Analyzing Outcome Scores

Chuanpu Hu, PhD

Senior Scientific Director and Fellow

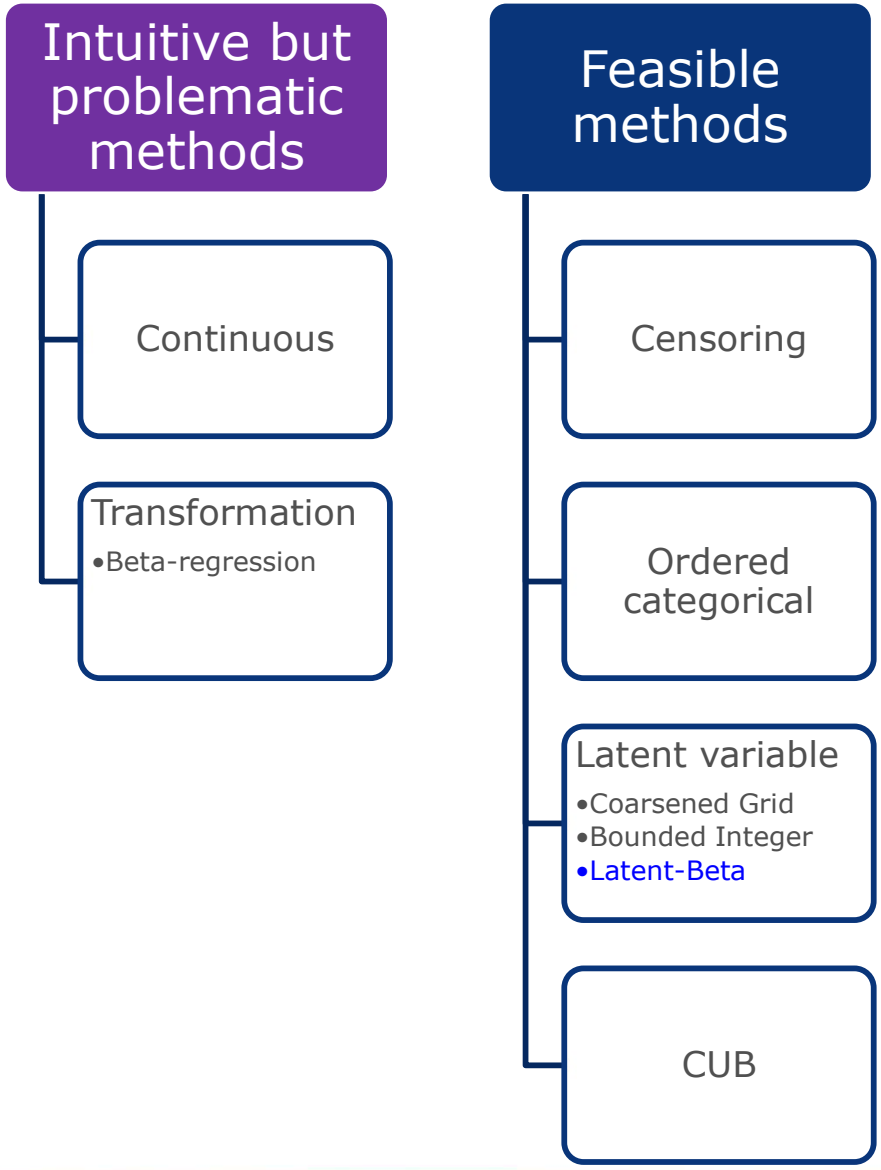
Clinical Pharmacology and Pharmacometrics

Janssen R&D, LLC

May 25, 2022

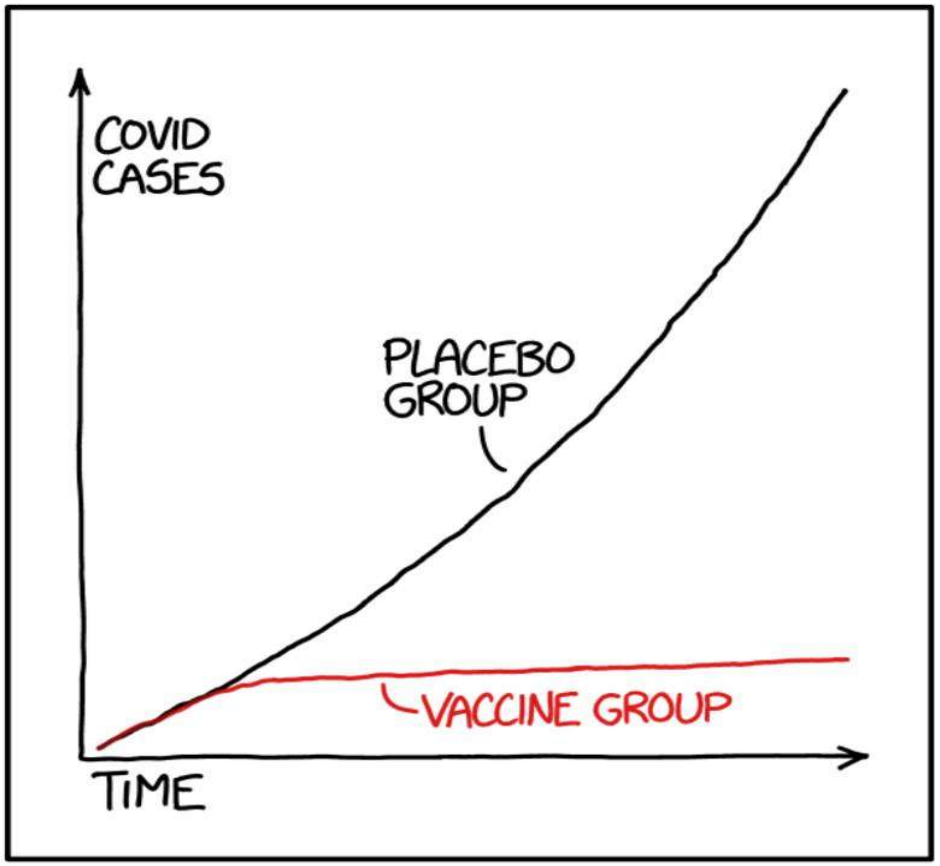
Outline

- (Bounded) Outcome Scores (BOS)
  - What, why should you care?
- Analysis methods
  - Ideas
  - Confusions
- Practical recommendations



Analyzing Data

- Often must make best use of data
  - Reduce bias
  - Increase efficiency
  - Smaller trials / better decisions
- A large class of data: (bounded) outcome scores
  - Emerging area in pharmacometrics, especially in past ~3 years



STATISTICS TIP: ALWAYS TRY TO GET DATA THAT'S GOOD ENOUGH THAT YOU DON'T NEED TO DO STATISTICS ON IT

Bounded Outcome Scores (BOS)

- Take restricted values within boundary
- Composite scores measuring disease severity
  - Used in many disease areas – immunology, neuroscience, etc.
  - Primary clinical trial endpoints, or used to derive them
- Example:
  - Psoriasis Activity Severity Index (PASI) score: 0 – 72, with 0.1 increments
- For notation, may standardize data as integers 0, 1, 2, …, n
  - Alternatively, onto closed interval [0,1]: for PASI score, [0, 1/721, 2/721, …, 1]
- Ordered categorical endpoints in nature (with many categories)

Derived Endpoints: Higher Bar for BOS

- Achieving a level, change from baseline, or combination of both (reaching certain threshold)
  - Often used as clinical trial primary endpoints
- Example:
  - Psoriasis: PASI 75/90/100: achieving 75, 90, or 100% improvement from baseline
    - Achieve PASI 100 $\Leftrightarrow$ PASI score = 0
- Model may describe often original scores but rarely derived endpoints
  - Describing derived endpoints requires that of the distribution of the original scores
    - Very difficult!

Common (Folklore?) Thinking

- If small # of categories (e.g., <6), analyze as ordered categorical
- "Intermediate" (e.g., >6 but <10) ??
- If "large" # of categories (e.g., >10), analyze as continuous
  - Problems:
    - Predicting data outside original range
    - Difficulty with skewed data distributions

