

食品安全委員会評価技術企画ワーキンググループ

(第7回) 議事録

1. 日時 平成29年8月31日(木) 10:00～11:34

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) ベンチマークドーズ(BMD)法に関する現状と課題について
- (2) その他

4. 出席者

出席専門委員

(評価技術企画ワーキンググループ専門委員)

川村座長、広瀬座長代理、赤堀専門委員、岡田専門委員、小坂専門委員、小関専門委員

(食品安全委員会)

佐藤委員長、山添委員、山本委員、吉田委員

(事務局)

川島事務局長、小平事務局次長、吉田評価第一課長、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、井上課長補佐、山原評価専門官、柳澤係長、比留間技術参与、石神技術参与

5. 配布資料

資料1 「用量-反応関係分析の課題」

資料2 「ベンチマークドーズ(BMD)法の最近の動向について」

6. 議事内容

○川村座長 定刻になりましたので、ただいまから第7回「評価技術企画ワーキンググループ」を開催いたします。

先生方にはお忙しい中、今日は少し雨模様の中お集まりいただきまして、どうもありがとうございます。

本日は評価技術企画ワーキンググループに所属する専門委員6名の先生方に御出席をい

ただいております。

また、食品安全委員会からは4名の委員が御出席です。

議事に入る前に事務連絡がありますので、事務局、よろしく申し上げます。

○井上課長補佐 事務局より2点ございます。

まず、本ワーキンググループにおきまして、先生方におまとめいただきました(Q) SAR及びRead acrossの報告書「新しい時代に対応した評価技術の検討～化学物質の毒性評価のための(Q) SAR及びRead acrossの利用～」につきましては、7月25日開催の第659回食品安全委員会において報告され、内容が承認されましたので報告申し上げます。

このことにつきまして、佐藤委員長から御礼申し上げます。

○佐藤委員長 おはようございます。

先生方には評価技術企画ワーキンググループの最初の議事だったわけですが、(Q) SAR及びRead acrossを取り上げていただいて、これらの評価方法の食品健康影響評価での活用に向けて御議論いただいたわけです。

その経過を報告書にまとめていただいたわけですが、先ほど井上から話がありましたように、先月の委員会において報告され、承認されました。

今後、食品安全委員会では報告書で提案された内容に基づいて、(Q) SAR及びRead acrossの活用に向けた取り組みを、スピード感を持って進めてまいりたいと思っております。

先生方には引き続き御協力いただけたらと思っております。

報告書の取りまとめまでの間、川村座長を初め、専門委員の先生方の皆様には多大なる御尽力をいただきまして、委員長からも、また、食品安全委員会からも御礼を申し上げる次第でございます。どうもありがとうございました。

○井上課長補佐 次に、事務局の人事異動について報告いたします。

7月9日付で事務局次長の東條が退職いたしまして、後任として7月10日付で小平が着任しております。

○小平事務局次長 おはようございます。よろしく申し上げます。

○井上課長補佐 また、7月11日付で評価第一課長の関野が異動いたしまして、後任として吉田が着任しております。

○吉田評価第一課長 どうぞよろしくお願いたします。

○井上課長補佐 また、7月10日付で評価第二課長の鋤柄が異動いたしまして、後任として吉岡が着任いたしております。

○吉岡評価第二課長 よろしくお願いたします。

○井上課長補佐 事務局からの連絡は以上でございます。

○川村座長 ありがとうございます。

先ほど佐藤委員長から身に余るお言葉を頂戴いたしましたけれども、報告書をまとめたとはいえ、まだ本当に緒に就いたばかりですし、その後でおっしゃったスピード感を持つてというところを肝に銘じまして、きちんと進めてまいりたいと思っております。

それでは、議事に入ります。本日の議題は「ベンチマークドーズ（BMD）法に関する現状と課題について」です。

開催通知等で連絡いたしましたように、本日の会議は公開で行いますので、よろしくお願いたします。

事務局より資料の確認をお願いいたします。

○井上課長補佐 お手元に議事次第、座席表、専門委員名簿のほか、資料1「用量－反応関係分析の課題」、資料2「ベンチマークドーズ（BMD）法の最近の動向について」を先生方の机上に用意してございます。こちらは近日中にホームページに掲載されます。

配付資料の不足等はありませんでしょうか。何かありましたら事務局までお申し出ただけたらと思います。

また、机上配付資料の「今後の審議における主な論点（案）」につきまして、専門委員の先生方にお配りしております。

なお、内閣府では5月1日からクールビズを実施しておりますので、御理解・御協力のほどよろしくお願いいたします。

○川村座長 ありがとうございます。

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を行ってください。

○井上課長補佐 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いたします。

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、平成15年10月2日委員会決定に規定する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ついて、相違はございませんでしょうか。よろしいですね。

それでは、議事1に入らせていただきます。「ベンチマークドーズ（BMD）法に関する現状と課題について」です。

ベンチマークドーズ法は、用量－反応関係をモデル化し、反応の大きさから無毒性相当値を決める方法です。食品安全に関わる国際機関や欧米のリスク評価機関では、汚染物質やかび毒等の毒性評価に活用しています。また、本食品安全委員会においても、グリシドールやアクリルアミド等の非意図的に汚染される物質の食品健康影響評価において、既に活用している技術です。

ベンチマークドーズ法では、動物実験やヒトの疫学調査などから得られた用量－反応評価の結果から得られる無毒性量相当値を決定するために、15ほどの数理モデルによって推計された数値の中から最適なものを選択する必要がある、食品健康影響評価では、調査審議の都度、適切な選択基準を議論してきました。

私自身も、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きた経験があ

りますけれども、従来からNOAEL/LOAEL法やベンチマークドーズ法の統計学的合理性、それから、レギュラトリーサイエンスの中での活用の意義といったものを十分に理解し、共通の基盤に立つということが重要ではないかと感じたこともありました。

ベンチマークドーズ法は1980年代に提唱されて、また、各国で活用されてきた経験もありますけれども、本法の統計学的な考え方がリスク評価機関によって異なっている部分がありまして、同じデータを用いても異なった基準が出されるといった問題があり、専門家による統計学的な議論が始まっております。この中で、こういった問題に対応するために複数のモデルをそれぞれの適合度から重みづけをして平均し、単一のBMDあるいはBMDL10を得るという「モデルの平均化」を重視すべきではないかといったガイダンスが、最近EFSAから提唱されているところであります。

このほかにも幾つかの検討すべき事項がありますけれども、こういった背景があることから、この食品安全委員会でベンチマークドーズ法を活用するに当たって、統計学的な側面を中心として、科学的な裏づけを持ってベンチマークドーズを選択できる環境を整備することが重要ではないかと考えます。

そのため、座長としては、まず、用量-反応解析において用いられるNOAEL/LOAEL法とBMD法について、その特徴を統計学を中心とした科学の視点で整理し、共有するというところから始めてはどうかと思っておりますし、また、BMD法については食品安全委員会において、その都度最適な方法を考えるのではなくて、一貫性を持って適用できるかどうかという観点から検討してはどうかということを考えています。

そういうことを考えますと、このワーキンググループで今、検討できることと、さらなる情報収集が必要なことがあるかもしれないと考えますので、このあたりの整理も必要かと考えています。

今後の審議の中で、取りまとめられる内容は取りまとめていくということも必要で、先生方の意見を伺いながら、今後の進め方自体を決めていきたいと思っておりますので、よろしくお願いたします。

本日は、ベンチマークドーズ法に関する最初の審議となりますので、まずはベンチマークドーズ法の現状と課題を先生方と共有したいと思います。

最初は私からです。通常活用しているNOAEL/LOAEL法と、今後の審議対象でありますベンチマークドーズ法の特徴をお示しし、統計学的な観点から現状と課題について整理したいと思います。

