



ベンチマークドーズ法(BMD)で使用する数理モデルの意味合い的由来とモデル選択について

食品安全委員会第8回WG
2017年10月20日

西浦博
北海道大学
nishiurah@med.hokudai.ac.jp

1

西浦 博

北海道大学大学院医学研究院
社会医学分野衛生学教室 教授
(平成28年4月より)



医学部：平成14年 宮崎医科大学
博士：平成18年 広島大学大学院
(専門)

理論疫学、生物統計学、感染症疫学
特に感染症を中心とした伝播・体内動態の機構的モデルを研究

(経歴)

平成16年 ロンドン大学インペリアルカレッジ (英国) 客員研究員
平成17-19年 チュービンゲン大学 (ドイツ) 研究員
平成19-22年 ユトレヒト大学 (オランダ) 研究員
平成23-24年 香港大学 (中国) Assistant Professor
平成25年-28年 東京大学大学院医学系研究科 准教授
平成29年より 北海道大学大学院医学研究科 (院) 教授

2

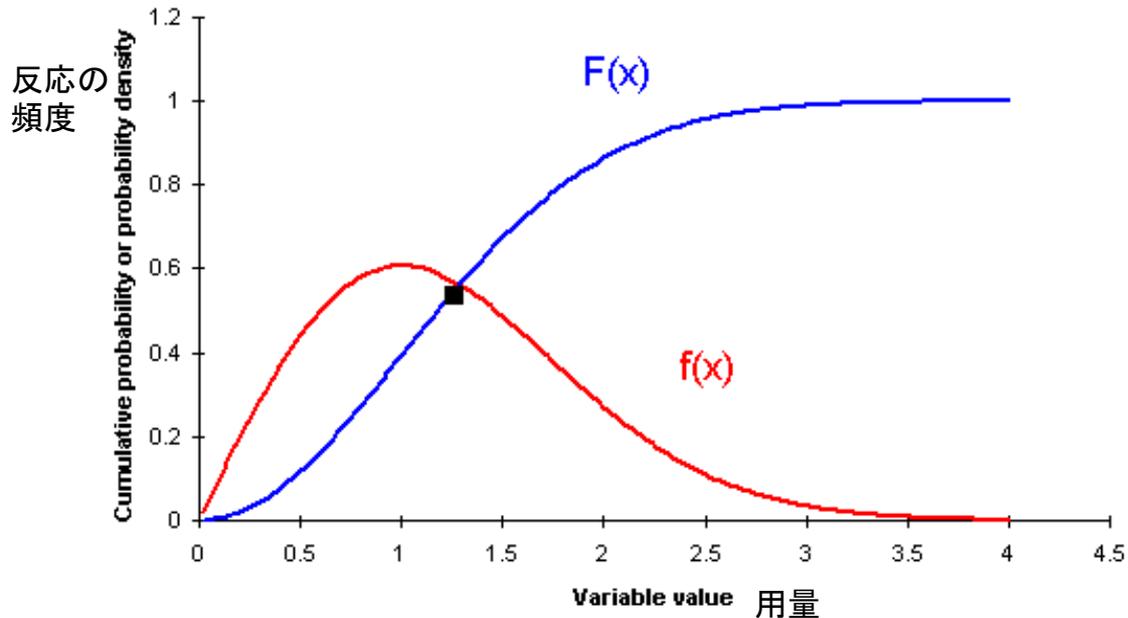
Contents

- ①モデルの成り立ち、内容について
- ②モデルの選択について:特に、使用モデルが決まらない場合
- ③モデルの平均化の妥当性について
- ④BMRの設定について、Restriction ONに対する考え方

Contents

- ①モデルの成り立ち、内容について
- ②モデルの選択について:特に、使用モデルが決まらない場合
- ③モデルの平均化の妥当性について
- ④BMRの設定について、Restriction ONに対する考え方

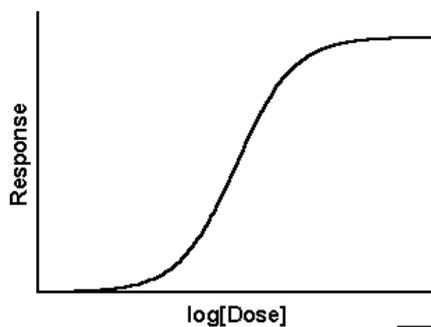
確率密度関数 $f(x)$ と累積分布関数 $F(x)$



数多くの $F(x)$ を用量反応関係に「あてはめている」

[https://www.vosesoftware.com/riskwiki/Probabilitydensityfunction\(pdf\).php](https://www.vosesoftware.com/riskwiki/Probabilitydensityfunction(pdf).php)

用量反応曲線 (dose response curve)



特定のDoseが被験者に投与され、
反応を観察する

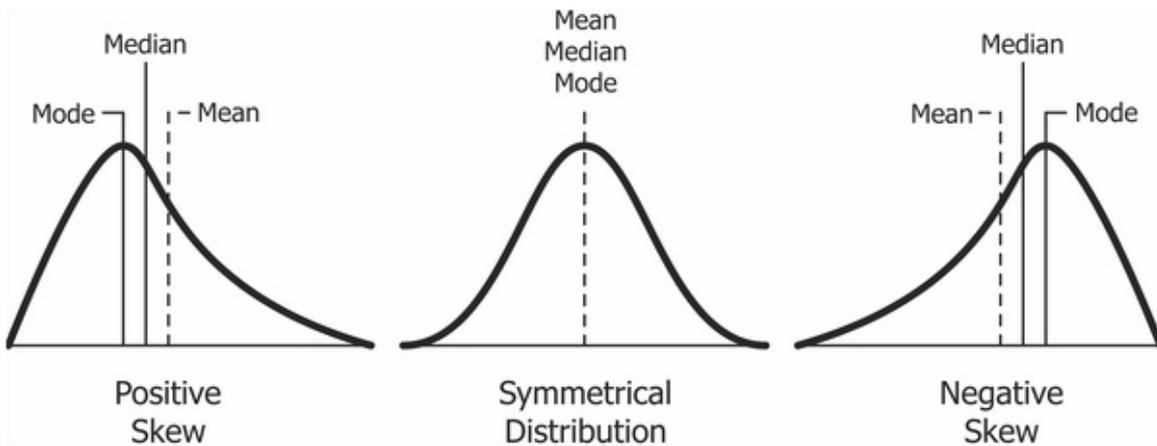
同じ反応の観察：

ある者は低いDoseで反応する

他の者は極めて高いDoseで反応する

カーブがねじまがる=>本来的に「理想的」でない
S字曲線: Popularなだけ。閾値付き曲線とかでもいい

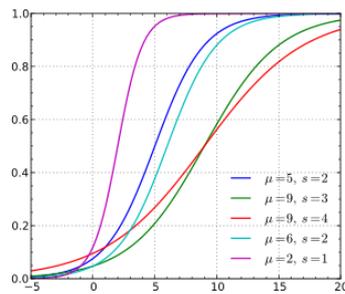
望ましい形



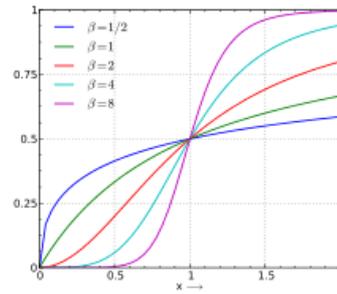
- Pureに形のみ: SkewnessとKurtosis(低次モーメントと言います)で当てはまりが良い
 - パラメータが少ないけれどFlexibleなものが良い
 - Dose responseと現象の対応があると良い
- 【勝手な想い】Low doseを含めた説明が可能であれば、なおさら良い

