Issues in the Statistical Analysis of Vaccine Efficacy in COVID-19 Vaccine Efficacy Trials

Peter Gilbert

ID-PRISM Center Seminar – May 20, 2020

Acknowledgements: Fred Hutch and UW Biostatistics Faculty, Staff, Students Co-Working on COVID-19 VE Trials

- **Fred Hutch/UW HVTN+HPTN+IDCRC:** Many individuals and growing; e.g. Lindsay Carpp, Tom Fleming, Yunda Huang, Holly Janes, Michal Juraska, **Alex Luedtke**
- **NIAID Biostatistics:** Dean Follmann, Martha Nason

NIH has launched the COVID-19 Prevention Network (Corey, Mascola, Fauci, Collins, Science, May 2020)

COVID-19 Prevention Network Objective

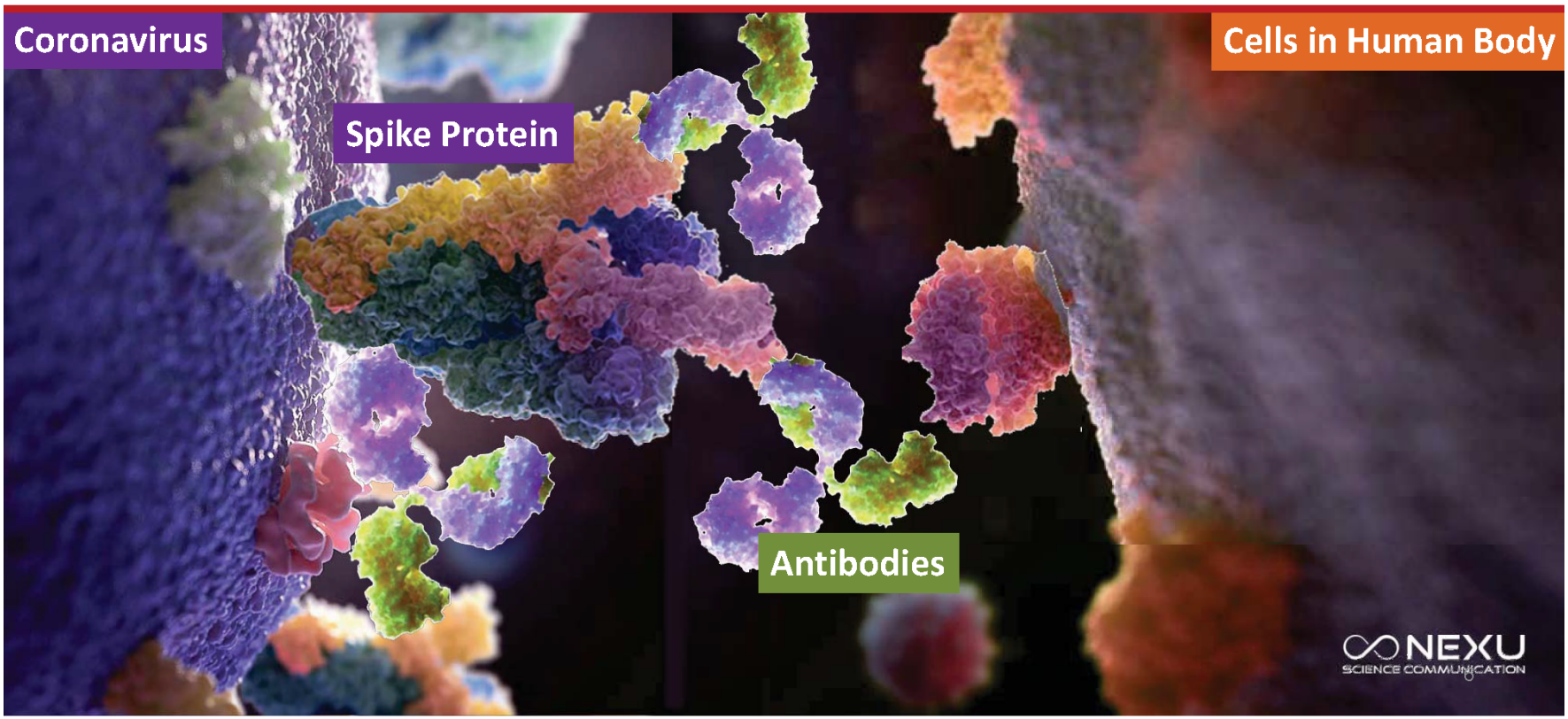
- Evaluate as rapidly as we can, with uncompromising veracity, candidate COVID-19 vaccines. 7 billion people on the planet need vaccination.
- Goal: to advance many effective vaccines
  - Global adult population 4.4 billion
  - US adult population 220 million
  - Increasingly, data suggesting childhood vaccination would be beneficial
  - Special groups may require diversity of approaches; very elderly, pregnant women, immunocompromised

