

12 **Abstract**

13 **Background:** The relationship between antibody levels (more generally, a scalar measure of
14 immune protection) at the time of exposure to infection (so-called exposure-proximal correlates
15 of protection) and the risk of infection given exposure is of central interest in evaluating the
16 evolution of immune protection conferred by prior infection and/or vaccination. A version of the
17 test-negative study design (TND), adapted from vaccine effectiveness studies, has been used to
18 assess this relationship. However, the conditions under which such a study identifies the
19 relationship between immune measurements and protection have not been defined.

20 **Objective:** To evaluate the conditions for TNDs to estimate the relationship between antibody
21 levels or a similar scalar measurement of immunity (hereafter exposure-proximal correlates of
22 protection, COP) and the relative incidence rate of infection given exposure.

23 **Method:** Individual-based transmission models, linking infection risk linearly and nonlinearly
24 with COP value and accounting for waning immunity post-vaccination and -infection, were used.
25 Simulations were performed of a TND with sampling on predetermined dates. Data from either
26 one or multiple simulation days were analyzed using logistic regression and generalized additive
27 models.

28 **Result:** A correctly specified logistic regression model provided an unbiased estimate of the
29 effectiveness of specific COP levels (analogous to vaccine effectiveness). Aggregating data
30 across different simulation dates with incidence-density sampling also provided reliable
31 estimates of protection. When, as is generally the case, the functional form relating COP level to
32 protection is unknown, generalized additive models offer a more flexible alternative to
33 traditional logistic regression approaches.

34 **Conclusion:** A TND can validly estimate the relative effect of an immune COP at the time of
35 exposure on the incidence rate of infection via logistic regression if the functional form of the
36 effect is known and appropriately modeled or unknown a semiparametric approach. Future
37 research should further examine the dynamics of immunity waning and boosting for more
38 reliable inference.

39 **1. Introduction**

40 Correlates of protection (COP) are measurable quantities such as binding or neutralizing
41 antibody concentrations that predict the degree of protection against incidence of an infectious
42 disease. These markers provide valuable insights into the immune system's response to
43 pathogens and vaccines,¹ which is essential for advancing the understanding of immune
44 mechanisms, as well as facilitating estimates of levels of protection in the population over time
45 and informing the evaluation of new vaccines. For example, hemagglutination-inhibition
46 antibody titers have been identified as a COP for influenza.^{2,3} COP are particularly useful for
47 estimating vaccine effectiveness by linking the magnitude of an immune response to levels of
48 protection, especially in scenarios where direct measures of effectiveness are not available.^{4,5}

49 Recent studies highlight the importance of post-immunization antibody titers as effective
50 COP for COVID-19 vaccines.^{6,7} Notable research efforts have investigated the use of these
51 correlates to forecast absolute risks (AR) and relative risks (RR) of infection in randomized
52 vaccine efficacy trials,^{4,6,8-10} using the antibody concentration measured at a fixed time post-
53 vaccination. More recently, an observational study using a test-negative design (TND) was
54 employed to estimate “exposure-proximal” COP, that is, how the COP level around the time an
55 individual may be exposed to infection affects their risk of becoming infected.^{11,12}

56 The use of the TND for exposure-proximal COP studies builds on a longstanding
57 tradition of using these studies, which compare vaccination histories of those who test positive
58 for a condition (e.g., COVID-19) with those experiencing the same symptoms but testing
59 negative for the condition, to evaluate vaccine effectiveness. If the vaccine provides all-or-
60 nothing protection and there are no unmeasured confounding with respect to infection or test-
61 seeking, such as due to heterogeneous vaccination decisions or characteristics leading to varying

62 susceptibility among participants, the odds ratio (OR) for vaccination among test-positive vs.
63 test-negative participants in a TND is an unbiased estimator of the incidence rate ratio (IRR) in
64 the population, such that one minus the OR estimates vaccine effectiveness ($1-IRR$). These
65 assumptions are strong,¹³ and methodological work has highlighted that they may be violated in
66 practice.^{14–20}

67 However, to our knowledge, no study has yet evaluated the validity of the TND for
68 monitoring the relationship between a continuous exposure-proximate COP and protection. In
69 contrast with using the TND for vaccine effectiveness, the goal is not to estimate the causal
70 effect, as interventions on the correlate may not be well defined, but rather to estimate the
71 (predictive) relationship between COP-level and infection over time. In this case, our main
72 concern is the potential for selection bias due to the sampling scheme of the TND.

73 Here, we introduce a novel, simplified simulation model that mimics disease transmission
74 within a community, taking into account vaccination and the waning immunity from previous
75 infections, to explore whether the TND is able to recover the relationship between COP and
76 infection and to identify the correct transformations and statistical models for accurately linking
77 COP to infection. We are specifically interested in situations where assessing vaccine
78 effectiveness in TND may otherwise introduce bias such as when the vaccine exhibits leaky
79 protection. When the relationship between COP, the predictor, and infection risk is linear, we
80 show that a model linking the $\ln(IRR)$ of infection to either the logarithm of one minus the
81 linear predictor of incidence rate (parametric) or a flexible function of that predictor
82 (semiparametric) can effectively recover the correct relationship between that predictor and the
83 incidence rate ratio of infection. When the relationship is unknown, we show semiparametric
84 methods are superior.

