Correlates of Protection through multidimensional immune modelling across respiratory viruses

David Hodgson | Dec 2025 | SWIM 2025
Charité — Universitätsmedizin Berlin

Charité Center
for Global Health

TALK AIM

"To systematically identify and compare correlates of protection across multiple biomarkers for respiratory viruses using rigorous statistical methods"

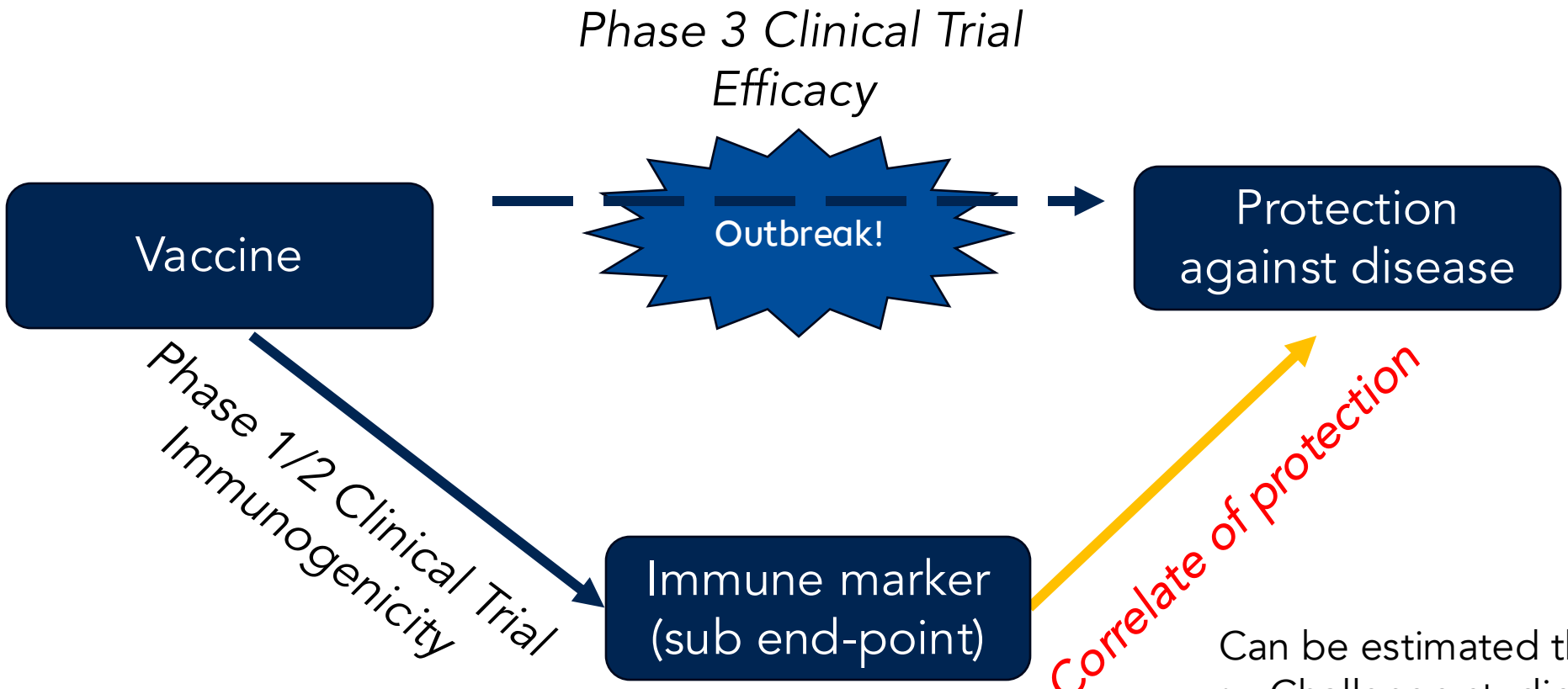
Specific objectives

1. Develop a framework for determining CoP in natural history cohort studies
   - Estimate CoR and CoP using serological and infection history data
   - Address key challenge: cannot directly measure exposure in real-world setting

2. Identify the "best" single biomarker CoP of biomarkers
   - Compare predictive capacity across multiple serum and mucosal biomarkers
   - Apply rigorous statistical criteria (AUC, out of sample prediction)

3. Assess value of combined biomarker CoP models
   - Test whether combining serum and mucosal markers improves predictive capacity
   - Quantify added benefit beyond single biomarkers

WHAT IS A CORRELATE OF PROTECTION?



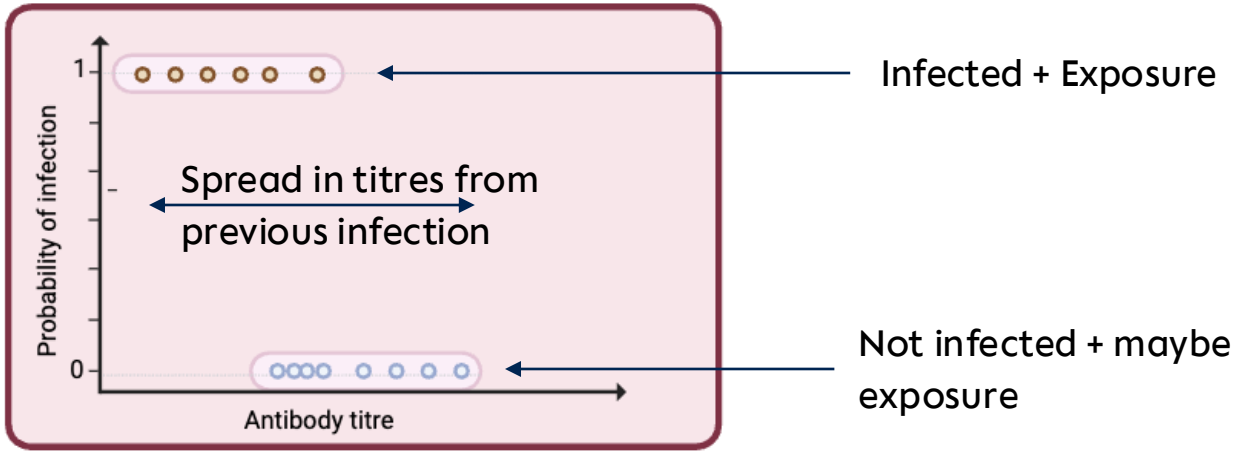
Can be estimated through:

- Challenge studies (spenny, ethics?)
- Seroepi-studies

NATURAL HISTORY STUDIES (NOT RCT)

Use of cohort studies

- In cohort studies, we can identify correlates by studying people who've been naturally infected previously



Limitation

- Don't know who's exposed
- No randomisation, hard to say anything truly causal as about these correlates of protection
- Thus, a correlate of risk and correlate of protection in this context has literal interpretation

STATISTICAL METHODS

Assume a continuous relationship between titre and infection

We fit a (bayesian) generalised logistic curve to the CoR/infection risk with:

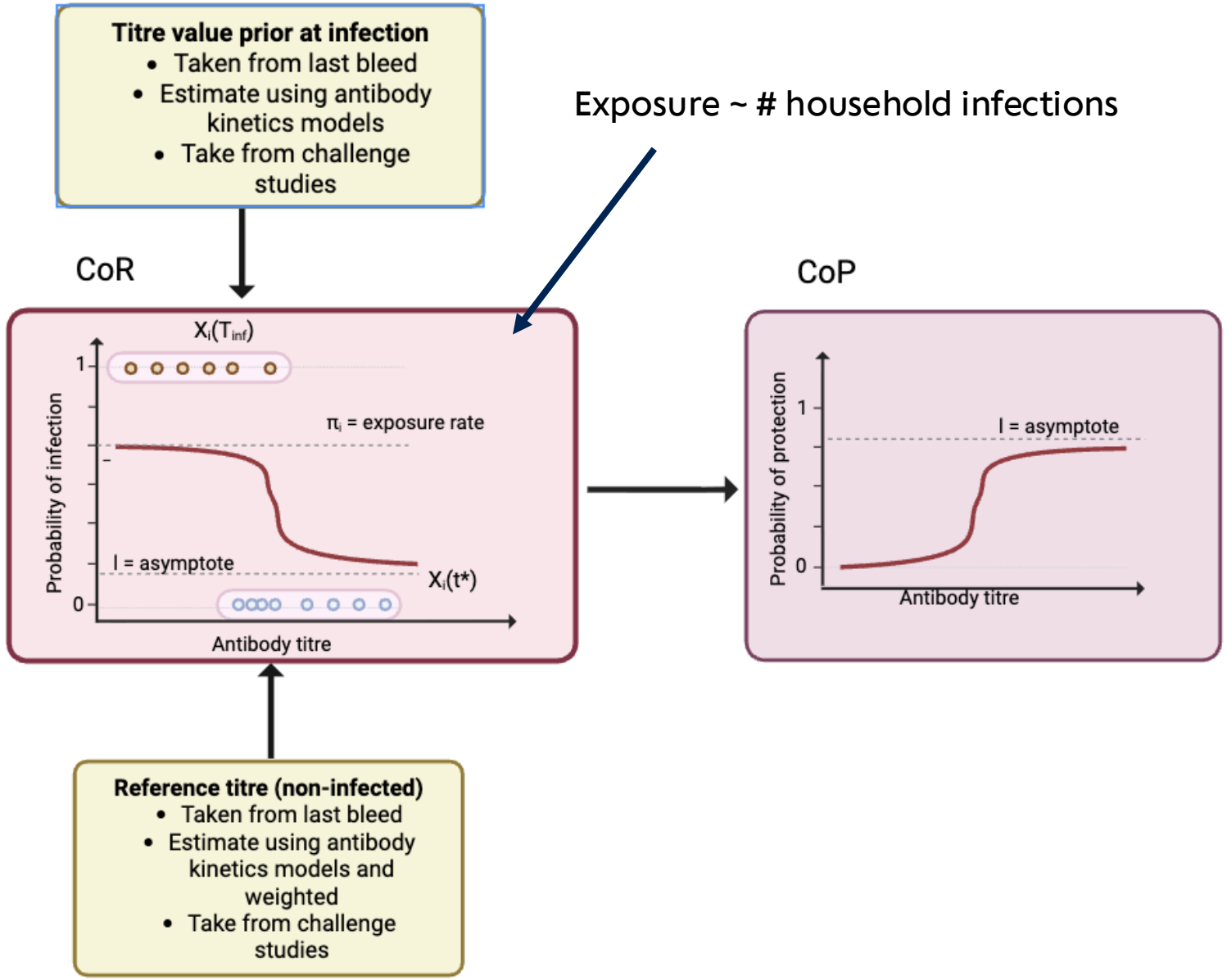
- Upper asymptope = exposure rate
- Lower asymptope = antibody doesn't provide full protection

To get CoP, marginalise out the exposure and find the inverse.

In maths:

COR := $\pi[1 - f(\mathbf{x}, \boldsymbol{\beta})]$ <- we fit this

COP := $f(\mathbf{x}, \boldsymbol{\beta})$



STATISTICAL METHODS

We fit a logistic curve to the CoR with

- Upper asymptope = exposure rate
- Lower asymptope = antibody doesn't provide full protection

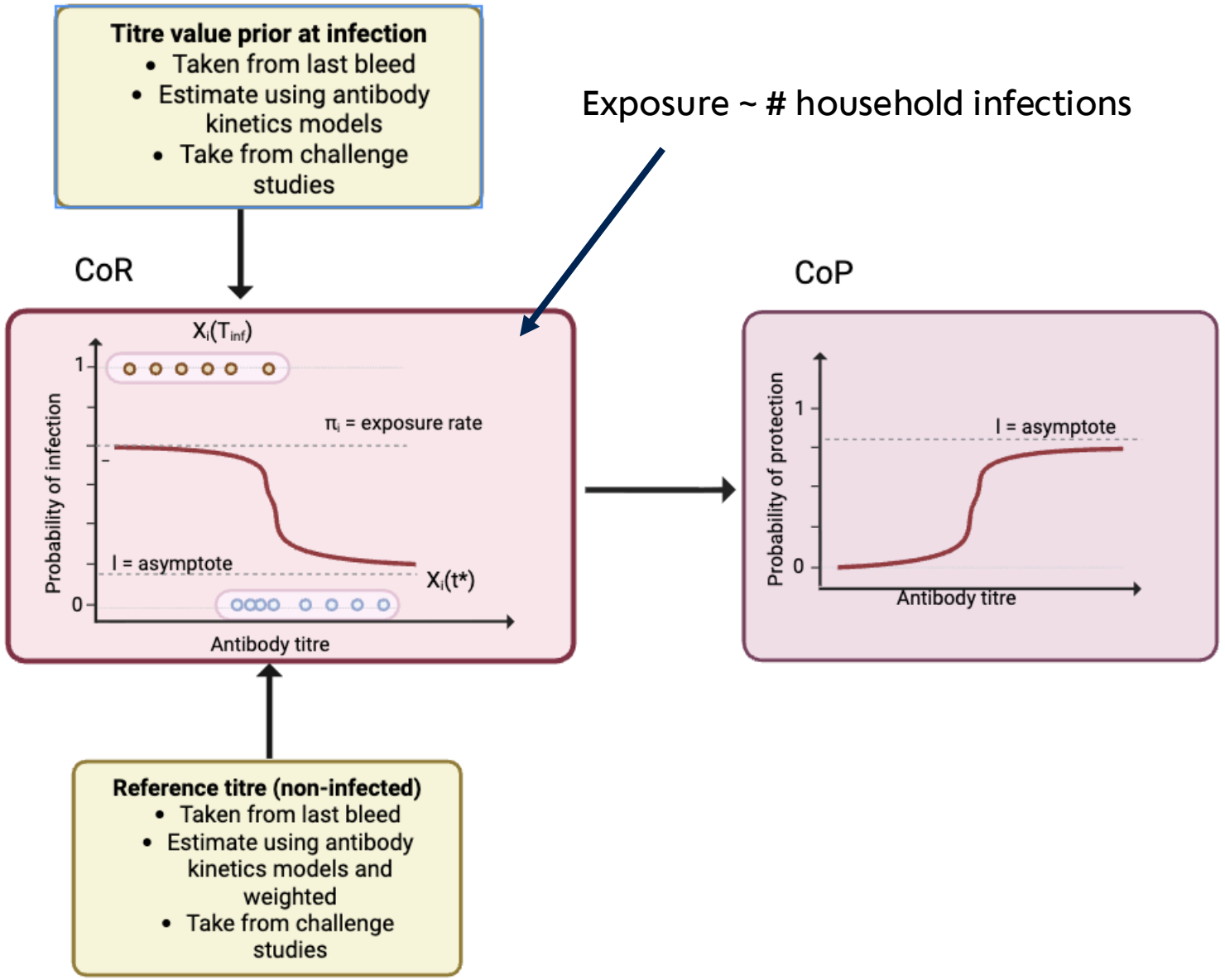
To get CoP, marginalise out the exposure and find the inverse.