続いて、広瀬専門委員からベンチマークドーズ法の最近の動向について発表させていただきます。

座長と座長代理が演台に立つというか、ここ（自席）でしゃべるのですけれども、そういう状況なので、進行が乱れないよう十分に留意して進めていきたいと思っております。

それでは、最初に私から、資料1にあります「用量-反応関係の分析の課題」についてお示ししたいと思います。

進行を同じ人がするのは差しさわりがありますので、座長代理の広瀬先生に進行をお願いいたします。

○広瀬座長代理 それでは、最初にベンチマークドーズ法の統計学的な課題と整理ということで「用量－反応関係分析の課題」というタイトルで、川村先生から御発表いただきたいと思います。

よろしく願いいたします。

○川村座長 それでは、最初の前提といたしますか、課題をお話しさせていただきます。

スライドを準備しましたので、それに沿って進めます。

(PP)

まず、これからの話は座長としてではなくて一専門委員として話をしますので、一つの考え方と、その考え方をするための根拠を示しますけれども、いろいろな御意見や御批判もお待ちしております。

まず、基本的に方法、我々が指標値を決めるときの手法としては、動物実験もしくはヒトにおける疫学研究をもとにします。

多段階のばく露をさせる。実験的にはばく露量を設定するということになりまして、それから、ヒトを対象とした疫学研究では、多くは自然のばく露、人災もありますけれども、それも含めてばく露量というのはコントロールされていませんので、連続的に分布するものをカテゴリー分けして評価するということになります。

エンドポイントとも言いますがけれども、アウトカムについてはイベント、つまり、ある事象が起きるか起きないかという二値、ダイコトマスなもの、検査値のように数量で示されるもの、それも連続量であるものと離散量であるもの等がありますけれども、数値で示されるものがあります。

このイベントか検査値かというのは、用量－反応関係とか用量－影響関係という言葉で区別されていたり、確率的影響、確定的影響という言葉が使われたりしますけれども、我々としては二値のイベントか、連続的な数量かという観点で仕分けすることになると思います。

いずれにしても、総称して用量－反応関係をグラフ化して、生体に重要な悪影響が出始める点あるいは、出ない上限値をもって指標値とするという共通の考え方があります。

そのやり方に具体的に統計学的な検定を使う方法。ここではNOAEL法と言いますがけれども、そういった検定に基づくものと、数理モデルを使って反応の大きさから評価するベンチマークドーズ法、BMD法というものが今、検討されている主要なものであります。

(PP)

まずは大半の評価で使われているNOAEL/LOAEL法について説明します。

このNOAEL法/LOAEL法の一番かなめになるところは、統計学的有意で指標値を判断することになります。右上の図を見ていただくと、これは毒性評価の図ではありませんけれども、血圧と突然死との関連を示したわかりやすいデータということでこの例を使っ

ていますが、横軸に、ばく露の量の程度をあらわして、縦軸にイベントの相対的な発生確率を示しています。

これも一種の用量-反応関係なのですが、影響の大きさというものは絶対値あるいはゼロレベルとの差であらわしたり、オッズ比やハザード比のように比であらわしたりするというものです。

これ(縦軸)はアウトカムの基準カテゴリーに対する大きさです。表現がそれですので、図の書き方はそれによって変わってきます。

問題は、統計的有意で判断するのですが、有意をあらわす指標であるP値、あるいは信頼区間、confidence interval (CI) と呼ばれるものは、実は得られた結果の確かさを示す指標であるというのが一つの肝になります。

影響の大きさは、その差か比であらわすし、その得られた結果の確かさをP値あるいは信頼区間であらわします。

確かさというのはどういうことかということ、どの程度得られた結果に基づいて一般論を強く主張できるかということになります。

後で説明しますが、我々が調べるのは標本であって、我々が議論したいのは標本の話ではなくて母集団、真実の世界の話をしたいわけです。

標本から母集団を推測するということなので、標本を見て真実のことをどれだけ強く、確かに言えるかという意味の確かさということです。

後から図を含めて、説明します。

この問題点は、ばく露の区切りレベルを変えると有意になる水準も動くということで、これは動物実験でも疫学研究でもそうですけれども、カテゴリーのとり方、ばく露量をどこで区切るか。通常、動物実験ですと3倍とか10倍とかというオーダーで階段状に設定しますし、疫学研究ですと連続的なものを区切るわけですが、そのカテゴリーの幅の設定によって、あるいは結果が出てから統合したりすると、有意になる水準は簡単に動いてしまうということがあります。

私の経験では、ヒ素の評価をしているとき、あるいは放射線の評価のときにもそういう疑問があったりして、これは恣意的にぼかしているという気配を感じ取って採択しなかった経験があるので、そういうふうに恣意的に動く部分があるということが問題になります。

それから、より本質的だと思いますけれども、統計学的に有意であることと、臨床的あるいは生物学的に有意であることは異なるということです。

統計学的有意は、あくまでも統計学、数学上の話ですので、対象動物数、対象者数に依存しまして、数が多くなれば有意が非常にしやすいわけですが、本来はどこから問題とすべきかというのは、臨床的に意味があるレベル。その臨床的に意味がある最小の差というのはminimal clinical important difference (MCID) と言うのですが、これの設定が必要である。

設定が必要であるといっても、実際に設定するのはなかなか容易ではないのですけれど

も、こういうもので生物学的に、あるいは臨床医学的に意味がある基準をつくるのが本来の姿であると思います。

(PP)

NOAEL法の特長としては、数学的なものに依存しますので、例えばデータが少ないとか、ばらつきが大きい研究だと信頼区間は広がってしまって、有意差が出にくいということになります。

例示として、スライドの図にありますように、もし信頼区間が大きくなったとすると有意水準が動くという一つのシミュレーションを示しています。これだと140~160というカテゴリで2倍のオッズ比があるのですが、エラーバーで示した信頼区間は、1から外れているので有意になります。エラーバーが大きくなって信頼区間が広がると、これは1をまたいでしまうと有意でなくなって、基準値としては1つ上のカテゴリになってしまうという現象を示しています。

そういうことで、有意差が出にくくなると指標値が上がってしまうということになります。これはレギュラトリーサイエンスとしては問題がある面もあります。

しかしながら、NOAEL法は動物実験ではほとんど問題にならない交絡 (confounding) の調整が可能というところがあります。

交絡については後ほど、もう一回説明します。

(PP)

そもそもP値というのは何だろうということなのですが、先ほど言いましたように我々は標本を見るのだけれども、標本を見て母集団を推測するわけです。それが推測統計学でありまして、だけれども母集団は見えない。見えないのだけれども、得られた標本から推測するということです。

古典的なNeyman-Pearsonの統計学の考え方は、実は標本から母集団に向かっているのではなくて、母集団である状況を仮定したときに、標本にどのように投影されるかというのが古典的な統計学の考え方です。ですので、母集団である値を仮定して、それが標本にはどうやって反映されるものなのか、どの範囲に反映されるかを確率で示したものということになります。

したがって、例えば仮定した真の値、通常は帰無仮説といって差がないという仮説を立てるわけですが、差がゼロになる仮説を立てて、標本のほうで、例えば1という観察値を観察した場合、これが母集団のゼロという仮定から標本にはどの範囲に投影されるかということを考えて、1という観察値が、もうこれが普通の分布図の範囲から外れている。図でいえば、観察値①となれば、これは仮定した値からは通常出てこないはずだと。だから最初の差が無いという仮説は間違っていて、差があるという対立仮説を採用するということになりますし、観察値②を得たとすると、差が1であるとしても、これは最初に仮定した差がゼロという仮説から、幾らでも存在し得る確率であるということであれば、最初の仮定、差がゼロであるという仮説は捨てることができない。それはそういうこともあるねと

いうことになって、最初の仮定を容認せざるを得ない。そういう考え方がP値の考え方です。

(PP)

