAELとの関係については更なる検討を要していた。本項目の研究では、NOAELとBMDL<sub>10</sub>との間の関係について検討を行った。

### 【研究成果】

NOAELとBMDL<sub>10</sub>の比較においては理論的研究を実施し、2者間の比較検討を実施した。Wilsonスコア信頼区間を利用すると、全ドーズを通じたサンプルサイズMを利用して、BMDLにおける反応率qが以下で与えられる：

$$q_i = \frac{p + \frac{z^2}{2M}}{1 + \frac{z^2}{M}} - \frac{z}{1 + \frac{z^2}{M}} \sqrt{\frac{p(1-p)}{M}}$$

ここでpは超過リスクであり、例えばBMR10%であればp=0.10と定義するものである。Zは標準正規分布からの統計量であって、両側95%信頼区間であればz=1.96のように数値で置き換えられる。ここで、BMDの逆関数f<sup>-1</sup>があるとすると、BMDL自体が以下となる：

$$BMDL_i = f^{-1} \left( \frac{p + \frac{z^2}{2M}}{1 + \frac{z^2}{M}} - \frac{z}{1 + \frac{z^2}{M}} \sqrt{\frac{p(1-p)}{M}} \right)$$

これにより、BMDLとその反応率qの関係が1つの閉鎖系（closed form solution）として記述可能となった。また、BMDLにおける超過リスクqが明らかであるということは2項分布に従うシミュレーションが可能になるということを意味しており、これを通じてBMDLの分位（超過リスク）別の値とNOAELとの関係を定量的にも解析的にも検討可能にすることができた。

図1に数値計算結果をまとめる。縦軸がX=0つまり、2値反応を認める動物が0匹である確率（すなわちNOAELである確率）、横軸はベンチマーク反応（BMR）に相当する。M=12、つまり、1ドーズあたり4匹が3点観察（あるいは1ドーズあたり3匹が4点観察）される場合、BMDL<sub>10</sub>がNOAELである確率は90%以上であった。しかし、M=20、つまり、1ドーズあたり5匹が4点観察（あるいは1ドーズあたり4匹で5点観察）だとNOAELである確率は85%程度

となる。サンプルサイズが増えると同じBMDLでもNOAELである確率は低くなる結果となった。

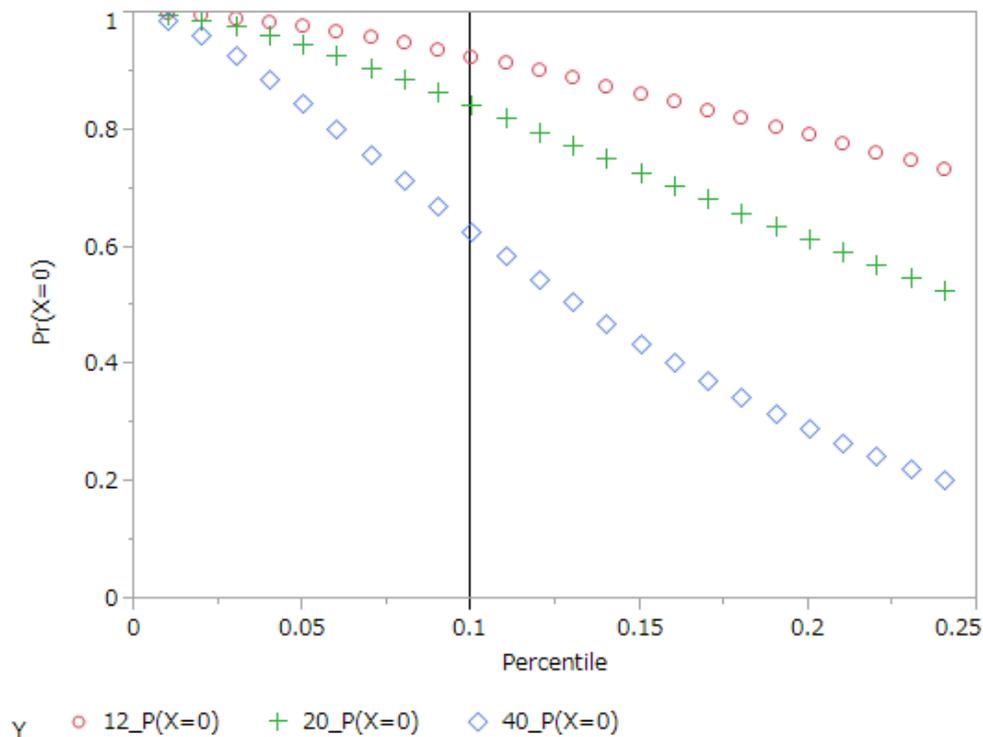


図1. 1群あたりの動物数の違いによるBMRと2値反応を認める動物が0匹である確率の関係

つまり、NOAELとBMDL<sub>10</sub>が近いと言われる根拠はサンプリングエラーに帰することが大きく、第1の理由は、サンプル数が少ない実験が多いからであると考えられた。1ドーズあたりで3匹や5匹の検討は、偶然ながら「ほどよく」80%以上の確率でNOAEL条件を満たす。また、同様の理由で、「ほどよくBMDより下にある偶然」が生じていることが同様の理由で明白である。つまり、もしもBMD<sub>10</sub>を規制値とすると、理論的にはそのドーズで10匹の実験を実施すると1匹が陽性となるはずである。BMDL<sub>10</sub>は、それよりも相当に低いDoseであり、BMRとして1/10より十分に小さい反応率をBMDL<sub>10</sub>が与えることとなるのである。

## 2) 個別課題名：BMD法のモデル選択基準に関する検討

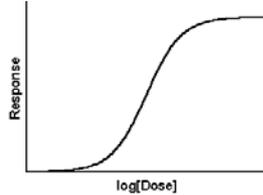
(研究担当者：広瀬明彦 (国立医薬品食品衛生研究所)、西浦博 (北海道大学))

### 【研究内容・方法】

シミュレーション技術を利用して、同じデータが与えられた場合のモデル選択基準について比較検討を実施した。用量反応関係に関する動物実験データは、統計学的な観点からするとサンプル数が不足しがちであり、実験データのNOAELも極めて限られた数の動物で反応が見られなかった用量を指し示しているに過ぎず、評価対象となる母集団の本質的なNOAELを捉えている保証があるわけではない。他方、シミュレーション技術を利用するとバイアスのないBMDL (unbiased BMDL) を定義した上で観察データをランダムに生成することができる。本研究では、それに基づいて仮想の実験データを作成し、BMDL等を推定し、本来的には推定できない「真の用量反応関係」を基にすることでモデル選択等の妥当性を評価した。

# シミュレーション手順

「真」の用量反応関係



1. 観察データ選択  
既存の動物実験データ: 毒性試験3点観察(1点につきn=10)、発がん性試験3点(1点につきn=50)  
=>  
・各2物質を任意に選択  
・観察誤差以外の混入ないデータと想定

