

# Choosing Models for Zero-Heavy, Bounded Outcomes: A Comparison Using WPAI Absenteeism in NHWS

Authors: Adam Jauregui

Affiliations: Oracle Life Sciences, Austin, Texas, United States

## Background

WPAI absenteeism (% work time missed due to health in the past 7 days) is difficult to model because it is bounded (0–100), typically reported as integers, and shows a large mass at 0 with spikes at 50 and 100.

In real-world analyses it is often modeled with negative binomial or zero-inflated negative binomial regressions, despite neither being a true count nor being a positively unbounded outcome. Alternative models for bounded, zero-heavy endpoints may better align with the outcome's structure.

## Objective

Compare the fit and predictive performance of alternative models for WPAI absenteeism (past 7 days), a bounded, zero-heavy outcome. Using train/test splits, assess RMSE, MAE, and predictive R<sup>2</sup> and evaluate binned calibration of predicted versus observed mean absenteeism. For each model, estimate adjusted marginal means by GAD-7 anxiety severity level while controlling for clinically relevant covariates.

## Methods

NHWS respondents with complete WPAI absenteeism data were analyzed. The outcome was WPAI absenteeism (% work time missed due to health in past 7 days; 0–100). Data were randomly split into training and test sets. Five models were fit: Gaussian, negative binomial (NB), zero-inflated NB (ZINB), fractional logit (scaled to 0–1), and ordered beta (scaled to 0–1). Test-set performance was summarized using RMSE, MAE, and predictive R<sup>2</sup>; calibration used binned predicted-versus-observed mean plots. Adjusted marginal means by GAD-7 severity (none / mild / moderate / severe) were estimated controlling for PHQ-9, age, Charlson Comorbidity Index (2011), sex, income, and education.

References:  
1. Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Maechler M, Bolker BM. *GlimmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling*. The R Journal. 2017;9(2):378–400. doi:10.32614/RJ-2017-066.  
2. R Core Team (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.  
3. Reilly MC, Zbrozek AS, Duker EM. *The validity and reproducibility of a work productivity and activity impairment instrument*. *Pharmacoeconomics*. 1993;4(5):353–365. doi:10.2165/00019053-199304050-00006.  
4. Robert L. Spitzer, Kurt Kroenke, Williams JBW, Löwe B. *A brief measure for assessing generalized anxiety disorder: the GAD-7*. *Arch Intern Med*. 2006;166(10):1092–1097. doi:10.1001/archinte.166.10.1092.  
5. Kurt Kroenke, Robert L. Spitzer, Williams JBW. *The PHQ-9: validity of a brief depression severity measure*. *Gen Intern Med*. 2001;16(9):606–613. doi:10.1046/j.1525-1497.2001.016009606.x.  
6. Quan H, Li B, Couris CM, et al. *Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using ICD-10 data*. *Am J Epidemiol*. 2011;173(6):676–682.

## Results

Among **39,080** employed NHWS respondents, **WPAI absenteeism exhibited a strongly zero-heavy and heaped distribution** (69% at 0%, with visible spikes at 50% and 100%; distribution of absenteeism). In test-set comparisons (**Table 1**), ZINB, Gaussian, and fractional logit showed similar predictive accuracy, with RMSE ≈ 19.4, MAE ≈ 12.5, and predictive R<sup>2</sup> ≈ 0.19–0.20. Negative binomial performed notably worse (RMSE 21.17; R<sup>2</sup> 0.044), while ordered beta performed poorest overall (RMSE 23.46; MAE 20.14; R<sup>2</sup> <0).

Table 1 Model metrics test-set predictive performance by model

| Model             | RMSE  | MAE   | R2    |
|-------------------|-------|-------|-------|
| Zero-Inflated NB  | 19.37 | 12.46 | 0.20  |
| Gaussian          | 19.41 | 12.75 | 0.20  |
| Fractional logit  | 19.44 | 12.51 | 0.19  |
| Negative binomial | 21.17 | 13.08 | 0.04  |
| Ordered beta      | 23.46 | 20.14 | -0.17 |

Worst performance

Density overlays stratified by observed outcome highlighted practical differences not captured by overall metrics.

In the observed-zero subgroup, ZINB, fractional logit, and NB produced tightly concentrated low predictions, whereas Gaussian generated a nontrivial mass of implausible negative predictions; ordered beta shifted predictions upward relative to the other models (**Figure 2, Table 2**).