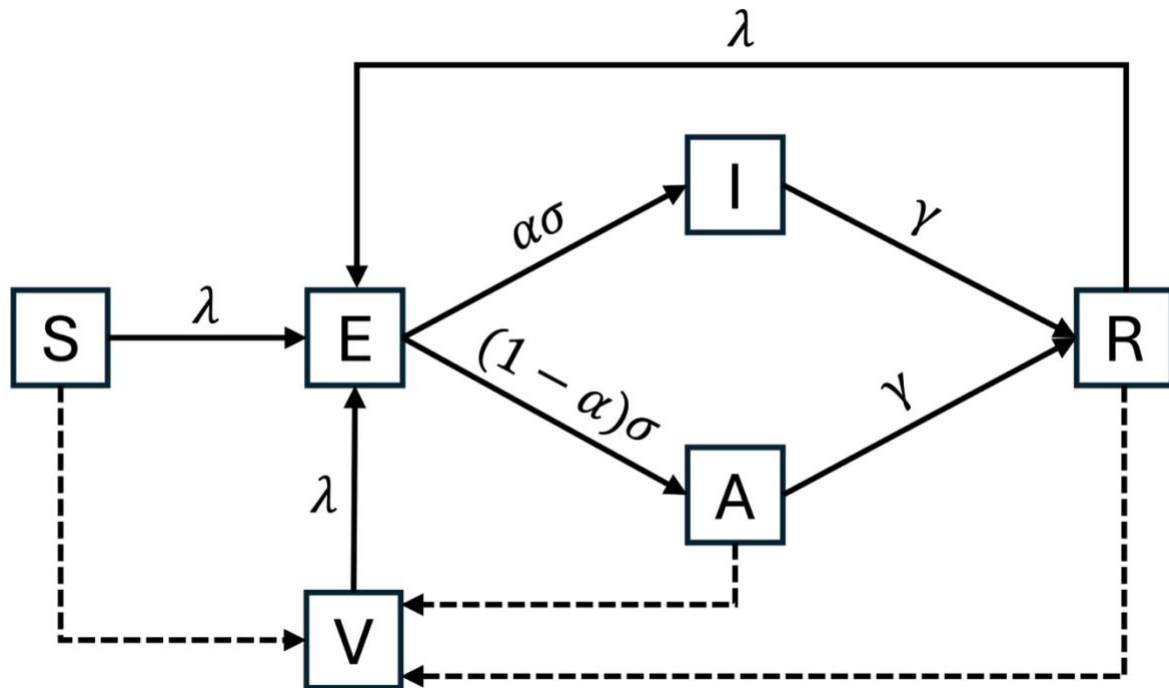
85 2. Methods

86 2.1 Simulation model and assumptions

87 We employed an individual-based transmission model, involving susceptible, exposed,
88 asymptotically infectious, symptomatically infectious, vaccinated, and recovered individuals,
89 to simulate disease transmission within a community (**Figure 1**), with detailed parameters listed
90 in **Table 1**. Individuals were tracked because each individual had a level of a scalar measure of
91 immunity, denoted as X , that varied over time and affected their risk of becoming infected. The
92 model made several key assumptions: The risk of infection decreases linearly as the individual's
93 COP level increases, calculated as $\beta * (I - X) * (I + A)/N$, where β is the transmission
94 coefficient, X is a rescaled COP level ranging from 0 to 1, I is the number of symptomatically
95 infectious individuals, A is the number of asymptotically infectious individuals, and N is the
96 total population. One third of the infectious individuals are assumed to be symptomatic and this
97 proportion did not vary over time or across subgroups.²¹ We assume COP provides equal
98 protection against asymptomatic and symptomatic infection. Individuals cannot receive
99 vaccinations if they are symptomatically infectious. All susceptible individuals start with a COP
100 level at 0 units, and a first-time exposure, including recovery after infection or vaccination, will
101 raise the COP level from 0.00 to 0.75 units. With subsequent exposures, this level will boost
102 from current level to 1 unit. These specific values act as simplified indicators to assess the degree
103 of protection conferred against infections following the first^{22,23} and subsequent²⁴⁻²⁶ exposures,
104 respectively. Additionally, all uninfected individuals experience a linear immunity decline at a
105 rate of 0.01 units per day, modeled as a simplified waning mechanism, whereas exposed and
106 infectious individuals will remain their antibody level unchanged until they recover.²⁷⁻³⁰ Data on
107 each individual's antibody level and infection status is recorded on predetermined simulation

108 dates. In the simulation, 0.5% of the eligible population will be vaccinated every 15 days from
109 day 1 to day 600, totaling 40 rounds. By the end, 20% of the eligible population will receive the
110 vaccine.

111



112

113 **Figure 1. Schematic representation of the model^a**

114 ^aFor detailed descriptions of the parameters used, refer to **Table 1**.

115 **State:** S, Susceptible; E, Exposed; I, Symptomatically Infectious; A, Asymptomatically
116 Infectious; R, Recovered; V, Vaccinated.

117 **Arrow:** Solid arrows represent continuous transitions between states; dashed arrows represent
118 the discrete vaccination schedule as detailed in Methods.

119

120 **Table 1. Model parameters**

Parameter	Description	Value
N	Population size	50,000
t_{latent}	Latent period	5 ³¹
$t_{\text{infectious}}$	Infectious period	7 ³²
α	Symptomatically infectious proportion	1/3 ²¹
σ	Rate from exposed to infectious	1 / t_{latent}
γ	Rate from infectious to recovered	1 / $t_{\text{infectious}}$
R_0	Basic reproduction number	2 ³³
β	Transmission coefficient	$R_0 / t_{\text{infectious}}$
λ	Per capita rate of infection	β * no. of infectious people
w	Antibody level waning rate	0.01
Initial S	No. of initial susceptible cases	48,950
Initial E	No. of initial exposed cases	25
Initial I	No. of initial infectious cases	25
Initial R	No. of initial recovered cases	1000

121

122 **2.2 Data sampling scheme**

123 In our simulation, we implemented a data sampling scheme modeled after a typical TND study.³⁴

124 Symptomatically infectious individuals were identified as cases, and each case was matched with

125 four controls who were susceptible, recovered, or exposed. This sampling scheme implies

126 selection of controls is independent of COP level on that day. It was assumed that all

127 symptomatically infectious individuals tested positive, and the matched controls tested negative.