<https://www.safaribooksonline.com/library/view/clojure-for-data/9781784397180/ch01s13.html>

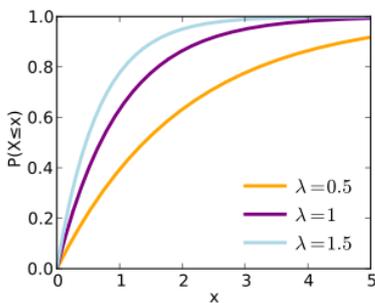
それぞれ、どんな形か？



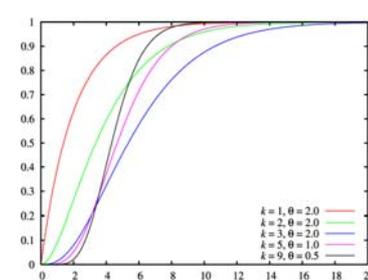
ロジスティック(Probitも似ている)



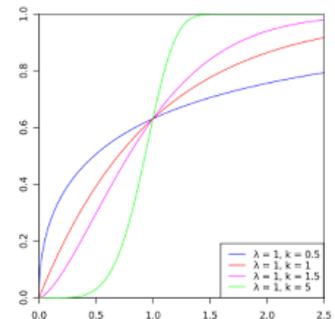
対数ロジスティック(対数Probitも似ている)



Quantal-Linear(指数分布)



ガンマ分布



ワイブル分布

各分布の英語版Wikipediaから引用

モデル	数式	パラメータの数	Restriction
Logistic	$\frac{1}{1 + e^{-(\alpha + \beta \text{dose})}}$	2	-
Probit	$\phi(\alpha + \beta \text{dose})$		原点付近の傾きが無限大になることを防ぐための制限
Log-logistic	$\gamma + \frac{1 - \gamma}{1 + e^{-(\alpha + \beta \ln(\text{dose}))}}$	3	β (slope) ≥ 1
Log-probit	$\gamma + (1 - \gamma)\phi(\alpha + \beta \ln(\text{dose}))$	3	β (slope) ≥ 1
Weibull	$\gamma + (1 - \gamma) \cdot (1 - e^{-\beta \text{dose}^\alpha})$	3	α (power) ≥ 1
Gamma	$\gamma + (1 - \gamma) \cdot \frac{1}{\Gamma(\alpha)} \int_0^{\beta \text{dose}} t^{\alpha-1} e^{-t} dt$	3	α (power) ≥ 1
Quantal linear	$\gamma + (1 - \gamma) \cdot (1 - e^{-\beta \text{dose}})$		単調増加でない曲線になることを防ぐための制限
Multistage two-stage (n=2) three-stage (n=3)	$\gamma + (1 - \gamma) \cdot (1 - e^{-\sum_{j=1}^n \beta_j \text{dose}^j})$ n = degree of polynomial	3 4	Betas ≥ 0 Betas ≥ 0

- ✓ EPA BMDSでは、実測データをモデルにフィッティングさせる段階で、パラメータに制限 (Restriction) を設けるオプションを選択することが可能。
- ✓ 生物学的に説明できない用量反応曲線にならないように、Restriction onとoffの両方でフィッティングすることを推奨。
(EPAテクニカルガイダンス2.3.3.3より)

9

第7回WG 広瀬先生スライドより

最初に気づくBMDモデルの仕掛け

ロジスティック・プロビット: 2パラメータのまま

他モデル: 左Tailのリスク最小値が γ (前頁パラメータ)になるよう設定 (γ はBackground doseと呼ばれる)

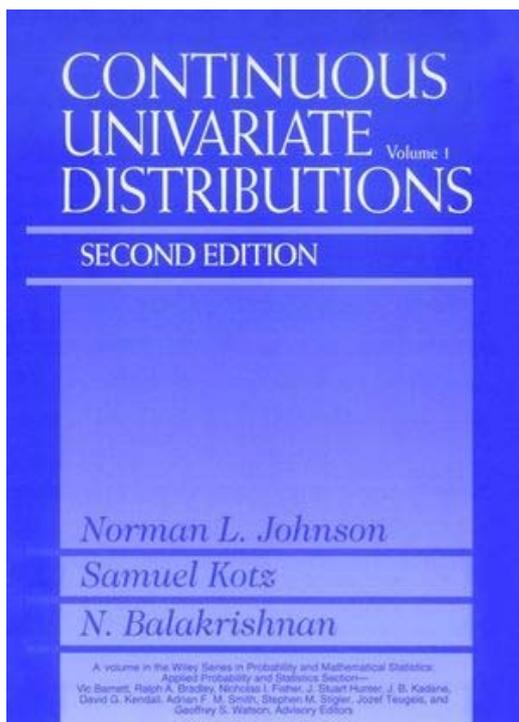
=> γ を用いた6つのモデルではBMDLの決定で γ の識別性 (γ を識別した上で得られるBMDLの信頼区間)が極めて重要になる

モデル	最小モデルのパラメータ数	左裾の最小値が γ
ロジスティック	2	No
プロビット	2	No
対数ロジスティック	2	Yes
対数プロビット	2	Yes
ワイブル	2	Yes
ガンマ	2	Yes
指数・ワンヒット	1	Yes
マルチステージ	≥ 2	Yes

定量的リスクアセスメントに使用するモデルの大まかな分類

1. 外挿モデル (BMRで本日カバー予定は1.1と1.2の計8モデル)
 - 1-1. **確率分布に基づくモデル**
プロビット、ロジスティック、ワイブルなど
 - 1-2. **発病機構に基づくモデル**
マルチステージモデル
 - 1-3. 用量詳細 (暴露時間など) を考慮したモデル
 - 1-4. 生物学的機構に基づくモデル
2. PB-PKモデル (体内でのADMEを考慮するモデル)

11



Johnson NL, Kotz S, Balakrishnan N. (1994),

Continuous univariate distributions. Vol. 1, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics (2nd ed.),

