

Article

A New Approach to Logistic Regression: Using the von Bertalanffy Equation as a Link Function

Kürşad Nuri Baydili ^{1,*}  and Mehmet Gürcan ² 

¹ Department of Biostatistics and Medical Informatics, Hamidiye Faculty of Medicine, University of Health Sciences Turkey, 34668 İstanbul, Türkiye

² Department of Statistics, Faculty of Science, Firat University, 23119 Elazığ, Türkiye

* Correspondence: kursadnuri.baydili@sbu.edu.tr

Abstract

Binary logistic regression relies on the symmetric logit link, whose fixed inflection point cannot adapt to asymmetric response mechanisms and so may yield miscalibrated probabilities when the true response curve is asymmetric—a problem that can be particularly consequential in rare event or imbalanced applications. This study proposes a flexible, asymmetric link derived from the von Bertalanffy growth equation and tests a single, calibration-oriented hypothesis: estimating an asymmetric link improves the calibration of the predicted probabilities under asymmetric response mechanisms while leaving discrimination unchanged. The single shape parameter ν (the reparameterized von Bertalanffy shape, $\nu = m - 1$; the growth curve baseline parameter is absorbed into the intercept and is not separately identifiable) is estimated jointly with the regression coefficients by L2-penalized maximum likelihood, giving a coherent generalized linear model rather than a post hoc re-thresholding of logistic scores; the logistic link is nested exactly at $\nu = 1$ ($m = 2$), so the model reduces to ordinary logistic regression whenever the logit is adequate. The link was benchmarked on identical out-of-sample partitions against four competitors—logistic regression at the 0.5 cut-off, threshold-optimized logistic regression, ridge-penalized logistic regression at the same penalty (which separates the link from the regularization it requires), and Stukel's generalized logistic model—across a correctly specified logistic mechanism and two asymmetric mechanisms (a complementary log–log mechanism and a von Bertalanffy-type best-case reference), three class imbalance ratios (0.10, 0.25, and 0.50) and five sample sizes (100–10,000), for 45,000 iterations in total. Paired differences were summarized by the Hodges–Lehmann median difference, its 95% confidence interval, and the rank-biserial correlation, with Benjamini–Hochberg control of multiplicity. The results supported the hypothesis. At moderate-to-large sample sizes, and increasingly as prevalence approached 0.50, the proposed link achieved lower Brier score and Log–Loss values than the three logistic competitors under both asymmetric mechanisms—for example, a median Log–Loss reduction against the matched-penalty ridge model of about 0.004 at balanced prevalence and the largest sample (95% confidence interval excluding zero; rank-biserial ≈ -1)—an advantage small in absolute size but, at the larger sample sizes, consistent across essentially every replication, and increasing with sample size and prevalence. Discrimination showed no practically important differences at moderate-to-large sample sizes: the area under the ROC curve and overall accuracy were essentially the same across the five methods, so the flexible link improves the quality of the probabilities without degrading class separation. Under the correctly specified logistic mechanism, the method showed no material deterioration, which is consistent with its exact nesting of the logistic model. The contribution is a growth curve-motivated asymmetric link that im-



Academic Editor: Jie Yang

Received: 16 June 2026

Revised: 15 July 2026

Accepted: 16 July 2026

Published: 29 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

proves probability calibration under response asymmetry while preserving discrimination and recovering the logistic model when it is adequate.

Keywords: logistic regression; von Bertalanffy growth curve; flexible link function; probability calibration; class imbalance; penalized maximum likelihood; Brier score

1. Introduction

Multivariate statistical analyses encompass a diverse range of methodologies designed to elucidate the complex interactions among multiple explanatory variables and their collective impact on a specific response variable [1]. Depending on the characteristics of the dependent variable and the distributional properties of the independent variables, numerous multivariate techniques can be employed for both risk factor identification and predictive modelling. While early cornerstones of the classification literature include cluster analysis [2] and discriminant analysis [3], these methods often impose strict distributional assumptions. To overcome these limitations, Berkson [4] introduced logistic regression, a robust technique that shifted the focus toward probabilistic classification without requiring assumptions such as multivariate normality or homogeneity of variances.

Owing to these properties, logistic regression has become a workhorse across many applied fields, including health and fertility research [5,6], geographically referenced modelling [7,8], and optimization-driven model construction [9]; its discriminatory performance has itself been the object of methodological study [10].

However, despite its widespread adoption, the calibration capacity of traditional logistic regression is strictly bound to the structural properties of its underlying link function. Given a linear predictor η of the explanatory variables, the standard logistic equation is written as follows:

$$\pi(\eta) = [1 + \exp(-\eta)]^{-1}$$

Conventionally, if the estimated probability $\pi(\eta)$ is greater than 0.5, the observation is classified as “1”; otherwise, it is classified as “0”. The graph of this logistic function is an S-shaped sigmoid curve, initially utilized by Verhulst [11]. While this symmetric curve facilitates classification, its fixed inflection point forces the probability transitions to be perfectly symmetrical. In real-world multifactor studies—especially those involving rare events or imbalanced classes—this structural symmetry often fails to capture the true underlying asymmetric mechanisms, leading to poorly calibrated risk estimates [12].

The idea of relaxing the fixed shape of the logit is itself well established within the theory of generalized linear models [13], and several parametric families have been proposed specifically to embed the logit within a richer, shape-controlled class of links. Prentice [14] generalized the probit and logit through a family based on the log-F distribution; Aranda-Ordaz [15] introduced one-parameter symmetric and asymmetric transformation families that contain the logit as a limiting case; Pregibon [16] developed goodness-of-link diagnostics together with a two-parameter link that perturbs the logit locally; and Stukel [17] proposed a generalized logistic model in which two shape parameters independently modify the two tails of the curve, with the probit and complementary log–log links arising as special cases. These proposals share the premise that a mis-specified link degrades both discrimination and calibration and that estimating shape parameters jointly with the regression coefficients can recover the lost performance.

The present study is in the same spirit but adopts a different construction. Rather than perturbing the logit algebraically, it replaces the symmetric logit with a link obtained as

the closed-form solution of the von Bertalanffy growth differential equation [18], yielding a single interpretable shape parameter m that governs the asymmetry of the approach to the upper asymptote and reduces exactly to the logistic link at $m = 2$. Its contribution relative to the existing flexible link literature is therefore a growth curve-motivated, monotone, and numerically stable asymmetric link whose single shape parameter carries a direct substantive interpretation and is estimated within a penalized-likelihood generalized linear model. The generalized link function used for coefficient estimation is:

$$\pi(\eta) = [1 - c \cdot \exp(-\eta)]^{1/(1-m)}, c = 1 - y_0^{(1-m)}$$

Here, $m > 1$ is a shape parameter and y_0 a growth curve baseline; as shown in Section 2.1, in the presence of a regression intercept, the baseline is absorbed and the response is governed by a single identified shape parameter, $v = m - 1$, which reduces exactly to the logistic link at $v = 1$. Over its admissible range, the map is strictly monotone in η and bounded in $(0, 1)$, with $\pi(\eta) \rightarrow 1$ as $\eta \rightarrow \infty$ and $\pi(\eta) \rightarrow 0$ as $\eta \rightarrow -\infty$, so it is a valid probability response. Figure 1 illustrates how the shape parameter deforms the curve relative to the symmetric logistic link.

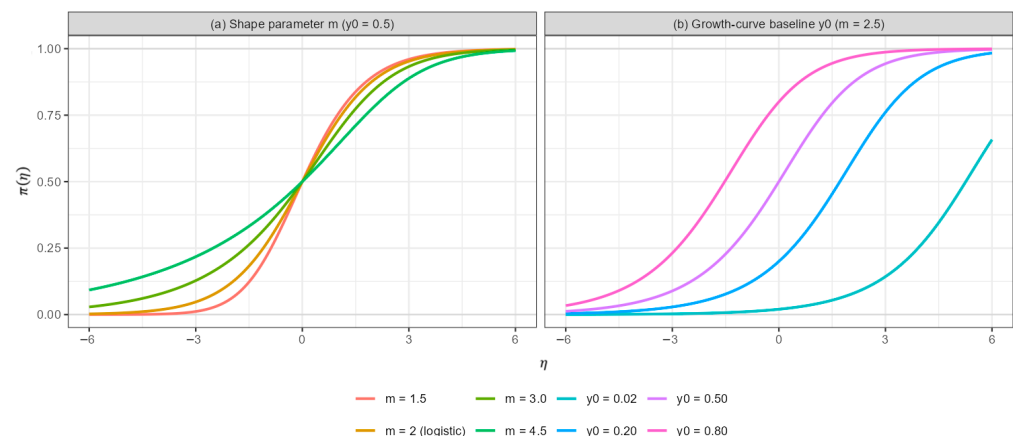


Figure 1. The von Bertalanffy link function compared with the standard logistic link, plotted as the fitted probability $\pi(\eta)$ against the linear predictor η . Panel (a) varies the shape parameter (larger values give a more asymmetric approach to the upper asymptote); panel (b) shows that varying the growth curve baseline merely translates the curve horizontally—a shift that a regression intercept absorbs. In every case, the curve is monotone and bounded within $(0, 1)$, and the standard logistic link is recovered exactly at $v = 1$ ($m = 2$).

The aim of this study is to determine whether estimating an asymmetric link derived from the von Bertalanffy growth curve improves the calibration of the predicted probabilities when the response mechanism is asymmetric while preserving discrimination and doing no harm when the ordinary logit is adequate. The study is deliberately framed around calibration. We neither expect nor claim gains in discrimination—in the area under the ROC curve or in overall accuracy—because a monotone reparameterization of the link cannot, on its own, be expected to improve the separation of the classes; the objective is better-calibrated probabilities, with discrimination expected to remain essentially unchanged because the proposed response is a smooth, monotone transformation of the linear predictor. Calibration is the primary target because, in the applications that motivate flexible links—clinical risk prediction, screening, credit scoring, and reliability analysis—decisions are driven by the estimated risk itself and not only by the ranking of cases [12,19]. To ensure that any calibration advantage is attributable to the link rather than to the mild regularization it requires, the proposed link is benchmarked not only against ordinary logistic regression at the 0.5 cut-off and at an accuracy-optimized threshold, but

also against a ridge-penalized logistic regression carrying an identical penalty and against Stukel's generalized logistic model as the closest flexible link competitor. All evaluation is strictly out-of-sample and is summarized with effect sizes and confidence intervals rather than p -values alone.

2. Materials and Methods

2.1. An Asymmetric Inverse Link from the von Bertalanffy Growth Curve

The proposed model replaces the symmetric logit with an inverse-link (response) function derived from the von Bertalanffy growth curve [18,20]. In its growth curve form, with $\eta = \beta_0 + \mathbf{x}^T \boldsymbol{\beta}$ as the linear predictor, the response is $\pi(\eta) = [1 - c \cdot \exp(-\eta)]^{(1/(1-m))}$ with $c = 1 - y_0^{(1-m)}$, $m > 1$, and $0 < y_0 < 1$. Writing $A = y_0^{(1-m)} - 1 > 0$ gives the equivalent form $\pi(\eta) = [1 + A \cdot \exp(-\eta)]^{(-1/(m-1))}$, and because $A \cdot \exp(-\eta) = \exp(-(\eta - \log A))$, the factor A is absorbed into the intercept through $\beta_0 \rightarrow \beta_0 - \log A$. The growth curve baseline parameter y_0 is therefore not separately identifiable in the presence of an intercept: any value of y_0 is reproduced exactly by an intercept shift, so it carries no independent interpretation, and the identified model has a single shape parameter, $\nu = m - 1 > 0$. We accordingly work throughout with the identified one-parameter response

$$\pi(\eta) = [1 + \exp(-\eta)]^{-1/\nu}, \nu > 0.$$

Differentiating this response yields the growth differential equation it satisfies,

$$d\pi/d\eta = \pi(1 - \pi^\nu)/\nu,$$

Therefore, at $\nu = 1$, the equation reduces to the logistic form $d\pi/d\eta = \pi(1 - \pi)$, and the response is the symmetric logit. The logistic model is thus nested exactly at the single interior point $\nu = 1$ (equivalently $m = 2$) and is recovered whenever the logit is adequate. As ν increases, the curve approaches its upper asymptote more gradually; as $\nu \rightarrow 0^+$, it tends toward an extreme value (Gompertz-type) shape related to the complementary log–log link. The family therefore moves smoothly between the symmetric logit and a range of asymmetric forms while remaining a valid probability response that is monotonic in η and bounded in $(0, 1)$, providing the flexibility that a fixed symmetric link lacks. This one-parameter asymmetric response belongs to the generalized logistic (Richards-type) family that also contains the Gompertz and von Bertalanffy growth curves; the contribution here is its penalized binary regression implementation and its evaluation for probability calibration under link mis-specification, not a new curve family.

2.2. Estimation

The parameters—the regression coefficients $\boldsymbol{\beta}$ (including the intercept) and the shape parameter ν —are estimated by minimizing the L2-penalized negative log-likelihood

$$-\ell(\boldsymbol{\beta}, \nu) + \lambda \sum_{j \geq 1} \beta_j^2,$$

With the penalty $\lambda = 0.05$ applied to the slope coefficients only; the intercept is not penalized, and no parameter is tuned to any classification metric. The ridge term is a mild, conventional device to prevent coefficient inflation and complete separation divergence in small samples; its value was fixed a priori at a small level rather than tuned, and a sensitivity check over $\lambda \in \{0.01, 0.05, 0.10\}$ left the qualitative conclusions unchanged. Optimization used the L-BFGS-B algorithm (up to 500 iterations, at the default convergence tolerance) from several starting values for the shape parameter, together with the ordinary logistic fit as the starting value for $\boldsymbol{\beta}$, retaining the fit with the lowest penalized objective. In the

implementation, the growth curve parameters (m, y_0) were carried explicitly; because y_0 is absorbed into the intercept (Section 2.1), this is equivalent to estimating the identified one-parameter model, and every reported performance metric—which depends on the fitted probabilities alone—is invariant to the reparameterization. The redundant baseline direction does leave a flat ridge in the objective along which the optimizer can drift; this contributes to the small sample boundary contact reported in Section 3.5 and is a property of the redundant parameterization rather than of the identified model, further motivating the one-parameter form.

2.3. Simulation Design

Each simulated data set contained four predictors—two continuous (a lognormal and an exponential variable) and two binary Bernoulli variables—and a binary response. The continuous predictors were standardized using the training fold mean and standard deviation to ensure numerical stability during penalized estimation. Their marginal distributions and true coefficients are given in Table 1. In each replication, the intercept was determined by root-finding on the marginal outcome mean so that the mean true event probability equalled the target prevalence for every mechanism. The 70/30 split within each replication was random rather than stratified; replications yielding a single-class training or test fold were discarded and not replaced, so the most extreme rare event cells rest on slightly fewer than 1000 replications, and non-converged fits (rare) were likewise excluded from the summaries.

Table 1. Predictor variables and true coefficients used in the simulation.