128 To increase sample size by aggregating data across multiple days, we used an incidence density
129 sampling approach, matching cases and controls based on the day of sampling.

130 **2.3 Statistical analysis**

131 Our analysis considered four regression models: two logistic regression models (with and
132 without transformation of the independent variable, the COP measurement X) and two
133 generalized additive models (GAMs, with and without transformation of X). COP levels were
134 assessed at the time of sampling, reflecting the exposure-proximate COP — the levels measured
135 when individuals tested positive or negative.

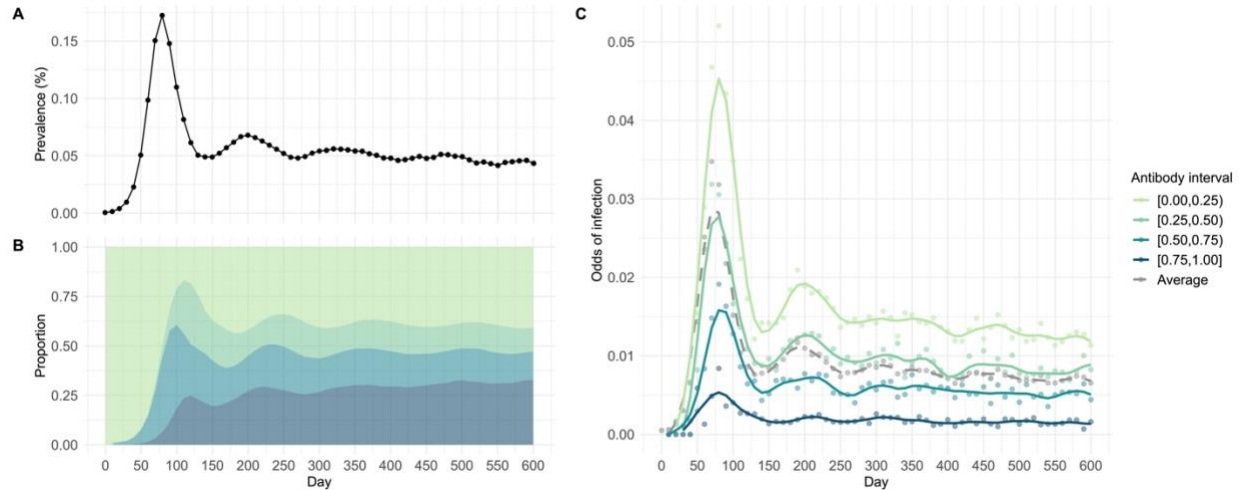
136 Untransformed models modeled $\text{logit}(p) \sim X$ for logistic regression and
137 $\text{logit}(p) \sim \text{GAM}(X)$ for the GAM, while transformed models considered $\text{logit}(p) \sim \ln(1 - X)$
138 for the transformed logistic regression and $\text{logit}(p) \sim \text{GAM}(\ln(1 - X))$ for the transformed
139 GAM. The rationale for the transformation was as follows: The transformation applied to these
140 levels was the natural logarithm of $1 - \text{COP level}$, $\ln(1 - x)$, which is used to reflect the log of
141 IRR expression in our model setting when comparing the infection incidence rate at a specific
142 COP level to the rate at zero COP level. The derivation is as follows: $\ln(\text{IRR}) = \ln\{\lambda * I(t) * U(t) * (1 - X_{X=x}) / [\lambda * I(t) * U(t) * (1 - X_{X=0})]\} = \ln(1 - x)$ where λ is the force of
143 infection, $I(t)$ is the number of infectious individuals at time t , and $U(t)$ is the number of
144 uninfected individuals at time t . Following the logic of TND vaccine studies, in which the OR
145 provides an unbiased estimate of the IRR of infection among vaccinated versus unvaccinated
146 individuals,¹³ we posited that the log of IRR is linearly related to the log OR comparing
147 individuals with a given level of COP to those who have a COP value of zero and hence the log
148 odds.
149

150 Further, the two different infection risk functions—squared ($risk = \beta * (I - X^2) * I/N$)
151 and cubic ($risk = \beta * (I - X^3) * I/N$) transformations of the COP level—were sensitivity
152 analyses used to assess the robustness of the approach to estimating the shape of the relationship
153 between the COP value and incidence when the relationship is more complex and nonlinear.^{35,36}
154 For each of the parametric and semiparametric methods, three models are applied: an
155 untransformed misspecified model, a transformed misspecified model, and a transformed
156 correctly specified model. In both the misspecified and correctly specified transformed models, a
157 natural logarithmic transformation is used. The misspecified transformed models apply linear
158 COP levels, consistent with the primary analysis, while the correctly specified models use
159 squared and cubic COP levels to align with the corresponding infection risk functions.

160 **3. Results**

161 The model simulated the spread of infection over 600 days within a community of 50,000
162 individuals. **Figure 2** shows the temporal dynamics of infectious-and-symptomatic prevalence,
163 COP level proportions, and odds of infection stratified by COP levels. Four complete waves of
164 infection occurred. Notably, declines in the number of individuals with higher antibody levels,
165 due to waning immunity, preceded the emergence of subsequent infection waves. As expected,
166 the simulation shows lower mean and variability in infection odds and lower amplitudes among
167 those with higher COP levels.