New York: John Wiley & Sons, ISBN 978-0-471-58495-7

12

ワイブル分布

技術者Weibullが考案・改良した分布. 歴史は浅い:1939年
「**故障の分布**」として名高い。

発端:「材料の破壊強度データは正規分布にうまくフィットしない。局所的な強度の定義が正規分布しようとも。。。」

理由:「破壊は材料の最も弱い箇所に発生する。**最も弱い箇所で一気に進む。**」

Weibull:「1つの鎖の強度は、構成物であるリング1つひとつの強度の平均値だとか分散では決定しない。そうでなくて、**最も弱いリングの特性で決まるはずだ!**」

13

Jiang, R.; Murthy, D.N.P. (2011). Reliability Engineering & System Safety. **96** (12): 1619–26

ワイブル分布(続き)

リングがn個の1つの鎖。加重xでリング1つが切れる確率G(x)は

$$G(x) = 1 - \exp(-\phi(x))$$

鎖のどこかが加重xで切断される確率F(x)は

$$1 - F(x) = 1 - (1 - G(x))^n = 1 - \exp(-n\phi(x))$$

$n\phi(x)$ の最良は $(x-g/a)^m$ であると経験的に見つけて以下を得た:

$$F(x) = 1 - \exp\left(-\left(\frac{x-g}{a}\right)^m\right)$$

14

ワイブル分布とRisk assessment

イベント発生までの危険率が経過時間で変化する

⇒ プロセスに「記憶」がある:

ハザード関数
$$\lambda(x) = \frac{m}{a} \left(\frac{x-g}{a} \right)^{m-1}$$

プロセスに「記憶」がない

指数分布のハザード関数
$$\lambda(x) = \frac{1}{a}$$

「古い車ほど壊れやすい」ことをモデル化できる

Jiang, R.; Murthy, D.N.P. (2011). Reliability Engineering & System Safety. **96** (12): 1619–26 15

ガンマ分布

発端: **「指数分布で済む話を繰り返したらどうなるのか」と数理統計者が考えた**

例題: 「西浦君は平均3個/年でグラスを割ってしまいます。ある日、グラスを1ダース(12個)買いました。西浦君が1ダース全て割るまでにかかる時間はどのような分布に従いますか？」

考え方: ひとつのコップが割れるイベントをランダムでメモリのないプロセスだとすると、イベント発生の時間は指数分布に従う。

$$f(x) = \lambda \exp(-\lambda x)$$

Dubey, Satya D. (December 1970). "Compound gamma, beta and F distributions". *Metrika*. **16**: 27–31. 16

ガンマ分布(続き)

2個が割れる。

$X_1 + X_2$ (ランダム変数の和は「**畳み込み**」)

$$\int f(x-y)f(y)dy = \lambda^2 x \exp(-\lambda x)$$

n個が割れる

$X_1 + X_2 + X_3 + \dots + X_n$

$$g(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x)$$

17

Dubey, Satya D. (December 1970). "Compound gamma, beta and F distributions". *Metrika*. 16: 27–31.

ガンマ分布とRisk assessment

1. **ポアソン待ち時間**

ポアソン過程に従うランダムイベントの待ち時間
(洪水がポアソン過程に従う場合は向いている)

2. ポアソン密度のランダムな変動に利用

数理的に以下が成立することが知られていて便利

$$\text{Poisson}(\text{Gamma}(a, b)) = \text{NegBin}(a, 1/(b+1))$$

3. ベイズ推定における共役事前分布

4. 正規分布のベイズ推定における事前分布

ガンマ分布＝マルチヒットモデル

Environmental Health Perspectives
Vol. 42, pp. 23-27, 1981

Modeling of Dose-Response Relationships

by Bernard Altshuler*

The Multi-hit Response

The gamma distribution function

$$P(d) = \frac{1}{\Gamma(k)} \int_0^{\theta d} x^{k-1} e^{-x} dx \quad (2)$$

can also provide a good fit to data (6).

It can be derived from the assumption that there is a discrete change, called a hit, which has to occur several times in order to produce the response. The expected number of hits is propor-

観察イベントが起こるまでに「ヒット」が何度も起こる。

Doseがdだと、 θd 回のヒット

k回ヒットが起こる確率は $(k-1)!$

理論的起源：Nording (1953) 発がん年齢からモデルを仮説

適用事例：エチルニトロソ尿素と消化管発がん

Communications in Statistics - Theory and Methods, 1995, 24, 2621-33.

ロジスティック分布

「**ロジスティック関数**」に由来する。**s字曲線のパイオニア**

$$f(x) = \frac{L}{1 + \exp(-k(x - x_0))}$$

Verhulst (1844) の人口増加モデル

$$\frac{dP}{dt} = rP \left(1 - \frac{P}{K} \right) \quad \rightarrow \quad P(t) = \frac{KP_0 \exp(rt)}{K + P_0 (\exp(rt) - 1)}$$

ロジスティック分布とRisk assessment

- ・正規分布に似ている。左右対称。正規分布よりも kurtosisが高い (縦に長くなれる)
- ・個体群生態学と経済モデルで頻用される (Growth curveと解釈できるし、ピーク1つで正規分布に似ていて便利だから)
- ・リスク解析では「使いやすいから」用いられているだけで、実は頻繁に出てこない (数理的にDose 0でリスクゼロから開始しない、単なる当てはめで正規分布に近い等が理由)

21

N., Balakrishnan (1992). Handbook of the Logistic Distribution. Marcel Dekker, New York.

対数ロジスティックモデル

(馴れ初め)

Log(X)がロジスティック分布に従うなら、Xは対数ロジスティック分布に従う

(似ている: 正規分布と対数正規分布の関係)

$$F(x; \alpha, \beta) = \frac{1}{1 + \left(\frac{x}{\alpha}\right)^{-\beta}}$$

応用: 経済学におけるFiskモデル (富の分布が似ている、 $1/\beta$ がGini係数を与える)

気象学における降水量との最大値の分布 (数理的に近似なしに観察データにフィットできるので便利)

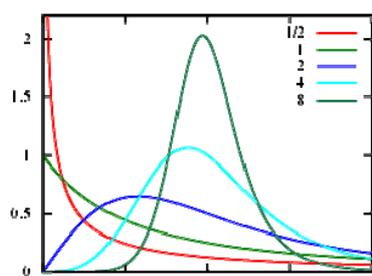
Diekmann, A. (1992), J Scientific and Industrial Research 51, 285.

