

II. 令和元年～令和3年度 分担研究報告

課題 3. 総合的な摂取量評価における推定値の精緻化および信頼性の向上に関する研究

研究分担者 鈴木美成

食品や環境からの農薬等の摂取量の推計と国際標準を導入するための研究
課題 3. 総合的な摂取量評価における推定値の精緻化および信頼性の向上に関する研究

研究分担者 鈴木美成 国立医薬品食品衛生研究所 食品部第四室長

研究要旨

本研究では、国民が食品を介して摂取する農薬の量をより精緻に推定するために、1) 確率論的な摂取量推定を行うため、喫食量の確率密度分布の推定、2) 不検出値を含むデータに対する統計的妥当性の高い推定法、3) モンテカルロシミュレーションを用いた確率論的な摂取量推定について検討を行った。

令和元年度は、確率論的な摂取量推定を行うため、喫食量の確率密度分布の推定をおこなった。喫食量には 0 にマスを持つ正の値であるので、ゼロ過剰を表現できる Tweedie 分布、ゼロ過剰ガンマ (ZIG) 分布、ゼロ過剰対数正規 (ZILN) 分布の確率密度分布を検討した。Tweedie 分布が最適だったのは、2 群, 8 群, 9 群, 10 群, 12 群であった。ZIG 分布が最適だったのは、1 群, 4 群, 5 群, 6 群, 7 群, 11 群であり、ZILN 分布が最適だったのは 3 群と 13 群であった。

令和 2 年度は、不検出例 (ND) を含むデータにおける統計妥当性の高い推定法として、従来もちいられてきた ND に LOQ の 0.2 倍の値を代入する方法 (代入法) と 最尤推定 (MLE) 法とベイズ (B)E 法を比較した。MLE 法による 1 日農薬摂取量の平均値は、全て BE 法による平均値の事後予測分布の四分位範囲内に収まっていた。しかしながら、代入法では一部の農薬が BE 法による平均値の事後予測分布の四分位範囲から外れていた。下限値が 0 のデータの割合が多く、相対標準偏差が大きいと、BE 法と MLE 法の推定値の相対的な差異が大きくなる傾向が認められた。最尤推定法は、サンプルサイズが小さく、歪度が大きいデータが不得意であることが知られている。この点を踏まえると、本研究の結果は、ベイズ推定の方が最尤推定よりも妥当な推定結果であることを示唆していると考えられた。

令和 3 年度は、令和 2 年度に引き続き、不検出例を含むデータにおける統計妥当性の高い推定法としてベイズ推定法を用いた摂取量評価を行った。さらに、真値に 0 を含むデータにおける統計妥当性の高い推定法に関する数値シミュレーション、および二次元モンテカルロシミュレーションを用いた摂取量分布の推定に関して検討を行った。数値シミュレーションの結果、MLE 法よりも BE 法の方が、推定の真度・精度・妥当性が良好であった。さらに、この傾向は事前分布の設定に寄らず、無情報事前分布を用いた場合でも BE 法の方が MLE 法よりも良好な真度・精度・妥当性を示した。以上のことから、ZILN 分布に従う不検出例を含むデータの解析においても、BE 法が有効であることが示された。これらの結果を基にゼロ過剰モデルも考慮して、各食品群中の農薬濃度分布を BE 法で推定し、2D-MCS による残留農薬摂取量分布を推定した。ADI を超過する確率は、アセフェートの 0.001% が最大であったことから、検討したいずれの農薬も健康リスクは小さいと示唆された。さらに、2D-MCS から推定した結果と、摂取量のデータから BE 法で推定した結果を比較したところ、農薬摂取量の平均値の事後予測分布が良い一致を示した。以上のことから、いずれの推定方法も妥当であると考えられた。