これは最近出たアメリカ統計学会のStatementで、これは統計学者が強く力説しているところなのですが、P値というものはいかに仮定したモデルから外れているかということを検証するものであって、決して仮定した仮説が真実である確率であるとか、偶然そういう結果が得られる確率を示したものではありませんということ踏まえて、赤色の部分ですね、単に特定のP値だけをもって判断することは適切ではないということだし、ましてや、P値の大きさは効果の大きさ、あるいは、結果の重要性を示すものではありませんという、改めてこういうStatement、注意信号がアメリカの統計学会から出されたということで、統計学を利用する人はこれを改めて肝に銘じる必要があるということになります。

(PP)

それにかわる新しい指標としてBMD法が提唱されているわけですが、これは最近の話ではなくて、1984年に論文が出ています。論文を提唱した人はCrumpさんという人で、この写真の方だと思います。

この方が1984年に提唱したのですが、そのときの論文を読むと、既に現在のBMDの概念がほとんど完成していますし、BMDの持つ課題もかなりそこで浮かび上がってきています。

そういう課題もあってでしょうけれども、いい方法というか、一つの注目すべき方法が提唱されていながらすぐには普及しているわけではないですし、現在もまだ一部で使用されているにとどまるわけです。

やはり解決しなければいけない問題があるのですけれども、それは何かということこれから述べていきます。

(PP)

最初にBMD法の特長ですけれども、まず、用量-反応関係を調べるということは一緒ですが、それを特定の категорияにおいて有意かどうかを見るのではなくて、全体を一つの関数と見るわけです。

関数を見て、これから問題視するベンチマークレスポンス (BMR)、反応量、反応の程度、その反応の程度に達するベンチマークドーズ、そして、その持つ不確実さを考慮して、ベンチマークドーズの信頼下限をとるというプロセスを経て、基準値といいますか、指標値を決めてまいります。

全体を一つの関数化することから、個々の categoria の反応にとらわれない。NOAEL/LOAEL法ですと各 categoria を一つずつ切り離して考えるので、特定の categoria でぶれが大きいとどうしても結果がぶれるわけですが、これは研究全体を一つの関数にするので、一つ一つの categoria で多少乱れたとしても、全体に影響を与えるといえども与えまじくても、影響は比較的少ない。

その実験に用いた用量でない結果が得られるということも十分にあり得るわけです。所

定の反応水準に達するレベルということで、実際の実験に使わなかった、あるいはカテゴリーに区切らなかった数値が得られるということもあります。

大事な点は、影響の大きさを判断する。先ほどのNOAEL法の「確かさ」ではなくて、影響の「大きさ」で判断するというのが一番の肝になります。

さらに、信頼区間が広い研究、つまり信頼度の低い、実験動物数が少ないとか、対象者数が少ないという研究であったり、ぶれが大きい研究になりますと信頼区間が広がる。図を見ていただくとわかるように、信頼区間が点で示したところですけども、そこで赤色で両端に矢印で示しておりますように、両側に広がると、この下限値が下のほうに行く。つまり、下限値をとるので、指標値が低いほうにずれる。先ほどのNOAEL法だと、不確かさが増すと右のほうにずれるというのは対称的で、これは低いほうにずれる。安全サイドに立つという基本的なポリシーにはかなうことになります。

(PP)

問題は、幾つかの課題が未解決というところを説明していきます。

ベンチマークレスポンス (BMR)、つまり所定の反応量というのをどこに設定するのが合理的かという問題です。

今は10%、5%、1%、あるいは量的な指標であれば1SDというようなものが言われているわけですけども、そもそも、それは何に対する割合かということです。

基本的にはばく露ゼロのバックグラウンドのレベルと言われてはいますが、ばく露量ゼロで、もし反応がゼロだったら、10%とか5%は意味を成さないわけです。

ですから、それに対応するものとしては、例えば最小ばく露と最大ばく露との変動幅という考え方もあるのですが、実際にこれは採用されているわけではありません。何パーセントがいい、合理的かというはまだ決まっていないということです。

先ほど言いましたように、臨床的に意味のある最小の値 (MCID) と一致するかという議論が本来は必要なのですが、こういうものについてきちんと議論されたという痕跡は余り見当たらないということです。ですので、未解決な点がある。

また、質的転帰、つまり、1、0であらわされるイベントの有無というのと、検査値のように量的にあらわされるものが一緒の割合でいいのかどうか。1SDというのは量的な転帰に限りますけれども、そういったアウトカムの種類によって異なるのではないか。あるいは、ヒトの疫学研究と少数の動物実験で違うのではないかとということが考えられるのですけれども、これは大きな課題として残っております。

(PP)

2番目が数理モデルの選択ということです。

一つの関数にするのであるのですが、その関数がいろいろ提唱されて、One-hitとかMultistage、Weibull、Probit/Logit、さらにそういったものを対数変換したものとか、いろいろなモデルが提唱されております。

この算定というか、採択に先立って、論理に基づいて本来決める、つまり、アプリアリ

に決めるべきものだろうと思うのですけれども、今のところは食品安全委員会の経験でも、いろいろやってみてというところで、必ずしも論理が先行しているわけではない。頭にはあるのですけれどもやってみて決める。経験の乏しさもありますし、論理がきちんと構築されているか。標的細胞、遺伝子への作用のほかに、生身の体ですので、吸収や分布、代謝、排泄などの問題も考慮しないといけないとなりますと、論理自体も大変複雑になる可能性があるということから、今のところ、まだこれ一本というふうには決め切れないうところがあります。

発がんにおいてはある程度狭めることはできるかもしれないですけれども、それにしてもさまざまな因子を総合的に考えるとどうかということ、まだ解決し切れていないと思います。

それから、適合度の問題も取り扱いの中で出てきます。適合度の問題はgoodness-of-fit testです。フィッティングの問題ですけれども、これも実際の実験の結果、あるいは疫学研究の結果で適合度を評価していますけれども、これで問題なのは、母集団で仮定できるかということです。

標本でこのモデルが成立するかどうかという話ではなくて、先ほどから言っているように、我々は標本しか見られないのですけれども、母集団でどうなのかということが本当は大事であるということは先ほど言いました。ですので、母集団で仮定できるかどうかです。

標本で多少形が歪んでいたとしても、母集団で合っていればいいわけで、だけれども、母集団が見えないのでどうするかというと、標本からは明らかな逸脱だけを排除する。goodness-of-fit testの検定は御存じかもしれませんが、逸脱を評価している。P値が0.05より小さいと、逸脱しているから外すのですけれども、P値が0.05より大きければ、あるいは0.1より大きければ採択してしまう。つまり、適合度と言いながら非適合のものを排除しているだけで、実際に当てはまりのよさを評価しているわけではないわけです。

それはどうしてかということ、母集団のことはわからないから、母集団でどうなのかということを推定しているにすぎない。逸脱は除外するけれども、先ほどの検定と一緒になのです。仮定したものを否定するかどうかということだけで、我々は当てはまりのよさを評価することはできない。当てはまりの極端な逸脱だけを評価することができるので、そうすると、明らかな逸脱だけを排除するということで、その後は許容する。母集団で否定できなければ許容するという考え方になります。その辺がまだ共通の認識にないかもしれない。

あとは縦軸、横軸のスケールによってモデルも変わってくるのですが、生の値が本質的なのか、対数値が本質的なのかというのは、むしろバイオロジカルな判断になります。

例えば、トランスアミナーゼであれば対数値をとったものが自然の分布を示すであろうと思いますし、身長は生の値で自然な分布だと思います。

(PP)

課題の3番目なのですが、最尤値と信頼区間の取扱いです。これは異論というか反論を呼ぶ恐れがあるのですけれども、あえて強目の意見を言いますが、あくまでも解析の結果

は、結果としては最尤値一本です。最尤値の確かさを評価するのがP値／信頼区間です。

最尤値と信頼区間というのは、データを示すときの平均値と標準偏差と同じような書き方をしますけれども、これは全く異質です。

生データの要約については、平均値と標準偏差にしる、中央値とパーセンタイル値にしる、これはどちらも同等の意味があつて、代表値とばらつきという指標です。これは両者並列であつて、もとのデータの特質を示すものなのですが、計算によって得られたオッズ比とかハザード比とか、あるいは差分とかというものは、最尤値があつて、最尤値の確かさをP値／信頼区間で示す。ですので、表記の仕方も厳密には一緒ではない。書き方は異なりますし、持つ意味が全然違う。これはデータの要約についてはどちらも同じ程度の重みがあるのに対して、この計算で得られた最尤値と信頼区間については圧倒的に最尤値に重みがあつて、信頼区間は補助値であるということが本質的な違いです。