Figure 2 Test set (observed 0%): model-implied predicted absenteeism

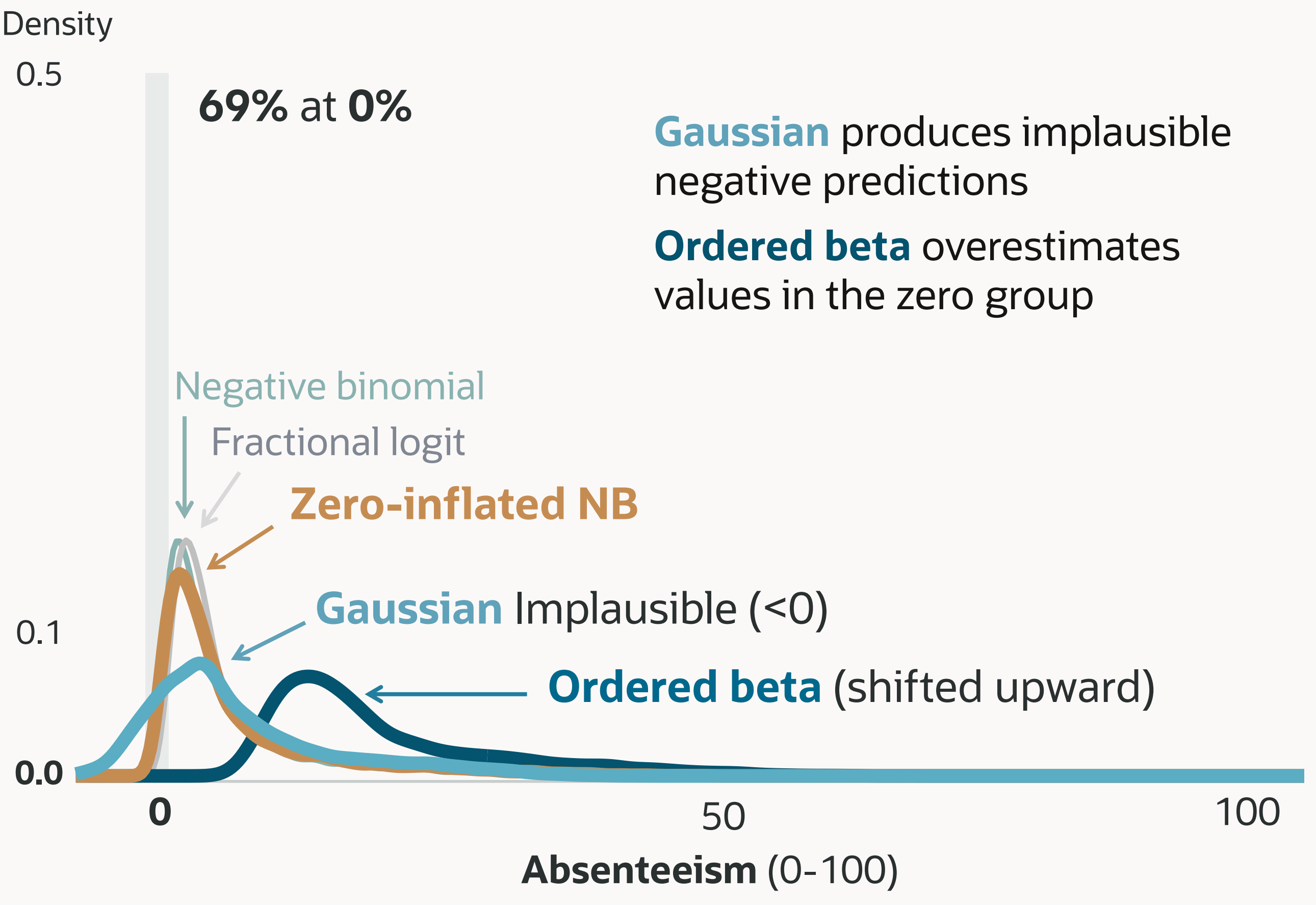


Table 2 Predictions among observed 0% (test set); N=8062

| Model             | Mean  | SE   | Median | Pct outside (0,1) |
|-------------------|-------|------|--------|-------------------|
| Zero-Inflated NB  | 7.77  | 0.08 | 4.88   | 0.00%             |
| Gaussian          | 7.84  | 0.09 | 5.70   | 10.21%            |
| Fractional logit  | 7.80  | 0.08 | 5.04   | 0.00%             |
| Negative Binomial | 8.16  | 0.12 | 4.45   | 0.10%             |
| Ordered beta      | 19.98 | 0.10 | 16.99  | 0.00%             |

In the observed-nonzero subgroup, all models tended to underpredict the mean level of absenteeism, but **ordered beta most closely tracked the nonzero distribution** (**Figure 3, Table 3**).

