Supplementary Note 1

Simulation structure for linear scenarios

True zoonotic spillover counts at site *i* in month *t* were drawn from a Poisson distribution:

$$Y_{it}^{true} \sim Poisson(\lambda_{it})$$

With log-mean:

$$log\ \lambda_{it} = \beta_0 + \alpha_i + \beta_{temp} \cdot Temp_{it} + \epsilon_{it}$$

Where β_0 is a scenario-specific intercept, $\alpha_i \sim N(0, 0.01^2)$ is a site-level random intercept, β_{temp} is the scenario-specific effect of temperature on spillover risk, and $\epsilon_{it} \sim N(0, 0.1^2)$ is observation-level noise on the log scale.

Site-month temperature (℃) was generated as:

$$Temp_{it} = 15 + \delta \cdot t + \nu_{it}$$

Where δ is the scenario-specific monthly temperature trend and $\nu_{it} \sim N(0, {\sigma_{temp}}^2)$ is independent with noise, with σ_{temp} varying by scenario.

Surveillance capacity at site i in month t, denoted W_{it} varied by scenario. For the increasing and decreasing detection scenarios:

$$W_{it} = max(0, \mu_W + \gamma \cdot \frac{t}{12} + \eta_i + \zeta_{it}$$

For the steady detection scenario:

$$W_{it} = max(0, \mu_W + \eta_i + \zeta_{it})$$

Where $\eta_i \sim N(0, 0.5^2)$ is a site-level surveillance intercept, and $\zeta_{it} \sim N(0, {\sigma_W}^2)$ is site-month noise. Scenario-specific values of μ_W, γ, γ', and σ_W are given in Table X.

Observed cases were selected by thinning the Poisson results. Detection probability at site i in month t was modeled on the logit scale:

$$logit(p_{it}) = \theta_0 + \theta_1 \cdot W_{it} + \zeta_{it}, \qquad \zeta_{it} \sim N(0, {\sigma_p}^2)$$

Where θ_0 and θ_1 are scenario-specific intercept and slope parameters, and σ_p varies by scenario (Table X). Observed cases are then drawn as:

$$Y_{it}^{obs} \sim Binomial(Y_{it}^{true}, p_{it})$$

Simulation structure for oscillating scenarios

The oscillating scenarios share the same data generation models as the linear scenarios. The oscillation is produced in the temperature process. Site-month now includes a seasonal component:

$$Temp_{it} = 15 + \delta \cdot t + A \cdot sin\left(\frac{2\pi t}{12}\right) + \nu_{it}\ , \qquad \nu_{it} \sim N(0,1)$$

Where δ is the scenario-specific monthly linear trend, and $A = 3$ is the oscillation amplitude fixed across all oscillating scenarios, which results in a twelve month seasonal cycle with peak-to-trough range of $2A = 6$℃.

Simulation structure for distance scenarios

The distance scenarios share the same structure as the linear scenarios. However, to generate the distance bias the pure site random effect is replaced by a distance-to-testing term. At the start of each simulation replicate, a distance (km) to the testing facility was drawn independently for each site:

$$d_i \sim Uniform(1{,}100)$$

This was normalized to a [0,1] proximity score, where higher values indicate closer proximity to testing facility:

$$d_i' = 1 - \frac{d_i - min_i(d_i)}{max_i(d_i) - min_i(d_i)}$$

Surveillance capacity combined the scenario-specific temporal baseline with the proximity effect and site-level noise:

$$W_{it} = max(0, B_{it} + 2d_i' + \phi_i)$$

Where B_{it} is the scenario-specific temporal baseline (identical in form to the linear scenario surveillance formulas), d_i' is the scaled testing facility proximity so that closer sites receive up to two additional units of surveillance capacity, and $\phi_i \sim N(0{,}0.3^2)$ is a site-level noise term drawn once per site per replicate.

Simulation structure for shared-driver scenarios

The shared-driver scenario, unlike the previous scenarios, has multiple covariates simultaneously influencing both the true spillover process and detection probability. At the start of each simulation replicate, the following site-level baseline parameters were drawn independently for each site i:

$Elevation \sim Uniform(0{,}3000)$, standardized to e_i'

$BaseTemp_i \sim Uniform(5{,}25)$, standardized to b_i'

$$BaseWildlife_i \sim Uniform(0.5, 2.5)$$

$\rho_i{}^W \sim N(3.5{,}1.5^2)$, $\kappa_i{}^W \sim N(0, 0.8^2)$ wildlife growth rate and nonlinear term

$$BaseSurv_i \sim Uniform(0.05{,}0.25)$$

$\rho_i{}^S \sim N(4.0{,}1.8^2)$, $\kappa_i{}^S \sim N(0{,}1.0^2)$ surveillance growth rate and nonlinear term

$$BaseResources_i \sim Uniform(-1{,}1)$$

Let $t' = (t-1)/(T-1) \in [0{,}1]$ denote time scaled to the unit interval.