BOS May Have Skewed Distributions

The AAPS Journal (2020) 22:61
DOI: 10.1208/s12248-020-00441-4

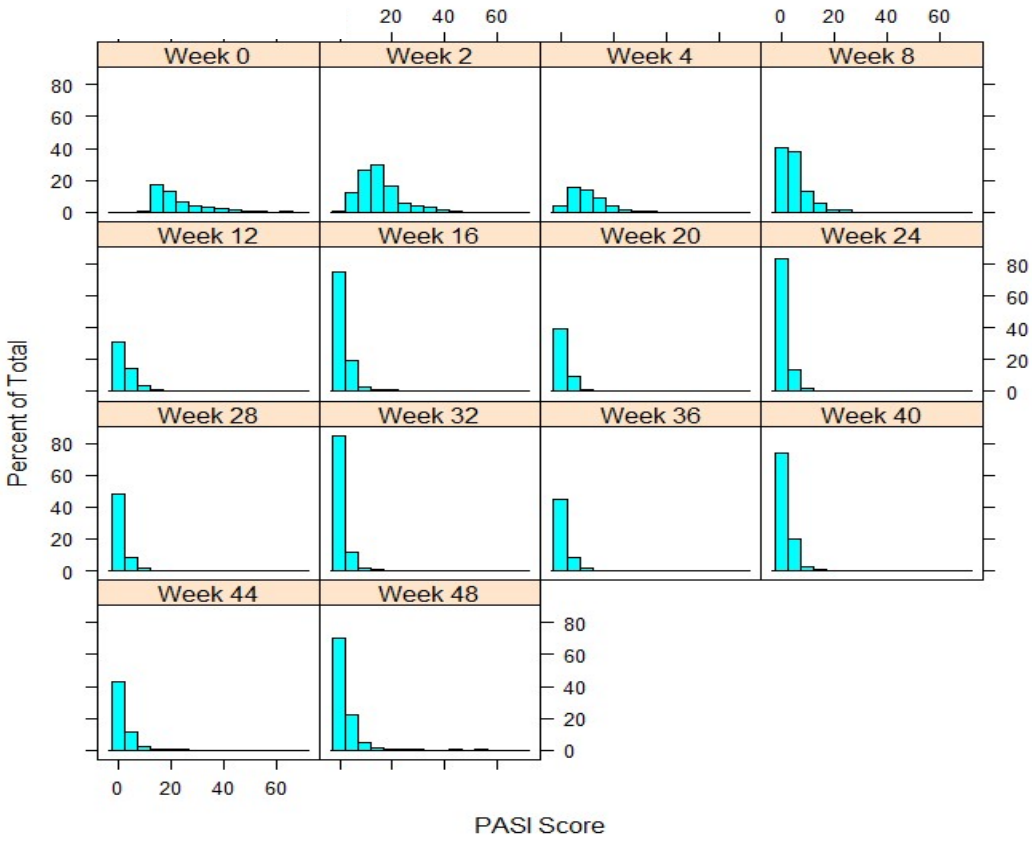
Check for updates

Research Article

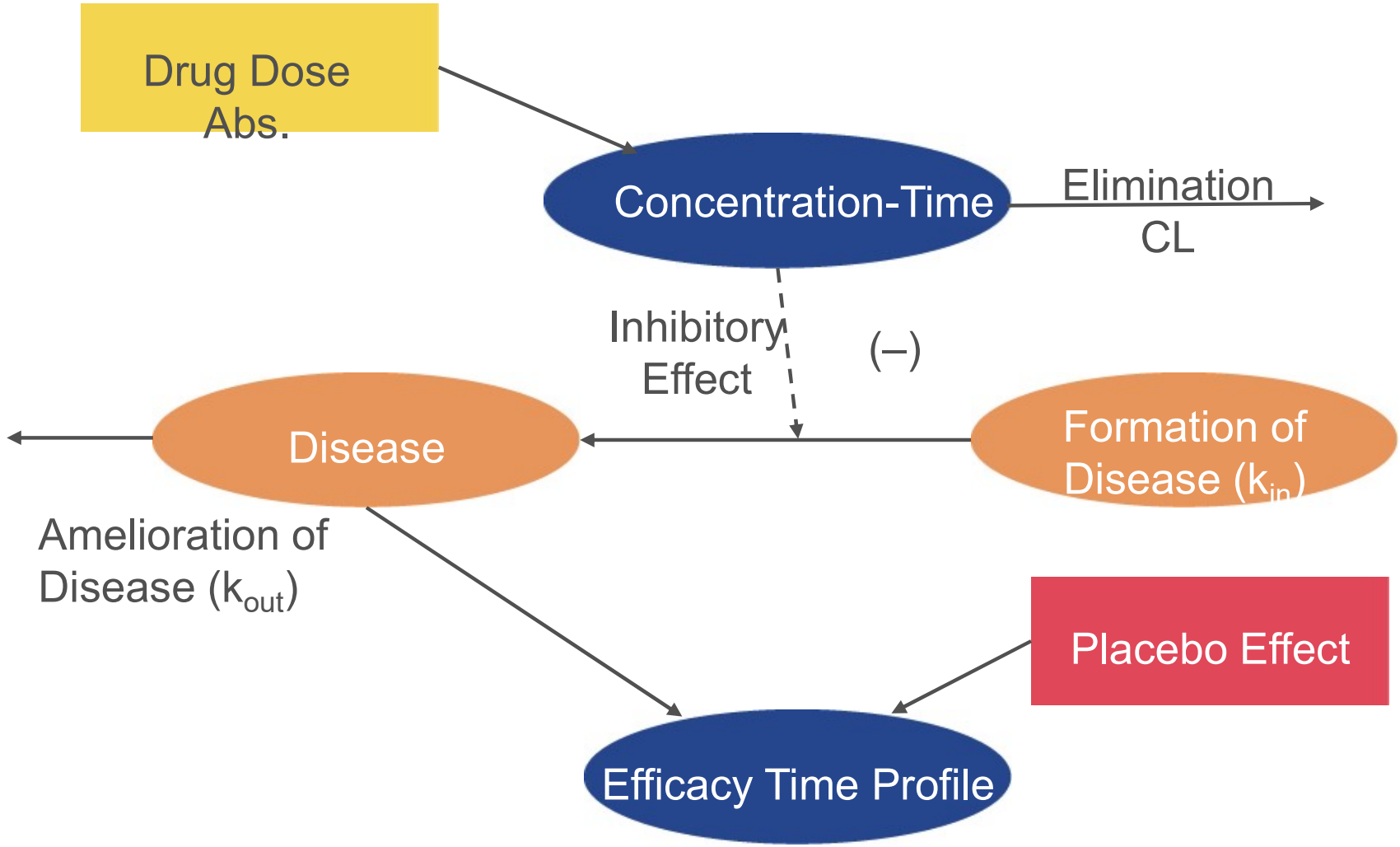
Applying Beta Distribution in Analyzing Bounded Outcome Score Data

Chuanpu Hu,[1,2] Honghui Zhou,[1] and Amarnath Sharma[1]

- Data Example: guselkumab Psoriasis
  - 2 Phase 3 studies, placebo-controlled, 48 weeks
- PASI score histograms at all 14 visits skewed to left
  - More PASI score = 0 over time



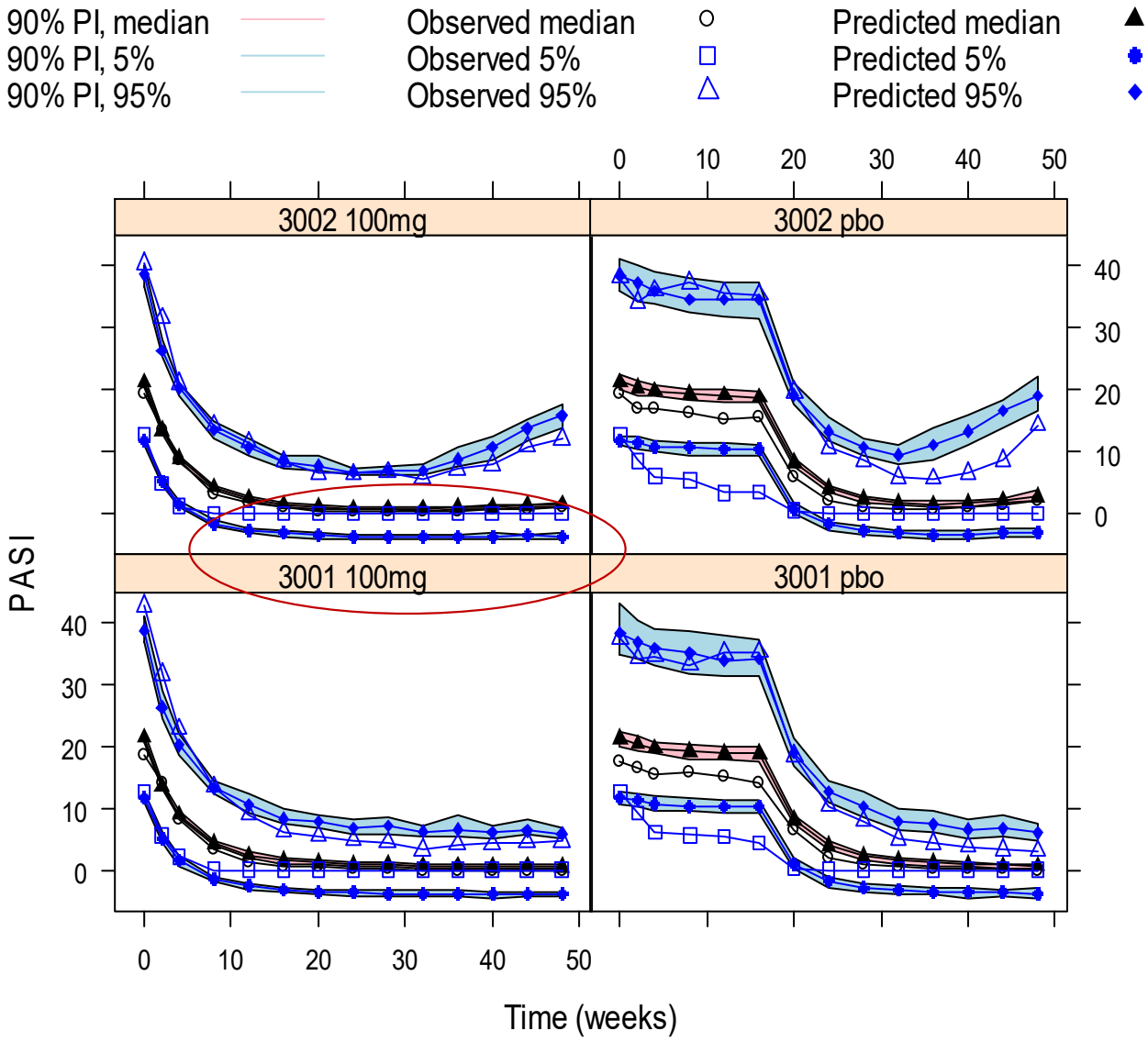
Longitudinal Model Diagram



- Sufficient PK: sequential PK/PD analysis fixing individual Posthoc PK parameter estimates

