June 17th, 2026

A robustness assessment of the latent variable framework for composite endpoints: With application to late-stage trials

Paul Newcombe[1], Jasna Cotic[1], David Whitney[1], Lindsey Schader[2], Jane Bentley[3], Aris Perperoglou[4], James Wason[5] & Dave Lunn[6]

[1]GSK, London, UK, [2]GSK, Denver, Colorado, USA, [3]GSK, Stockley Park, UK, [4]GSK, Stevenage, UK, [5]Newcastle University, Newcastle, UK, [6]GSK, Brentford, UK

Utilising novel statistical methods to design trials in industry

Seeking adoption of the LVF framework we are currently focused on:

1. Evaluating **Robustness to mis-specification**
   - The multivariate normal (MVN) assumption on the continuous components is core to the precision gain
   - It is not required in a traditional responder analysis
   - Demonstrating robustness is expected to be **critical for regulatory acceptance**

2. Both **Bayesian** and **frequentist** implementations
   - Bayesian: for internal decision making following Ph 2, evidence synthesis. Bayesian reasoning widely used for internal decision making[1] and use of historical controls increasingly adopted[2]
   - Frequentist: for designing pivotal Phase 3 trials

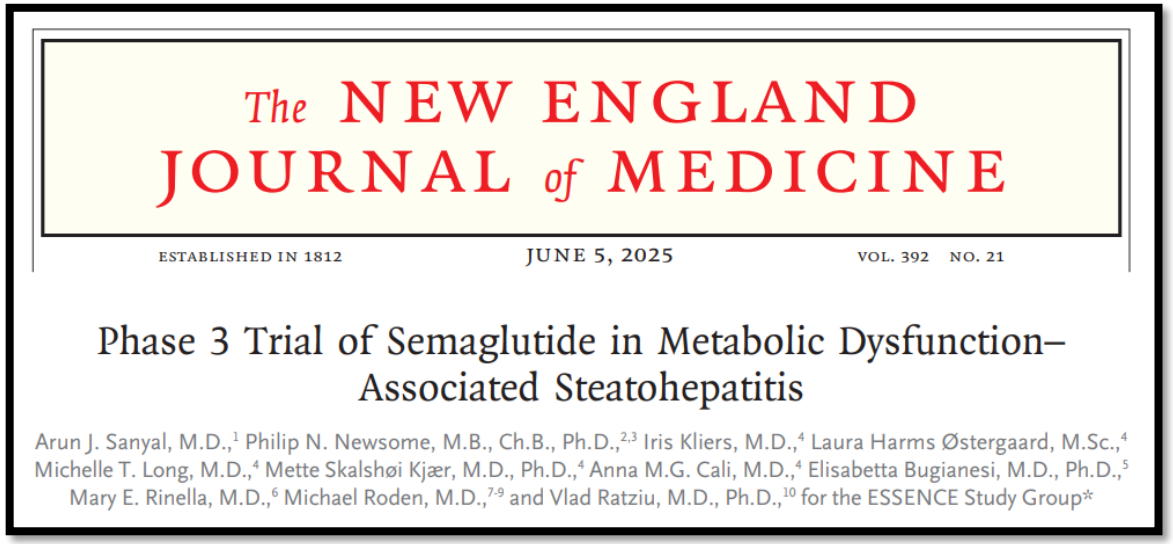
3. Showcasing **application to late-stage trials**
   - To our knowledge, this methodology has never been used to design a pivotal trial
   - Will present first application to two large GSK Ph 3 trials: Of belimumab in SLE

GSK

[1]Temple and Robertson. 2021. "Conditional assurance: the answer to the questions that should be asked within drug development". *Pharmaceutical Statistics*
[2]Best et al. 2023. "Beyond the classical type I error: Bayesian metrics for Bayesian designs using informative priors". *Statistics In Biopharmaceutical Research*

Robustness Evaluation: Hepatology motivating example

Responder endpoint built from two continuous biomarkers

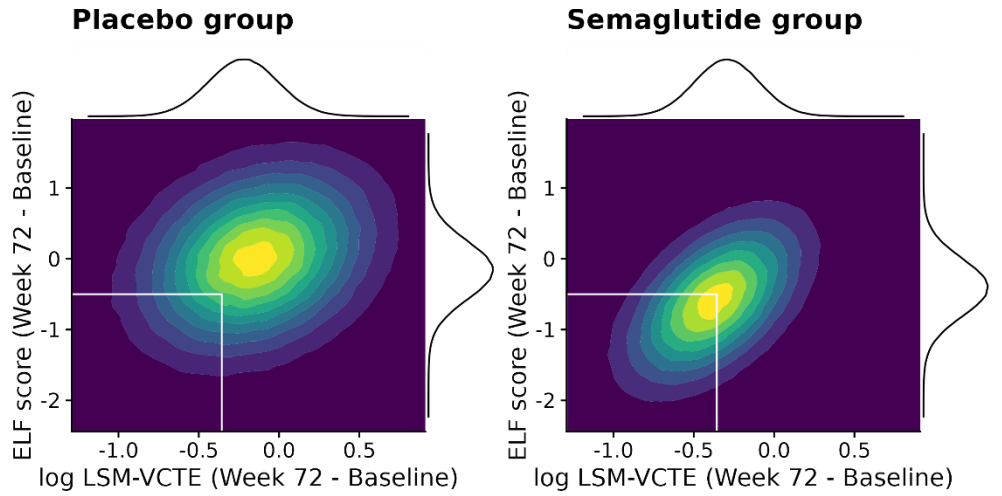


The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 JUNE 5, 2025 VOL. 392 NO. 21