Temperature variation over time was calculated by:

$$Temp_{it} = b_i' + 0.08sin\left(\frac{2\pi t}{12}\right) + \nu_{it}{}^{temp}, \qquad \nu_{it}{}^{temp} \sim N(0, 0.4^2)$$

Wildlife density follows a site-specific nonlinear trajectory:

$$W_{it}{}^{wild} = BaseWildlife_i + \rho_i{}^W t' + \kappa_i{}^W (t'^2 - 0.5) + 0.05sin\left(\frac{2\pi t}{12}\right) + \nu_{it}{}^W,$$

$$\nu_{it}{}^W \sim N(0, 0.8^2)$$

Surveillance infrastructure follows an analogous nonlinear trajectory:

$$W_{it}{}^{surv} = BaseSurv_i + \rho_i{}^S t' + \kappa_i{}^S (t'^2 - 0.5) + \nu_{it}{}^S, \qquad \nu_{it}{}^S \sim N(0{,}0.7^2)$$

Resource availability follows a multi-period oscillation:

$$R_{it} = BaseResources_i + 0.5sin\left(\frac{2\pi t}{36}\right) + 0.3sin\left(\frac{2\pi t}{18}\right) + \nu_{it}{}^R, \qquad \nu_{it}{}^R \sim N(0{,}1.0^2)$$

True cases were drawn from a Poisson distribution with log-mean:

$$log\lambda_{it} = \beta_0 + \beta_e e_t' + \beta_{temp} \cdot Temp_{it} + \beta_W W_{it}{}^{wild} + \beta_R R_{it} + \epsilon_{it}{}^{\lambda} \ , \qquad \epsilon_{it}{}^{\lambda} \sim N(0{,}0.06^2)$$

Scenario-specific values of β_0 and β_W are given in Table X; remaining coefficients are fixed across scenarios at $\beta_e = 0.05$, $\beta_{temp} = 0.06$, $\beta_R = 0.08$.
Detection probability was modeled on the logit scale as:

$$logit(p_{it}) = \theta_0 + \theta_{temp} \cdot Temp_{it} + \theta_W {W_{it}}^{wild} + \theta_S {W_{it}}^{surv} + \theta_R R_{it} + {\epsilon_{it}}^{p}, \quad {\epsilon_{it}}^{p} \sim N(0, 0.2^2)$$

Scenario-specific values of θ_0 and θ_W are given in Table X; remaining coefficients vary by surveillance trajectory.
Observed cases were then drawn as:

$${Y_{it}}^{obs} \sim Binomial({Y_{it}}^{true}, p_{it})$$

Supplementary Table 1. Shared parameter values across all scenarios. Not all parameters are used in all scenarios.

Parameter	Value
Sites (S)	70
Months (T)	120
Simulations per scenario	200
Site intercept SD (σ_a)	0.01
Log-mean noise SD (σ_ϵ)	0.10
Oscillation amplitude (A)	3
Distance to testing facility (d_i)	Uniform(1,100)