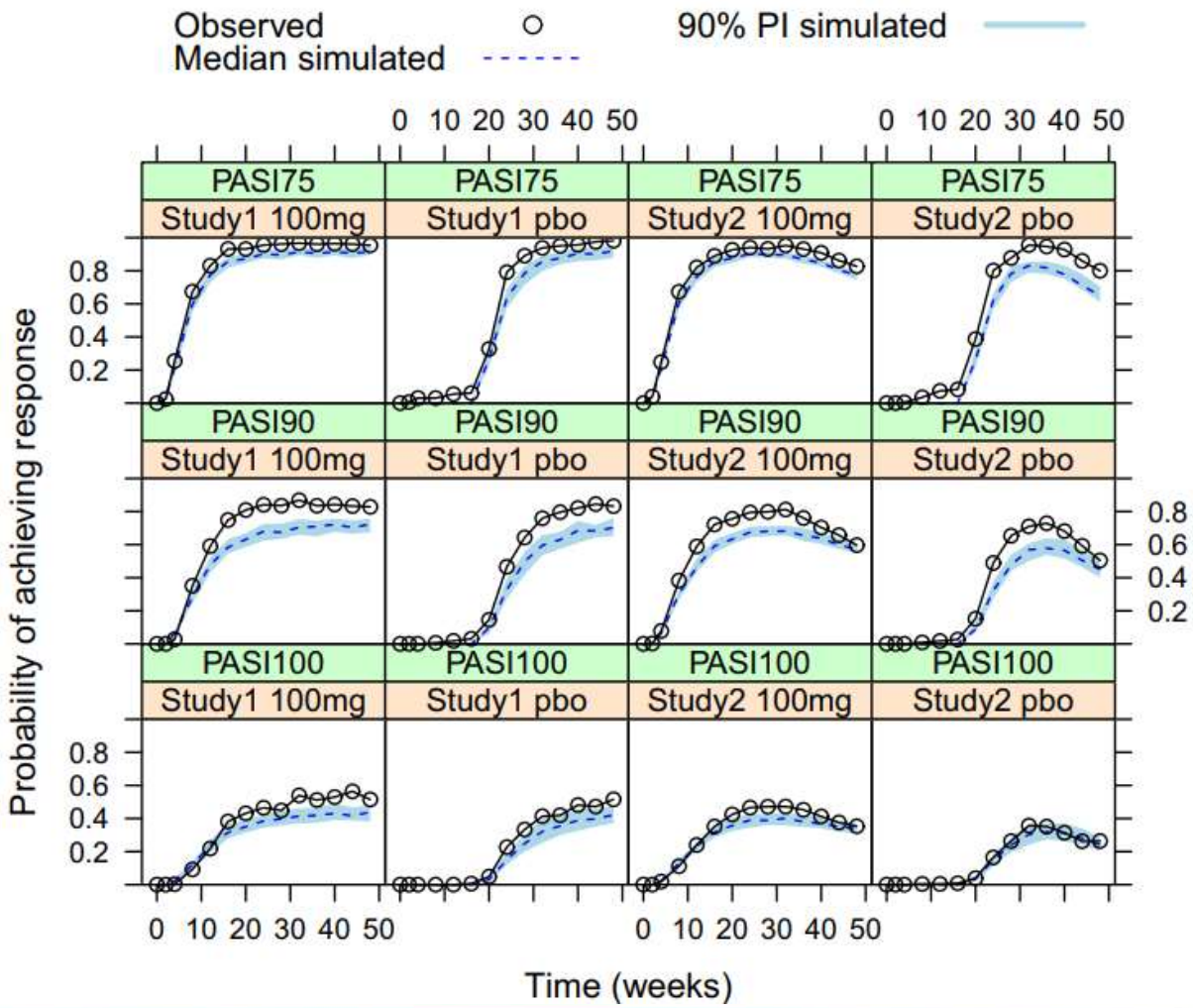
Model Data as Continuous: VPC of PASI Score Biased at Low End

- As treatment progresses:
  - Observed median and 5% percentiles both become near 0
    - Cannot be achieved by continuous model predictions
  - Model predicted median is fine, but 5% percentile outside data range (<0)



VPC of Derived Endpoints: PASI 75/90/100

- Notably underpredicting PASI 90, among others



More Biased VPC of A Derived Endpoint in a highly sensitive subpopulation

- Continuous model significantly underpredicted PASI 100 in a highly sensitive subpopulation
  - Achieved PASI 100 at Week 28, then taken off drug

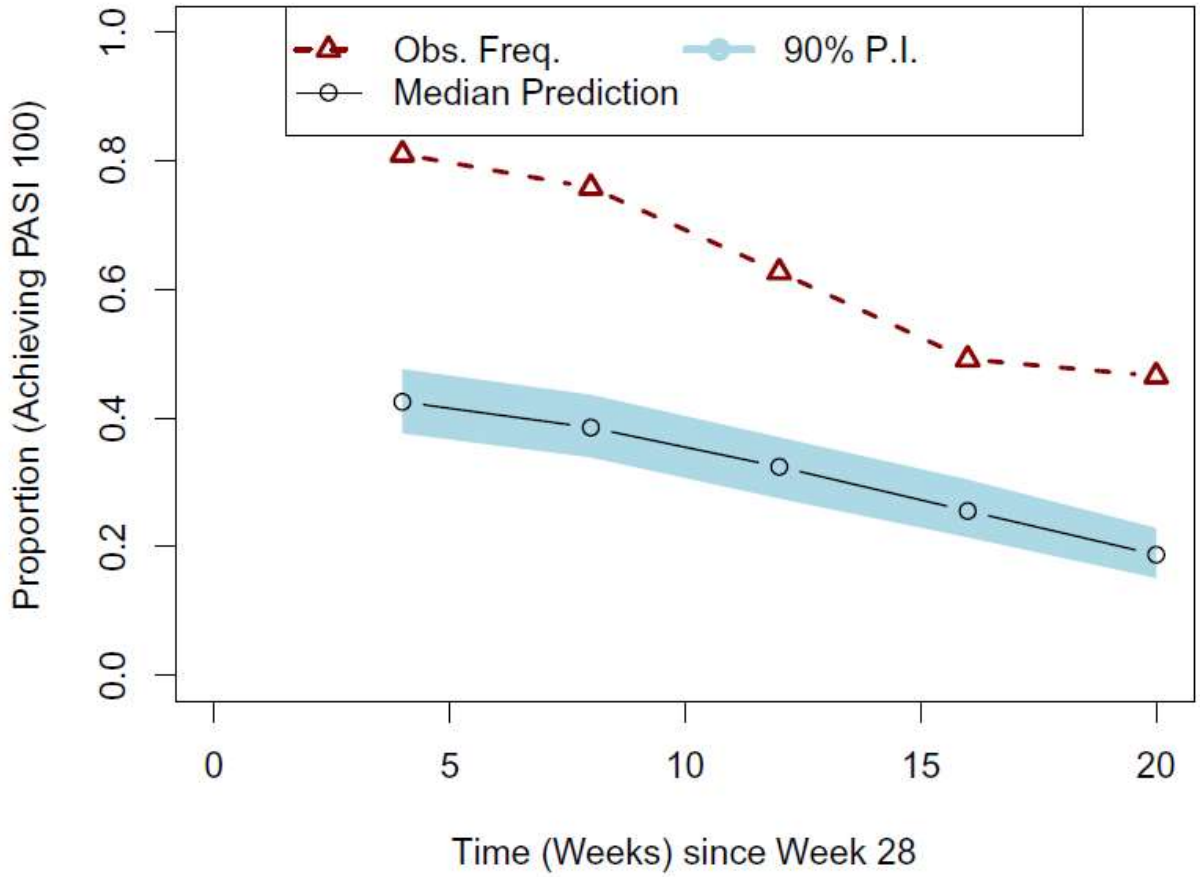


Fig. 3. The continuous analysis model predicted and observed PASI 100 responder rates when active treatment was withdrawn after week 28

Transformation and Related Approaches

- Common transformations, e.g., log or logit, cannot be directly applied
  - At boundary (0), log(0) = ∞

J Pharmacokinet Pharmacodyn (2017) 44:437–448
DOI 10.1007/s10928-017-9531-3

CrossMark

ORIGINAL PAPER

Improvement in latent variable indirect response modeling of multiple categorical clinical endpoints: application to modeling of guselkumab treatment effects in psoriatic patients

Chuanpu Hu[1] · Bruce Randazzo[2] · Amarnath Sharma[1] · Honghui Zhou[1]

- Proportional/(additive + proportional) error model Ill-behaved
  - At boundary (0), likelihood → ∞

Transform Data to within Boundary?
Beta-regression

Psychological Methods
2006, Vol. 11, No. 1, 54–71

1082-989X/06/$12.00 DOI: 10.1037/1082-989X.11.1.54

A Better Lemon Squeezer? Maximum-Likelihood Regression With Beta-Distributed Dependent Variables

Michael Smithson
The Australian National University

Jay Verkuilen
University of Illinois at Urbana–Champaign

- Originated from psychological literature
  - Spread to statistical literature (occasionally but not recently) and pharmacometrics
- Beta-distribution: on open interval (0,1), with density: $f(x) \sim x^a(1-x)^b$
- Linearly transform original score to (0,1)
  - (Left) boundary must be transformed to some value:
  - Transform data to $[\varepsilon, 1)$
  - Arbitrary fudge factor ε: often $\varepsilon = 0.01$
- Intuition: small change does not matter

Problem with Beta-regression

J Pharmacokinet Pharmacodyn (2017) 44:437–448
DOI 10.1007/s10928-017-9531-3

ORIGINAL PAPER

Improvement in latent variable indirect response modeling of multiple categorical clinical endpoints: application to modeling of guselkumab treatment effects in psoriatic patients

Chuanpu Hu[1] · Bruce Randazzo[2] · Amarnath Sharma[1] · Honghui Zhou[1]

- Statistically ill-behaved
  - Boundary observations become arbitrarily influential with smaller ε, e.g., 10^{-6}
- Small change does matter at highly-sensitive location!
  - "Butterfly effect"

Censoring

- Motivated from modeling BQL PK
- Separate data on/within boundary:
  - Model data within boundary as continuous, with transformations to handle data skewness, if needed
  - Model boundary data as censored, like "BQL"
  - Need additional "LLOQ" parameters for boundary data
- Removes previous problems due to achieving boundary

Research Article

Statistics in Medicine

Received 30 June 2009, Accepted 25 October 2010 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4155

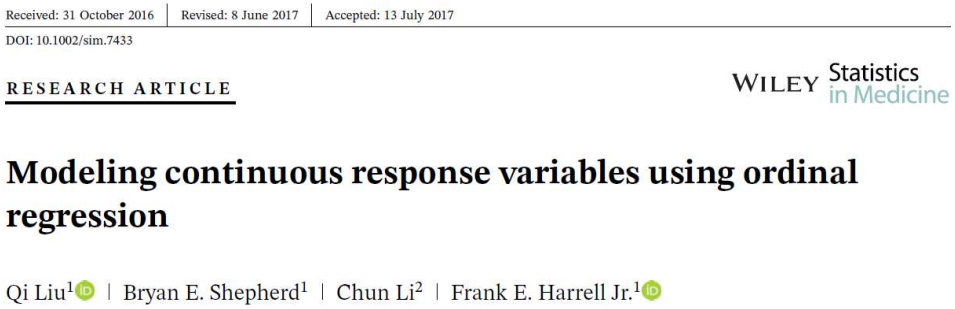
Estimating transformations for repeated measures modeling of continuous bounded outcome data

Matthew M. Hutmacher,[a*†] Jonathan L. French,[b] Sriram Krishnaswami[b] and Sujatha Menon[b]

- Esthetics: boundary and within-bounds data have the same nature; why should they be treated differently ("BQL" vs continuous)?