Phase 3 Trial of Semaglutide in Metabolic Dysfunction–Associated Steatohepatitis

Arun J. Sanyal, M.D.,[1] Philip N. Newsome, M.B., Ch.B., Ph.D.,[2,3] Iris Kliers, M.D.,[4] Laura Harms Østergaard, M.Sc.,[4] Michelle T. Long, M.D.,[4] Mette Skalshøi Kjær, M.D., Ph.D.,[4] Anna M.G. Cali, M.D.,[4] Elisabetta Bugianesi, M.D., Ph.D.,[5] Mary E. Rinella, M.D.,[6] Michael Roden, M.D.,[7-9] and Vlad Ratziu, M.D., Ph.D.,[10] for the ESSENCE Study Group*



- Recent Ph 3 included post hoc analysis of dichotomised change in two continuous biomarkers
- Summary data shared characterising component variances and correlation by treatment arm
- Opportunity to inform a “realistic” simulation to explore how robust the MVN assumption is when in-correct

Table S14. Post-hoc analysis of total Number of Participants Achieving Enhanced Liver Fibrosis Score Decrease of ≥0.5 and Liver Stiffness Decrease of ≥30% (Full Analysis Set, Interim)

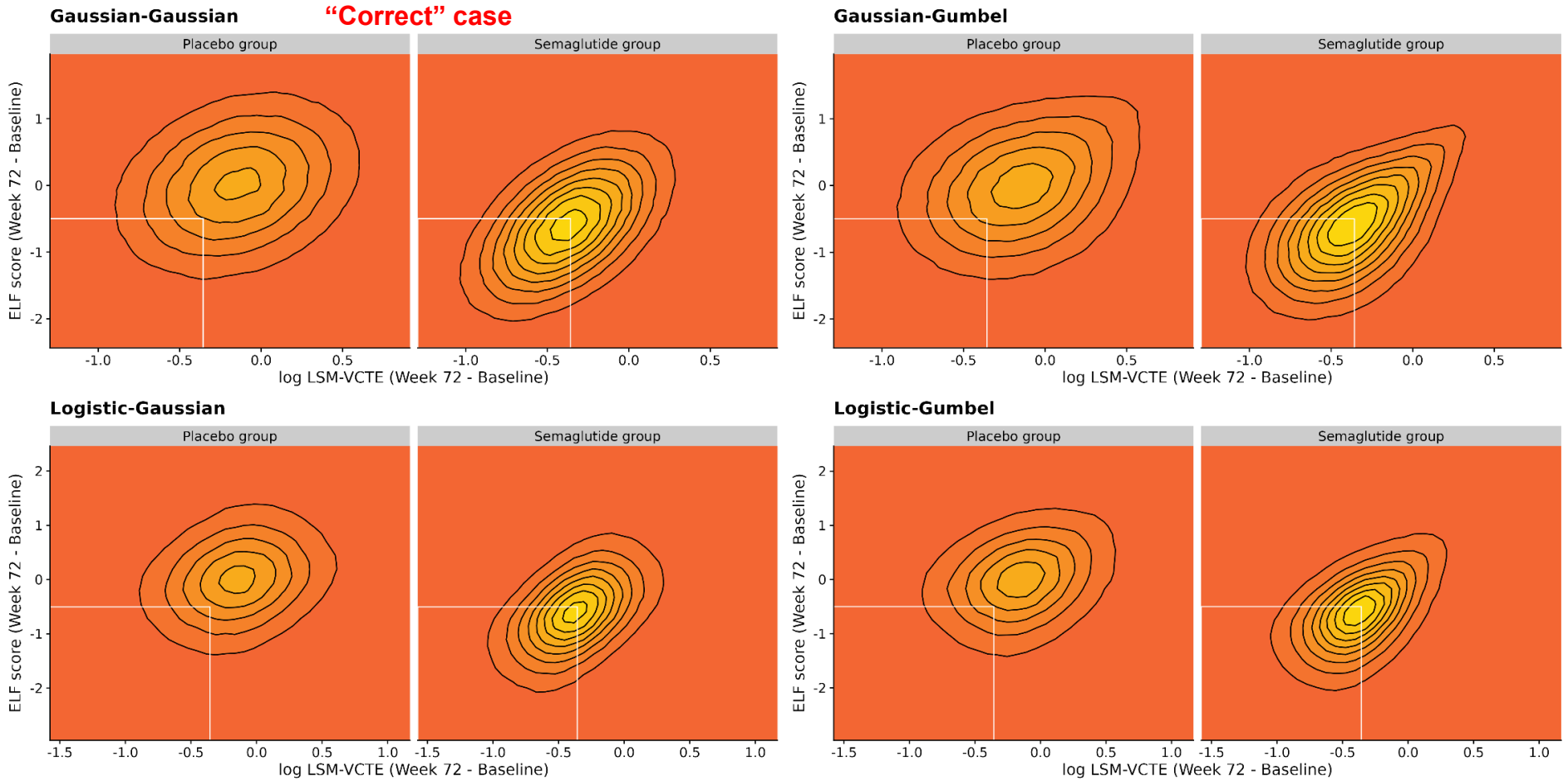
Measure	Semaglutide 2.4 mg (N=534)	Placebo (N=266)
Participants with enhanced liver fibrosis score ≥0.5 and liver stiffness ≥30% at baseline, N (%)	417 (100.0)	216 (100.0)
Enhanced liver fibrosis Score ≥0.5 and liver stiffness ≥30%, N (%)		
Yes	138 (33.1)	20 (9.3)
No	229 (54.9)	167 (77.3)

Number of participants with missing evaluations for semaglutide/placebo were 50/417 and 29/216, respectively. Liver stiffness data are from sites with available VCTE equipment only. VCTE, vibration-controlled transient elastography.