Figure 3 Test set (observed >0%): model-implied predicted absenteeism

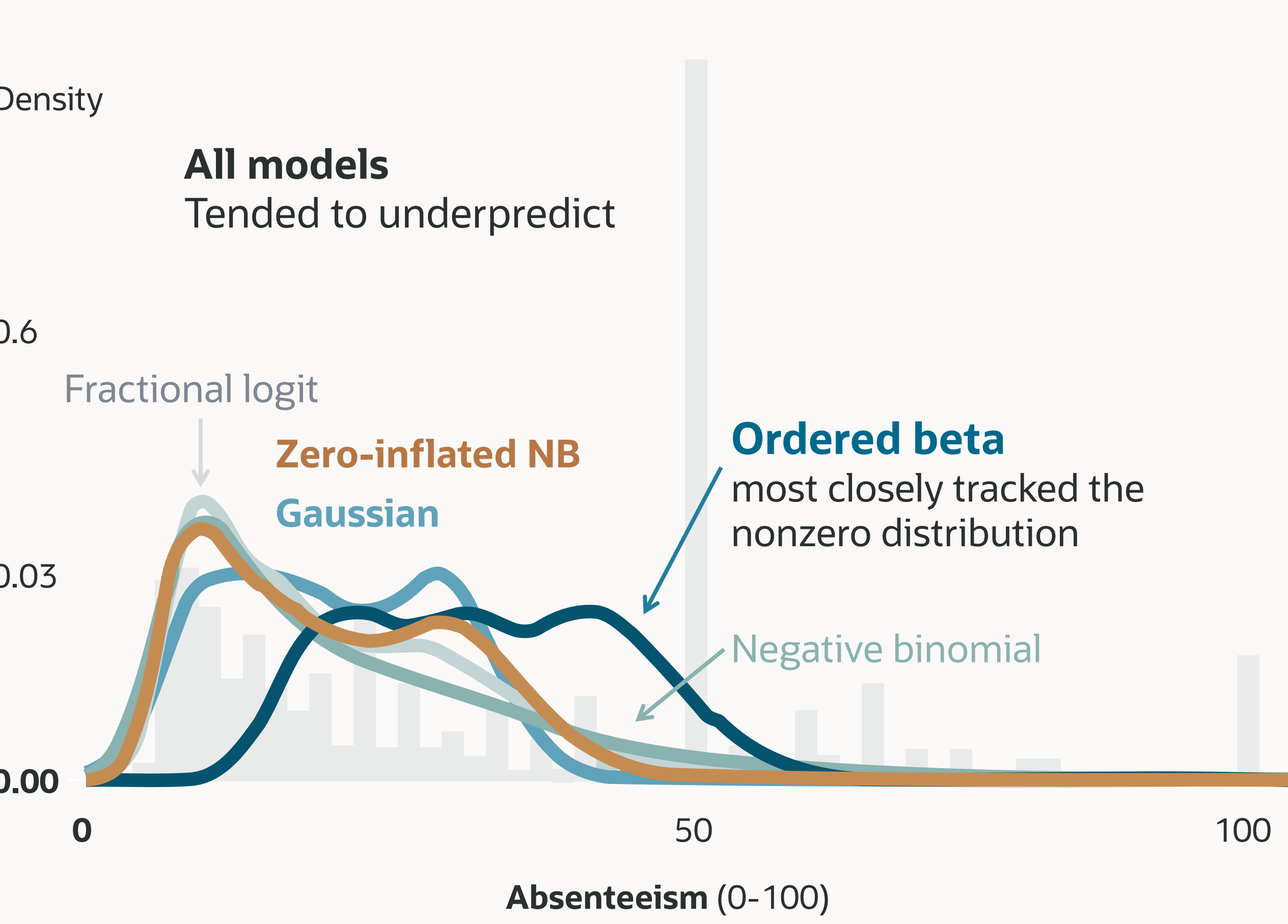


Table 3 Predictions among observed >0% (test set); N=3662

| Variable           | Mean  | SE   | Median | Pct outside (0,1) |
|--------------------|-------|------|--------|-------------------|
| Actual absenteeism | 34.71 | 0.43 | 33     | -                 |
| Zero-Inflated NB   | 18.12 | 0.19 | 16.35  | 0.00%             |
| Gaussian           | 18.04 | 0.17 | 17.53  | 1.26%             |
| Fractional logit   | 18    | 0.2  | 15.31  | 0.00%             |
| Negative Binomial  | 20.93 | 0.31 | 15.3   | 0.76%             |
| Ordered beta       | 32.23 | 0.2  | 31.78  | 0.00%             |