以上の検討により、食品に含まれる残留農薬を対象とした統計的推定において BE 法は有効であることが示された。

A. 研究目的

食品を介した農薬の摂取量評価は、残留農薬基準値といった健康リスクの管理を目的とする規格値策定等の行政施策の検討、及び効果検証のための科学的根拠となる。

食品中の農薬濃度については打ち切り (censoring) 問題が生じる。打ち切りは、ある値よりも大きい、小さい、またはその両方の値を非表示にする。切り捨てられた (truncated) データとは異なるのは、打ち切られたデータポイントの数がわかっている点である。より具体的に残留農薬分析の場合について言及すると、検出限界 (LOD) あるいは定量下限値 (LOQ) 未満のデータは、妥当な数値を割りてすることは困難となる。とくに、本研究課題の様に複数の研究機関が参加した調査の場合には、ある機関の LOQ が他の機関の LOQ よりも高い場合に ND となった場合には影響が大きくなる。

これまでの厚生労働省が主体となって行ってきた農薬摂取量評価において、LOD 未満・および LOD 以上 LOQ 未満のデータは、ND として扱い、ND には LOQ の 0.2 倍した値を代入し平均的な摂取量を評価してきた。しかしながら、代入法は問題のある方法として認識されて来っており、限定された場合にのみ使用を推奨されるようになってきている。代入法に代わり、最尤推定法などの代替法の使用が推奨されて来ている。しかしながら、サンプル数が少ない場合は、検出値と検出された割合のみを報告すべきとの提言もある。

このような場合に、ベイズ推定が有効となる可能

性がある。ベイズ推定は尤度分布にこれまでの知識を用いて重み付けしたものと解釈できる。適切な事前分布を用いることで、最尤推定よりも安定した推定値を得ることができると期待されている。そこで、2019 年から 2021 年度の間に得られた農薬摂取量のデータに対して、ベイズ法による推定を行い、代入法および最尤推定法との相違について評価した。

ND の問題と関連して、非天然型農薬は、使用履歴が無ければ、検体中の残留農薬含有量が 0 である可能性もありうる。近年になって、真値に 0 を含むようなデータの解析に関する報告がなされてきた。既存の報告は仮定したモデルの相違に注目した検討であり、推定法の差異については検討していない。また、真値が 0 である割合も最大で 60% であり、より広い範囲での検討が必要である。そこで、本研究では真値が 0 となる割合および不検出割合について幅広い組合せで、ベイズ推定 (BE) 法と最尤推定 (MLE) 法による推定結果の真度・精度・妥当性について評価を行った。また、BE 法については事前分布の影響についても検討した。さらに、TD 試料中の農薬濃度について適用を試み、モンテカルロシミュレーション (MCS) を用いた農薬の摂取量推定を行った。

B. 研究方法

1. 喫食量分布の推定

2014 年から 2016 年の国民健康・栄養調査のデータを解析に使用した (2014 年: 8047 件, 2015 年: 7456 件, 2016 年: 30820 件)。コード化された情報

(都道府県名, 食品分類) を変換後、無効レコードを削除したものを (約 10%が無効) 解析に使用した。データベース内の数字は適宜換算を行った。

喫食量は非負値の連続変数であり、0 にマスを持っていた。そこで、0 過剰を表現できる確率密度分布である Tweedie 分布、zero-inflated gamma 分布、zero-inflated log-normal 分布を検討した。

解析には R (3.4.0) を用いた。ベイズモデルの計算は rstan パッケージ (2.16.2) を用いた。作成したベイズモデルの妥当性は、下記の式で示した広く使える情報量規準 [widely applicable information criterion (WAIC)] を用いて WAIC が最も低いモデルを採用した。

2. 不検出例を含んだ残留農薬摂取量の推定

データには、2019 年から 2021 年に本研究班で行ったトータルダイエツスタディー調査による、残留農薬摂取量の下限値および上限値を用いた。ただし、下限値がすべて 0 となった農薬及び、サンプル数が 2 未満の農薬は除いた。

ここで、解析の基本となる尤度について基本的な数式を示す。 N 個の観測値 Y がパラメーター θ を持つ確率密度関数 $f(Y|\theta)$ に従うと仮定したときの尤度 (L) は、以下の式で算出することができる。

$$L(\theta|Y) = \prod_{i=1}^N f(Y_i|\theta) \quad \dots \text{Eq. 1}$$

打ち切りのあるデータの場合、定量できたレコードには確率密度関数 $f(Y|\theta)$ から確率密度を、打ち切りとなったレコードには累積確率関数 $F(L, U|\theta) = P(L \leq Y \leq U|\theta)$ から累積確率を使用して、尤度関数次のように求めることができる。

$$L(\theta|Y) = \prod_{i=1}^{N_{\text{obs}}} f(Y_i|\theta) \cdot \prod_{j=1}^{N_{\text{cen}}} F(L_j, U_j|\theta) \quad \dots \text{Eq. 2}$$