22

対数ロジスティックモデルとRA

対象イベントがたくさんの変数で起こっているとするならば、対数正規分布のほうが自然な分布

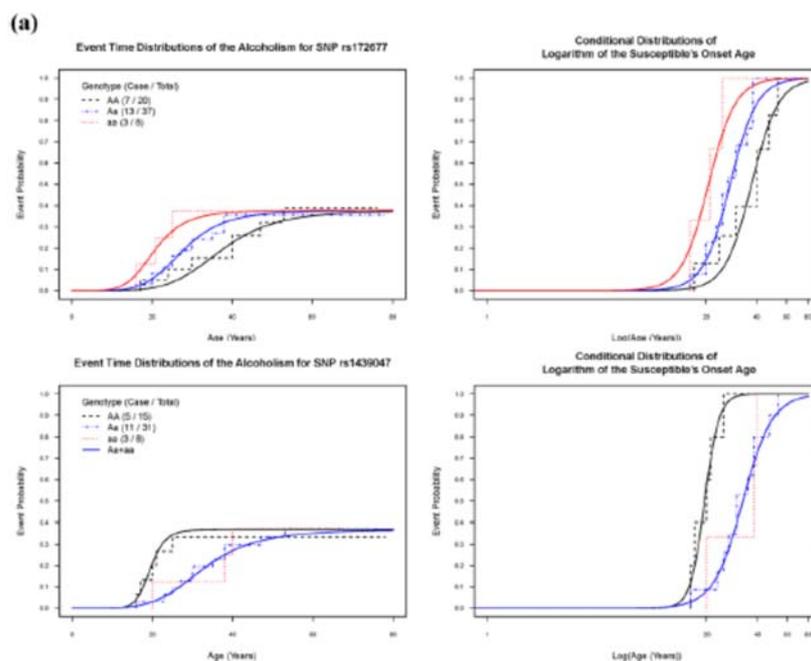
多数の変数の中で、1つか2つの要因がdominant或いは相関している場合、対数正規分布よりも狭い分布になりがち => 対数ロジスティックがより適切



https://en.wikipedia.org/wiki/Log-logistic_distribution

Diekmann, A. (1992), J Scientific and Industrial Research 51, 285.

23



Logistic-accelerated failure time location-scale mixture regression model. To facilitate the genotype analysis procedures, we used the web-based statistical software system EHA-RiskFree with a user friendly interface, which was developed according to the logistic-AFT location-scale mixture regression model¹². We herein summarized the formulation, inference, and properties of the model.

We assumed that T of an alcoholism susceptible case follows a log-logistic distribution⁴⁸ in the AFT location-scale regression submodel; in brief, error term ε has the logistic distribution of the density $f(\varepsilon) = e^\varepsilon / (1 + e^\varepsilon)^2$. We fit a logistic regression submodel and an AFT log-logistic regression submodel to simul-

プロビットモデル

2分変量(イベント発生の有無)が説明変数(Doseなど)を入力情報として利用した場合に正規分布で与えられる:

$$\Pr(Y = 1; x) = \Phi(\alpha x + \beta)$$

β はパラメータ。

1930年代に薬物処方量と回復率のリンクが研究され、発展した

現在では、ロジスティックモデルと対比されるシンプルな関数の1つ(Excelで計算できるためか好まれる)

Bliss, C. I. (1935). "The calculation of the dosage-mortality curve". *Annals of Applied Biology*. 22: 134–167. ²⁵

プロビット vs ロジスティック

プロビット $\Pr(Y = 1; x) = \Phi(x\beta + \varepsilon)$

ロジスティック $\Pr(Y = 1; x) = \frac{1}{1 + \left(\frac{x}{\alpha}\right)^{-\beta}}$

ロジスティックのほうがTailが厚い(高いDoseでゆっくり反応)、プロビットのほうが早く上がりきる

Logitのほうが数理的には解釈がマシ、という違い

対数プロビットモデル

他と比べて、大変リスク解析寄りに発展してきたモデル

個々の動物は「これ以下だと反応が起こらない」という用量(閾値)を持つはずで、それより上だとイベントが起こる(Finney, 1971)

鍵となる想定: 閾値ドーズの対数は正規分布に従う

$$\Pr(Y = 1; x) = \Phi(\alpha \log_{10} x + \beta)$$

D.J.FINNEY: Probit Analysis. Cambridge: Cambridge University Press, 1971 (3. edition). 27

対数プロビットモデルとRA

最も反応が起こりやすい用量はED50である

=>以後、ED50の計算には対数プロビットが頻用される
イベント発生頻度はLog(dose)が平均Log(dose)からどれくらい離れるかで、正規分布で決まる

発がんまでの時間の分布が対数正規分布に従うとき、対数プロビットの用量反応関係になる(Chand & Hoel)

Diekmann, A. (1992), J Scientific and Industrial Research 51, 285.
Fishbein L. (2011) Potential Industrial Carcinogens and Mutagens 28

Quantal-Linear モデル

“quantal response”: イベント現象の観察頻度(2値データ)に基づくモデル化

Quantal-linearモデル=>指数分布のこと

$$F(x) = 1 - \exp(-\beta x)$$

詳しいBMDではQuantal quadraticモデルも用いる

$$F(x) = 1 - \exp(-\beta x^2)$$

後者は「適合が良いから」

29

指数分布モデル

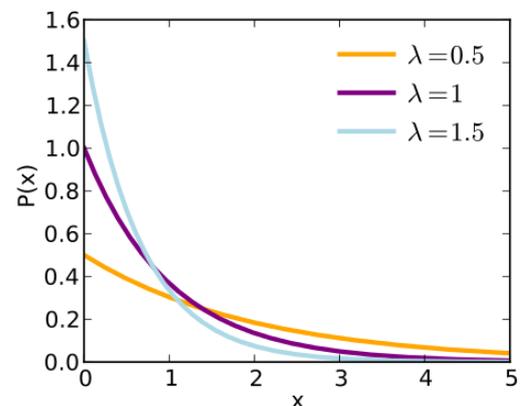
興味のあるイベントまでの時間がポアソン過程(ランダム)

例) 次の地震までの時間

放射線量の減衰

電話時間の分布

Risk assessmentでの意義: 短い時間に1回のみ起こるか否か興味があり、単位時間当たりのイベント発生ハザードが一定である、という場合



ワンヒットモデル＝指数分布モデル

発がん物質が遺伝子に衝突し、損傷を受けた遺伝子ががん細胞に転化する。

解釈:

1. 毒性物質はランダムに分布しており、その摂取量はポアソン分布に従う。
2. 毒物量当たりのイベント確率は一定
3. 1つの細胞または分子に発がん物質がヒットする確率は指数分布に従う

D. F. Schmidt and E. Makalic, "Universal Models for the Exponential Distribution", IEEE Transactions on Information Theory, Volume 55, Number 7, pp. 3087–3090, 2009

31

マルチステージモデル

発がんは「多段階」で「加速度的に」起こるというメカニズムを捉えるためのアイデア

細胞——>イニシエート——>腫瘍——>浸潤・転移
initiator promotor progressor

癌の状態に近いほど、何度もDoseの影響が重なる
線形化多段階モデル(Linearized multistage model)

$$F(x) = 1 - \exp(-\beta_1 x - \beta_2 x^2 - \beta_3 x^3 \cdots - \beta_k x^k)$$

発がんスロープファクター

線形化多段階モデル(Linearized multistage model)

$$F(x) = 1 - \exp(-\beta_1 x - \beta_2 x^2 - \beta_3 x^3 \cdots - \beta_k x^k) \\ \approx \beta_1 x$$