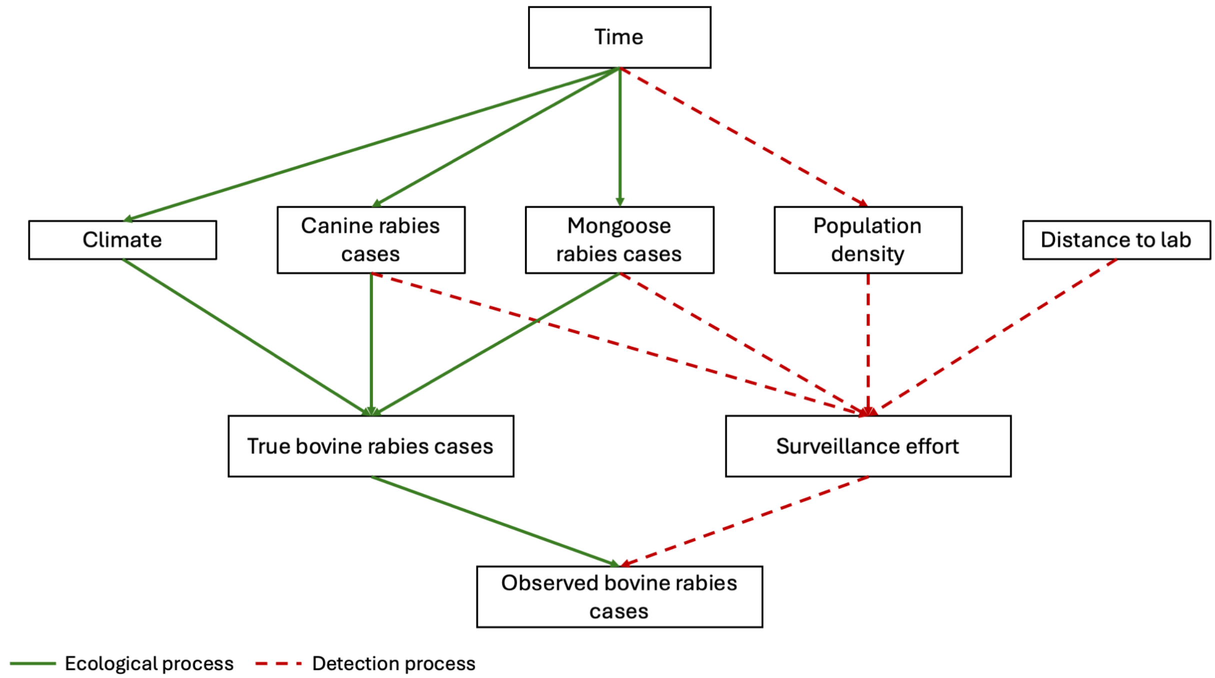
Supplementary Table 2. Parameter values across all scenarios varying by surveillance trajectory.

Surveillance trajectory	μ_W	γ	θ_0	θ_1	σ_p
Increasing	0.5	0.55	-1.0	0.5	0.01
Steady	2.5	0	-1.0	0.5	0.01
Decreasing	10.0	0.05	-2.0	0.4	0.12

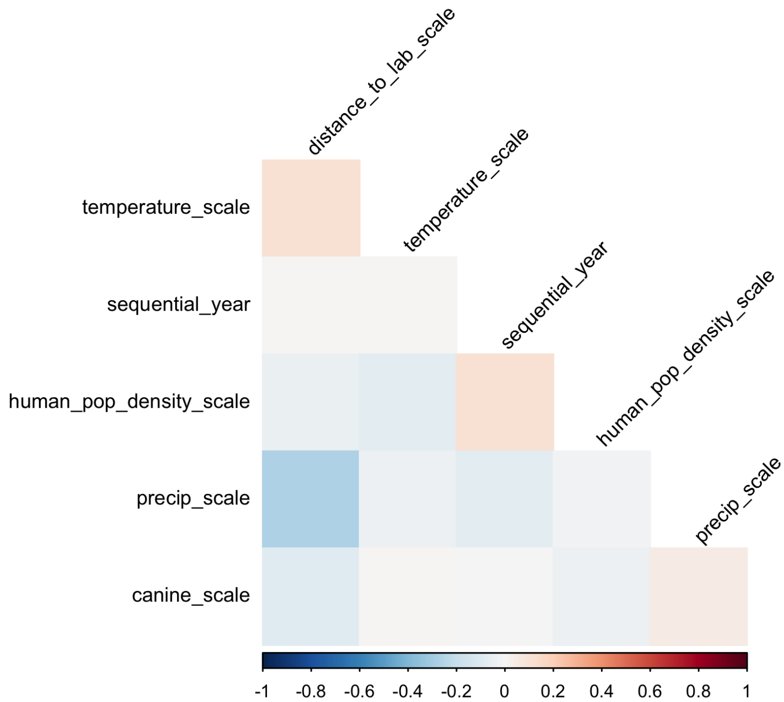
Supplementary Table 3. Scenario-specific parameter values varying by spillover direction and prevalence. Values given as high/ medium/ low prevalence were applicable. β_0 and β_{temp} were jointly calibrated to achieve target expected case counts at each prevalence level for linear, oscillating, and distance families; β_0 and β_W serve the same purpose in the shared-driver family. Fixed shared-driver coefficients across all scenarios: $\beta_e = 0.05$, $\beta_{temp} = 0.06$, $\beta_R = 0.08$.