ここで、 N_{obs} は定量値が得られたサンプル数、 N_{cen} は定量値未満のサンプル数、 L_j はサンプル j の下限値、 U_j は上限値を示す。Eq.2 で示した尤度

関数を用いて、MLE 法および BE 法を用いた推定を行った。

解析には R (3.4.0) を用いた。ベイズモデルの計算は rstan パッケージ (2.16.2) を用いた。

作成したベイズモデルの選択は、下記の式で示した広く使える情報量規準 [widely applicable information criterion (WAIC)] を用いて WAIC が最も低いモデルを採用した。

3. 真値に 0 を含むデータの解析方法に関する数値シミュレーション研究

以下の①～⑥に従って、MLE 法および BE 法で分布パラメーターの推定を行った。

- ① サンプルサイズ=200、幾何平均値=1、幾何標準偏差=3 の乱数を生成した。
- ② 真値が 0 のレコードとなる割合 ω (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9) を決め、200 のサンプルから ω の割合で 0、 $(1-\omega)$ の割合で 1 を生成し、①と掛け合わせることで ZILN 分布に従う乱数を得た。
- ③ 不検出割合 φ (0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95) を決めた。ただし、 $\varphi > \omega$ を満たす。
- ④ 不検出とした最大値—検出された最小値間の一様分布から、不検出となる基準 (reporting limit, RL) を決めた。不検出となった測定値の下限値を 0、上限値を RL とした。
- ⑤ 無情報、弱情報、強情報事前分布を用いた BE 法および MLE 法を用いて、分布パラメーター等を推定した。その際には Eq.2 で示した尤度関数を用いた。ただし、 L の値には 0 を、 U の値には RL を用いた。
- ⑥ ①～⑤を 1000 回 (N_{sim}) 繰返す。

BE 法においては、3 つの事前分布を用いた; 無情報事前分布、弱情報事前分布 [gsd~cauchy(3,1), mean_est~cauchy(mean^{MB}, mean^{MB})], 特定情報事前分布 [$\omega \sim N(\omega, \omega/10)$, gm~N(1,0.1), gsd~N(3, 0.3)]。ここで、mean^{MB} は不検出となった観測値に LOQ の半値を代入して算出した平均値である。

推定した分布パラメーターおよび期待値等につ

いて、以下の項目で評価した。

$$\text{真度: rMSE}(\hat{\theta}) = \sqrt{\frac{\sum(\hat{\theta} - \theta)^2}{N_{\text{sim}}}}$$

精度: RSD($\hat{\theta}$)

妥当性: 平均値の 95%推定区間に母平均が含まれる確率

平均値の 95%推定区間は、BE 法では事後予測分布の 95%信用区間を用い、MLE 法ではブートストラップ法 (500 回) を用いて計算した平均値の 95%区間を用いた。

解析には R (4.1.0) を用いた。ベイズモデルの計算は cmdstanr パッケージ (0.4.0) を用いた。

4. 二次元モンテカルロシミュレーションを用いた農薬摂取量分布の推定

TD 試料中の残留農薬濃度データには、2019 年から 2021 年に本研究班で行った TD 試料中の残留農薬濃度の下限値および上限値を用いた。各食品群の喫食量には、2014 年から 2016 年の国民健康・栄養調査のデータを用いた。

解析には R (3.4.0) を用いた。ベイズ推定には rstan パッケージ (2.16.2) を用いて、2000 個のモンテカルロサンプルを事後予測分布として得た。得られた事後予測分布を分布パラメータとする乱数を 50 個発生させ、計 10 万 (2000×50) 個の乱数を残留農薬濃度及び喫食量について生成した。

C. 研究結果及び考察

1. 喫食量分布の推定

BE 法を用いて各食品群の喫食量 (g/person) を解析したところ、Tweedie 分布が最適だったのは、2 群、8 群、9 群、10 群、12 群であった。ZIG 分布が最適だったのは、1 群、4 群、5 群、6 群、7 群、11 群であり、ZILN 分布が最適だったのは 3 群と 13 群であった。