ただし、モデル妥当性が過去に議論(100-1000倍の違い)

1. いくつかの癌で閾値が見られない
2. 適合性のみで下向きの外挿が厳しいから。
3. 変異原試験では低用量まで直線性を認める
4. 閾値なしの仮定 + 安全側の推定をするべき。

33

機構的・数理的由来と特性まとめ

モデル	分類	由来	リスクアセスメント特性
ワイブル	機構	故障.	プロセスに記憶がある
ガンマ	機構	指数分布の繰り返し	ポアソン待ち時間(洪水)
ロジスティック	分布	人口増加. 左右対称	Growth curveが「当てはまりやすいから」
対数ロジスティック	分布	変数の対数がロジスティック(数理的由来)	多数の複合メカニズムの1-2個がDominantな場合
プロビット	分布	薬物処方量と回復	ロジスティックよりTailが薄い
対数プロビット	分布	リスク解析(低用量で無反応のドーズ探索)	ED50に基づく計算が容易
指数・ワンヒット	機構	ポアソン過程1回	地震のイメージ
マルチステージ	機構	多段階発がん	線形化多段階モデルの低用量での活用経験

34

Contents

- ①モデルの成り立ち、内容について
- ②モデルの選択について:特に、使用モデルが決まらない場合
- ③モデルの平均化の妥当性について
- ④BMRの設定について、Restriction ONに対する考え方

35

BMDモデル選択について(おさらい)

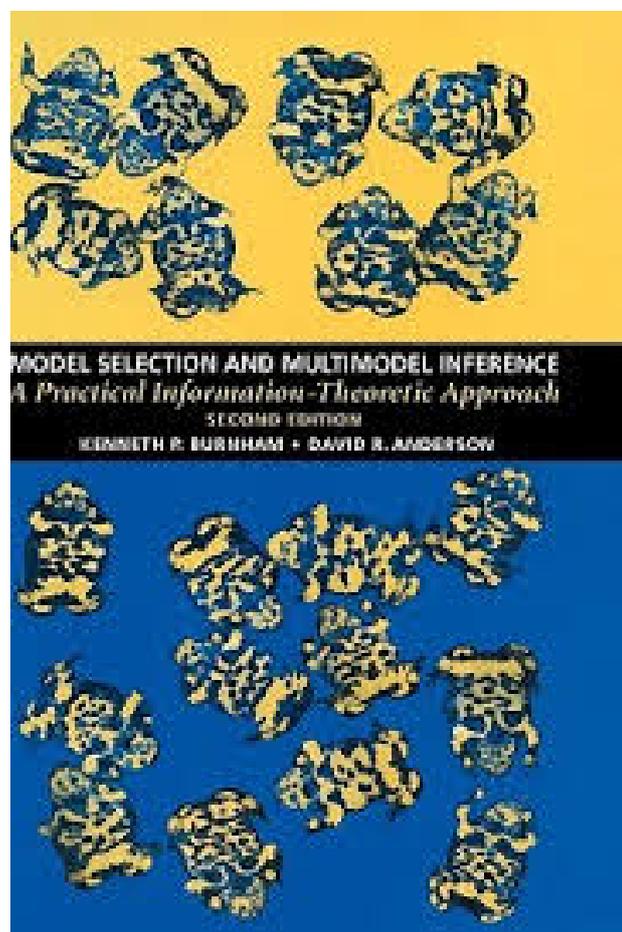
今日のBMD法における2つのモデル選択プロセス

- 1.EPAのBMDSなどを利用する際、以下棄却基準を設定
適合度検定でp値が0.1未満のモデルを却下
BMD/BMDL比が10以上(NOAELから外れる)却下
2. Survivedの(フィットした)モデル
「BMDLが最も低いものを取る」
か、「モデル依存性がなければ最もAICが低いもの」
か、Averagingなどを検討(後述)

36

Kenneth P. Burnham, David R. Anderson

Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach
Springer; 2nd ed. 2002. Corr. 3rd printing 2003



37

モデル選択とは

モデル選択の科学=>「**予測を重視する**」: 同じサンプリングを繰り返しても当てはまりが良い

手元にある限定的な観測データに基づいて予測のためのモデルを構築

目的: 予測が良い。真のモデルでないが、予測がうまくいくモデル(簡単なモデル)

AIC、DIC(モデル選択規準)・・・上記のモデル選択を簡便にすませてもらおう

38

Akaike Information Criterion

複数の候補モデルがある場合、最も好まれるモデルは次のAICが最小値となるモデル:

$$AIC = 2k - 2 \ln(L)$$

AICは適合度(尤度で評価)し、また、同時に必要とするパラメータ数のペナルティも取り込んだ指標。オーバーフィッティングを避ける

最も当てはまりの良いモデル: パラメータ数の多いモデル

AIC: 「当てはまりだけでなく、パラメータ数も勘案した比較」

Akaike, H. (1974), "A new look at the statistical model identification",
IEEE Transactions on Automatic Control, 19 (6): 716–723, ³⁹

BMDソフトウェア(EPAなど)でのモデル選択での注意

2段階の評価をしている

【1段階目】

適合度検定(p値): 有意に乖離しているモデルを捨てるため。
p<0.1などのモデルは「フィットさえしていない」ので捨てたい。
捨てた後は、(残ったモデルの)p値を以後参照しない。

【2段階目】

残ったモデルでAICの比較、BMDLの保守的な解釈、モデルの意味論(後述)でモデル選択(p値は同時に参照しない)

続き：よくある誤解「モデル選択時に検定もする」

- ・検定はNeyman-Pearson的な仮説検定の考え方。
 - ・モデル選択の情報量規準とは考え方が異なる
- 「検定すべきモデルをモデル選択で選ぶ」というのもおかしい

別の注意点： 選択されたモデルのパラメータの95%信頼区間に注意するのは重要。「モデル選択云々の前に、パラメータの信頼区間が広すぎると適合が怪しい。」

Burnham and Anderson, 2002
41

AICで選択したモデルとは

1. たくさんあるモデルたちの中で**最良の予測が得られるモデル**
=>もう1度、ランダムに同じようなデータをサンプリングしても、たくさんあるモデルの中で最もよく適合すると期待される
2. **真のモデルを選んでいるわけではない**
3. モデル選択は**サンプルサイズに影響されている**ことに注意
n数が少ないデータ=>単純なモデルがAICに好まれやすい
n数が多いデータ=>複雑なモデルが選ばれやすい