Figure 4 Test set calibration (deciles): mean predicted vs. mean observed absenteeism

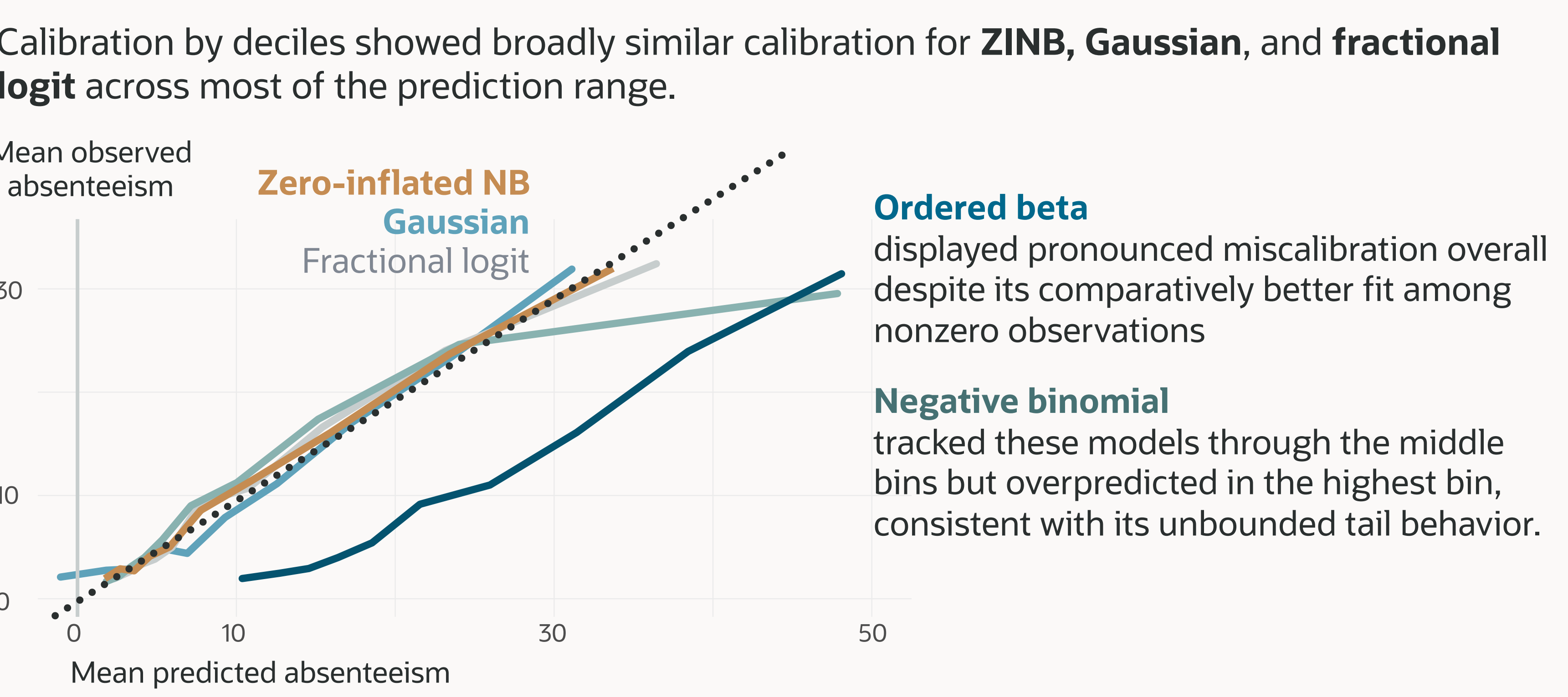
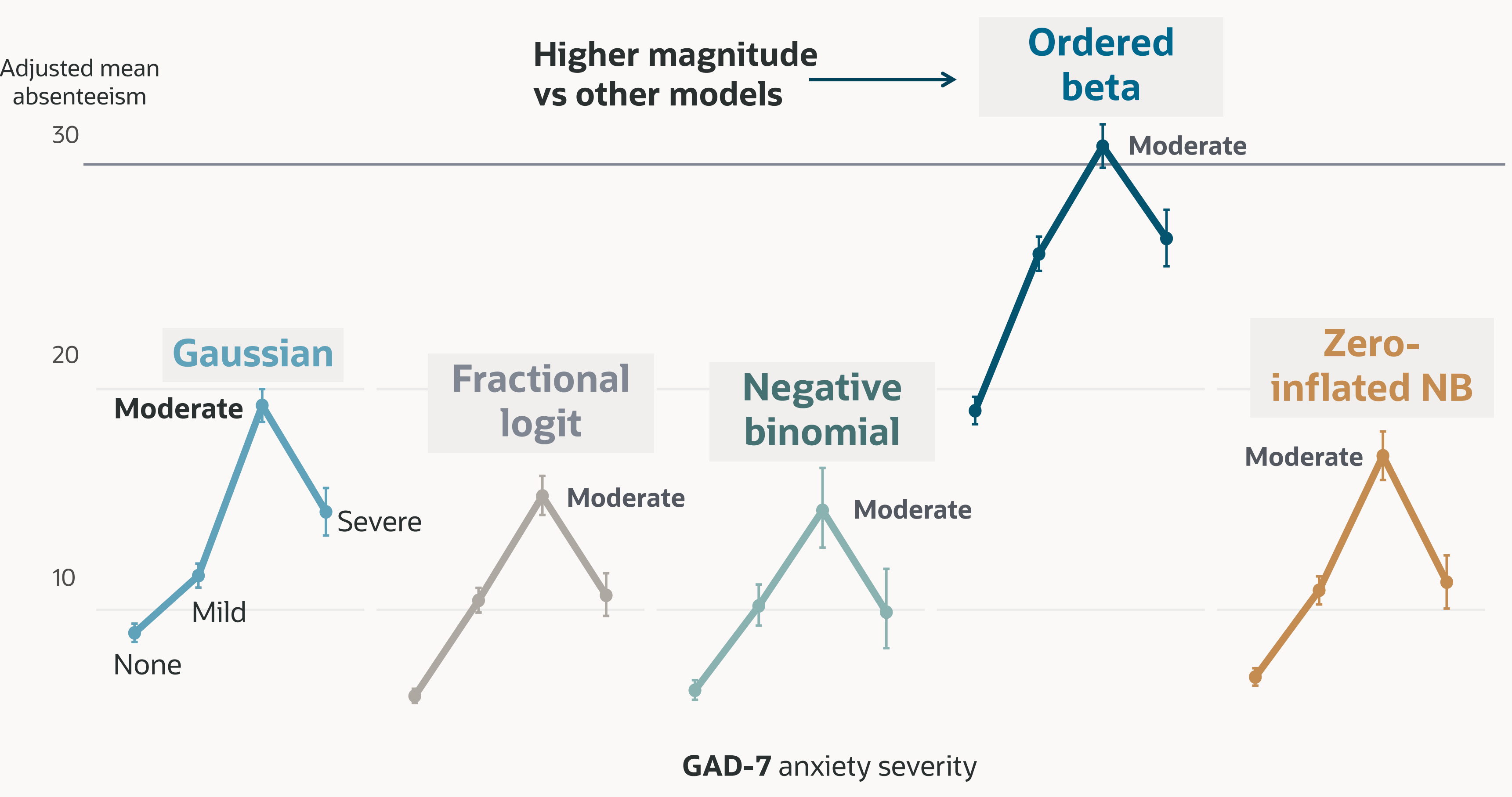


Figure 5 Adjusted mean WPAI absenteeism by GAD-7 severity

Adjusted mean absenteeism increased from none to moderate anxiety and attenuated at severe across all models, with differences in magnitude. **Consistent pattern across models: absenteeism peaks at moderate anxiety**



Note: With large sample size, statistical significance is expected; emphasis is on magnitude and agreement across models.

## Conclusion

WPAI absenteeism is bounded and highly zero-heavy, making model choice consequential. In test-set comparisons, ZINB, Gaussian, and fractional logit performed similarly (RMSE ≈ 19.4; MAE ≈ 12.5; predictive R<sup>2</sup> ≈ 0.19–0.20), while negative binomial and ordered beta performed worse overall.

Diagnostics highlighted practical differences beyond aggregate error: Gaussian produced implausible negative predictions, NB generated >100 values, and ordered beta modeled the nonzero population better than the others. Given comparable accuracy, fractional logit is an attractive, parsimonious bounded-outcome alternative with a single coefficient set and straightforward marginal means, while ZINB may be preferred when a distinct zero-generating mechanism is assumed.