Variable	Distribution	Type	True Coefficient
X_1	Lognormal (0, 1)	Continuous	1.5
X_2	Exponential (rate = 1)	Continuous	−2.0
X_3	Bernoulli (0.5)	Binary	0.8
X_4	Bernoulli (0.5)	Binary	−1.2

Three data-generating mechanisms (DGPs) are reported in the main text; a fourth, Stukel-type mechanism was also evaluated and is provided in full in the deposited output (see the Data Availability Statement). The first is a correctly specified symmetric logistic mechanism, under which the ordinary logit is the true model and nothing but the operating threshold can be improved; it serves as the null (“do-no-harm”) benchmark. The remaining two are asymmetric. The complementary log–log mechanism is a standard, independent asymmetric link that is not one of the fitted models, and therefore provides a neutral test of whether a flexible link helps under a generic asymmetry it does not match. The von Bertalanffy-type mechanism (shape $v = 3.5$, equivalently $m = 4.5$) is the proposed link’s own family and is retained only as a best-case reference—a setting in which the proposed model is correctly specified by construction—and is labelled as such throughout. Simulations were run across three class imbalance ratios (prevalences 0.10, 0.25, and 0.50) and five sample sizes (100, 500, 1000, 5000, and 10,000), giving 45 scenarios and, at 1000 replications each, 45,000 iterations in total (Table 2).

Table 2. Design of the simulation study.

Data-generating mechanisms	Logistic (null); complementary log–log (asymmetric); von Bertalanffy-type, $m = 4.5$, $y_0 = 0.02$ (asymmetric, best-case reference)
Class priors (prevalence)	0.10, 0.25, 0.50

Table 2. Cont.

Sample sizes (n)	100, 500, 1000, 5000, 10,000
Replications per scenario	1000
Total scenarios/iterations	45/45,000
Train/test split	70%/30%, per replication
Estimation	L2-penalized MLE ($\lambda = 0.05$, slopes only); L-BFGS-B, multi-start
Evaluation	All metrics computed on the held-out 30% test fold

2.4. Competing Methods

Five classifiers were evaluated on identical training and test partitions. All model fitting—the coefficients and shape parameters of the flexible link, the ridge and Stukel coefficients, and the optimized threshold—used the training fold only, and every reported metric was computed on the untouched 30% test fold. Throughout, the suffix “05” marks a method evaluated at the fixed 0.5 probability cut-off, and “Opt” marks a threshold optimized on the training fold; the intercept is never penalized in any of the penalized fits.

LR05: This classifier is the ordinary logistic regression at the conventional 0.5 cut-off.

LROpt: This classifier is the ordinary logistic regression whose classification threshold was chosen to maximize accuracy on the training fold—the optimal accuracy threshold, as distinct from a sensitivity-plus-specificity (Youden) criterion—and then applied to the test fold.

Ridge05: This classifier is the L2-penalized (ridge) logistic regression at the same penalty as the proposed model ($\lambda = 0.05$, slopes only), evaluated at the 0.5 cut-off. This baseline is the crux of the comparison: because the flexible link is itself ridge-penalized, contrasting it with an unpenalized logistic model would confound the link with the penalty. Ridge05 holds the penalty fixed, so any residual difference is attributable to the link alone. To keep the penalty on an identical scale, Ridge05 was fitted by the same penalized-likelihood optimization as the proposed model rather than by a package with a different internal scaling.

Stukel05: This classifier is Stukel’s [17] generalized logistic model, whose two shape parameters modify the tails independently fitted by the same penalized maximum likelihood. It is the closest existing flexible link competitor and is included so that the question of what the proposed formulation adds beyond the established literature can be answered empirically rather than only in prose.

Bert05: This classifier is the proposed generalized asymmetric link, estimated by L2-penalized maximum likelihood as described in Section 2.2 and evaluated at the standard 0.5 cut-off, so that its threshold-dependent metrics reflect the link itself rather than any post hoc threshold adjustment.

Contrasting LROpt with LR05 isolates the pure effect of threshold tuning, which leaves the ranking of observations—and hence the AUC, Brier score, and Log-Loss—unchanged. Contrasting Bert05 with Ridge05 isolates the effect of the link at a fixed penalty. Because a monotone re-thresholding of a fixed logistic score cannot change the AUC, any AUC difference between Bert05 and the logistic models is direct evidence that the fitted model, not merely its operating point, has changed.

2.5. Evaluation Metrics

Model performance was assessed along two complementary dimensions—discrimination and calibration—using metrics computed exclusively on the held-out test fold. Discrimination, the ability to separate the two classes, was quantified at the 0.5 cut-off by overall accuracy (the

proportion of correctly classified observations), balanced accuracy (the mean of sensitivity and specificity, which corrects the optimism of overall accuracy under class imbalance), sensitivity and specificity (the true-positive and true-negative rates), and the F1-score (the harmonic mean of precision and recall, a summary suited to imbalanced problems in which the positive class is of primary interest) [10]. Because all of these depend on the chosen cut-off, they were complemented by the area under the receiver operating characteristic curve (AUC): the probability that a randomly chosen positive case receives a higher predicted risk than a randomly chosen negative case [21]. As a threshold-independent, rank-based measure, the AUC is unaffected by relocating the operating point and therefore serves as a clear diagnostic of whether the intrinsic ordering of observations has changed—the property that separates a genuine change in the fitted model from a mere shift of the classification threshold.

Calibration, the agreement between predicted probabilities and observed event frequencies, was evaluated by the Brier score and the logarithmic loss (Log–Loss). The Brier score is the mean squared difference between the predicted probability and the binary outcome [22]; it is a strictly proper scoring rule and admits the classical decomposition into a calibration (reliability) component and a discrimination (refinement) component [23], so that an improvement in it under fixed discrimination isolates an improvement in calibration. The Log–Loss, or cross-entropy, is the negative mean log-likelihood of the observed outcomes under the predicted probabilities; it too is strictly proper [24] and penalizes confident but incorrect predictions especially severely, making it sensitive to exactly the over- and under-confidence that a mis-specified link induces. Calibration is a distinct and, in many applied settings, important property: a model can rank cases correctly yet assign systematically distorted probabilities, and in domains such as clinical risk prediction, diagnostic screening, credit scoring, and reliability analysis it is the estimated probability itself—the risk on which a treatment, loan, or maintenance decision is based—and not merely the ranking that governs the quality of the decision [12,19]. Because the Brier score and Log–Loss are threshold-free, they cannot be improved by re-thresholding, so they are both the appropriate place to seek a genuine benefit of a flexible link and the metrics on which the proposed link showed its advantage. Discrimination and calibration are separable dimensions, and a model may improve one while leaving the other essentially unchanged [12,19].

2.6. Statistical Analysis

For each metric and scenario, the proposed link was compared with each competitor over the 1000 replications using the paired Wilcoxon signed-rank test on the per-replication differences. Marginal performance is reported as the median and interquartile range (IQR) of each method. Because marginal medians do not reveal the paired effect—the median of the differences is not the difference of the medians—the paired contrast between the proposed link and the matched-penalty ridge baseline is additionally summarized, for every scenario, by the Hodges–Lehmann median paired difference [25], its 95% confidence interval, and the rank-biserial correlation [26] as a standardized effect size; Figures extend these paired comparisons to all four competitors.

Two safeguards govern interpretation. First, with 1000 paired replications, the signed-rank test has very high power and will flag negligible differences as significant [27]; significance is therefore read jointly with the magnitude of the Hodges–Lehmann difference and the rank-biserial correlation, and a significant but numerically trivial difference is treated only as evidence of a consistent operational direction. Second, given the large number of contrasts, the *p*-values were adjusted for multiplicity by the Benjamini–Hochberg false-discovery-rate procedure [28]; the Bonferroni-adjusted values were also computed and led to the same qualitative conclusions. All computations were carried out in R 4.3.3 [29];

logistic models used the glm routine, the ROC and AUC used the pROC package, and the penalized von Bertalanffy and Stukel models were fitted by numerical maximization of the penalized log-likelihood with the L-BFGS-B algorithm. The complete code and the fixed random seed are provided (see the Data Availability Statement) so that all results are exactly reproducible. Cells at $n = 100$ and prevalence 0.10 are dominated by complete-separation instability (see Section 3.4) and are reported for completeness only; they are not interpreted as evidence for any method.

3. Results

Before interpreting the magnitudes, there is an internal check, visible throughout Tables 3–11, that verifies the assessment is behaving correctly: in each scenario, LR05 and LROpt yield similar AUC, Brier score, and Log–Loss values, differing only in threshold-dependent metrics (accuracy, balanced accuracy, sensitivity, specificity, and F1 score). This is exactly what must happen if threshold optimization relocates the operating point without changing the ranking or the predicted probabilities. Against this backdrop, the results are organized around a single question: once the penalty is held fixed by the ridge baseline, where does the flexible link genuinely help? The answer is calibration.

3.1. Marginal Performance Across Scenarios

Tables 3–11 report the median (IQR) of all eight metrics for the five methods, grouped by mechanism and prevalence. Under the two asymmetric mechanisms, the proposed link (Bert05) most often holds the best Brier score and Log–Loss values, while discrimination metrics are near-identical across methods; under the logistic mechanism, the five methods are interchangeable up to negligible, offsetting differences.

Under the von Bertalanffy-type mechanism (Tables 3–5), where the proposed model is correctly specified by construction, Bert05 holds the best or joint-best Brier score and Log–Loss values in the majority of cells, and its margin over the three logistic variants widens as prevalence moves toward 0.50 and as n grows. On the threshold-dependent metrics, the pattern is an exchange rather than a dominance: relative to LR05 and Ridge05, Bert05 is typically slightly more specific and slightly less sensitive, so balanced accuracy and F1-score differ only marginally, and LROpt—free to move its cut-off—recovers most of the sensitivity that the fixed 0.5 threshold costs Bert05. Stukel05 tracks Bert05 closely throughout, as expected of a second flexible link, and the AUC is effectively tied across all five methods at every sample size.

Table 3. The von Bertalanffy-type (asymmetric) mechanism—best-case reference, class prior = 0.10.

Metric	Bert05	LR05	LROpt	Ridge05	Stukel05
$n = 100$					
AUC	76.79 (25.37)	75.86 (26.64)	75.86 (26.64)	76.79 (28.11)	76.00 (27.93)
Acc	90.00 (6.67)	90.00 (6.67)	90.00 (6.67)	90.00 (6.67)	90.00 (6.67)
BalAcc	60.58 (25.00)	60.00 (25.00)	63.46 (25.00)	60.00 (24.00)	60.00 (23.21)
Sens	25.00 (50.00)	25.00 (50.00)	33.33 (50.00)	20.00 (50.00)	25.00 (50.00)
Spec	100.00 (3.70)	100.00 (3.70)	96.43 (7.14)	100.00 (3.57)	100.00 (3.70)
F1	33.33 (57.14)	33.33 (57.14)	40.00 (57.14)	33.33 (57.14)	33.33 (50.00)
Brier	0.075 (0.057)	0.076 (0.060)	0.076 (0.060)	0.073 (0.059)	0.079 (0.060)
Log–Loss	0.301 (0.266)	0.300 (0.253)	0.300 (0.253)	0.280 (0.200)	0.350 (0.642)

Table 3. Cont.

Metric	Bert05	LR05	LR0pt	Ridge05	Stukel05
n = 500					
AUC	78.91 (9.90)	78.89 (9.88)	78.89 (9.88)	78.93 (9.78)	78.66 (10.58)
Acc	92.00 (3.33)	92.00 (3.33)	92.00 (3.33)	92.00 (3.33)	92.00 (2.67)
BalAcc	63.33 (8.55)	63.51 (8.57)	64.64 (9.23)	63.33 (8.55)	62.50 (8.33)
Sens	28.57 (16.36)	28.57 (16.84)	30.77 (18.57)	27.78 (16.36)	25.00 (15.69)
Spec	99.26 (1.47)	99.26 (1.48)	99.23 (2.21)	99.26 (1.47)	99.28 (1.42)
F1	40.00 (20.00)	40.00 (19.42)	42.11 (19.96)	40.00 (19.42)	38.10 (19.43)
Brier	0.067 (0.023)	0.067 (0.022)	0.067 (0.022)	0.067 (0.022)	0.068 (0.024)
Log-Loss	0.248 (0.073)	0.248 (0.072)	0.248 (0.072)	0.248 (0.072)	0.262 (0.100)
n = 1000					
AUC	79.35 (6.93)	79.33 (6.87)	79.33 (6.87)	79.33 (6.90)	79.23 (6.93)
Acc	92.33 (2.00)	92.33 (2.00)	92.17 (2.33)	92.33 (2.00)	92.33 (2.33)
BalAcc	63.78 (6.01)	63.81 (6.03)	64.10 (6.50)	63.78 (6.00)	62.91 (5.98)
Sens	28.12 (11.79)	28.57 (11.69)	29.17 (13.03)	28.12 (11.79)	26.47 (11.76)
Spec	99.26 (0.76)	99.26 (0.76)	99.25 (1.12)	99.26 (0.76)	99.62 (0.76)
F1	41.67 (13.91)	41.86 (13.98)	42.11 (14.71)	41.86 (13.89)	40.00 (13.80)
Brier	0.066 (0.016)	0.066 (0.016)	0.066 (0.016)	0.066 (0.016)	0.066 (0.016)
Log-Loss	0.243 (0.050)	0.244 (0.050)	0.244 (0.050)	0.244 (0.050)	0.253 (0.064)
n = 5000					
AUC	79.41 (3.18)	79.40 (3.10)	79.40 (3.10)	79.40 (3.10)	79.39 (3.14)
Acc	92.27 (0.93)	92.27 (0.93)	92.20 (0.93)	92.27 (0.93)	92.27 (0.93)
BalAcc	63.78 (2.53)	63.96 (2.63)	63.97 (3.03)	63.95 (2.63)	63.41 (2.66)
Sens	28.23 (5.13)	28.61 (5.18)	28.66 (6.21)	28.57 (5.21)	27.45 (5.34)
Spec	99.33 (0.30)	99.27 (0.37)	99.32 (0.58)	99.27 (0.37)	99.41 (0.31)
F1	42.05 (5.95)	42.31 (6.02)	42.34 (6.71)	42.29 (6.00)	41.31 (6.13)
Brier	0.066 (0.007)	0.066 (0.007)	0.066 (0.007)	0.066 (0.007)	0.066 (0.007)
Log-Loss	0.244 (0.022)	0.246 (0.022)	0.246 (0.022)	0.246 (0.022)	0.245 (0.022)
n = 10,000					
AUC	79.52 (2.12)	79.50 (2.14)	79.50 (2.14)	79.50 (2.14)	79.49 (2.13)
Acc	92.23 (0.67)	92.20 (0.67)	92.20 (0.63)	92.20 (0.67)	92.20 (0.67)
BalAcc	63.77 (1.81)	64.02 (1.76)	63.81 (2.25)	64.02 (1.76)	63.60 (1.90)
Sens	28.24 (3.62)	28.77 (3.64)	28.38 (4.60)	28.77 (3.64)	27.86 (3.84)
Spec	99.34 (0.26)	99.29 (0.25)	99.34 (0.37)	99.29 (0.25)	99.37 (0.26)
F1	42.05 (4.28)	42.46 (4.15)	42.04 (5.05)	42.46 (4.12)	41.75 (4.43)
Brier	0.066 (0.005)	0.066 (0.005)	0.066 (0.005)	0.066 (0.005)	0.066 (0.005)
Log-Loss	0.243 (0.015)	0.245 (0.015)	0.245 (0.015)	0.245 (0.015)	0.244 (0.016)

Values are the median (interquartile range) over 1000 test fold replication. LR05: logistic regression at the 0.5 cut-off; LR0pt: logistic regression at the optimal accuracy threshold; Ridge05: ridge-penalized logistic regression ($\lambda = 0.05$) at the 0.5 cut-off; Stukel05: Stukel's generalized logistic model; Bert05: the proposed von Bertalanffy link. AUC: area under the ROC curve; Acc: accuracy; BalAcc: balanced accuracy; Sens: sensitivity; Spec: specificity; F1: F1-score; Brier: Brier score.