Potential Beneficial Effects of a Vaccine

- Reduce COVID-19 disease, morbidity and mortality
- Reduce SARS-CoV-2 transmissibility (e.g., herd immunity, vaccine-reduction of viral shedding)
- Reduce acquisition of SARS-CoV-2 infection
- Reduce acquisition of COVID-19 disease

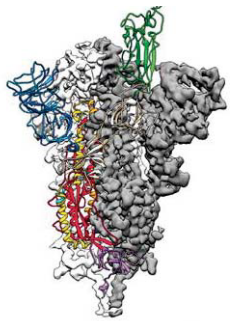
Vaccines Cause Production of Antibodies

- **Antibodies** are produced after a person gets vaccinated.
- **Antibodies** bind **coronavirus spike protein** to block infection and protect against COVID-19 disease.



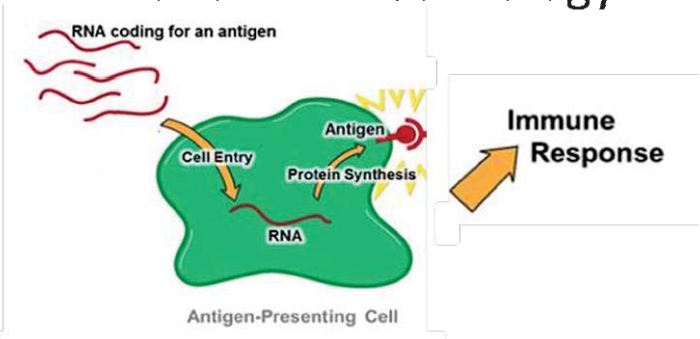
Platform Vaccine Technologies

- Protein vaccines

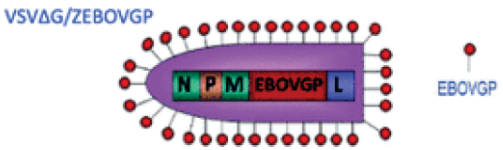


- Viral vector vaccines
  - Ad26 vector
  - ChAdOx1

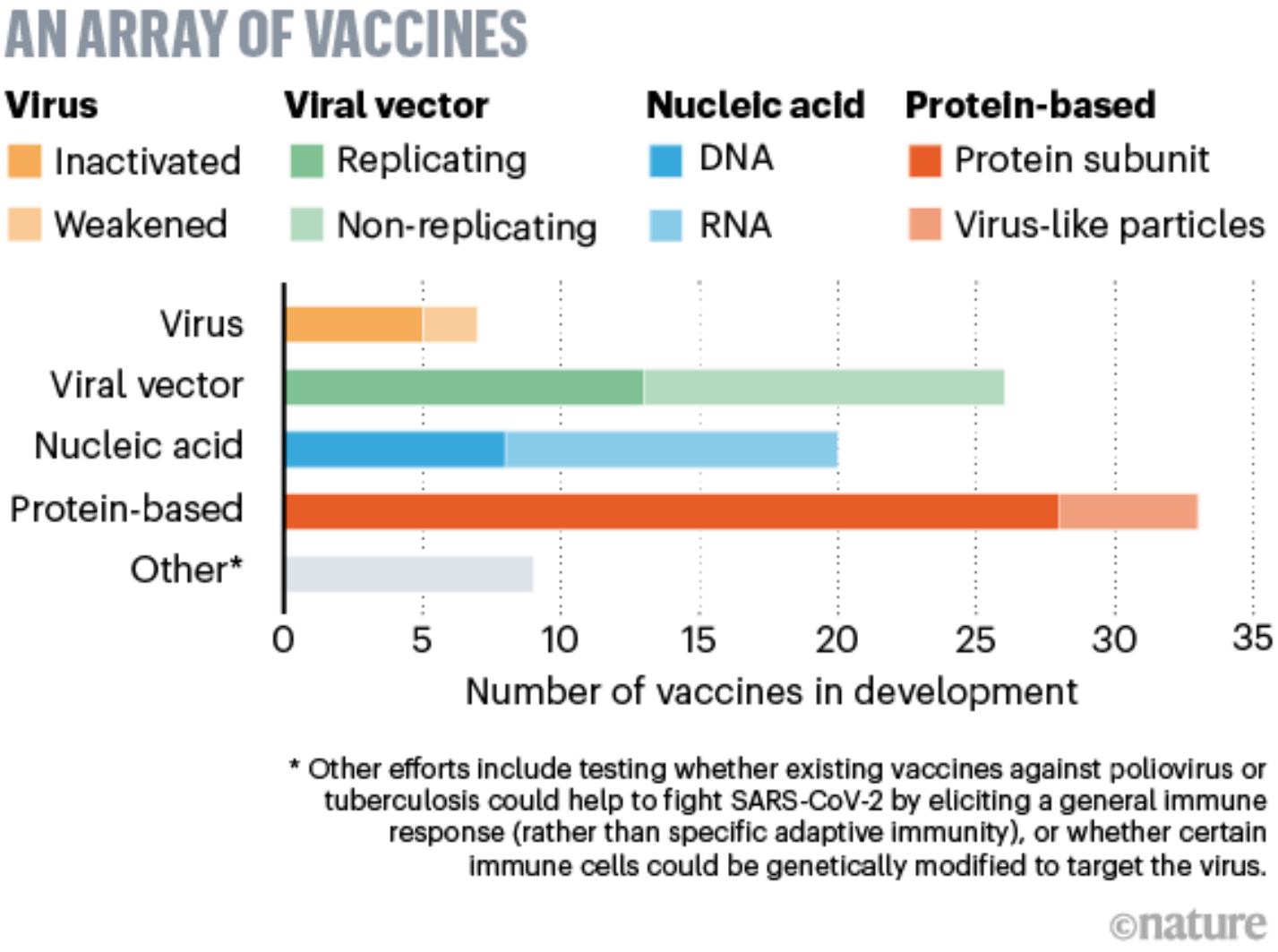
- RNA and DNA technology



  - VSV vector

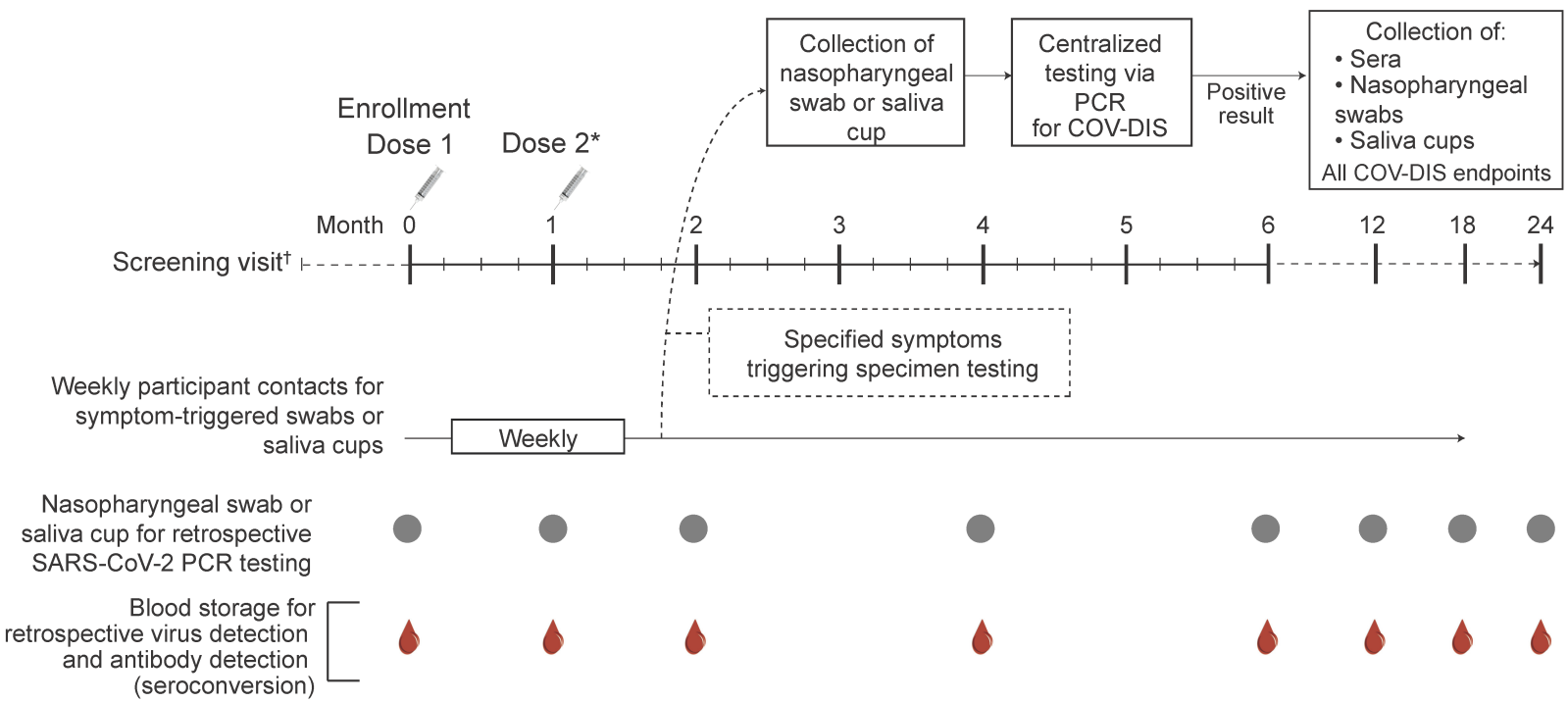


COVID-19 Vaccine Pipeline (April 28, 2020)