Ideally, Respect Data Nature

- BOS data are in fact ordered categorical
- Treating the data as such (using logit/probit regression) is superior, when can be done



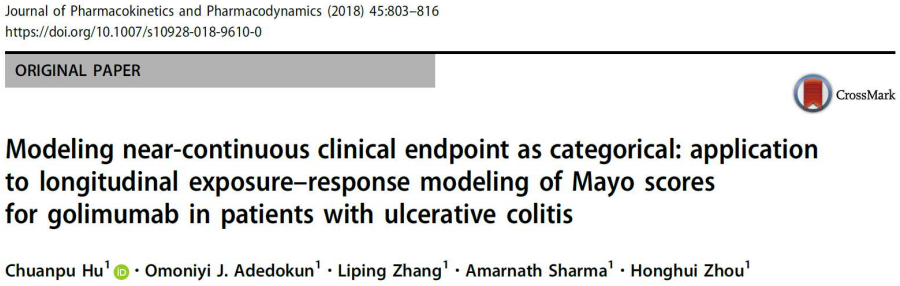
Received: 31 October 2016 | Revised: 8 June 2017 | Accepted: 13 July 2017
DOI: 10.1002/sim.7433

RESEARCH ARTICLE

WILEY Statistics in Medicine

Modeling continuous response variables using ordinal regression

Qi Liu[1] | Bryan E. Shepherd[1] | Chun Li[2] | Frank E. Harrell Jr.[1]



Journal of Pharmacokinetics and Pharmacodynamics (2018) 45:803–816
https://doi.org/10.1007/s10928-018-9610-0

ORIGINAL PAPER

Modeling near-continuous clinical endpoint as categorical: application to longitudinal exposure–response modeling of Mayo scores for golimumab in patients with ulcerative colitis

Chuanpu Hu[1] · Omoniyi J. Adedokun[1] · Liping Zhang[1] · Amarnath Sharma[1] · Honghui Zhou[1]

- If too many intercepts to estimate: can they be "fixed"?

Recent Advances: Latent Variable Approaches

- Most natural
- Underlies standard categorical data analysis approaches of logit/probit regression
- Observed data occurs when underlying latent variable crosses certain thresholds
- The thresholds correspond to intercepts
- Idea:
  - Fix the thresholds "naturally"
  - Model latent variable with flexible distributions to handle skewness
    - Instead of transforming the original score

Approach 1: Coarsened Grid

Biostatistics (2007), **8**, 1, *pp.* 72–85
doi:10.1093/biostatistics/kxj034
Advance Access publication on April 5, 2006

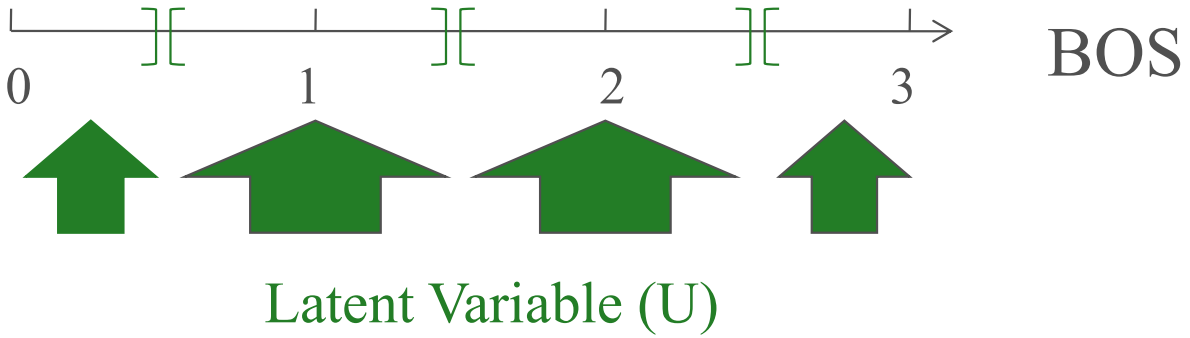
The logistic transform for bounded outcome scores

EMMANUEL LESAFFRE*, DIMITRIS RIZOPOULOS, ROULA TSONAKA

Biostatistical Centre, Catholic University of Leuven,
U.Z. St. Rafaël, Kapucijnenvoer 35, B–3000 Leuven, Belgium
emmanuel.lesaffre@med.kuleuven.be

- Established BOS concept

Coarsened Grid Illustration

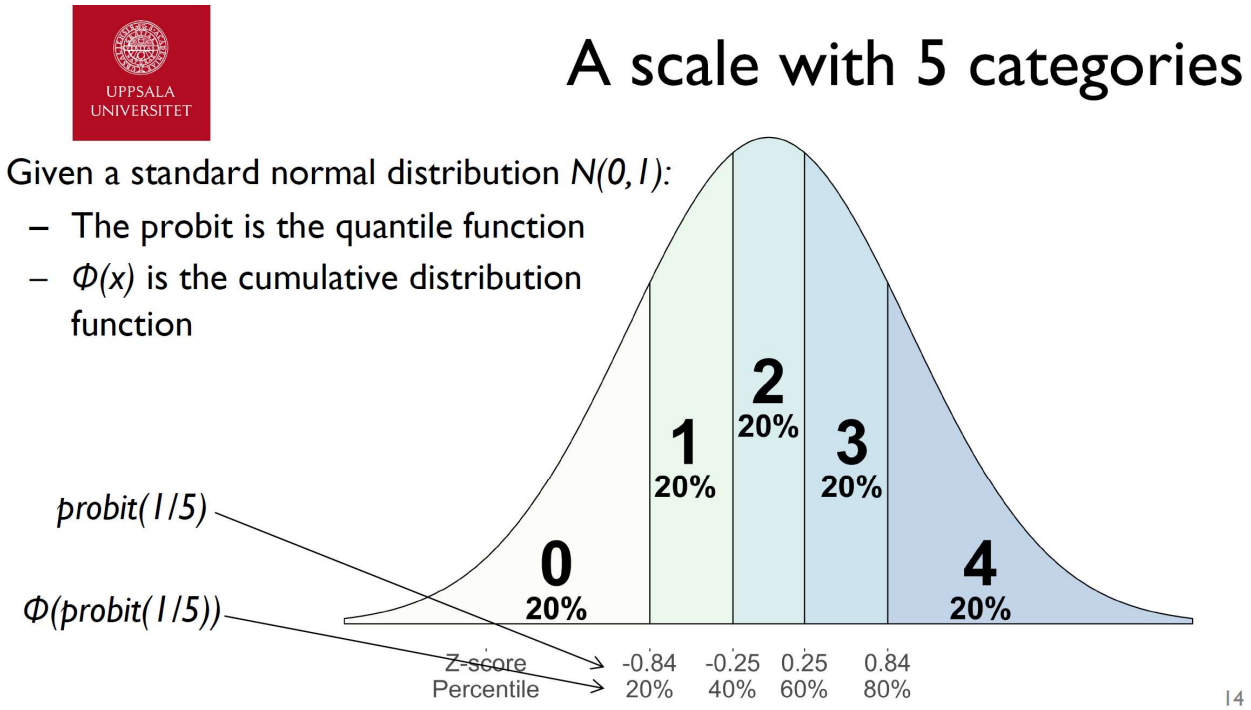


- Motivation: BOS value k = 0,1, ...m occurred by rounding of continuous latent variable U on interval (0, m) to integer
  - Latent variable thresholds fixed at middle points: 0.5, 1.5, etc.
- Scale U to (0,1), then model with logit-normal distribution
  - logit(U) = pred + ε, with $\varepsilon \sim N(0, \sigma^2)$

Approach 2: Bounded Integer

- Motivated differently: Wellhagan et al, PAGE 2018

- Split a normal distribution
  - To # of intervals as BOS categories
  - With equal probability
- Model mean and sd



14

- Should be close to Coarsened Grid, especially if # of category is large

More on Coarsened Grid and Bounded Integer

The AAPS Journal (2019) 21:74
DOI: 10.1208/s12248-019-0343-9

Check for updates

Research Article

A Bounded Integer Model for Rating and Composite Scale Data

Gustaf J. Wellhagen,[1] Maria C. Kjellsson,[1] and Mats O. Karlsson[1,2]