AICの兄弟分：AICc(小サンプル), DIC, BIC (Schwarz Criterion),

その他：Adjusted R2

Burnham and Anderson, 2002
42

AICがほんの少しでも小さければ。。。

- ・原則、最小AICのモデルが良い

- ・AICにもばらつきがある。

AICの差が0-2の間だと「より母集団に近いモデル」が正しく選択されないリスクがある。4-7だと「ちょっと危険」、10以上だと「大丈夫」

- ・どのようなときでも「近い」AICの2番手・3番手候補も列挙・比較すべき

- ・前述の通り、パラメータ値の信頼区間を比較すべき

Burnham and Anderson, 2002
43

モデルの機構的解釈による選択

- ・現象の構造の本質を捉えたモデルが良い

- ・AICによるモデル選択では対象とするシステムの構造の科学的証拠を特定できない

- ・他方、重要なDrawback

「がんと言えはMulti-stageのみ」=>誤り。モデルメカニズムの仮説通りに推定が当てはまっていなければ意味が失われかねない

他でも、データが発生するメカニズムが不明なら、それは“Model uncertainty”であって複数のモデルで分析して比較すべき

「Quantal linearがいつも問題だから外す」=>誤り。除外理由に恣意性を入れたくない。

機構的モデルが優先的に選択される とき

西浦の考え

1. 物質レベルで生命現象と共に実証経験がある場合

例：Doseが増えて多段階発がんの線形モデルが生体内データを基に実証されている病気を対象にする場合

2. 経験的に特定分布が特定現象に適合しやすいエビデンスがある場合

例：「急性感染症の潜伏期間は対数正規分布に従う」

「腸管感染症のDose response curveはBeta Poisson分布が最適である」

・生体内現象の説明が既に実施されて「どうして対数正規分布か」が一定の数理的レベルで説明可能である

・興味の対象である暴露・アウトカムの組合せについて経験的に何度も適合されてきた

注意：「宇多田ヒカルのCD売り上げは対数正規分布に。。。」

45

Lowest BMDL vs AIC(同一データが与えられたとき)

Lowest BMDL

(利点) **ConservativeなBMDLの決定**が可能

良い実験(試験数↑)へのインセンティブが働く

(欠点) **サンプルサイズに依存して、すごく低いDoseになる可能性**

妥当性不明でフィットが悪いモデル(パラメータ γ)の広い信頼区間

AIC

(利点) **当てはめの最も良いモデルが選択**可能

Low dose付近での判断に対して恣意性が入りにくい

(高用量=>低用量の予測をしているので、どんな場合も恣意的)

(欠点) **パラメータの信頼区間は気にしていない**

サンプルサイズが小さければAICも誤ることがある

46

Contents

- ①モデルの成り立ち、内容について
- ②モデルの選択について:特に、使用モデルが決まらない場合
- ③モデルの平均化の妥当性について
- ④BMRの設定について、Restriction ONに対する考え方

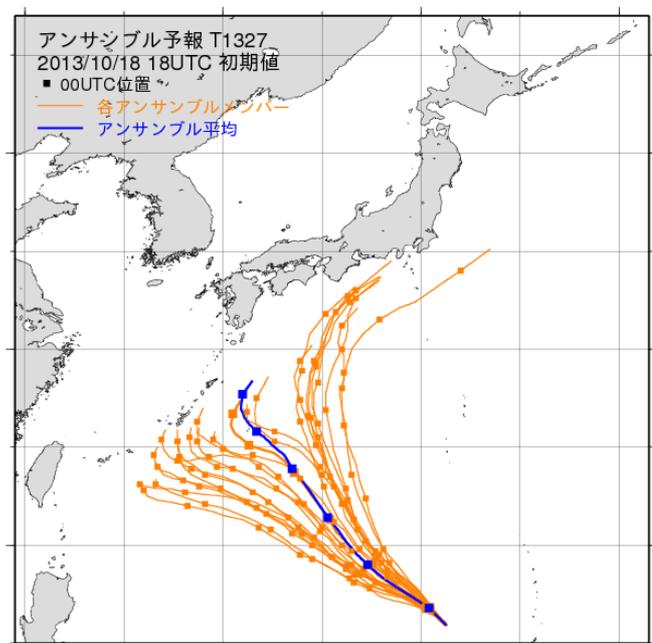
47

アンサンブル予測という考え方

モデル予測の先駆者:天気予報

数値予報誤差:初期値に含まれる誤差拡大、数値予報モデルが完全ではないこと

アンサンブル予報:異なる初期値・条件を多数用意して、多数の予報を行う。その平均やバラつきを利用する



48

AICやBICでモデル平均化話が出る理由

選択後推定量 (Post model selection estimator; PMSE)の問題:

- ・AICやBICでのモデル選択自体も不確実である
- ・その不確実性が推定量の性質に及ぼす影響を考慮しないと、何かと誤った推論が行われがち
- ・例: モデルから求まる信頼区間が誤ったものになる! ?

(BMDLに関わる)

BMDではWheelerらが提唱:

「真のモデル」からは相当に離れて、「予測が良い」ことを追求し、
パラメータの区間推定値を含めて各所で追及したもの

49

Wheeler MW, Bailer AJ. Properties of model-averaged BMDLs: a study of model averaging in dichotomous response risk estimation. Risk Anal. 2007;27(3):659-70.

BMD法におけるモデル平均化の狙い

全てのモデル推定値を統合。モデルの不確実性を考慮

1つのモデルに集中するより「最もらしい」に特化

$$\widehat{BMDL}_{MA} = \sum_k w_k \widehat{BMDL}_k$$

$$w_j = \frac{\exp(-I_j / 2)}{\sum_{i=1}^K \exp(-I_i / 2)}$$

ただし、個々の指標(BMDLなど)そのものの平均化は信頼性に乏しい(完全にad hocなため、再現性は担保されない)とわかった

50

Wheeler MW, Bailer AJ. Properties of model-averaged BMDLs: a study of model averaging in dichotomous response risk estimation. Risk Anal. 2007;27(3):659-70.

Dose response 曲線を平均化

$$\pi_{MA}(d) = \sum_{i=1}^K \pi_i(\boldsymbol{\theta}, d) \cdot w_i$$

$$BMR = [\pi_{MA}(d)_i - \pi_{MA}(0)] / [1 - \pi_{MA}(0)].$$

BMDLは、ブートストラップ法(もしくはベイズ推定)で、設定された推定値の95%信頼区間から計算

信頼性(再現度)が良いことを証明。**BMDの不確実性を正確に捉えたBMDLの決定が可能である=>BMDL決定のみだと最良のモデルを1つ選ぶよりも優れる**

(BMDL決定の際、Quantal-linearモデルは平均化の信頼性が低い傾向)

51

Wheeler MW, Bailer AJ. Properties of model-averaged BMDLs: a study of model averaging in dichotomous response risk estimation. Risk Anal. 2007;27(3):659-70.