COVID-19 Vaccine Efficacy Trial Design

- Population: Adults at high predicted risk for SARS-CoV-2 acquisition and COVID-19 disease
- Randomize participants to vaccine vs. placebo
- Primary endpoint: Virologically-confirmed symptomatic COVID-19 disease



Vaccine Efficacy (VE)

- VE is the effect of the vaccine on biological susceptibility to acquire the primary endpoint
- Estimated by comparing a rate of the endpoint between the vaccine vs. placebo groups
- Trial is randomized and double-blinded
  - Ensures that the VE estimate measures a vaccine effect on biological susceptibility (Schaper, Fleming, Self, Rida, 1995, *Ann Review Public Health*)

Two Kinds of VE Parameters

- **PH VE** $\quad \mathrm{VE} = 1 - e^{\beta} = 1 - \frac{\lambda_1(t)}{\lambda_0(t)}$
- **Cumulative VE** $\quad \mathrm{VE(t)} = 1 - \frac{F_1(t)}{F_0(t)}$
  $t =$ a fixed time after first vaccination (e.g., 6 months)

Notation

- $A =$ randomized treatment assignment ($A = 1$ Vaccine; $A = 0$, Placebo)
- $F_a(t) = P(T \leq t | A = a)$ for $a = 1, 0$
- $T =$ time from enrollment to endpoint
- $X =$ participant covariates at baseline
- $N_a =$ number randomized to $A = a$; total sample size $N = N_1 + N_0$

Primary Endpoint and Success Criteria Based on COVID-19 FDA Guidance Document (May 12, 2020)* and WHO Target Product Profile

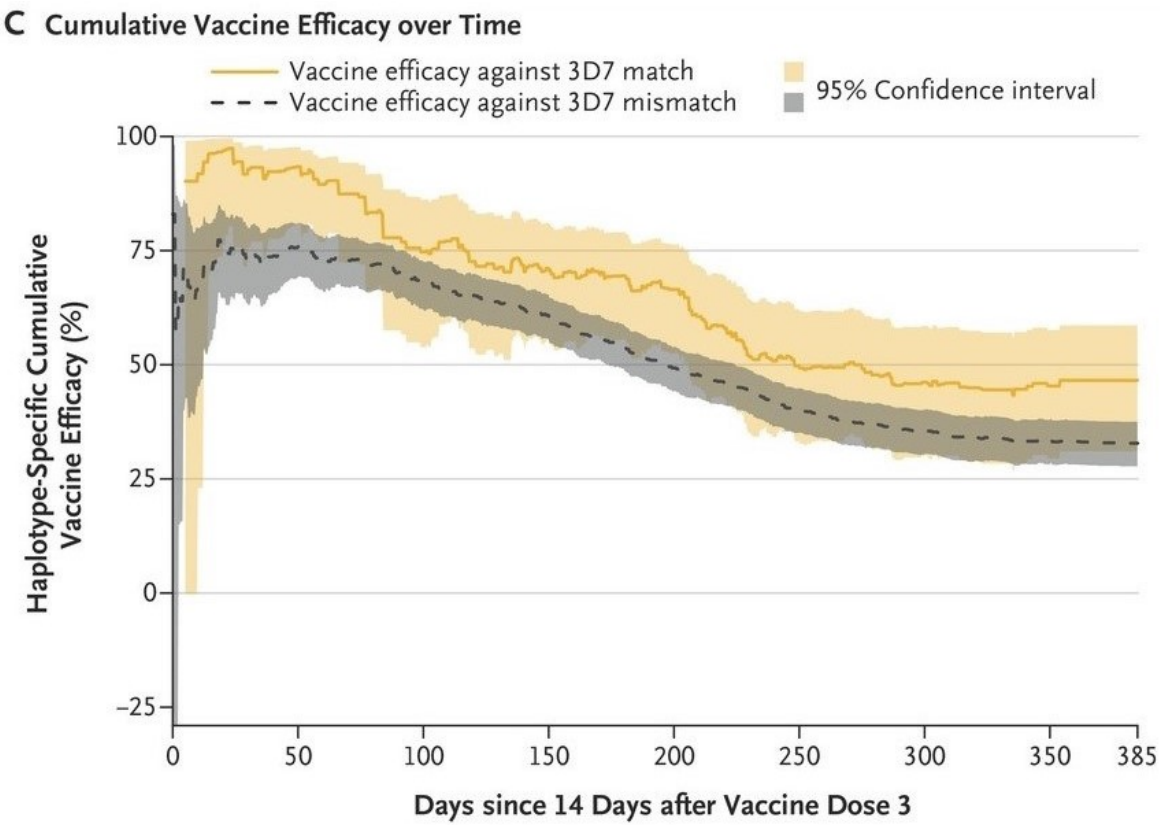
"*In prevention trials, the primary endpoint should be the occurrence of laboratory-confirmed SARS-CoV-2 infection (with or without symptoms) or SARS-CoV-2 infection with symptoms (i.e., COVID-19) through a prespecified time point."