したがって、その信頼区間を指標値に用いているBMDLという信頼区間の下限値を最初から使うということは、非常に問題がある。それは統計学者が言っていたように、統計学会のStatementにあるように、P値と信頼区間は本質的に一緒ですので、P値をプライマリーに置くということは問題があると思います。

ただ、そうは言っても、不確かさは残る。これをどう取り扱うかというのがやはり問題で、どうしても少数の標本である。真実の母集団の状態がわかればいいのだけれども、わからないから標本でやっているのだけれども、標本というのは母集団からランダムにとってきたものだから、当然ぶれるでしょうと。そのぶれの部分を規制値、レギュラトリーサイエンスとしてはどう反映するのかということです。これは不確か係数に似ていて、全然レベルが違うといえは違うのですけれども、動物実験の結果を人間に援用する場合に、種の差ですとか、急性と慢性の差ですとか、そういったことで不確か係数を使うという現実があるわけですが、それと似ていて、ランダムにとってきた標本であるけれども、標本はあくまでも抽出したものであつて、それは偶然のぶれというものがある。その偶然のぶれをどう反映するかというところで、そのぶれの程度を信頼区間なりP値で示しているわけです。それを取り込まないと、実際のレギュレーションにおいては、そこまで最尤値がリジッドなものではないとすると、それを入れ込まないと危ないのではないかと、ぶれが大きければ社会的に問題が起きる可能性があるのではないかという意見に対して応えようとすると、そのぶれを考慮に入れないといけないということになるのです。

ただ、入れるにしても、まずは最尤値ありきで、その確かさが信頼区間なのだという視点は持つ必要がある。これをデータの要約と同じような考えで、初めに下限値ありきと考えるのは問題があると考えています。

(PP)

そもそも信頼区間とは何でしょうということですが、先ほどのP値の図も一緒ですけれども、こういうのは普通の統計の教科書には載っておりませんで、自分の頭で考えたことなのです。信頼区間はなかなか説明しにくいのですが、もともと母集団があつて、

母集団の仮定した値が標本にどう投影されるかという基本的な考え方は一緒です。

信頼区間というのは、まず仮定値。一番低いところに仮定を置いて、その95%の分布範囲の一番上限の値が観察値に来るように仮定値を置いてみるということです。

反対に、最も大きな見積もりをしてみて、その分布範囲の最小のところが、95%ですからもちろん分布は無限に続くのですけれども、95%なら95%の点がたまたま観察値に来るような仮定を置いてみる。その間にあれば、必ずこの観察値は95%の山の中に入りますという指標である。その最小の仮定値と最大の仮定値の間が信頼区間であるという説明をしています。

信頼区間とは何だろうということがなかなか説明しづらいし、真実の値があるかもしれない値であるというマインドはそうなのですけれども、真実の値が存在する確率という、いや、真実には確率は存在しない、分布は存在しないという前提に背くので、こういうような、真実はあくまでも仮定を置くだけで、標本に対して確率の分布が存在するということになります。

(PP)

それを整理しますと、母集団に確率はありません。標本に確率が存在します。信頼区間というのは、真の値が95%の確率で存在する範囲ではありません。真実を仮定する範囲です。何も95%である必要はなくて、90%ならこの幅になります、99%ならこの幅になりますという指標であって、どの値を使っても別に構わないものなのです。

最尤値の確かさの指標であった、あくまでも添えの数字ですということから、くどいですが、信頼区間の両端値を単独で用いるべきではないということになります。

(PP)

次に4番目になりますが、交絡の話です。交絡とは何か。英語ではconfoundingと言いますけれども、これは動物実験では基本的に発生しません。つまり、ピュアな状態。各群に分けたとしても、それぞれ割と純粋な動物実験を使うので、ほかの要素が絡むということは基本的にない。偶然の要素は多少あるかもしれないけれども、ほかの要素が系統的に絡むということはないわけですが、生身の人間で研究をすると、ある注目している因子と、そのアウトカムの関連を調べるに当たって、第三の因子が邪魔をすることがある。例えば、お酒と肺がんの関係を考えようとするすると、ここに介在するのはたばこという因子があります。お酒を飲む人はたばこもよく吸う傾向があるので、お酒と肺がんとの関連を研究しようとするすると、ここにたばこが絡んで、お酒をよく飲む人はたばこも吸いがちであって、そのたばこの影響が肺がんに出してしまうと、見かけ上お酒と肺がんが関連しているように見えてしまうという関係がある。これは誰でもたばこ肺がんのことを知っているの、それから、たばこお酒の関連も薄々は気づいているので、想像はできるわけですが、知らない間に第3の因子が介在するということはよくあります。特に生身の人間の実際の生活では、介在することがむしろ普通であって、したがって、こういうのを調整するという試みがあります。

これがNOAEL/LOAEL法ですと、多変量解析を使ったりして、あるいは、最近ですと propensity score、あるいはinstrumental variable、そういった新しい指標を使って、できるだけ交絡を取り除こうということをやっているのですけれども、もちろんそれは限界があるのです。一応NOAEL法では多変量解析を中心として調整は可能ということになっています。

一応つけたのは、限界があるからです。調整できるのは既知の因子で、かつ、既知で測定されている因子のみですので、未知の因子や未測定のもの調整できません。

しかし、ベンチマークドーズは、そもそもそれ用にできていない。ソフトウェアがあるのだけれども、交絡の調整はできなくはないのだろうけれども、今のところ普遍的に使う状況にはないということです。

ですから、ここはやり方を考えないといけない。動物実験のときには余り問題はないのだけれども、ヒトの疫学研究をやるときには、ひょっとすると各測定値を補正しないといけないかもしれないのです。このやり方はまだ確立していないので、こうやってやるのかという私の思いを書いただけですけれども、いずれにしても対策は練らないといけないということになります。

また、交互作用というのがある。例えば、性別というのもよく交互作用を起こす因子になってきて、男性における用量-反応関係と、女性における用量-反応関係がもし違うとすると、あるいは、年齢によって、子供の用量-反応関係と大人の用量-反応関係が違うとすると、性別や年齢と原因物質との間に交互作用があることになる。その取り扱いをどうするかという問題もあります。

こういった幾つかの、まだ十分解明されていない問題があるので、そこを考えていかないといけないということになります。

(PP)

最後の2枚は付録でありまして、問題点を直接指摘するものではありません。毒性に関する研究としては、ヒトを対象とした研究のものと、動物実験があります。ヒトを対象とした研究の中には、いわゆる分析疫学研究、通常いろいろな公害ですとか、原爆の被爆者コホートとかで代表されるように、多数の人たちの集団をつかまえて、長期にフォローするというコホート研究とか、病気を起こした人と起こさない人のかつてのばく露状況はどうであったかという症例対照研究のような研究スタイルであるヒトの観察研究、疫学研究が使われますけれども、ヒトでも動物実験と同じような介入研究は行われることがあります。

特に薬剤の評価は圧倒的にこれでやっておりまして、介入研究、臨床試験というもので、しかも薬事法上の承認に係るようなものを治験といいますけれども、こういったものを使い得るのですけれども、毒性の評価にはなかなか人間での介入試験というのは使われていません。

動物実験とかヒトの介入試験のいいところは、ばく露量をコントロールすることができ

ますということで、等間隔で設定できます。

交絡は基本的にないということなのですが、ただ、残念なことに、対象者あるいは対象動物が少数であってというところが問題です。

動物実験は一応ルールがあるのですけれども、人間については、実は数字は出ていなくて、数字を算定しなさい、サンプルサイズを算定しなさいということが求められています。

ヒトによってわざわざ介入試験、実験を行うのは、通常は効果を証明するためですので、その有効性を証明するために必要な最小限の数を、何せヒトで実験するのですから倫理的な問題、要請として、最小の数を算定しなさいという、サンプルサイズの算定が強く求められます。