Spillover	δ	β_0**for linear, oscillating, and distance scenarios**	β_0**for shared-driver scenarios**	β_{temp} **for linear, oscillating, and distance scenarios**	β_W**for shared-driver scenarios**
Increasing	0.05	−0.20 / −2.85 / −5.10	1.45 / −0.982 / -3.35	0.12 / 0.15 / 0.15	0.150 / 0.256 / TBD
Steady	0	1.00 / −0.80 / −3.00	1.70 / −0.067 / -2.54	0.05 / 0.05 / 0.05	0.005 / 0.005 / TBD
Decreasing	0.05	3.95 / 2.56 / −0.20	2.50 / 0.684 / -1.69	−0.11 / −0.15 / −0.15	−0.150 / −0.256 / TBD

Supplementary Figure 1. Rabies system DAG



Supplementary Figure 2. Correlation between bovine rabies model covariates.



Supplementary Table 4. Variance inflation factors (VIF) between bovine rabies model covariates for GLMM and GAMMs.

Covariate	GAMM VIF	GLMM VIF
Distance to lab	1.01	1.01
Human population density	1.01	1.03
Year	1.01	1.03

Supplementary Note 2

Using the same NIMBLE configuration as the case study (4 parallel chains, 500,000 iterations, 100,000 burn-in, thinning 200) with weakly informative priors on the abundance and detection coefficients, the detection parameters failed to converge. In a pilot run of ten iterations of the steady spillover, increasing detection, high prevalence simulation, effective sample size (ESS) of the detection intercept was 46 to 180. This behavior is consistent with the well-documented weak separability of abundance and detection in N-mixture models fit to count data, which is most severe for the sparse, zero-heavy data generated here (Barker et al. 2018; Kéry 2018). Convergence and approximate trend recovery were achievable only when we imposed strongly informative priors centered near the data-generating values. Because those priors encode the very quantity the simulation is designed to estimate, the resulting recovery is partly an artifact of the prior rather than evidence that the model extracts the trend from biased counts; an informative prior centered on an incorrect value would bias the estimate toward that value just as readily. In an applied setting the true trend is unknown, so such priors cannot be specified without effectively assuming the answer. We therefore judged the N-mixture model uninformative as a simulation benchmark due to the high correlation between predictors and year, and restricted it to the case study where it serves as one of four methods.

Supplementary Table 5. Full simulation results for high prevalence scenarios. Counts of simulations classified into each recovery category for each true-trend condition, excluding three GLMMs with convergence failures. Percents in each cell are the column percent correct.

Model Recovered Trend		True trend		
		Decreasing	Steady	Increasing
Linear scenario				
GLMM	Decreasing	600 (100%)	11 (2%)	0 (0%)
	Steady	0 (0%)	570 (95%)	0 (0%)
	Increasing	0 (0%)	19 (3%)	600 (100%)
GAMM	Decreasing	600 (100%)	11 (2%)	0 (0%)
	Steady	0 (0%)	571 (95%)	0 (0%)
	Increasing	0 (0%)	18 (3%)	600 (100%)
DML	Decreasing	565 (94%)	9 (2%)	0 (0%)
	Steady	35 (6%)	582 (97%)	18 (3%)
	Increasing	0 (0%)	9 (2%)	582 (97%)
Distance scenario				
GLMM	Decreasing	600 (100%)	14 (2%)	0 (0%)
	Steady	0 (0%)	572 (95%)	0 (0%)
	Increasing	0 (0%)	14 (2%)	600 (100%)
GAMM	Decreasing	600 (100%)	11 (2%)	0 (0%)
	Steady	0 (0%)	575 (96%)	0 (0%)
	Increasing	0 (0%)	14 (2%)	600 (100%)
DML	Decreasing	581 (97%)	6 (1%)	0 (0%)
	Steady	19 (3%)	582 (97%)	13 (2%)
	Increasing	0 (0%)	12 (2%)	587 (98%)
Oscillating scenario				
GLMM	Decreasing	600 (100%)	1 (0%)	0 (0%)
	Steady	0 (0%)	212 (35%)	0 (0%)
	Increasing	0 (0%)	387 (65%)	600 (100%)
GAMM	Decreasing	600 (100%)	1 (0%)	0 (0%)
	Steady	0 (0%)	213 (36%)	0 (0%)
	Increasing	0 (0%)	386 (64%)	600 (100%)
DML	Decreasing	600 (100%)	1 (0%)	0 (0%)
	Steady	0 (0%)	191 (32%)	0 (0%)
	Increasing	0 (0%)	408 (68%)	600 (100%)
Shared-driver scenario				
GLMM	Decreasing	6 (1%)	7 (1%)	4 (1%)
	Steady	543 (91%)	556 (93%)	545 (91%)
	Increasing	51 (9%)	34 (6%)	51 (9%)
GAMM	Decreasing	31 (5%)	31 (5%)	21 (4%)
	Steady	472 (79%)	473 (79%)	471 (79%)
	Increasing	97 (16%)	96 (16%)	108 (18%)
DML	Decreasing	32 (5%)	11 (2%)	2 (0%)
	Steady	542 (90%)	571 (95%)	544 (91%)
	Decreasing	26 (4%)	18 (3%)	54 (9%)