WHO Target Product Profile (April 9, 2020) Minimal Success Criteria

- Estimated VE $\geq$ 50% for at least 6 months with sufficient precision (e.g., lower 95% confidence bound for VE $>$ 30%)

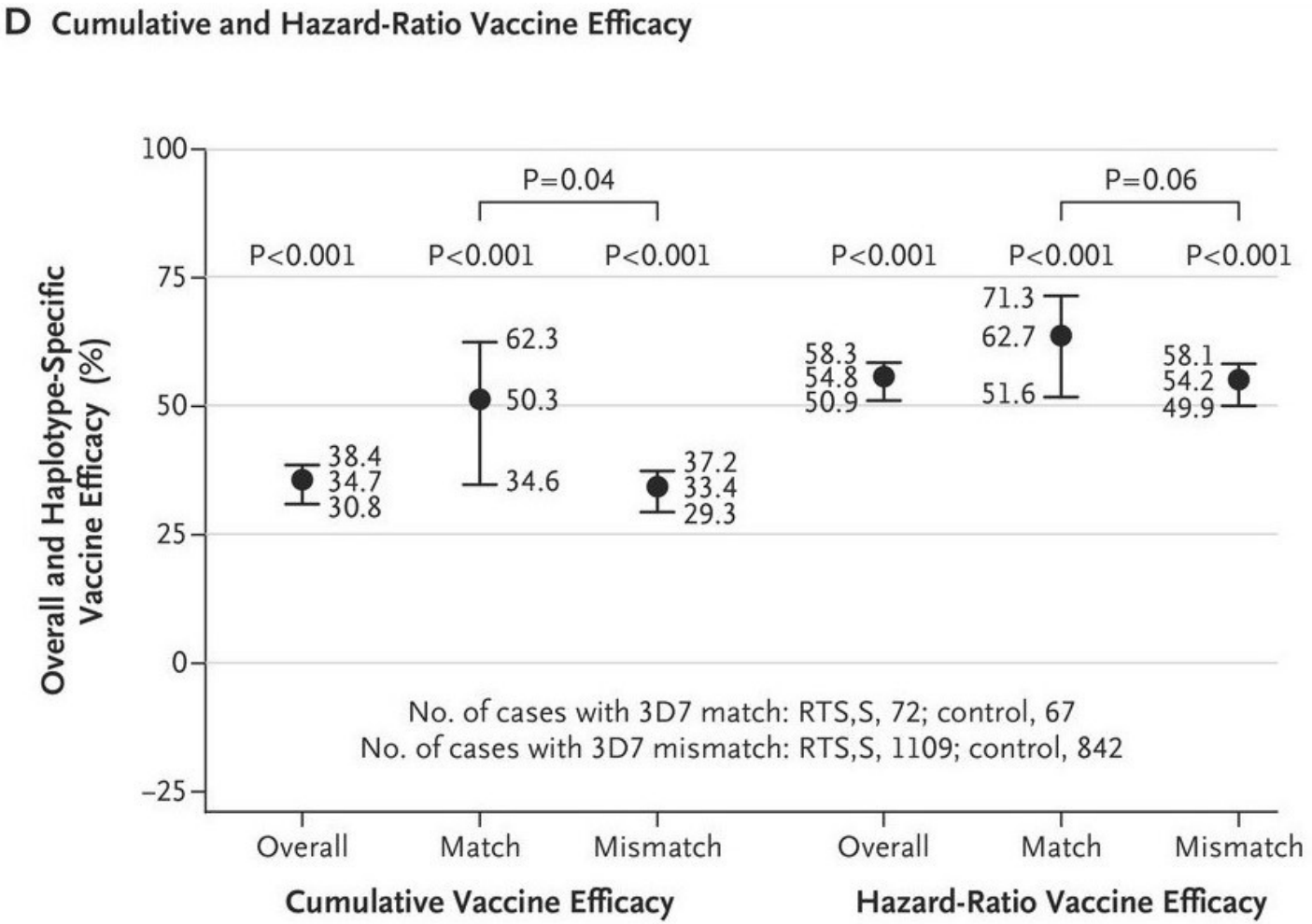
*COVID-19: Developing Drugs and Biological Products for Treatment or Prevention Guidance for Industry (May 12, 2020)