In maths:

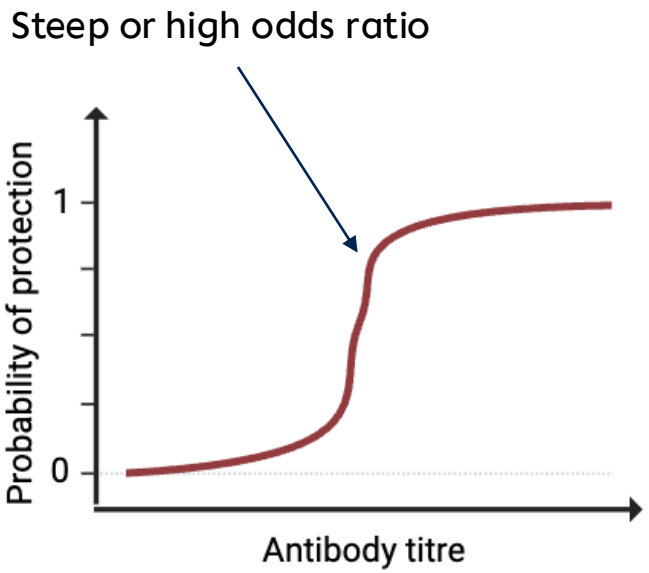
COR := $\pi[1 - f(x, \boldsymbol{\beta})]$ <- we fit this

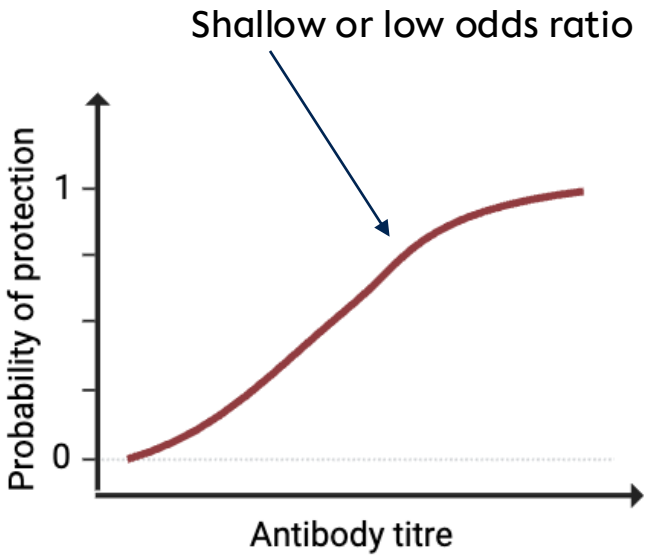
COP := $f(\mathbf{x}, \boldsymbol{\beta})$

-> 2 biomarkers, multidimensional logistical regression

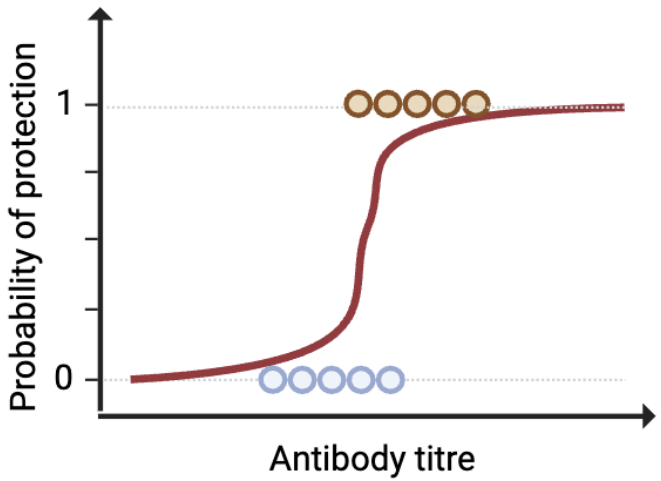


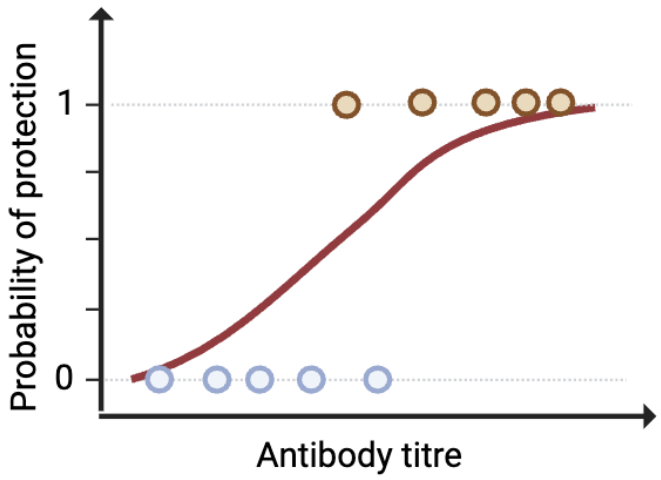
ASSESSMENT OF "GOODNESS" OF COP



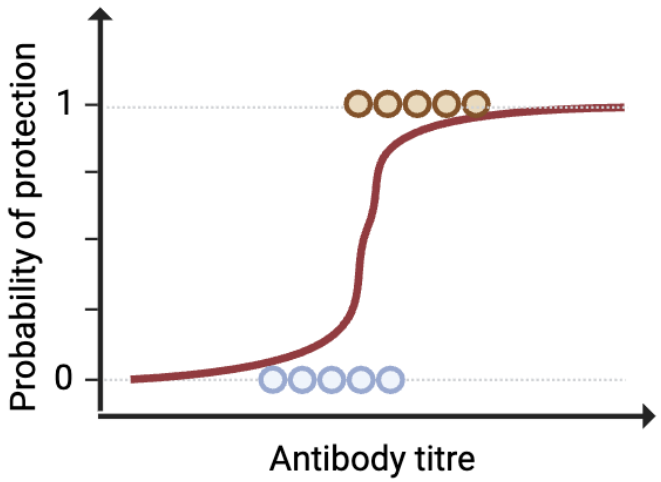


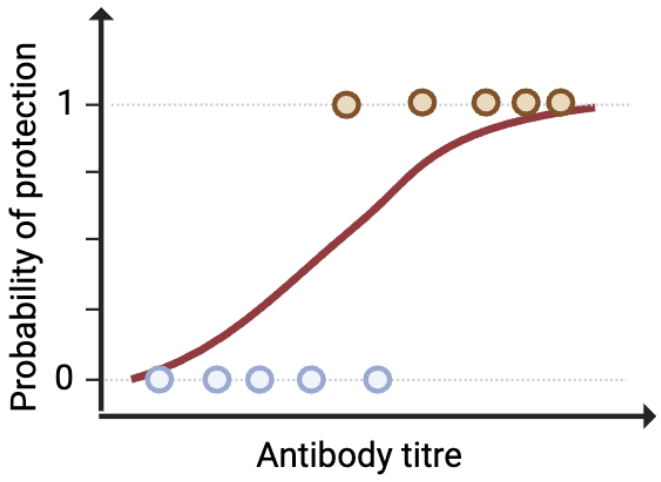
ASSESSMENT OF "GOODNESS" OF COP





ASSESSMENT OF "GOODNESS" OF COP

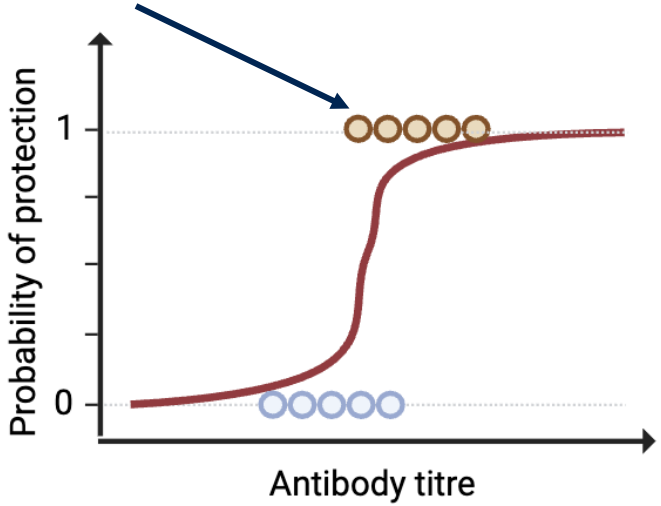




If you are making inferences using fitted model which isn't able to discrimate between those who are infection and those who are not, then it has limited practical use as a CoP => more of an association of protection

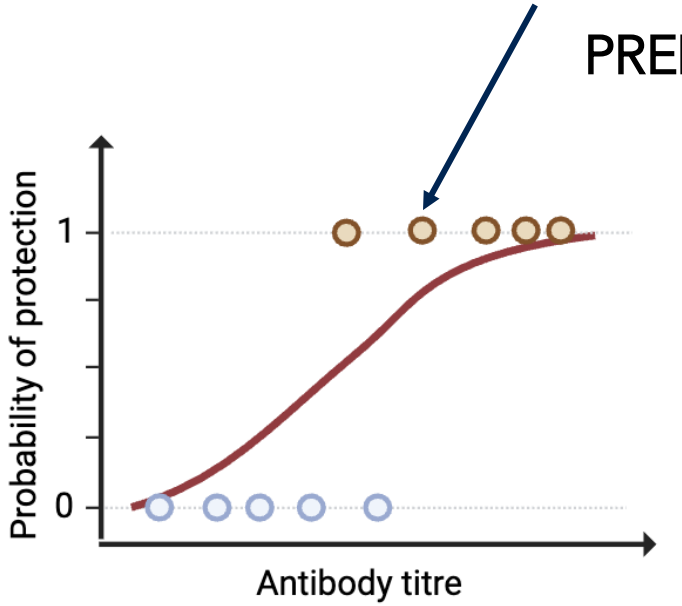
ASSESSMENT OF "GOODNESS" OF COP

Could have poor predictive performance



Could have better predictive performance

PREFERRED!!



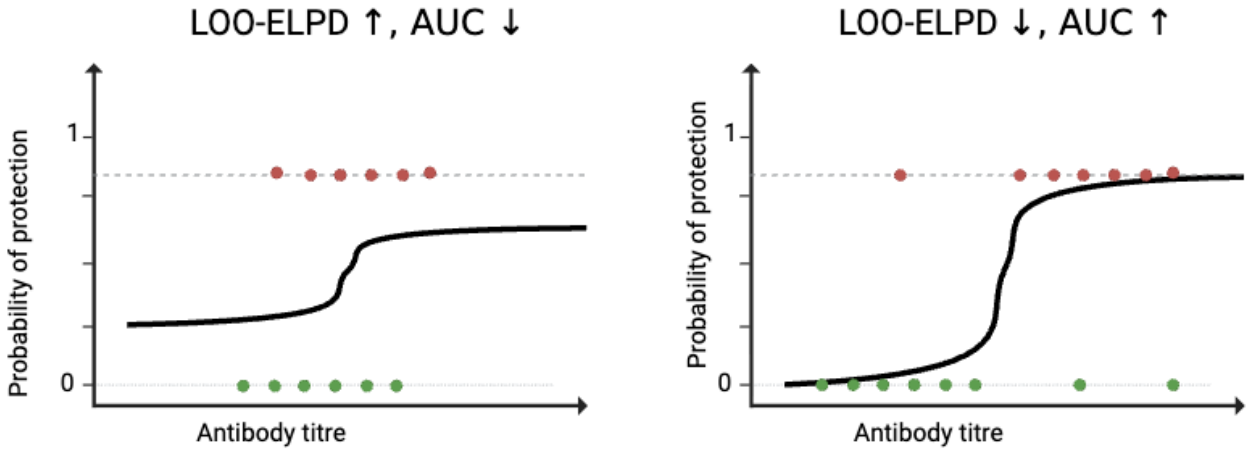
PLAN:

- Compare predictive performance of the fitted curve for each correlate
- Choose the biomarker with the best predictive performance -> better support for causality
- SIDE NOTE: generally in 1D [odds ratio + p-value] ≈ performance, but not true in higher dimensions

METRICS USED

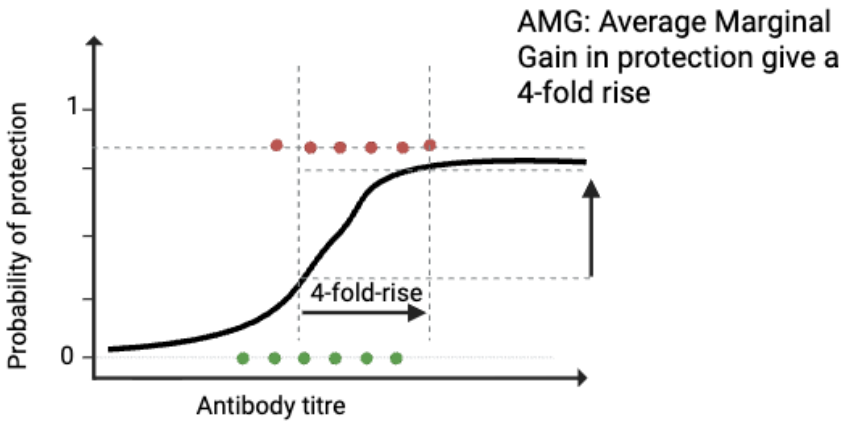
Predictive performance

- Discrimination: AUC
- Out-of-sample predictive fit: LOO-ELPD



Protection impact and applicability

- Impact: AMG, β
- Applicability: Coverage



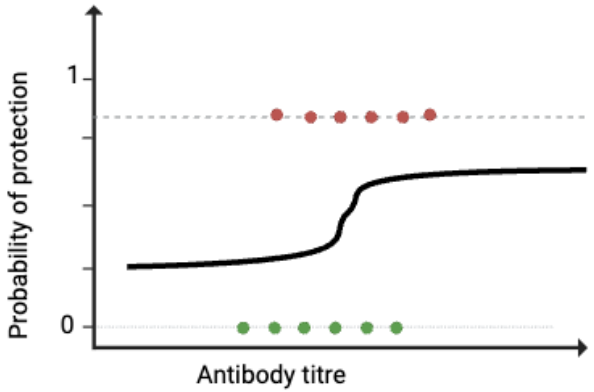
METRICS USED

Predictive performance

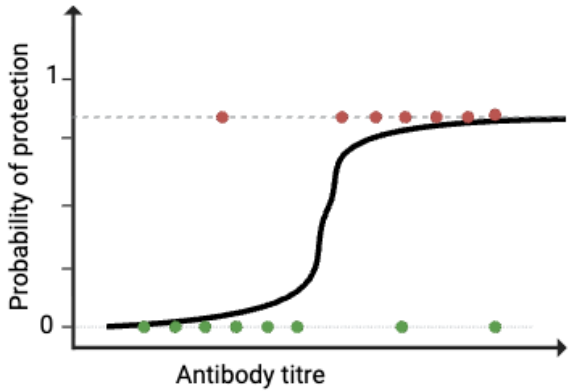
- Discrimination: AUC
- Out-of-sample predictive fit: LOO-ELPD

Trans-dimensionally comparable

LOO-ELPD ↑, AUC ↓



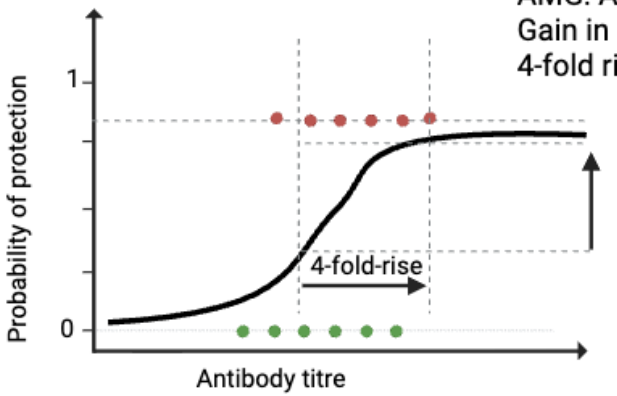
LOO-ELPD ↓, AUC ↑



Protection impact and applicability

- Impact: AMG, β
- Applicability: Coverage

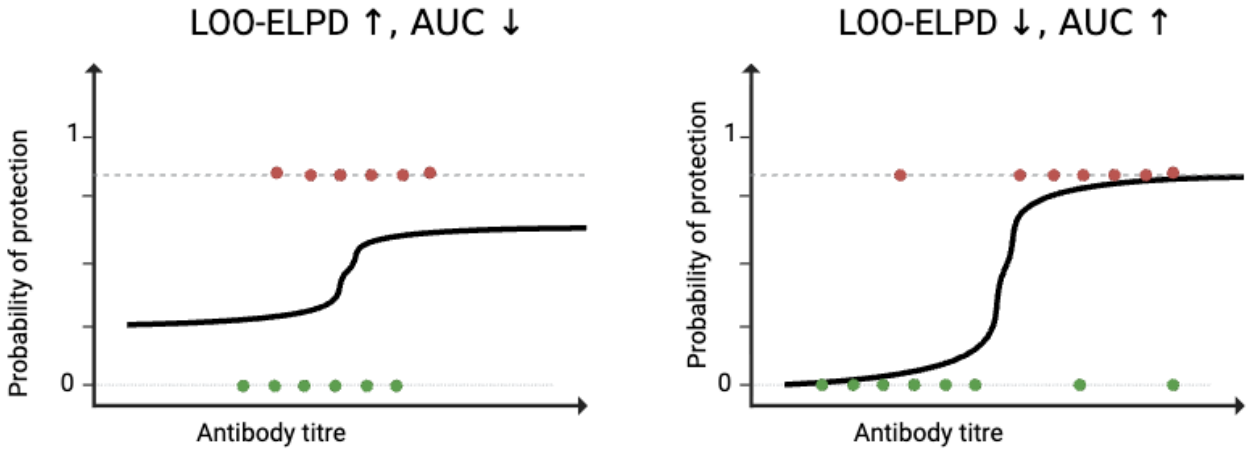
AMG: Average Marginal Gain in protection give a 4-fold rise



METRICS USED

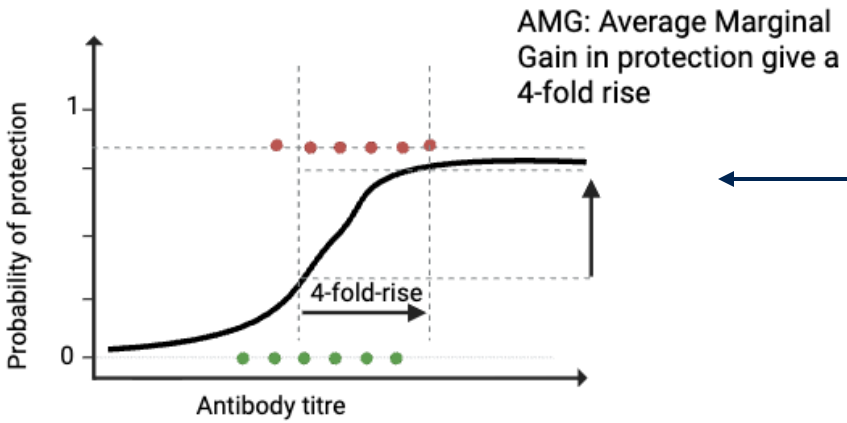
Predictive performance

- Discrimination: AUC
- Out-of-sample predictive fit: LOO-ELPD



Protection impact and applicability

- Impact: AMG, β
- Applicability: Coverage



If we boosted everyone's pre-exposure titre by 4-fold, how much more protected would the population be?

