

Leveraging immunogenicity data and logistic regression for detection of covariate effects on vaccine efficacy

Julie Dudasova^{1,2}; Zdenek Valenta³; Garrett T. Nieddu⁴; Alexander D. Becker⁴; Jeffrey R. Sachs⁴

¹Quantitative Pharmacology and Pharmacometrics, MSD Czech Republic, Prague, Czech Republic; ²First Faculty of Medicine, Charles University, Prague, Czech Republic; ³Institute of Computer Science of the Czech Academy of Sciences, Prague, Czech Republic; ⁴Quantitative Pharmacology and Pharmacometrics, Merck & Co., Inc., Rahway, NJ, USA

Background

- Vaccine efficacy (VE) is defined as a proportional reduction in risk of disease for vaccinated subjects compared to control subjects and is often assessed by counting disease cases and non-cases in randomized controlled clinical trials¹
- VE can be affected by covariates (demographic and clinical characteristics of enrolled subjects), eg, age, gender, or prevaccination serostatus
- Statistical significance of covariate effects on the binary clinical outcome (diseased versus nondiseased) is typically evaluated by multiple logistic regression
- In most efficacy trials, immune response postvaccination (immunogenicity) is measured in addition to the primary clinical endpoint. An immunogenicity biomarker that reliably predicts protection is a correlate of protection (CoP)²
- It has been shown that CoP-based VE prediction is more precise than the case-count-based VE estimate.³ Several approaches have been proposed to model the relationship between immunogenicity and probability of disease (PoD)³⁻⁵

Motivation

Determine if vaccine efficacy is impacted by any characteristics of enrolled subjects

Use an immune response biomarker data to determine statistical and clinical significance of demographic and clinical characteristics

Use immune response biomarker data to make predictions of vaccine efficacy in demographically and clinically defined subgroups

Objectives

Evaluate accuracy and precision of immunogenicity-based efficacy predictions using logistic function as a PoD curve

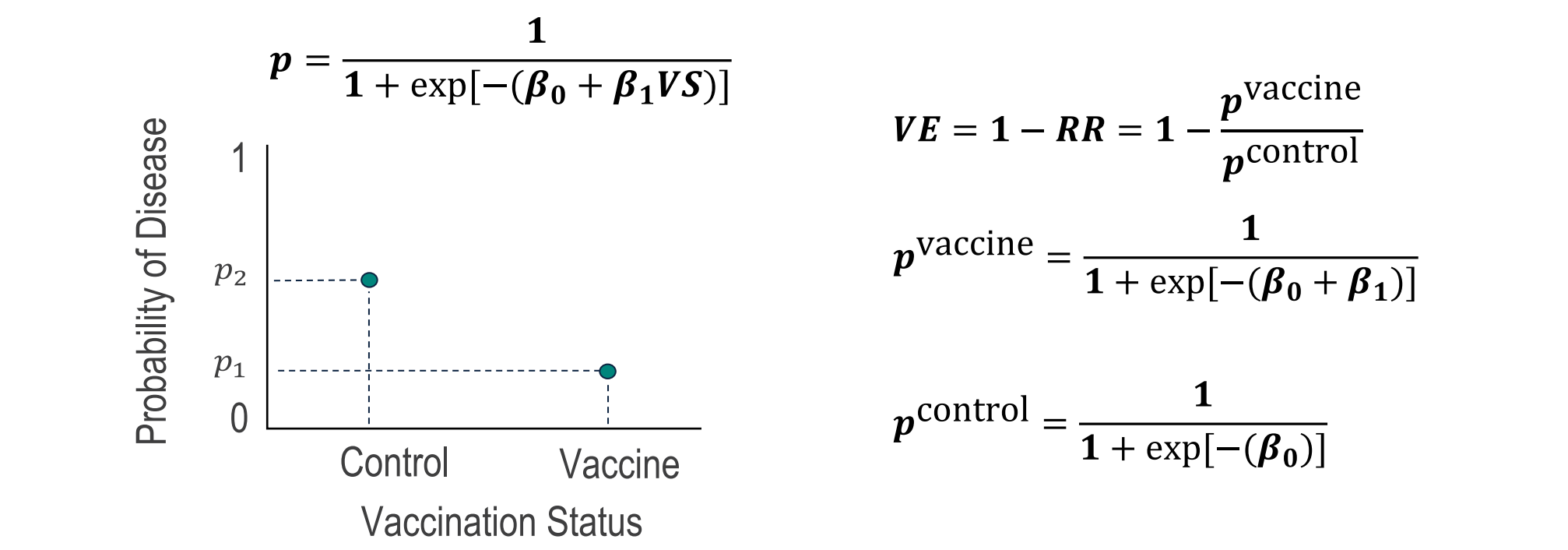
Evaluate whether inclusion of immunogenicity data in multiple logistic regression increases efficiency in finding impactful covariates