Supplementary Table 6. Full simulation results for medium prevalence scenarios. **Counts of simulations classified into each recovery category for each true-trend condition, excluding 139 GLMMs with convergence failures. Percents in each cell are the column percent correct.**

Model Recovered Trend		True trend		
		Decreasing	**Steady**	**Increasing**
Linear scenario				
GLMM	Decreasing	588 (100%)	8 (1%)	0 (0%)
	Steady	0 (0%)	563 (97%)	0 (0%)
	Increasing	0 (0%)	10 (2%)	592 (100%)
GAMM	Decreasing	600 (100%)	11 (2%)	0 (0%)
	Steady	0 (0%)	578 (96%)	0 (0%)
	Increasing	0 (0%)	11 (2%)	600 (100%)
DML	Decreasing	360 (60%)	13 (2%)	0 (0%)
	Steady	240 (40%)	572 (95%)	230 (38%)
	Increasing	0 (0%)	15 (3%)	370 (62%)
Distance scenario				
GLMM	Decreasing	587 (100%)	10 (2%)	0 (0%)
	Steady	0 (0%)	565 (97%)	0 (0%)
	Increasing	0 (0%)	10 (2%)	585 (100%)
GAMM	Decreasing	600 (100%)	11 (2%)	0 (0%)
	Steady	0 (0%)	578 (96%)	0 (0%)
	Increasing	0 (0%)	11 (2%)	600 (100%)
DML	Decreasing	405 (68%)	11 (2%)	0 (0%)
	Steady	195 (33%)	586 (98%)	212 (35%)
	Increasing	0 (0%)	3 (1%)	388 (65%)
Oscillating scenario				
GLMM	Decreasing	598 (100%)	2 (0%)	0 (0%)
	Steady	0 (0%)	437 (77%)	0 (0%)
	Increasing	0 (0%)	132 (23%)	600 (100%)
GAMM	Decreasing	600 (100%)	2 (0%)	0 (0%)
	Steady	0 (0%)	466 (78%)	0 (0%)
	Increasing	0 (0%)	132 (22%)	600 (100%)
DML	Decreasing	600 (100%)	3 (0%)	0 (0%)
	Steady	0 (0%)	241 (40%)	0 (0%)
	Increasing	0 (0%)	356 (59%)	600 (100%)
Shared-driver scenario				
GLMM	Decreasing	11 (2%)	5 (1%)	12 (2%)
	Steady	541 (92%)	574 (96%)	549 (93%)
	Increasing	36 (6%)	19 (3%)	27 (5%)
GAMM	Decreasing	18 (3%)	13 (2%)	11 (2%)
	Steady	541 (90%)	556 (93%)	552 (92%)
	Increasing	41 (7%)	31 (5%)	37 (6%)
DML	Decreasing	16 (3%)	9 (2%)	6 (1%)
	Steady	568 (95%)	579 (97%)	546 (91%)
	Increasing	16 (3%)	12 (2%)	48 (8%)

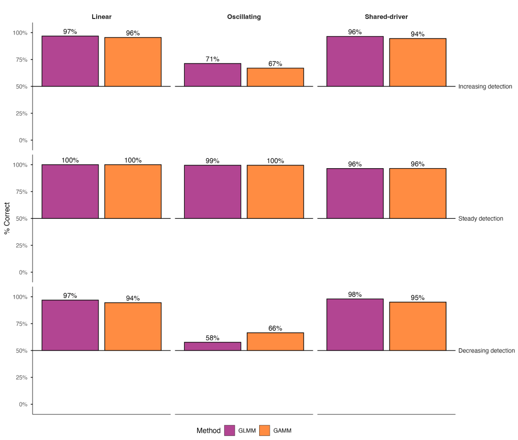
Supplementary Table 7. Full simulation results for low prevalence scenarios. **Counts of simulations classified into each recovery category for each true-trend condition, excluding 243 GLMMs with convergence failures. Percents in each cell are the column percent correct.**

