Simple Approach to Calculate Random Effects Model Tolerance Intervals to Set Release Limits and Shelf-life Specification Limits of Pharmaceutical Products

Richard O. Montes (Hospira, a Pfizer company)*

Richard K. Burdick (Elion Labs)

David J. LeBlond (CMC Stats)

Joint Research Conference
Santa Fe, New Mexico
13 Jun 2018

Pfizer ESSENTIAL HEALTH

Presentation Outline

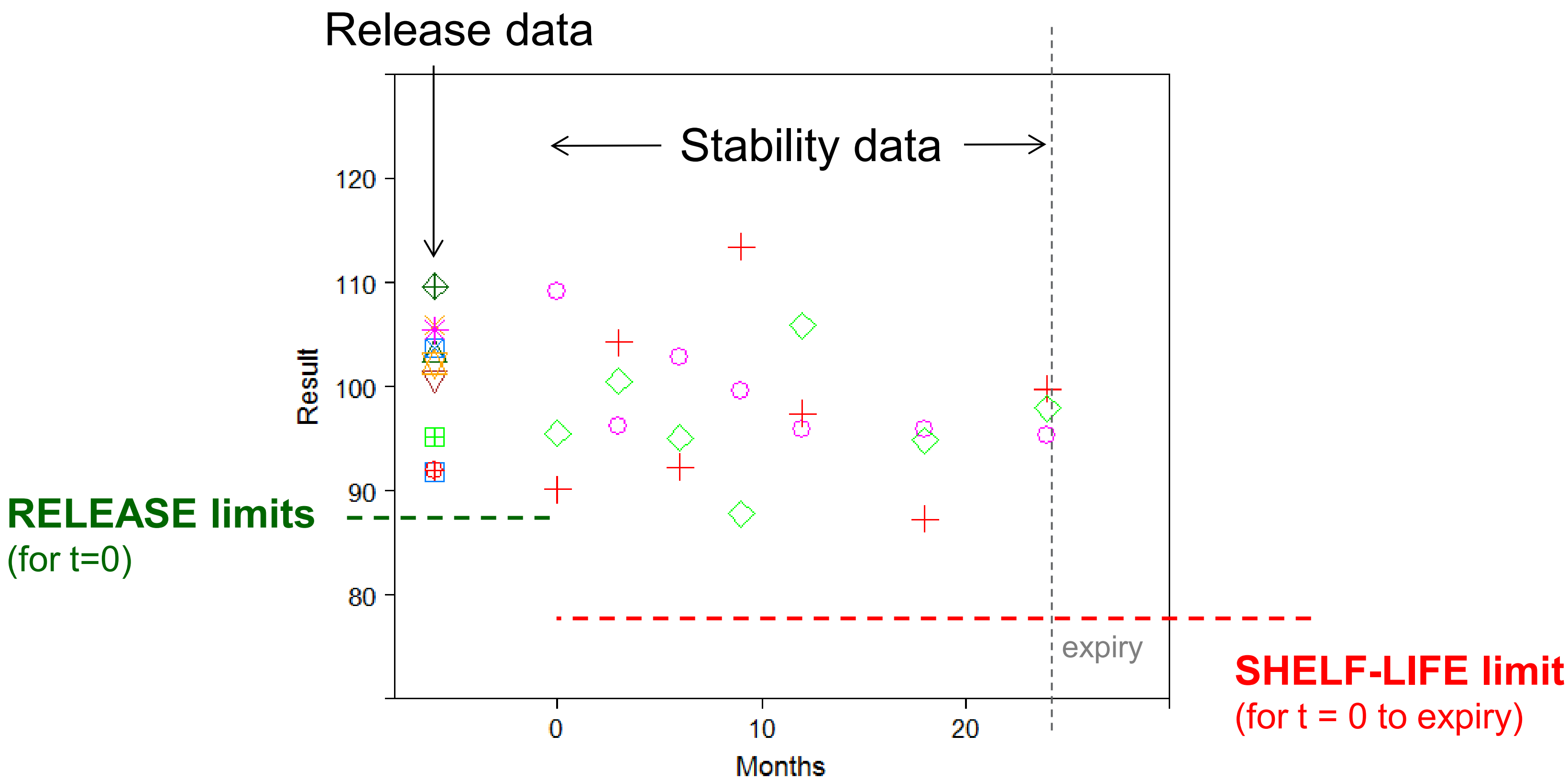
- What are specification limits
  - Levels: Release and Shelf-life
  - Justification of Specifications
  - Survey of current statistical methods
- Study objective
  - Propose a tolerance interval method for random effects model that can be implemented by applied statisticians and quality practitioners
- Describe and evaluate proposed method (H and HK1)
  - simulation studies
  - compare versus more complex methods (GPQ, Bayesian)
- Conclusion / Recommendations

What is specification?

Source: ICH Q6B – Specifications: Test Procedures and Acceptance Criteria for Biotechnological / Biological Products

- A specification is defined as a list of tests, references to analytical procedures, and appropriate acceptance criteria which are numerical limits, ranges, or other criteria for the tests described. It establishes the **set of criteria** to which a new drug substance or new drug **product should conform** to be considered acceptable for its intended use.