Methods and tests with simulations

Probability of disease as a function of vaccination status

Figure 1. Typical approach

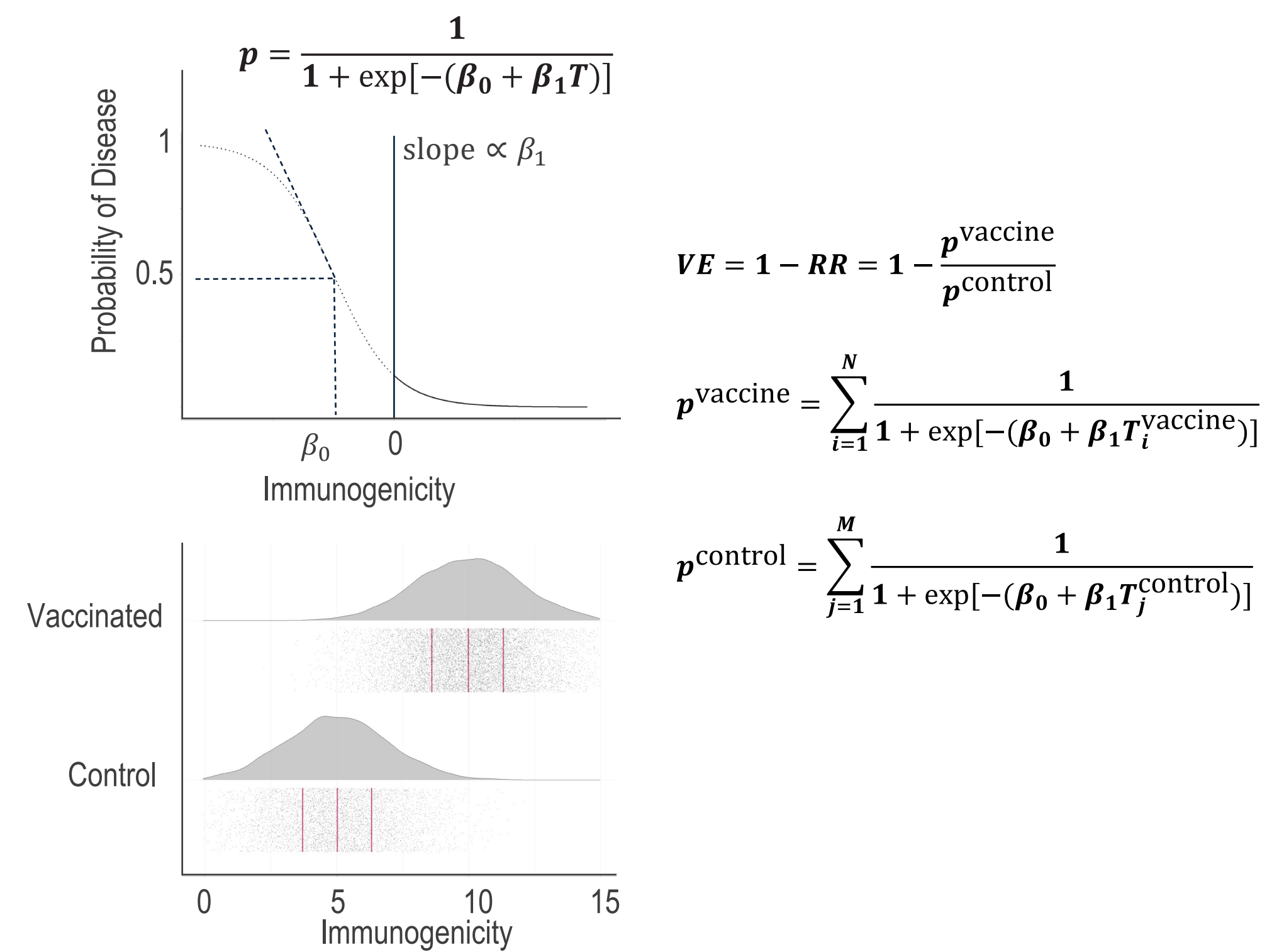
Probability of disease and vaccine efficacy (VE) as a function of vaccination status, VS.