Robustness Evaluation: Simulation Scenarios

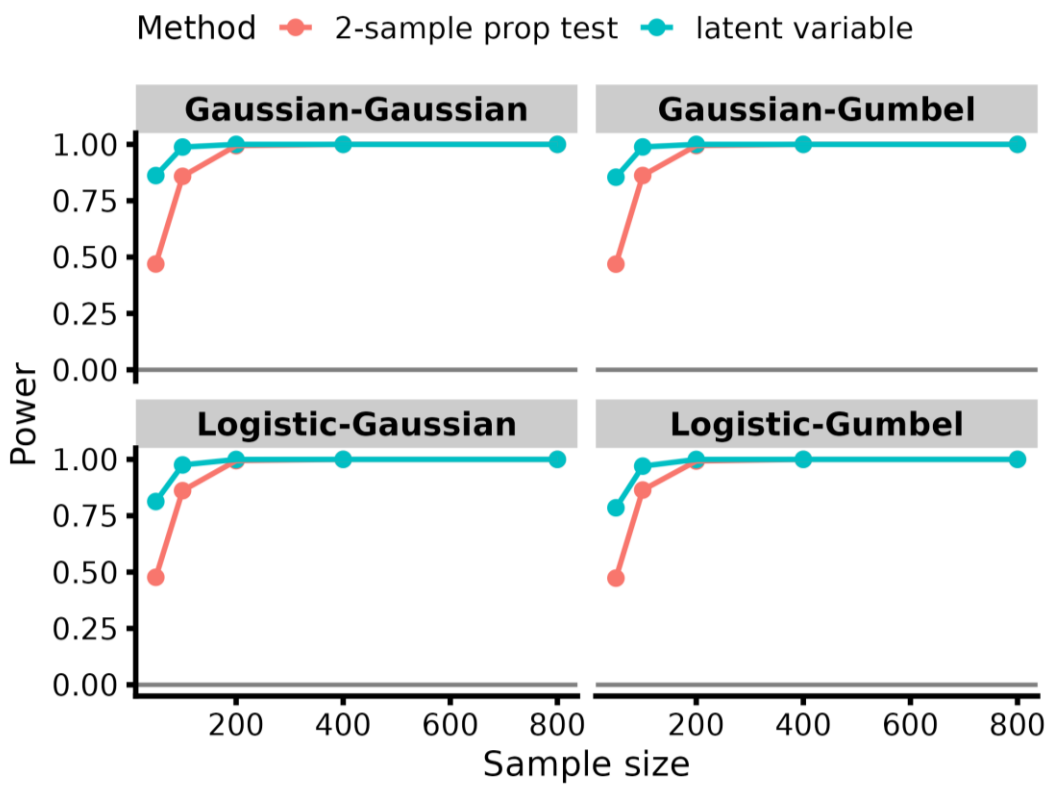
Based on Semaglutide Ph3 supplement

- Four scenarios, varying marginal choices (Gaussian/Logistic/Gumbel). 2 Coppulas.
- Bi-variate DGMs calibrated to NIT components published in Sanyal et al