2. 最も一般的な分布を適合  
閾値(原点)を持つ4パラメータあるいは5パラメータで構成されるデルタ分布、一般化ガンマ分布或いは一般化ワイブル分布

$$f(D - D_0; \theta)$$

閾値(原点)  $D_0$  は理論的NOAELを与える  
パラメータ  $\theta$  の不確実性情報は使わない

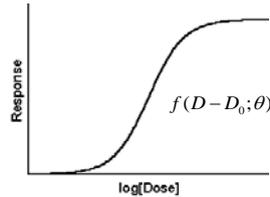


図2. 真のBMDLを仮定したベースラインの構築

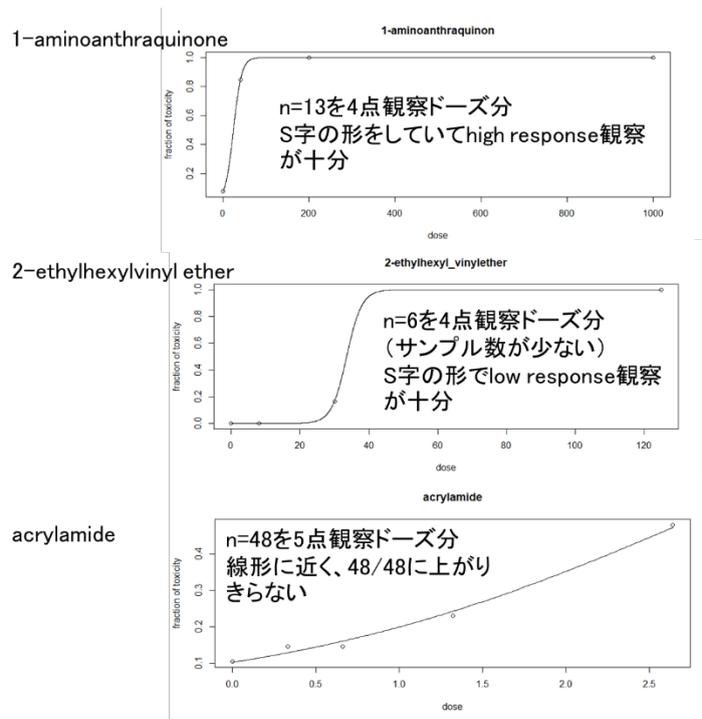


図3 3つの異なる観測パターン

データ量の多い既存のデータに基づき、実装成果に近い真のシミュレーション設定に基づいてモンテカルロ計算を実施し、多数の実験データを人為的に生成した。その際、実験ではミニマムで3点の計測が可能だとしたとき、3つの異なる観測パターンが考えられる(図3)。それらに対応して観察を仮定した個々の実験データにBMD法を1つひとつ適用し評価を行った。具体的には、まず観察データに対し、logistic, probit, loglogistic, logprobit, gamma, Weibull, multistage(polynomial degree = 2), multistage(polynomial degree = 3), quantal-linearの9つのモデルを当てはめた。その中でAIC(赤池情報量規準)が最も低い(=最も観察データにフィットしている)モデルとそのBMDL<sub>10</sub>値を「真の用量反応曲線」・「真のBMD<sub>10</sub>、BMDL<sub>10</sub>値」として求めた。次に、AICが最小のモデルの最尤値と分散共分散行列から1000セットの仮想実験データ

を得た。これら1000セットの仮想データそれぞれに9つのモデルをあてはめ、合計9000回のモデルのパラメータ推定を行った。それぞれの推定で得られたモデルに対してBMD法を適用し、カットオフ値（BMR＝毒性発現の閾値）に対するBMD、BMDLなどの値を計算した。

多くの事例において、用量反応曲線は高用量の情報を基に検討を加えたものが多く、低用量領域の暴露に関する情報が少ないにも関わらずBMD法のような統計モデルを用いて規制すべきドースを決定することが多かった。その具体的基準として以下が挙げられる：

- a) 8つ以上の統計モデルを適合し、その中で最小のBMDL<sub>10</sub>を与えるモデルを選択する。
- b) 8つ以上の統計モデルを適合し、その中で最小のBMD<sub>10</sub>を与えるモデルを選択する。
- c) 考え得る統計モデルの中から最小値のAICを与えるモデルを選択する。
- d) 一定のモデルを除外し、妥当と考え得るモデルを複数選択してモデル平均化を行う。

上記のいずれも低用量領域の適合あるいは予測が良いという保証はない。その場合は安全サイドに立ったモノの考え方をすることが多く、例えばa)の選択基準に基づく考え方は、安全性を重視した保守的なモデル選択方針と言える。しかし、それが真のBMDL未満か否かということや、BMDLより低すぎる値を与えていないかという点は疑問として残るため、シミュレーションを通じてこれらの点に関して一定の見解を提供すべく計算作業に取り組んだ。さらに、モデル平均化と他の選択基準の比較取り纏めを行い、ガイダンス案の基礎となる知見を充実させるべく研究作業を行った。特に、モデル平均化の中でもAICのトップ3の平均化のみを行う場合であったり、最小AICとの差がプラスマイナス3程度の範囲にある場合などを、全収束モデルを平均化した場合との比較も実施した。

また、近年はオランダのRIVMが開発したPROASTや米国EPAが開発したBMDSなどのソフトウェアもモデル平均化手法を積極的に取り入れており、本研究で開発してきた平均化手法との間でBMDL<sub>10</sub>値の比較検討を行った。

一方、シミュレーションに基づくモデル選択基準の検討結果に関しては原著論文として外部査読を受けた国際誌に掲載できるよう、その成果の創出に取り組んだ。

## 【研究成果】

3つの異なる観測パターンが考えられる3つの物質に関するデータを用いて上記の比較検討結果を表にして示す（表1-3）。

モデル選択基準について、以下のように比較検討を行った。シミュレーションデータから作られた9000個のモデルを用いて、以下のモデル選択基準により妥当性や信頼性などの検討を行った。モデル平均化については、全モデルのモデル平均化を行ったモデル、およびAIC上位3モデルによるモデル平均化を行ったモデル、最小AICとの差がプラスマイナス3未満のモデルによる平均化モデルを用いた。

- **LowestBMDL**：9つの統計モデルを適合し、その中で最小のBMDL<sub>10</sub>を与えるモデル
- **LowestBMD**：9つの統計モデルを適合し、その中で最小のBMD<sub>10</sub>を与えるモデル
- **LowestAIC**：9つの統計モデルを適合し、その中で最小のAICを与えるモデル
- **MA-all**：9つの統計モデルすべてのAICを考慮し、モデル平均化を行ったモデル
- **MA-3**：9つの統計モデルのうち、AIC上位3モデルを選択し、モデル平均化を行ったモデル
- **MA AIC<3**：9つの統計モデルのうち、最小AICとの差がプラスマイナス3未満となるモデルを選択し、モデル平均化を行ったモデル

妥当性と信頼性について、次のような点を検討した。

**信頼性**：1000セットの仮想データそれぞれに対し、各モデル選択基準により選択された1000個のモデルのBMDL<sub>10</sub>値と「真のBMDL<sub>10</sub>値」との平均距離およびその中央値を算出した。

**妥当性**：1000セットの仮想データそれぞれに対し、各モデル選択基準により選択された1000個のモデルのBMDL<sub>10</sub>値が、「真のBMD<sub>10</sub>値」より小さい割合を算出した。

これら基本的指標に加えて収束率と真の用量曲線復元率を計算した。前者は、ランダムシミュレーションの中で特定の選択基準を用いると解が収束しない場合の割合を指す。例えば収束率が80%というのは1000回のシミュレーションを実施した場合には800回だけ参照用量が得られる、という場合を指す。後者の真の用量曲線復元率とはシミュレーション前に利用した真と仮定される統計モデルがベストの結果として使用される割合を指す。例えば、ワイブル分布を真のモデルと仮定して数値計算に使用した際に、ワイブル分布が選択されたシミュレーションの割合がそれに相当する。シミュレーションに基づく研究を実施している際には、真と仮定したモデルが過度に選択されすぎて特定の選択基準が良いと判断されてはならないため、この基準を加味して検討した。

表1. 1-aminoanthraquinoneの観察データを基にしたシミュレーション結果

除外	モデル選択	信頼性	順位	妥当性 (%)	順位	収束 (%)	BMDL 計算可 (%)	除外込み 収束 (%)	除外込みBMDL 計算可 (%)	真用量曲線の復元 (%)
なし	lowestBMDL	0.39	5	100	1	95.6	95.6	95.5	95.5	0.1
	lowestBMD	0.28	2	100	1					0.1
	lowestAIC	121	15	88.4	15					34.1
	MA-all	6.20	9	99.6	8					-
	MA -3	4.73	7	99.8	7					-
	MA AIC <3	8.96	11	98.8	11					-
KS	lowestBMDL	0.39	5	100	1	95.6	95.6	95.5	95.5	0.1
	lowestBMD	0.28	2	100	1					0.1
	lowestAIC	121	15	88.4	15					34.1
	MA -all	6.13	8	99.6	8					-
KS, BMD/BMDL	lowestBMDL	0.34	4	100	1	95.6	95.6	79.1	79.1	0.5
	lowestBMD	0.23	1	100	1					0.5
	lowestAIC	121	17	88.4	15					38.4
	MA -all	6.29	10	99.3	10					-
KS, BMDU/BMDL	lowestBMDL	27.1	13	91.0	12	95.6	95.6	49.0	49.0	18.4
	lowestBMD	27.1	12	91.0	12					18.4
	lowestAIC	148	18	79.4	18					35.3
	MA-all	33.9	14	90.2	14					-