Probability of disease as a function of immunogenicity

Figure 2. Immunogenicity-based approach

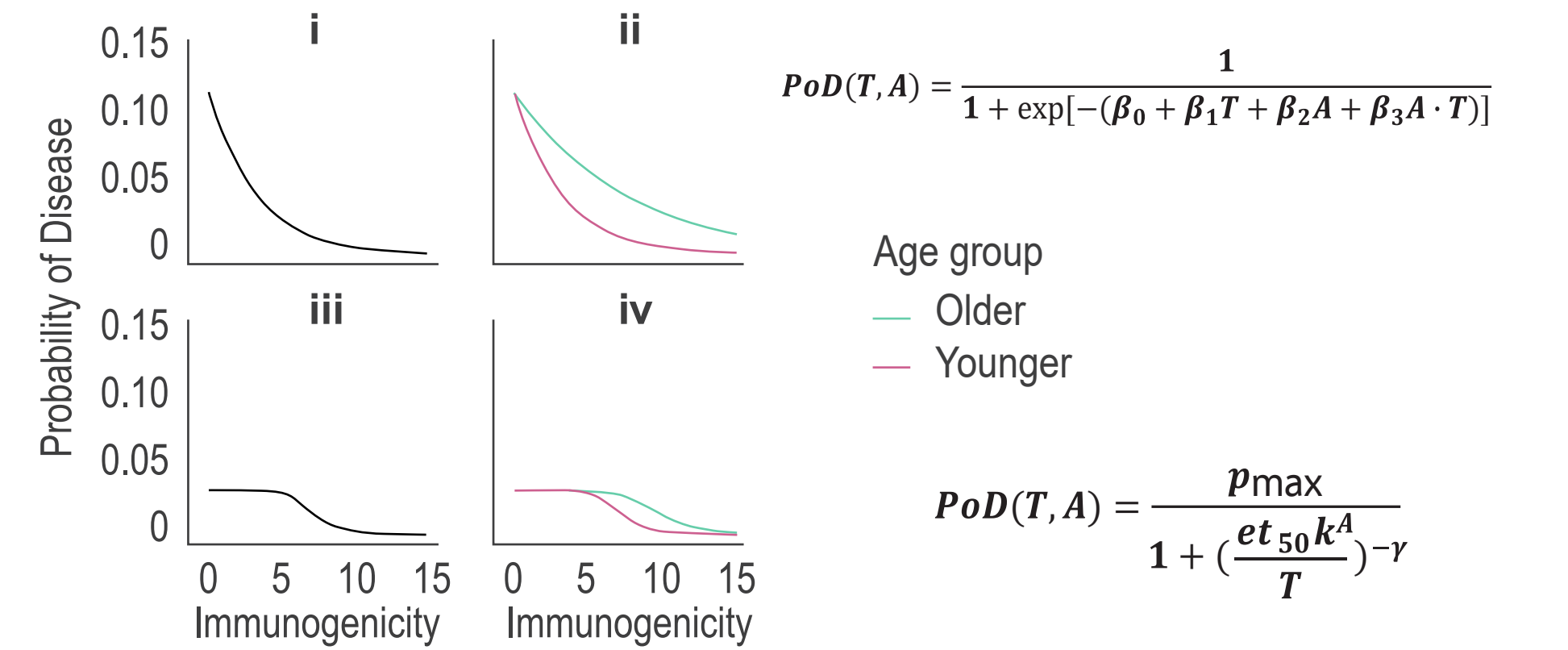
Probability of disease and vaccine efficacy (VE) as a function of immunogenicity, T.