Two levels of specification: Release and Shelf-life

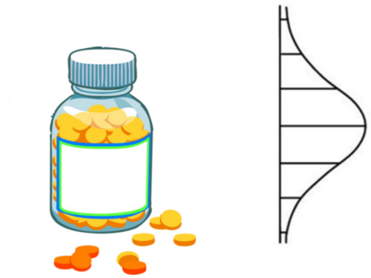


RELEASE limits are set such that if met at time of manufacture, there will be a high assurance that the attribute will remain within the SHELF-LIFE limits throughout the expiry of the product.

ESSENTIAL HEALTH

Justification of specifications: Clinical experience vs. Process Capability?

CLINICAL EXPERIENCE

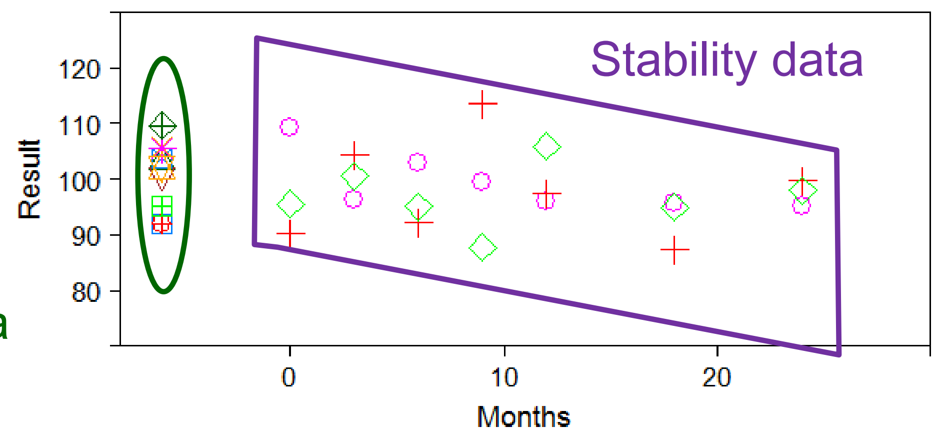


Quality attribute

safe

efficacious

ANALYSIS OF BATCHES MANUFACTURED



Release data

Assume specification limit derived from batch data analysis

Study Motivation and Goal

- Previous methods set Release limit assuming Shelf-life limit is already established
  - Practical, “industry” method (Allen, Dukes, Gerger, 1991)
  - Bayesian approach (Shao and Chow, 1991)
- In current exercise, both Release and Shelf-life limits need to be established jointly from available Release and Stability data
- Tolerance Interval is a logical basis for establishing limits
- CHALLENGE: Available methods for tolerance interval with random effects require complex formula, often involving simulations
- GOAL: propose a simple, closed-form tolerance interval for applied statisticians, QC/QA personnel, stability scientists

Statistical model for Release and Stability data

$$Y_{ij} = \mu + A_i + \beta t_{ij} + E_{ij}$$

Y_{ij}	measured response for lot *i* $(i = 1, \ldots, I\ lots)$ at time *j*, $(j = 1, \ldots, J_i\ timepoints\ for\ lot_i)$
μ	overall mean
A_i	Random lot effect on intercept, $A_i \sim N(0, \sigma_A^2)$
β	Slope (assumed common among lots)
E_{ij}	Random error, $E_{ij} \sim N(0, \sigma_E^2)$

A_i *and* E_{ij} *are jointly independent*

$Y_{ij}\ \ at\ t_0\ of\ interest\ \sim N(\mu + \beta t_0, \sigma_A^2 + \sigma_E^2)$

Calculate Tolerance Interval for Y_{ij} evaluated at:

- $t_0 = 0$ (set as Release limit)
- $t_0 = expiry$ (set as Shelf−life limit)

Tolerance Interval for Random Effects Model

$$Y_{ij} = \mu + A_i + \beta t_{ij} + E_{ij}$$

No stability trend, $\beta = 0$

- Simplifies to a One-Way Random Effect
- Widely studied, e.g.:
  - Hoffmann and Kringle, 2005 (**HK**) – TWO-sided
  - Hoffman, 2010 (**H**) – ONE-sided

With stability trend, $\beta \neq 0$

- Computationally more complex
  - longitudinal
  - hierarchical
  - unbalanced
- Available methods
  - Generalized Pivotal Quantity, (**GPQ**, Liao et al, 2005)
  - Bayesian Posterior Predictive (**BayesPP**)

Proposed Simplified Tolerance Interval Method

$$Y_{ij} = \mu + A_i + \beta t_{ij} + E_{ij}$$

No stability trend, $\boldsymbol{\beta = 0}$

- Simplifies to a One-Way Random Effect
- Widely studied, e.g.:
  - Hoffmann and Kringle, 2005 (**HK**) – TWO-sided
  - Hoffman, 2010 (**H**) – ONE-sided