モデル平均化は「妥当なのですか？」

西浦の答え:

1. どんなモデルも「誤り」です。「妥当ですか？」はUnfair Question
2. **BMD法におけるモデル化哲学**では、「真のモデル」を探求する努力を否定しませんが、「**良い予測モデル**」を追求することを**第一にしてリスク評価**をしています
3. だから、「真のモデル」という意味では、平均化前も後も決して妥当ではありません
4. 「**当てはめの良いモデル**」という意味では一定の検証を経た方法に基づいており、**アクセス可能データに基づく最も「数値的に妥当な」BMDLの決定**ができるようです。

52

Wheeler MW, Bailer AJ. Properties of model-averaged BMDLs: a study of model averaging in dichotomous response risk estimation. Risk Anal. 2007;27(3):659-70.

Contents

- ①モデルの成り立ち、内容について
- ②モデルの選択について:特に、使用モデルが決まらない場合
- ③モデルの平均化の妥当性について
- ④BMRの設定について、Restriction ONに対する考え方

53

BMRの設定について

- BMDはBMRで推定される反応の増加にともなって推定されるものでドーズdが以下を満たす

$$BMR = [\pi_d - \pi_0] / [1 - \pi_0] \text{ or}$$

$$BMR = \pi_d - \pi_0$$

- BMRは頻繁に次のいずれかである。1%, 5%, 10%.

いずれも恣意的なもので、Conventionallに10%がソフトウェアで設定されていることが多い(疫学データのみ1%とか)

- BMDLとはBMDの $100(1-\alpha)\%$ の下側の信頼区間を指す。

54

BMRの設定(続き)

Table 1: Illustrations of upper bounds^(a) of effect at NOELs related to 10 substances evaluated pr

Substance (source +year)	Endpoint	Quantal data	Continuous data
		Upper bound extra risk (%) ^(b)	Upper bound percent change (%) ^(c)
Thiodicarb (JMPR, 2000)	Splenic extramedullary haematopoiesis	21	
Carbaryl (JMPR, 2001)	Vascular tumours	15	
Spinosad (JMPR, 2001)	Thyroid epithelial cell vacuolation	2.7	
Flutolanil (JMPR, 2002)	Erythrocyte volume fraction		9
	Haemoglobin concentration		9.7
	Mean corpuscular haemoglobin		3
	Decreased cellular elements in the spleen	30	
Metalaxyl (JMPR, 2002)	Serum alkaline phosphatase activity		260
	Serum AST		100
Cyprodinil (JMPR, 2003)	Spongiosis hepatitis	5.1	
Famoxadone (JMPR, 2003)	Cataracts	29	
	Microscopic lenticular degeneration	29	
Tributyltin (EFSA, 2004)	Testis weight		9.1
Fumonisin (EFSA, 2005)	Nephrosis	8.6	
Deoxynivalenol (EFSA, 2004)	Body weight		10.5
Ethyl lauroyl arginate (EFSA, 2007)	White blood cell counts		23

・「リスク10%の超過」で閾値決定。
Backgroundで15%、追加リスクが8.5%なら超過リスクは $8.5/(100-15)=10\%$

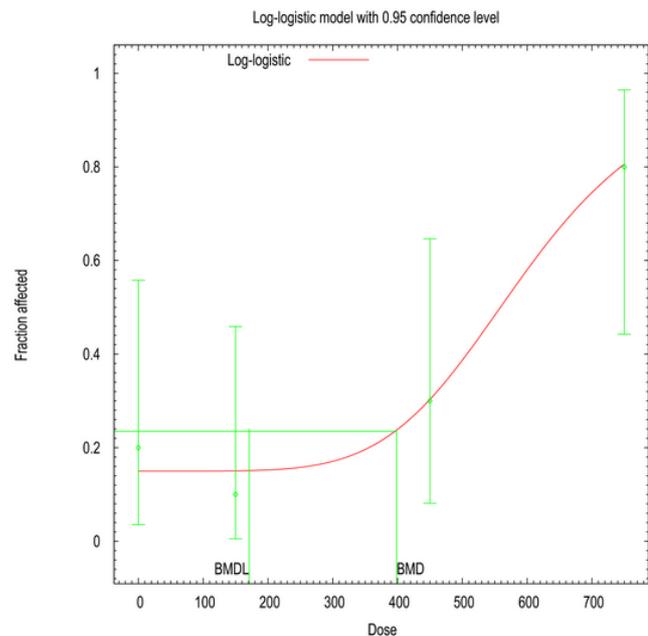
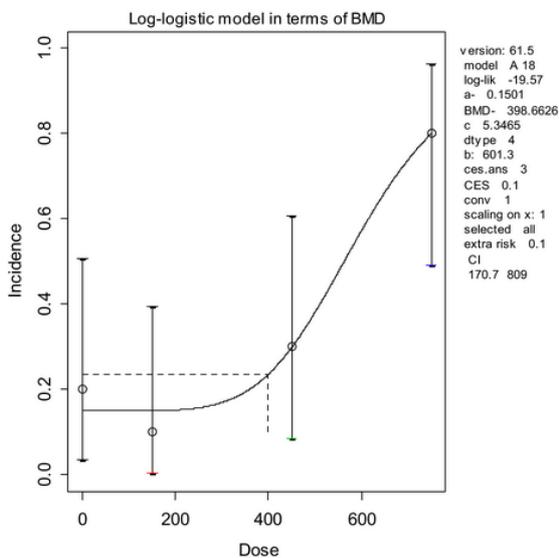
・BMD10「Defaultでない」

・NOAELより下を保障したい。

経験的に10%でBMDを定義すると大丈夫なことが多い。連続データだと5%

EFSA Journal. 2017;15(1):e04658

Incidenceを利用したときはExtraをとる



EFSA Journal. 2017;15(1):e04658

BMRの設定に対するEPAの姿勢

・2値データ

BMR: 10%はStandardである(化学物質やEndpointを比較する上において)

10%は癌バイオアッセイのほとんどおよび非癌バイオアッセイのいくつかにおける感度限界のあたりに相当

統計学的あるいは生物学的な理由でLower BMRも使う

ただし、BMC10は常にプレゼンすべき

・連続データ

BMR: エンドポイントの変化レベルに応じて生物学的有意性を根拠に決めるべきだ。

そうでなければ2値のデータに変換してもいい。

アイデアがなければ1SDの変化が平均に認められれば。。

EPA. Benchmark Dose Technical Guidance, 2012
57

今日におけるBMR設定の問題点

サンプルサイズの決定手法はFormulateされていない

サンプルサイズ小さい=>BMDLが低くなりがち

最適サンプルサイズ=>用量反応曲線によって異なる

(特に低用量の反応によって)