The AAPS Journal (2019) 21:102
DOI: 10.1208/s12248-019-0370-6

Check for updates

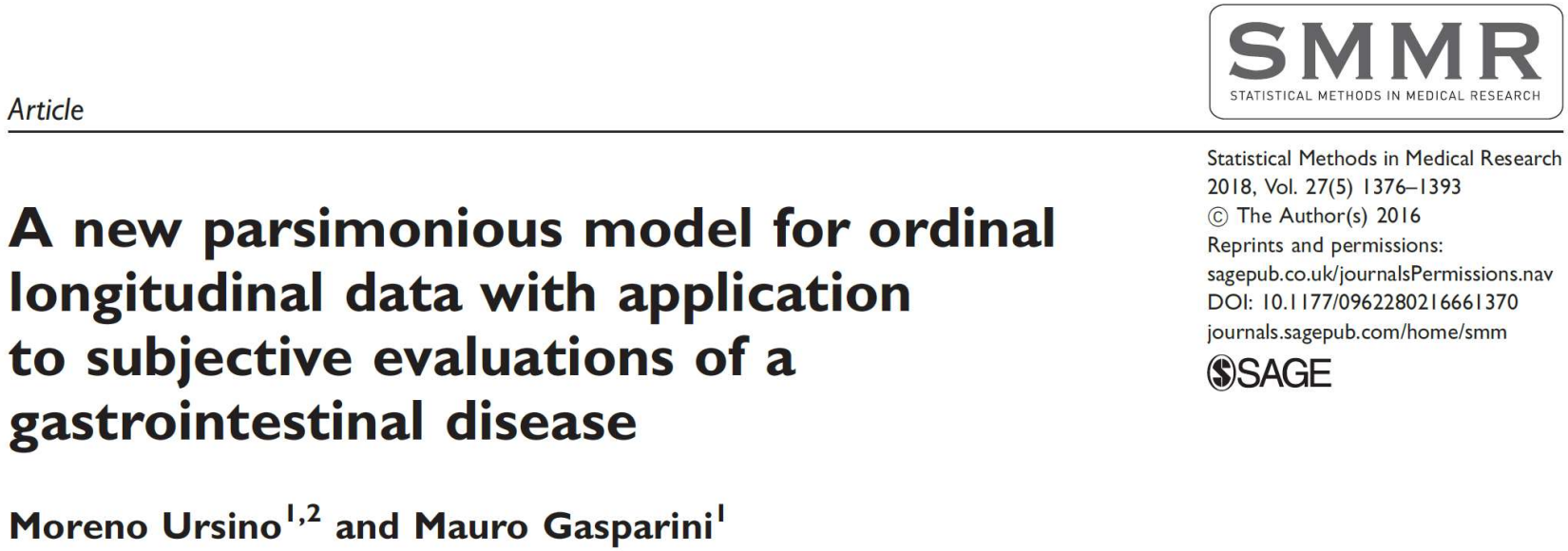
Commentary

On the Comparison of Methods in Analyzing Bounded Outcome Score Data

Chuanpu Hu[1,2]

- Both approaches may have difficulties with skewness with underlying normal distribution

Approach 3: Latent-beta



Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
2018, Vol. 27(5) 1376–1393

DOI: 10.1177/0962280216661370
journals.sagepub.com/home/smm
SAGE

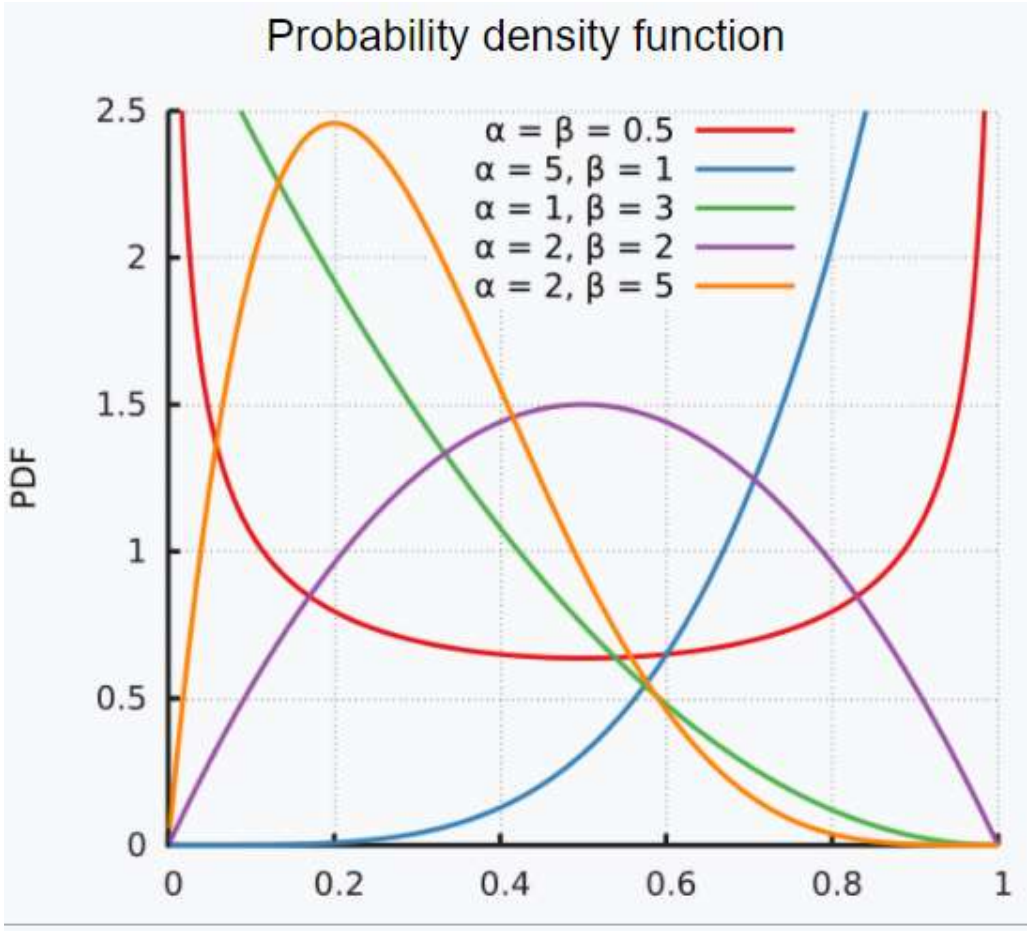
A new parsimonious model for ordinal longitudinal data with application to subjective evaluations of a gastrointestinal disease

Moreno Ursino[1,2] and Mauro Gasparini[1]

- The newest and brightest

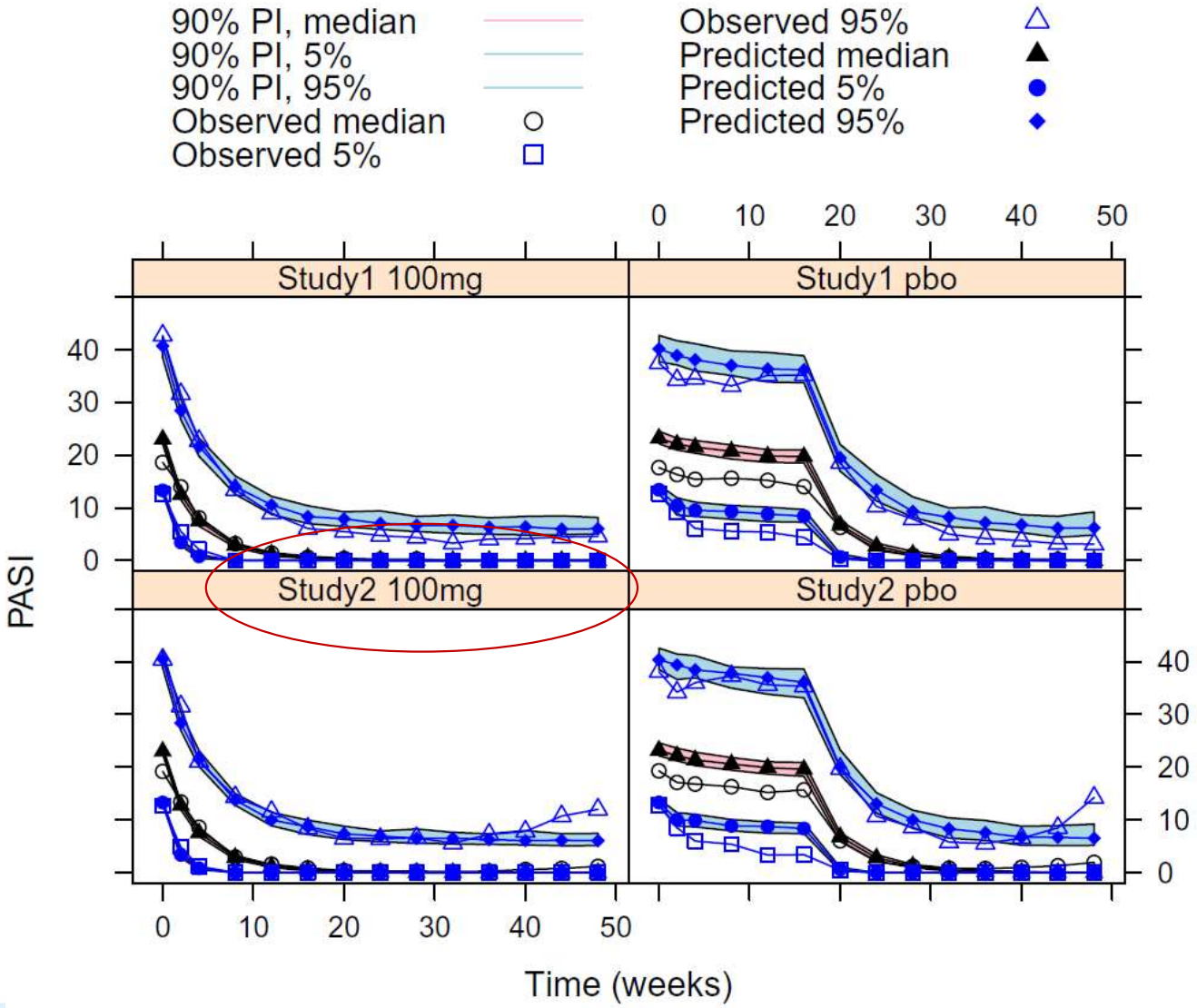
Latent-beta

- Use beta distribution instead of (logit)normal
- Wikipedia:
  - Beta distribution shapes
- High flexibility to model skewed distributions
- Resolves all previous issues with other approaches