Table 4. The von Bertalanffy-type (asymmetric) mechanism—best-case reference, class prior = 0.25.

Metric	Bert05	LR05	LROpt	Ridge05	Stukel05
n = 100					
AUC	76.80 (15.94)	76.80 (15.52)	76.80 (15.52)	76.92 (15.58)	75.94 (16.24)
Acc	80.00 (13.33)	80.00 (13.33)	80.00 (10.00)	80.00 (13.33)	80.00 (10.00)
BalAcc	65.00 (14.00)	65.00 (14.07)	66.00 (13.07)	64.83 (14.09)	64.29 (13.43)
Sens	37.50 (25.00)	37.50 (25.00)	40.00 (29.87)	36.36 (27.14)	33.33 (27.78)
Spec	95.24 (10.00)	95.24 (10.00)	92.00 (10.44)	95.45 (9.52)	95.45 (9.09)
F1	46.15 (26.67)	46.15 (26.67)	50.00 (23.64)	46.15 (26.67)	44.44 (26.37)
Brier	0.150 (0.064)	0.151 (0.063)	0.151 (0.063)	0.150 (0.062)	0.156 (0.070)
Log–Loss	0.478 (0.192)	0.474 (0.183)	0.474 (0.183)	0.470 (0.173)	0.546 (0.455)
n = 500					
AUC	76.92 (6.12)	76.98 (6.15)	76.98 (6.15)	76.96 (6.22)	76.84 (6.28)
Acc	81.33 (4.67)	81.33 (4.67)	80.67 (4.00)	81.33 (4.67)	80.67 (4.67)
BalAcc	65.78 (5.34)	66.14 (5.45)	65.79 (5.69)	66.12 (5.50)	64.60 (5.53)
Sens	35.56 (11.67)	36.59 (11.03)	36.11 (13.16)	36.36 (11.09)	32.26 (11.41)
Spec	96.43 (3.16)	96.26 (3.02)	96.33 (4.47)	96.30 (2.94)	97.41 (2.98)
F1	48.15 (11.42)	49.06 (11.38)	48.15 (11.71)	48.98 (11.44)	45.66 (12.40)
Brier	0.142 (0.025)	0.142 (0.025)	0.142 (0.025)	0.142 (0.025)	0.143 (0.025)
Log–Loss	0.448 (0.067)	0.449 (0.065)	0.449 (0.065)	0.449 (0.064)	0.459 (0.084)
n = 1000					
AUC	77.88 (4.11)	77.90 (4.08)	77.90 (4.08)	77.90 (4.09)	77.82 (4.10)
Acc	81.33 (3.00)	81.00 (3.00)	81.00 (3.33)	81.00 (3.00)	81.00 (3.00)
BalAcc	66.02 (3.92)	66.34 (3.94)	66.04 (4.42)	66.28 (3.91)	65.16 (3.88)
Sens	35.62 (7.97)	36.76 (8.09)	35.62 (10.16)	36.62 (8.11)	33.33 (8.51)
Spec	96.60 (2.10)	96.19 (2.07)	96.51 (3.01)	96.23 (2.07)	97.33 (2.06)
F1	48.65 (7.99)	49.18 (7.96)	48.62 (9.15)	49.15 (7.96)	46.91 (8.60)
Brier	0.140 (0.018)	0.141 (0.018)	0.141 (0.018)	0.140 (0.018)	0.141 (0.018)
Log–Loss	0.441 (0.044)	0.442 (0.047)	0.442 (0.047)	0.442 (0.047)	0.445 (0.048)
n = 5000					
AUC	77.89 (1.95)	77.86 (1.98)	77.86 (1.98)	77.86 (1.98)	77.90 (1.94)
Acc	81.33 (1.40)	81.23 (1.40)	81.20 (1.48)	81.27 (1.40)	81.33 (1.40)
BalAcc	65.80 (1.73)	66.31 (1.71)	65.75 (2.01)	66.30 (1.71)	65.66 (1.80)
Sens	34.69 (3.58)	36.44 (3.63)	34.77 (4.96)	36.44 (3.61)	34.41 (3.75)
Spec	96.85 (0.88)	96.24 (0.88)	96.87 (1.56)	96.25 (0.88)	96.95 (0.91)
F1	48.24 (3.61)	49.24 (3.47)	48.11 (4.30)	49.25 (3.46)	47.93 (3.73)
Brier	0.140 (0.008)	0.140 (0.008)	0.140 (0.008)	0.140 (0.008)	0.140 (0.008)
Log–Loss	0.439 (0.021)	0.441 (0.021)	0.441 (0.021)	0.441 (0.021)	0.439 (0.021)

Table 4. *Cont.*

Metric	Bert05	LR05	LROpt	Ridge05	Stukel05
n = 10,000					
AUC	77.90 (1.33)	77.87 (1.35)	77.87 (1.35)	77.87 (1.35)	77.89 (1.32)
Acc	81.30 (0.97)	81.27 (1.00)	81.27 (0.97)	81.27 (1.00)	81.30 (1.00)
BalAcc	65.72 (1.25)	66.26 (1.20)	65.73 (1.50)	66.26 (1.20)	65.70 (1.28)
Sens	34.61 (2.43)	36.39 (2.39)	34.67 (3.69)	36.39 (2.39)	34.59 (2.67)
Spec	96.87 (0.60)	96.20 (0.66)	96.85 (1.20)	96.21 (0.65)	96.88 (0.62)
F1	48.03 (2.62)	49.20 (2.46)	48.04 (3.17)	49.19 (2.46)	48.00 (2.69)
Brier	0.139 (0.006)	0.140 (0.005)	0.140 (0.005)	0.140 (0.005)	0.139 (0.006)
Log–Loss	0.438 (0.014)	0.440 (0.014)	0.440 (0.014)	0.440 (0.014)	0.438 (0.014)

Values are the median (interquartile range) over 1000 test fold replications. LR05: logistic regression at the 0.5 cut-off; LROpt: logistic regression at the optimal accuracy threshold; Ridge05: ridge-penalized logistic regression ($\lambda = 0.05$) at the 0.5 cut-off; Stukel05: Stukel's generalized logistic model; Bert05: the proposed von Bertalanffy link. AUC: area under the ROC curve; Acc: accuracy; BalAcc: balanced accuracy; Sens: sensitivity; Spec: specificity; F1: F1-score; Brier: Brier score.

Table 5. The von Bertalanffy-type (asymmetric) mechanism—best-case reference, class prior = 0.50.

Metric	Bert05	LR05	LROpt	Ridge05	Stukel05
n = 100					
AUC	78.22 (12.76)	78.28 (12.73)	78.28 (12.73)	78.28 (12.65)	77.78 (13.56)
Acc	70.00 (10.00)	70.00 (13.33)	70.00 (10.00)	70.00 (13.33)	70.00 (10.00)
BalAcc	69.86 (11.79)	69.91 (12.17)	69.00 (12.50)	69.68 (12.14)	68.08 (11.48)
Sens	69.23 (22.11)	69.23 (21.80)	70.59 (25.21)	69.23 (21.80)	63.16 (26.26)
Spec	71.43 (21.25)	71.43 (20.71)	70.00 (26.10)	71.43 (20.71)	76.92 (23.86)
F1	68.97 (14.14)	68.97 (13.46)	68.97 (13.95)	68.97 (13.46)	66.67 (15.00)
Brier	0.196 (0.056)	0.197 (0.056)	0.197 (0.056)	0.196 (0.055)	0.205 (0.062)
Log–Loss	0.573 (0.149)	0.574 (0.145)	0.574 (0.145)	0.570 (0.139)	0.648 (0.651)
n = 500					
AUC	79.21 (4.71)	79.07 (4.59)	79.07 (4.59)	79.11 (4.60)	79.15 (4.70)
Acc	71.33 (4.67)	70.67 (5.33)	70.67 (4.67)	70.67 (5.33)	70.67 (5.33)
BalAcc	71.16 (4.96)	70.81 (4.98)	70.46 (4.78)	70.83 (4.89)	70.72 (5.05)
Sens	67.13 (9.47)	68.83 (9.08)	67.07 (15.51)	68.83 (9.05)	64.71 (10.29)
Spec	75.00 (9.33)	73.21 (8.78)	75.00 (13.50)	73.24 (8.71)	77.40 (10.28)
F1	69.80 (5.94)	70.03 (6.00)	69.21 (6.85)	70.07 (6.00)	68.66 (6.65)
Brier	0.186 (0.020)	0.186 (0.020)	0.186 (0.020)	0.186 (0.019)	0.186 (0.020)
Log–Loss	0.542 (0.052)	0.545 (0.050)	0.545 (0.050)	0.545 (0.050)	0.547 (0.058)
n = 1000					
AUC	79.63 (3.46)	79.54 (3.42)	79.54 (3.42)	79.54 (3.42)	79.61 (3.48)
Acc	71.33 (3.33)	71.00 (4.00)	71.00 (3.67)	71.00 (4.00)	71.33 (3.67)
BalAcc	71.32 (3.57)	71.20 (3.79)	71.10 (3.68)	71.18 (3.73)	71.26 (3.60)
Sens	66.67 (6.69)	69.13 (6.70)	66.67 (10.52)	69.13 (6.67)	65.70 (7.12)

Table 5. Cont.

Metric	Bert05	LR05	LROpt	Ridge05	Stukel05
Spec	76.32 (6.45)	73.78 (6.52)	75.54 (10.22)	73.78 (6.58)	77.38 (6.91)
F1	69.75 (4.53)	70.47 (4.23)	69.55 (4.96)	70.47 (4.23)	69.41 (4.55)
Brier	0.183 (0.014)	0.184 (0.014)	0.184 (0.014)	0.184 (0.014)	0.183 (0.014)
Log–Loss	0.536 (0.036)	0.539 (0.036)	0.539 (0.036)	0.539 (0.036)	0.539 (0.038)
n = 5000					
AUC	79.61 (1.49)	79.57 (1.47)	79.57 (1.47)	79.58 (1.47)	79.62 (1.48)
Acc	71.53 (1.53)	71.40 (1.60)	71.40 (1.60)	71.40 (1.60)	71.47 (1.55)
BalAcc	71.51 (1.52)	71.40 (1.57)	71.41 (1.58)	71.40 (1.56)	71.50 (1.54)
Sens	65.85 (3.11)	69.03 (2.91)	66.05 (6.37)	69.03 (2.92)	65.89 (3.08)
Spec	77.00 (2.91)	73.73 (2.96)	76.95 (6.00)	73.74 (2.96)	76.95 (3.01)
F1	69.78 (1.93)	70.69 (1.87)	69.77 (2.59)	70.69 (1.87)	69.80 (1.91)
Brier	0.182 (0.006)	0.183 (0.006)	0.183 (0.006)	0.183 (0.006)	0.182 (0.006)
Log–Loss	0.534 (0.015)	0.537 (0.015)	0.537 (0.015)	0.537 (0.015)	0.535 (0.016)
n = 10,000					
AUC	79.57 (1.10)	79.50 (1.12)	79.50 (1.12)	79.50 (1.12)	79.57 (1.11)
Acc	71.47 (1.11)	71.30 (1.20)	71.33 (1.13)	71.28 (1.20)	71.43 (1.07)
BalAcc	71.46 (1.07)	71.29 (1.19)	71.34 (1.13)	71.29 (1.19)	71.47 (1.05)
Sens	65.74 (2.17)	68.97 (2.10)	65.76 (5.08)	68.96 (2.11)	65.97 (2.19)
Spec	77.10 (2.01)	73.64 (2.07)	76.92 (4.86)	73.64 (2.08)	76.85 (2.06)
F1	69.72 (1.30)	70.58 (1.31)	69.60 (1.99)	70.59 (1.31)	69.79 (1.34)
Brier	0.183 (0.005)	0.183 (0.005)	0.183 (0.005)	0.183 (0.005)	0.183 (0.005)
Log–Loss	0.535 (0.012)	0.538 (0.012)	0.538 (0.012)	0.538 (0.012)	0.535 (0.012)

Values are the median (interquartile range) over 1000 test fold replications. LR05: logistic regression at the 0.5 cut-off; LROpt: logistic regression at the optimal accuracy threshold; Ridge05: ridge-penalized logistic regression ($\lambda = 0.05$) at the 0.5 cut-off; Stukel05: Stukel's generalized logistic model; Bert05: the proposed von Bertalanffy link. AUC: area under the ROC curve; Acc: accuracy; BalAcc: balanced accuracy; Sens: sensitivity; Spec: specificity; F1: F1-score; Brier: Brier score.

Under the complementary log–log mechanism (Tables 6–8)—an asymmetry the proposed link does not itself generate, and therefore the more demanding test—the same calibration ordering holds: Bert05 attains the lowest Brier score and Log–Loss values among the logistic competitors, most clearly at prevalence 0.50 and large n , while remaining level with Stukel05. Discrimination is again indistinguishable across methods, so the benefit is confined to the quality of the probabilities rather than their ordering. The $n = 100$ /prevalence 0.10 cells are an exception to be read with caution: there the logistic models show interquartile ranges an order of magnitude larger than elsewhere (for example, a ridge Log–Loss IQR above 2), which is the signature of complete separation in a training fold of roughly seventy observations with about seven events, so the apparent gap in those cells reflects instability rather than a dependable advantage.

Table 6. Complementary log–log (asymmetric) mechanism, class prior = 0.10.