2. 不検出例を含んだ残留農薬摂取量の推定

BE 法では、対数正規分布が 12 農薬で、ガンマ分布が 4 農薬で、ワイブル分布が 4 農薬で最も適

していると判断された。

対数正規分布の形状パラメータは平均して 3.14 ± 1.20、ガンマ分布の形状パラメータは 1.63、ワイブル分布の形状パラメータは平均して 0.97 ± 0.51 であった。これらの値は、設定した事前分布との乖離は小さかった。この結果から、ベイズ推定に用いた形状パラメータの事前分布は妥当であったと判断できた。

BE 法と MLE 法で異なる確率密度分布と推定された農薬が散見された。MLE 法でガンマ分布が最適と判断されたもののうち、フルアジホップブチル・フルアジホップブチル代謝物・フラ時ホップブチルは、BE 法ではワイブル分布が最適であると判断された。

1 日農薬摂取量の理論平均値に関する事後予測分布と LOQ の 0.2 倍の値を代入した推定値および MLE 法と比較した。ND に LOQ の 0.2 倍の値を代入して推定した値よりも BE 法による推定値の方が低かったのは、アセフェート・メタミドホスであった。MLE 法よりも BE 法による推定値の方が低かったのは、アセフェート・クロチアニジン・クロルピリホス・チアクロプリド・ノバルロン・ビフェントリン・ヘキサジノン・メタミドホスであった。インドキサカルブ・ビフェントリン・ブプロフェジン・フルベンジアミド代謝物の BE 法による平均値の事後予測分布の四分位範囲内には、0.2LOQ を代入して推定した値が含まれていなかった。MLE 法による 1 日農薬摂取量の平均値は、全て BE 法による平均値の事後予測分布の四分位範囲内に収まっていた。

MLE 法と BE 法の差異に与える要因を解析するために、BE 法と MLE 法で推定した平均値の相対的な差 $[2(\hat{\mu}^{\text{BE}} - \hat{\mu}^{\text{MLE}})/(\hat{\mu}^{\text{BE}} + \hat{\mu}^{\text{MLE}})]$ について解析を行った。

すべての食品群で不検出となったサンプルの割合が高い場合、推定した分布の σ/μ 比が大きい場合、用いたデータのサンプルサイズが小さい場合に、BE 法と MLE 法による推定値の相対的な差が大きくなる傾向が認められた。L=0 となるサンプルの割合は、有効なサンプルサイズに相当すると考えられる。また、本研究で仮定した確率密度分布

の場合、 σ/μ が大きくなると歪度が大きくなることを意味する。これらの結果より、BE法とMLE法の差異が大きくなった要因は、歪みの大きい確率密度分布であること、サンプルサイズが小さいこと、および $L=0$ となるサンプルの割合が大きかったことの影響が大きいと考えられた。

MLE法は、有効なサンプルサイズが少なくなるほど、また分布の歪みが大きくなるほど、推定の精度が低くなることが報告されている。この点を踏まえ、MLE法が不得意なデータにおいて、事前分布を用いることでBE法がMLE法の弱点をカバーしているとも捉えることができる。つまり、BE法による推定結果の方がMLE法による推定結果よりも妥当である可能性が示唆された。

3. 真値に 0 を含むデータの解析方法に関する数値シミュレーション研究

BE法の点推定値として事後分布の中央値(MED)を用いて、真度・精度・妥当性を評価した。MLE法は、 ω よりも φ による影響を強く受け、 φ が大きくなるほど、推定の真度および精度が悪くなる傾向があった。一方で、BE法は φ よりも ω による影響の方が強い傾向にあった。また、MLE法よりもBE法の方が、推定の真度・精度・妥当性が良好であった。さらに、この傾向は事前分布の設定に寄らず、無情報事前分布を用いた場合でもBE法の方がMLE法よりも良好な真度・精度・妥当性を示した。MLEとBE(NIP)を比較すると、 ω と gm の真度・精度はBE法が広い φ と ω の組合せでMLE法よりも良い推定結果となる傾向を示したが、 gsd に関してはそれほど大きな差異は認められなかった。この結果は、事前分布に形状パラメータである gsd の情報を追加することで推定結果が向上する可能性を示唆するものと考えられた。実際に、弱事前分布として、 gsd に関する事前分布を用いることで、推定の真度・精度・妥当性が向上することが示された。以上のことから、ZILN分布に従う不検出例を含むデータの解析においても、BE法が有効であることが示された。また、事前分布の利用は推定の真度・精度・妥当性を向上させるが、通常の利用では真値は不明であるため、弱情報事