With stability trend, $\boldsymbol{\beta \neq 0}$

- Computationally more complex
  - longitudinal
  - hierarchical
  - unbalanced

Adaptation using confidence intervals in regression models with unbalanced one-fold nested error structures ***(Park and Burdick, 2003)***

Formula for proposed Tolerance Interval method

H will be shown to be conservative; customization from **HK**

NOTATION	FORMULA
Lower and Upper TWO-sided bounds **(HK Method)**	$[L_2, U_2] = \hat{Y} \mp Z_{\frac{1+P}{2}} \sqrt{1+\frac{1}{n_E}} \sqrt{U}$
Lower and Upper bound on ONE-sided tolerance bound **(H Method)**	$[L_1, U_1] = \hat{Y} \mp \left(Z_P \sqrt{U} + Z_{1-\alpha} \sqrt{\hat{V}(\hat{Y})} \right)$
Lower and Upper ONE-sided bounds **(HK1 Method)**	$[L_1, U_1] = \hat{Y} \mp Z_P \sqrt{1+\frac{1}{n_E}} \sqrt{U}$
Upper bound on variance of Y	$U = S_L^2 + S_E^2 \left[1-\frac{1}{J_H}\right] + \sqrt{C_1^2 S_L^4 + C_2^2 \left(1-\frac{1}{J_H}\right)^2 S_E^4}$
Effective Sample Size	$n_E = \frac{\hat{V}(Y)}{\hat{V}(\hat{Y})}$
Estimated variance of Y	$\hat{V}(Y) = S_L^2 + S_E^2 \left(1-\frac{1}{J_H}\right)$
Estimated variance of $\hat{Y}$ at t_0	$\hat{V}(\hat{Y}) = \frac{S_L^2}{I} + \frac{S_E^2 \left(t_0 - \bar{t}^*\right)^2}{S_{ttw}}$

Formula (cont'd)

NOTATION	FORMULA
Constants used to compute U where $\chi^2_{\alpha:s}$ is a chi-squared percentile with area α to the left and s degrees of freedom	$C_1 = s/\chi^2_{\alpha:s} - 1; C_2 = r/\chi^2_{\alpha:r} - 1$
Error degrees of Freedom	$r = \sum_{i=1}^{I} J_i - I - 1$
Lot Degrees of Freedom	$s=I-1$
Harmonic mean of the number of values in each lot	$J_H = \dfrac{I}{\sum_{i=1}^{I} \frac{1}{J_i}}$
Predicted value of Y at t_0	$\hat{Y} = \overline{Y}^* + \hat{\beta}\left(t_0 - \overline{t}^*\right)$
Mean Squared Error	$S_E^2 = \dfrac{S_{yyw} - \frac{\left(S_{tyw}\right)^2}{S_{ttw}}}{r}, \quad S_{yyw} = \sum_{i=1}^{I}\sum_{j=1}^{J_i}\left(Y_{ij} - \overline{Y}_i\right)^2$
Variance of Predicted Lot Means	$S_L^2 = \dfrac{\sum_{i=1}^{I}\left(\overline{Y}_i - \overline{Y}^* + \hat{\beta}\left(\overline{X}^* - \overline{X}_i\right)\right)^2}{I-1}$ $\overline{Y}^* = \dfrac{\sum_{i=1}^{I}\overline{Y}_i}{I}, \quad \overline{X}^* = \dfrac{\sum_{i=1}^{I}\overline{X}_i}{I}$
Within lot estimate of slope	$\hat{\beta} = \dfrac{S_{tyw}}{S_{ttw}}, \quad S_{tyw} = \sum_{i=1}^{I}\sum_{j=1}^{J_i}\left(t_{ij} - \overline{t}_i\right)\left(Y_{ij} - \overline{Y}_i\right)$ $S_{ttw} = \sum_{i=1}^{I}\sum_{j=1}^{J_i}\left(t_{ij} - \overline{t}_i\right)^2, \overline{Y}_i = \dfrac{\sum_{j=1}^{J_i} Y_{ij}}{J_i}, \quad \overline{t}_i = \dfrac{\sum_{j=1}^{J_i} t_{ij}}{J_i}$

Evaluate Proposed Method via Simulation

FACTOR	LEVELS STUDIED
Trend with time	• **CHANGES**, slope ($\beta = -0.15$)
Sample size	• **SMALL**: [4(0),3(6),3(12)] *[# of lots (max. Months of lots)]* • **LARGE**: [10(0),10(12),10(24)]
Shelf-life (or expiry)	• **12** months for Small sample size • **24** months for Large sample size
Intraclass Correlation Coefficient	$\rho = \sigma_A^2 / (\sigma_A^2 + \sigma_E^2) = [\mathbf{0.2, 0.8}]$
Proportion, P	0.9, 0.95, 0.99, 0.9973
Data used	• **RS** = both Release + Stability data • **S** = Stability data only
Method used	• **H** = Hoffman (2010) one-sided limits • **HK1** = Hoffman and Kringle (2005) one-sided limits modified from Hoffman and Kringle (2005) two-sided

Simulation Performance Metrics (from 10,000 iterations)

- Average **one-sided** LRL and LSL, normalized to the true limit

The CLOSER to without going above 1, the BETTER.