Metric	Bert05	LR05	LROpt	Ridge05	Stukel05
n = 100					
AUC	98.77 (6.17)	96.55 (13.46)	96.55 (13.46)	91.51 (50.00)	98.15 (12.50)
Acc	96.67 (3.33)	93.33 (6.67)	93.33 (6.67)	93.33 (6.67)	96.67 (3.33)
BalAcc	84.62 (25.00)	84.62 (24.93)	87.50 (23.08)	66.67 (37.50)	83.65 (24.75)
Sens	75.00 (50.00)	75.00 (50.00)	75.00 (50.00)	33.33 (75.00)	75.00 (50.00)
Spec	100.00 (3.70)	100.00 (3.85)	96.43 (7.14)	100.00 (0.00)	100.00 (3.85)
F1	75.00 (31.75)	66.67 (35.71)	66.67 (35.71)	50.00 (80.00)	75.00 (35.71)
Brier	0.035 (0.040)	0.042 (0.040)	0.042 (0.040)	0.053 (0.072)	0.039 (0.047)
Log–Loss	0.120 (0.187)	0.252 (0.679)	0.252 (0.679)	0.221 (2.206)	0.173 (0.732)
n = 500					
AUC	97.47 (3.62)	97.47 (3.47)	97.47 (3.47)	97.52 (3.48)	97.28 (3.70)
Acc	96.00 (2.00)	96.00 (2.00)	96.00 (2.67)	96.00 (2.00)	96.00 (2.00)
BalAcc	86.37 (7.96)	86.44 (8.35)	86.89 (9.06)	85.93 (8.27)	84.58 (8.93)
Sens	75.00 (16.82)	75.00 (17.35)	75.00 (18.33)	73.33 (16.54)	70.00 (17.84)
Spec	98.57 (1.43)	98.56 (1.44)	98.54 (1.51)	99.23 (0.82)	99.26 (1.48)
F1	80.00 (12.27)	80.00 (12.27)	78.79 (11.95)	80.00 (12.99)	78.26 (13.19)
Brier	0.029 (0.015)	0.029 (0.015)	0.029 (0.015)	0.029 (0.015)	0.029 (0.015)
Log–Loss	0.110 (0.064)	0.108 (0.060)	0.108 (0.060)	0.106 (0.056)	0.122 (0.104)
n = 1000					
AUC	97.21 (2.38)	97.22 (2.38)	97.22 (2.38)	97.22 (2.36)	97.19 (2.36)
Acc	96.33 (1.33)	96.33 (1.33)	96.33 (1.67)	96.33 (1.33)	96.33 (1.67)
BalAcc	86.20 (5.94)	86.36 (5.97)	86.11 (6.38)	86.24 (5.85)	84.43 (6.74)
Sens	73.43 (11.66)	74.07 (12.14)	73.33 (13.33)	73.53 (11.59)	69.57 (13.70)
Spec	98.89 (0.78)	98.88 (0.79)	98.88 (1.12)	98.89 (0.78)	99.28 (0.74)
F1	80.00 (7.88)	80.00 (7.96)	79.41 (8.33)	80.00 (7.77)	78.69 (9.17)
Brier	0.029 (0.010)	0.029 (0.010)	0.029 (0.010)	0.029 (0.010)	0.029 (0.011)
Log–Loss	0.105 (0.038)	0.105 (0.039)	0.105 (0.039)	0.105 (0.037)	0.114 (0.056)
n = 5000					
AUC	97.08 (1.03)	97.09 (1.02)	97.09 (1.02)	97.09 (1.03)	97.09 (1.04)
Acc	96.33 (0.73)	96.33 (0.68)	96.33 (0.67)	96.33 (0.67)	96.40 (0.73)
BalAcc	85.89 (2.61)	86.31 (2.60)	85.72 (3.14)	86.29 (2.65)	85.11 (2.68)
Sens	72.73 (5.19)	73.77 (5.15)	72.44 (6.48)	73.63 (5.26)	71.04 (5.39)
Spec	99.03 (0.38)	98.89 (0.44)	99.05 (0.59)	98.90 (0.44)	99.20 (0.38)
F1	80.14 (3.75)	80.14 (3.79)	79.73 (3.88)	80.14 (3.76)	79.55 (3.85)
Brier	0.028 (0.005)	0.028 (0.005)	0.028 (0.005)	0.028 (0.005)	0.028 (0.005)
Log–Loss	0.104 (0.017)	0.104 (0.017)	0.104 (0.017)	0.104 (0.017)	0.104 (0.017)

Table 6. *Cont.*

Metric	Bert05	LR05	LR0pt	Ridge05	Stukel05
n = 10,000					
AUC	97.05 (0.71)	97.05 (0.72)	97.05 (0.72)	97.05 (0.71)	97.05 (0.71)
Acc	96.40 (0.47)	96.37 (0.43)	96.37 (0.47)	96.37 (0.43)	96.40 (0.47)
BalAcc	85.81 (1.75)	86.35 (1.73)	85.66 (2.15)	86.32 (1.74)	85.35 (1.83)
Sens	72.58 (3.60)	73.75 (3.49)	72.26 (4.49)	73.73 (3.53)	71.53 (3.72)
Spec	99.04 (0.29)	98.92 (0.27)	99.08 (0.44)	98.92 (0.27)	99.18 (0.26)
F1	80.20 (2.47)	80.39 (2.43)	80.07 (2.46)	80.39 (2.44)	80.00 (2.43)
Brier	0.028 (0.003)	0.028 (0.003)	0.028 (0.003)	0.028 (0.003)	0.028 (0.003)
Log–Loss	0.103 (0.011)	0.104 (0.011)	0.104 (0.011)	0.104 (0.011)	0.103 (0.011)

Values are the median (interquartile range) over 1000 test fold replications. LR05: logistic regression at the 0.5 cut-off; LR0pt: logistic regression at the optimal accuracy threshold; Ridge05: ridge-penalized logistic regression ($\lambda = 0.05$) at the 0.5 cut-off; Stukel05: Stukel's generalized logistic model; Bert05: the proposed von Bertalanffy link. AUC: area under the ROC curve; Acc: accuracy; BalAcc: balanced accuracy; Sens: sensitivity; Spec: specificity; F1: F1-score; Brier: Brier score.

Table 7. Complementary log–log (asymmetric) mechanism, class prior = 0.25.

Metric	Bert05	LR05	LR0pt	Ridge05	Stukel05
n = 100					
AUC	93.65 (8.34)	93.60 (8.16)	93.60 (8.16)	93.65 (8.48)	92.61 (9.34)
Acc	90.00 (10.00)	86.67 (10.00)	86.67 (6.67)	86.67 (10.00)	86.67 (10.00)
BalAcc	83.33 (12.55)	83.33 (12.51)	83.33 (12.35)	81.37 (12.50)	81.37 (12.50)
Sens	71.43 (26.19)	71.43 (26.19)	75.00 (25.71)	66.67 (24.44)	70.00 (27.78)
Spec	95.45 (9.09)	95.24 (9.52)	94.74 (12.12)	95.65 (8.33)	95.45 (9.09)
F1	75.00 (16.67)	75.00 (16.67)	75.00 (16.44)	72.73 (21.79)	72.73 (19.36)
Brier	0.086 (0.051)	0.089 (0.056)	0.089 (0.056)	0.085 (0.051)	0.091 (0.054)
Log–Loss	0.295 (0.216)	0.301 (0.210)	0.301 (0.210)	0.278 (0.156)	0.380 (0.671)
n = 500					
AUC	93.74 (3.19)	93.74 (3.21)	93.74 (3.21)	93.77 (3.19)	93.76 (3.22)
Acc	89.33 (3.33)	89.33 (3.33)	89.33 (4.00)	89.33 (3.33)	89.33 (3.33)
BalAcc	83.41 (5.34)	83.59 (5.41)	83.34 (5.62)	83.52 (5.37)	81.72 (5.41)
Sens	71.43 (10.20)	71.43 (10.58)	71.05 (12.04)	71.43 (10.24)	66.35 (11.27)
Spec	95.74 (3.04)	95.61 (3.10)	95.80 (3.52)	95.73 (2.88)	97.39 (2.07)
F1	77.14 (7.71)	77.19 (7.22)	76.92 (7.52)	77.30 (7.31)	76.06 (7.85)
Brier	0.077 (0.021)	0.078 (0.021)	0.078 (0.021)	0.077 (0.021)	0.078 (0.021)
Log–Loss	0.255 (0.065)	0.255 (0.066)	0.255 (0.066)	0.255 (0.063)	0.272 (0.095)
n = 1000					
AUC	93.76 (2.24)	93.73 (2.24)	93.73 (2.24)	93.72 (2.25)	93.75 (2.23)
Acc	89.67 (2.33)	89.67 (2.33)	89.67 (2.33)	89.67 (2.00)	89.67 (2.33)
BalAcc	83.47 (3.88)	83.78 (3.75)	83.07 (4.15)	83.69 (3.74)	82.12 (4.09)
Sens	70.70 (7.98)	71.52 (7.72)	69.89 (9.94)	71.25 (7.50)	67.08 (8.41)
Spec	96.13 (2.05)	95.89 (2.04)	96.38 (2.82)	95.96 (1.94)	97.33 (1.83)

Table 7. *Cont.*

Metric	Bert05	LR05	LROpt	Ridge05	Stukel05
F1	77.47 (5.38)	77.61 (5.44)	76.82 (5.56)	77.59 (5.20)	76.39 (5.73)
Brier	0.076 (0.015)	0.076 (0.015)	0.076 (0.015)	0.076 (0.014)	0.076 (0.015)
Log–Loss	0.252 (0.045)	0.253 (0.047)	0.253 (0.047)	0.253 (0.046)	0.255 (0.049)
n = 5000					
AUC	93.68 (0.97)	93.69 (0.97)	93.69 (0.97)	93.69 (0.97)	93.69 (0.97)
Acc	89.80 (1.13)	89.73 (1.13)	89.73 (1.13)	89.73 (1.13)	89.80 (1.13)
BalAcc	83.09 (1.78)	83.57 (1.78)	82.99 (2.03)	83.56 (1.77)	82.69 (1.84)
Sens	69.74 (3.66)	71.27 (3.55)	69.28 (4.87)	71.23 (3.48)	68.53 (3.78)
Spec	96.49 (0.91)	95.88 (0.90)	96.64 (1.50)	95.89 (0.88)	96.90 (0.90)
F1	77.27 (2.51)	77.60 (2.39)	77.17 (2.59)	77.59 (2.41)	77.02 (2.58)
Brier	0.076 (0.007)	0.076 (0.007)	0.076 (0.007)	0.076 (0.007)	0.076 (0.007)
Log–Loss	0.250 (0.019)	0.252 (0.019)	0.252 (0.019)	0.252 (0.019)	0.250 (0.019)
n = 10,000					
AUC	93.74 (0.72)	93.73 (0.73)	93.73 (0.73)	93.73 (0.73)	93.74 (0.72)
Acc	89.83 (0.77)	89.77 (0.80)	89.77 (0.77)	89.77 (0.80)	89.83 (0.74)
BalAcc	83.06 (1.21)	83.66 (1.20)	83.00 (1.50)	83.66 (1.21)	82.82 (1.25)
Sens	69.58 (2.56)	71.43 (2.28)	69.46 (3.70)	71.39 (2.31)	68.77 (2.53)
Spec	96.60 (0.61)	95.88 (0.60)	96.67 (1.08)	95.89 (0.60)	96.86 (0.61)
F1	77.32 (1.78)	77.77 (1.74)	77.24 (1.85)	77.77 (1.77)	77.15 (1.78)
Brier	0.075 (0.005)	0.076 (0.005)	0.076 (0.005)	0.076 (0.005)	0.075 (0.005)
Log–Loss	0.249 (0.014)	0.251 (0.014)	0.251 (0.014)	0.251 (0.014)	0.249 (0.014)

Values are the median (interquartile range) over 1000 test fold replications. LR05: logistic regression at the 0.5 cut-off; LROpt: logistic regression at the optimal accuracy threshold; Ridge05: ridge-penalized logistic regression ($\lambda = 0.05$) at the 0.5 cut-off; Stukel05: Stukel's generalized logistic model; Bert05: the proposed von Bertalanffy link. AUC: area under the ROC curve; Acc: accuracy; BalAcc: balanced accuracy; Sens: sensitivity; Spec: specificity; F1: F1-score; Brier: Brier score.

Table 8. Complementary log–log (asymmetric) mechanism, class prior = 0.50.

Metric	Bert05	LR05	LROpt	Ridge05	Stukel05
n = 100					
AUC	92.89 (6.39)	92.87 (6.39)	92.87 (6.39)	92.89 (6.63)	92.55 (6.94)
Acc	83.33 (6.67)	83.33 (10.00)	83.33 (10.00)	83.33 (6.67)	83.33 (10.00)
BalAcc	83.33 (8.54)	83.48 (9.72)	83.33 (8.82)	83.48 (9.11)	83.33 (8.81)
Sens	85.71 (14.66)	84.62 (15.84)	84.62 (17.31)	84.62 (15.36)	81.82 (17.14)
Spec	83.33 (16.67)	84.62 (15.38)	84.21 (17.86)	84.62 (15.38)	86.67 (13.33)
F1	83.87 (9.74)	83.33 (9.97)	82.76 (10.08)	83.33 (9.97)	82.76 (10.58)
Brier	0.115 (0.051)	0.116 (0.054)	0.116 (0.054)	0.114 (0.049)	0.117 (0.057)
Log–Loss	0.356 (0.170)	0.362 (0.178)	0.362 (0.178)	0.350 (0.137)	0.435 (0.598)

Table 8. Cont.

Metric	Bert05	LR05	LR0pt	Ridge05	Stukel05
n = 500					
AUC	93.17 (2.74)	93.17 (2.76)	93.17 (2.76)	93.17 (2.74)	93.11 (2.82)
Acc	84.67 (4.00)	84.67 (4.67)	84.00 (4.00)	84.67 (4.67)	84.67 (4.00)
BalAcc	84.43 (4.33)	84.45 (4.46)	84.33 (4.02)	84.53 (4.43)	84.39 (4.17)
Sens	84.10 (6.44)	84.62 (6.62)	82.81 (8.76)	84.71 (6.62)	80.26 (8.02)
Spec	85.44 (6.27)	84.91 (6.11)	86.80 (8.31)	85.00 (6.23)	88.73 (6.33)
F1	84.28 (4.63)	84.42 (4.65)	83.92 (4.63)	84.43 (4.68)	83.67 (4.69)
Brier	0.107 (0.023)	0.107 (0.023)	0.107 (0.023)	0.107 (0.022)	0.107 (0.023)
Log–Loss	0.333 (0.063)	0.334 (0.065)	0.334 (0.065)	0.333 (0.063)	0.339 (0.079)
n = 1000					
AUC	93.00 (2.01)	93.00 (2.05)	93.00 (2.05)	92.99 (2.04)	92.98 (2.01)
Acc	84.33 (3.00)	84.33 (3.33)	84.33 (3.00)	84.33 (3.33)	84.33 (3.00)
BalAcc	84.36 (2.98)	84.32 (3.11)	84.27 (2.93)	84.32 (3.02)	84.45 (2.90)
Sens	83.10 (4.42)	84.23 (4.36)	82.41 (6.30)	84.21 (4.36)	81.07 (4.82)
Spec	85.90 (4.66)	84.62 (4.57)	86.60 (6.57)	84.62 (4.62)	87.99 (4.64)
F1	84.05 (3.30)	84.19 (3.23)	83.83 (3.20)	84.18 (3.24)	83.87 (3.24)
Brier	0.107 (0.016)	0.108 (0.016)	0.108 (0.016)	0.108 (0.016)	0.108 (0.015)
Log–Loss	0.332 (0.043)	0.335 (0.044)	0.335 (0.044)	0.335 (0.044)	0.333 (0.046)
n = 5000					
AUC	93.11 (0.84)	93.10 (0.84)	93.10 (0.84)	93.10 (0.83)	93.11 (0.83)
Acc	84.60 (1.33)	84.53 (1.33)	84.53 (1.27)	84.53 (1.33)	84.60 (1.33)
BalAcc	84.62 (1.31)	84.53 (1.33)	84.54 (1.29)	84.53 (1.32)	84.62 (1.28)
Sens	82.06 (2.13)	84.26 (1.95)	82.07 (3.43)	84.26 (1.94)	81.68 (2.17)
Spec	87.27 (2.14)	84.81 (2.06)	87.21 (3.47)	84.79 (2.06)	87.65 (2.15)
F1	84.21 (1.46)	84.45 (1.38)	84.11 (1.53)	84.45 (1.37)	84.16 (1.44)
Brier	0.106 (0.007)	0.106 (0.007)	0.106 (0.007)	0.106 (0.007)	0.106 (0.007)
Log–Loss	0.327 (0.019)	0.331 (0.019)	0.331 (0.019)	0.331 (0.019)	0.327 (0.019)
n = 10,000					
AUC	93.09 (0.61)	93.07 (0.60)	93.07 (0.60)	93.07 (0.60)	93.09 (0.61)
Acc	84.57 (0.87)	84.43 (0.90)	84.50 (0.84)	84.43 (0.90)	84.57 (0.87)
BalAcc	84.58 (0.86)	84.44 (0.90)	84.53 (0.86)	84.44 (0.90)	84.58 (0.85)
Sens	81.62 (1.46)	84.14 (1.35)	81.82 (2.75)	84.13 (1.37)	81.59 (1.49)
Spec	87.60 (1.42)	84.82 (1.47)	87.36 (2.85)	84.82 (1.47)	87.65 (1.37)
F1	84.11 (0.91)	84.41 (0.95)	84.06 (0.98)	84.41 (0.96)	84.10 (0.92)
Brier	0.106 (0.005)	0.106 (0.005)	0.106 (0.005)	0.106 (0.005)	0.106 (0.005)
Log–Loss	0.327 (0.013)	0.331 (0.014)	0.331 (0.014)	0.331 (0.014)	0.326 (0.013)

Values are the median (interquartile range) over 1000 test fold replications. LR05: logistic regression at the 0.5 cut-off; LR0pt: logistic regression at the optimal accuracy threshold; Ridge05: ridge-penalized logistic regression ($\lambda = 0.05$) at the 0.5 cut-off; Stukel05: Stukel's generalized logistic model; Bert05: the proposed von Bertalanffy link. AUC: area under the ROC curve; Acc: accuracy; BalAcc: balanced accuracy; Sens: sensitivity; Spec: specificity; F1: F1-score; Brier: Brier score.