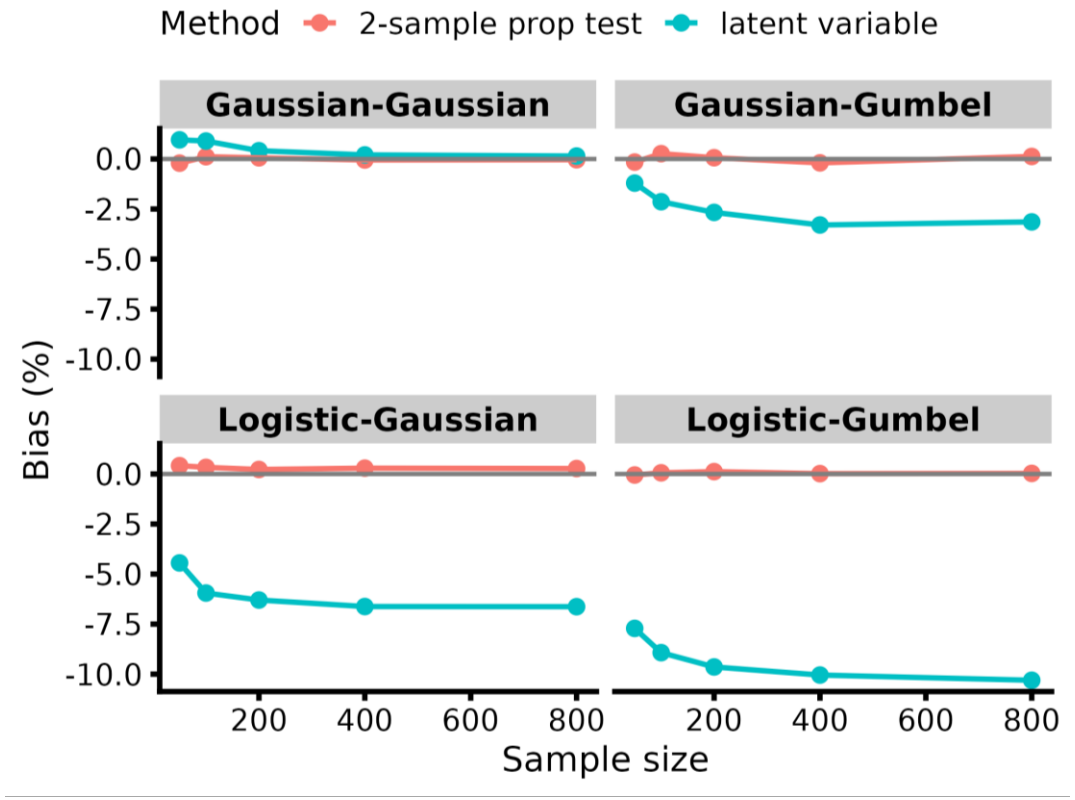


Power gain holds, but mis-specified marginals introduce bias

Power (alternative)



Bias (alternative)

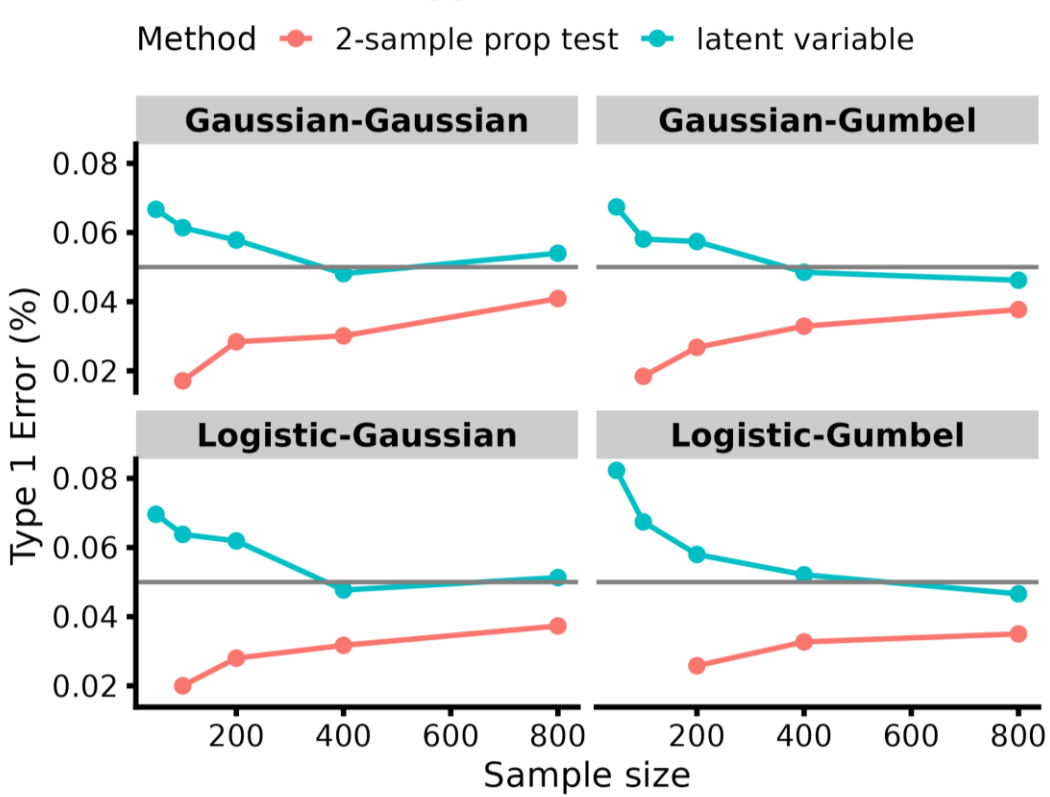


- LVF retains efficiency gain even under mis-specification
- LVF has non-trivial downwards bias under mis-specification, that does not diminish with N