2

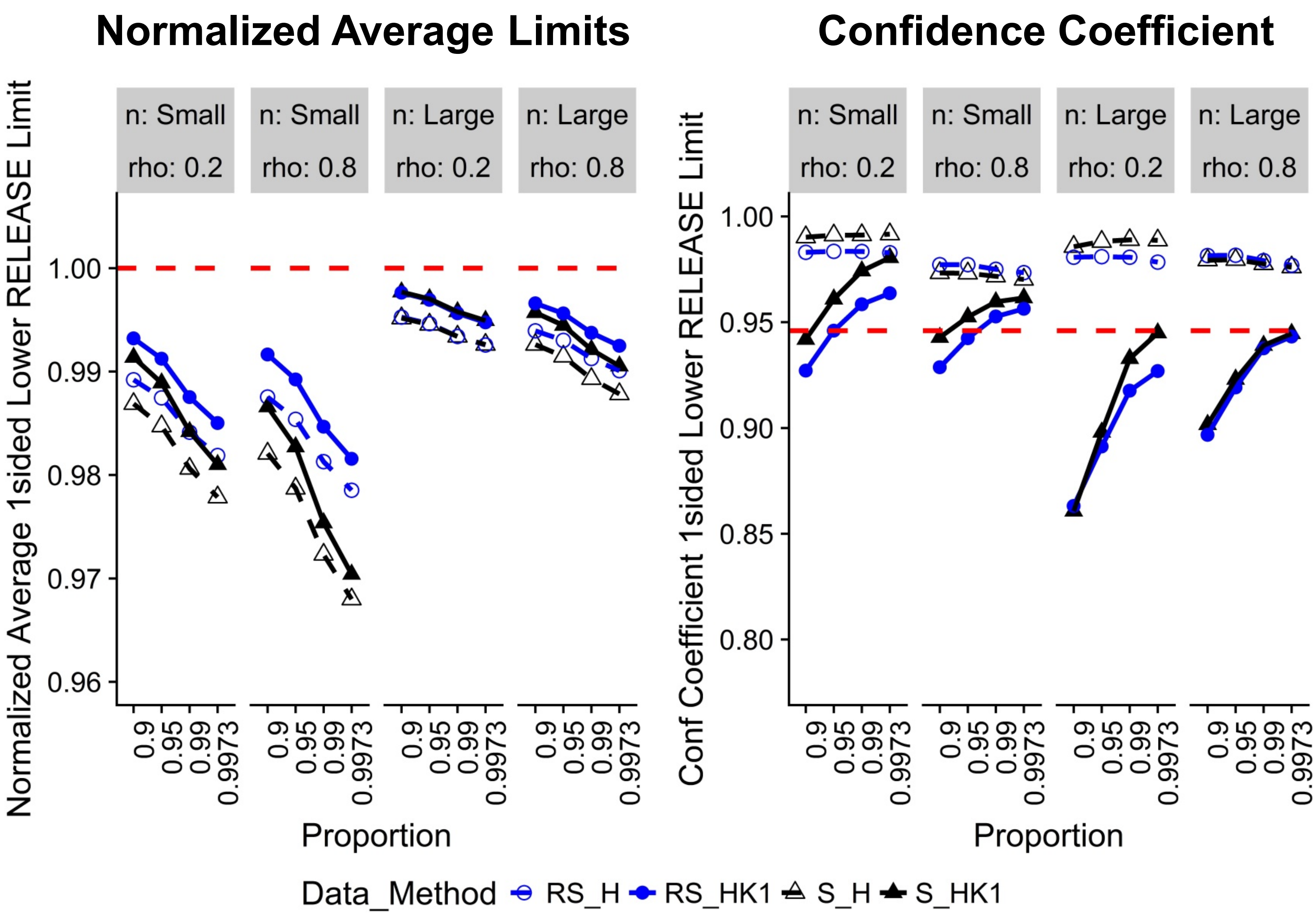
- CONFIDENCE COEFFICIENT - fraction of iterations where:
  - calculated **one-sided** LRL and LSL bracketed the true bounds.

The CLOSER the confidence coefficient to without dropping below the nominal level (i.e., 0.95), the BETTER.