したがって、通常は有効性を証明するために研究を行うのでありまして、この食品安全委員会で必要な有害事象の検証のために介入研究というのは事実上行えないのです。ただ、運よくというか悪くというか、有効性を証明するつもりの研究が有害性の証明になった研究というのは少数あります。

代表的なのはβカロチンと肺がんの関係です。ヘビースモーカーに対してβカロチンを投与したら、本当は肺がんが減るということを証明したかっただけすけれども、結果は反対で、βカロチンを投与したら肺がんがふえたというアンラッキーな結果になったような研究は、有害性の証明になってしまったものでもあるわけです。

ただし、ヒトではなかなか多段階のばく露設定は少なく、薬剤などが少量、多量ぐらいの2段階の設定をすることはあるのですけれども、3倍、10倍で何段階かやるということは非常にまれです。

そうすると、やはり観察研究です。自然の成り行きでばく露したのを見るというのがほとんどになりますが、そうすると、ばく露量というのは連続的に分布するので、しかも濃密、粗密があるわけです。ばく露量の固まって多いところと、非常に少数しかいないところがあるということなので、カテゴリー区分というのが非常に問題になります。

交絡という第3の因子の介在は避けられない。多変量調整はするのだけれども限界があるということになります。

対象者は一般的に多数なので、動物実験などと違ってまれな現象も起こり得るわけで、しかも有害事象も評価できることになります。

(PP)

あとは、それらを総覧して、動物実験と疫学研究の問題点を整理して終わります。

動物実験の中で問題なのは、まれな事象は評価できないということになります。最大ばく露カテゴリー、一番グラフの右端に来るところで、10分の1の確率で起きる事象です。50匹であれば5匹発生する事象は、1群50匹では検証できないわけです。発症確率0.1の事象の検定に必要なサンプルサイズは、サンプルサイズの必要な方程式があるのですが、計算すると1群140匹必要ということになります。10分の1の確率ですらこれだけ要るとなると、100分の1の確率で起きる事象は、これよりまた1桁多くて、1,400匹要る計算になり

ます。

そうなる、もう現実的にはやれないので、これについて対策を考える必要がある。これはBMD法から逸脱するのですけれども、メタアナリシスとかプールド・アナリシスとか、ヒトの疫学研究では今、世界はそれに向かって動いています。ビッグデータの時代なので、データの統合が行われているのですが、動物実験もそういう考え方がきつとあるのではないかと思うのですけれども、そういうことも検討していかないといけないかもしれません。

それから、ヒトの疫学研究については、ばく露カテゴリーをどうするかという問題があって、等間隔で分類する。それは等差数列であったり等比数列であったりしますが、等間隔的なものの考え方と、もう一つ、分位とって、3分位とか4分位とか、人数が等しくなるようにする、2つの方法があります。

等間隔だとわかりやすいのですけれども、各カテゴリーの人数が異なってきて、検出力がばらばらになる。それがNOAEL法の問題でもあったわけですが、検出力がばらつく。したがって、下位のカテゴリーより上位のカテゴリーのほうが反応は大きいのに有意にならないということがあったわけです。それで統合が行われたりして、作為的になってしまうということもあります。

分位だと人数が等しいので、検出力はそろうのですけれども、切れ目、区切りの値が研究ごとに異なってきて、これはレギュレーションとしてはどこに基準値を置くかというのはなかなか難しい問題になってしまう。我々はベンチマークドーズをここから議論していきますけれども、動物実験で最初にやるとしても、ヒトの疫学研究もベンチマークドーズの対象になるので、この両方の特性を知った上で議論を進めていかないといけないと思っていますので、問題点を整理してみたところです。

(PP)

そういうことで、まとめですが、NOAEL法というのは反応の確かさで指標値を決めています。ベンチマークドーズは反応の大きさに判定するということで、こちらが今、注目をされているのですけれども、ただ、残念なことに、まだ解決されていない問題、実際の取扱い上、取り決めをしないといけない問題があるので、この評価技術のワーキンググループで合理的な措置と手順を検討していくべきであろうと考えて、これをまとめとさせていただきます。

以上です。

○広瀬座長代理 川村先生、どうもありがとうございました。

それでは、今の先生の発表について、質問がありましたら受け付けたいと思いますので、よろしくをお願いします。

どなたかおりませんか。

ないでしょうか。

佐藤委員長、どうぞ。

○佐藤委員長 BMD法において、BMDLを一つの指標として使うのは問題だというのは、確か

にそうだと思うのです。

たしか『Proceedings of the National Academy of Sciences』でメチル水銀の評価をやったときには、BMDとBMDLの比をとって一番小さいところをやったというのは、多分そういうところに問題があるからだと理解しています。

Crump自身がニュージーランドのデータを分析したときには、どういう値をとっていたのでしたでしょうか。先生、覚えておられますか。私はちょっと記憶がないのだけれども。

○川村座長 彼の論文によればBMDLを採用するということが書いてあります。彼も今までの毒性評価の基盤の上に考えているので、BMDLの不確実性を盛り込まないといけないという考え方を持っているようで、BMDLを採用はしていますので、本チャンの統計学者もやはりそういう視点は持っているのだけれども、ただ、振り返って見ると、最初からBMDLに着目してしまうと問題が起きるといえるのは、例えば先生がおっしゃったように、BMDとBMDLにすごく解離がある場合というのは信頼区間が広くて不確実な研究になるので、積極的にBMDLの低いのを採用すると、信頼度の低い研究を積極的に採用することになってしまうから、まずはBMDを見て、最終的にとるのはBMDLの値であっても、まず、見る目はBMDで見えないといけないのではないかというのが、この前の実際、食品安全委員会で検討したときに私が申し上げた主張ではあります。

○佐藤委員長 逆に言えばあれですね。不確実係数に反映させてもいいのかもしれないですね。

○川村座長 おっしゃるとおりです。不確実さをレギュレーションの現場でどう用いるかというのは、一つのポリシーの問題だと思いますので、統計学そのものではなくて、それをアプリケーション上の課題と考えることはできるかと思います。

○広瀬座長代理 ほかによろしいでしょうか。

なければ、またその後でも時間があるかもしれませんので、次の私の発表に移りたいと思います。

川村先生、どうもありがとうございました。

○川村座長 それでは、続きまして「ベンチマークドース (BMD) 法の最近の動向について」を、広瀬専門委員から御発表いただきます。

お願いいたします。

○広瀬座長代理 私のほうからは、川村先生からは統計学の基本に立ち返ってベンチマークドースの意義といったこと、あるいは課題について問題点を指摘していただいたのですが、私のほうから、どちらかというとベンチマークドースを実際にそういう規制値といったらあれですけども、きりのいいエリアを設定するときに使ってきた経験、あるいは、その中で今、国際的かというと、いろいろな問題点がある。実使用における問題点について、統計学的な裏づけまで解説し切れないのですけれども、その問題点について紹介させていただきたいと思っています。

(PP)

これはどこにでも書いてありますけれども、ベンチマークドーズ法を利点として使うことをこちらに示しています。基本的な用量反応評価の中での使い方については、毒性学的な有意な発現率というのをある程度設定して、経験的なところで、動物実験の場合のを中心にして今日はお話ししますが、そういうことの検出に対して、それに対する数理モデル、動物実験のモデルを設定して、実際にはNOAELが求められない実験は結構ありますので、そういったときにTDIを設定するための出発点です。通常はNOAELに不確実係数を適用して、ADIとかTDIを設定するわけですが、そういう実際のNOAELというのは、動物実験をするときの計画で立ててしまうので、やってみた後では、それが実際には無毒性量を想定して実験をするわけですが、できないことがあります。