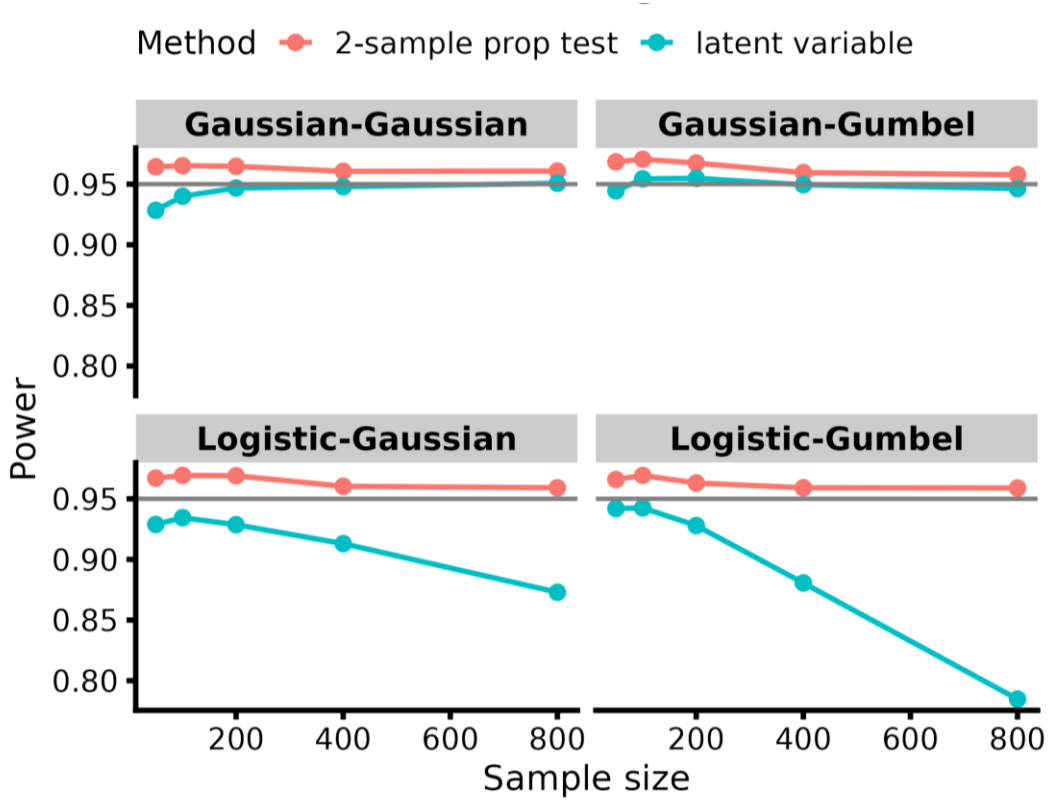
Work by David Whitney

Type 1 error preserved / Coverage drops when margins mis-specified

Type 1 error rate (null)



Coverage (alternative)



- Type 1 error rate inflated at lower sample sizes, but controlled at nominal 5% level across all scenarios for larger trials
- Coverage drops below nominal under mis-specification: gets worse with sample size

Work by **David Whitney**

Utilising novel statistical methods to design trials in industry

Seeking adoption of the LVF framework we are currently focused on:

1. Evaluating **Robustness to mis-specification**
   - The multivariate normal (MVN) assumption on the continuous components is core to the precision gain
   - It is not required in a traditional responder analysis
   - Demonstrating robustness is expected to be **critical for regulatory acceptance**

2. Both **Bayesian** and **frequentist** implementations
   - Bayesian: for internal decision making following Ph 2, evidence synthesis. Bayesian reasoning widely used for internal decision making[1] and use of historical controls increasingly adopted[2]
   - Frequentist: for designing pivotal Phase 3 trials

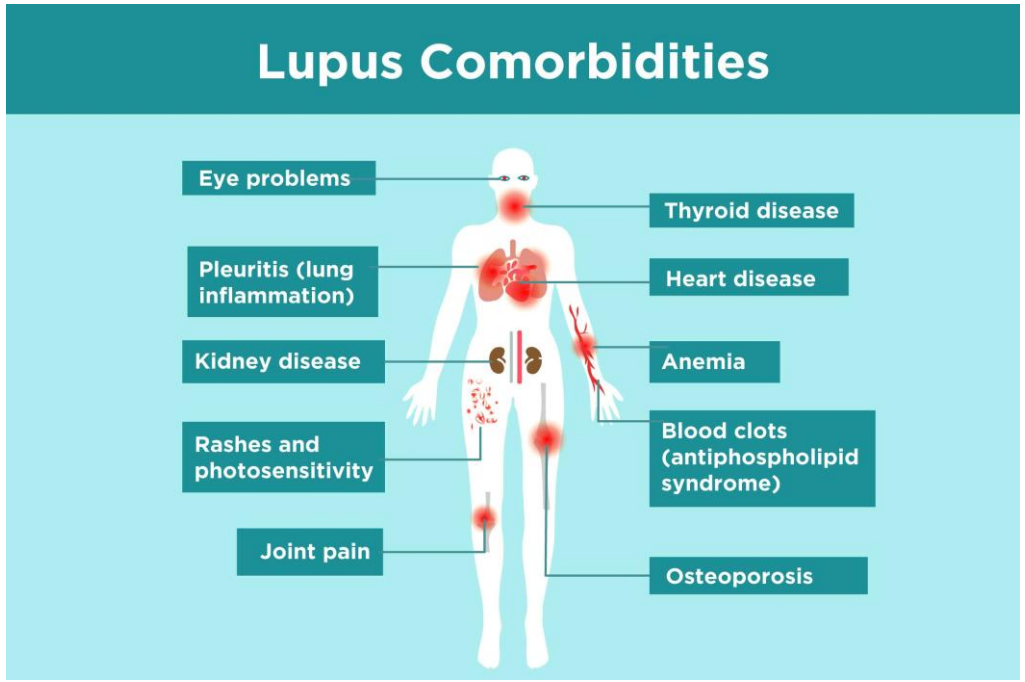
3. Showcasing **application to late-stage trials**
   - To our knowledge, this methodology has never been used to design a pivotal trial
   - Will present first application to two large GSK Ph 3 trials: Of belimumab in SLE

GSK

[1]Temple and Robertson. 2021. "Conditional assurance: the answer to the questions that should be asked within drug development". *Pharmaceutical Statistics*
[2]Best et al. 2023. "Beyond the classical type I error: Bayesian metrics for Bayesian designs using informative priors". *Statistics In Biopharmaceutical Research*