Model Recovered Trend		True trend		
		Decreasing	Steady	Increasing
Linear scenario				
GLMM	Decreasing	337 (59%)	11 (2%)	1 (0%)
	Steady	239 (41%)	563 (96%)	214 (38%)
	Increasing	0 (0%)	11 (2%)	349 (62%)
GAMM	Decreasing	355 (59%)	12 (2%)	1 (0%)
	Steady	245 (41%)	576 (96%)	217 (36%)
	Increasing	0 (0%)	12 (2%)	382 (64%)
DML	Decreasing	238 (40%)	13 (2%)	1 (0%)
	Steady	362 (60%)	580 (97%)	365 (61%)
	Increasing	0 (0%)	7 (1%)	234 (39%)
Distance scenario				
GLMM	Decreasing	432 (75%)	11 (2%)	0 (0%)
	Steady	143 (25%)	567 (96%)	125 (22%)
	Increasing	0 (0%)	11 (2%)	439 (78%)
GAMM	Decreasing	463 (77%)	12 (2%)	0 (0%)
	Steady	137 (23%)	577 (96%)	129 (22%)
	Increasing	0 (0%)	11 (2%)	471 (79%)
DML	Decreasing	237 (40%)	13 (2%)	2 (0%)
	Steady	361 (60%)	574 (96%)	369 (62%)
	Increasing	2 (0%)	13 (2%)	229 (38%)
Oscillating scenario				
GLMM	Decreasing	546 (94%)	4 (1%)	0 (0%)
	Steady	37 (6%)	551 (94%)	30 (5%)
	Increasing	0 (0%)	32 (5%)	546 (95%)
GAMM	Decreasing	561 (94%)	5 (1%)	0 (0%)
	Steady	39 (7%)	562 (94%)	34 (6%)
	Increasing	0 (0%)	33 (6%)	566 (94%)
DML	Decreasing	576 (96%)	2 (0%)	0 (0%)
	Steady	24(4%)	523 (87%)	14 (2%)
	Increasing	0 (0%)	75 (13%)	586 (98%)
Shared-driver scenario				
GLMM	Decreasing	15 (3%)	8 (1%)	15 (3%)
	Steady	554 (95%)	565 (96%)	551 (95%)
	Increasing	16 (3%)	17 (3%)	17 (3%)
GAMM	Decreasing	17 (3%)	8 (1%)	14 (2%)
	Steady	567 (95%)	575 (96%)	569 (95%)
	Increasing	16 (3%)	17 (3%)	17 (3%)
DML	Decreasing	27 (5%)	10 (2%)	9 (2%)
	Steady	561 (94%)	571 (95%)	572 (95%)
	Decreasing	12 (2%)	19 (3%)	19 (3%)

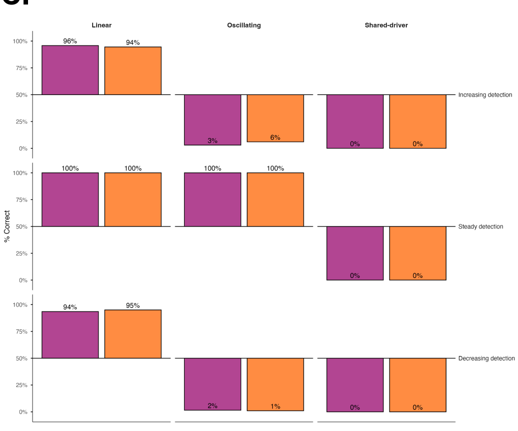
Supplementary Figure 3. Point recovery success for generalized linear mixed models (GLMMs) and generalized additive mixed models (GAMMs) for linear, oscillating, and shared-driver scenarios at medium prevalence. Panel A shows scenarios where the spillover is increasing over time, B is steady spillover over time, and C is decreasing spillover.

A.