Under the correctly specified logistic mechanism (Tables 9–11), the five methods are interchangeable. At moderate-to-large samples, Bert05, Ridge05, LR05, and LROpt differ only in the third decimal of the Brier score and Log–Loss and by at most a few hundredths of a percentage point in the AUC (the $n = 100$ cells being unstable, as elsewhere), and the only systematic differences are the mutually cancelling sensitivity–specificity exchange at the 0.5 cut-off and the artificial penalty that threshold tuning imposes on LROpt under balance. This is the behaviour required by exact nesting: when the logit is the true link, the generalized model returns to it, so no method can achieve a gain and, crucially, the proposed link does not lose.

Table 9. Logistic (linear, correctly specified) mechanism, class prior = 0.10.

Metric	Bert05	LR05	LROpt	Ridge05	Stukel05
n = 100					
AUC	96.55 (9.62)	94.64 (15.73)	94.64 (15.73)	93.14 (49.04)	95.19 (15.26)
Acc	93.33 (6.67)	93.33 (6.67)	93.33 (6.67)	93.33 (6.67)	93.33 (6.67)
BalAcc	81.25 (26.43)	80.00 (26.43)	81.61 (25.00)	73.21 (33.33)	81.48 (26.43)
Sens	66.67 (59.29)	66.67 (60.00)	66.67 (50.00)	50.00 (66.67)	66.67 (50.00)
Spec	100.00 (3.70)	96.55 (3.85)	96.43 (7.14)	100.00 (3.45)	100.00 (3.85)
F1	66.67 (30.00)	66.67 (30.00)	66.67 (30.00)	50.00 (75.00)	66.67 (30.00)
Brier	0.042 (0.041)	0.054 (0.044)	0.054 (0.044)	0.053 (0.048)	0.048 (0.043)
Log–Loss	0.154 (0.211)	0.250 (0.651)	0.250 (0.651)	0.199 (0.660)	0.240 (0.741)
n = 500					
AUC	96.35 (3.72)	96.35 (3.82)	96.35 (3.82)	96.39 (3.77)	96.25 (3.92)
Acc	95.33 (2.67)	95.33 (2.67)	95.33 (2.00)	95.33 (2.67)	95.33 (2.67)
BalAcc	81.98 (9.38)	81.82 (9.29)	83.10 (10.00)	81.54 (9.50)	80.19 (10.35)
Sens	66.67 (19.94)	64.71 (18.75)	68.75 (19.88)	64.29 (18.81)	61.90 (21.43)
Spec	98.53 (1.48)	98.53 (1.49)	98.50 (2.19)	98.54 (1.48)	99.24 (1.55)
F1	72.22 (13.70)	72.73 (13.33)	72.73 (14.24)	72.73 (15.17)	71.43 (14.94)
Brier	0.036 (0.016)	0.036 (0.016)	0.036 (0.016)	0.036 (0.016)	0.037 (0.016)
Log–Loss	0.130 (0.060)	0.129 (0.057)	0.129 (0.057)	0.128 (0.055)	0.150 (0.113)
n = 1000					
AUC	96.23 (2.45)	96.24 (2.44)	96.24 (2.44)	96.24 (2.43)	96.20 (2.46)
Acc	95.33 (1.67)	95.33 (1.67)	95.33 (1.67)	95.33 (1.67)	95.33 (1.67)
BalAcc	82.64 (6.29)	82.60 (6.32)	83.36 (6.96)	82.57 (6.06)	81.12 (6.54)
Sens	66.67 (13.08)	66.67 (13.08)	68.57 (15.04)	66.67 (12.73)	63.64 (13.37)
Spec	98.58 (1.11)	98.56 (1.11)	98.51 (1.49)	98.85 (1.11)	98.91 (0.86)
F1	74.19 (9.12)	74.19 (9.30)	74.07 (9.34)	74.07 (9.08)	73.02 (9.87)
Brier	0.035 (0.011)	0.035 (0.011)	0.035 (0.011)	0.035 (0.011)	0.035 (0.011)
Log–Loss	0.127 (0.040)	0.126 (0.038)	0.126 (0.038)	0.126 (0.038)	0.133 (0.049)
n = 5000					
AUC	96.02 (1.23)	96.01 (1.23)	96.01 (1.23)	96.01 (1.23)	96.02 (1.23)
Acc	95.47 (0.73)	95.47 (0.73)	95.40 (0.73)	95.47 (0.73)	95.47 (0.73)
BalAcc	82.58 (2.81)	82.58 (2.83)	82.86 (3.53)	82.56 (2.86)	82.37 (2.96)

Table 9. *Cont.*

Metric	Bert05	LR05	LROpt	Ridge05	Stukel05
Sens	66.46 (5.65)	66.45 (5.63)	67.12 (7.52)	66.45 (5.69)	66.00 (5.94)
Spec	98.73 (0.46)	98.73 (0.50)	98.61 (0.73)	98.74 (0.50)	98.75 (0.46)
F1	74.45 (4.28)	74.53 (4.23)	74.44 (4.44)	74.49 (4.30)	74.28 (4.41)
Brier	0.035 (0.005)	0.035 (0.005)	0.035 (0.005)	0.035 (0.005)	0.035 (0.005)
Log–Loss	0.127 (0.017)	0.127 (0.018)	0.127 (0.018)	0.126 (0.017)	0.127 (0.018)
n = 10,000					
AUC	96.03 (0.75)	96.03 (0.75)	96.03 (0.75)	96.03 (0.75)	96.03 (0.75)
Acc	95.47 (0.53)	95.47 (0.53)	95.43 (0.53)	95.47 (0.53)	95.47 (0.53)
BalAcc	82.50 (1.87)	82.52 (1.83)	82.57 (2.48)	82.50 (1.85)	82.39 (2.00)
Sens	66.33 (3.80)	66.32 (3.70)	66.50 (5.32)	66.31 (3.68)	66.01 (3.98)
Spec	98.71 (0.33)	98.71 (0.33)	98.67 (0.52)	98.71 (0.33)	98.74 (0.33)
F1	74.43 (2.76)	74.45 (2.69)	74.33 (3.03)	74.45 (2.71)	74.37 (2.71)
Brier	0.035 (0.004)	0.035 (0.004)	0.035 (0.004)	0.035 (0.004)	0.035 (0.004)
Log–Loss	0.126 (0.011)	0.126 (0.012)	0.126 (0.012)	0.126 (0.011)	0.126 (0.011)

Values are the median (interquartile range) over 1000 test fold replications. LR05: logistic regression at the 0.5 cut-off; LROpt: logistic regression at the optimal accuracy threshold; Ridge05: ridge-penalized logistic regression ($\lambda = 0.05$) at the 0.5 cut-off; Stukel05: Stukel's generalized logistic model; Bert05: the proposed von Bertalanffy link. AUC: area under the ROC curve; Acc: accuracy; BalAcc: balanced accuracy; Sens: sensitivity; Spec: specificity; F1: F1-score; Brier: Brier score.

Table 10. Logistic (linear, correctly specified) mechanism, class prior = 0.25.

Metric	Bert05	LR05	LROpt	Ridge05	Stukel05
n = 100					
AUC	91.43 (9.67)	90.99 (9.74)	90.99 (9.74)	91.30 (9.64)	90.40 (10.78)
Acc	86.67 (10.00)	86.67 (10.00)	86.67 (10.00)	86.67 (10.00)	86.67 (10.00)
BalAcc	78.57 (14.15)	78.00 (14.26)	79.19 (14.11)	77.50 (14.29)	77.50 (13.29)
Sens	62.50 (27.78)	62.50 (27.78)	66.67 (29.31)	60.00 (29.55)	62.50 (25.00)
Spec	95.00 (9.64)	95.24 (9.52)	91.67 (9.47)	95.45 (8.70)	95.24 (10.13)
F1	66.67 (20.63)	66.67 (20.63)	70.00 (20.63)	66.67 (22.22)	66.67 (19.78)
Brier	0.103 (0.053)	0.105 (0.058)	0.105 (0.058)	0.103 (0.053)	0.108 (0.057)
Log–Loss	0.333 (0.189)	0.341 (0.185)	0.341 (0.185)	0.327 (0.152)	0.405 (0.713)
n = 500					
AUC	91.45 (3.61)	91.48 (3.55)	91.48 (3.55)	91.48 (3.54)	91.40 (3.71)
Acc	86.67 (4.00)	86.67 (4.00)	86.67 (4.00)	86.67 (4.00)	86.67 (4.00)
BalAcc	79.38 (5.71)	79.31 (5.72)	79.47 (6.10)	79.07 (5.87)	77.95 (6.42)
Sens	63.64 (11.66)	63.30 (11.48)	64.71 (13.85)	63.04 (11.50)	60.00 (13.37)
Spec	95.00 (2.85)	95.17 (2.89)	94.59 (4.19)	95.41 (2.75)	96.19 (2.85)
F1	71.08 (8.41)	70.83 (8.67)	71.05 (8.54)	70.77 (8.33)	69.57 (9.51)
Brier	0.094 (0.023)	0.094 (0.024)	0.094 (0.024)	0.094 (0.024)	0.096 (0.023)
Log–Loss	0.305 (0.070)	0.305 (0.068)	0.305 (0.068)	0.304 (0.067)	0.316 (0.090)

Table 10. Cont.

Metric	Bert05	LR05	LROpt	Ridge05	Stukel05
n = 1000					
AUC	91.60 (2.32)	91.62 (2.34)	91.62 (2.34)	91.63 (2.34)	91.57 (2.34)
Acc	87.33 (2.67)	87.17 (2.33)	87.00 (2.67)	87.33 (2.33)	87.00 (2.33)
BalAcc	79.36 (3.97)	79.25 (4.04)	79.43 (4.32)	79.25 (3.85)	78.78 (4.08)
Sens	63.53 (7.61)	63.26 (7.50)	64.14 (10.08)	63.16 (7.70)	62.12 (8.24)
Spec	95.14 (2.33)	95.26 (2.30)	94.89 (3.37)	95.39 (2.28)	95.59 (2.45)
F1	71.32 (5.89)	71.15 (5.84)	71.13 (6.21)	71.19 (5.83)	70.77 (6.03)
Brier	0.093 (0.014)	0.093 (0.015)	0.093 (0.015)	0.093 (0.015)	0.093 (0.015)
Log–Loss	0.300 (0.041)	0.300 (0.041)	0.300 (0.041)	0.299 (0.041)	0.303 (0.045)
n = 5000					
AUC	91.56 (1.15)	91.56 (1.16)	91.56 (1.16)	91.56 (1.16)	91.56 (1.15)
Acc	87.33 (1.20)	87.30 (1.20)	87.27 (1.13)	87.33 (1.20)	87.33 (1.20)
BalAcc	79.34 (1.84)	79.33 (1.80)	79.41 (2.17)	79.33 (1.81)	79.25 (1.85)
Sens	63.50 (3.72)	63.39 (3.64)	63.79 (5.26)	63.36 (3.63)	63.16 (3.89)
Spec	95.26 (0.95)	95.27 (0.99)	95.16 (1.84)	95.28 (0.96)	95.34 (0.99)
F1	71.34 (2.79)	71.32 (2.73)	71.34 (2.90)	71.32 (2.73)	71.22 (2.79)
Brier	0.092 (0.007)	0.092 (0.007)	0.092 (0.007)	0.092 (0.007)	0.092 (0.007)
Log–Loss	0.298 (0.020)	0.298 (0.020)	0.298 (0.020)	0.298 (0.020)	0.298 (0.020)
n = 10,000					
AUC	91.50 (0.76)	91.50 (0.76)	91.50 (0.76)	91.50 (0.76)	91.50 (0.76)
Acc	87.30 (0.77)	87.30 (0.80)	87.27 (0.83)	87.30 (0.80)	87.30 (0.80)
BalAcc	79.30 (1.23)	79.30 (1.28)	79.38 (1.66)	79.30 (1.27)	79.27 (1.21)
Sens	63.24 (2.52)	63.26 (2.46)	63.34 (3.99)	63.25 (2.44)	63.14 (2.46)
Spec	95.34 (0.68)	95.34 (0.69)	95.30 (1.24)	95.36 (0.68)	95.37 (0.72)
F1	71.38 (1.82)	71.37 (1.80)	71.40 (2.16)	71.37 (1.80)	71.33 (1.84)
Brier	0.092 (0.005)	0.092 (0.005)	0.092 (0.005)	0.092 (0.005)	0.092 (0.005)
Log–Loss	0.298 (0.014)	0.298 (0.014)	0.298 (0.014)	0.298 (0.014)	0.298 (0.014)

Values are the median (interquartile range) over 1000 test fold replications. LR05: logistic regression at the 0.5 cut-off; LROpt: logistic regression at the optimal accuracy threshold; Ridge05: ridge-penalized logistic regression ($\lambda = 0.05$) at the 0.5 cut-off; Stukel05: Stukel's generalized logistic model; Bert05: the proposed von Bertalanffy link. AUC: area under the ROC curve; Acc: accuracy; BalAcc: balanced accuracy; Sens: sensitivity; Spec: specificity; F1: F1-score; Brier: Brier score.

Table 11. Logistic (linear, correctly specified) mechanism, class prior = 0.50.

Metric	Bert05	LR05	LROpt	Ridge05	Stukel05
n = 100					
AUC	88.69 (8.65)	88.69 (8.61)	88.69 (8.61)	88.84 (8.72)	87.96 (8.84)
Acc	80.00 (10.00)	80.00 (10.00)	76.67 (10.00)	80.00 (10.00)	76.67 (10.00)
BalAcc	78.73 (10.15)	79.17 (10.15)	77.78 (10.12)	79.17 (10.15)	77.78 (10.83)
Sens	81.25 (15.28)	80.00 (16.08)	82.35 (20.24)	80.00 (16.23)	78.57 (17.92)
Spec	78.57 (16.48)	80.00 (15.24)	76.92 (19.05)	80.00 (15.24)	80.00 (16.91)

Table 11. Cont.