168



169

170 **Figure 2. Trends in infectious prevalence, COP proportions, and odds of infection^a from**
171 **simulation days 10 to 600 with an increment of 10**

172 ^a The odds of infection curves are fitted using cubic splines with 15 degrees of freedom.

173 **Panel A:** the prevalence of infectious individuals. **Panel B:** the distribution of the population

174 across different antibody intervals. **Panel C:** the odds of incident infection by COP intervals.

175

176 TND sampling from the simulation showed that when using data collected from a single

177 day, the transformed logistic model accurately recovered the linear relationship between COP

178 level and IRR. In contrast, the untransformed logistic model failed to do so due to

179 misspecification of the functional form (**Figure 3**). Although both semiparametric GAM models

180 generally captured the relationship, they often deviated from the true pattern, particularly at

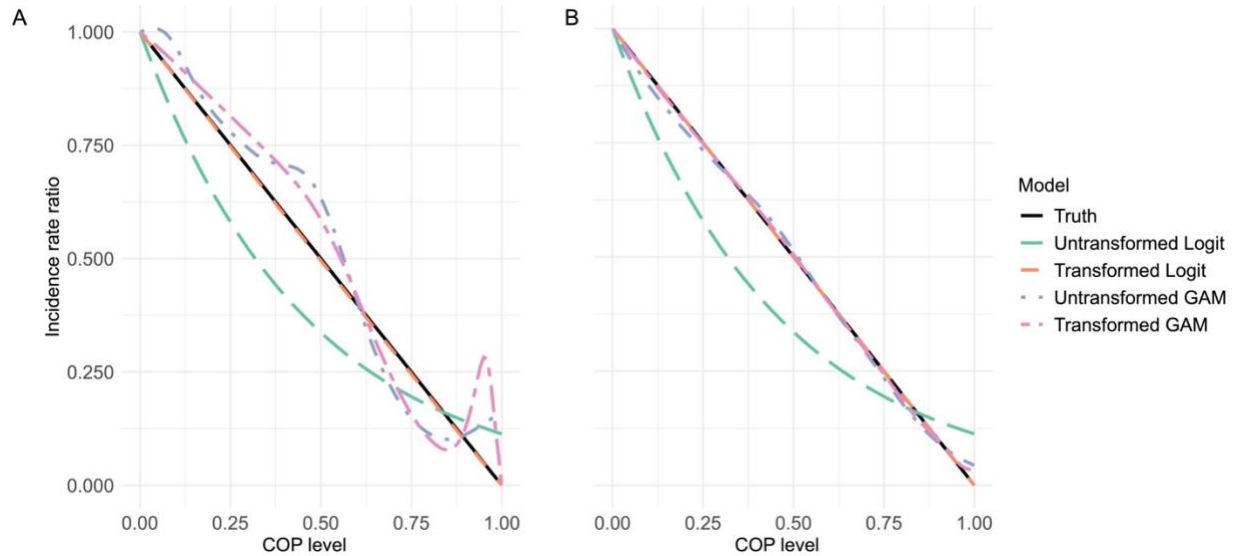
181 extreme COP values (**Figure 3A**). **Figure 3B** illustrates the relationships predicted by these

182 models using compiled incidence density data sampled over multiple days, showing that both

183 GAM models produced results more closely aligned with those of the transformed logistic

184 model.

185



186

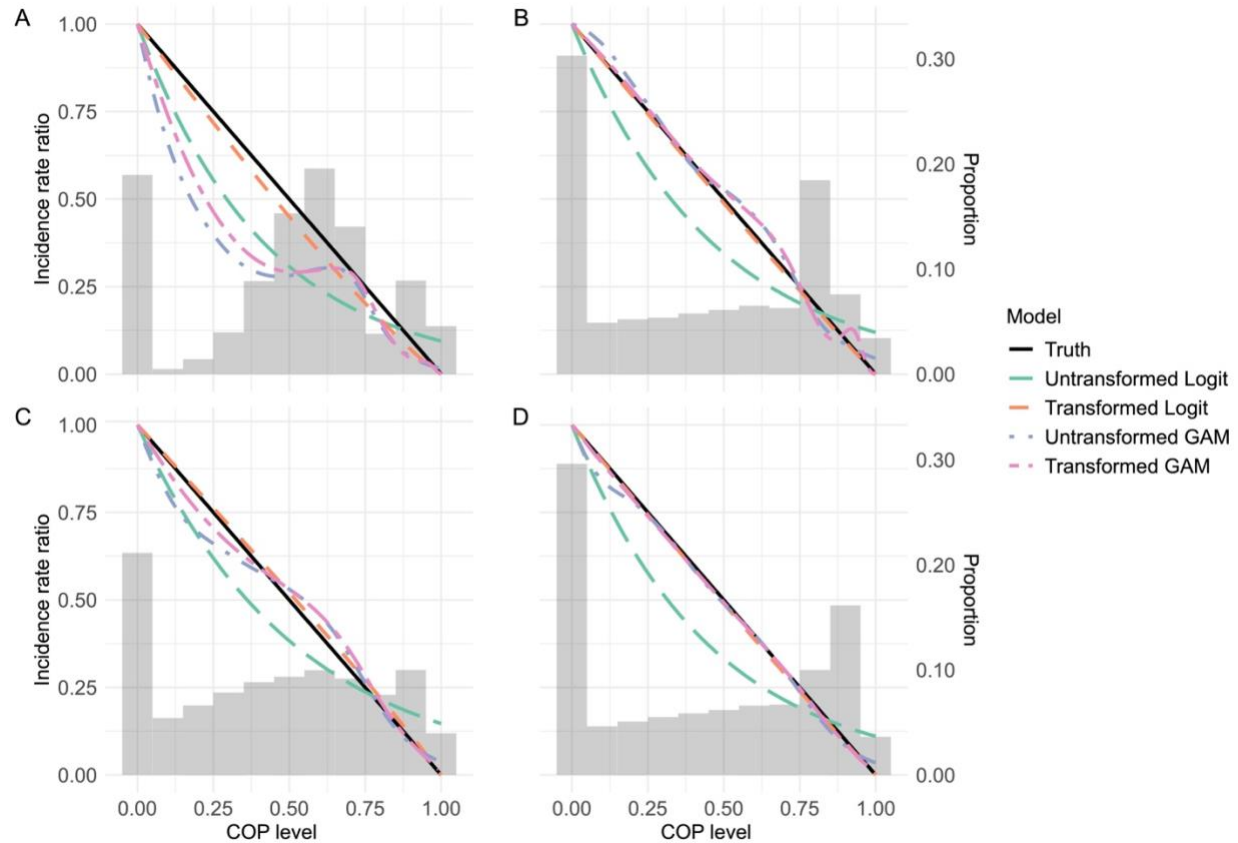
187 **Figure 3. COP level and infection incidence rate ratio sampling on a single simulation day**
188 **and compiled simulation days**