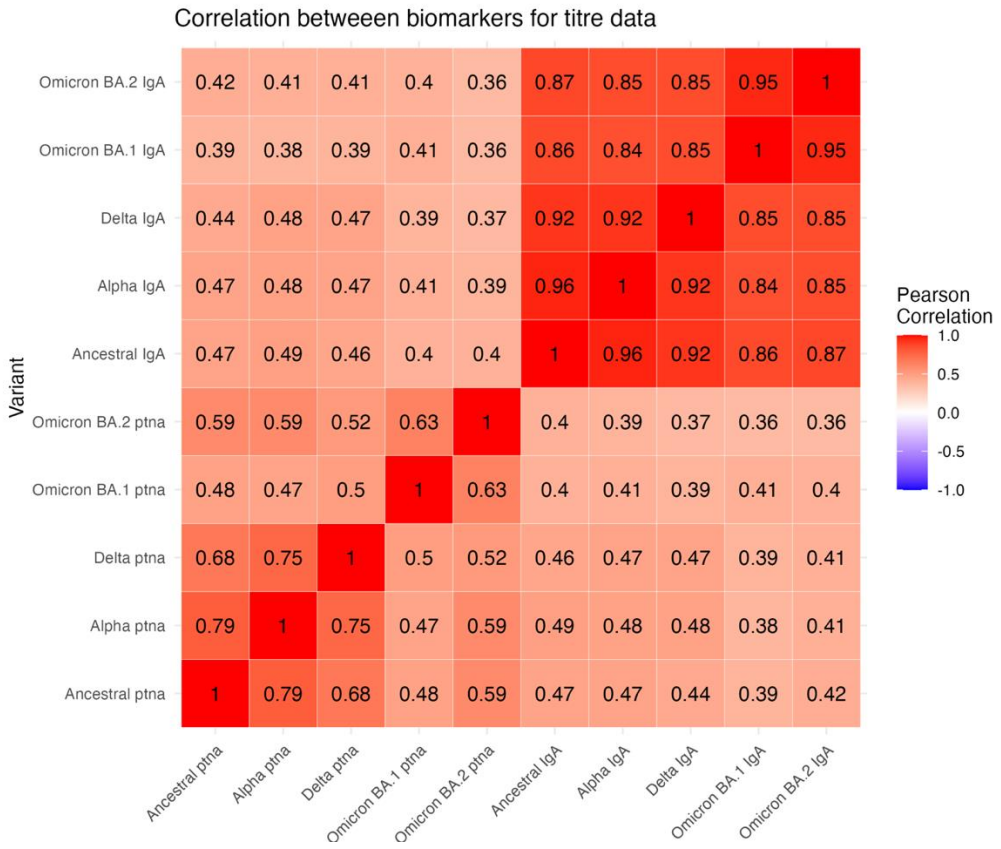
CASE 1: SARS-CoV-2 in The Gambia

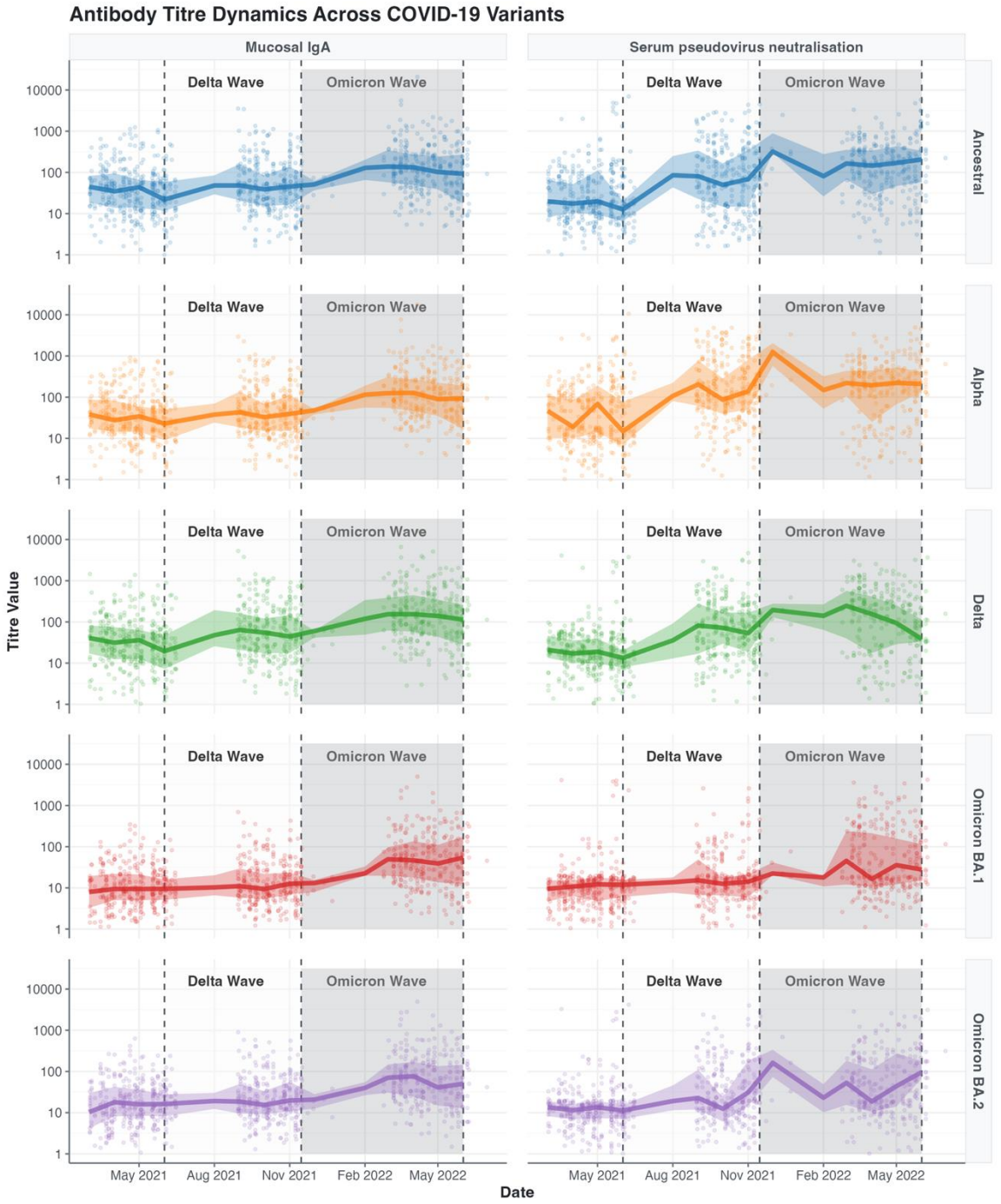
TRANSVIR Study (vaccine naïve)

N = 256 people, 308 days, 2 bleeds person,

Two wave; Delta wave and Omicron BA.1 wave

PCR swabbing weekly, CoP against infection (~70% asymptomatic)





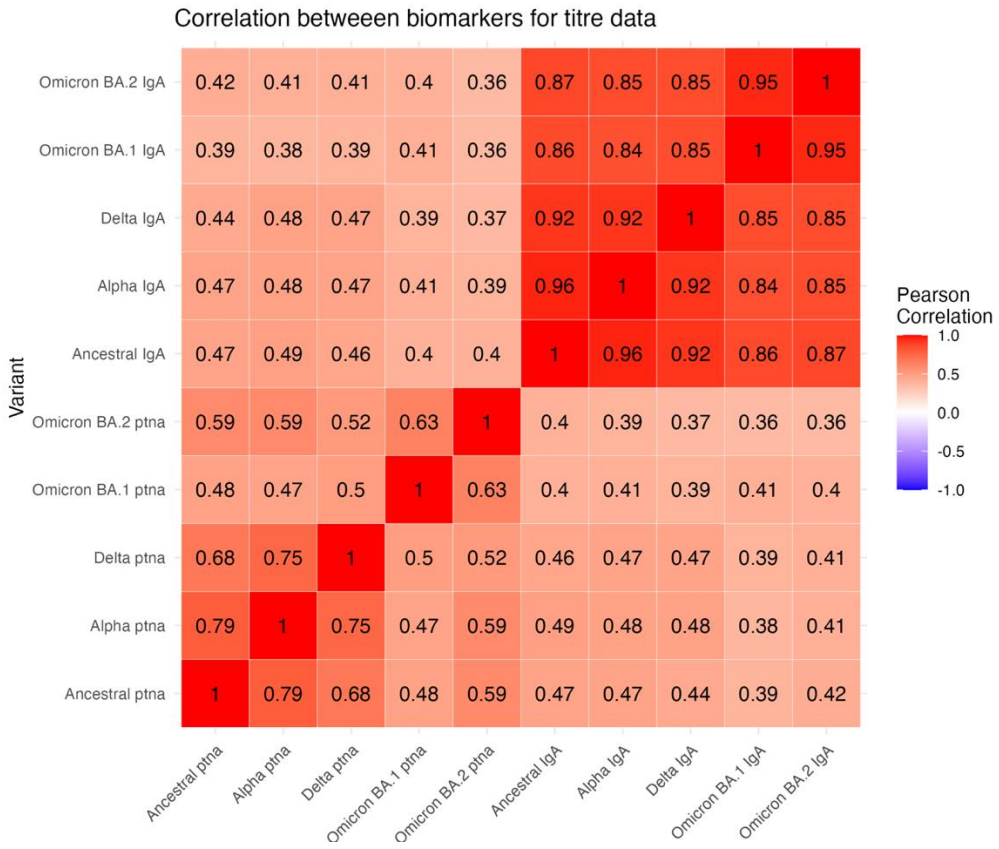
CASE 1: SARS-CoV-2 in The Gambia

TRANSVIR Study (vaccine naïve)

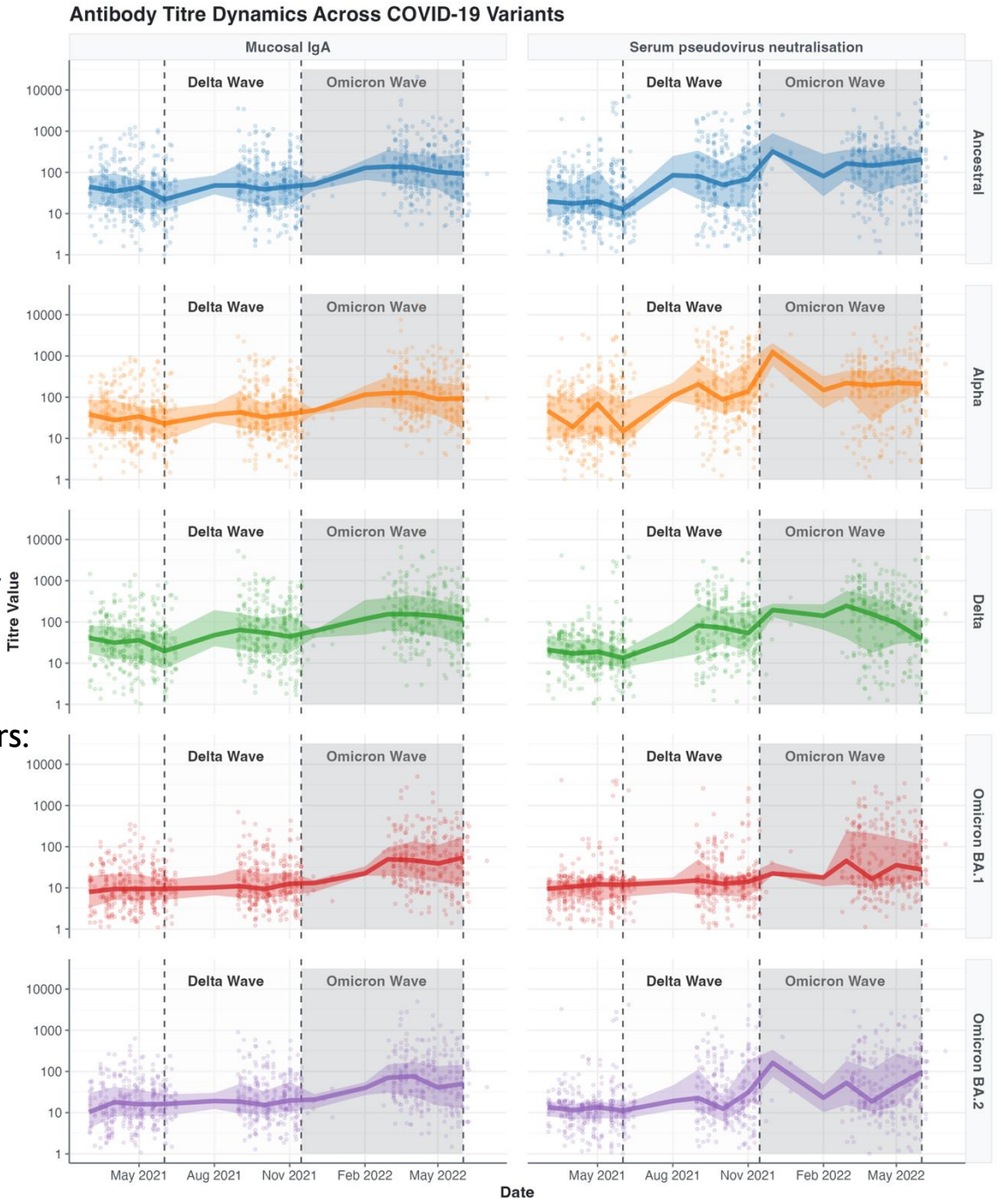
N = 256 people, 308 days, 2 bleeds person,

Two wave; Delta wave and Omicron BA.1 wave

PCR swabbing weekly, CoP against infection (~70% asymptomatic)



10 biomarkers:



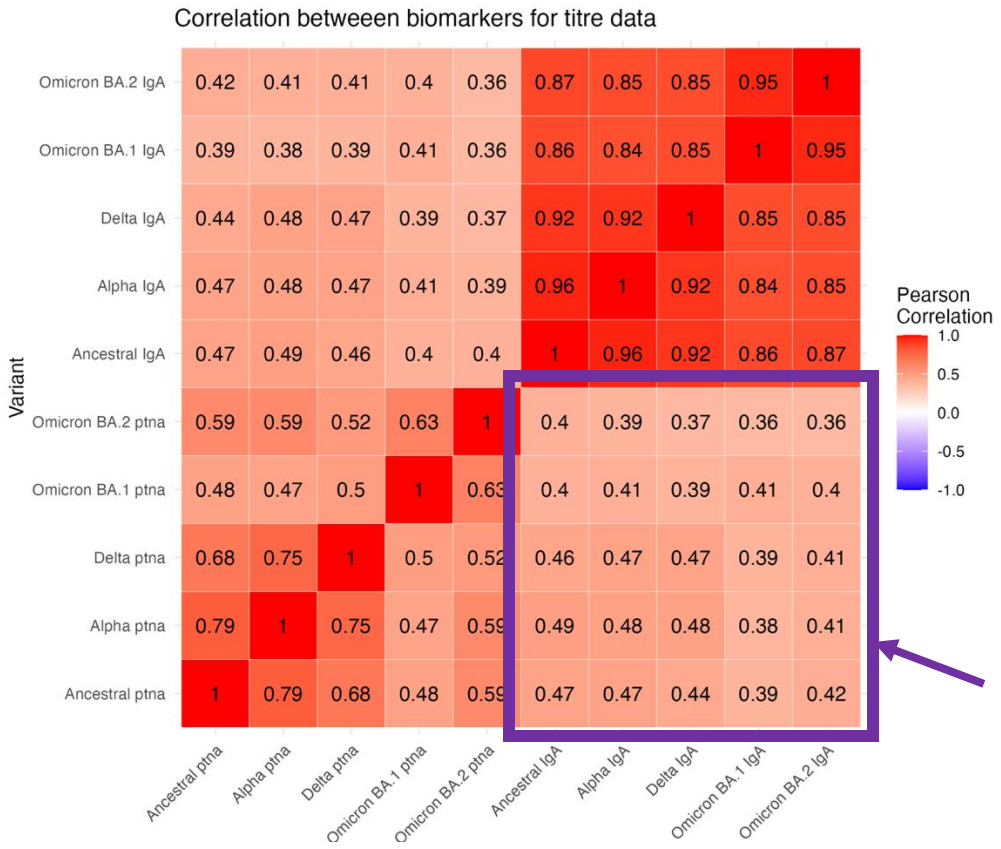
CASE 1: SARS-CoV-2 in The Gambia

TRANSVIR Study (vaccine naïve)