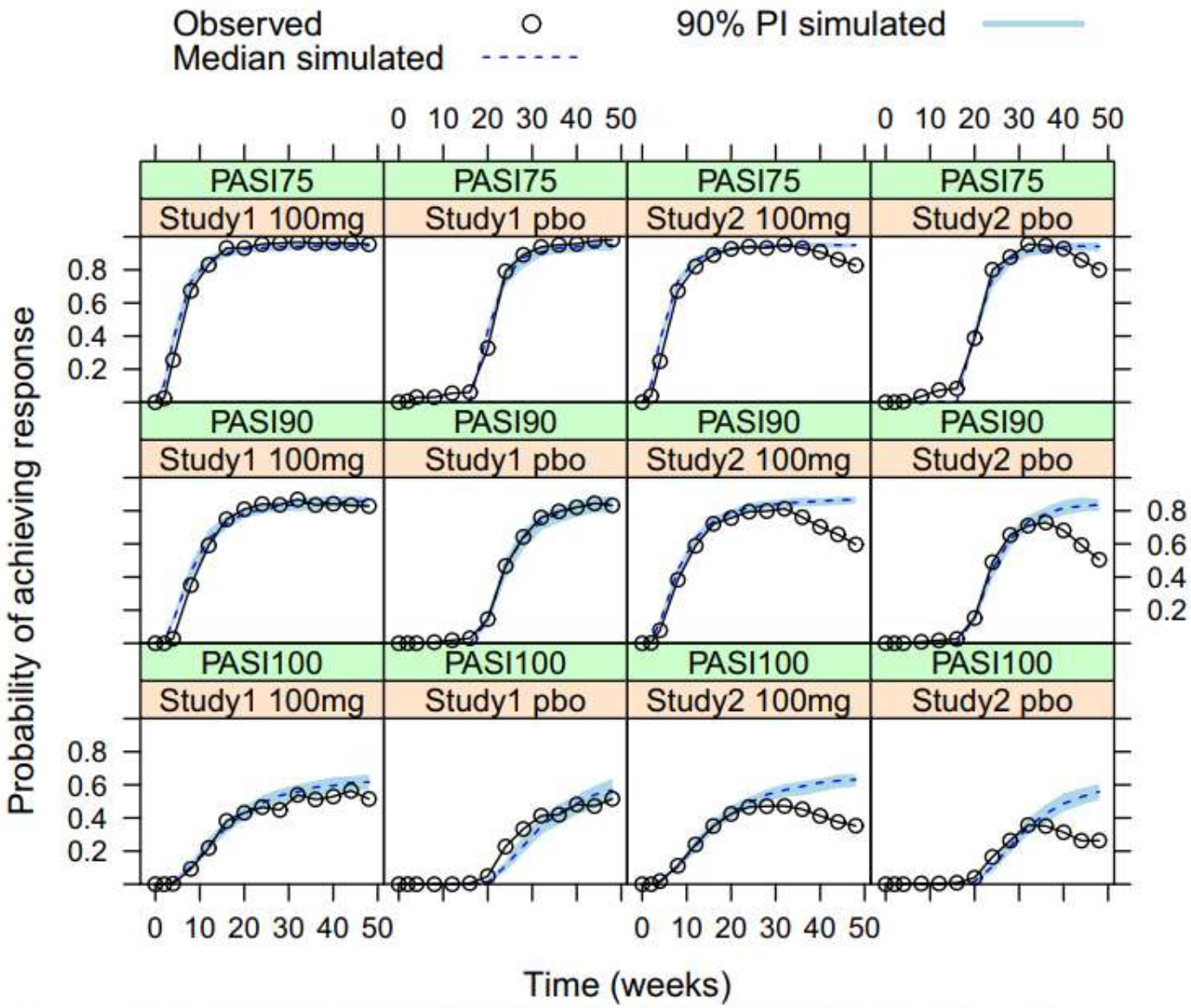
Applying Latent-beta to Guselkumab Psoriasis

- Model reasonably predicted observed PASI scores
  - Minor problem remained at baseline median
    - Baseline distribution appeared more skewed
  - Reasonable prediction of treatment effect



Latent-beta VPC of Derived Endpoints: PASI 75/90/100

- Model reasonably described data
- Later time points of Study 2 are nuanced due to treatment optimization
- Much improved from the continuous approach



Reasonable VPC of Derived Endpoint

- Latent-beta model reasonably predicted PASI 100 in a highly sensitive subpopulation
  - Achieved PASI 100 at Week 28, then taken off drug
- Much improved from the continuous model

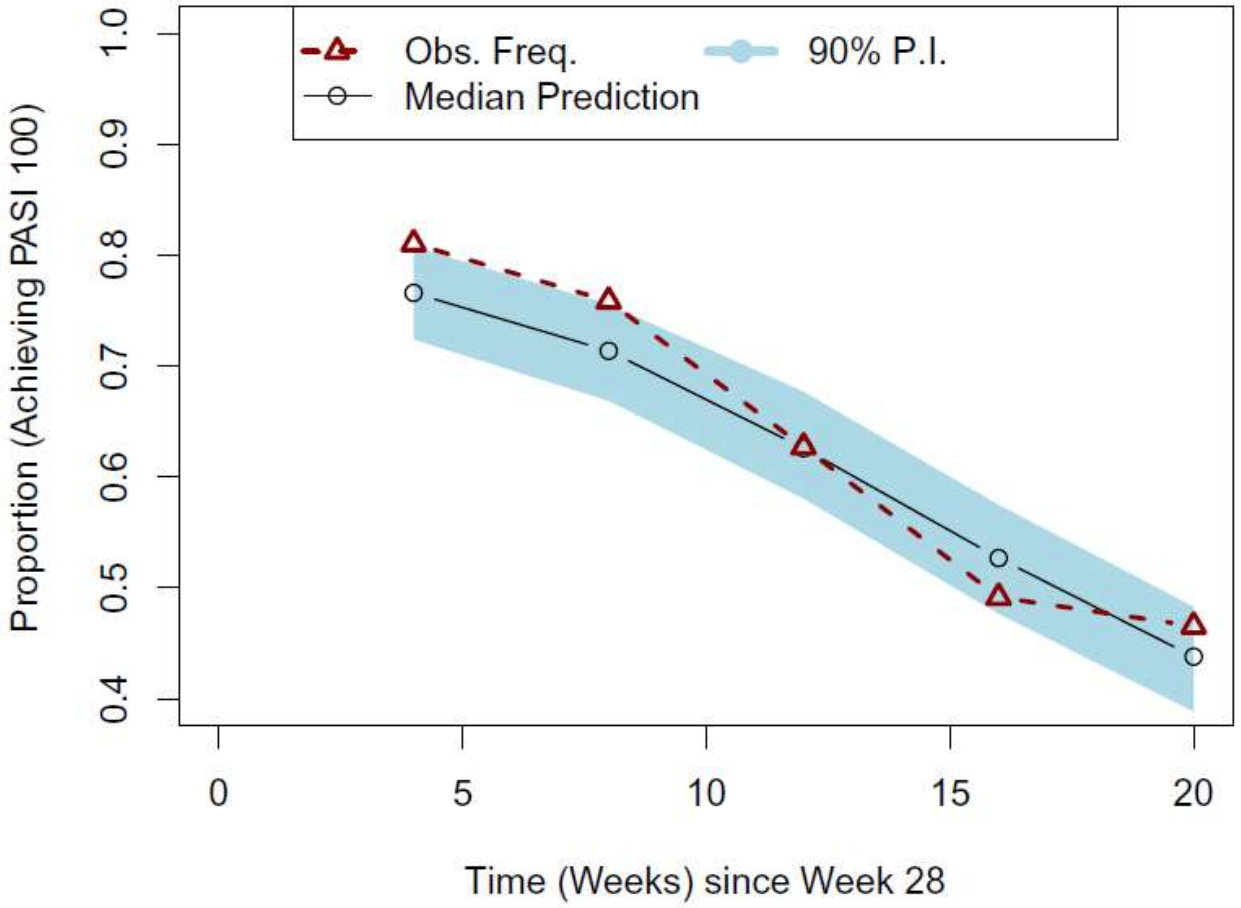


Fig. 6. The BOS analysis model predicted and observed PASI 100 responder rates when active treatment was withdrawn after week 28

Complexities of Describing Derived Endpoints in Skewed Distributions:

- Brief History of PASI Score Modeling
  - Few publications pre 2020 – suggesting its difficulty
- Unpublished (Hutmacher ~2016): Censoring described PASI scores and PASI 75, but biased in PASI 90/100
- Success in describing PASI scores and PASI 75/90/100 not achieved until 2020, with Latent-beta
  - Even in a highly sensitive subpopulation!
- 2021: latent-beta success in PASI score and PASI 75/90/100 repeated, and another method applied
- 2022: latent-beta success repeated in a highly sensitive subpopulation

Complexities of Describing Reponses in Highly Sensitive Subpopulations

- Response-adaptive study designs
  - Induction-maintenance paradigm:
    - Responders to initial treatment re-randomized to different "maintenance" treatments
- Few successes in modeling "maintenance" data

Journal of Pharmacokinetics and Pharmacodynamics (2022) 49:283–291
https://doi.org/10.1007/s10928-021-09796-3

ORIGINAL PAPER

Improving categorical endpoint longitudinal exposure–response modeling through the joint modeling with a related endpoint

Chuanpu Hu[1] · Honghui Zhou[1]