189 **Panel A** used data collected from simulation day 600, **panel B** used data from simulation days
190 500-600 with an increment of 10.

191

192 The impact of population-level antibody distribution and the aggregation of data from
193 multiple days on the accuracy of regression models in predicting infection IRR were
194 investigated. Besides the transformed logistic regression, both GAMs accurately estimate the
195 IRR when sufficient data are aggregated from multiple days using incidence-density sampling
196 (**Figure 4D**). This accuracy is achieved during periods when the pandemic has stabilized after an
197 extended onset, resulting in a diverse and adequately distributed range of COP levels. In
198 situations with more limited numbers of participants with certain COP levels (**Figure 4A, C**) or
199 insufficient data (**Figure 4A, B**) from TND sampling, semiparametric approaches tend to
200 produce more complex estimated relationships, likely due to overfitting the noise in the data. In
201 contrast, the transformed logistic regression consistently recovers the linear relationship.

202



203

204 **Figure 4. Impact of antibody distribution^a and data aggregation on infection incidence rate**
205 **ratio predictions by regression models**

206 ^a Gray areas in the plots indicate the proportions of COP levels within the population.

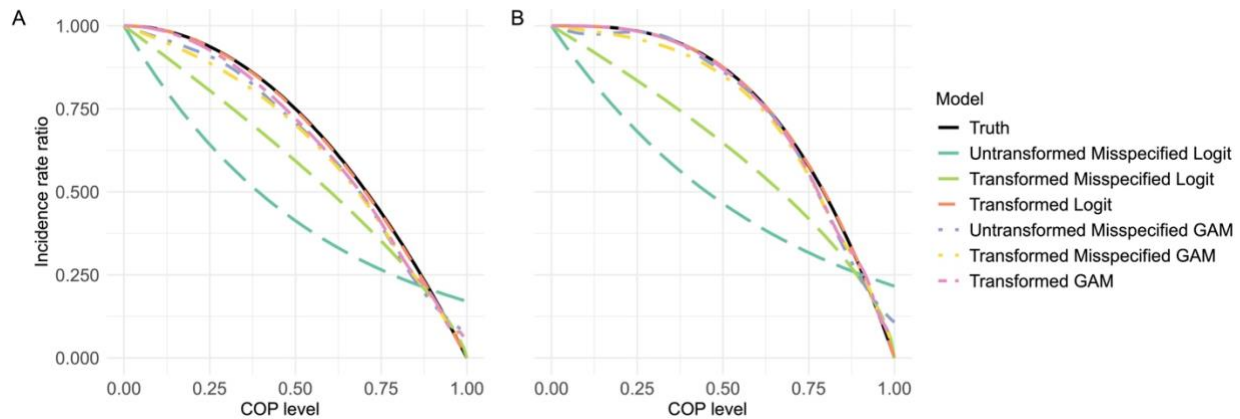
207 **Panel A** used data collected from simulation day 50, **panel B** used data collected from
208 simulation day 450, **panel C** used data from simulation days 50-150 with an increment of 10,
209 and **panel D** used data from simulation days 400-500 with an increment of 10.

210

211 The sensitivity analyses examined the predictability of models on the nonlinear
212 relationships between infection risk and COP levels, including squared and cubic, with sufficient
213 data sampled. The curves generated from the correctly specified transformed logistic regression
214 models are capturing the true relationships, as the primary analysis did, which further shows the
215 robustness of the model selection and transformation function. Although the semiparametric
216 models (whether transformed, correctly specified, or not) did not capture the exact relationship at

217 all levels of COP, they closely approximate the true relationship with only slight fluctuations
218 around the actual lines (**Figure 5**).

219



220

221 **Figure 5. Sensitivity Analysis – COP level and infection incidence rate ratio sampling on**
222 **compiled simulation days 500 to 600 with an increment of 10**

223 **Panel A** used data from simulation where infection risk correlates with squared COP level, and
224 **panel B** used data from simulation where infection risk correlates with cubic COP level.