Cumulative VE vs. PH VE: Can Make a Big Difference! (Neafsey, Juraska et al., 2015, *NEJM*)



*Cumulative VE(t) = 1 - $\frac{F_1(t)}{F_0(t)}$ assessed by TMLE methodology (David Benkeser, Marco Carone, Peter Gilbert papers)

Cumulative VE vs. PH VE: Can Make a Big Difference! (Neafsey, Juraska et al., 2015, *NEJM*)



Primary Analyses of VE in HIV Vaccine Efficacy Trials

VE Trial	Years	Primary VE Parameter
Vax004	1998–2003	PH VE
Vax003	1999–2003	PH VE
HVTN 502/Step	2005–2008	PH VE
HVTN 503/Phambili	2006–2008	PH VE
RV144	2003–2009	PH VE
HVTN 505	2009–2013	PH VE
HVTN 702	2017–2020	Cm. VE(t) (t=24, 36)
HVTN 705	2018–	Cm. VE(t) (t=24, 36)
HVTN 706	2019–	Cm. VE(t) (t=24, 36)

Pros and Cons of PH VE

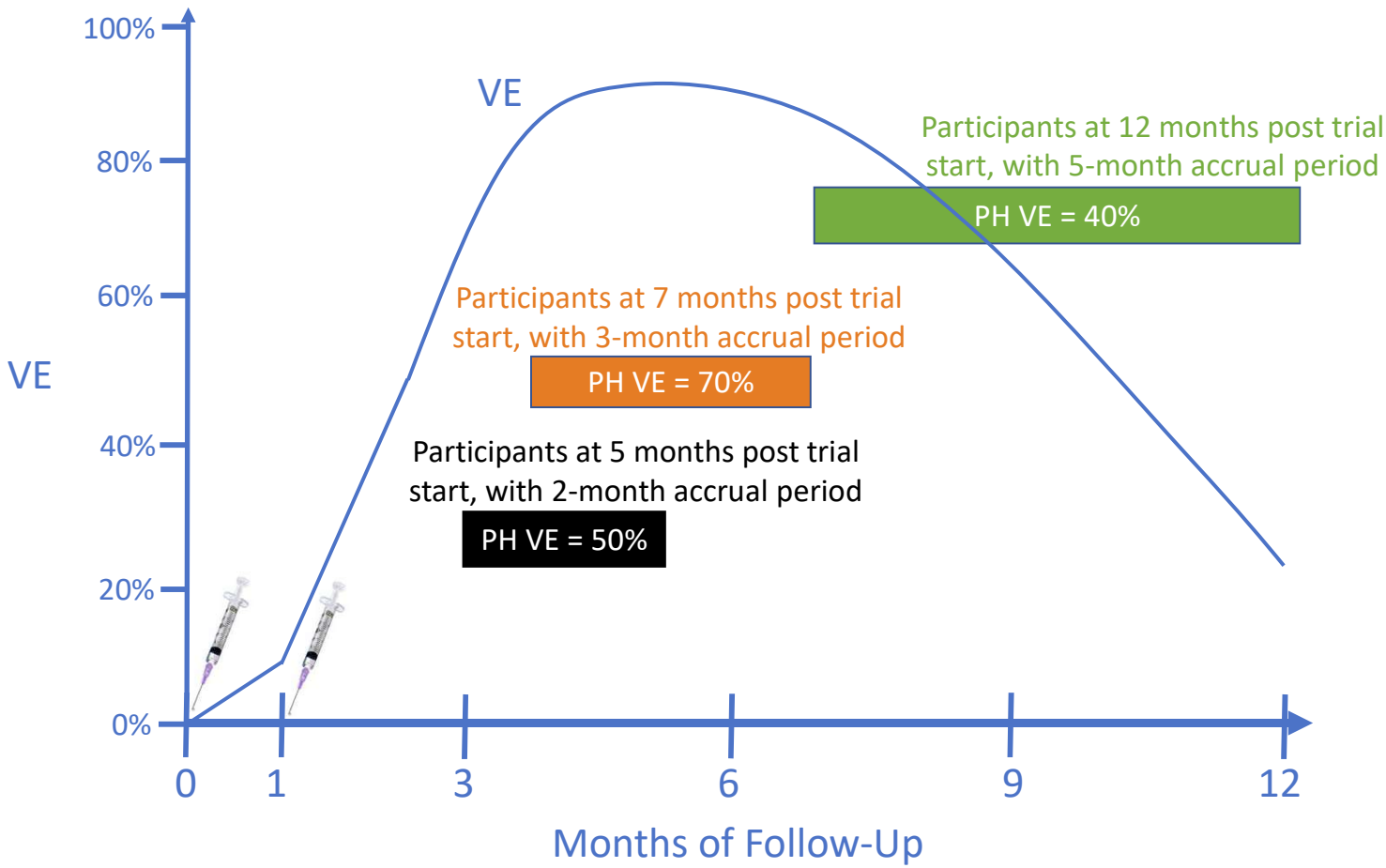
- **Pros:**
  - Simple trial design: power calculations and group sequential monitoring boundaries determined by total number of events
  - Simple interim analyses: for any amount of follow-up can easily analyze VE counting all events
  - Fairly interpretable if PH approximately holds (HR $\approx$ time-averaged relative risk)
  - Nearly efficient if PH approximately holds (and use a covariate-adjusted version)