そういったデータを、また動物実験をやり直せばいい話ではありますが、それは経済的な問題、あるいは、最近では動物愛護の問題でそうそうできない。そういったときに得られたデータから、とりあえずはそういった無毒性量、出発点を求める代替手段として開発。それだけではないのですけれども、そういうことをメインに開発してきたということがあります。

そういったフィッティングをしたときに、先ほどから少し問題になってはいますが、実際にはレギュレーション的にはBMDLという値を出発点として採用する。そのBMDLというのはどういった値かという、そのフィッティングさせたときのモデルの95%信頼限界上限値と、毒性学的な発現率との交点、そのときの下限容量をBMDLとして定義して使っていくというのが、基本的な使い方となっています。

(PP)

それがどういった利点があるかというのを、これは簡単に図で示しただけですので、そんなに大きな意味はないのですけれども、この緑の濃い点が実際に得られた実験データで、本来ならもう一点、下の実験の容量をとればNOAELは求められたかもしれないのですけれども、報告にはこのデータはない。そういった状態で、実際に今までのベンチマークドーズを使わないと、このLOAELからNOAELをある程度保証するための不確実係数にさらに通常の種差、個体差の100を掛けて、実際にはLOAELの、簡単に言うと1,000分の1ぐらいのTDIを設定するというのが、ベンチマークドーズを使わないとこういうことになります。

ただ、それだと、この点の位置にもよりますが、かなり過剰に低い値を算定することになりますので、そういったときにベンチマークドーズ法を活用すると、このNOAELはわかりませんが、ベンチマークドーズのある値から、TDIの設定の出発点とする。これが経験的な解析をするとLOAELに近いということがわかっていますので、それに対して、通常はNOAELからLOAELに掛ける10を使わないで、そこから不確実係数100を掛けると、この辺の値。この緑の線はあくまで仮定なので、このような線があるかどうかかわからないのですけれども、そういったところに多分近い値が算出できるだろうといったことで、ベンチマークドーズのNOAELが求められない場合の利点はこういったところにあるのかと考えるために示した図です。

(PP)

ベンチマークドーズはそういったところで使えるという利点以外にも、実は統計学的なフィッティングをかけることができるということで、実は副次的な効果も幾つかあります。

例えば、同一の設定、投与量で、NOAELが同じ実験が2つあったとしても、反応率が違う場合には、感覚的には反応の強弱は違うと思うのですけれども、実験データからすると、NOAELは一つになってしまっただけで同じになります。

そういったときのベンチマークドーズで解析すると、用量-反応関係からBMDLが別に求められますので、そういったことで比較すると、NOAELが同じ実験の2つを、毒性の強弱を比較できる可能性があることにも使えますし、統計学的手法で信頼区間を持ったデータとして、BMDLをどう扱うかという問題はありますけれども、動物数が低くなると信頼区間は大きくなる。必然的にBMDLの値が低くなるということで、それをレギュレーションの値として採用するとすれば、信頼が低い場合は安全側の低い値を採用する。信頼性の高い場合は、よりBMDに近い値になるということで、同じNOAEL、あるいは同じBMDLであっても、統計学的な信頼性の値を数字に反映させることができるという特徴があります。

過去のデータからの特徴としては、NOAEL、LOAELに近い値になって、これは多分経験的というよりも、検出力の問題がありまして、何匹の動物を使った場合の検出力というのが、BMRが5%、10%というときに、NOAELと同等になったという結果であると思います。

もう一つの特徴は、実験用量のフィッティングなので、計算式によるBMDLの値に、モデルはいろいろありますけれども、そんなに違いはないという、これはちょっと唐突なのですけれども、これはもともと、この後に少し説明しますけれども、発がん性の低用量外挿にもいろいろなモデルが今まで使われてきた経緯があります。低用量外挿というのは、実験要件よりもはるかに低いところの、10のマイナス5乗、6乗のリスクを予測してきたわけですけれども、それはモデルによってかなりばらつきがあるということで、今日は説明しませんが、最近では直線外挿をするということがあります。そのときの出発点をBMDLに設定すると、モデルによる定量外挿の違いは、低用量外挿では何千倍、何百倍違う値が、BMDLからの直線外挿の場合は数倍あたりの違いに収まる。そういう意味で、いろいろなモデルを使うということにメリットがあるというのが、ここで特徴として挙がっています。

ただ、この後説明しますけれども、実はこのBMDLに違いが少ないはなく、実際は結構違っています。それは何千、何万と、桁までは違わないのですけれども、1桁近くの違いがあるということが実は、最終的にはADI、TDIに反映させたとき、あるいは基準値に反映させたときに、1桁違ってくることが起きてしまう。こういったところが実は適用の問題点として今、あるということがあります。

あとは、やはり用量-反応曲線は一つの値に対しての2次元のグラフの中で展開していきますので、例えば病理のデータは、発現率プラス重篤度が反映されたデータがあります。こういったときには、実はそのままのデータを使うことができないということがあります。

(PP)

例えばなのですけれども、動物で肝肥大の発現する発現率が用量相関で増えていく。ただ、(用量が)上がっていくと同時に、実は重篤度も一緒に上がっていくので、これだけを見ると簡単な2次元のグラフに展開することはできませんけれども、その発症する動物をとる。これはさまざまな、用量-反応曲線でこう使えば異論があるかもしれませんが、とにかくこういうことをすれば、制限がありますけれども、ベンチマークドーズを病理の所見にも使うことができる。ただ、そのままでは重篤を示したデータを使うことができないということがある。

欠点もありますけれども、そういった利点としてベンチマークドーズを計算して使っていきます。

この後、実際、ベンチマークドーズの中で値がぶれるということがどうして起きるかということ、因子の解説と一緒にしたいと思います。

(PP)

その前に、そもそもベンチマークドーズを計算するとは言っていますけれども、私もそれほど統計の専門家ではないのですけれども、実際にはCrumpさんが最初に開発してきたBMDSという、現在では米国の環境保護庁で維持されて、アップデートされてきていますけれども、そういったプログラムを使うことで、実は手軽に計算するようになっています。

最もよく使われているのは、このBMDSといわれるソフトウェアで、その中に発現率の数字を入れると、自動的にそれぞれのモデルに対してのBMDとBMDLが使える。インターフェースもWindowsで使えるようになってきたので、普及率も高く国際的によく使われているということになっています。

計算方法は違うのですけれども、似たようなソフトはオランダのほうでも開発されていますし、最初のほうはそれほど使い勝手がよくないというか、Windowsで簡単に素人が使えるようなソフトではなかったもので、そんなに普及はしていませんけれども、原理的には同じ開散結果が得られるソフトでして、幾つかのモデルに対してのベンチマークドーズを計算することはできる。

最近はこのソフトのモデルの、後で問題になりますけれども、どのモデルがいいかという選択性的問題点を少し解消するために、それぞれのベンチマークドーズで出たBMDLを平均化して、ある程度一つのモデルを計算するというソフトも開発されてきています。

アメリカのNIOSHの研究者が2007年ぐらいに開発して公開しているのですけれども、こういったソフトもありますし、それに対して、実はヨーロッパのほうも、このPROASTを基本原理として用いて、それを平均化する開発のソフトを今、開発中です。

こういうふうにならば、BMDSからPROAST、MADr、幾つかのソフトウェアで使うことができるという状態になっています。

(PP)

実際にそうしたら、ベンチマークドーズのソフトを使ったときには、先ほど1980年代からベンチマークドーズの概念は確立しているのですけれども、実際に普及し始めたという

のもあれなのですけれども、国際機関で本格的に使い始めたのは2010年の第72回のJECFAなのですけれども、この回で汚染物質をいろいろ扱いました。

このときに6物質の汚染物質全てベンチマークドーズを適用しまして、ADIあるいはMargin of Exposureの計算を、定量的なリスク評価に本格的に使い始めました。