1

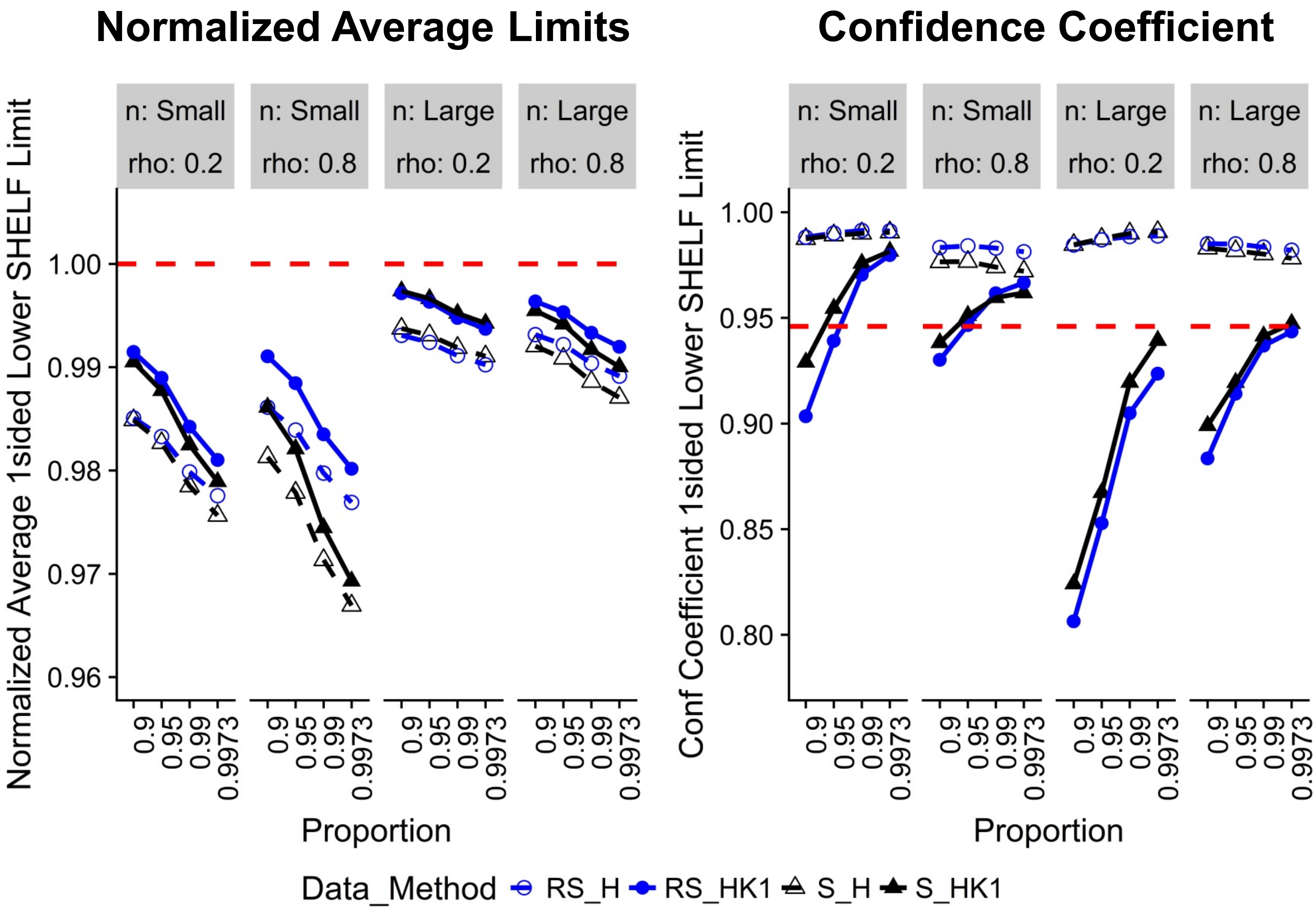
LRL = Lower Release Limits
LSL = Lower Shelf-life Limits

WITH Change with Time, ONE-sided Lower Release Limits (t_0=0)



ESSENTIAL HEALTH

WITH Change with Time, ONE-sided Lower Specification Limits (t_0=expiry)



Summary from simulation study to evaluate proposed method

- Combined Release and Stability (RS) data provides more precise limits than using only Stability (S) data.
- H method is too conservative (confidence coefficients well above the nominal level).
- HK1 vs. H methods comparison:
  - for targeted proportion, $P \geq 0.99$, HK1 is more precise and less conservative than H (while still meeting nominal confidence) limits
  - for targeted proportion, $P < 0.99$, HK1 does not meet nominal confidence. Conservative H method is a “safer” choice.
- The relative performances of the methods are approximately the same for the one-sided LRL and the one-sided LSL.
  - expected because Release and Shelf-life limits are jointly set

Is there any loss of performance using simplified proposed HK1 and H methods?

Benchmark comparison vs. more complex alternative methods

Generalized Pivotal Quantity

One- and Two-Sided Tolerance Intervals for General Balanced Mixed Models and Unbalanced One-Way Random Models

C. T. LIAO

Division of Biometry
Institute of Agronomy
National Taiwan University
Taipei, Taiwan
(*ctliao@ntu.edu.tw*)

T. Y. LIN

Division of Biometry
Institute of Agronomy
National Taiwan University
Taipei, Taiwan
(*d89621201@ntu.edu.tw*)

H. K. IYER

Department of Statistics
Colorado State University
Fort Collins, CO 80523
(*hari@stat.colostate.edu*)

Bayesian Posterior Predictive

- Wolfinger, 1998
  - TI for variance components
- Gelman and Hill, 2007
  - Hierarchical model
- Kruschke, 2015
  - Hierarchical model, JAGS

Generalized Pivotal Quantity (GPQ)

$$R_{\hat{Y},GPQ} = \hat{Y} - Z\sqrt{V(\hat{Y})}$$

$$TI_{1sided,GPQ} = R_{\hat{Y},GPQ} + Z_{\beta}\sqrt{R_{\sigma_A,GPQ} + R_{\sigma_E,GPQ}}$$

- R's are solutions to generalized pivotal quantities.
- Requires 10,000 simulations of normal and chi square random variates.
- Requires solving for root of nonlinear equation to obtain $R_{\sigma_A,GPQ}$.
- Output $\frac{\alpha}{2}th$ quantile of the 10,000 simulations.