Pros and Cons of PH VE

- **Cons:**
  - Not a causal effect unless PH holds and the null hypothesis is true
  - PH VE depends on nuisance parameter information (censoring distribution)
  - VE typically varies over time (low at beginning, peaks after vaccinations, then wanes), violating the PH assumption*
  - A focus on reporting the single estimate $\widehat{VE} = 1 - e^{\widehat{\beta}}$ can obscure the fact that the interpretation depends on the period of follow-up
  - Conducting PH VE analysis triggered by a target number of events implies the question addressed changes over time

*The Cox model can readily add time-effects: $\lambda_1(t) = \lambda_0(t)e^{\beta(t)}$

Event-Driven PH VE: Question Addressed is a Random Variable (3 Examples each at a Fixed Number of Endpoints, e.g., 100)



Pros and Cons of Cumulative VE

- **Pros:**
  - Scientific question fixed [e.g., VE(6) naturally addresses WHO criteria "VE for at least 6 months"]
  - A causal effect without specifying any parametric assumptions
  - Transparently studies how VE(t) changes over time
  - Models of population impact of VE show durability is a key parameter
    - Need to input time-dependent VE into the models
  - Robust and efficient methods with covariate-adjustment are well understood
    - Many papers including Moore and van der Laan (2009); papers by David, Marco, Peter; and a grant proposal by David and Alex

FDA Support for Covariate-Adjusted VE Analysis in COVID-19 FDA Guidance Document (May 12, 2020)*

"*To improve the precision of treatment effect estimation and inference, sponsors should consider adjusting for prespecified prognostic baseline covariates (e.g., age, baseline severity, comorbidities) in the primary efficacy analysis and should propose methods of covariate adjustment."

*COVID-19: Developing Drugs and Biological Products for Treatment or Prevention Guidance for Industry (May 12, 2020)

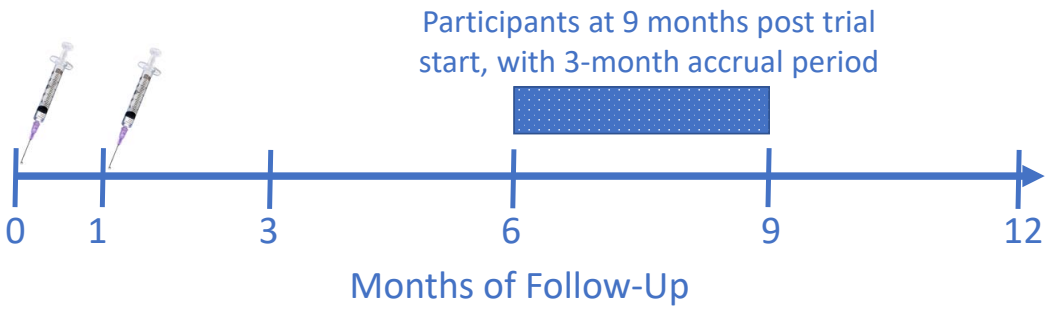
Pros and Cons of Cumulative VE

- **Cons:**
  - More complicated trial design: power and group sequential monitoring boundaries depends on additional features besides event totals*
    - Numbers enrolled, accrual pattern, incidence over time, VE over time
  - More complicated for handling the available follow-up at a given analysis time
    - E.g., Time-dependent Kaplan-Meier monitoring (Brittain, Follmann, Yang, 2008, *Biometrics*)