前分布の使用を検討するのが現実的であろう。事前分布に利用できるデータベースを整備することが有効であると考えられる。

4. 二次元モンテカルロシミュレーションによる農薬摂取量分布の推定

ゼロ過剰モデルも考慮に入れ、TD試料中の残留農薬濃度を、BE法を用いて推定した。解析の対象としたのは、14群全ての農薬が検出され、データサンプルサイズが6以上のものとした。ここでは、サンプルサイズの最も大きかったチアマトキサムの結果をTable 4に示す。14群中13群でゼロ過剰モデルが最適なモデルであると推定され、ZILN分布は11の食品群で、ゼロ過剰ワイブル分布が2つの食品群で最適な分布であった。検出率が最も高かった8群でもゼロ過剰モデルが採用されたことから、ゼロ過剰モデルが適切かどうかは検出率だけで決定されるものではないと示唆された。

農薬濃度分布の推定の不確かさを考慮するために、BE法で得られた2000個の事後予測サンプルから50個の乱数を生成し、計10万個の乱数を生成した。別途国民健康・栄養調査のデータを解析して得られた体重当たりの喫食量分布の乱数と掛け合わせ、摂取量分布を推定した。チアマトキサムの場合、5, 25, 50, 75, 96パーセンタイル値は、0.0005, 0.0024, 0.0060, 0.0148, 0.0586 $\mu\text{g/kg/day}$ であり、ADI (18 $\mu\text{g/kg/day}$) を超過する確率は0%であった。検討した農薬についての推定摂取量の平均値や各パーセンタイル値等をTable 7に示す。ADIを超過する確率は、アセフェートの0.001%が最大であったことから、検討したいずれの農薬も健康リスクは小さいと示唆された。

農薬摂取量の平均値について、2D-MCSから推定した結果と、摂取量のデータからBE法で推定した結果を比較した。異なるアプローチで算出した平均値の分布が良い一致を示したことから、いずれの推定方法も妥当であると考えられた。

D. 結論

食品中残留農薬濃度のような不検出例の多いデータに対する統計的推定方法について検討を行

った。推定法間の比較・推定法間の差異に影響を及ぼす要因の解析・数値シミュレーションの結果から、BE法の有用性が示された。

E. 研究発表

1. 論文発表

- 1) Suzuki Y, Tanaka N, Akiyama H.; Attempt of Bayesian Estimation from Left-censored Data Using the Markov Chain Monte Carlo Method: Exploring Cr(VI) Concentrations in Mineral Water Products. Food Safety. 8(4):67-89 (2020).

2. 学会発表

- 1) 鈴木美成, 穠山浩; トータルダイエツスタディによる農薬摂取量の推定におけるベイズモデルを用いた精緻化, 日本食品衛生学会第116回学術講演会 2020年11-12月.
- 2) 鈴木美成, 青柳光敏, 戸田英汰, 伊藤

功一, 福光徹, 萩尾真人, 林孝子, 新宅沙織, 井原紗弥香, 川崎恭寛, 中島安基江, 佐藤環, 飛石和大, 堀就英, 穠山浩; ベイズ法を用いた食品を介した残留農薬摂取量の推定の試み, 第29回環境化学討論会 2021年6月.

- 3) 鈴木美成, 青柳光敏, 戸田英汰, 伊藤功一, 福光徹, 萩尾真人, 林孝子, 新宅沙織, 井原紗弥香, 川崎恭寛, 中島安基江, 佐藤環, 飛石和大, 堀就英, 穠山浩; 不検出値を含むデータを用いたベイズ推定による残留農薬摂取量の精緻化の試み, 第58回全国衛生化学技術協議会年会 2021年12月.

F. 知的財産権の出願・登録状況

なし