ただ、そのときからもう既に少し問題になったのですけれども、こういうモデルの中でベンチマークドーズ、例えばBMDSの計算結果が出ていますけれども、この中で各モデルについてBMD、BMDLというのは全て出てきます。この中でどれが最も適切であるかというところが、実は議論されることになりまして、それぞれは大体AICといわれる適合度のある統計学的手法の小さいものを選定する。そういった選択方法でやってきています。

ただ、統計のわからない我々が、もしこのデータを使ってしまうと、一体どの数字を使っていいかというのが実際には、専門家のそういった知識が要るところで、実際に計算は簡単なのですけれども、実際の使用には少し壁があるのではというのを感想としては持っています。

(PP)

そうは言っても、そういった使う流れがある中で、食品安全委員会としてもそういったことで、実は、同じぐらいの2005年から2010年にかけて、その時々で必要性はありますけれども、NOAELを求めにくい、あるいはLOAELしかない場合。そういった場合について使われてきたわけです。

ただ、これらの例はそれぞれの委員会で、それぞれの専門家の判断、あるいは国際的状況をあわせて考慮して、実態はケース・バイ・ケースでそれぞれやってきたのが現状かと思えます。

(PP)

そういった、どれを選ぶかということについては、やはりそれぞれの機関は、それぞれのガイダンスなりポリシーを持ってやってきています。

これは少し古いのですけれども、2012年ぐらい、現在になるのですけれどもEPAのガイダンス。EFSAは2009年に、後で2回目の説明をしますけれども、1回目の使用のガイダンスを発表しますし、似たような時期にWHOのほうもそういったベンチマークドーズ。これはベンチマークドーズとは言っていないけれども、用量-反応評価にどういったやり方がいいかということで、その中でベンチマークドーズに言及するところがあります。

その中で、それぞれの、先ほど言った8つぐらいのモデルからどれがいいか。どういうふうを選択するかという基準を、ここに一覧表として示しています。

詳しいことをここでやると時間がないので説明しませんけれども、例えば、そのモデル適合度の、川村先生からも説明があったP値で、適合しないのを排除するときのP値の閾値をどこにするかというのは、実は評価機関によって違います。0.05と0.1とか、あるいはモデルを選択するときの基準も、安全側に立つという基準に立つ場合、BMDLの一番低く検査されるモデルが、安全側という意味でいいのかという考え方もあれば、AICといった評価

基準で選定するということがあります。

(PP)

ここで簡単に、川村先生も既に解説されたわけですがけれども、P値、AICはどういった概念か。P値は先ほど川村先生からもお話がありましたけれども、0.1の10%の確率で、もともと乖離しているとは言えないというものは、一応適合したと。逆説的ですがけれども、そういう値として設定される値ということで、そこを0.1で切るか、0.05で切るかということが、このP値の値になっています。

もう一つはAIC。先ほど最尤値を計算すると言っていますけれども、その最尤値はサンプリングのモデルの点にどのくらい近いかというのを示す指標として、統計学的なモデルの適合度であります。その中で今、最もよく使われているのは赤池情報量基準、AICという基準でして、これは実験測定値に最も近ければ低い値になっていくわけですがけれども、実はパラメータをふやすと、後で説明しますが、適合するときのパラメータの数をふやすと、どのような点でも完全にフィットさせることができるのです。ですから、パラメータの数がふえていくと、今度は適合度は高くないというふうに、相反する数字を足すことによって、適合のバランスを評価するというふうにした計算値になっています。

これの値の小さいほうが、いいモデルだろうということで、今は評価するといった考え方になっています。

(PP)

それを、例えばこれはBMDS2.6.0.1に収載されている8つぐらいのモデルですがけれども、それぞれ経験識あるいは生物学的な反応をある程度加味した幾つかのモデルということで、これのどれにフィットするかということでフィッティングが行われています。

ここでこの後、さらに評価機関でずれる大きな原因となるところが、このRestrictionという、これはソフトを使っていくときに設定するための言葉なので、統計学的用語ではありませんけれども、それぞれのモデリングについて、パラメータにある制限をかけるといった操作が選択できるようになります。これが素人からするとすごく迷う部分で、ソフトの中ではオンとかオフとか制限できるのですが、それはどちらが、使ったほうがいいのか、使わないのがいいのかという指示はないわけです。それが逆に評価機関によって、これをオンにする、オフにするというところの選択によって、実は値が大きく変わるようなことがあります。

特にこの原点の傾きに対して、傾きが最初から、原点からいきなり反応曲線が立ち上がるというモデルを防ぐという、フィッティング上そういった係数の制限をかけるということで、実はベンチマークドーズの値が大きく変わる。ちょっと見にくいかもしれませんが。

(PP)

これが一つの例でして、Log-Probitというものに対して幾つか、高用量は切れていますけれども、4点の用量-反応曲線に対して、ベンチマークドーズのフィッティングが出ます。

Restrictionをオンにするということは、濃度がいきなり原点で立ち上がります。シグモイドになるように制限をかけるといった意味で、フィッティングをかけます。

同じようなフィッティングは適合するのですけれども、Restrictionをオフにすると、もっと適合する値として、モデルとして、いきなり原点から用量-反応が立ち上がるような、これが生物学的にあり得るかあり得ないかという議論は決着がついていないのですけれども、それについて、オフにするとこういうふういきなり立ち上がります。そうすると、BMDLの10%発現の値は、オンにするところの値。オフにするともっと低いところに行ってしまうというふうに、この制限をかけるかかけないかでBMDLが大きく違うというのが問題提起としてあります。

(PP)

もう一つは、これは先ほど言いましたけれども、この4点に対して、実はパラメータの数をふやすと、この変な曲線でフィッティングをかけるということは数学的にできてしまいます。

これはやはり用量-反応曲線が、通常は単調増加ということを考えて、こういったことはあり得ないわけですし、こういった群のモデルというのは排除する。要するに、パラメータの数は少なく抑えるといった、単調増加にない曲線をつくることを防ぐような制限。これはちょっと別な意味がありますけれども、そういった制限をかけるかかけないかということで、モデルの適合度、BMDLの値が変わってきています。

(PP)

そういった問題点を受けまして、私のほうでは平成20~24年で、健康影響評価の技術研究の研究費をいただきまして、こういったモデルは既に2010年ぐらいから問題になっていましたので、どういったやり方がいいか。統計学的な妥当性もさることながら、実際のデータを使ってみて、どういったパラメータのセッティングをしたらいいのではないかと、いうことを提案させていただいた成果があります。

その中で、例えばベンチマークドーズで未知のデータを使う場合は、おおむね10%使うと、今、あるいろいろなデータの中の多くはこういったものでいいのではないかと。あるいは、Restrictionのオンとオフについては、今のところ決着がつかないので両方見てやるかとか、そういったことを提案させていただいたというのがあります。

(PP)

そういうことも受けて、アクリルアミドのときも私の研究班の考え方はベースではありますが、どういった方向でBMDLを設定したらいいかということをし少し検討しまして、アクリルアミドのときには実際にはこの選定方法を使いました。この場合は適合、生物学的な妥当性、あるいは国際的な指標の、この場合はデータ自体が良い実験データをモデル化したので、Restrictionをオンとオフで、実際にはBMDLの値はほとんど変わりませんでした。

なので、Restriction ONを優先して使いました。さらに適合度の検定については、この

辺は変わりません。

もう一つ、川村先生の問題提起の中の、多分指摘の中の一つであった、BMDをとるかBMDLをとるかということで、最尤値が最も低いということは、安全側になるといった判定、議論、検討していく中で、BMDの一番小さいモデルを選定するという一方で、ベンチマークドーズの選定の基準は、このときは設定しました。

(PP)

具体的には、マウスのハーダー腺の場合はこのデータを使ったわけですが、Restrictionがオンの上半分のモデルの中から、BMDの一番小さい値で、それに相当するBMDLを、とりあえずはMargin of ExposureのPODとして設定するというやり方を使いました。

(PP)