Bayesian method: Likelihood and Priors

FIXED Slope

$$Y_{ij} = \mu + A_i + \beta \times t_{ij} + E_{ij}$$

$$\mu \sim N\left(100, 10^2\right)$$

$$A_i \sim N\left(0, \sigma_A^2\right), \qquad \sigma_A \sim \frac{U(0,\ 30)}{\sqrt{\Gamma(1,\ 0.5)}}$$

$$\beta \sim N\left(0, 2^2\right)$$

$$E_{ij} \sim N\left(0, \sigma_E^2\right), \qquad \sigma_E \sim \frac{U(0, 30)}{\sqrt{\Gamma(1,\ 0.5)}}$$

$$\left(\mu^{[s]}, \sigma_A^{[s]}, \beta^{[s]}, \sigma_E^{[s]}\right),$$
$$s = 1, \ldots, 20K$$

RANDOM Slope

$$Y_{ij} = \mu + A_i + \beta_i \times t_{ij} + E_{ij}$$

$$\mu \sim N\left(100, 10^2\right)$$

$$A_i \sim N\left(0, \sigma_A^2\right), \quad \sigma_A \sim \frac{U(0,\ 30)}{\sqrt{\Gamma(1,\ 0.5)}}$$

$$\beta_i \sim N\left(b, \sigma_b^2\right), \quad b \sim N\left(0, 2^2\right), \quad \sigma_b \sim \frac{U(0, 2)}{\sqrt{\Gamma(1,\ 0.5)}}$$

$$E_{ij} \sim N\left(0, \sigma_E^2\right), \quad \sigma_E \sim \frac{U(0, 30)}{\sqrt{\Gamma(1,\ 0.5)}}$$

$$\left(\mu^{[s]}, \sigma_A^{[s]}, b^{[s]}, \sigma_b^{[s]}, \sigma_E^{[s]}\right),$$
$$s = 1, \ldots, 20K$$

Simulate 20,000 future values from joint posterior distribution of parameters; output, p quantile.

Some distinctives of Bayesian method

- BayesPP method is conditioned on the observed data
  - confidence properties depend on chosen priors and not expected to provide nominal frequentist confidence in general
- Type of tolerance interval
  - H, HK1, and GPQ are P-**content** type
  - BayesPP is a P-**expectation** type (i.e., narrower than content type)

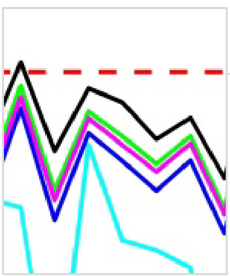
Frequentist (H, HK1, and GPQ) and Bayesian approaches (BayesPP) are conceptually different, but comparison of methods is still useful.