1-aminoanthraquinoneの結果解釈は以下のようにまとめられる：

- ・信頼性はlowest BMD/lowest BMDLを利用したものがベスト。Lowest AICが最悪、MAは7位以降。MAの中ではMA-3が良い。除外は方法によってまちまち。
- ・妥当性はlowest BMD/lowest BMDLを利用したものがベスト。Lowest AICが最悪、MAは7位以降。MAの中ではMA-3が良い。
- ・収束はMA-3やMA-AIC<3が最も良い。除外かつ収束でもMA-3とMA-AIC<3が良い。
- ・lowest AICだと元々の真のモデルが選ばれてしまう頻度が高い

表2. 2-ethylhexylvinyl etherの観察データを基にしたシミュレーション結果

除外	モデル選択	信頼性	順位	妥当性 (%)	順位	収束 (%)	BMDL 計算可 (%)	除外込み 収束 (%)	除外込みBMDL 計算可 (%)	真用量曲線の復元 (%)
なし	lowestBMDL	17.8	7	100	1	85.2	85.2	85.2	85.2	0
	lowestBMD	17.8	7	100	1					0
	lowestAIC	20.7	15	66.6	9					33.4
	MA-all	6.17	6	66.6	9					-
	MA -3	3.55	1	66.6	9					-
	MA AIC <3	5.11	2	66.6	9					-
KS	lowestBMDL	17.8	7	100	1	85.2	85.2	85.2	85.2	0
	lowestBMD	17.8	7	100	1					0
	lowestAIC	20.7	15	66.6	9					33.4
	MA -all	6.15	5	66.6	9					-
KS, BMD/BMDL	lowestBMDL	17.8	7	100	1	85.2	85.2	85.2	85.2	0
	lowestBMD	17.8	7	100	1					0
	lowestAIC	20.7	15	66.6	9					33.4
	MA -all	6.12	4	66.6	9					-
KS, BMDU/BMDL	lowestBMDL	17.8	7	100	1	85.2	85.2	85.2	85.2	0
	lowestBMD	17.8	7	100	1					0
	lowestAIC	20.7	15	66.6	9					33.4
	MA-all	6.11	3	66.6	9					-

2-ethylhexylvinyl etherの検討結果は以下のようにまとめられる：

- ・信頼性はMAが良く、中でもMA-3が良い。MA-AIC<3がそれに次ぐ。Lowest BMD/lowest BMDLは7位以降、lowest AICが最悪。除外の影響はほとんどない。除外によってMAの信頼性は軽微に改善。
- ・妥当性はlowest BMDL/lowest BMDが良い。MAとlowest AICは9位。
- ・収束はMA-3やMA-AIC<3が最も良い。除外かつ収束でもMA-3とMA-AIC<3が良い。
- ・lowest AICだと元々の真のモデルが選ばれてしまう頻度が高い

表3. Acrylamideの観察データを基にしたシミュレーション結果

除外	モデル選択	信頼性	順位	妥当性 (%)	順位	収束 (%)	BMDL 計算可 (%)	除外込み 収束 (%)	除外込みBMDL 計算可 (%)	真用量曲線の復元 (%)
なし	lowestBMDL	0.39	5	100	1	95.6	95.6	95.5	95.5	0.1
	lowestBMD	0.28	2	100	1					0.1
	lowestAIC	121	15	88.4	15					34.1
	MA-all	6.20	9	99.6	8					-
	MA -3	4.73	7	99.8	7					-
	MA AIC <3	8.96	11	98.8	11					-
KS	lowestBMDL	0.39	5	100	1	95.6	95.6	95.5	95.5	0.1
	lowestBMD	0.28	2	100	1					0.1
	lowestAIC	121	15	88.4	15					34.1
	MA -all	6.13	8	99.6	8					-
KS, BMD/BMDL	lowestBMDL	0.34	4	100	1	95.6	95.6	79.1	79.1	0.5
	lowestBMD	0.23	1	100	1					0.5
	lowestAIC	121	17	88.4	15					38.4
	MA -all	6.29	10	99.3	10					-
KS, BMDU/BMDL	lowestBMDL	27.1	13	91.0	12	95.6	95.6	49.0	49.0	18.4
	lowestBMD	27.1	12	91.0	12					18.4
	lowestAIC	148	18	79.4	18					35.3
	MA-all	33.9	14	90.2	14					-

acrylamideの検討結果は以下のようにまとめられる：

- ・信頼性はMAが良く、中でもMA-3が良い。lowest AICがそれに次ぐ。Lowest BMDLが最悪。除外の影響はほとんどない。除外によってMAの信頼性は軽微に改善。
- ・妥当性はlowest BMDLを利用したものがベスト。Lowest AICが最悪、MAは9位以降。MAの中ではKS検定とBMD/BMDL比を利用して除外した後のMAが良い。
- ・収束はほとんど変わらない。わずかにMA-3やMA-AIC<3が最も良い。
- ・lowest AICだと元々の真のモデルが選ばれてしまう頻度が高い（シミュレーション研究の影響か）

以上を総合すると、妥当性と信頼性に関しては、選択条件によって異なる結果が得られた。Lowest BMDL/lowest BMDが良い場合もあるが、それは過度に保守的であるため、最悪の結果を与える場合もある。Lowest AICに関しては、真のモデルが高い頻度で選ばれているので（現在の結果でさえ）結果が良く出てしまっている可能性がある。他方、MA-3を利用した場合、際立って悪い場合は認めず、また、他のMAよりも良い傾向が明らかであった。

PROASTやBMDS等の海外の平均化手法を用いたBMDソフトウェアとの比較については、平成22-24年度の「用量反応性評価におけるベンチマークドース法の適用に関する研究（研究課題番号1007）」において、解析に用いた検証用データのうち、NOAELが得られている172の二値データを用いて、NOAELに対するBMDL<sub>10</sub>の比を算出しソフトウェア間の比較を行った。PROAST（Version67.0）は、EFSAの統計モデルを公開しているWebページ（<https://shiny-efsa.openanalytics.eu/app/bmd>）の中で計算を行った。BMDSはV3.1.1をダウンロードして算出に用いた。さらにIndiana大学で公開しているBMDの計算サーバー（Bayesian Benchmark Dose (BBMD) Analysis System）上で計算を行った。因みにBMDSとBBMDについては、BMDLの算出手法としてベイズ法が使われている。ソフトウェア間のパフォーマンスの比較としては、NOAEL値に対する各ソフトウェアが算出した平均化手法によるBMDL<sub>10</sub>値の比を集計することによって行った。

表4. 各ソフトウェアによるBMDL値のNOAEL値に対する比率の集計結果

BMDL <sub>10</sub> /NOAEL	PROAST <sub>v67</sub> /NOAEL	BBMD /NOAEL	BMDS <sub>v3.1.1</sub> /NOAEL	MA-3 /NOAEL
Average	2.97	4.47	1.45	1.67
SD	5.69	4.29	1.79	1.35

NOAEL値が必ずしも理論的なBMDL<sub>10</sub>に近いことを示すわけではないが、本比較解析では、実験的に得られたNOAEL値に対して、各ソフトウェアがどの程度の違いとしてBMDL値を算出する傾向があるかについての比較を行うことができる。我々の手法（MA-3）と同様の頻度論的手法によるモデリングを採用しているPROASTはMA-3によるBMDL値に比べて概ね2倍程度高い値を算出する方向あることが判明した。一方ベイズ法を用いたモデリング手法を用いるBMDSとの比較において、MA-3はBMDSと同レベルのBMDL値を算出していた。しかし、同じベイズ法を用いる手法でもBBMDにより算出されるBMDLはBMDSより約3倍も高い値を示しており、平均化手法を用いた場合でもその設定次第で大きな違いを生み出すことが判明した。全体としてMA-3とBMDSによる平均化によるBMDLの値は平均的にNOAELの1.5倍程度の値を算出しており、より安全側の値を算出する傾向があることが示された。また、頻度論的手法やベイズ法を用いた何れの場合でも平均化手法を適用する際には初期設定等の条件設定の妥当についての検討が今後必要であることも示された。