Simulation study

Figure 3. Definition of scenarios i, ii, iii, and iv by shapes of the true PoD curve

i: Logistic function, no age effect on PoD curve and VE; ii: Logistic function, age effect causes 30% difference in VE between age groups; iii: Hill function, no age effect on PoD curve and VE; iv: Hill function, age effect causes 30% difference in VE between age groups. T represents immunogenicity; A represents age group.



Simulation study

Table 1. Parameter values used to set the truth for simulated data

All scenarios: 25% younger and 75% older subjects in the trial population, 2:1 ratio vaccinated to control.

		Scenarios							
		i		ii		iii		iv	
		a	b	a	b	a	b	a	b
Parameters	Total number of subjects	15000	7500	15000	7500	15000	7500	15000	7500
	True PoD model	Logistic		Logistic		Hill		Hill	
	Age effect	No		Yes		No		Yes	
	True VE	Overall		80%		80%		72%	
		Younger		80%		80%		80%	
		Older		80%		80%		50%	
	Vaccinated	mean(x^vaccine)		10		10		10	
		sd(x^vaccine)		2		2		2	
		mean(x^control)		5		5		5	
	Control	sd(x^control)		2		2		2	
		p_max		—		0.033		0.033	
		e^t_50		—		7.12		7.12	
	gamma	—		—		7		7	
		k		—		1		1.361	
		beta_0		-2		—		—	
	beta_1	-0.33		-0.33		—		—	
		beta_2		0		—		—	
		beta_3		0		0.184		—	

Results

Accuracy and precision of VE estimation

Figure 4. Immunogenicity-based VE estimation using logistic regression is accurate when true PoD is logistic

Age group-specific distributions of VE point estimates for each simulated scenario (5,000 simulated trials).