N = 256 people, 308 days, 2 bleeds person,

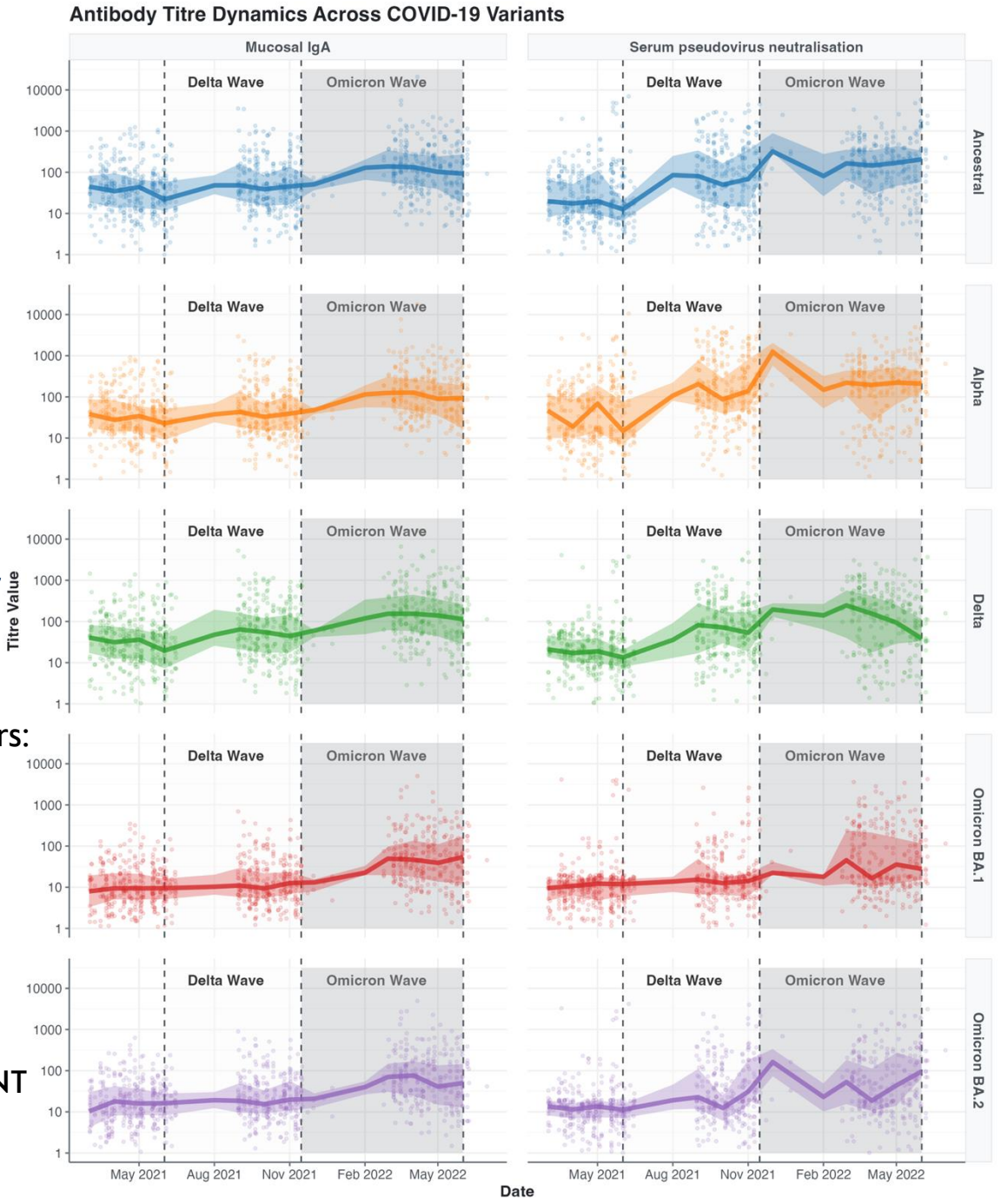
Two wave; Delta wave and Omicron BA.1 wave

PCR swabbing weekly, CoP against infection (~70% asymptomatic)

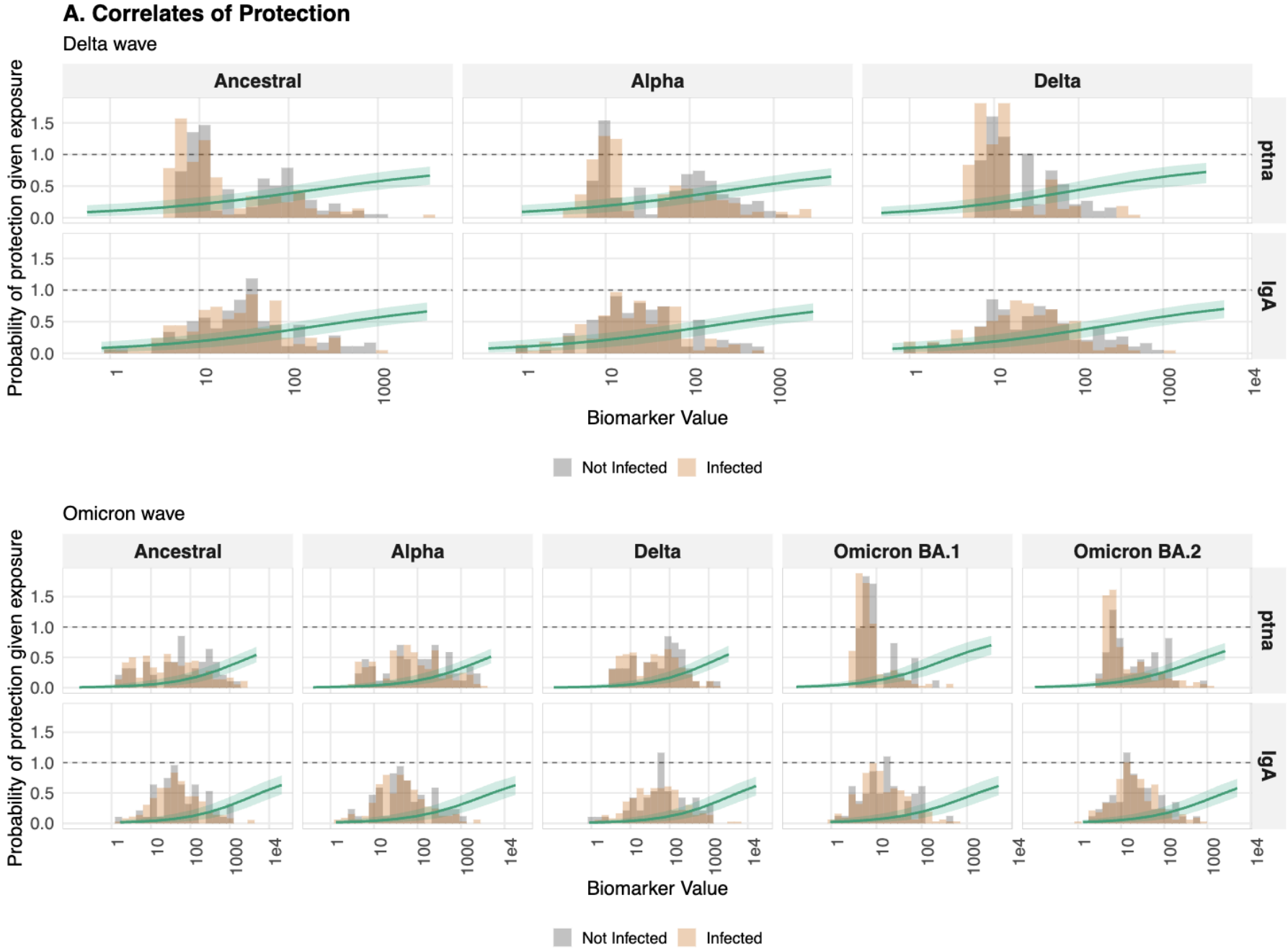


10 biomarkers:

Medium correlation between pVNT and mIgA



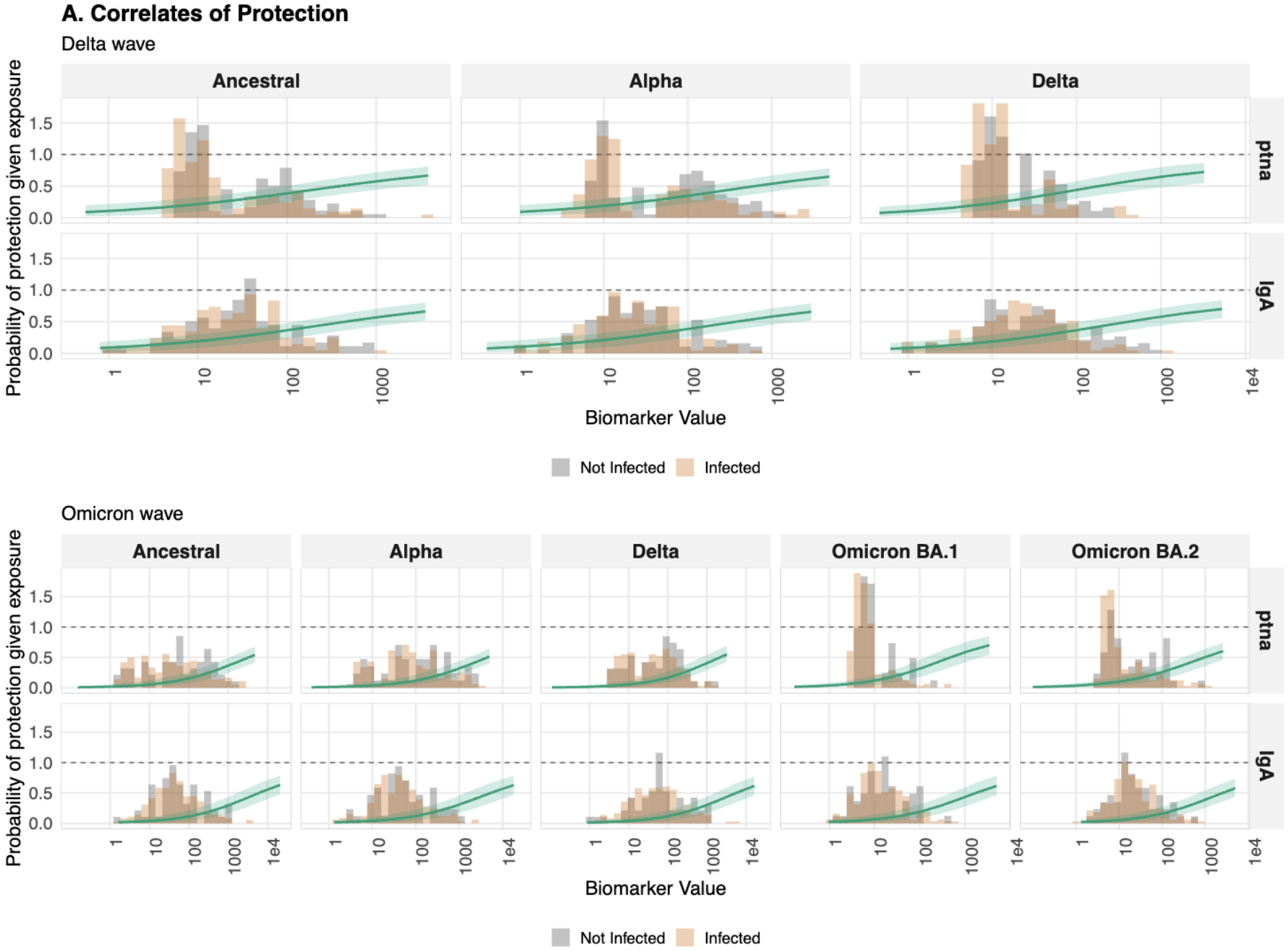
RESULTS FOR SARS-CoV-2



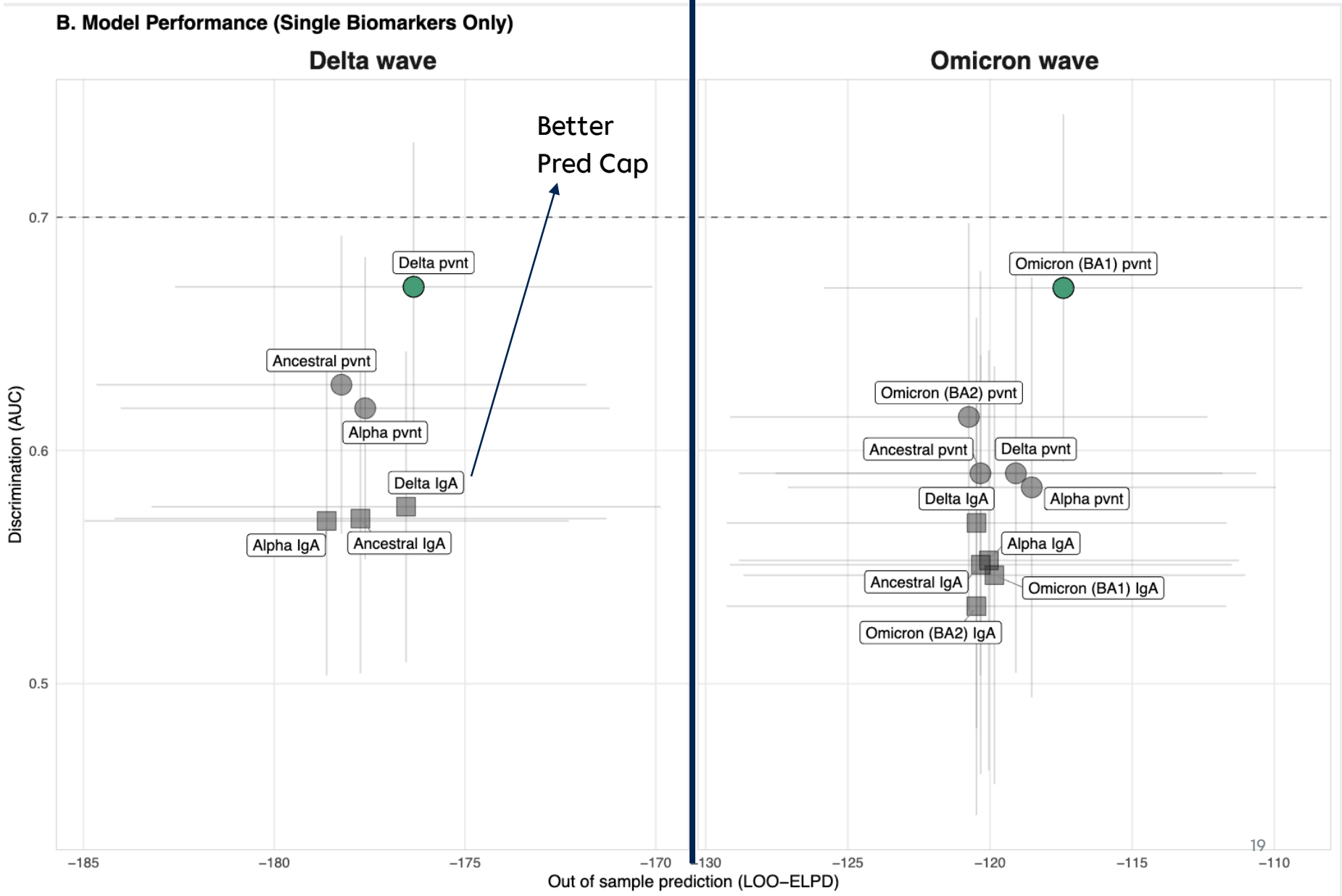
RESULTS FOR SARS-CoV-2

Which of these biomarkers is the best COP?