### 3) 個別課題名：BMD法におけるモデル除外基準の考案

（研究担当者：広瀬明彦（国立医薬品食品衛生研究所）、西浦博（北海道大学））

#### 【研究内容・方法】

BMD法に基づくモデル選択を実施するにあたって、計算手順をガイダンスとして公表することを念頭に置くと、妥当な除外基準を事前に構築しておくことが役立つと考えられる。例えば、以下のようなモデル除外基準が考え得る：

- a) BMD<sub>10</sub>/BMDL<sub>10</sub>

b) LOAEL/BMDL<sub>10</sub>

c) BMDU<sub>10</sub>/BMDL<sub>10</sub>

このうち、モデル除外基準a)とc)はモデルの不確実性に関連するものであるが、モデル除外基準b)は実測値とモデルの下限値とのズレに関する情報である。それぞれに性格が異なるものであるが、いわゆる（質的に）「ひどい適合」のモデルを除外するために十分なのかどうかは検討の余地が残る。今年度は、上記で検討した選択基準に先立ってモデル除外を実施した場合の推定パフォーマンスの改善について取り纏めを行い、ガイドランス案の基となる知見を充実させるべく作業を実施した。また、前年度のシミュレーションに基づく除外基準の検討結果に関して原著論文として外部査読を受けた国際誌に掲載できるよう、その成果の創出に取り組んだ。

#### 【研究成果】

表1-3に示すようにシミュレーション研究では以下の3つの除外を検討した。

- **KS**:p-value ( $p>0.10$ )
- **KS, BMD/BMDL**:p-valueと BMD/BMDL比 ( $p>0.10$ と比10未満)
- **KS, BMDU/BMDL**:p-value と BMDU/BMDL比 ( $p>0.10$ と比10未満)

アクリルアミドを除き、Kolmogorov Smirnov検定やBMD/BMDL比の導入では妥当性の改善を認めなかった。また、3つの物質を通じて、信頼性に関しても顕著な改善を認めたものはなかった。

少なくとも、検討したような3つの実験データのような用量反応関係の理論的基盤から質的に大きく逸脱しない場合において除外の介在がモデルのパフォーマンスに大きく影響することがなかったものと考えられた。他方、本研究で検討していないような挙動を示すデータではその限りではないため、その点において注意することが必要であると考えられた。

#### 4) 個別課題名：BMD法におけるパラメータ制約の有用性の検討

(研究担当者：広瀬明彦（国立医薬品食品衛生研究所）、西浦博（北海道大学）)

#### 【研究内容・方法】

これまで、ワイブル分布やマルチヒットモデルなどに関してパラメータ制約を実施した場合とそうでない場合のモデルの妥当性が検討されてきた。本課題もちろんシミュレーションで一定の見解を提供することが可能であるが、他方、数理的分析によって用量に対する反応の起こり方が用量の微細な変化に伴って起こっているか否かなど、元々の観察データによって大きく異なる可能性が高い。そのため、前年度は本課題に限定して、上記の3点-5点計測の同じ実験データだけでなく、ワイブルやマルチヒットモデルなどの収束でピットフォールが生じ得る観察データを追加で仮定するなど、手作業での比較検討を実施してきた。特に、今年度は、パラメータ制約に関して欧州や米国で議論されてきた詳細について再度精査を実施し、その統計学的要点について非専門家にもわかりやすく説明できる状態を達成すべく研究作業に取り組んだ。パラメータ制約に関してはシミュレーションでの検討対象には入れておらず、また、本研究班に先立つ過去の研究班で制約の有る・無し両方を検討することが推奨されてきたので、同推奨が妥当なものであるかどうかについて、上記文献情報を基に議論し、ガイドラインに適切にその成果が反映できるよう検討を行った。

#### 【研究成果】

文献精査の結果、次の3点が明らかとなった。ワイブル分布のパラメータ制約は、低用量領域の反応に関するモデル化とそれが捕捉すべき反応とBMDL<sub>10</sub>の間に関する詳細な

議論であり、それは他の化学物質等へ広く一般化可能なものではないものと考えられた。マルチヒットモデルに関してはその想定複雑さと観察データの質的パターンの齟齬に伴うものであり、こちらも観察データ毎に検討すべき事象であることが明らかであった。今後、モデル平均化を実施するとき、理論的にはパラメータ制約を実施したモデルも平均化の対象とすることが可能であるが、他方、パラメータ制約とは特定のモデルが特定用量で反応する様子を子細にモデルで捉えることを目的として実施しているものであり、それと平均化の理論的哲学が必ずしも融和するものではない。上記の理由より、当面はソフトウェアに収録されたモデルを利用したモデル平均化を実施するときにはパラメータ制約を実施しないモデルのみを対象に実施すれば良いものと考えられる。

## (2) 研究項目名：病理学的データにBMD法を適用する際の留意点

### 1) 個別課題名：病理学的データにBMD法を適用する際の留意点

(研究担当者：梅村隆志 (ヤマザキ動物看護大学)、広瀬明彦 (国立医薬品食品衛生研究所))

本研究課題では、病理所見データにBMD法を適用する際の病理学的留意点に関して取り纏めを行うことに加えて、病理所見データを定量解析するための検討として、現在までに実際のグレーディングが付された毒性病理所見に対して、画像解析装置を用いた定量解析を行い、BMD法適用への適正化の可能性を探る検討を実施した。

### 【目的】

病理所見データにBMD法を適用する際には、複雑な毒性反応の結果としての病理所見名による分類を単純な二値化データとして処理することにより、BMD法を適用しているのが現状であるが、その際の限界と留意点を明らかにする。また、動物を用いた一般毒性試験で得られる病理所見データを定量化し、BMD法の適用範囲を広げる試みが模索されているが、病変の程度を観察者の印象によりグレード付けするなどの病理所見の特殊性が大きな障害となっている。そこで、いくつかの病理所見に対して、ヘマトキシリン・エオジン染色 (HE染色) 標本を用いて常法に従ったグレード付けによる解析、画像解析装置を用いた切片面積当たりの切片上の病変数の解析、さらには病変特異的蛋白を検出する免疫染色標本を用いた画像解析処理による解析結果を比較して、病理所見のデジタル化における問題点を検討することを目的とした。

### 【方法】

病理所見データにBMD法を適用する際の限界と留意点の検討については、JECFA等の国際的な評価機関でグレーディングのある所見を用いた例、複数の毒性試験等で得られた同一病理所見の発生頻度を合算した例、続発する一連の病理変化を採用した事例などを検鏡し、その課題について取り纏めを行った。また、いくつかの事例について、その病理所見の採用の妥当性と研究項目 (1) で開発したモデル平均化手法であるMA-3によるBMDL<sub>10</sub>値の妥当性を検証した。

グレーディングが付された毒性病理所見の画像解析による定量解析とBMD法の適用に関しては、アセタミドを0 (Control), 1.25%, 2.5%, 5.0%の用量で雄性F344系ラットに4週間混餌投与試験 (5匹/群) して得たホルマリン固定パラフィン包埋した肝臓のHE染色標本 (1切片/個体) について鏡検した。病理組織学的所見として、「封入体」、「アポトーシス」、「有糸分裂像」の3つの病変についてそれぞれグレード付けによる所見の収集を行った。また鏡検は盲検により群構成の情報未知のまま実施した。グレード付けは、病変が多い場合には+++、中間の場合には++、少ない場合には+とした。また、その後の処理として、+++を3、++を2、+を1として数値化し、群ごとにその平均値を求めた。一方、デジタル化データの収集は病変ごとに標本上のすべての病変をカウントし、

画像解析装置で計測した切片の総面積で除して、群内の個体ごとに数値を求め、群ごとにその平均値と標準偏差を求めた。さらに、細胞周期のG1後期からS期にかけて発現するタンパク質として知られるKI-67の免疫染色を実施した。計測値の算出は、肝臓各葉から無作為に3視野を選び、KI-67陽性細胞を数え、接眼メーターを用いて肝細胞3切片3箇所、計9つの総面積を測定後、観察面積当たりの病変数を算出した。