Metric	Bert05	LR05	LROpt	Ridge05	Stukel05
F1	78.79 (12.12)	78.57 (11.89)	78.95 (11.48)	78.79 (11.95)	77.60 (12.44)
Brier	0.144 (0.055)	0.144 (0.056)	0.144 (0.056)	0.142 (0.051)	0.150 (0.060)
Log–Loss	0.441 (0.166)	0.445 (0.159)	0.445 (0.159)	0.438 (0.142)	0.529 (0.696)
n = 500					
AUC	88.95 (3.56)	88.95 (3.56)	88.95 (3.56)	88.96 (3.55)	88.91 (3.59)
Acc	80.00 (4.67)	80.00 (4.67)	79.33 (4.00)	80.00 (4.67)	79.33 (4.67)
BalAcc	79.76 (4.54)	79.84 (4.49)	79.37 (4.27)	79.83 (4.49)	79.48 (4.43)
Sens	80.30 (7.50)	79.71 (7.29)	80.56 (10.03)	79.71 (7.33)	79.45 (8.43)
Spec	79.43 (7.31)	80.26 (7.00)	79.14 (9.95)	80.26 (7.02)	80.28 (8.00)
F1	79.74 (4.83)	79.72 (4.88)	79.47 (4.92)	79.73 (4.85)	79.47 (4.82)
Brier	0.137 (0.022)	0.137 (0.022)	0.137 (0.022)	0.137 (0.022)	0.138 (0.022)
Log–Loss	0.419 (0.060)	0.418 (0.059)	0.418 (0.059)	0.418 (0.059)	0.425 (0.069)
n = 1000					
AUC	89.10 (2.48)	89.11 (2.50)	89.11 (2.50)	89.11 (2.48)	89.10 (2.47)
Acc	80.00 (3.33)	80.00 (3.33)	79.67 (3.33)	80.00 (3.33)	80.00 (3.33)
BalAcc	79.97 (3.22)	79.97 (3.29)	79.69 (3.26)	79.92 (3.27)	80.00 (3.33)
Sens	80.19 (5.47)	79.62 (5.21)	80.56 (7.73)	79.72 (5.21)	79.67 (5.69)
Spec	80.25 (5.43)	80.54 (5.17)	80.00 (8.25)	80.54 (5.17)	80.54 (5.62)
F1	80.00 (3.56)	79.87 (3.47)	79.76 (3.71)	79.87 (3.48)	79.87 (3.63)
Brier	0.135 (0.016)	0.135 (0.016)	0.135 (0.016)	0.135 (0.016)	0.135 (0.016)
Log–Loss	0.412 (0.044)	0.411 (0.044)	0.411 (0.044)	0.411 (0.044)	0.413 (0.045)
n = 5000					
AUC	89.23 (1.04)	89.23 (1.04)	89.23 (1.04)	89.23 (1.03)	89.23 (1.04)
Acc	80.13 (1.40)	80.13 (1.40)	80.07 (1.40)	80.13 (1.40)	80.13 (1.40)
BalAcc	80.15 (1.34)	80.15 (1.38)	80.08 (1.38)	80.15 (1.38)	80.15 (1.39)
Sens	79.78 (2.30)	79.77 (2.26)	79.73 (4.25)	79.78 (2.28)	79.69 (2.29)
Spec	80.57 (2.32)	80.64 (2.24)	80.55 (4.07)	80.64 (2.24)	80.66 (2.33)
F1	80.08 (1.57)	80.08 (1.54)	80.06 (1.62)	80.08 (1.56)	80.05 (1.57)
Brier	0.134 (0.007)	0.134 (0.007)	0.134 (0.007)	0.134 (0.007)	0.134 (0.007)
Log–Loss	0.408 (0.018)	0.408 (0.018)	0.408 (0.018)	0.408 (0.018)	0.409 (0.018)
n = 10,000					
AUC	89.22 (0.77)	89.23 (0.77)	89.23 (0.77)	89.23 (0.77)	89.22 (0.77)
Acc	80.17 (0.97)	80.20 (0.97)	80.17 (0.98)	80.20 (0.97)	80.17 (1.00)
BalAcc	80.18 (0.98)	80.20 (0.96)	80.15 (0.99)	80.20 (0.96)	80.17 (0.99)
Sens	79.80 (1.69)	79.80 (1.62)	79.87 (3.32)	79.80 (1.62)	79.77 (1.72)
Spec	80.57 (1.73)	80.63 (1.63)	80.61 (3.37)	80.63 (1.62)	80.61 (1.71)
F1	80.09 (1.03)	80.11 (1.03)	80.08 (1.19)	80.11 (1.04)	80.09 (1.05)

Table 11. Cont.

Metric	Bert05	LR05	LROpt	Ridge05	Stukel05
Brier	0.134 (0.005)	0.134 (0.005)	0.134 (0.005)	0.134 (0.005)	0.134 (0.005)
Log–Loss	0.408 (0.013)	0.408 (0.013)	0.408 (0.013)	0.408 (0.013)	0.408 (0.013)

Values are the median (interquartile range) over 1000 test fold replications. LR05: logistic regression at the 0.5 cut-off; LROpt: logistic regression at the optimal accuracy threshold; Ridge05: ridge-penalized logistic regression ($\lambda = 0.05$) at the 0.5 cut-off; Stukel05: Stukel's generalized logistic model; Bert05: the proposed von Bertalanffy link. AUC: area under the ROC curve; Acc: accuracy; BalAcc: balanced accuracy; Sens: sensitivity; Spec: specificity; F1: F1-score; Brier: Brier score.

3.2. Calibration Is Where the Flexible Link Helps

The advantage of the proposed link is concentrated in probabilistic calibration, and it holds against every logistic competitor. For every scenario, Table 12 reports these paired differences in full for the matched-penalty ridge baseline, while Figures 2–4 plot the Hodges–Lehmann median differences, with 95% confidence intervals, against each of the four competitors, for the AUC, Brier score, and Log–Loss. Under both asymmetric mechanisms, the proposed link attained lower Brier score and Log–Loss values than the ordinary logistic regression at the 0.5 cut-off (LR05), the threshold-optimized logistic regression (LROpt), and the matched-penalty ridge model (Ridge05). Because threshold optimization does not alter the predicted probabilities, LR05 and LROpt are calibrated identically, and at large samples the mild ridge penalty becomes negligible, so all three logistic variants converge: at prevalence 0.50 and $n = 10,000$ the flexible link improved Log–Loss by 0.0044 against each of the three logistic variants under the complementary log–log mechanism (Brier 0.0006) and by 0.0030 under the von Bertalanffy-type reference (Brier 0.0007). The effect was not large in any single replication but almost perfectly consistent across them (rank-biserial correlations near -1), and it grew with both sample size and prevalence—smallest under severe imbalance, largest at balance. This is the core result: a systematic, penalty-adjusted gain in the proper scoring rules over logistic regression that cannot be an artefact of re-thresholding because the Brier score and Log–Loss are threshold-free. Since these scores reward both calibration and refinement, and the latter (discrimination) was essentially unchanged (Section 3.3), the gain is attributable to the calibration component.

Table 12. Paired effect sizes, proposed link versus matched-penalty ridge logistic regression (Bert05 – Ridge05).

DGP	Prev	n	AUC	Brier Score	Log–Loss
von Bertalanffy (best-case)	0.10	100	−0.1027 (−0.4121, +0.1365) [−0.04]	+0.0010 * (+0.0007, +0.0012) [+0.33]	+0.0126 * (+0.0090, +0.0172) [+0.45]
	0.10	500	+0.0061 (−0.0172, +0.0334) [+0.02]	−0.0001 * (−0.0001, +0.0000) [−0.11]	−0.0004 * (−0.0006, −0.0002) [−0.16]
	0.10	1000	+0.0138 (−0.0009, +0.0290) [+0.07]	−0.0002 * (−0.0002, −0.0001) [−0.42]	−0.0011 * (−0.0012, −0.0010) [−0.54]
	0.10	5000	+0.0102 * (+0.0035, +0.0168) [+0.11]	−0.0003 * (−0.0003, −0.0002) [−0.81]	−0.0013 * (−0.0014, −0.0012) [−0.89]
	0.10	10,000	+0.0182 * (+0.0132, +0.0232) [+0.26]	−0.0003 * (−0.0003, −0.0003) [−0.94]	−0.0014 * (−0.0015, −0.0014) [−0.98]
	0.25	100	−0.0605 (−0.1944, +0.0105) [−0.07]	+0.0007 * (+0.0004, +0.0009) [+0.21]	+0.0040 * (+0.0028, +0.0053) [+0.30]
	0.25	500	+0.0262 * (+0.0060, +0.0476) [+0.10]	−0.0003 * (−0.0003, −0.0002) [−0.28]	−0.0013 * (−0.0016, −0.0010) [−0.33]
	0.25	1000	+0.0279 * (+0.0151, +0.0408) [+0.16]	−0.0004 * (−0.0005, −0.0003) [−0.51]	−0.0019 * (−0.0021, −0.0017) [−0.65]
	0.25	5000	+0.0367 * (+0.0300, +0.0434) [+0.38]	−0.0005 * (−0.0005, −0.0005) [−0.91]	−0.0023 * (−0.0024, −0.0022) [−0.96]
	0.25	10,000	+0.0414 * (+0.0361, +0.0467) [+0.52]	−0.0005 * (−0.0005, −0.0005) [−0.97]	−0.0023 * (−0.0024, −0.0022) [−0.99]
	0.50	100	−0.0077 (−0.2061, +0.0001) [−0.07]	+0.0009 * (+0.0006, +0.0012) [+0.24]	+0.0036 * (+0.0025, +0.0048) [+0.26]
	0.50	500	+0.0447 * (+0.0267, +0.0633) [+0.17]	−0.0005 * (−0.0006, −0.0004) [−0.38]	−0.0022 * (−0.0025, −0.0019) [−0.46]
	0.50	1000	+0.0495 * (+0.0373, +0.0632) [+0.27]	−0.0005 * (−0.0006, −0.0005) [−0.53]	−0.0025 * (−0.0027, −0.0022) [−0.66]
	0.50	5000	+0.0633 * (+0.0568, +0.0699) [+0.60]	−0.0007 * (−0.0007, −0.0007) [−0.90]	−0.0030 * (−0.0031, −0.0029) [−0.97]
	0.50	10,000	+0.0618 * (+0.0569, +0.0669) [+0.73]	−0.0007 * (−0.0007, −0.0007) [−0.98]	−0.0030 * (−0.0031, −0.0029) [−1.00]

Table 12. Cont.

DGP	Prev	n	AUC	Brier Score	Log-Loss
C-log-log	0.10	100	+24.6894 * (+24.3827, +25.6174) [+0.81]	−0.0159 * (−0.0206, −0.0109) [−0.36]	−0.7094 * (−0.7494, −0.5330) [−0.46]
	0.10	500	−0.0382 * (−0.0522, −0.0186) [−0.26]	+0.0002 * (+0.0001, +0.0002) [+0.32]	+0.0013 * (+0.0010, +0.0016) [+0.42]
	0.10	1000	−0.0005 (−0.0032, +0.0010) [−0.03]	+0.0000 (+0.0000, +0.0000) [−0.03]	+0.0000 (+0.0000, +0.0001) [+0.00]
	0.10	5000	+0.0011 * (+0.0004, +0.0017) [+0.12]	−0.0002 * (−0.0002, −0.0001) [−0.77]	−0.0011 * (−0.0011, −0.0010) [−0.96]
	0.10	10,000	+0.0004 * (+0.0001, +0.0008) [+0.08]	−0.0002 * (−0.0002, −0.0002) [−0.88]	−0.0013 * (−0.0013, −0.0012) [−0.99]
	0.25	100	−0.1818 * (−0.3370, −0.0501) [−0.18]	+0.0018 * (+0.0014, +0.0021) [+0.37]	+0.0143 * (+0.0107, +0.0188) [+0.43]
	0.25	500	−0.0014 (−0.0099, +0.0003) [−0.07]	+0.0000 (+0.0000, +0.0000) [+0.01]	+0.0000 (−0.0002, +0.0001) [−0.02]
	0.25	1000	+0.0042 * (+0.0014, +0.0067) [+0.12]	−0.0002 * (−0.0002, −0.0001) [−0.46]	−0.0012 * (−0.0013, −0.0010) [−0.58]
	0.25	5000	+0.0054 * (+0.0040, +0.0068) [+0.27]	−0.0004 * (−0.0004, −0.0004) [−0.85]	−0.0025 * (−0.0026, −0.0024) [−0.99]
	0.25	10,000	+0.0044 * (+0.0034, +0.0054) [+0.31]	−0.0004 * (−0.0004, −0.0004) [−0.93]	−0.0027 * (−0.0027, −0.0026) [−1.00]
	0.50	100	−0.0060 (−0.1105, −0.0001) [−0.10]	+0.0015 * (+0.0011, +0.0019) [+0.35]	+0.0078 * (+0.0060, +0.0098) [+0.39]
	0.50	500	+0.0000 (−0.0003, +0.0087) [+0.01]	−0.0001 * (−0.0001, −0.0001) [−0.17]	−0.0008 * (−0.0010, −0.0006) [−0.28]
	0.50	1000	+0.0067 * (+0.0044, +0.0111) [+0.16]	−0.0004 * (−0.0005, −0.0004) [−0.60]	−0.0027 * (−0.0029, −0.0025) [−0.83]
	0.50	5000	+0.0087 * (+0.0071, +0.0104) [+0.36]	−0.0006 * (−0.0007, −0.0006) [−0.88]	−0.0042 * (−0.0043, −0.0041) [−1.00]
	0.50	10,000	+0.0080 * (+0.0068, +0.0093) [+0.45]	−0.0006 * (−0.0007, −0.0006) [−0.96]	−0.0044 * (−0.0045, −0.0043) [−1.00]
Logistic (null)	0.10	100	+21.9788 * (+20.5356, +22.9286) [+0.55]	−0.0012 * (−0.0023, −0.0004) [−0.12]	−0.0096 * (−0.0195, −0.0051) [−0.20]
	0.10	500	−0.0070 (−0.0224, +0.0015) [−0.07]	+0.0001 * (+0.0001, +0.0002) [+0.17]	+0.0008 * (+0.0005, +0.0011) [+0.28]
	0.10	1000	+0.0000 (−0.0035, +0.0035) [+0.00]	+0.0000 * (+0.0000, +0.0001) [+0.20]	+0.0005 * (+0.0004, +0.0006) [+0.34]
	0.10	5000	+0.0000 (−0.0005, +0.0003) [−0.01]	+0.0000 * (+0.0000, +0.0000) [+0.18]	+0.0001 * (+0.0001, +0.0001) [+0.30]
	0.10	10,000	+0.0000 (−0.0001, +0.0001) [+0.01]	+0.0000 * (+0.0000, +0.0000) [+0.23]	+0.0000 * (+0.0000, +0.0001) [+0.31]
	0.25	100	−0.0579 (−0.1654, +0.0000) [−0.08]	+0.0011 * (+0.0008, +0.0014) [+0.27]	+0.0076 * (+0.0051, +0.0104) [+0.30]
	0.25	500	−0.0026 (−0.0125, +0.0012) [−0.05]	+0.0001 * (+0.0000, +0.0002) [+0.19]	+0.0006 * (+0.0004, +0.0008) [+0.24]
	0.25	1000	−0.0003 (−0.0035, +0.0027) [−0.01]	+0.0001 * (+0.0001, +0.0001) [+0.17]	+0.0004 * (+0.0003, +0.0005) [+0.30]
	0.25	5000	−0.0004 (−0.0009, +0.0000) [−0.07]	+0.0000 * (+0.0000, +0.0001) [+0.27]	+0.0001 * (+0.0000, +0.0001) [+0.33]
	0.25	10,000	−0.0002 (−0.0004, +0.0000) [−0.06]	+0.0000 * (+0.0000, +0.0001) [+0.20]	+0.0000 * (0.0000, +0.0000) [+0.30]
	0.50	100	+0.0000 (−0.0052, +0.0093) [+0.03]	+0.0007 * (+0.0004, +0.0010) [+0.17]	+0.0036 * (+0.0022, +0.0052) [+0.20]
	0.50	500	−0.0001 (−0.0090, +0.0007) [−0.03]	+0.0002 * (+0.0001, +0.0002) [+0.22]	+0.0010 * (+0.0007, +0.0012) [+0.32]
	0.50	1000	−0.0045 * (−0.0089, −0.0022) [−0.12]	+0.0001 * (+0.0001, +0.0001) [+0.21]	+0.0004 * (+0.0003, +0.0005) [+0.30]
	0.50	5000	−0.0004 (−0.0008, +0.0000) [−0.07]	+0.0000 * (+0.0000, +0.0001) [+0.24]	+0.0001 * (+0.0001, +0.0001) [+0.32]
	0.50	10,000	−0.0001 (−0.0003, +0.0001) [−0.03]	+0.0000 * (+0.0000, +0.0000) [+0.15]	+0.0000 * (0.0000, +0.0000) [+0.27]

Each cell reports the Hodges–Lehmann median paired difference, its 95% confidence interval, and the rank-biserial correlation in square brackets; * denotes significance after Benjamini–Hochberg correction. For the AUC, a positive value favours the proposed link; for the Brier score and Log-Loss, a negative value favours the proposed link. Cells at $n = 100$ and prevalence 0.10 are dominated by complete-separation instability (see Section 3.4) and are reported for completeness only; they are not interpreted as evidence for any method.