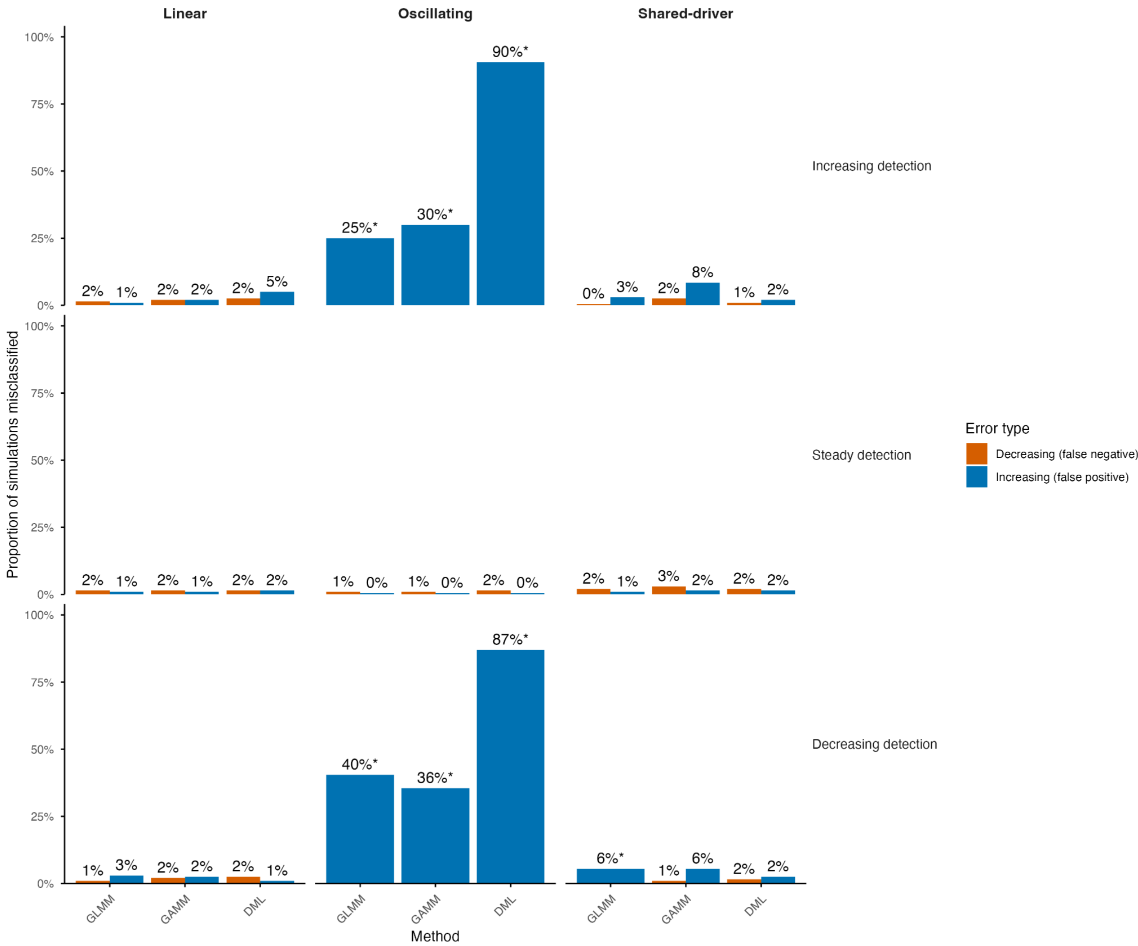
B.