## 【結果】

病理所見データを直接用いてBMD法を適用する際の限界と留意点の検討

a) グレーディングのある所見のBMD法適用に関する留意点：

グレードのついた病理所見を用いてBMD法を適用した例を検索したが、現在のところ公表資料等で見つけることはできなかった。しかし、腫瘍病変の場合を考えると、病理組織診断能力を認証されている病理学者の診断であれば、良性または悪性であるかの判断は概ね合意が得られることから、誘発率に基づいてBMD法を適用することは可能ではないかと考える。一方、一般毒性所見については多くの問題点を含んでいると考えられた。特に非腫瘍性の毒性所見のグレーディングには様々な方法が混在している。病変の種類によって、その数を判断基準にする場合やその病変の広がりや判断基準にする場合など多様である。また、組織切片上でより高いグレードの病変を記録した場合、多くの場合、同じ切片上に認められるそれより低いグレードの病変は、毒性学的意義は無いまたは低いとみなされ、記録されないなど、グレーディングの程度を定量的に扱うには問題がある。以上より、グレードのついた毒性病理所見データを直接BMD法に適用するには、そのグレーディングの詳細についてどの所見レベルを定量化する必要があるかを検討すべきであり、個別の組織切片等の検索を行い詳細なデータなどを取り直す必要が出てくる。しかし、特に論文などで公表された過去の試験について、保存試料等を用いてグレード付けの再検索をすることはかなり困難であり、再検索によりグレーディングを正確に取り入れたBMD法の適用は、実際上ほぼ不可能である。従って、病理組織学的検索を開始する際にBMD法などによる解析を意識して用量相関性を示すことができるグレーディング基準等を設定することが望ましい。

b) 複数の毒性試験等で得られた同一病理所見の発生頻度の合算によるBMD法の適用に関する留意点：

複数の毒性試験等で得られた同一病理所見の発生頻度を合算してBMD法を適用した例を検索した結果、JECFAで評価が実施されたフモニシンで、遺伝的背景の異なるマウスの病変の発生頻度を合算している例が報告されている。具体的には、p53ヘテロ欠損マウスとその野生型で認められた巨大肝細胞の発生頻度が合算され、BMD法を適用していた。しかし、系統が異なる動物では、その病変の発生機序が全く同等・同質であることを担保するための情報が示されていないので、このような場合の発生頻度の合算は、食品安全委員会では不適として採用しなかった。JECFAならびにJMPRで評価されたデフルベンズスロンで、試験期間が異なる同用量の群の発生頻度を合算している例が報告されている。この場合も、試験（投与）期間が異なることから、同じ病理所見名が付されている病変であるとしても、その程度や随伴病変の有無などの違いを検討した結果として、同質の病変であるという保証は示されていないこともあり、その合算については詳細な検討が必要であると考えられる。一方、JECFAで評価が実施されたXanthan gumで、同じ動物施設、同じ環境で同ロットのXanthan gumを投与した追加試験のデータを合算してNOAELを求めた例が認められた。それぞれの試験で対照群を設定しており、且つ対照群間の差が認められないなどの両試験の生物学的な同等性を付加したものとなっており、発生頻度の合算は容認できるケースであると考えている。

c) その他、続発する病変を扱う場合や自然発生病変を取り扱う際の留意点：

JECFAにおける3-MCPDの評価では、ラットの慢性腎症と尿細管の再生像について、

慢性腎症はラット特有の病変であり、ヒトへの外挿性はないとの理由から尿細管の再生像をBMD法に適用する際のエンドポイントとして採用している。しかし、尿細管の再生は慢性腎症に続発する病変でもあることから、病理学的には両所見を切り離すことなく用量相関性を解析することが必要であると考えられる。また、JECFAにおけるアクリルアミドの評価では、その多種類の発がん性データから乳腺腫瘍やハーダー腺腫瘍の発生頻度がBMD法に適用されている。しかし、いずれの腫瘍も自然発生性腫瘍として考えられている腫瘍であり、その発生頻度の上昇に遺伝毒性機序の関与が疑わしいと考えられた。この事例のように、腫瘍頻度にBMD法を適用する場合、その腫瘍タイプの背景データを考慮する必要があると考えられる。

d) 具体的事例に基づく毒性病理学的観点からのBMD法適用に関する検討：（別添参照）

グレーディングが付された毒性病理所見の画像解析による定量解析とBMD法の適用：

病理組織学的検査のグレード付けでは、封入体はControl群（0%）で1.2、1.25%群で2、2.5%群で2、5%群で2.6、アポトーシスはControl群で1、1.25%群で1、2.5%群で2.4、5%群で1.2、有糸分裂像はControl群で1.2、1.25%群で1.2、2.5%群で2.8、5%群で1となった（表5）。個体ごとの病変の数をカウントし、画像解析ソフトで切片の面積で除した結果、封入体はControl群で $0.23 \pm 0.36$ 、1.25%群で $1.09 \pm 0.77$ 、2.5%群では $3.97 \pm 0.83$ 、5%群では $7.27 \pm 3.57$ 、アポトーシスはControl群では $0.04 \pm 0.01$ 、1.25%群では $0.08 \pm 0.03$ 、2.5%群では $2.05 \pm 1.28$ 、5%群では $0.43 \pm 0.54$ 、有糸分裂像はControl群では $0.12 \pm 0.06$ 、1.25%群では $0.09 \pm 0.12$ 、2.5%群では $1.54 \pm 0.29$ 、5%群では $0.08 \pm 0.11$ となった（表6）。また、単位面積当たりのKI-67陽性細胞の数はControl群で $18.7 \pm 6.08$ 、1.25%群で $13.5 \pm 3.79$ 、2.5%群では $274 \pm 54.5$ 、5%群では $82.8 \pm 53.7$ となった（表7）。

表5

用量(%)	0	1.25	2.5	5
封入体	1.2	2	2	2.6
アポトーシス	1	1	2.4	1.2
有糸分裂像	1.2	1.2	2.8	1

表6

用量(%)	0	1.25	2.5	5
封入体	$0.23 \pm 0.36$	$1.09 \pm 0.77$	$3.97 \pm 0.83$	$7.27 \pm 3.57$
アポトーシス	$0.04 \pm 0.01$	$0.08 \pm 0.03$	$2.05 \pm 1.28$	$0.43 \pm 0.54$
有糸分裂像	$0.12 \pm 0.06$	$0.09 \pm 0.12$	$1.54 \pm 0.29$	$0.08 \pm 0.11$

表7

用量 (%)	0	1.25	2.5	5
KI-67 陽性率	$18.7 \pm 6.08$	$13.5 \pm 3.79$	$274 \pm 54.5$	$82.8 \pm 53.7$

## 【考察】

基本的に、評価対象物質により誘発された病理所見のうち、最も毒性学的意義が高い所見の発生頻度に対しBMD法を適用すべきであるが、毒性試験の詳細な報告書を入手できる場合は、当該所見の発生頻度に関する集計データだけでなく、関連病変の発生状況や機序、対照群にも発生している場合は同一試験施設における背景データを確認し、用量相関性の有無を含めてBMD法適用に妥当な所見及び発生頻度かを判断することが求められる。具体的事例に基づく毒性病理学的観点からのBMD法適用に関する検討結果（別添）からは、評価対象物質の投与により誘発され、当該物質の毒性を評価する上で重要な所見について、病変の程度（グレード）毎に発生頻度データがある場合は、BMD法適用時に、総発生数を適用すべきか、ある程度以上の発生頻度に対して適用すべきか等、毒性学的意義や用量相関性等を考慮しながら検討すべきであることが示された。