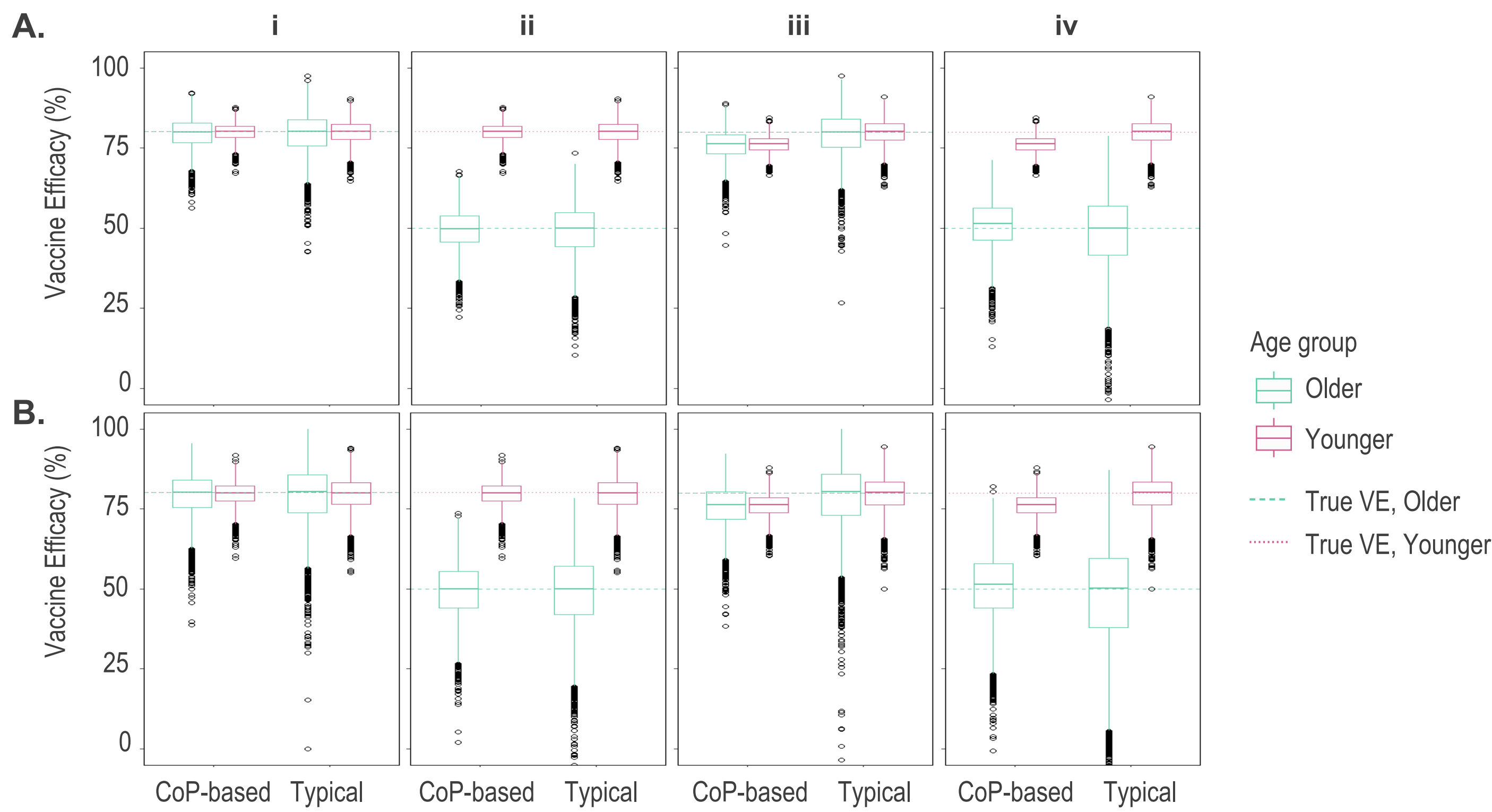
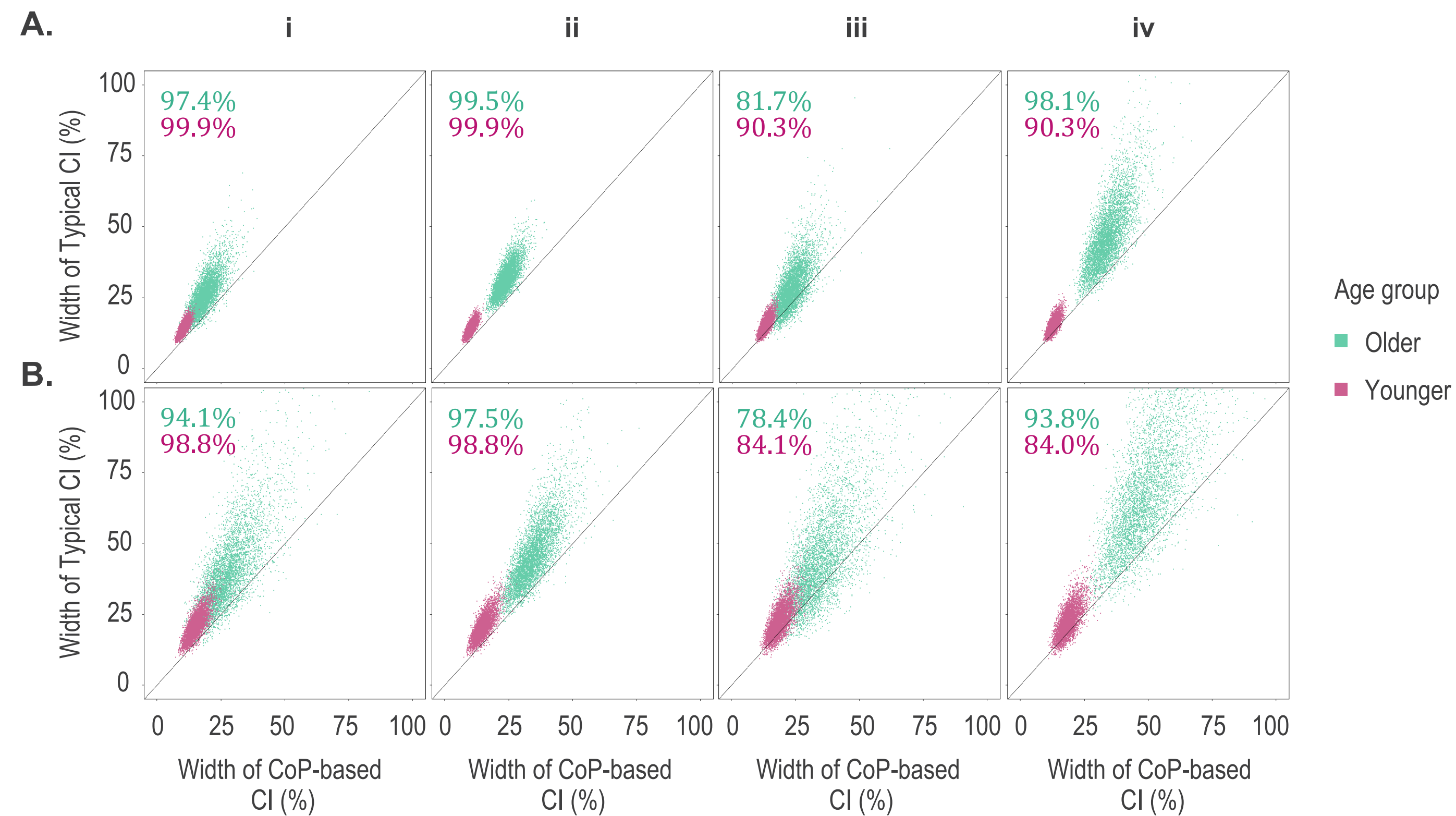