Against Stukel’s generalized logistic model the proposed link was neither better nor worse to any material degree: the two coincided on calibration (for instance, Log-Loss differing by about 0.0004 in Stukel’s favour under the complementary log-log mechanism and 0.0001 in the proposed link’s favour under the von Bertalanffy-type reference, at prevalence 0.50 and $n = 10,000$; the green series in Figures 2–4). The proposed link is therefore competitive with, not superior to, the closest established flexible link, and its advantage over Stukel lies in its interpretable growth curve parameterization and its exact nesting of the logistic model rather than in calibration. Discrimination told a uniform story of parity across all five methods: Figure 5 overlays their ROC curves for every mechanism and prevalence at the largest sample, and the curves are visually indistinguishable, matching the near-zero AUC differences in Table 12. Figure 6 makes the calibration difference visible directly, with the von Bertalanffy predictions lying closer to the diagonal of the reliability diagram than the matched-penalty ridge predictions.

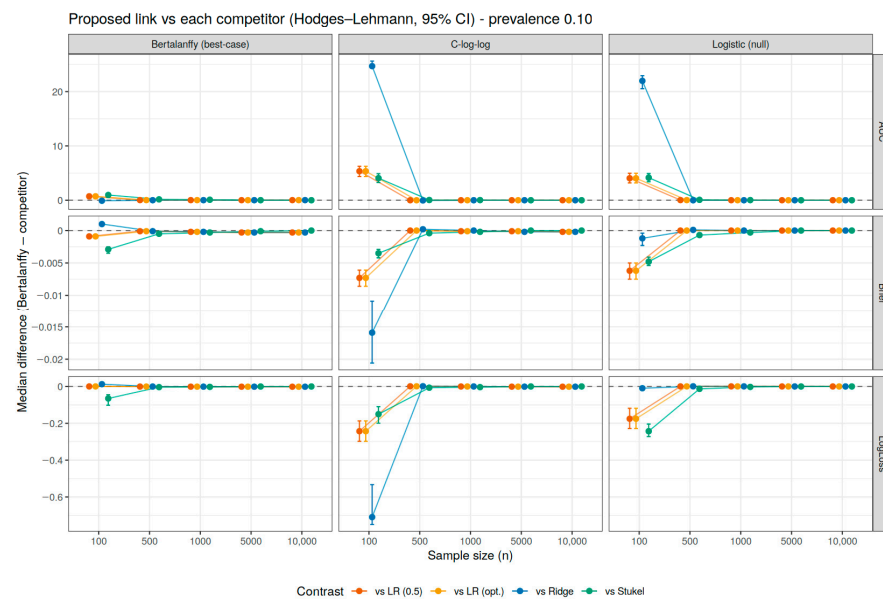


Figure 2. Paired median differences (Hodges–Lehmann, with 95% confidence intervals) between the proposed link and each of the four competitors—logistic regression at the 0.5 cut-off, threshold-optimized logistic regression, matched-penalty ridge logistic regression, and Stukel’s model—for the AUC, Brier score, and Log–Loss across sample sizes, at prevalence 0.10. For the AUC, a positive value favours the proposed link; for the Brier score and Log–Loss, a negative value favours the proposed link. The three logistic variants (orange/blue) are separated from the proposed link on calibration, while Stukel (green) overlaps it (Figure 2). The same pattern is visible at prevalence 0.25 (Figure 3).

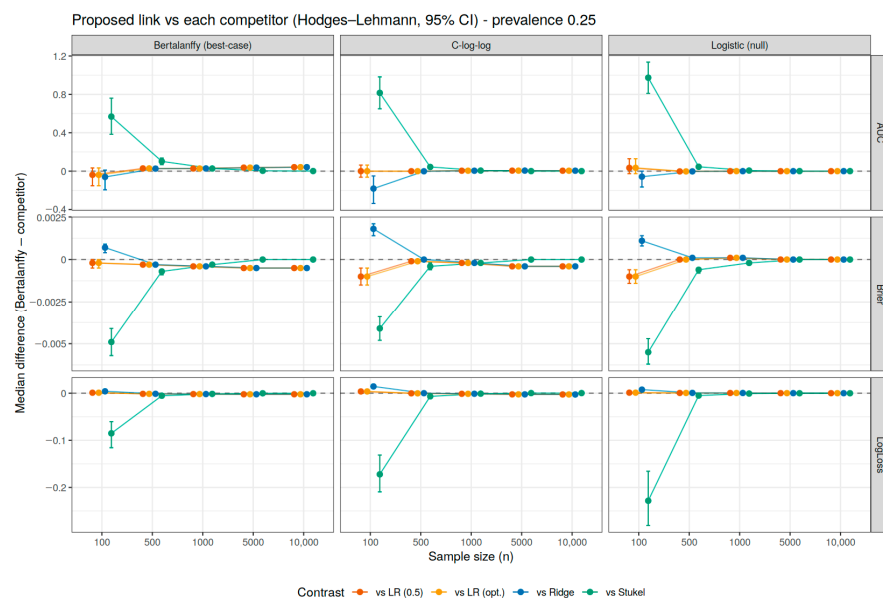


Figure 3. Paired median differences between the proposed link and each competitor at prevalence 0.25 (details as in Figure 2).

3.3. Discrimination Is Preserved

Discrimination was, by design, essentially unchanged. Because a monotone reparameterization of the link cannot on its own reorder the observations to any material degree, no gain in the area under the ROC curve or in overall accuracy was expected, and none was observed: across all five methods, the AUC values coincided to within a fraction of a single point in every moderate-to-large-sample scenario (with the smallest sample cells being unstable, as noted in Section 3.4), and the overlaid ROC curves of Figure 5 are visually indistinguishable. Under the correctly specified logistic mechanism, the AUC difference

between the proposed link and the matched-penalty ridge model was at the fourth decimal place and not consistently signed; under the complementary log–log mechanism, it ranged from about +0.0004 to +0.008 points across prevalences at the largest sample; and even under the von Bertalanffy-type best-case reference, it reached only about +0.02 to +0.06 points. The value of these differences is not their magnitude but their sign and existence: because threshold optimization leaves the AUC exactly unchanged (Section 3.1), the fact that the flexible link produces any non-zero AUC difference confirms that estimating the shape parameter within the likelihood genuinely re-estimates the fitted model rather than merely relocating its decision threshold. In short, the flexible link improves the probabilities without disturbing the ranking—the behaviour the study was designed to test.

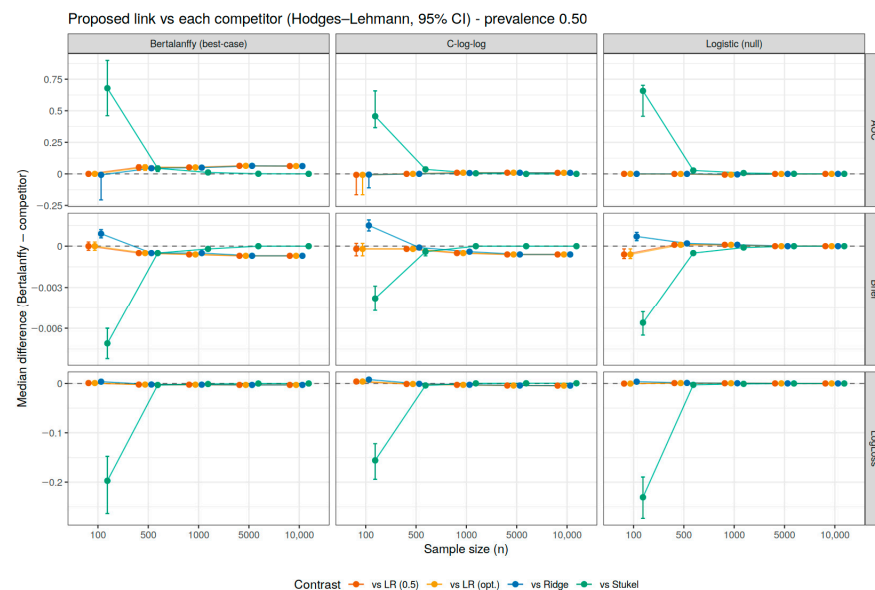


Figure 4. Paired median differences between the proposed link and each competitor at prevalence 0.50 (details as in Figure 2). The calibration separation from the three logistic variants is largest here, which is consistent with its growth with prevalence (Figure 4).

3.4. Do-No-Harm, Operating-Point Trade-Off, and Small Samples

Under the logistic mechanism, where the ordinary logit is the true model, the five methods are equivalent for practical purposes: the AUC, Brier score, and Log–Loss differences between Bert05 and Ridge05 are essentially zero at every prevalence and sample size, differing at most at the fourth decimal and in offsetting directions. This is the behaviour guaranteed by exact nesting: when the logistic link is adequate, the generalized model reduces to it, so it can neither help nor harm materially.

At the 0.5 cut-off, the flexible link shifts the operating point rather than dominating: under asymmetry, it is slightly more specific and slightly less sensitive than the ridge logistic model (for instance, at prevalence 0.25 and $n = 10,000$, sensitivity is lower by roughly 1.5–2.0 points while specificity is higher by roughly 0.6–0.9 points), an offsetting exchange that leaves balanced accuracy little altered. These threshold-dependent differences are properties of where the 0.5 cut-off falls, not of the ranking, and should not be read as discrimination gains.

Behaviour in the smallest samples is governed by separation rather than by the choice of link. At $n = 100$ under the heaviest imbalance (prevalence 0.10), the marginal tables show very large interquartile ranges for the logistic models—for example, ridge logistic AUC and Log–Loss IQRs of roughly 50 and above 2 under the complementary log–log mechanism—which are the signature of complete separation in a training fold of about seventy observations with roughly seven events. In this regime, the mild ridge penalty does

not fully stabilize the fit and estimates vary widely across replications, so these cells are unreliable and are not interpreted as evidence for any method. With the penalty matched by the ridge baseline, the proposed link shows no distinct advantage at this sample size, which is consistent with the view that stabilization in scarce data is a property of the penalty rather than of the link.

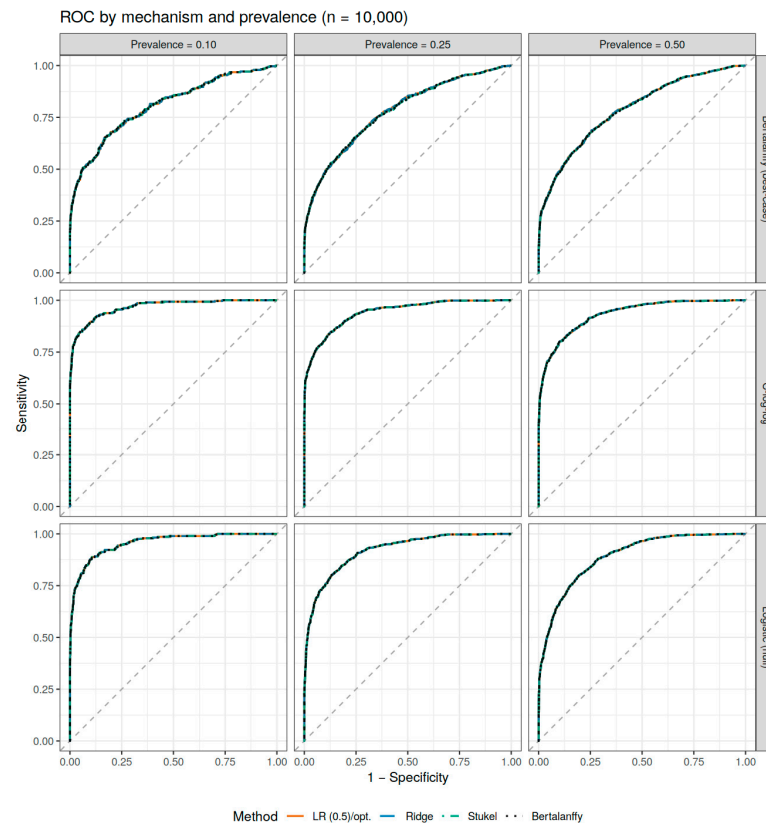


Figure 5. ROC curves for all five methods across every mechanism and prevalence at the largest sample size ($n = 10,000$). The dashed diagonal represents the performance of a random classifier ($AUC = 0.5$).

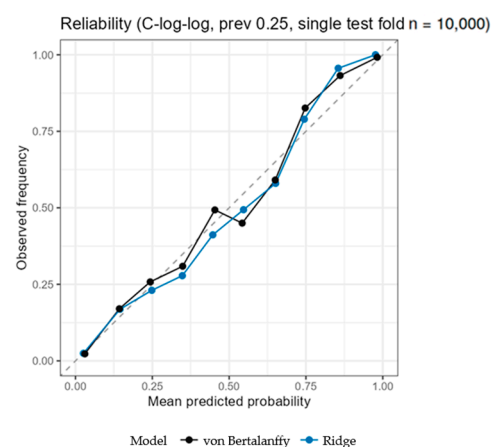


Figure 6. Reliability diagram computed on a single held-out test fold (complementary log-log mechanism, prevalence 0.25, $n = 10,000$). The dashed diagonal represents perfect calibration (predicted probability = observed frequency); points closer to it indicate better-calibrated probabilities.

3.5. Convergence, Boundary Behaviour, and Identifiability

Table 12 (paired effects) is complemented by Table 13, which reports the optimizer's behaviour by scenario. Non-convergence of the penalized fit was rare (below 1% in

almost all cells). Boundary contact for the shape parameter was, however, appreciable in two regimes: at $n = 100$, the shape parameter reached a box constraint in a sizeable fraction of replications, reflecting both the redundant growth curve baseline direction of the original parameterization (Section 2.2) and genuinely weak information about curvature when events are few; under the von Bertalanffy-type best-case mechanism at the largest samples, the shape parameter approached its upper bound in a minority of replications because the true value ($m = 4.5$) sits near the operational ceiling of the parameterization. These rates are a genuine property of the method and are reported transparently rather than suppressed; they reinforce that inference should rest on the calibration effect under generic asymmetry (the complementary log–log mechanism, where boundary contact vanishes at large n) rather than on the best-case reference.

Table 13. Optimizer diagnostics by scenario: percentage of replications with non-convergence or shape parameter boundary contact for the proposed link, and non-convergence for the Stukel model.