一方、グレード付け所見に対して、グレードを単純に数値化して、その平均値を群の代表値とする考え方がある。今回はその方法に従ってグレードのついた病理組織所見を数値化したところ、封入体では低用量と中間用量間では差は認められなかったが、これを画像解析装置による切片面積当たりの個数でみると中間用量群は低用量群の約4倍の増加が認められた。また、アポトーシスのグレード付け所見の単純数値化では対照群と低用量群では差異は認められていないが、デジタル化数値では約2倍の増加が認められた。有糸分裂像では、単純数値化とデジタル数値化の何れにおいても中間用量群のみで増加が認められ、その増加の程度は単純数値化では約2倍にとどまっていたが、デジタル数値化では約17倍の増加となった。また、今回認められた有糸分裂像の増加は細胞増殖活性の増加を意味していることから、細胞周期特異的に発現するKI-67タンパク質の免疫染色によりKI-67陽性細胞の発現増加をデジタル数値化したところ、他の方法と同様に中間用量群で顕著な増加を示し、その増加の程度は約20倍となり、HE染色標本を用いた有糸分裂像発現率のデジタル数値とほぼ同程度となった。しかし、他の方法では認められなかった高用量群においても約6倍の上昇が観察された。以上より、これまで提案されているグレード付け所見の単純数値化では病変の程度を正確に反映することはできないことが観察標本面積当たりの病変数として求めたデジタル数値化との比較から明らかとなった。病理所見ではその所見名と生物学的反応が1対1の関係であるとは限らないが、今回認められた有糸分裂像の増加は細胞周期異常によるものではなく、単純な細胞増殖活性の増加であることから、細胞周期のG1後期からS期にかけて発現するKI-67のたんぱく発現を標本上で定量化した。その結果は有糸分裂像のデジタル数値化結果と一部は良く一致した。しかし、高用量群における変化を免疫染色標本による解析がより高感度に検出した。従来のHE染色標本による病理所見に対して、その原因となる生物学的反応の詳細が明らかな場合の免疫染色の利用の有用性が示唆された。つまり、デジタル化された画像解析結果は、病理所見をBMD法に適用するための代替手段としてではなく、毒性発現に関連する独立した有害性指標として扱い、免疫染色や画像解析による定量的な数値に用量反応評価としてBMD法を適用することが相応しいかどうかという観点での議論が必要であることが示された。

### (3) 研究項目名：疫学データに基づくBMD法の適用

- 1) 個別課題名：既存の疫学データに対するBMD法適用の事例検討  
(研究担当者：西浦博（北海道大学）)

#### 【研究内容・方法】

今年度は、疫学データにBMD法が適用された事例を集積するとともに、Bayesian BMD法のような疫学データに多い連続値に適用されるモデルの技術的問題点について事例を通じて提示可能になるよう文献検討とその系統的な解析を実施した。その際、疫学サンプルの代表性や閾値の定義についての動向を明らかに提示できるよう問題点の整理に取り組んだ。前年度からの事例検討を取り纏めて総説などが創出できるよう、その成果の

創出に取り組む。

#### 【研究成果】

動物実験データと疫学データの差異についてまとめることができた（表8）。ヒトのデータでは事件記録などを毒性評価に用いることができる極めて貴重なデータであると考えられ、また、連続反応が多いことが明確である。困難を伴うのは、ほぼ観察研究しかないため、無限に交絡因子が存在することである。よく陥りやすいピットフォールとして対象の代表性が問題になることが挙げられた。

表8. 動物実験データと疫学データの大きな違い

動物実験データ	疫学データ
ヒトへの外挿は困難 ヒトでできないことが検討できる	ヒトの毒性反応に基づく評価 事件など有用に利用できる
二値反応（発癌ある/なし）が多い BMRの設定にデフォルトがある	連続反応（血圧、臨床評価）が多い BMR設定のデフォルトに異論ある場合がある
理想的環境で暴露を繰り返せる 暴露量も調整可能である	ほぼ観察研究しかできない 無限に「交絡因子」が存在する 対象の代表性が良く問題になる
ミニマムサンプルで実施される	観察前のサンプル数計画が困難 なことも多い

特にドーズサンプリングの問題として、疫学データでは、高曝露群ほど標本数が小さいこと、標本数をn分割すると用量は等間隔にならないこと、低曝露域付近のみに集まり、高曝露域にある臨界濃度の推定が難しくなること、など、必ずしも理論的に用量反応関係を明確にする上で有益でないことが重なりやすいことが明らかとなった。

以上のような知見を中心にとりまとめ、BMD法を疫学データに適用する場合においては必ず次の事項が問題となるものと考えられた：

1. 代表性（臨界濃度付近をよく含んでいて、かつ、全人口に適應できるのか、という問題）、観察された情報（ドーズ）がどのあたりなのか
2. 交絡因子（観察研究である限り無限に存在する）
3. カットオフをどこにするのか、BMRをどのように設定するのか（リスクの種類、値）

以上までの知見をとりまとめた成果は、食品安全委員会評価技術企画ワーキンググループで発表する機会を得て、論点を含めて食品安全委員会委員らとディスカッションする機会を得ることができた。また、各事項に対する具体的な検討は次の2)個別課題において取り纏めた。現在までに、ベイズ法を利用したBMD法の疫学データへの適用に関して検討を続けており、レビュー論文の創出に関してもそれら問題意識を踏まえた上で専門家同士で共同執筆ができるよう引き継ぐことができた。

- 2) 個別課題名：疫学研究デザイン別のBMD法の使用実績に関する文献研究と技術的ピットフォールの特定（研究担当者：西浦博（北海道大学））

## 【研究内容・方法】

今年度は、疫学データにBMD法が適用された事例検討等を通じて、その適用の問題点をまとめたリスト化を実施すべく精査作業に取り組んだ。特に、代表性以外の問題点として交絡因子への対処や閾値付近の観察データの充実などについてのピットフォールを明確にできるよう務めた。また、前年度から着手した結果を取り纏めて総説などが創出できるよう、その成果の創出に取り組む。

中でも、細目課題として全体計画の通り、ハイブリッド法など既存の方法論の長短所に関する検討を実施した。諸外国では、連続値および不連続値の両方のアウトカムに対応した方法論が整理されており、それはハイブリッド法を含めて疫学研究領域で日本国内を含めて紹介されたことがある。他方、Bayesian BMD法などの登場で連続値モデルが大きく変化しようとしている。本検討では、ハイブリッド法やBayesian BMD法などを通じた推定技術について調査するとともに長所短所およびモデル選択におけるピットフォール等に関して検討した。既存の疫学研究における用量反応関係の評価結果について文献的検討を重ねつつ検討が進められるよう工夫しつつ作業を行った。

## 【研究成果】

疫学データにBMD法を適用する上で、特にHybrid法やBayesian BMD法の台頭を踏まえて、以下の項目を要点としてまとめた：

- ・疫学データはヒトの観察に基づく活用した用量反応評価を可能にする。そのため、ヒトの暴露量が推測可能なデータで用量反応関係が評価可能な場合は出来るだけその成果を加味することが望ましい。
- ・ほぼ観察研究（つまり、ランダム割り付けなどない研究）に基づくため、交絡因子は無限に存在し得る。そのことから、以下のような検討が求められる。

### 【分析時の検討】

・観察研究データを使用する場合は交絡因子の影響を拭い切れない。少なくとも観察情報として存在する交絡因子と未観察の可能性のある交絡因子を羅列し、前者の場合は調整を行った上でデータを分析することを検討するべきである。

### 【使用可能性の検討】

- ・因果推論に立脚し、DAG (Directed acyclic graph) など因果ダイアグラムなども駆使してアウトカムと関係する交絡を特定し、その生物学的メカニズムが想定される場合は調整を検討する。
- ・観察研究のサンプルの偏りなど、アウトカムに影響を及ぼす可能性のある観察可能な交絡因子に関しては、常に記述統計の情報として（用量反応関係の検討に先立って）整理することを習慣にする。
- ・傾向スコアマッチングなど交絡因子の調整に関する新しい因果推論の方法論が発展途上にあるため、そういった方法を利用した用量反応関係の推定について、今後研究が求められる。
- ・サンプル数が多くても高用量曝露領域が観察されにくい傾向がある。その場合、BMD法を適用したリスク評価を実施すること自体が難しくなる。そういった代表性に関する問題に対しては以下のような対処が求められる。

### 【分析時の確認】

臨界濃度の付近相当と考えられる曝露量は、動物実験から体重および種差の換算によって推測することが可能である。そのため、一義的には毒性学的検討を要する疫学研究が実施されており、そのデータに内在する用量反応関係を分析する際には、想定される臨界濃度付近の暴露があったものが観察対象に含まれるか否かを確認することが可能である。この点に関して必ずまずは検討を実施する。

### 【使用可能性の検討】

・ 臨界濃度に相当する曝露を伴わない場合、観察されたデータを用量反応関係の検討に使用すべきかどうかを検討する。例えば、ある化学物質の曝露量が十分でないことが全てのサンプルに共通しており、臨界濃度でなく一般的曝露量とアウトカムの因果関係のみを目的とするような研究は用量反応関係の検討対象とはならない。