225 Formulas used for **Panel A** include:

226 Untransformed Misspecified Logit: $\text{logit}(p) \sim X$

227 Transformed Misspecified Logit: $\text{logit}(p) \sim \ln(1 - X)$

228 Transformed Logit: $\text{logit}(p) \sim \ln(1 - X^2)$

229 Untransformed Misspecified GAM: $\text{logit}(p) \sim \text{GAM}(X)$

230 Transformed Misspecified GAM: $\text{logit}(p) \sim \text{GAM}(1 - X)$

231 Transformed GAM: $\text{logit}(p) \sim \text{GAM}(1 - X^2)$

232

233 4. Discussion

234 Antibody levels have been used in prior studies to predict the infection risk of influenza and

235 COVID-19, and are considered one of the appropriate COP for investigating the relationship

236 between infection and individual immunity. However, apart from one recent paper designed for

237 cohort studies,¹¹ there has been little work to define approaches for estimating exposure-

238 proximate COP, that is how an individual's instantaneously measured level of immunity predicts

239 their susceptibility to infection at that moment. While prospective cohorts have many

240 advantages, case-control designs such as the TND are widely employed due to their

241 comparatively low cost and feasibility. Thus, we sought to understand the conditions under
242 which a TND would validly estimate the relationship between a COP level and the degree of
243 protection offered by that level.

244 Our study demonstrates that a TND-style approach can identify the shape of a predictive
245 relationship between a correlate of protection measured near the time of exposure and the risk of
246 infection when using a semiparametric model or when using a correctly specified parametric
247 model with appropriate transformation of the value of the correlate. We have phrased this in
248 terms of prediction rather than causal inference because it is difficult to envision even a
249 hypothetical intervention that would set an individual's COP value at a certain level, and because
250 for many practical purposes, prediction is the question of interest: how well protected is an
251 individual, given a particular COP value? In a causal setting one would also have to consider
252 confounding in which a predictor of COP value had a causal effect on the likelihood of infection.
253 For example, if occupation were predictive of vaccination (and thus COP level) and outcome, or
254 in vaccine campaigns where elderly or immunocompromised individuals are prioritized for early
255 uptake and exhibit lower protection at the same COP value. We have used a model in which such
256 common causes do not exist.

257 Borrowing from the theory of TNDs, the OR estimates the IRR for a particular covariate
258 value. In the simulation, we define the risk of infection as one minus the immunity level.
259 Therefore, in a logistic regression framework, the right-hand side, $\ln(IRR)$ or $\ln(1 - COP)$,
260 should estimate the left-hand side $\ln(odds)$, and our simulations confirm that using this
261 functional form produces estimates that are indistinguishable from the input to the simulation.
262 Alternatively, using a generalized additive model can approximate this relationship if one does
263 not know the proper functional form for the relationship of the COP value to the IRR.

264 In TND, estimating the ‘vaccine direct effect’ for leaky vaccines—those that confer only
265 partial protection to all recipients—is problematic and tends to show declining protection over
266 time. The bias arises because vaccinated individuals continue to experience infections at a
267 reduced rate, while unvaccinated individuals may gain immunity through natural infection.¹³
268 Over time, this dynamic narrows the infection rate gap between vaccinated and unvaccinated
269 groups, causing the OR to trend toward the null. This pattern fails to accurately reflect the
270 vaccine’s true effectiveness. In contrast, this study shows that examining immunity levels as a
271 predictor of infection incidence rate does not suffer from this bias. By analyzing infection rate at
272 varying levels of COP, this approach measures how immunity, regardless of its source,
273 influences infection likelihood. Since it does not rely solely on vaccination status, using COP
274 levels sidesteps the specific biases introduced by leaky vaccine effects in TND. This method,
275 therefore, offers a framework for understanding how incremental COP levels may influence
276 infection incidence rates, while minimizing the impact of vaccine-specific assumptions.

277 The utilization of simulation models stands out as a major strength of this study, allowing
278 us to emulate real-life pandemic scenarios with a degree of control over experimental variables
279 and pre-assumed infection relative risk functions that are not typically possible in field studies.
280 This approach enabled us to systematically test different infection IRR functions, both linear and
281 nonlinear, and to assess their impact on the relationship between immunity levels and infection
282 risks. Additionally, by employing several logistic regression models on the results obtained from
283 TND, we have enhanced the robustness and applicability of our findings to actual pandemic
284 conditions. Despite its strengths, our study is not without limitations. The infection IRR and risk
285 functions used were intentionally simplified, which might have affected the granularity and
286 generalizability of our findings. Similarly, the model simplifies assumptions about immunity

287 boosting and waning, which may not fully capture the complexities of immunity development
288 and decline in diverse populations. Prior studies suggest that immunity waning could be
289 nonlinear,^{35,36} indicating that the model might overlook important variations. Furthermore, the
290 assumption of homogeneity among agents—considering them to have identical susceptibility and
291 transmission characteristics—may not truly reflect the variability observed in real populations.³⁷

292 Future research should aim to incorporate more realistic infection IRR and risk functions
293 and more sophisticated mechanisms for modeling immunity waning and boosting. Additionally,
294 enhancing the model by calibrating it with real-world data that includes detailed agent
295 characteristics and distributions of COP among circulating variants of concern could
296 significantly improve the model's accuracy and relevance. Such advancements are important for
297 developing more effective epidemiological models, which, in turn, can inform public health
298 strategies and vaccination programs more accurately.