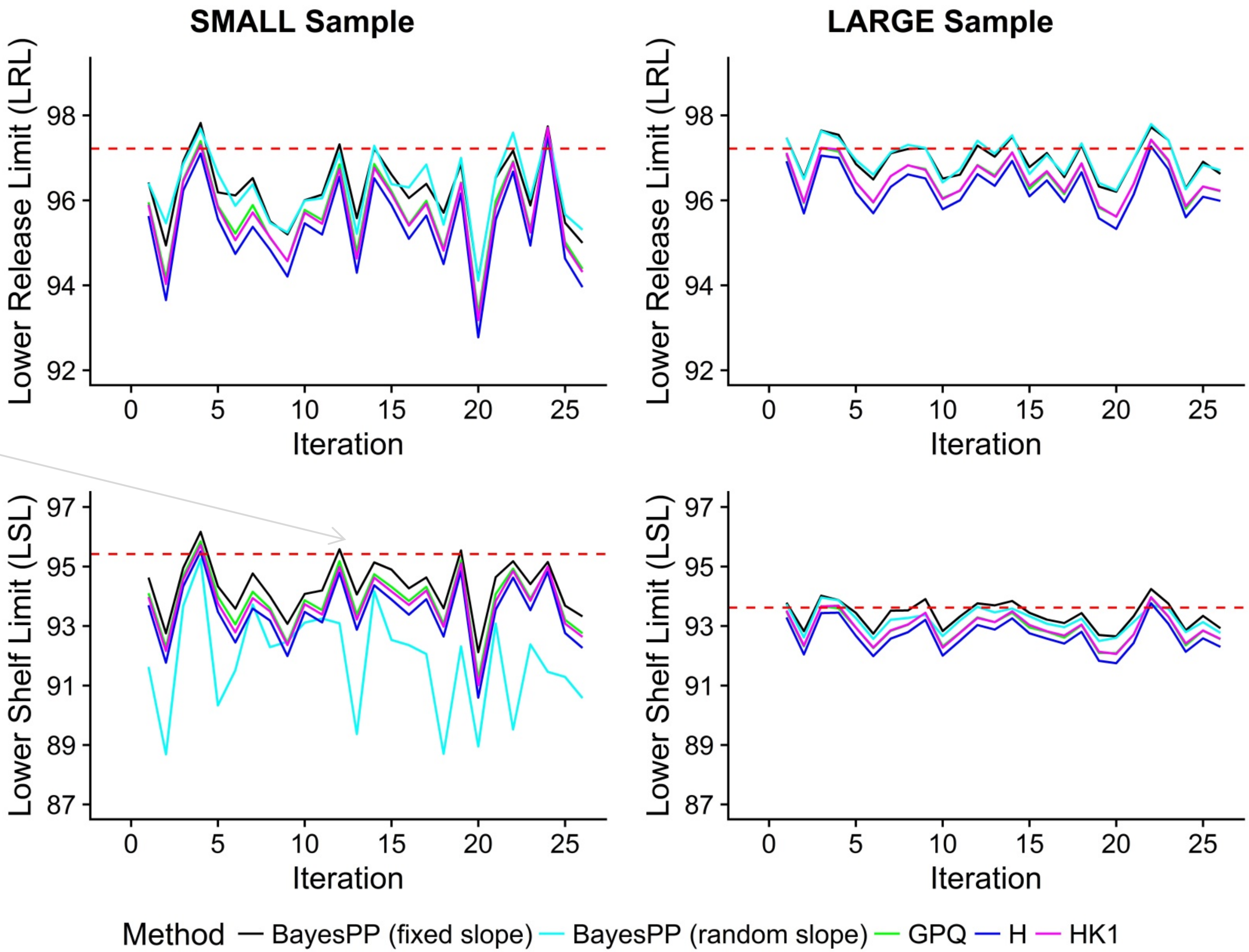
Computational constraints in comparison of proposed method with GPQ and BayesPP

Extent of Simulation	Compared Methods	# of Iterations	# of Simulated Future Values per Iteration
Comprehensive	H and HK1	10,000	not applicable (closed formula)
Limited	H and HK1	100	not applicable (closed formula)
	GPQ		10,000
	BayesPP		20,000 (from the pooled 4 MCMC chains of 5,000 each)

- *Perform only 100 iterations for proposed vs. GPQ and BayesPP comparison*
- *Scenario evaluated is β=-0.15, ρ=0.8, P=0.9973, Release + Stability data*