Figure 5. Immunogenicity-based VE estimation is more precise than case-counting (narrower confidence intervals)

Widths of age group-specific VE confidence intervals for each simulated scenario (5,000 simulated trials).



Assessment of age group effect

Table 2. Predictive values of immunogenicity-based (CoP-based) and vaccination status-based (typical) logistic regression models for each simulated scenario (5,000 simulated trials)

A. Number of subjects = 15,000 (~200 disease cases); B. Number of subjects = 7,500 (~100 disease cases) (True positive refers to percentage of simulated trials where age group was correctly identified as a significant covariate. True negative refers to percentage of simulated trials where age group was correctly identified as a non-significant covariate).

			Scenarios					
			i	ii	iii	iv		
A	Results	True positive (%)	CoP-based		—	100	—	99.62
			Typical		—	100	—	99.30
		True negative (%)	CoP-based		90.66	—	94.08	—
			Typical		90.62	—	91.12	—
		Positive predictive value (%)	CoP-based		91.46		94.39	
			Typical		91.42		91.79	
		Negative predictive value (%)	CoP-based		100		99.60	
			Typical		100		99.24	
		ROC AUC (%)	CoP-based		100		99.82	
			Typical		100		99.65	
				<i>P</i> -value of AUC difference		1		5.70e-12
			Scenarios					
			i	ii	iii	iv		
B	Results	True positive (%)	CoP-based		—	100	—	89.08
			Typical		—	100	—	88.32
		True negative (%)	CoP-based		90.54	—	94.26	—
			Typical		91.42	—	91.42	—
		Positive predictive value (%)	CoP-based		91.36		93.95	
			Typical		92.10		91.15	
		Negative predictive value (%)	CoP-based		100		89.62	
			Typical		100		88.67	
		ROC AUC (%)	CoP-based		99.99		97.56	
			Typical		99.99		96.12	
				<i>P</i> -value of AUC difference		0.20		2.20e-16

Conclusion

- Logistic regression can be reliably used to detect the effect of a binary covariate on VE
- In the simulated trials of a typical phase 3 study (15,000 subjects, ~200 disease cases), the typical approach was 46% more likely (100%-PPV: 8.2% vs 5.6%) than the immunogenicity-based approach to falsely predict difference between the two subgroups (with a Hill curve as the true PoD)
 - Potential impact: a false determination that a factor is a covariate could result in failure to license an appropriate indication
- This false prediction could result in failure to license an appropriate indication
- Immunogenicity-based performance is also better than that of the typical approach for smaller trials (7,500 subjects, ~100 disease cases)
- The difference in this performance is smaller if the true PoD curve is logistic

References

- Halloran ME, et al. *Design and Analysis of Vaccine Studies*. Springer, 2010.
- Plotkin SA, et al. *Plotkin's Vaccines*. Elsevier, 2017.
- Dudašová J, et al. *NPJ Vaccines*. 2021;6(1):133.
- Dunning AJ. *Stat Med*. 2006;25(9):1485-1497.
- Coudeville L, et al. *Hum Vaccin*. 2010;6(10):841-848.