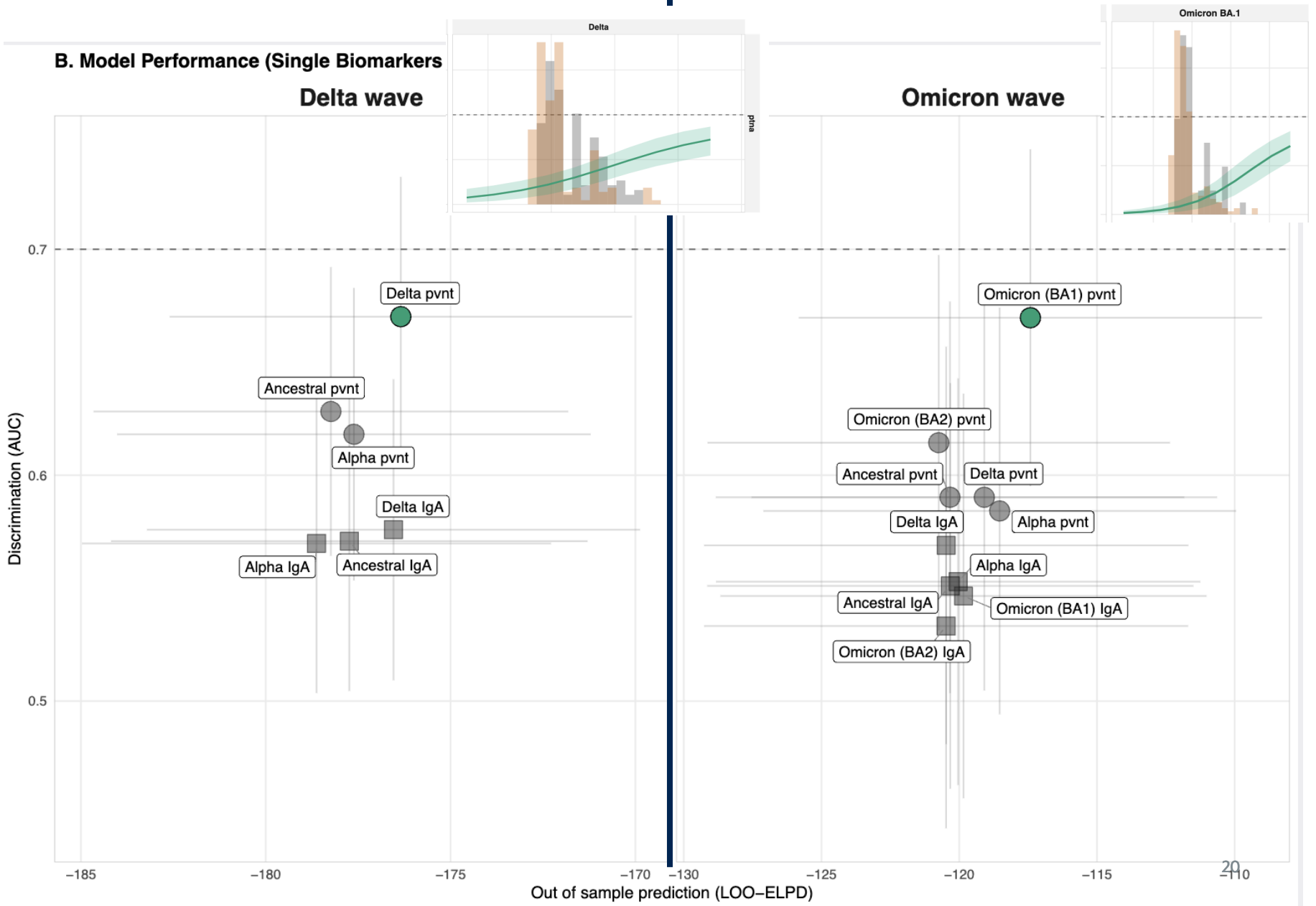
Determine predictive capacity!!



RESULTS FOR SARS-CoV-2

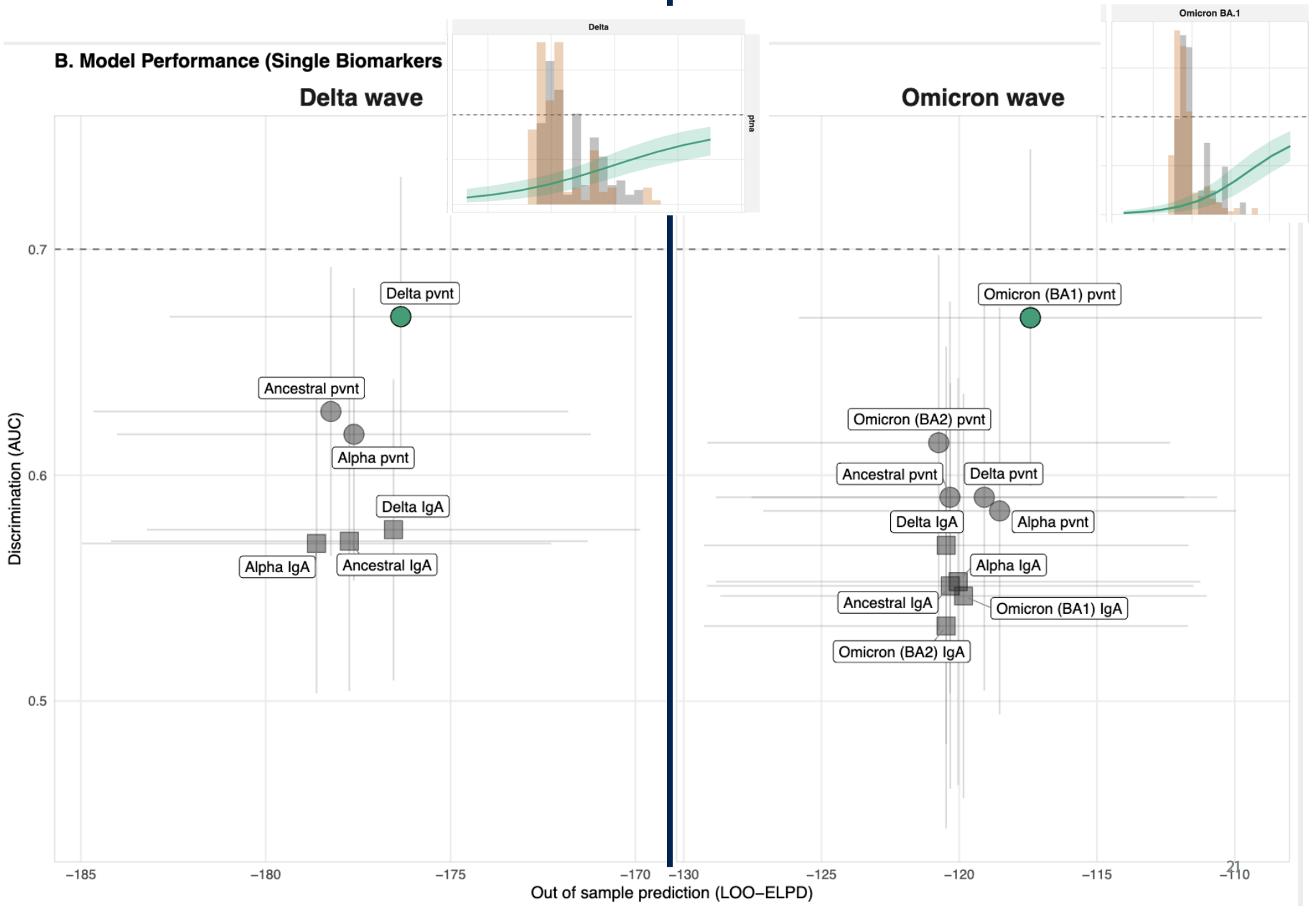


RESULTS FOR SARS-CoV-2



RESULTS FOR SARS-CoV-2

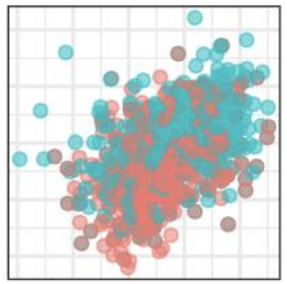
What happens if we look at combining Delta pNTA and Delta mIgA?



RESULTS FOR SARS-CoV-2

Adding mIgA to pTNA decreases predictive capacity!

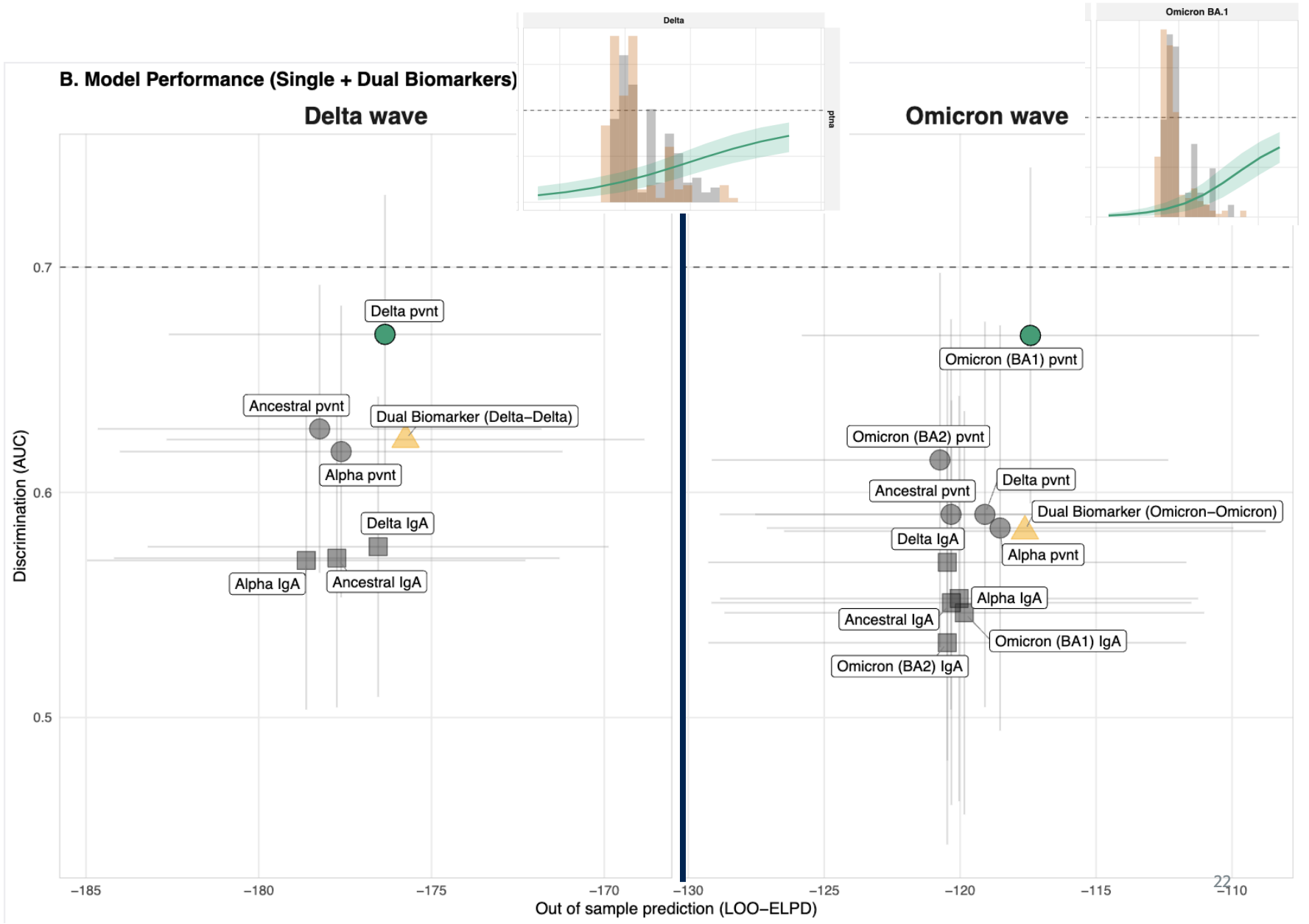
Why?



* Correlation between pTNA and mIgA
* mIgA very noisey

2 dimensional model overfits—pTNA only better

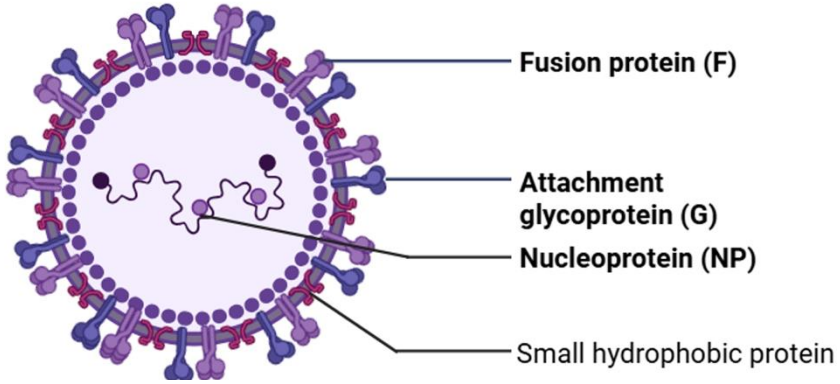
Best to stick to pTNA only
Make sense, pTNA is a functional measure, dominated binding assay

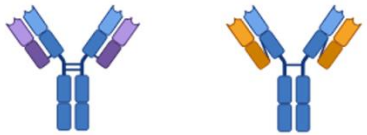


CASE 2: RSV in The Gambia

TRANSVIR Study (vaccine naïve)

N = 256 people, 308 days, 2–5 bleeds person





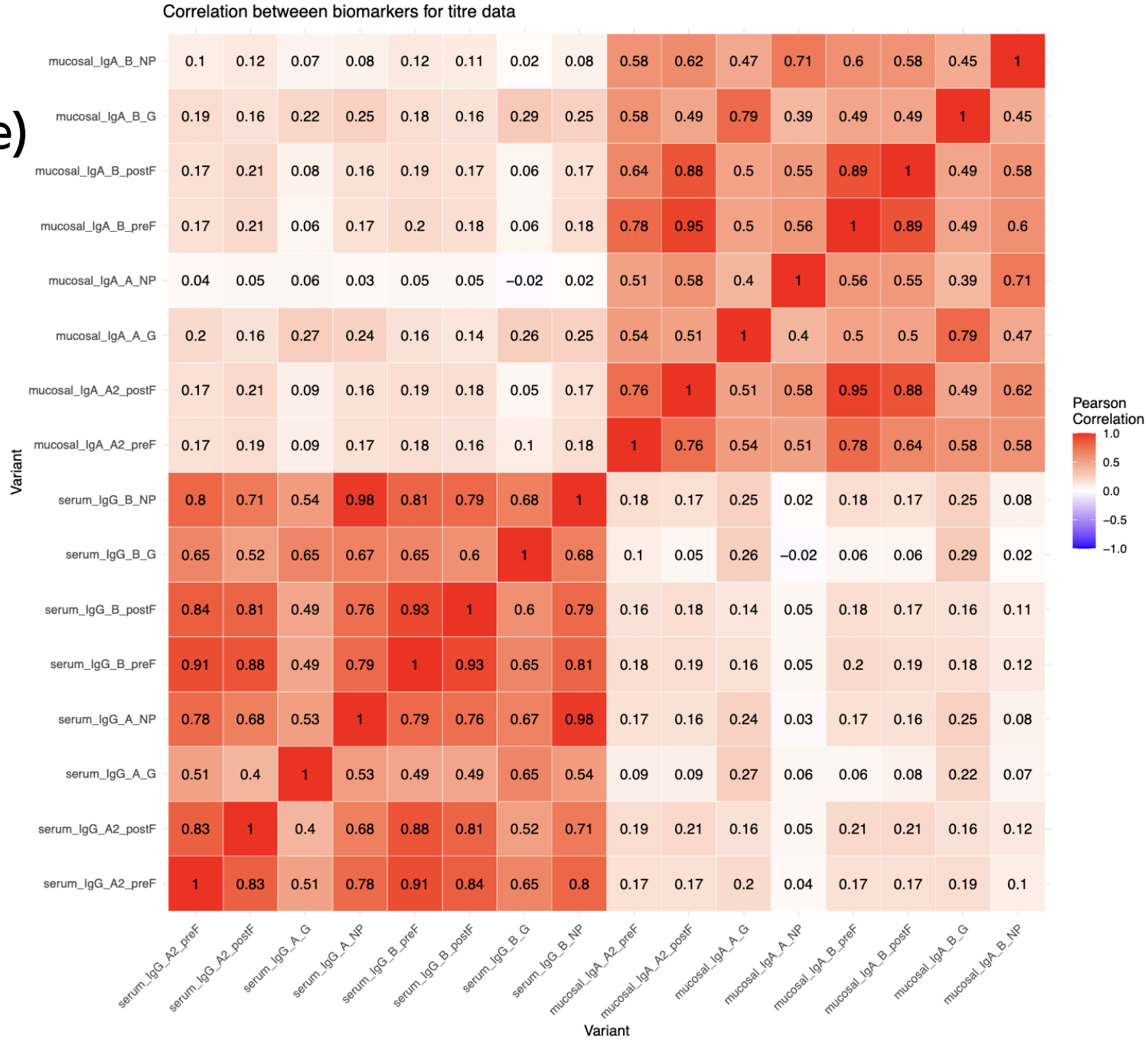
IgG levels in serum/ plasma

IgA levels in mucosal lining fluid

PreF, PostF, G, NP

- A and B serotype
- mIgA and **sIgG**

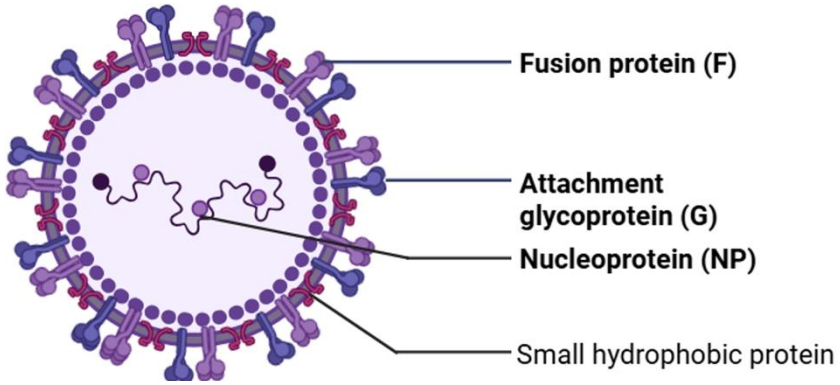
SH)