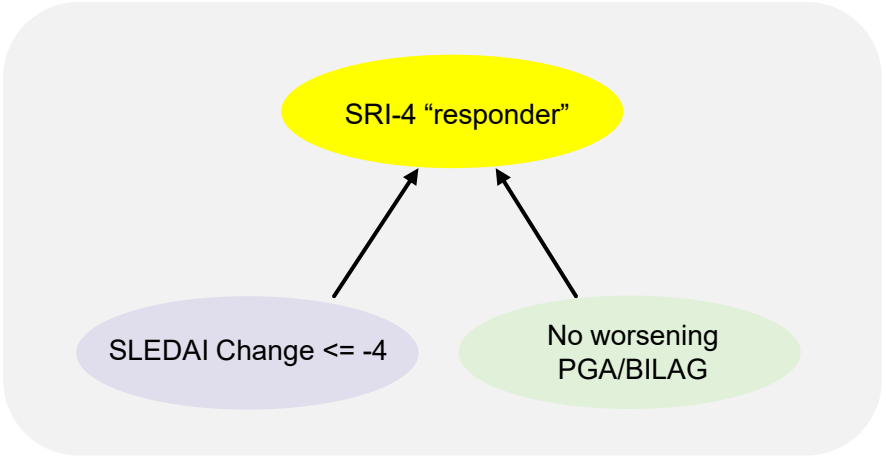
Motivating example: Two GSK Phase 3 SLE trials

- **Systemic Lupus Erythematosus** (SLE) is a lifelong disease with high rates of early and irreversible organ damage
- **Belimumab** is a B-cell modulator that targets the central immunopathogenic pathway in lupus by selectively inhibiting BLyS & reducing the autoreactive B cells that drive disease activity
- **BLISS-52** (NCT00424476) and **BLISS-76** (NCT00410384): Pivotal Phase III trials of Belimumab in active SLE
- Each trial had a 1:1:1 randomized design comparing two doses to placebo: We will focus on the approved 10 mg/kg dose



Source: CreakyJoints (creakyjoints.org)

Primary endpoint: SRI-4 Composite



GSK

Bayesian Implementation of the latent variable model

Analogous 2-step process to McMenamin et al

Step 1:

The same multivariate normal model for **endpoint components** is fitted in JAGs with vague priors

Joint model for treatment effect on endpoint components:

$$Y_{i1} = \alpha_0 + \alpha_1 T_i + \epsilon_{i1}$$ SLEDAI reduction

$$Y^*_{i2} = \psi_0 + \psi_1 T_i + \epsilon^*_{i2}$$ PGA/BILAG

$$(\epsilon_{i1}, \epsilon^*_{21}) \sim N\left(0, \Sigma = \begin{bmatrix} \sigma^2 & \sigma_{12} \\ \sigma_{12} & 1 \end{bmatrix}\right)$$

Priors

$\alpha_0, \alpha_1 \sim N(0, 10^6)$

$\psi_0, \psi_1 \sim N(0, 10^6)$

$\sigma_{12} \sim Unif(-1, 1)$

$\sigma \sim Unif(0, 20)$

Mapping continuous latent -> observed binary steroids

$$Y_{i2} = \begin{cases} 0 \ if \ y^*_{i2} < 0 \\ 1 \ if \ y^*_{i2} \geq 0 \end{cases}$$

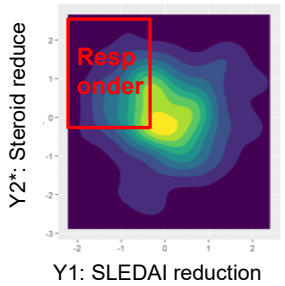
Step 2:

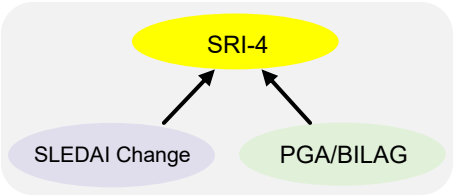
Posterior inference on the "responder" odds ratio:

1. Start with posterior sample *i* of joint model parameters
2. Infer responder probabilities for each patient by integrating the implied multivariate normal model
3. Estimate marginal "responder" odds ratio
4. Repeated for each posterior sample

NB: Computationally intensive, in practice use a posterior sub-sample

Bivariate endpoint distribution (for 1 patient)





Marginal Odds Ratio for Composite Response

$$= \frac{\left(\frac{\sum_{i=1}^{N} \hat{p}_{i1}}{N - \sum_{i=1}^{N} \hat{p}_{i1}}\right)}{\left(\frac{\sum_{i=1}^{N} \hat{p}_{i0}}{N - \sum_{i=1}^{N} \hat{p}_{i0}}\right)}$$