DGP	Prev	n	Bert. Non-Conv. (%)	Bert. Boundary (%)	Stukel Non-Conv. (%)
von Bertalanffy (best-case)	0.10	100	1.8	57.6	0.1
	0.10	500	0.0	4.7	0.0
	0.10	1000	0.2	1.3	0.0
	0.10	5000	0.3	12.7	0.0
	0.10	10,000	0.1	16.3	0.0
	0.25	100	0.1	43.1	0.0
	0.25	500	0.0	5.5	0.0
	0.25	1000	0.1	8.2	0.0
	0.25	5000	0.0	19.0	0.0
	0.25	10,000	0.0	17.9	0.0
	0.50	100	0.1	44.4	0.0
	0.50	500	0.0	15.8	0.0
	0.50	1000	0.0	17.5	0.0
	0.50	5000	0.0	22.1	0.0
	0.50	10,000	0.0	15.3	0.0
C-log–log	0.10	100	2.5	86.6	0.1
	0.10	500	0.6	13.2	0.0
	0.10	1000	0.0	0.6	0.0
	0.10	5000	0.0	0.0	0.0
	0.10	10,000	0.4	0.0	0.0
	0.25	100	0.3	70.3	0.0
	0.25	500	0.0	1.7	0.0
	0.25	1000	0.0	0.0	0.0
	0.25	5000	0.0	0.0	0.0
	0.25	10,000	0.0	0.0	0.0
	0.50	100	0.1	61.0	0.0
	0.50	500	0.1	1.0	0.0
	0.50	1000	0.0	0.0	0.0
	0.50	5000	0.0	0.1	0.0
	0.50	10,000	0.0	0.0	0.0

Table 13. Cont.

DGP	Prev	n	Bert. Non-Conv. (%)	Bert. Boundary (%)	Stukel Non-Conv. (%)
Logistic (null)	0.10	100	2.2	85.6	0.0
	0.10	500	0.5	18.6	0.0
	0.10	1000	0.0	3.7	0.0
	0.10	5000	0.0	0.0	0.0
	0.10	10,000	0.0	0.0	0.0
	0.25	100	0.6	67.4	0.1
	0.25	500	0.1	9.3	0.0
	0.25	1000	0.0	1.2	0.0
	0.25	5000	0.0	0.0	0.0
	0.25	10,000	0.0	0.0	0.0
	0.50	100	0.1	66.1	0.0
	0.50	500	0.0	11.3	0.0
	0.50	1000	0.0	2.9	0.0
	0.50	5000	0.0	0.0	0.0
	0.50	10,000	0.0	0.0	0.0

4. Discussion

The picture that emerges once a matched-penalty baseline is in place is narrower, and more defensible, than a discrimination-based reading would suggest. The proposed link's principal contribution is to probability calibration under link mis-specification. Under both asymmetric mechanisms, it produced a lower Brier score and Log-Loss than every logistic competitor, including a ridge logistic model penalized identically; the advantage was almost perfectly consistent across replications (rank-biserial correlations near -1) and increased with sample size and prevalence. Crucially, this benefit is threshold-free—the Brier score and Log-Loss cannot be improved by moving a cut-off—so it cannot be dismissed as re-thresholding. Its magnitude is modest, and we report it as such: a systematic direction rather than a large number [27,30].

Discrimination tells a complementary story. The AUC differences are negligible in size, but their mere existence is informative. Because threshold optimization left the AUC identical (LR05 and LROpt coincide on the AUC, Brier score, and Log-Loss in every scenario), the small non-zero AUC differences for the flexible link establish that estimating the shape parameter within the likelihood changes the fitted model rather than only its operating point. This resolves the central concern about flexible links masquerading as re-thresholding, while keeping the discrimination claim honest: on ranking the method, it is essentially at parity with logistic regression.

Behaviour in very small samples is a matter of numerical stability rather than of the link. When data are scarce and imbalanced, the ordinary logistic maximum-likelihood estimate diverges under complete separation, and it is a penalty—not a change in link—that repairs this; the separation problem is well recognized and is classically addressed by ridge-type penalization or by Firth's bias-reduced estimator [31]. In the present study, the ridge term plays exactly this stabilizing role for all penalized models alike, so at $n = 100$, the proposed link holds no distinct advantage over the matched-penalty ridge logistic model, and the estimates of both are, in any case, unstable. Whether a flexible link confers any specific benefit under severe separation is a separate question, left to future work.

Two further properties bear on practical use. First, the method's behaviour at the 0.5 cut-off is an operating-point shift—slightly higher specificity for slightly lower sensitivity under asymmetry—which is a matter of threshold placement, not discrimination, and can be tuned separately if a different sensitivity–specificity balance is desired. Second, the shape parameter is only weakly determined when events are few or the asymmetry is extreme, as the boundary contact rates in Table 13 show; the redundant growth curve baseline of the original parameterization—removed in the identified one-parameter form—further inflates this contact at small samples. We therefore anchor the substantive conclusion on the complementary log–log mechanism, a generic asymmetry the proposed link does not itself generate and where the optimizer sits comfortably in the interior at large samples, rather than on the von Bertalanffy-type best-case reference. That the effect appears under an asymmetry foreign to the model is the strongest evidence that it reflects genuine flexibility rather than self-recovery.

Relative to the established flexible link literature [14–17], the empirical comparison against Stukel's generalized logistic model shows the proposed link to be competitive rather than dominant: both improve calibration under asymmetry, and the growth curve construction's appeal lies in its interpretable, monotone, numerically stable parameterization and its exact nesting of the logistic model, with the latter ensuring the model reduces to ordinary logistic regression when the logit is adequate, so that it does no material harm, and, in principle, enabling a likelihood-ratio test of link adequacy against the one-parameter null $\nu = 1$ ($m = 2$), subject to the usual caveats for penalized-likelihood inference. Because the strengths are concentrated where classes are imbalanced or the boundary is asymmetric and where trustworthy probabilities matter, the method is best suited to fields such as clinical risk prediction, diagnostic screening, credit scoring, and rare event or reliability classification. In such settings, it can serve as a drop-in generalization of logistic regression: it recovers the standard model when the logit is adequate and supplies better-calibrated probabilities, at a modest computational cost, when it is not. Its principal limitations are the additional computation for the shape parameter, the small absolute size of the calibration gains, the weak identifiability just noted, and—as with the whole study—reliance on simulation; a validation on real data sets with asymmetric or imbalanced responses is the natural next step. A further open question is how the shape parameters, estimated here under an L2 (ridge) penalty, would interact with L1 (lasso) or elastic-net penalties in high-dimensional settings ($p > n$), such as genomic risk models, where variable selection is required alongside estimation.

5. Conclusions

This study proposed and tested a calibration-oriented asymmetric link for binary logistic regression, derived in closed form from the von Bertalanffy growth equation, whose single shape parameter is estimated jointly with the regression coefficients by L2-penalized maximum likelihood and which nests the logistic model exactly at $\nu = 1$ ($m = 2$). The evaluation was deliberately conservative: the link was compared not only with ordinary logistic regression at both fixed and optimized thresholds, but with a ridge-penalized logistic regression carrying an identical penalty—separating the link from the regularization it needs—and with Stukel's generalized logistic model, the closest existing flexible link competitor, all strictly out-of-sample and summarized by effect sizes with confidence intervals rather than by p -values that 45,000 replications would render uninformative.

The hypothesis was supported. At moderate-to-large sample sizes, and increasingly as prevalence approached 0.50, the proposed link produced better-calibrated probabilities than the three logistic competitors, with lower Brier score and Log–Loss whose Hodges–Lehmann median differences, though modest in absolute size, were highly consis-

tent across replications at the larger samples (rank-biserial correlations near -1 , confidence intervals excluding zero) and grew with sample size and prevalence. At the smallest sample size, the picture was mixed and dominated by separation, so no advantage is claimed there. Against Stukel's flexible model, it was competitive rather than superior, with the two coinciding on calibration. Equally important, and as intended by design, the method did not sacrifice discriminatory power: the area under the ROC curve and overall accuracy showed no practically important differences from the competing models at moderate-to-large sample sizes, so the improvement in calibration is obtained without any loss in classification performance. Under a correctly specified logistic mechanism, the method showed no material deterioration, which is consistent with its exact nesting of the logit. The practical conclusion is a link that improves probability calibration under response asymmetry while preserving discrimination, is competitive with the best existing flexible link, and reduces to ordinary logistic regression when the logit is adequate. Its interpretable growth curve parameterization and exact nesting—the latter supporting, in principle, a likelihood-ratio test of link adequacy against the one-parameter null $\nu = 1$ ($m = 2$), subject to the usual caveats for penalized-likelihood inference—are its distinguishing features. Priorities for future work include validation on real asymmetric and imbalanced data sets, cross-validated selection of the penalty, stratified resampling to further stabilize the smallest sample scenarios, and a formal test of link adequacy.

Author Contributions: Conceptualization, K.N.B. and M.G.; methodology, K.N.B. and M.G.; software, K.N.B.; validation, K.N.B. and M.G.; formal analysis, K.N.B.; investigation, K.N.B.; data curation, K.N.B.; writing—original draft preparation, K.N.B.; writing—review and editing, M.G.; visualization, K.N.B.; supervision, M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: This is a simulation study; no external or empirical data sets were used. The complete R 4.3.3 code required to reproduce the analyses and tables reported in this article—together with the fixed random seed (`set.seed(2324)`) that makes the simulated data, the train/test partitions, and the optimization paths reproducible—is openly available in a public GitHub (version: von-Bertalanffy-Logistic-Regression-2026) repository (<https://github.com/knuribaydili/von-Bertalanffy-Logistic-Regression>, accessed on 24 July 2026) and permanently archived on Zenodo (<https://doi.org/10.5281/zenodo.21498454>). The repository contains the data-generating, estimation, evaluation, and figure-producing scripts and covers all four data-generating mechanisms examined in the study (logistic, complementary log-log, Stukel-type, and von Bertalanffy-type). Three of these mechanisms are presented in the article; In addition to the scenarios presented in the article, the fourth set of results, relating to a Stukel-type mechanism, is also among the recorded outputs.

Acknowledgments: During the preparation of this article, the authors used generative AI assistants (Google Gemini-3.1 Pro and Anthropic Claude-Opus 4.8) for language editing, text enhancement, and figure generation from the results obtained by the authors. All statistical analyses, simulations, and numerical results were independently generated by the authors in the R environment; the authors reviewed and edited all AI-assisted outputs and are fully responsible for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Tabachnick, B.G.; Fidell, L.S. *Using Multivariate Statistics*; Allyn & Bacon: Boston, MA, USA, 2001.
2. Driver, H.E.; Kroeber, A.L. *Quantitative Expression of Cultural Relationships*; University of California Press: Berkeley, CA, USA, 1932; Volume 31, pp. 463–474.
3. Fisher, R.A. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **1936**, *7*, 179–188. [[CrossRef](#)]
4. Berkson, J. Application of the logistic function to bio-assay. *J. Am. Stat. Assoc.* **1944**, *39*, 357–365. [[CrossRef](#)]

5. Samura, S.K.; Kamara, J.A.; Kamara, A.A.; Kaikai, S.G. Determinants of fertility in Sierra Leone: A logistic regression model approach. *Sierra Leone J. Biomed. Res.* **2023**, *14*, 46–54. [\[CrossRef\]](#)
6. Novaliendry, D.; Yang, C.H.; Desnelita, Y.; Sanjaya, R.; Lizar, Y.; Ardi, N. Hemodialysis patient death prediction using logistic regression. *Int. J. Online Biomed. Eng.* **2023**, *19*, 4–17. [\[CrossRef\]](#)
7. Souisa, G.A.; Delsen, M.S.; Aulele, S.N. Geographically weighted logistic regression model with adaptive Gaussian weights. *AIP Conf. Proc.* **2023**, 2588, 020006. [\[CrossRef\]](#)
8. Atkinson, P.M.; German, S.E.; Sear, D.A.; Clark, M.J. Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression. *Geogr. Anal.* **2003**, *35*, 58–82. [\[CrossRef\]](#)
9. Bertsimas, D.; King, A. Logistic regression: From art to science. *Stat. Sci.* **2017**, *32*, 367–384. [\[CrossRef\]](#)
10. Gürçan, M.; Kaya, M.O.; Halisdemir, N. Evaluation of discrimination performance in logistic analysis. *Avrupa Bilim. Teknol. Derg.* **2019**, *17*, 1008–1013.
11. Verhulst, P.-F. Notice sur la loi que la population suit dans son accroissement. *Corresp. Math. Phys.* **1838**, *10*, 113–121.
12. van Calster, B.; McLernon, D.J.; van Smeden, M.; Wynants, L.; Steyerberg, E.W. Calibration: The Achilles heel of predictive analytics. *BMC Med.* **2019**, *17*, 230. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Nelder, J.A.; Wedderburn, R.W.M. Generalized linear models. *J. R. Stat. Soc. Ser. A (Gen.)* **1972**, *135*, 370–384. [\[CrossRef\]](#)
14. Prentice, R.L. A generalization of the probit and logit methods for dose response curves. *Biometrics* **1976**, *32*, 761–768. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Aranda-Ordaz, F.J. On two families of transformations to additivity for binary response data. *Biometrika* **1981**, *68*, 357–363. [\[CrossRef\]](#)
16. Pregibon, D. Goodness of link tests for generalized linear models. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **1980**, *29*, 15–24. [\[CrossRef\]](#)
17. Stukel, T.A. Generalized logistic models. *J. Am. Stat. Assoc.* **1988**, *83*, 426–431. [\[CrossRef\]](#)
18. von Bertalanffy, L. Quantitative laws in metabolism and growth. *Q. Rev. Biol.* **1957**, *32*, 217–231. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Steyerberg, E.W.; Vickers, A.J.; Cook, N.R.; Gerd, T.; Gonen, M.; Obuchowski, N.; Pencina, M.J.; Kattan, M.W. Assessing the performance of prediction models. *Epidemiology* **2010**, *21*, 128–138. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Seber, G.A.F.; Wild, C.J. *Nonlinear Regression*; Wiley: New York, NY, USA, 1989; pp. 325–365.
21. Hanley, J.A.; McNeil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, *143*, 29–36. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Brier, G.W. Verification of forecasts expressed in terms of probability. *Mon. Weather Rev.* **1950**, *78*, 1–3. [\[CrossRef\]](#)
23. Murphy, A.H. A new vector partition of the probability score. *J. Appl. Meteorol.* **1973**, *12*, 595–600. [\[CrossRef\]](#)
24. Gneiting, T.; Raftery, A.E. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **2007**, *102*, 359–378. [\[CrossRef\]](#)
25. Hodges, J.L.; Lehmann, E.L. Estimates of location based on rank tests. *Ann. Math. Stat.* **1963**, *34*, 598–611. [\[CrossRef\]](#)
26. Kerby, D.S. The simple difference formula: An approach to teaching nonparametric correlation. *Compr. Psychol.* **2014**, *3*, 2165–2228. [\[CrossRef\]](#)
27. Lin, M.; Lucas, H.C.; Shmueli, G. Too big to fail: Large samples and the p -value problem. *Inf. Syst. Res.* **2013**, *24*, 906–917. [\[CrossRef\]](#)
28. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **1995**, *57*, 289–300. [\[CrossRef\]](#)
29. The R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2024.
30. Sullivan, G.M.; Feinn, R. Using effect size—Or why the p value is not enough. *J. Grad. Med. Educ.* **2012**, *4*, 279–282. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Firth, D. Bias reduction of maximum likelihood estimates. *Biometrika* **1993**, *80*, 27–38. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.