・ 現時点で高用量曝露に対する反応率情報が存在しない中で、それを数理モデル等で推測しつつ補間する方法論は得られていない。今後、その実行可能性について検討を要する。

・ 他方、十分な曝露量を伴う事故の疫学データ等が得られる場合、用量反応関係の検討にはそういった臨床情報の集積に基づいて実施するべきである。また、そういった臨床情報と同等のアウトカム発生頻度に関して低用量曝露を伴う集団データとの結合（データ統合）による分析も今後検討を要する。

・ また、観察対象が一般的でない場合、そこにサンプル自体の代表性の問題が生じる。毒性学的には一部の地域住民の観察に基づく疫学情報が少なくないため、分析後の検討としてできるだけ議論しておくことが望ましい。

・ プールしたデータへのBMD法適用、モデル平均化、Hybrid法を超えたBayesian Impairment Thresholdなど技術的発展は続いており、今後のリスク評価に高い可能性を与える。加えて、既に系統的レビューの方法を用いて、参照用量のメタ分析を実施するような研究は出版されはじめており、複数のデータセットあるいは複数の研究が存在する場合の妥当な評価方法のガイドラインが求められる。

また、連続データへの適用が想定される各用量反応モデリング手法（関数外挿からのモデル選択若しくは平均化、又は、反応率等の分位を利用したHybrid法若しくは分位点回帰）について、下記の点に注意することが求められる。

#### 【分析時の検討】

・ 動物実験データはアウトカム反応量が2値であることが多いが、疫学データはアウトカムに連続量を選択し得る。分析にあたってはいずれの量をアウトカムにし得るのかを先立って整理して分析の戦略を立てることが求められる。連続量に関してはその分布を分位に変換して検討することの可否を戦略検討時に考慮するべきである。

・ 疫学データは動物実験データと比較してサンプル数が多い場合がほとんどである。動物実験で2値反応を扱う場合のBMRとして10%が採用されることが多いが、疫学データではそれに相当する量について定例的に受け入れられているBMRは10%よりも低い場合があり得る。そのため、こういった検討のために連続量におけるBMRの考え方に関して、分位に変換した場合のBMRのあり方について整理・研究しておくことが望ましい。

#### 【使用可能性の検討】

・ 連続量を分位に変換した場合のカットオフ値やBMRの選定においては動物実験による評価結果との一致性も踏まえて検討することが求められる。

・ 連続量データを分位として扱う場合、その分布に正規性を仮定することが可能ならばハイブリッドモデル、そうでなく偏った分布であったり、分位点そのものを利用した回帰をすることが求められたりする場合、ベイズ推定の手法を活用した分位点回帰などを検討することが求められる。

・ シミュレーションや理論的検討に基づくBMRの提案は重要な課題になり得る。疫学データにおいてBMRが1%や5%で検討されることがあるが、それらの持つメカニズムとしての意味が不明瞭な場合は、それら値を用いる場合の他データとの比較可能性や統計学的なパフォーマンスについて十分に分析しておくことが求められる。

・ 連続値データへの適用が想定される各用量反応モデリング手法（関数外挿、Hybrid法、分位点回帰等）の長所、短所

関数外挿法：数値計算が比較的容易であり、関数の意味論としてもメカニズムが既知の場合が多い。他方、データの連続量範囲でのバラつきや分位について十分に加味できない。

Hybrid法：連続量を分位として取り扱う最も簡便な方法であり、血圧や体重など正規性を仮定できるデータの取り扱いに向いている。計算量もベイズ法を用いる場合と比較して市販の計算機の処理能力で事足りる。しかし、正規性の仮定が必ずしも成立しない。

分位点回帰：理論的に曝露量による分位の変化や分位点ベースのリスク評価を可能にする点で、現在ある選択肢の中では最も理論的に頑健なモデルであると考えられる。ただし、ベイズ推定を用いた推定など計算には専門家の介入を要する。

・用量反応モデリング手法の選択に当たって考慮すべき情報

1次データへのアクセス可能性、観察されたサンプル数、連続量としてのデータの分布の仕方、計算環境（マルコフ連鎖モンテカルロ法が実装可能な状態か等）、推定専門家とのコミュニケーション体制、などを加味して分析手法の選択を行うことが求められる。

以上を踏まえて、食品安全委員会評価技術企画ワーキンググループでは、2019年10月にベイズ法とベイズ推定を活用したBMD法に関する勉強会を開催し、その場で発表する機会を得た。連続値データを中心として素早い動きで進展を認めるBayesian BMD法について食品安全委員会委員・専門委員らと共有することができ、今後検討すべき課題の整理を実施することができた。なお、ベイズ推定を活用したBMD法の適用条件については、別研究課題で検討を重ね、成果発表を行う予定である。

## （2）研究全体の成果、考察及び結論

研究実施前から計画してきたBMD法の動物実験データ（特に2値反応）への適用条件に関して、予定していたシミュレーションに基づく方法論の評価をまとめることができた。様々なモデル除外・選択基準のある中で、モデル平均化を実施することが妥当性と信頼性のそれぞれの点で優れていることを示すことができた。特に、全てのモデルを平均化することで不確実性をむしろ増大させてしまうことがないよう、適合度が良い順に3つのモデルを選定して平均化を行う、というオプションを選択した方がパフォーマンスが高くなることを示すことができた。これらの研究成果は、2019年10月29日に内閣府食品安全委員会から発表された「食品健康影響評価におけるベンチマークドース法の活用に関する指針」の下支えをする科学的基盤を提供することに繋がった。また、病理学的データの分析については、病理学的データをBMD法に適用する際にどのような論点があるかを示し、検討項目に対する対応案や事例紹介を行うことができた。さらに、画像解析により得られた定量的データにBMD法を適用する場合、毒性学的意義付けを明確にして用量反応性を表すものとして評価できるものであることを確認することが重要であることを確認できた。疫学データへの適用に関しては、まずは疫学データに特徴的な交絡因子の排除やデータの代表性の確保を行うことは定量評価として最低限必要であるが、さらに、BMD法を適用する際の影響指標のカットオフ値やBMRの選定について毒性学的意義付けのための検討が必要となる。本研究によりHybrid法やベイズ推定を用いた解析手法の開発に向けた課題や留意点等を抽出することができたため、新たな疫学データのためのBMD法研究、特に、ベイズ法を利用したBMD法を用いる連続値データ分析手法開発研究へと繋げることができた。

## III 本研究を基に発表した論文等

### 1 本研究を基に発表した論文と掲載された雑誌名のリスト

（投稿準備中）

### 2 本研究を基にした学会発表の実績

- Yoshiil K, Nishiural H, Inoue K, and Hirose A. Simulation-Based Assessment of Model Selection Criteria during the Application of Benchmark Dose Method to Quantal Response Data. Society of Toxicology (SOT) 58th Annual Meeting (SOT2019), (2019. 3. 10-14), Baltimore, USA. (The abstract is available from Toxicologist. (Toxicological Sciences. 173 (1), Suppl. 426, 20219))
- Keita Yoshii, Hiroshi Nishiural, Kaoru Inoue, Akihiko Hirose. Assessing model selection criteria during the application of benchmark dose method to quantal response data: Japanese perspectives. 15th International Congress of Toxicology (米国ワイキキ市、2019年7月15-18日)
- 吉井啓太、西浦博、井上薫、山口崇幸、広瀬明彦. BMD法適用におけるモデル選択手法の定量的評価. 第90回日本衛生学会学術総会 (岩手県民情報交流センター アイーナ、2020年3月26日(木) ~ 3月28日(土)) ポスター発表予定
- Yoshii K., Nishiura H., Inoue K., Hirose A. Evaluation of Model Selection Criteria Including Model Averaging during the Application of Benchmark Dose Method to Quantal Response Data (PS2253). Society of Toxicology (SOT) 59th Annual Meeting (SOT2020), (2020. 3. 14-18), Anaheim, USA (The meeting was cancelled by COVID-19 epidemic, but the abstract is available from Toxicologist. (Toxicological Sciences. 174 (1), Suppl. 299, 2020).)
- Hirose A, Yoshii K, Inoue K, Shigeta Y, Matsumoto M, Kawamura T, Yamaguchi T, Nishiura H. Performance Comparison of BMDL Calculation by the Model Averaging Methods for Quantal Dataset (PS1573). Society of Toxicology (SOT) 59th Annual Meeting (SOT2020), (2020. 3. 14-18), Anaheim, USA (The meeting was cancelled by COVID-19 epidemic, but the abstract is available from Toxicologist. (Toxicological Sciences. 174(1), Suppl. 131, 2020).)