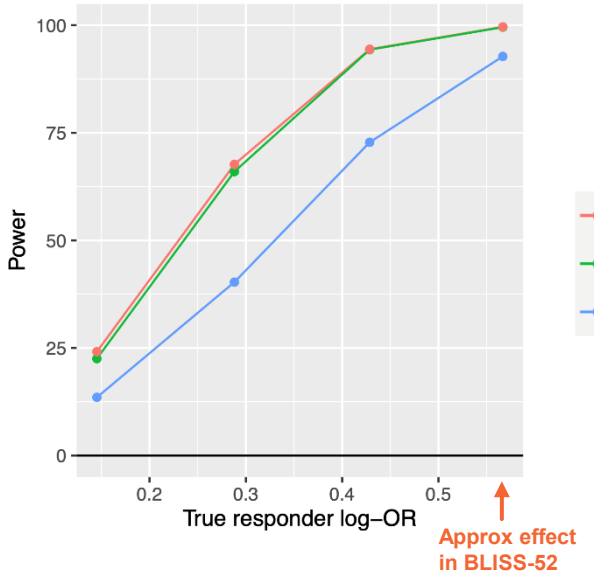
GSK

Simulations – Effect Size vs Power / Bias

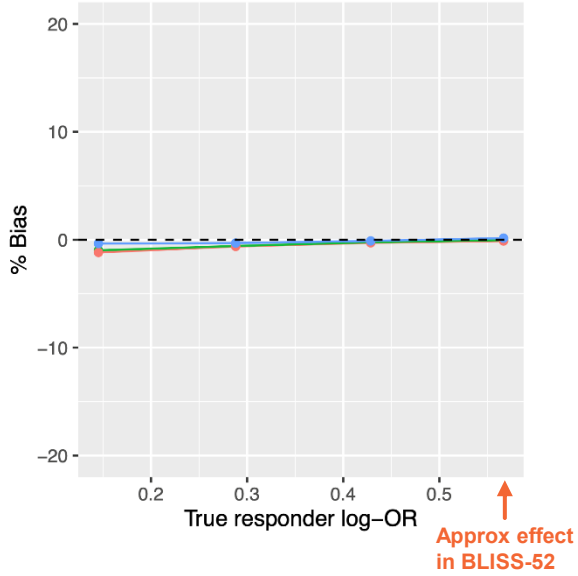
Bayesian implementation offers the same power boost and accuracy

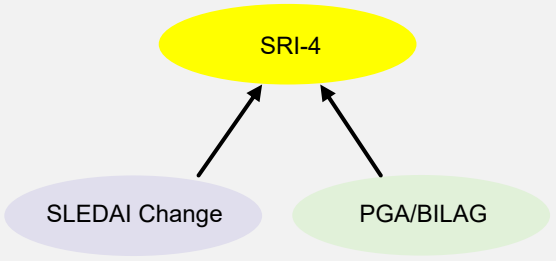
- 5,000 simulations (n = 300:300) based on parameters estimated in BLISS-52
- **Substantive power improvements** over range of effect sizes matching frequentist approach
- Bayesian **coverage and type 1 error slightly off** (~1%), driven by number of posterior sub-samples
- How does precision gain translate to smaller trials?

Power vs Effect Size



Bias vs Effect Size





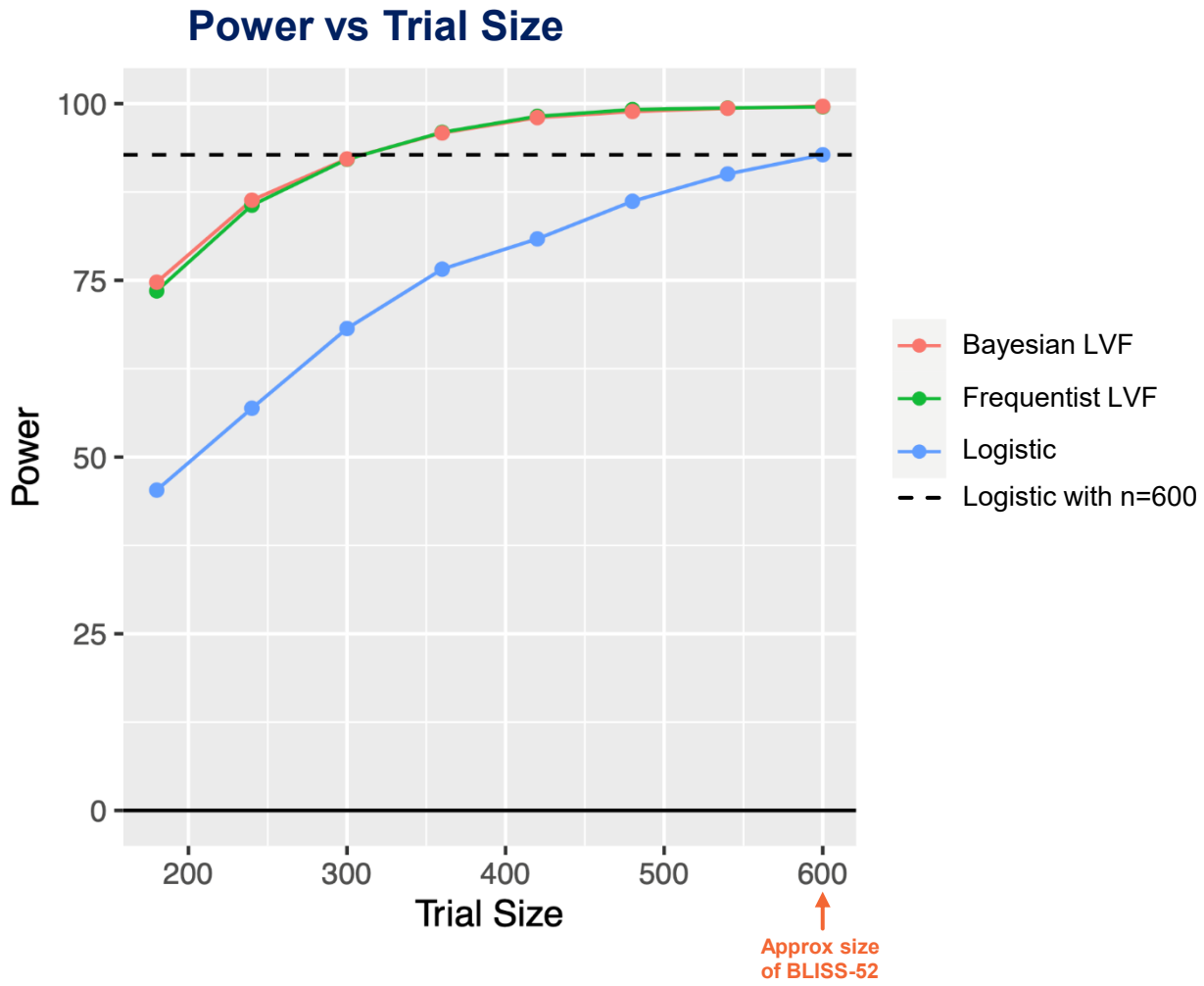
Type 1 error of different models (5,000 null simulations)