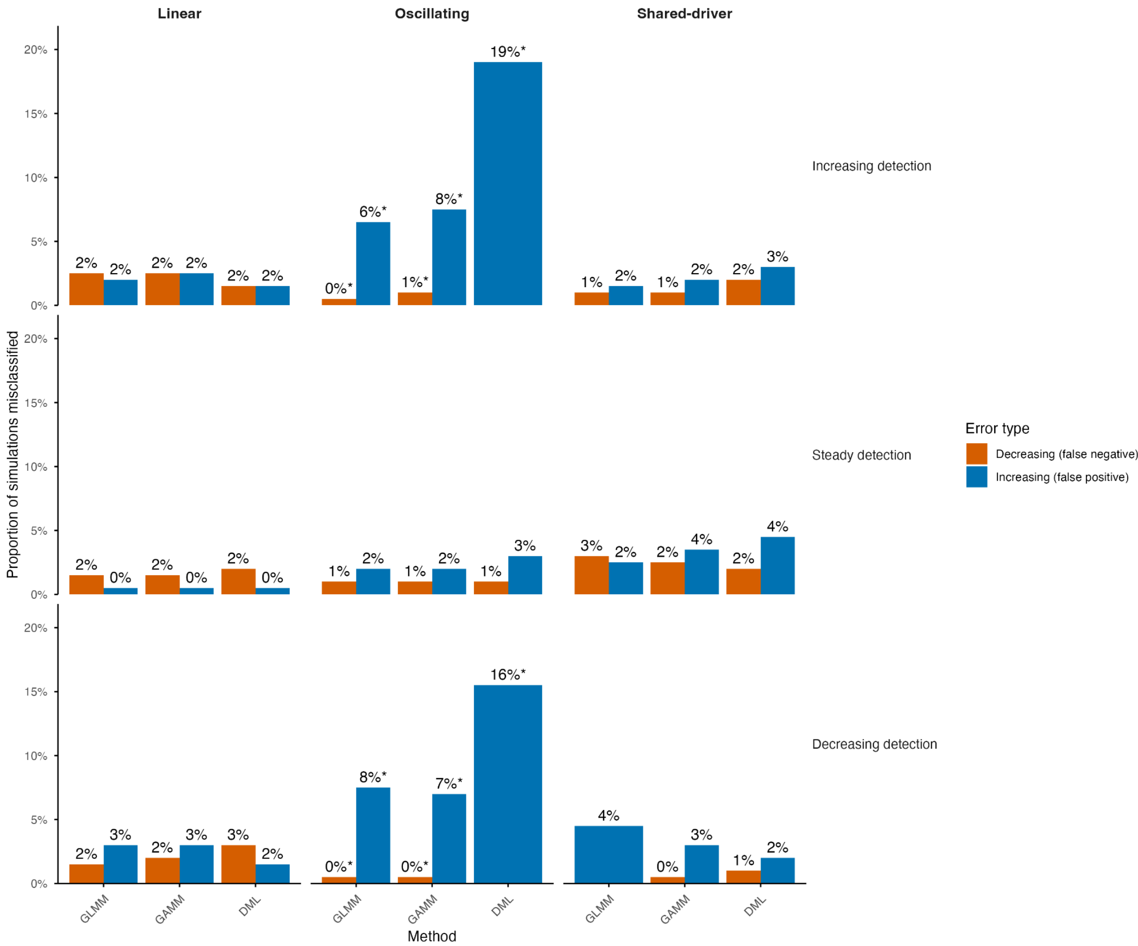
C.



Supplementary Figure 4. Misclassification errors for steady spillover trend in generalized linear mixed models (GLMMs), generalized additive mixed models (GAMMs), and double machine learning (DML) for linear, oscillating, and shared-driver scenarios at medium prevalence. Denominator (n=200) is total simulations per scenario cell. * significant asymmetry between positive and negative errors (binomial test, Bonferroni-corrected p < 0.05, only tested when ten or more errors occurred for a given model).

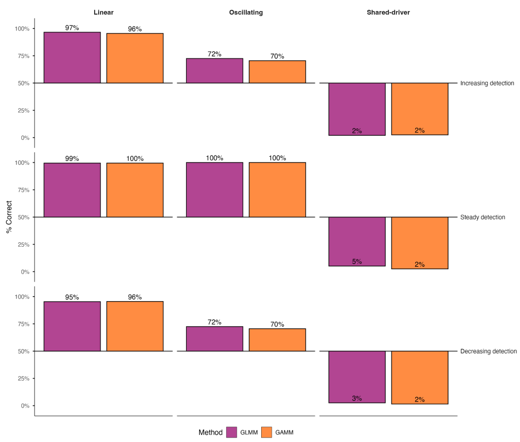


Supplementary Figure 5. Misclassification errors for steady spillover trend in generalized linear mixed models (GLMMs), generalized additive mixed models (GAMMs), and double machine learning (DML) for linear, oscillating, and shared-driver scenarios at low prevalence. Denominator (n=200) is total simulations per scenario cell. * significant asymmetry between positive and negative errors (binomial test, Bonferroni-corrected p < 0.05, only tested when ten or more errors occurred for a given model).

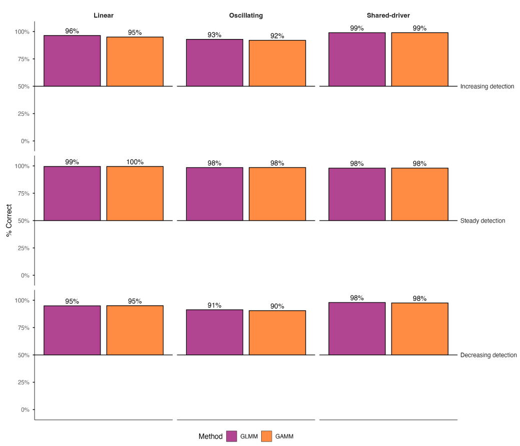


Supplementary Figure 6. Point recovery success for generalized linear mixed models (GLMMs) and generalized additive mixed models (GAMMs) for linear, oscillating, and shared-driver scenarios at low prevalence. Panel A shows scenarios where the spillover is increasing over time, B is steady spillover over time, and C is decreasing spillover.

A.



B.



C.