- Requires getting between-subject variabilities right

Another Family of Models: CUB

International Statistical Review (2018), 0, 0, 1–30 doi:10.1111/insr.12282

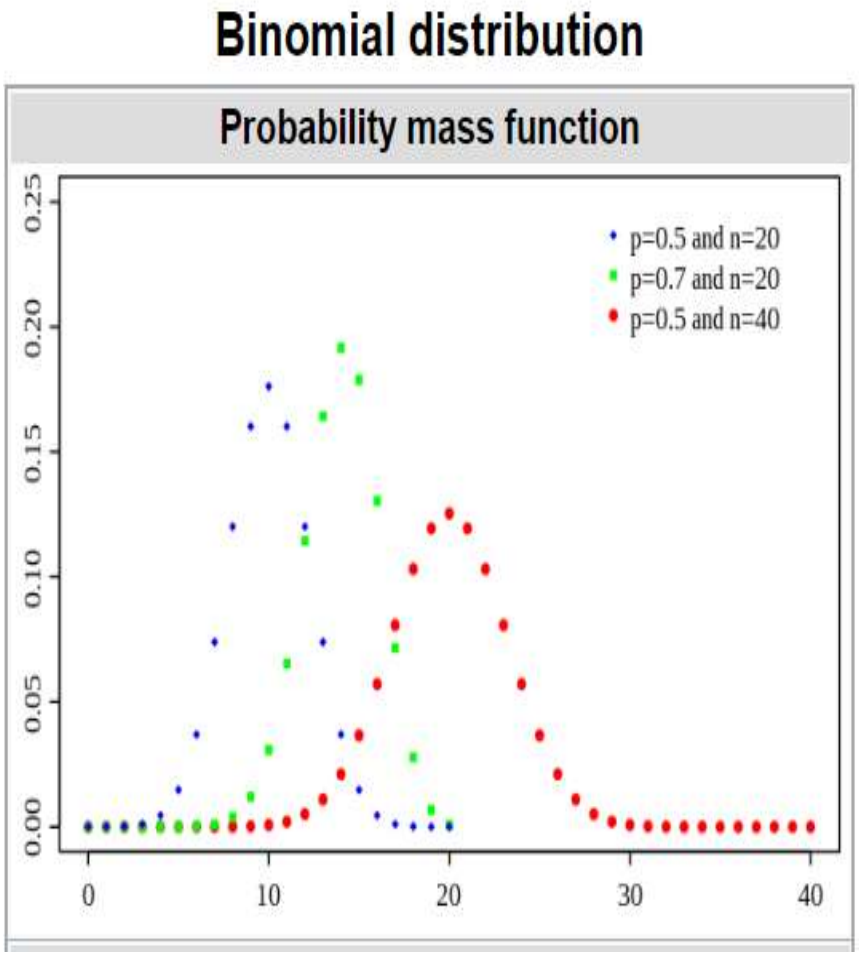
Cumulative and CUB Models for Rating Data: A Comparative Analysis

Domenico Piccolo, Rosaria Simone and Maria Iannario

- Popular in psychological rating data analysis
- Motivated from the Binomial distribution
  - Not a latent variable approach
  - But still respect data nature

CUB: Details

- Binomial distribution with total level = n:
  - Prob(score = k) = $\binom{n}{k} p^k (1-p)^{n-k}$
  - Distribution skewed (left, right) when p (<,>) 0.5
- Random noise, i.e., Uniform:
  - Prob(score = k) = 1/n
- Combine the distributions, with mixture probability π:
  - Combined Uniform Binomial (CUB)
    - Prob(score = k) = $\pi \binom{n}{k} p^k (1-p)^{n-k}$ + (1- π) /n



Latent-beta vs CUB

The AAPS Journal (2020) 22:95
DOI: 10.1208/s12248-020-00478-5

Check for updates

Research Article

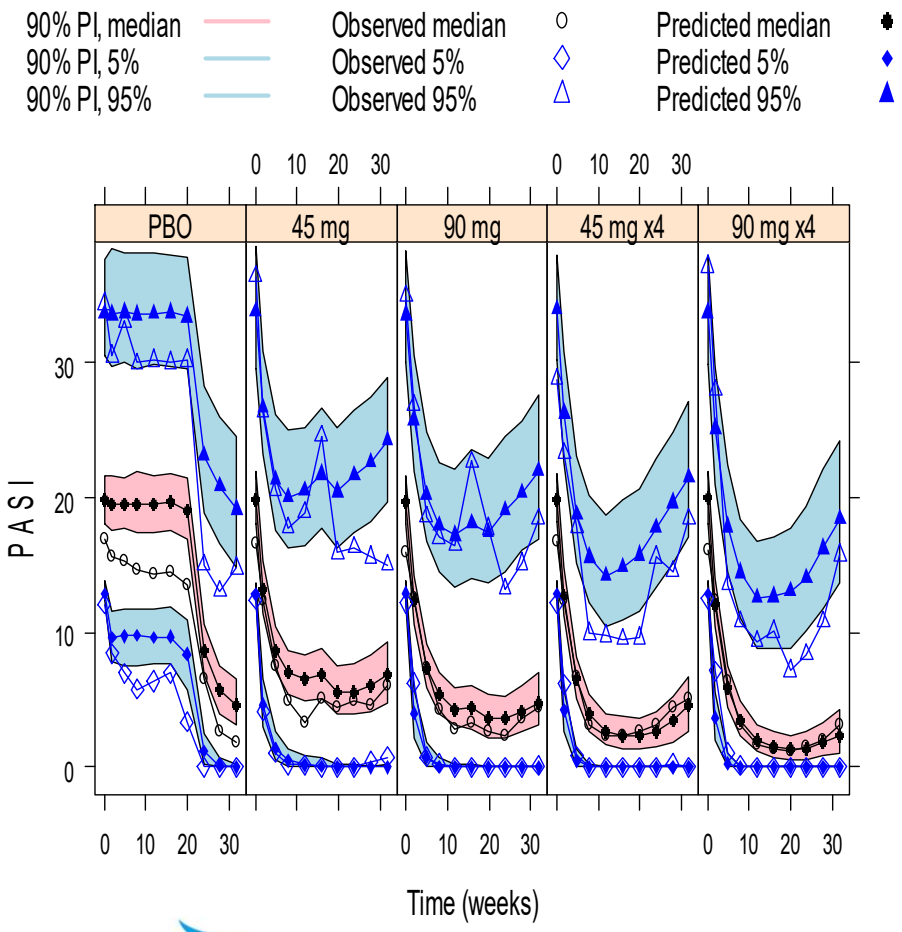
Application of Beta-Distribution and Combined Uniform and Binomial Methods in Longitudinal Modeling of Bounded Outcome Score Data

Chuanpu Hu,[1,2] Honghui Zhou,[1] and Amarnath Sharma[1]

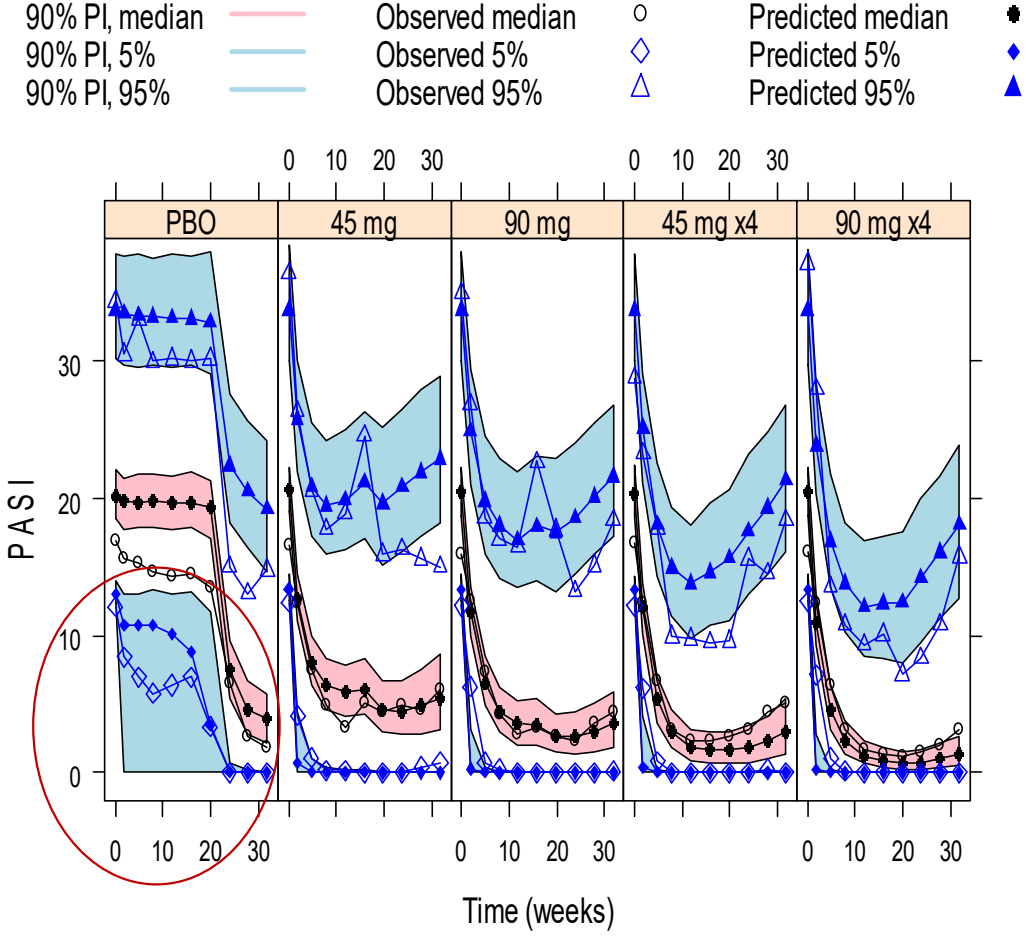
- Data: ustekinumab psoriasis Phase 2
- 320 patients randomized to receive SC injection in 5 arms:
  - PBO (till Week 20), 45 mg, 90 mg, (45 mg weekly x4), (90 mg weekly x4)
- Data collected ~q4w during Weeks 0-32