299 **5. Conclusion**

300 Antibody levels are vital in epidemiological research, serving as a key metric for evaluating how
301 these COP are associated with infections under various study designs. These insights are crucial
302 for assessing vaccine efficacy and guiding public health interventions. This study shows that
303 employing logistic regression models with natural logarithm transformations of infection IRR
304 function helps to model the relationship between infection incidence rate and antibody levels
305 more precisely, enabling the visualization of both linear and nonlinear effects. To enhance model
306 accuracy, it is essential to refine infection IRR and risk functions and integrate mechanisms of
307 immunity waning and boosting given vaccination or infection within these models. Calibration
308 with real-world data is crucial to confirm model accuracy and relevance. By transitioning from
309 basic theoretical frameworks to more sophisticated, data-driven models, researchers can more

310 effectively simulate the complex interplay between pathogen exposure, immune response, and
311 population health outcomes, advancing the understanding of immunity dynamics and improving
312 the capacity to predict and manage infectious diseases.

313 **References**

- 314 1. Plotkin SA. Correlates of Protection Induced by Vaccination. *Clin Vaccine Immunol.*
315 2010;17(7):1055-1065. doi:10.1128/CVI.00131-10
- 316 2. Black S, Nicolay U, Vesikari T, et al. Hemagglutination Inhibition Antibody Titers as a
317 Correlate of Protection for Inactivated Influenza Vaccines in Children. *Pediatr Infect Dis J.*
318 2011;30(12):1081-1085. doi:10.1097/INF.0b013e3182367662
- 319 3. Ohmit SE, Petrie JG, Cross RT, Johnson E, Monto AS. Influenza Hemagglutination-
320 Inhibition Antibody Titer as a Correlate of Vaccine-Induced Protection. *J Infect Dis.*
321 2011;204(12):1879-1885. doi:10.1093/infdis/jir661
- 322 4. Earle KA, Ambrosino DM, Fiore-Gartland A, et al. Evidence for antibody as a protective
323 correlate for COVID-19 vaccines. *Vaccine.* 2021;39(32):4423-4428.
324 doi:10.1016/j.vaccine.2021.05.063
- 325 5. Khoury DS, Schlub TE, Cromer D, et al. Correlates of Protection, Thresholds of Protection,
326 and Immunobridging among Persons with SARS-CoV-2 Infection. *Emerg Infect Dis.*
327 2023;29(2):381-388. doi:10.3201/eid2902.221422
- 328 6. Gilbert PB, Montefiori DC, McDermott AB, et al. Immune correlates analysis of the mRNA-
329 1273 COVID-19 vaccine efficacy clinical trial. *Science.* 2022;375(6576):43-50.
330 doi:10.1126/science.abm3425
- 331 7. Dimeglio C, Herin F, Martin-Blondel G, Miedougé M, Izopet J. Antibody titers and
332 protection against a SARS-CoV-2 infection. *J Infect.* 2022;84(2):248-288.
333 doi:10.1016/j.jinf.2021.09.013
- 334 8. Feng S, Phillips DJ, White T, et al. Correlates of protection against symptomatic and
335 asymptomatic SARS-CoV-2 infection. *Nat Med.* 2021;27(11):2032-2040.
336 doi:10.1038/s41591-021-01540-1
- 337 9. Goldblatt D, Fiore-Gartland A, Johnson M, et al. Towards a population-based threshold of
338 protection for COVID-19 vaccines. *Vaccine.* 2022;40(2):306-315.
339 doi:10.1016/j.vaccine.2021.12.006
- 340 10. Khoury DS, Cromer D, Reynaldi A, et al. Neutralizing antibody levels are highly predictive
341 of immune protection from symptomatic SARS-CoV-2 infection. *Nat Med.*
342 2021;27(7):1205-1211. doi:10.1038/s41591-021-01377-8
- 343 11. Nilles EJ, Paulino CT, de St Aubin M, et al. Tracking immune correlates of protection for
344 emerging SARS-CoV-2 variants. *Lancet Infect Dis.* 2023;23(2):153-154.
345 doi:10.1016/S1473-3099(23)00001-4
- 346 12. Huang Y, Follmann D. Exposure proximal immune correlates analysis. *Biostatistics.*
347 Published online August 14, 2024:kxae031. doi:10.1093/biostatistics/kxae031