最近はどういったRestrictionのオン、オフの問題、あるいはBMD、BMDLの問題ということをもう少しレギュレーションにも使いやすいということもあると思うのですが、そういった概念を持ってきて、いろいろなモデルの中の平均でやっというのを、EFSAのほうでは新しいガイダンスとして、昨年公表しました。

(PP)

それはこのBMDのモデルの平均化という手法を用いるわけですが、8つぐらいのモデルの中から、実際にはそれぞれのモデルはその線でフィッティングするわけです。この場合は多分どれも、全部適合しています。この中でどれが選択できるか。どれも多分同じ95%信頼区間で、P値は0.1を満たすものについては、多分全部適合します。この中でどれが一番いいか。今までですと、例えば一番低いBMDL、あるいは、最もこれにフィットしたものを選ぶといった判断基準があったわけですが、これを全て平均化する。そのときに、単純に平均化するのではなくて、適合度にウェイトをかけて平均化するという手法を持っていて、そういうことで、モデル選定手法の違いによって異なる値が出るのを避けようといったことが提案されてきています。

(PP)

これについてJECFAのほうも、こういったモデルの平均化そのものは問題点としては持っていなかったのですが、第83回のJECFAのときで、生物学的に受け入れないモデル。これはRestrictionをオンでとるか、オフでとるかという議論があった中で、JECFAのほうではRestrictionがオンで、モデルを選択するという選択をとりました。

ただ、問題点は認識してはいて、オフもやってみて、一応は計算したのですが、比較目的というか参照データとして使っただけでした。しかし、実際にはONもOFFも使ったということになります。この問題点を解決するために新しいワーキンググループをつくらなければいけないという認識になっています。

(PP)

話があちこちに行っていますけれども、それが端的に出たのは、3-クロロプロパン-1,2-ジオール(3-MCPD)のエステル類の評価が、実はBMDLはJECFAとEFSAで1桁も違

う。もともになった実験はここに示す表の、同じデータを使って、BMDLが実に1桁違う値が計算されています。

(PP)

その違いは何かと端的に言うと、このRestrictionがオンのモデルを使ったか、オフのモデルを使うかの違いです。

EFSAのほうはオフのモデルで一番小さい。これ全部を比較して、最も低いBMDLを採用しましたがけれども、JECFAのほうは基本的にはRestrictionがオンのモデルであったので、この中で一番低いモデルを使用する。結果として、これが10倍の違いを出すという、国際的な有名な評価機関でもこういった事態が実は起きているといったことが現状であります。

それを解消するために、EFSAのほうではこれらを平均化する手法というのを、今は提唱というか、実際にはそれを運用し始めているという状態になっています。

(PP)

そういった国際的状況を受けて、食品安全委員会でもこの後、このワーキンググループで検討していくという中で、こういった点の課題があるのではないかということで、これは二重にやりますので説明しませんが、こういったことを今後検討していけばいいのではないかという提案をさせていただきたいと思います。

済みません。長くなりました。

○川村座長 どうもありがとうございました。

今、世界の趨勢というか、国際的な評価機関の現況と、我々はどうするかという課題を整理していただきました。

ここは確認しておきたいというところはございませんでしょうか。

お願いします。

○赤堀専門委員 簡単な質問なのですが、EFSAの新しいガイダンスの中で、いわゆる生物学的に受け入れられないモデルの取り扱いの観点は、何かあるのですか。なぜオフを採用しているのかというのが少しあったので。

○広瀬座長代理 生物学的なあり得なさについては、実は用量-反応曲線はいきなり立ち上がっているとはいえ、多分そこはもっと低い濃度まで持って行ってグラフを描けば、またシグモイドが得られるだろう。だから、要するに、横軸の取り方で見かけ上は上に上がっているというだけで、低い用量にフォーカスすればそういうシグモイドも見られる場面もあるはずなので排除できない。いきなり上に上がっていくグラフも、EFSAの統計学をやっている専門家、EFSAのガイダンスをつくったエキスパートは、それは排除しないといったポリシーになっています。それは、そうではないと言っている人とはもう何十年も議論してきて、これは会議でも解決できないだろうということになっていました。なので、もうそれはポリシーの問題です。

EFSAはオフのモデルを平均化するといったモデルで、今は公開しています。

○赤堀専門委員 オフで平均化。

わかりました。ありがとうございます。

○佐藤委員長 2つあるのですけれども、1つは簡単な話で、AICなのですけれども、Crumpたちは確かに最初、log likelihoodというのを使っていたと思うのですけれども、AICというのはそれに、要するにパラメータの数による補正をかけたものだという理解でいいですか。

○広瀬座長代理 多分そうだと思います。パラメータが多いとこれから小さくなっていくので、それを補正した値だと思います。

○佐藤委員長 もう一点は、今のモデルの話なのですけれども、先生がここに書かれている、各数理モデルの統計学的背景の理解というのは必要なのでしょうか、多分WeibullにしてもProbitにしても、生物現象を見てこれがフィットするだろうとつくられたモデルだと思うのです。この中のモデルが全部ではないと思うのだけれども、そういう生物学的な背景も考えて、例えば、このエンドポイントはこういうものを見ているのだから、こうなるはずではないかみたいな考え方があっても、私はいいのだと思うのです。

単純ではなくて重みづけしているみたいだけれども、アベレージをとるという考え方はどうも、何か生物の反応から考えてみるとおかしな話で、そういう意味ではモデルをもっと生物学的な反応として考えていったほうがいいのではないかというふうに思っているのですけれども、これは素人の考え方なのです。

○広瀬座長代理 それは私も思うところではありまして、そうですね。本来はそうあるべきだと思います。ただ、そこは、そもそもベンチマークドーズはそういうエビデンスがない中でどうしたらいいかというのでも発展してきた部分もあるので、もともとのベンチマークドーズ法の考え方をどうするかという議論に戻るかもしれないのですけれども、そういった議論は必要なのだと思います。

○佐藤委員長 プラクティカルな議論とフィロソフィカルな議論は、多分切り離れたほうがいいように思うのです。これは全体で、世界的にもそういう話だろうと私は思っています。

○川村座長 ありがとうございます。

今、委員長からも御指摘がありましたようなことも含めて、やはり解決と言うと変ですけれども、取扱いを決めていかないといけない部分が多分にあるかと思っています。

広瀬先生、ありがとうございました。

まとめのほうに入っていきたいと思いますが、この2つの発表から、ベンチマークドーズ法の現状とか課題といったものが皆さんと共有できたのではないかと思います。

ベンチマークドーズ法を適用するとした場合に、どういった課題があるかとか、現在議論されている問題にどういったことがあるかとか、そういうことを把握した上で、そもそも採否の問題から始まって、採用するとしたらどのように取り扱っていくか。フィロソフィーの部分とプラクティカルな部分と、2つの側面を持ったものをどのようにしていくかということを考えていかななくてはいけないわけで、今後このワーキンググループにおいて、

今、指摘のあった種々の点を、例えばモデルの選択、あるいはRestrictionの問題、BMRの設定、モデルそのものを統合するような話も含めて、実際の取扱いに向けて合理性と実現性、フィージビリティも含めたことを検討していかないといけないのではないかと思います。

そういうことで、今後このワーキンググループの2つ目の課題、BMDを採用するとしたら、ということで議論を進めていかねばならないのではと考えているところです。

今日は、今までの課題の整理と、問題点がありますという投げかけでしたので、次回はこれをベースに議論を進めるようにしてまいりたいと思いますので、事務局のほうはこれに沿った形で準備をお願いしたいと思います。

○井上課長補佐 承知いたしました。

次回の議事及びお集まりいただく日時につきましては、日程調整の上、決まり次第お知らせいたします。

○川村座長 ありがとうございます。

投げかけなので、エンディングというのが収束して終わりではなくて、離散して終わるので、消化が悪いかもしれませんが、こういうサイエンスに基づく議論というのはそういうプロセスが必要だと思いますので、御理解いただきたいと思います。

今日予定した議事はこれだけでございますので、閉会いたします。ありがとうございます。