### 3 特許及び特許出願の数と概要

特になし。

### 4 その他 (各種受賞、プレスリリース、開発ソフト・データベースの構築等)

特になし。

## IV 研究開始時に申告した達成目標及び研究全体の自己評価

### 1 達成目標の自己評価

達成目標	評価結果	自己評価コメント
(1) 食品安全委員会のリスク評価において使用される動物実験データに対応したBMD法の適用に関して、最適なモデルの選択ガイダンス案の作成	5	モデル選択基準や平均化について進めることができ、ガイダンスの公表や論文化を達成できた。
(2) 病理所見データにBMD法を適用する際に考慮すべき課題の整理と、その対応案の策定	4	各種病理所見データの特徴に応じて、BMD法適用の可否についての判断法や、適用するための大まかな対応案について取りまとめることができた。

(3) 既存の疫学データに対するBMD法適用の事例の整理と疫学研究データにおけるサンプル集団を基にした推定のピットフォールの特定	5	疫学データへのBMD法の適用について論点を整理して総説論文の基盤を達成でき、また、ワーキンググループにて疫学データへの適用とベイズ推定の活用など複数回の報告・共有ができた。

注) 評価結果欄は「5」を最高点、「1」を最低点として5段階で自己採点すること。

## 2 研究全体の自己評価

項目	評価結果	自己評価コメント
(1) 研究目標の達成度	5	モデル選択基準において平均化手法を取り込むための細部の論点を詰めることができた 他、病理所見データへの定量化や疫学データの対応法に関する論点整理に見通しがついた。
(2) 研究成果の有用性	4	より最適化されたBMD法を用いて様々な用量反応データを客観的に評価できる手法の開発が期待できる。
総合コメント 今後、食品安全委員会のリスク評価においてBMD法を適用した客観的な用量反応評価指針等の作成に寄与できた		

注) 評価結果欄は、「5」を最高点、「1」を最低点として5段階で記述すること。

この報告書は、食品安全委員会の委託研究事業の成果について取りまとめたものです。本報告書で述べられている見解及び結論は研究者個人のものであり、食品安全委員会としての見解を示すものではありません。全ての権利は、食品安全委員会に帰属します。

(別紙) 研究成果の概要 (英文)

Title of research project	Research on application condition of the benchmark dose methods for health risk assessment
Research project number	1804
Research period	FY 2018 – 2019
Name of principal research investigator (PI)	Akihiko Hirose

### Abstract/Summary

Benchmark dose (BMD) method has been increasingly used for risk assessment practice, including the determination of the point of departure (POD) for toxicological assessment of chemical substances. Besides, model selection criteria for the BMD method have been highly variable by existing guidelines of different organizations. Also, inconsistent applications of pathological data to the BMD method are conducted in different organizations. As for application of BMD method to epidemiological data, some pitfalls exist. Therefore, it is necessary to establish guidance for application of BMD approach to the regulatory risk assessment.

In this study, we performed literature survey, conducting systematic review analysis to understand technical issues surrounding the application of BMD method, and evaluated model selection criteria by using simulation-based approach.

The mathematical exercise revealed that the similarity between the two are supported by limited sample size during animal experiments, coincidentally meeting the necessity of  $BMDL_{10}$  due to sampling error. Model selection criteria were evaluated, restricting our analysis to the application of BMD method to quantal response data using simulation-based approach. Candidate criteria included the use of 1) smallest  $BMDL_{10}$ , 2) smallest  $BMD_{10}$ , 3) smallest AIC, and 4) an option to take model averaging values. The model averaging with 3 models with the lowest AIC values (MA-3) returned reasonable results overall, while it did not always yield the best performance. We gathered the information of simulation-based performance by applying Kolmogorov-Smirnov test p-value and the ratios of  $BMD/BMDL$ , and  $BMDU/BMDL$ . No exclusion criteria we considered in the present study greatly improved the performance of evaluation, i.e. reliability and validity, in none of dose-response patterns. To assess the added values of parameter restriction or imposing statistical constraints for parameters, we have discussed its details proposed by literature, and concluded that they do not have to be applied.

Consideration on the application of BMD method to pathological data, we summarized issues and points of note for applying the BMD method to histopathological findings. Based on some case studies using published toxicity data, setting of clear criteria for grading in the histopathological analysis was needed to consider application of the BMD methods. To combine incidence data derived from different toxicity studies or findings, biological equivalence of the background or mechanism of the corresponding histopathological findings should be clarified. Quantitative data by image analysis revealed that the imaging data could be used as markers related to the toxicologically significant findings.

As for application of BMD method to epidemiological data, we have enumerated and compared technical characteristics of epidemiological data against those of animal experiment data. Epidemiological data tend to involve evaluation of continuous response data, and the incidence data can be usable for the analysis of BMD method. We noted that the representativeness, i.e. if the quantified outcome is applicable to the entire population, confounders, cut-off value, and quantiles are often problems, whenever the application of BMD method to epidemiological data is considered.

If a technical guidance of the BMD methodology with a single recommended criterion for applying animal experimental data is required, we considered that one may recommend

MA-3 method as the method to calculate BMDL<sub>10</sub> as a POD to the reference dose. When the pathological data or related imaging data are used for the BMD method, we should verify whether the quantitative data is appropriate for explaining toxicologically significant dose-responses. We found the possible points to be considered below at the application of BMD method to epidemiological data: \*The infinite number of confounders may be related since the epidemiological is usually based on observational researches (i.e. nothing such as randomization). \*Studies sometimes includes the smaller number of high-dose data. \*Quantiles of observed continuous data are sometimes unknown. The development of BMD methods (e.g. model averaging, hybrid model, and Bayesian-based BMD method) possibly changes the methods of risk assessment.

This report provides outcome of the captioned research programme funded by Food Safety Commission Japan (FSCJ). This is not a formal publication of FSCJ and is neither for sale nor for use in conjunction with commercial purpose. All rights are reserved by FSCJ. The view expressed in this report does not imply any opinion on the part of FSCJ.

1. List of papers published on the basis of this research

None

2. List of presentations based on this research

Yoshii K, Nishiura H, Inoue K, and Hirose A. Simulation-Based Assessment of Model Selection Criteria during the Application of Benchmark Dose Method to Quantal Response Data. Society of Toxicology (SOT) 58th Annual Meeting (SOT2019), (2019.3.10-14), Baltimore, USA. (The abstract is available from *Toxicologist*. (Toxicological Sciences. 173 (1), Suppl. 426, 20219))

Keita Yoshii, Hiroshi Nishiura, Kaoru Inoue, Akihiko Hirose. Assessing model selection criteria during the application of benchmark dose method to quantal response data: Japanese perspectives. 15th International Congress of Toxicology (Hawaii, USA), 2019.7.15-18.

Yoshii K., Nishiura H., Inoue K., Hirose A. Evaluation of Model Selection Criteria Including Model Averaging during the Application of Benchmark Dose Method to Quantal Response Data (PS2253). Society of Toxicology (SOT) 59th Annual Meeting (SOT2020), (2020.3.14-18), Anaheim, USA (The meeting was cancelled by COVID-19 epidemic, but the abstract is available from *Toxicologist*. (*Toxicological Sciences*. 174 (1), Suppl. 299, 2020).)

Hirose A, Yoshii K, Inoue K, Shigeta Y, Matsumoto M, Kawamura T, Yamaguchi T, Nishiura H. Performance Comparison of BMDL Calculation by the Model Averaging Methods for Quantal Dataset (PS1573). Society of Toxicology (SOT) 59th Annual Meeting (SOT2020), (2020.3.14-18), Anaheim, USA (The meeting was cancelled by COVID-19 epidemic, but the abstract is available from *Toxicologist*. (*Toxicological Sciences*. 174(1), Suppl. 131, 2020.)

3. The number and summary of patents and patent applications

None

4. Others (awards, press releases, software and database construction)

None