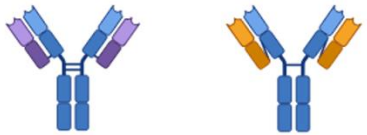


CASE 2: RSV in The Gambia

TRANSVIR Study (vaccine naïve)

N = 256 people, 308 days, 2–5 bleeds person





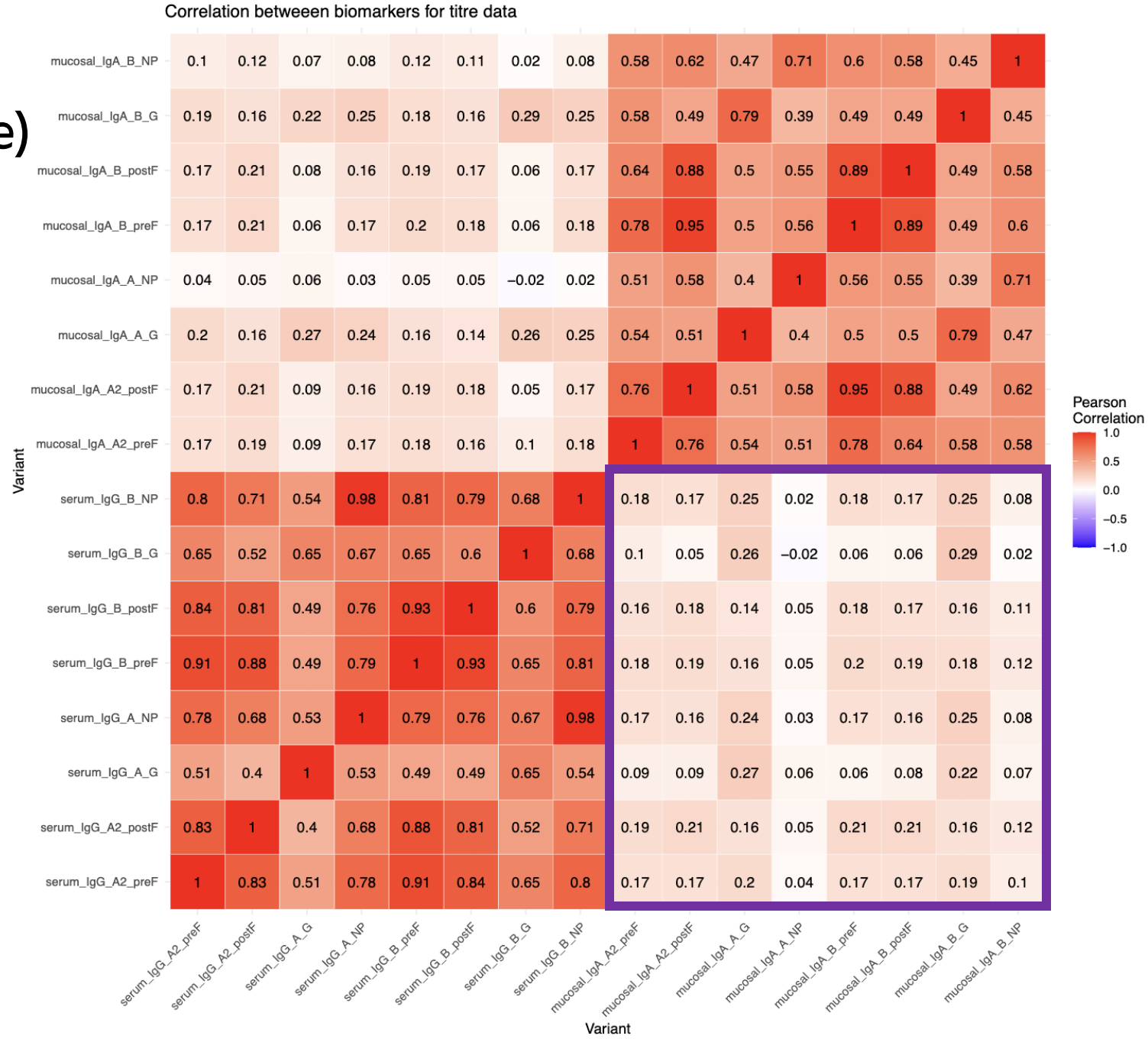
IgG levels in serum/ plasma

IgA levels in mucosal lining fluid

PreF, PostF, G, NP

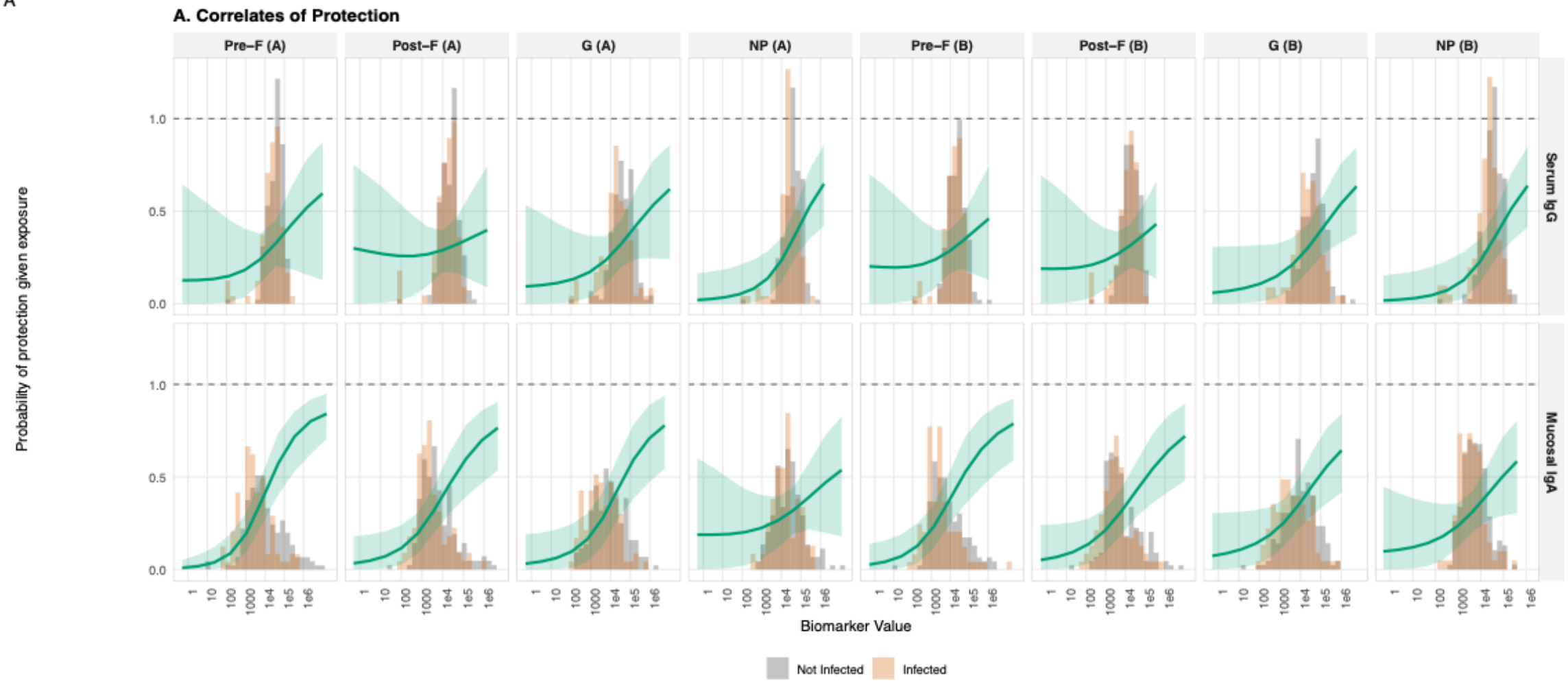
- A and B serotype
- mIgA and **sIgG**

SH)