- 348 13. Lewnard JA, Tedijanto C, Cowling BJ, Lipsitch M. Measurement of Vaccine Direct Effects
349 Under the Test-Negative Design. *Am J Epidemiol*. 2018;187(12):2686-2697.
350 doi:10.1093/aje/kwy163
- 351 14. Endo A, Funk S, Kucharski AJ. Bias correction methods for test-negative designs in the
352 presence of misclassification. *Epidemiol Infect*. 2020;148:e216.
353 doi:10.1017/S0950268820002058
- 354 15. Jackson ML, Rothman KJ. Effects of imperfect test sensitivity and specificity on
355 observational studies of influenza vaccine effectiveness. *Vaccine*. 2015;33(11):1313-1316.
356 doi:10.1016/j.vaccine.2015.01.069
- 357 16. De Smedt T, Merrall E, Macina D, Perez-Vilar S, Andrews N, Bollaerts K. Bias due to
358 differential and non-differential disease- and exposure misclassification in studies of vaccine
359 effectiveness. Hozbor DF, ed. *PLOS ONE*. 2018;13(6):e0199180.
360 doi:10.1371/journal.pone.0199180
- 361 17. Lipsitch M, Jha A, Simonsen L. Observational studies and the difficult quest for causality:
362 lessons from vaccine effectiveness and impact studies. *Int J Epidemiol*. Published online July
363 24, 2016:dyw124. doi:10.1093/ije/dyw124
- 364 18. Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative
365 Study Design for Assessment of Influenza Vaccine Effectiveness. *Am J Epidemiol*.
366 2016;184(5):345-353. doi:10.1093/aje/kww064
- 367 19. Westreich D, Hudgens MG. Invited Commentary: Beware the Test-Negative Design. *Am J*
368 *Epidemiol*. 2016;184(5):354-356. doi:10.1093/aje/kww063
- 369 20. Chua H, Feng S, Lewnard JA, et al. The Use of Test-negative Controls to Monitor Vaccine
370 Effectiveness: A Systematic Review of Methodology. *Epidemiology*. 2020;31(1):43-64.
371 doi:10.1097/EDE.0000000000001116
- 372 21. Oran DP, Topol EJ. The Proportion of SARS-CoV-2 Infections That Are Asymptomatic: A
373 Systematic Review. *Ann Intern Med*. 2021;174(5):655-662. doi:10.7326/M20-6976
- 374 22. Kojima N, Klausner JD. Protective immunity after recovery from SARS-CoV-2 infection.
375 *Lancet Infect Dis*. 2022;22(1):12-14. doi:10.1016/S1473-3099(21)00676-9
- 376 23. Wang Z, Muecksch F, Schaefer-Babajew D, et al. Naturally enhanced neutralizing breadth
377 against SARS-CoV-2 one year after infection. *Nature*. 2021;595(7867):426-431.
378 doi:10.1038/s41586-021-03696-9
- 379 24. Hall V, Foulkes S, Insalata F, et al. Protection against SARS-CoV-2 after Covid-19
380 Vaccination and Previous Infection. *N Engl J Med*. 2022;386(13):1207-1220.
381 doi:10.1056/NEJMoa2118691

- 382 25. Glück V, Grobecker S, Köstler J, et al. Immunity after COVID-19 and vaccination: follow-
383 up study over 1 year among medical personnel. *Infection*. 2022;50(2):439-446.
384 doi:10.1007/s15010-021-01703-9
- 385 26. Abbasi J. Study Suggests Lasting Immunity After COVID-19, With a Big Boost From
386 Vaccination. *JAMA*. 2021;326(5):376. doi:10.1001/jama.2021.11717
- 387 27. Goldberg Y, Mandel M, Bar-On YM, et al. Protection and Waning of Natural and Hybrid
388 Immunity to SARS-CoV-2. *N Engl J Med*. 2022;386(23):2201-2212.
389 doi:10.1056/NEJMoa2118946
- 390 28. Andrews N, Tessier E, Stowe J, et al. Duration of Protection against Mild and Severe
391 Disease by Covid-19 Vaccines. *N Engl J Med*. 2022;386(4):340-350.
392 doi:10.1056/NEJMoa2115481
- 393 29. Goldberg Y, Mandel M, Bar-On YM, et al. Waning Immunity after the BNT162b2 Vaccine
394 in Israel. *N Engl J Med*. 2021;385(24). doi:10.1056/NEJMoa2114228
- 395 30. Ferdinands JM, Rao S, Dixon BE, et al. Waning 2-Dose and 3-Dose Effectiveness of mRNA
396 Vaccines Against COVID-19—Associated Emergency Department and Urgent Care
397 Encounters and Hospitalizations Among Adults During Periods of Delta and Omicron
398 Variant Predominance — VISION Network, 10 States, August 2021–January 2022. *MMWR*
399 *Morb Mortal Wkly Rep*. 2022;71(7):255-263. doi:10.15585/mmwr.mm7107e2
- 400 31. Xin H, Li Y, Wu P, et al. Estimating the Latent Period of Coronavirus Disease 2019
401 (COVID-19). *Clin Infect Dis*. 2022;74(9):1678-1681. doi:10.1093/cid/ciab746
- 402 32. Ma S, Zhang J, Zeng M, et al. Epidemiological parameters of coronavirus disease 2019: a
403 pooled analysis of publicly reported individual data of 1155 cases from seven countries.
404 Published online March 24, 2020. doi:10.1101/2020.03.21.20040329
- 405 33. Locatelli I, Trächsel B, Rousson V. Estimating the basic reproduction number for COVID-19
406 in Western Europe. Khudyakov YE, ed. *PLOS ONE*. 2021;16(3):e0248731.
407 doi:10.1371/journal.pone.0248731
- 408 34. Feldstein LR, Self WH, Ferdinands JM, et al. Incorporating Real-time Influenza Detection
409 Into the Test-negative Design for Estimating Influenza Vaccine Effectiveness: The Real-time
410 Test-negative Design (rtTND). *Clin Infect Dis*. 2021;72(9):1669-1675.
411 doi:10.1093/cid/ciaa1453
- 412 35. Pérez-Alós L, Armenteros JJA, Madsen JR, et al. Modeling of waning immunity after SARS-
413 CoV-2 vaccination and influencing factors. *Nat Commun*. 2022;13(1):1614.
414 doi:10.1038/s41467-022-29225-4
- 415 36. Ghosh SK, Ghosh S. A mathematical model for COVID-19 considering waning immunity,
416 vaccination and control measures. *Sci Rep*. 2023;13(1):3610. doi:10.1038/s41598-023-
417 30800-y

418 37. Bushman M, Kahn R, Taylor BP, Lipsitch M, Hanage WP. Population impact of SARS-
419 CoV-2 variants with enhanced transmissibility and/or partial immune escape. *Cell*.
420 2021;184(26):6229-6242.e18. doi:10.1016/j.cell.2021.11.026

421