	Logistic	Bayes Logistic	McMenamin Latent	Bayes Latent
Type 1 Error	5.0	5.3	5.0	6.1
Coverage	95.0	94.7	95.0	93.9
Bias	-0.0038	-0.0039	-0.0025	-0.0023

Simulations – Trial Size vs Power

Logistic power matched with ~50% of the required trial size

- Repeated simulations under range of trial sizes smaller than BLISS-52
- Latent variable framework matches power of ”standard” logistic analysis of the composite with **~50% fewer patients**
- Whether using the **Frequentist or Bayesian** implementation

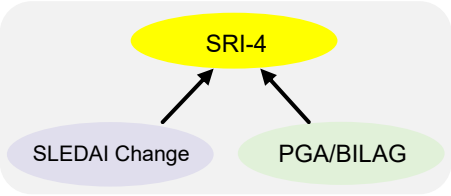


GSK

Post hoc application in two GSK Phase 3 trials

Latent variable analysis provides substantial precision increase

- Analysed 10 mg/kg Belimumab vs placebo effect in BLISS-52 and BLISS-76
- For composite **primary endpoint** SRI-4
- Precision gain corresponds to **45% – 51%** effective sample size saving



Trial	Method	OR	95% CI	SE	Effective trial size reduction
BLISS-52	Logistic	1.59	(1.11, 2.28)	0.18	
	Latent Variable Framework (Frequentist)	1.72	(1.34, 2.21)	0.13	51%
	Latent Variable Framework (Bayesian)	1.72	(1.34, 2.23)	0.13	48%
BLISS-76	Logistic	1.51	(1.03, 2.21)	0.19	
	Latent Variable Framework (Frequentist)	1.42	(1.07, 1.89)	0.14	44%
	Latent Variable Framework (Bayesian)	1.42	(1.06, 1.87)	0.15	45%

Completers only analysis for simplicity due to missing data in components: Results differ from published primary analyses. Work underway on handling intercurrent events.

analyses by **Jasna Cotic**

Summary

- Latent variable framework offers substantial efficiency gains for composite endpoint trials – smaller, faster trials, drugs to patients sooner

- **Robustness simulations** show:
  - Power gain holds under mis-specified marginals and copulas
  - Small type 1 error inflation at lower sample sizes, but preserved for larger trials (n>400) across all scenarios tested
  - Estimator bias is non-trivial when marginals are mis-specified

- Novel **Bayesian implementation** matches precision gains of existing **frequentist implementation**, enabling use at GSK to **design early stage** trials (incorporating into internal quantitative decision making, historical controls etc)

- First application of technique to **late stage trials** (BLISS-52 and BLISS-76 trials of belimumab for SLE) suggest a **~50% trial size reduction** may have been possible

- We are actively exploring application in **several disease areas** to support design of upcoming trials

Caveats:

- Primary endpoint power is not the **sole determinant** of trial size (safety database, secondary endpoint power requirements)
- Not yet known whether **regulators** would **accept this methodology** to design a trial

- **Future work**: More mis-specified DGMs, Intercurrent events, repeat measures, adaptive designs, component -> marginal prior mapping

Disclosures & Acknowledgements

Disclosures and funding

- **PN**, **JC**, **DW**, **LS**, **JB**, **AP** and **DL** are employed by GSK and hold financial equities in GSK. **JW** has received institutional grants from MRC and NIHR, consulting fees from Apollo Therapeutics, Boehringer-Ingelheim, Hemay, Mirobio, Nemysis, Novartis, Orbsen, UCB and Worg (all paid to institution), payment or honoraria to institution from Astrazeneca and Johnson and Johnson, support for EPSRC/GSK iCase PhD studentship from GSK and support for MRC iCase PhD studentship from UCB, and participated on a Data Safety Monitoring/Advisory Board from Medpace (paid to institution)
- Work presented in this presentation was funded by GSK

Acknowledgements

- We thank the co-authors for their contributions: Jasna Cotic, David Whitney, Lindsey Schader, Jane Bentley, Aris Perperoglou, James Wason and Dave Lunn
- We also thank Mike Sweeting, Tom Drury, Anne Hammer and Nicola Scott of GSK Biostatistics; and Ryan Tomlinson, Ciara O'Shea and Deepak Assudani of the GSK belimumab team

GSK