PASI Score VPC: CUB Somewhat Nuanced

- Latent-beta

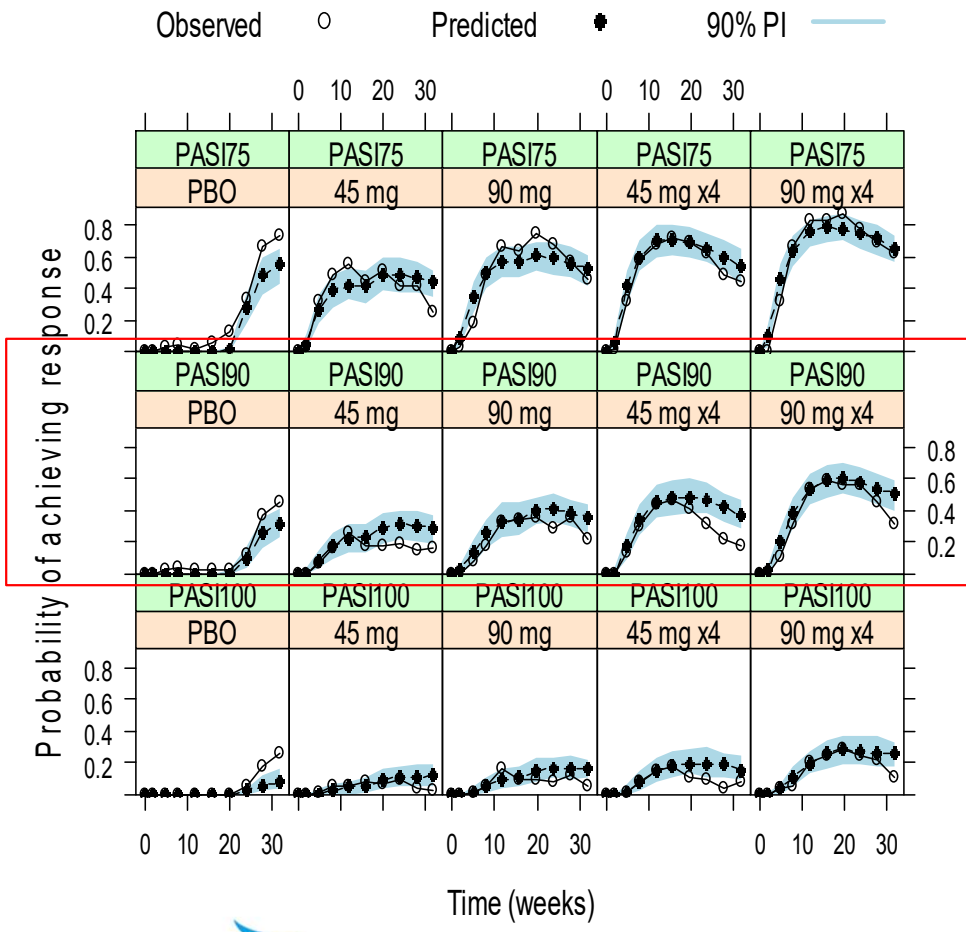


- CUB

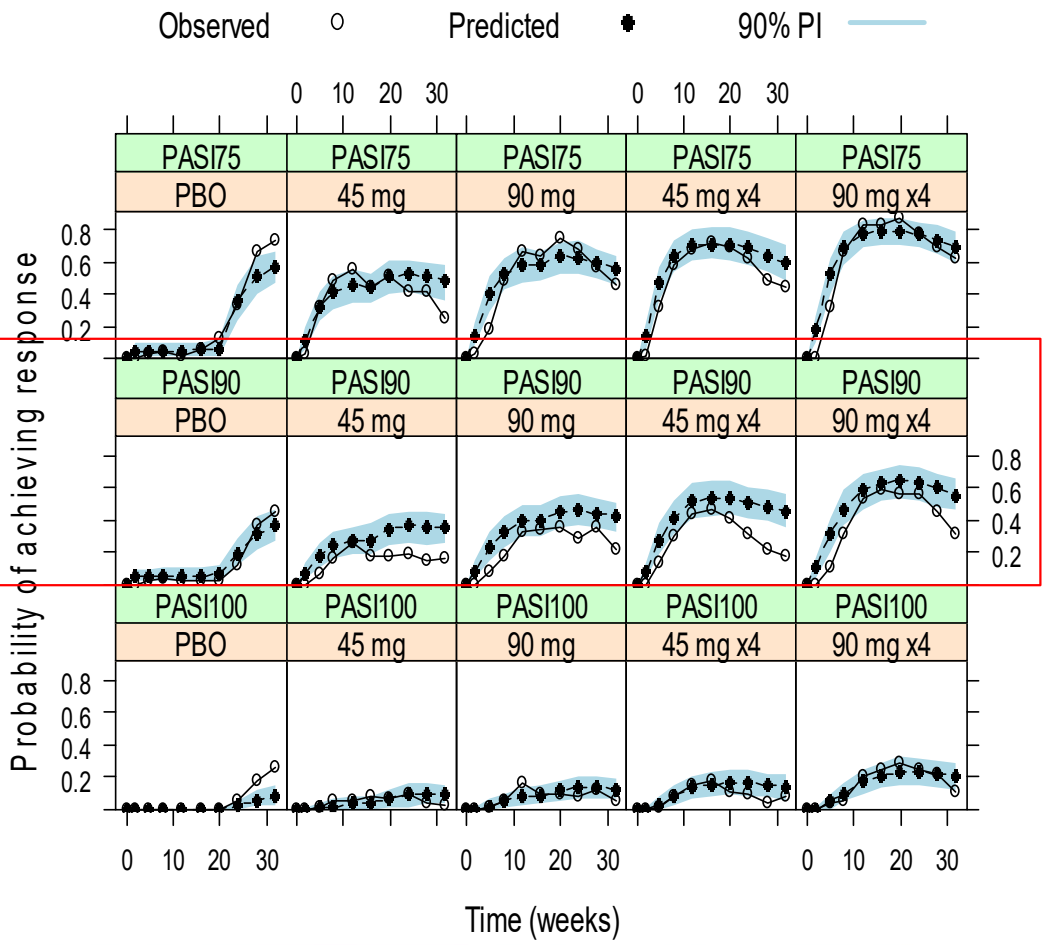


PASI 75/90/100 VPC: Latent-beta Somewhat better in PASI 90

- Latent-beta



- CUB



More Details

- Overall, VPC of latent-beta slightly better than CUB
  - In both PASI scores and PASI 75/90/100
- Latent-beta also has much better NONMEM OFV
  - Improvement >400 over CUB
- Uniform distribution in CUB too noisy?
  - May need future verification
- NONMEM implementation of latent-beta:

The AAPS Journal (2020) 22:61
DOI: 10.1208/s12248-020-00441-4

Check for updates

Research Article

Applying Beta Distribution in Analyzing Bounded Outcome Score Data

Chuanpu Hu,[1,2] Honghui Zhou,[1] and Amarnath Sharma[1]

Model Parameters - Complexity

- Continuous: (mean, sd)
- Censoring: (mean, sd, censoring limits at boundary) – 2 extra parameters
- Ordered categorical: (intercepts) – many parameters!
- Coarsened Grid: (mean, sd)
- Bounded Integer: (mean, sd)
- Latent-beta: (mean, precision)
- CUB: (p, π)

Which parameter to model?

- Modeling is typically done on the mean parameter
  - As function of dose/exposure, etc.
- How about the variance/precision parameter, e.g., should it be modeled as a function of the mean?
  - Like "proportional error" in pharmacokinetics
  - Used in BOS literature, though no clear evidence of need
- To avoid overfitting, likely best to keep variance/precision as a constant parameter for BOS, unless clear reasons supporting otherwise

Comparing the Methods: Confusion with AIC for Pharmacometricians

The AAPS Journal (2019) 21:102
DOI: 10.1208/s12248-019-0370-6

Check for updates

Commentary

On the Comparison of Methods in Analyzing Bounded Outcome Score Data

Chuanpu Hu[1,2]

- AIC/BIC cannot be used to compare Continuous with Censoring or categorical approaches
  - "Likelihood" not comparable with changed data
  - Same when **treating data differently** (numerical vs. categorical)!
    - Category levels have no numerical meaning: cannot calculate "Low" + "Mild"
    - Read Akaike (1974)
  - Confusion in pharmacometrics literature even to-date

Appropriate Method Comparisons

- Use AIC/BIC to compare only categorical approaches
  - i.e., latent variable approaches, and CUB
- To compare approaches treating data differently, e.g., continuous vs. categorical: Use VPC
  - In abstract:
    - Continuous scale will favor the continuous approach
    - Categorical scale (proportion of achieving category) will favor the Categorical approach
  - (Another indication that Continuous/Categorical approaches are not formally comparable)
  - Choose the quantity/scale of practical interest

Summary: Which Method to Use, When?

- Use Ordered Categorical when possible, even if >10 categories
- If not (i.e., too many intercepts to estimate):
  - OK to use Continuous, if
    - Symmetric data
    - Tight timelines
  - OK to use Censoring, if
    - Skewed data
    - Do not care about esthetics, or predicting outside data range
  - Can use Coarsened Grid / Bounded Integer, if
    - Near-symmetric data
  - (Should?) use Latent-beta
    - The only method shown to describe derived endpoints, in a highly sensitive subpopulation
  - Might consider CUB

References: See within

The AAPS Journal (2019) 21:102
DOI: 10.1208/s12248-019-0370-6

Check for updates

Commentary

On the Comparison of Methods in Analyzing Bounded Outcome Score Data

Chuanpu Hu[1,2]

The AAPS Journal (2020) 22:95
DOI: 10.1208/s12248-020-00478-5

Check for updates

Research Article

Application of Beta-Distribution and Combined Uniform and Binomial Methods in Longitudinal Modeling of Bounded Outcome Score Data

Chuanpu Hu,[1,2] **Honghui Zhou,**[1] **and Amarnath Sharma**[1]