Compare H and HK1 vs. GPQ and BayesPP limits





HK1 and GPQ overlap
H is wider
BayesPP (random) wider
BayesPP (fixed) narrower

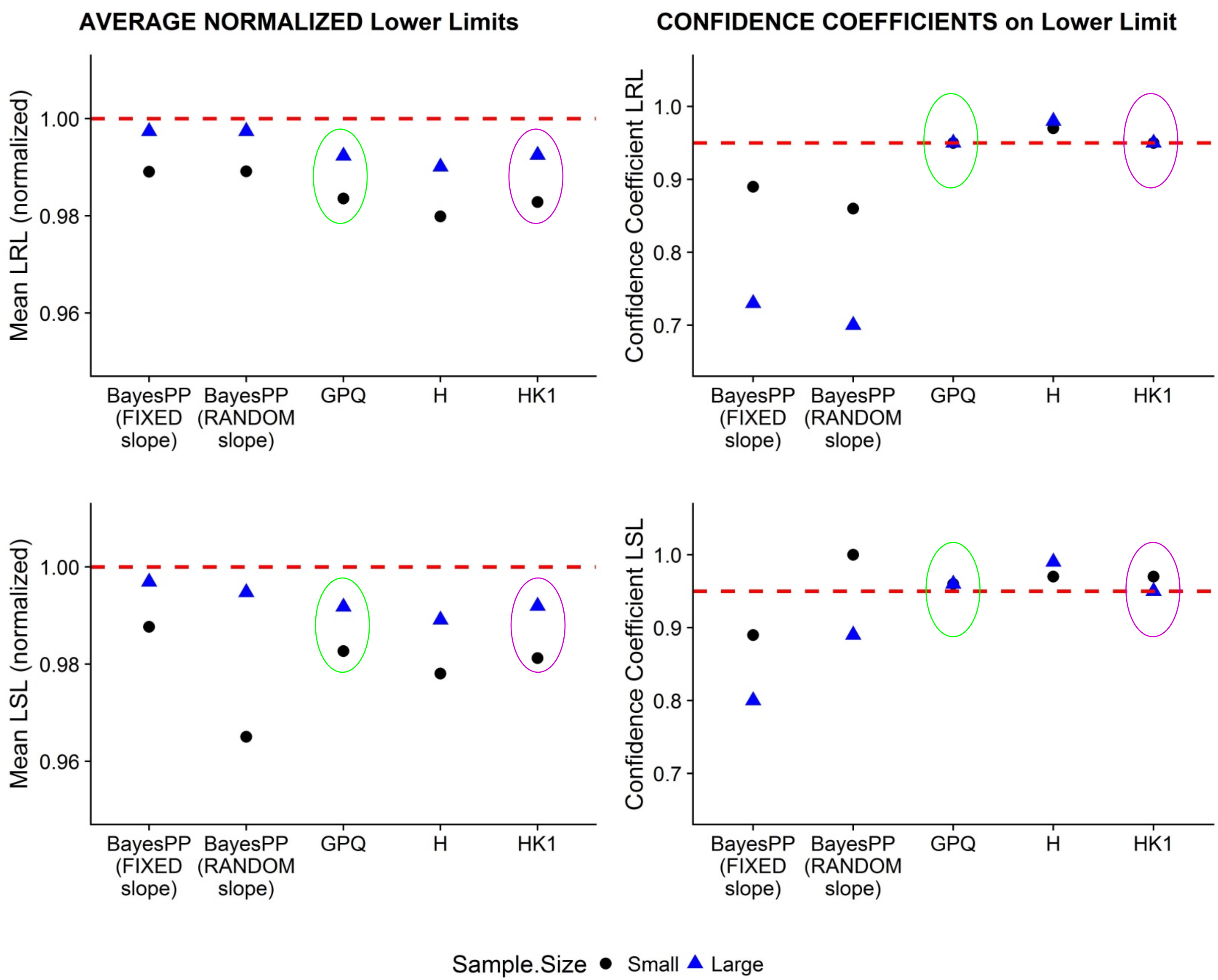
ESSENTIAL HE

Note: Only 25 of 100 iterations shown for clearer visualization of trends.

Metrics - H and HK1 vs. GPQ and BayesPP (for 100 iterations)

HK1 and GPQ has reasonable precision

HK1 and GPQ maintain nominal confidence



Summary of comparison vs. GPQ and BayesPP

- HK1 and GPQ methods have comparable performance, almost overlapping each other
  - maintain the nominal 0.95 confidence while more very closely approximating the true limits
- H method more conservative (wider) than HK1 and GPQ
- BayesPP (random slope) too conservative for the case considered
  - generates a non-zero posterior predictive distribution for σ_b
  - data generated using $\sigma_b = 0$
- BayesPP (fixed slope) does not maintain 0.95 confidence
  - Not expected to as not calibrated to do so, only conditioned upon data and prior selection; also, P-expectation tolerance interval type

Conclusions and Recommendations

- One-Way Random Effect model for NO change over time extended to WITH change over time
  - mixed effects linear regression with random intercepts and a fixed slope
  - Release and Shelf-life specification limits can be simultaneously set using the same equation evaluated at $t_0 = 0$ and $t_0 = expiry$.
- Proposed H and HK1 method relatively simpler to implement without any performance loss compared to more computationally complex GPQ
  - closed-form, no simulation involved
  - Excel spreadsheets can be used
  - accessible to applied practitioners (QA/QC, stability scientists)

References

- Allen, P. V., Dukes, G. R., Gerger, M. E. Determination of release limits, a general methodology. Pharmaceutical Research, 1991, 8(9), 1210-1213.
- Gelman A., Hill, J. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, New York, NY, 2007, 251-300.
- Hoffman, D. One-sided tolerance limits for balanced and unbalanced random effects models. Technometrics, 2010, 52(3), 303-312.
- Hoffman, D., Kringle, R. Two-sided tolerance intervals for balanced and unbalanced random effects models. Journal of Biopharmaceutical Statistics, 2005, 15(2), 283-293.
- ICH Q6B Specifications: Test Procedures and Acceptance Criteria for Biotechnological / Biological Products (1999). Available at https://www.fda.gov/ucm/groups/fdagov-public/@fdagov-drugs-gen/documents/document/ucm073488.pdf
- Kruschke, J.K., Doing Bayesian data analysis: a tutorial with R, JAGS, and Stan. 2nd ed. Academic Press, San Diego, CA, 2015.
- Liao, C.T., Lin, T.Y., Iyer, H.K. One- and two-sided tolerance intervals for general balanced mixed models and unbalanced one-way random models. Technometrics, 2005, 47(3), 323-335.
- Park, D.J., Burdick, R.K. Performance of confidence intervals in regression models with unbalanced one-fold nested error structures. Communications in Statistics - Simulation and Computation. 2003, 32(3), 717-732.
- Shao, J., Chow, S.C. Constructing release targets for drug products: A Bayesian decision theory approach, Journal of the Royal Statistical Society. Series C (Applied Statistics). 1991, 40(3), 381–390
- Wolfinger, R. Tolerance intervals for variance component models using Bayesian simulation. Journal of Quality Technology, 1998, 30(1), 18-32.