Questions

1. BMRのためにMinimallyに必要なとされるサンプルサイズとは?(物質別;OECDテストガイドラインを踏まえつつ)

2. 動物実験において1群20頭などをどのDoseにどうやって割り振るのか

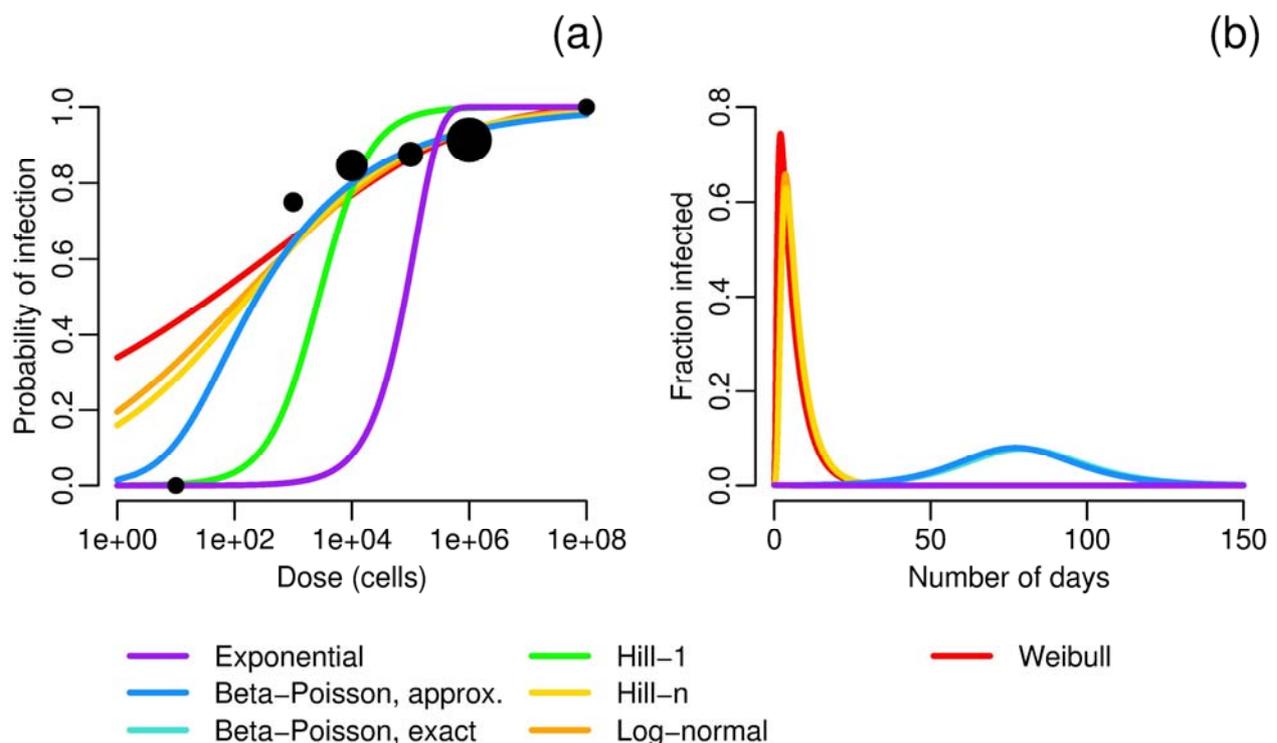
RestrictionのOn/Off設定について

Model	Number of model parameters	Model expression mean response (y) as function of dose (x)	Constraints
Quantal data			
Logistic	2	$y = 1/(1 + \exp(-a - bx))$	$b > 0$
Probit	2	$y = \text{CumNorm}(a + bx)$	$b > 0$
Log-logistic	3	$y = a + (1-a)/(1 + \exp(-\log(x/b)/c))$	$0 \leq a \leq 1, b > 0, c > 0$
Log-probit	3	$y = a + (1-a) \text{CumNorm}(\log(x/b)/c)$	$0 \leq a \leq 1, b > 0, c > 0$
Weibull	3	$y = a + (1-a) \exp((x/b)^c)$	$0 \leq a \leq 1, b > 0, c > 0$
Gamma	3	$y = a + (1-a) \text{CumGam}(bx^c)$	$0 \leq a \leq 1, b > 0, c > 0$
LMS (two-stage) model	3	$y = a + (1-a)(1 - \exp(-bx - \alpha^2))$	$a > 0, b > 0, c > 0$
Latent variable models (LVMs) based on the continuous models above ^(v)	Depends on underlying continuous model	These models assume an underlying continuous response, which is dichotomised into yes/no response based on a (latent) cut-off value that is estimated from the data	See continuous models

EFSA Journal. 2017;15(1):e04658

59

体育会系的に投げた例(コレラ)



パラメータ制約 (constraint) : 非現実的なフィットを避ける

- ・Primarily equality constraint, inequality constraint, integer constraint
- ・最適化プログラミングなどで詳しく研究
- ・数理的にはフィットが良いが明らかに「間違った」カーブで収束
- ・典型例:
 - Weibull with no restrictions on the power parameter
 - 100%未満のところまでPlateauに達してしまうモデル(元から対応していない)

EPA. Benchmark Dose Technical Guidance, 2012
61

Restriction ONの理由と議論

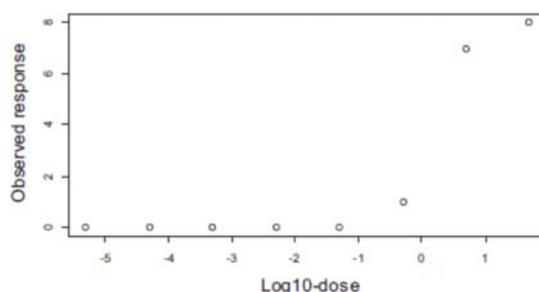
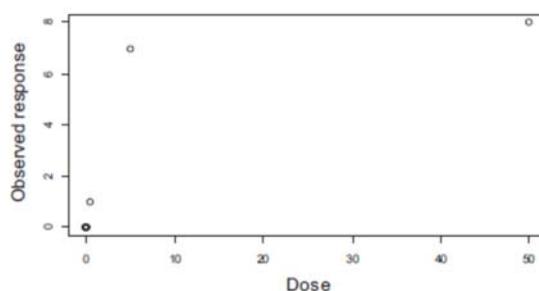
infinite slope at dose zeroを避けるべき

steepness parameterを1より大として対応してきた

必ずしも正しくない

(BMDLが人為的に高くなる)

EFSA Journal. 2017;15(1):e04658
Slob W and Setzer RW, 2014. Shape and steepness of toxicological dose-response relationships of continuous endpoints. Critical Reviews in Toxicology, 44, 270–297.



パラメータ制約 (constraint) に関して 西浦ラボでの大学院生Supervision

「Constraintがないと収束できない場合は、収束理由がキッチリ説明できるか、あるいは、誤りである」

=>わからないまま使ってはならない

(わからない場合は、ONをDefaultにして無批判に受けるよりOFFも同時に見る、という姿勢が必要)

再確認: BMD法にふさわしいデータとは

1. 観察データに量反応トレンド
2. イベントが見られる2点以上+コントロール1点が必須
3. イベントは低用量データが望ましい(高用量=>低用量予測ばかりしても、それはSpeculationにすぎない)