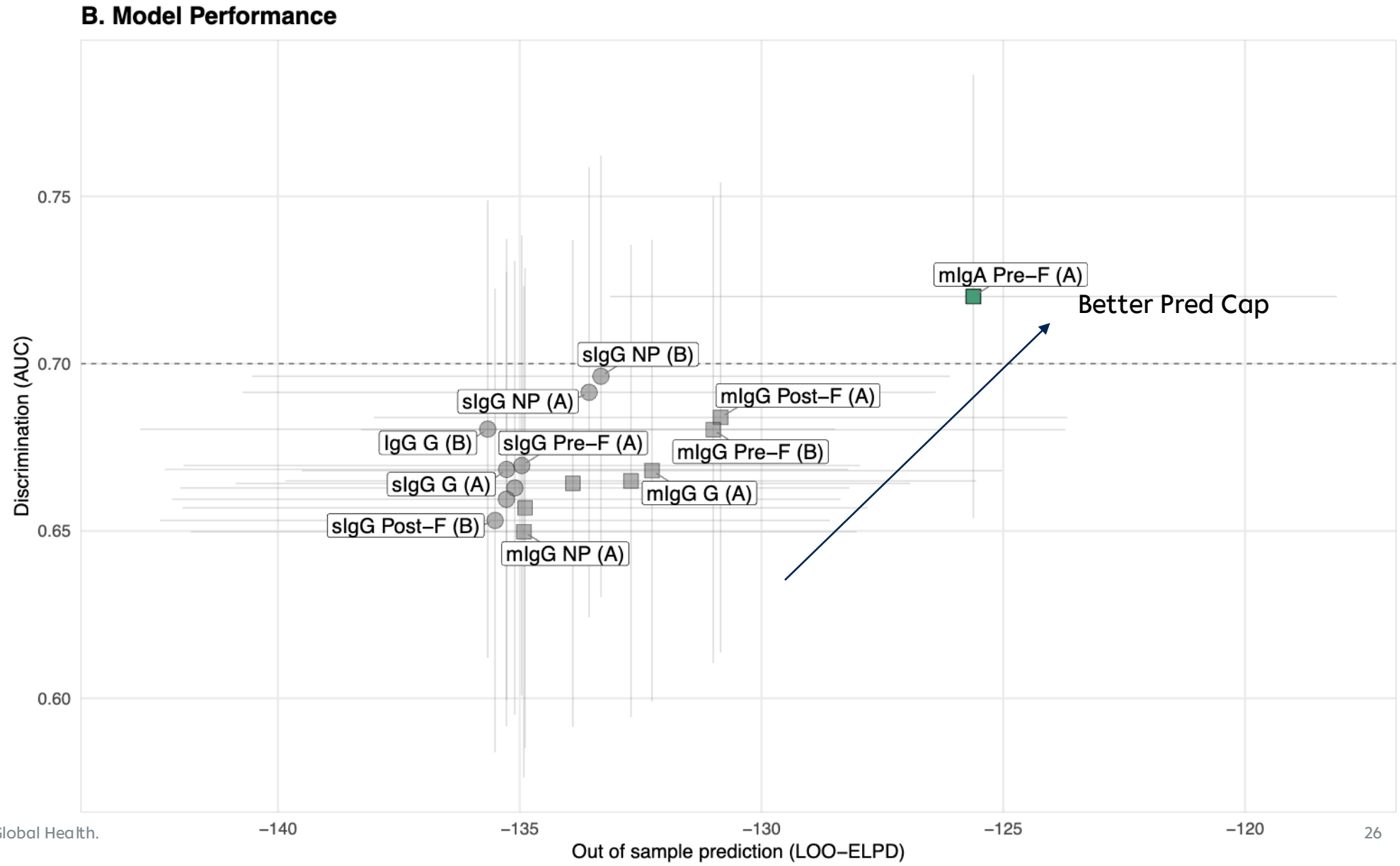
FITTED COP FOR RSV

A

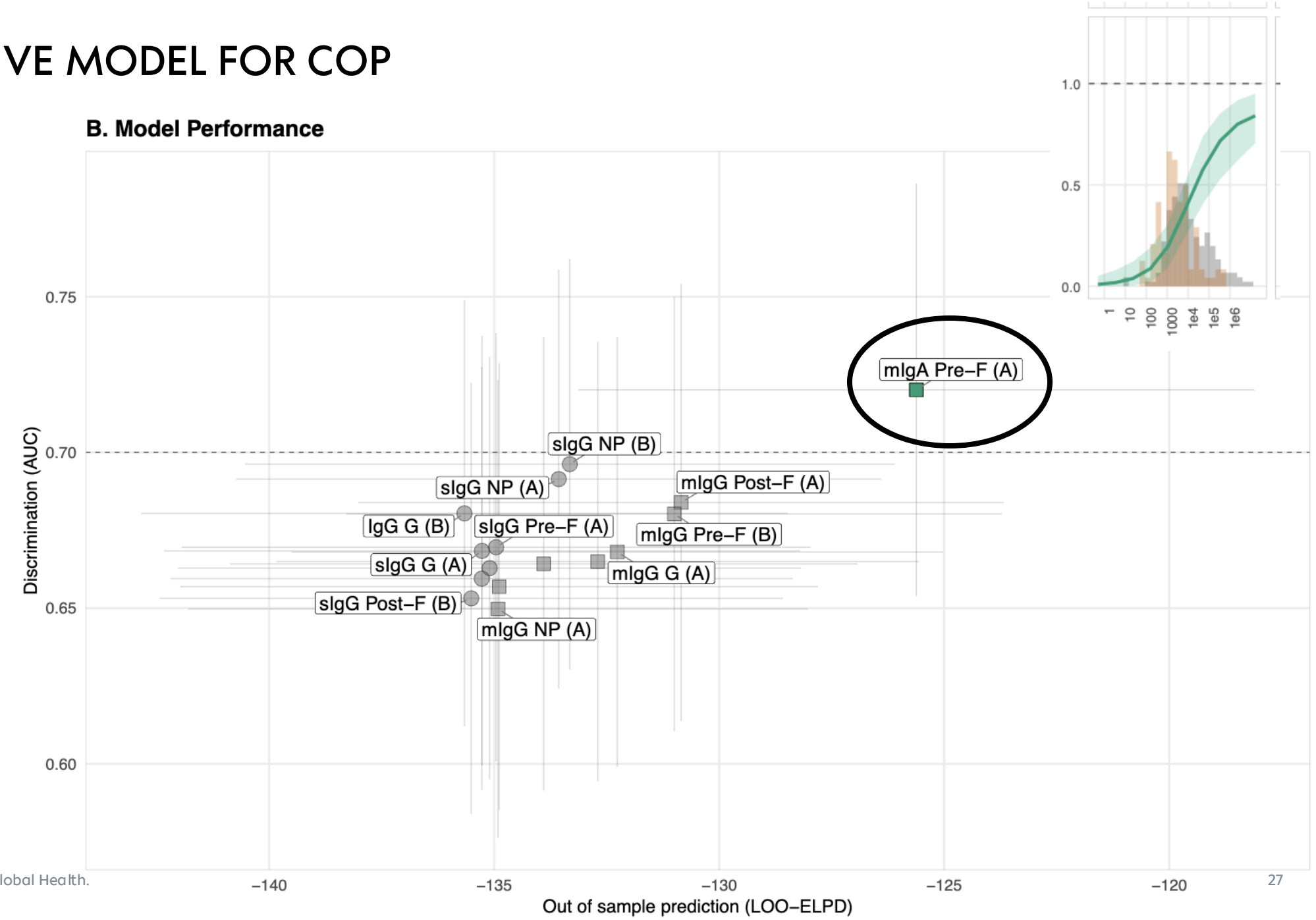


B

BEST PREDICTIVE MODEL FOR COP

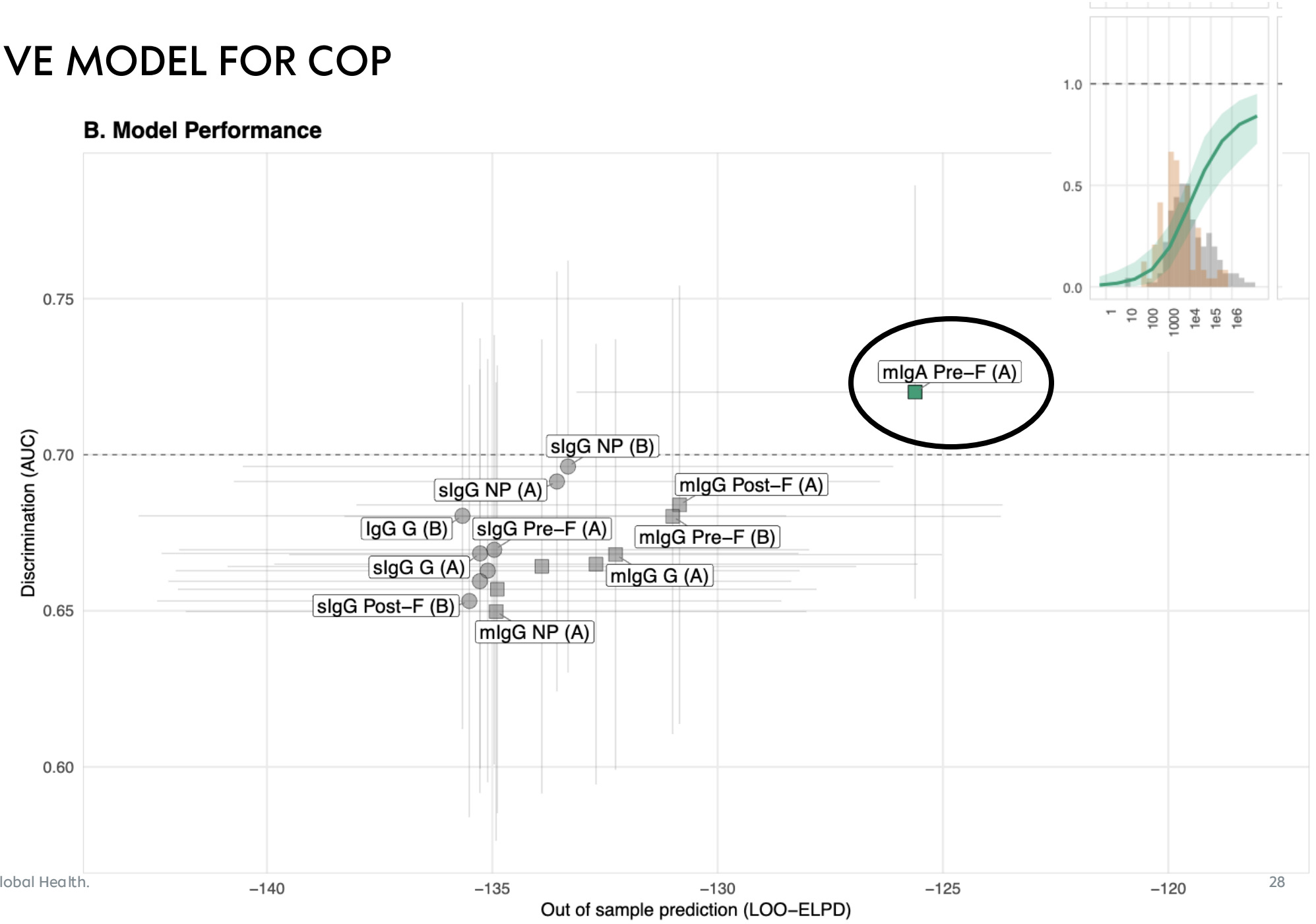


BEST PREDICTIVE MODEL FOR COP



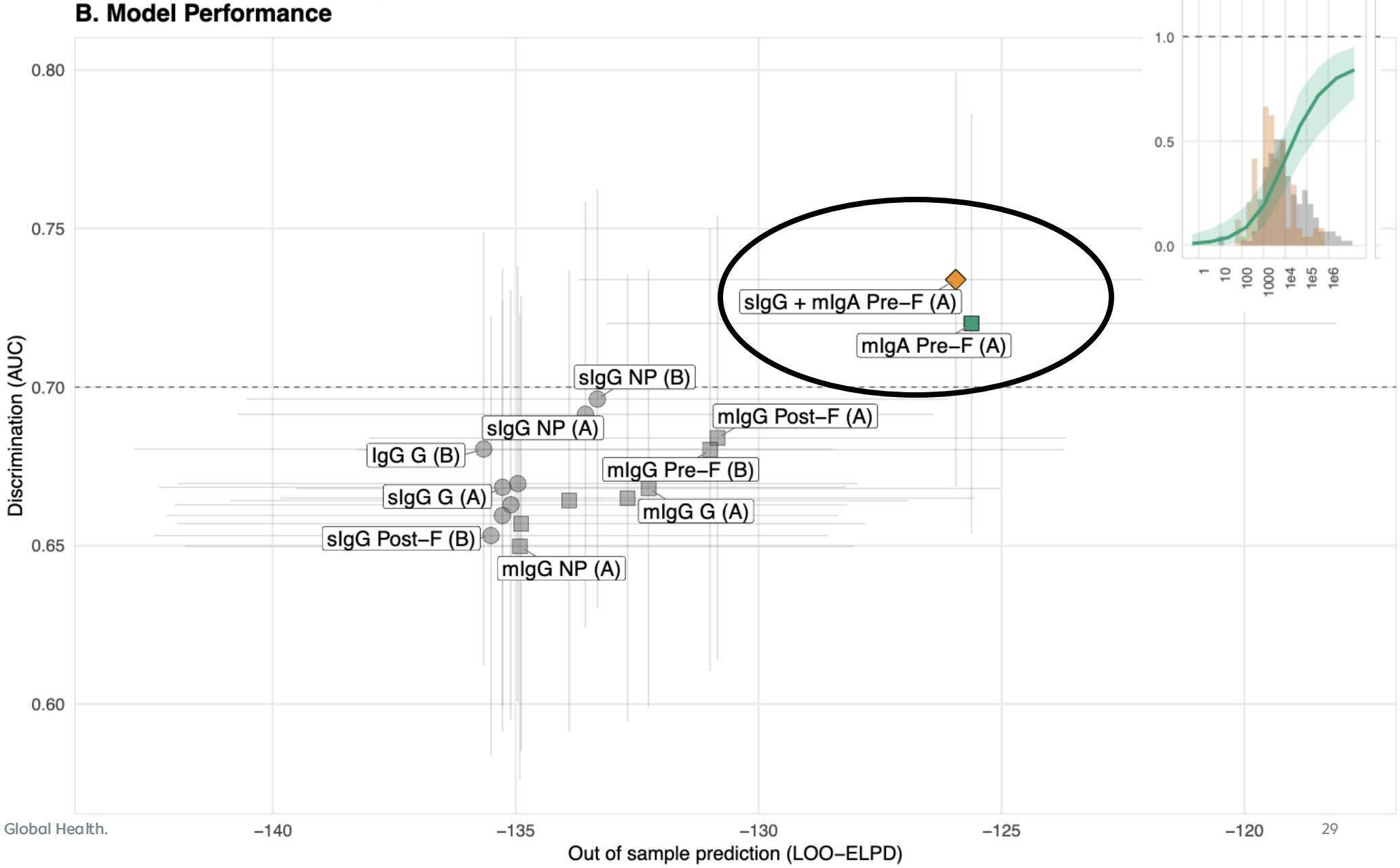
BEST PREDICTIVE MODEL FOR COP

What happens when we add sIgG to mIgA?



BEST PREDICTIVE MODEL FOR COP

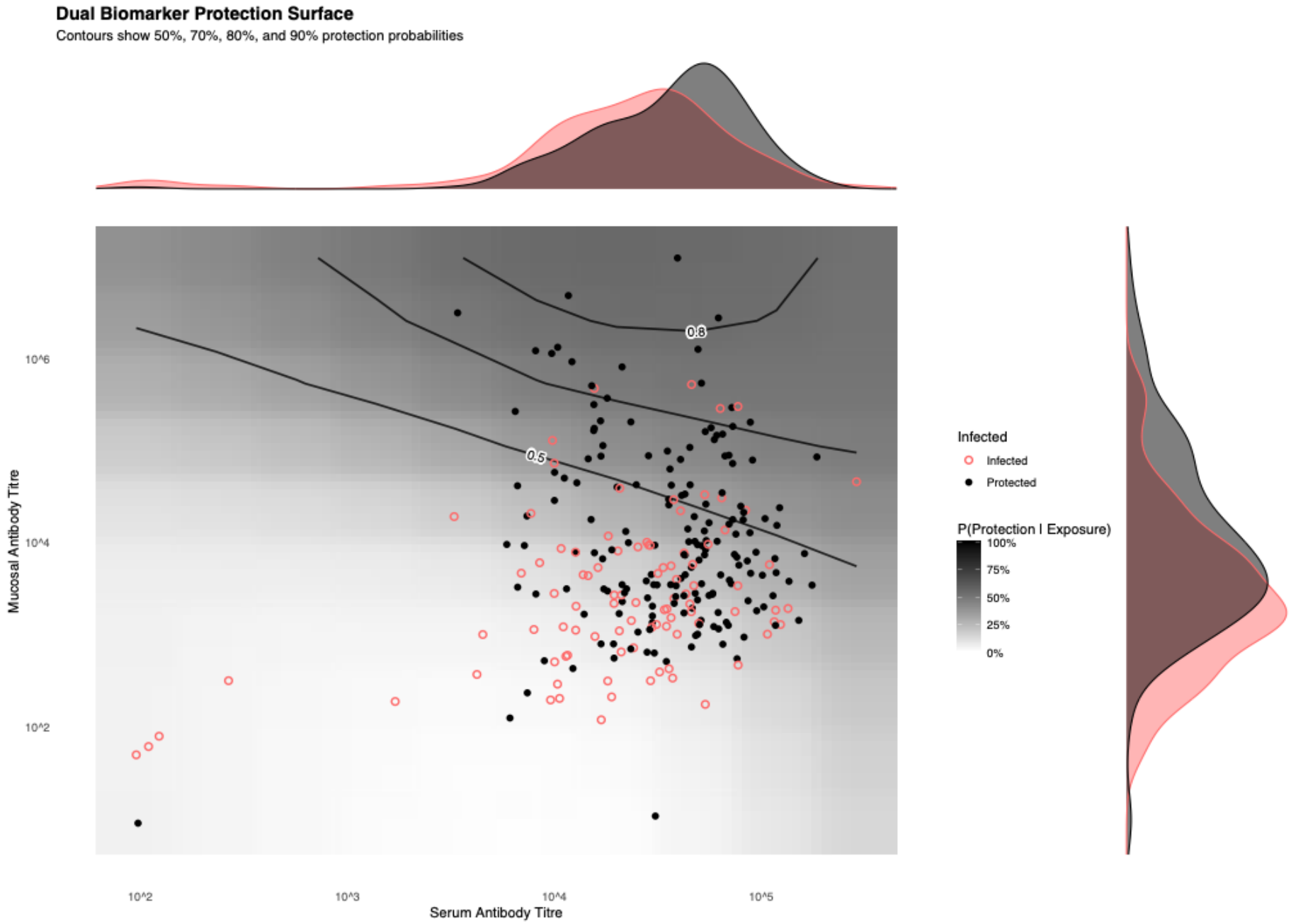
What does surface look like?



Cowling(?) CoP surface for RSV Pre-F

2D correlate of Protrection surface:

Questionable practical use?



DISCUSSION

IMPACT

- Developing robust statistical methods for establishing CoP from natural history studies important;
  - Not enough time/money to run a clinical trial in humans to determine a CoP causally (MoP)
  - A lot of pathogens have no vaccine -> can used as preliminary work to determine candidate CoP in clinical trials in humans/animals
  - Potential for better CoP using multiple biomarkers
    - Better discrimination + better counterfactual impact

EXTENSIONS

- Add hierarchical effects to logistic function to see how CoP varying across covariates (infection history and/or age)
- Similar stuff using ML; good at discovering unexpected patterns in complex data; blackbox-y so not good for regulatory-acceptable evidence

LIMITATIONS

- Setting and seasonal specific, unsure how well this generalises

CONCLUSIONS

1. We have developed a framework for for CoP; broad application
- Will be implemented as an *R package* and an online widget

Coming soon!

2. We identify the "best" single biomarker form lots of biomarkers
- ***SARS-CoV-2*:** Best single biomarker is serum pTNA to Delta for Delta wave, and Omicron BA.1 pTNA to Omicron wave
- ***RSV*:** Best single biomarker is mIgA PreF to infection

3. Assessed value of combined biomarker models
- ***SARS-CoV-2*:** Adding mIgA binding assay information has worse predictive power
- ***RSV:*** *Combing with sIgG to PreF has similar predictive capacity, but better AMG (ensuring both biomarker have a four-fold rise)*

ACKNOWLEDGEMENTS

Dr. James Hay
Dr. Sheikh Jarju
Dr.Dawda Jobe
Dr. Rhys Wenlock
Dr. Thushan I de Silva
Prof Adam J Kucharski

MRC Unit The Gambia

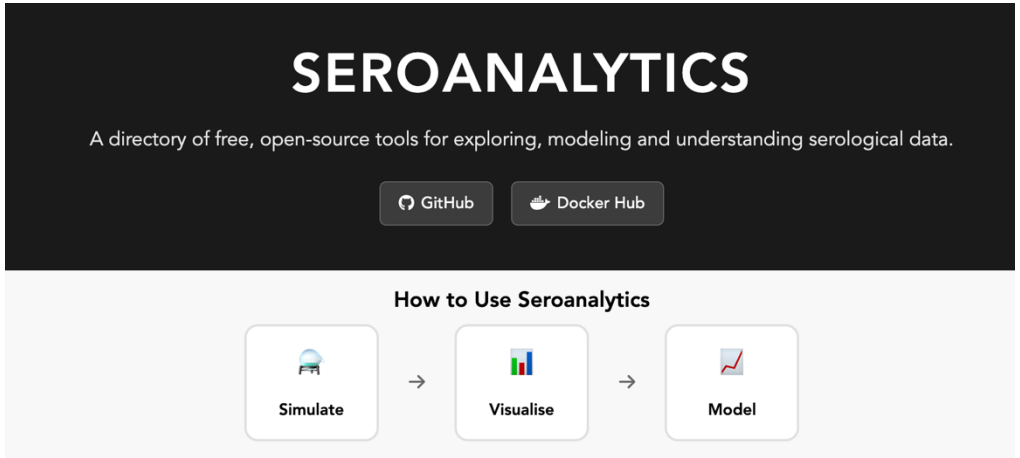
London School of Hygiene & Tropical Medicine

cmmid

University of Sheffield

Pandemic Sciences Institute

NIHR | National Institute for Health and Care Research



-> seroanalytics.org

FOLLOW ME!
david.hodgson@charite.de
LinkedIn: https://www.linkedin.com/in/dchodgson/
Bluesky: dchodge.bsky.social

CHARITÉ | Center for Global Health.

EXTRA SLIDES

MOTIVATION

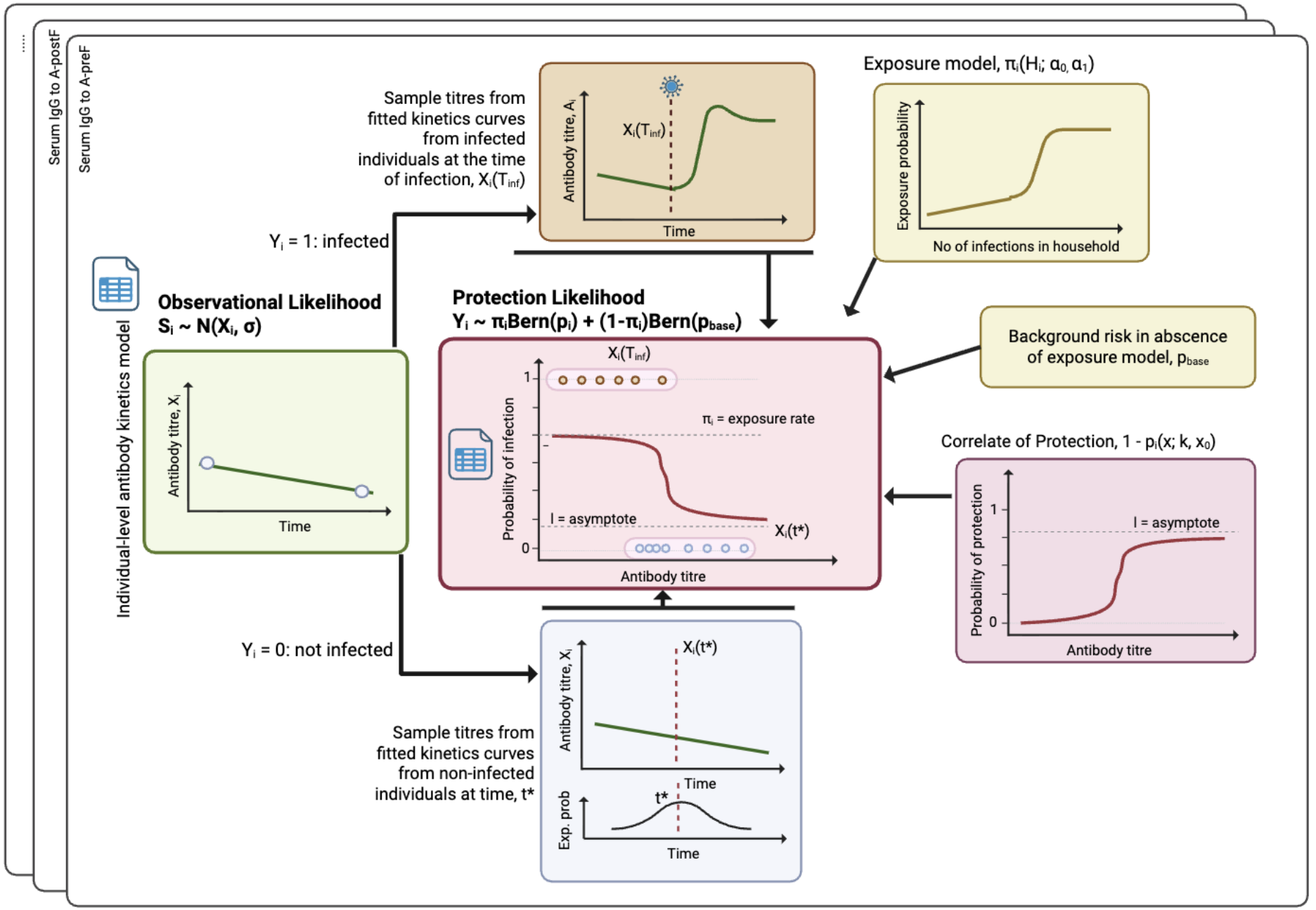
The Problem:

- Without CoPs => vaccine trials require large sample sizes, long follow-up periods, and are resource-intensive
- Current CoPs (assessed in vaccine trials) rely on single biomarkers (typically serum antibodies), which may miss important aspects of protective immunity.

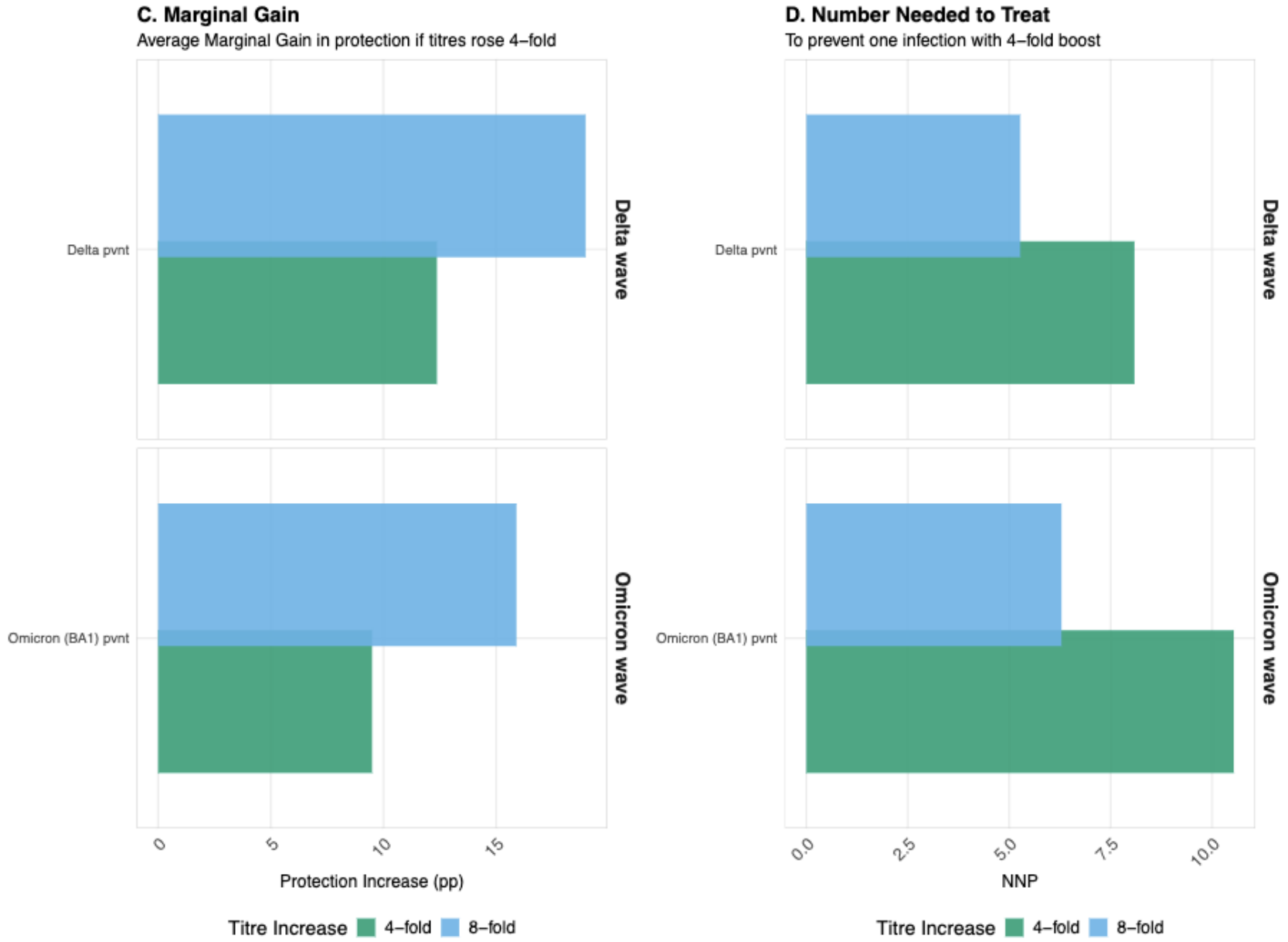
The Gap:

- Mucosal immunity is the frontline defence for respiratory pathogens, yet (historically) rarely measured in CoP studies
- Rigorous statistical framework needed to compare multiple biomarkers and identify the "best" CoP in a natural history setting
- Limited data on whether combining biomarkers improves prediction of protection

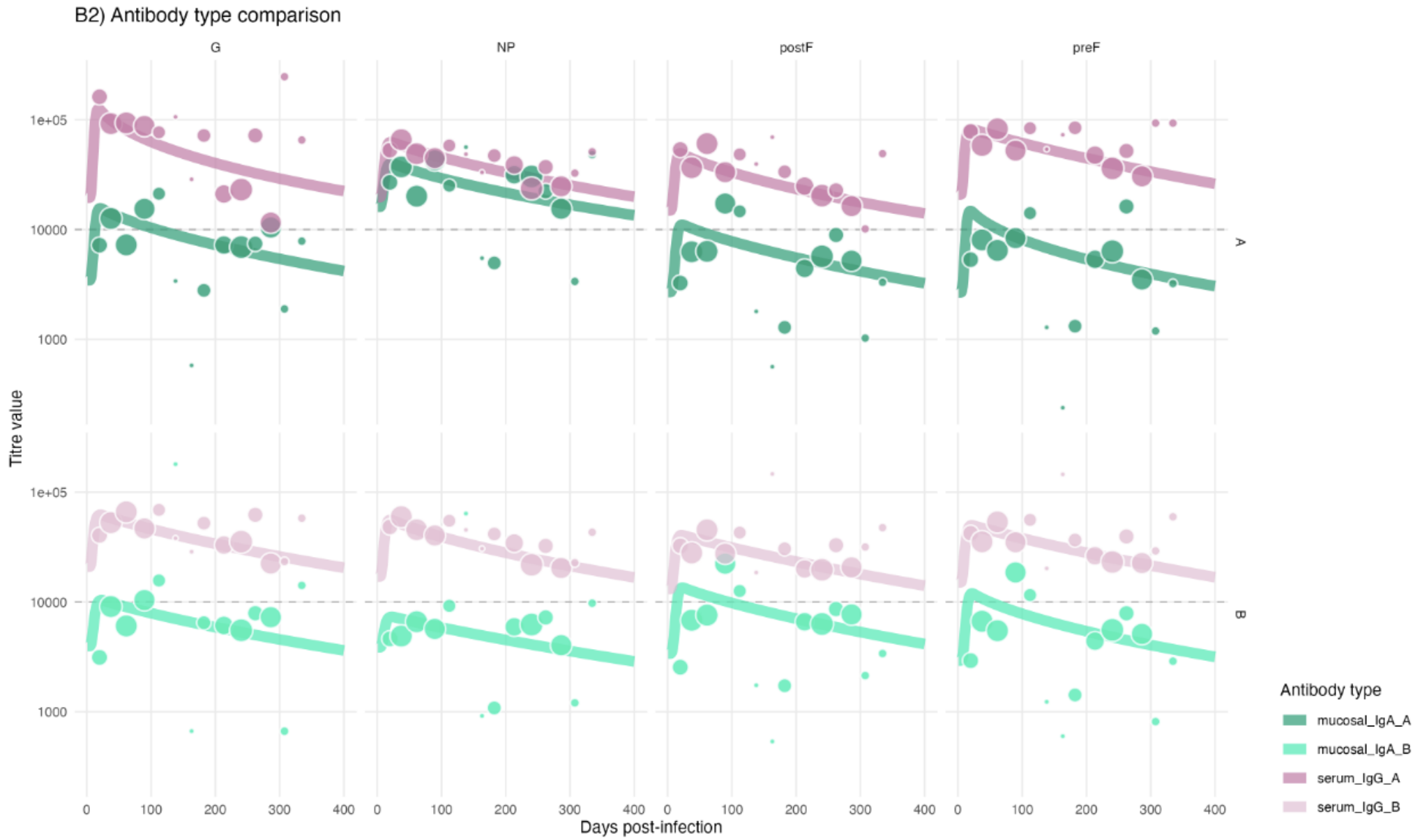
For each of G suptype, A and B, we have serum IgG and muscoal IgA to Pre-F, post-F, G, and NP BA2



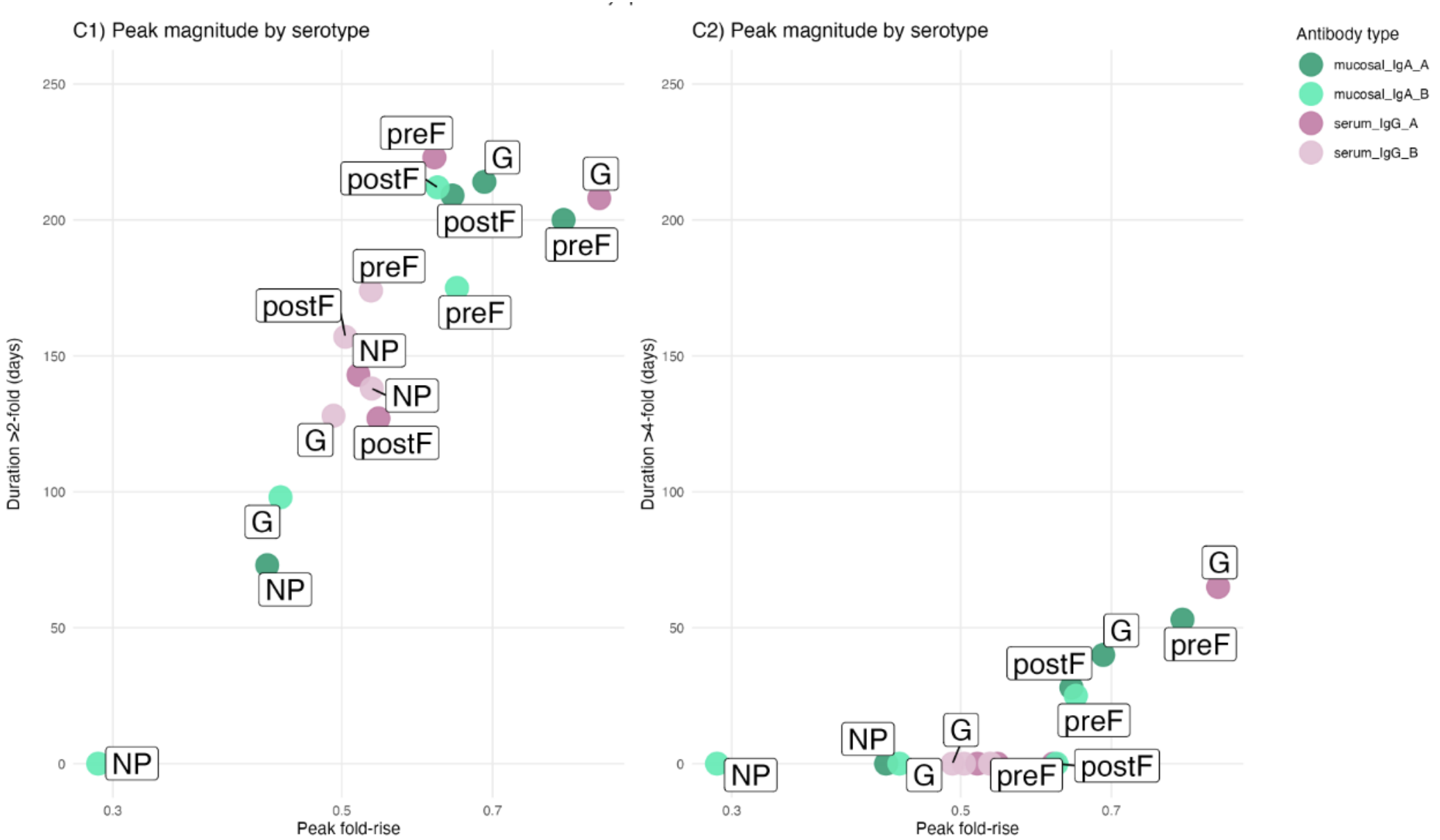
RESULTS FOR SARS-CoV-2



RSV kinetics

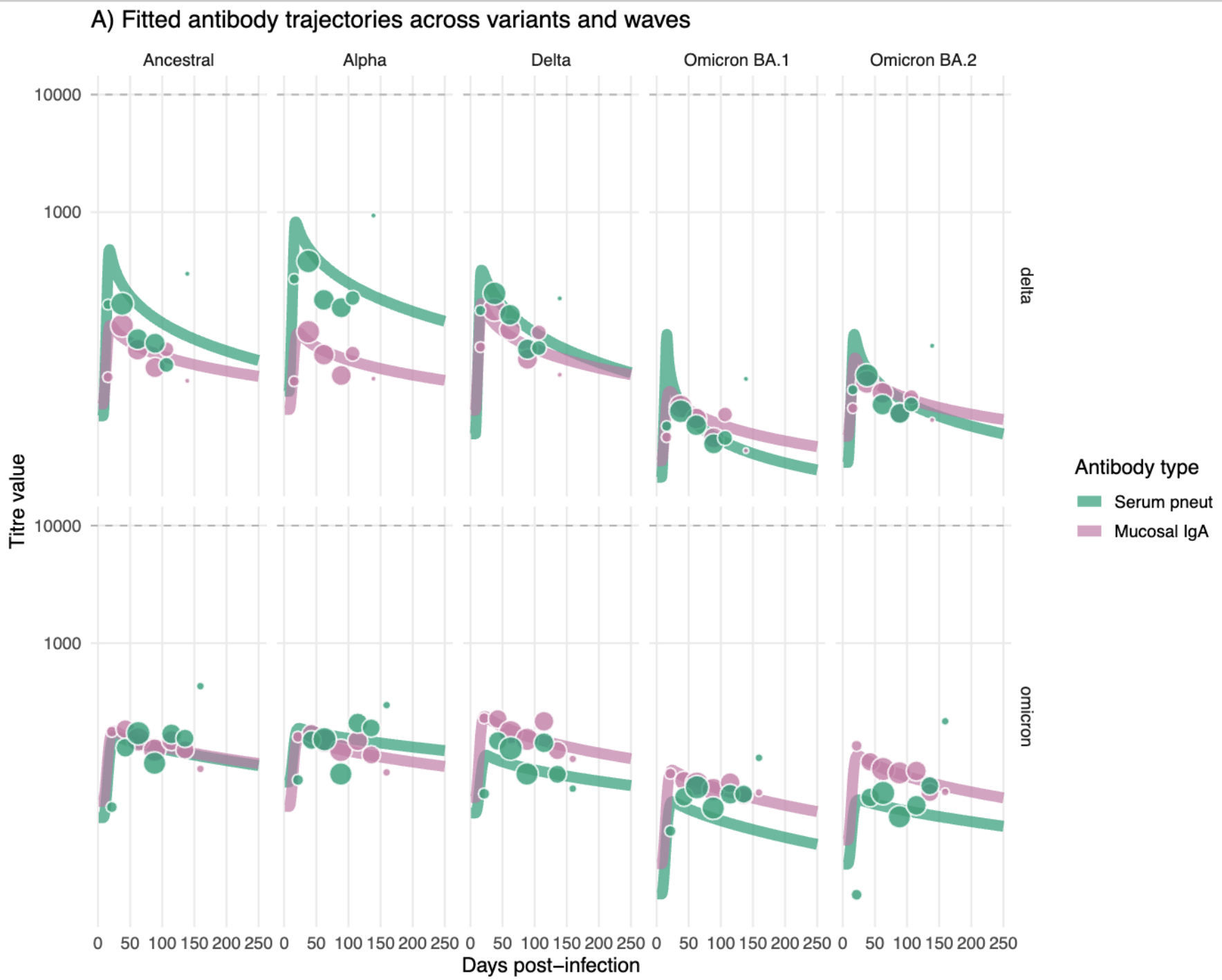


RSV kinetics



CHARITÉ

SARS-CoV-2 Kinetics



SARS-CoV-2 Kinetics