*HVTN uses the *seqDesign* R package led by Michal Juraska and Doug Grove, and RCTdesign by Scott Emerson

PH VE vs. Cm. VE in Light of Follow-Up

Success criterion: Estimated VE > 0.50 over **about 6 months** post first vaccination with some precision (e.g., lower 95% confidence limit $> 30\%$)



- PH VE approach: Count all events through month 9, such that $\text{VE} = 1 - \frac{\lambda_1(t)}{\lambda_0(t)}$ averaging over all follow-up
- Cm. VE approach: Estimate VE(6), only counting events through month 6

Which approach wins? Counterbalancing effects of including more events vs. waning VE after 6 months

Potential Approaches

1. **Cm. VE via simple PH model:** Based on Cox model fit including all events, estimate VE(6) by

$$\widehat{VE}(6) = 1 - \frac{\sum_{i=1}^{N_1} \widehat{F}(6|X_i, A_i = 1)}{\sum_{i=1}^{N_0} \widehat{F}(6|X_i, A_i = 0)}$$

Potential Approaches

1. **Cm. VE via simple PH model:** Based on Cox model fit including all events, estimate VE(6) by

$$\widehat{VE}(6) = 1 - \frac{\sum_{i=1}^{N_1} \widehat{F}(6|X_i, A_i = 1)}{\sum_{i=1}^{N_0} \widehat{F}(6|X_i, A_i = 0)}$$

2. **Cm. or PH VE via time-varying PH model:** Similar but use time-varying coefficients $\beta(t)$ such as $\beta(t) = \beta_0 + \beta_1 log(t)$, or more flexible (e.g., splines)

Potential Approaches

1. **Cm. VE via simple PH model:** Based on Cox model fit including all events, estimate VE(6) by

$$\widehat{VE}(6) = 1 - \frac{\sum_{i=1}^{N_1} \widehat{F}(6|X_i, A_i = 1)}{\sum_{i=1}^{N_0} \widehat{F}(6|X_i, A_i = 0)}$$

2. **Cm. or PH VE via time-varying PH model:** Similar but use time-varying coefficients $\beta(t)$ such as $\beta(t) = \beta_0 + \beta_1 log(t)$, or more flexible (e.g., splines)
3. **Cm. VE via doubly-robust method:** E.g., apply Superlearner and a TMLE bias adjustment to estimate $F^{(1)}(6|X_i)$ and $F^{(0)}(6|X_i)$ for each $i = 1, \cdots, n$, to obtain

$$\widehat{VE}(6) = 1 - \frac{\sum_{i=1}^{N} \widehat{F}^{(1)}(6|X_i)}{\sum_{i=1}^{N} \widehat{F}^{(0)}(6|X_i)}$$

Potential Approaches

1. **Cm. VE via simple PH model:** Based on Cox model fit including all events, estimate VE(6) by

$$\widehat{VE}(6) = 1 - \frac{\sum_{i=1}^{N_1} \widehat{F}(6|X_i, A_i = 1)}{\sum_{i=1}^{N_0} \widehat{F}(6|X_i, A_i = 0)}$$

2. **Cm. or PH VE via time-varying PH model:** Similar but use time-varying coefficients $\beta(t)$ such as $\beta(t) = \beta_0 + \beta_1 log(t)$, or more flexible (e.g., splines)
3. **Cm. VE via doubly-robust method:** E.g., apply Superlearner and a TMLE bias adjustment to estimate $F^{(1)}(6|X_i)$ and $F^{(0)}(6|X_i)$ for each $i = 1, \cdots, n$, to obtain

$$\widehat{VE}(6) = 1 - \frac{\sum_{i=1}^{N} \widehat{F}^{(1)}(6|X_i)}{\sum_{i=1}^{N} \widehat{F}^{(0)}(6|X_i)}$$

Standard implementations include all events through 9 months (approaches 1 and 2) or only through 6 months (approach 3)

Innovative Ways to Improve Doubly Robust Approach?

1. Estimate each $F^{(a)}(t = 9|X)$ using parametric models only in the Superlearner library, so that all endpoints are included in the estimate of $F^{(a)}(t = 6|X)$
2. Estimate each $F^{(a)}(t|X)$ for all $t \in [0, 9]$ months using TMLE with an arbitrary Superlearner library, report point and confidence bands for VE(t) for all $t \in (0, 9]$ months, and test

$$H_0 : VE(t) \leq 0.30 \text{ for all } t \in [6, 9] \text{ vs.}$$

$$H_1 : VE(t) > 0.30 \text{ for at least one } t \in [6, 9]$$

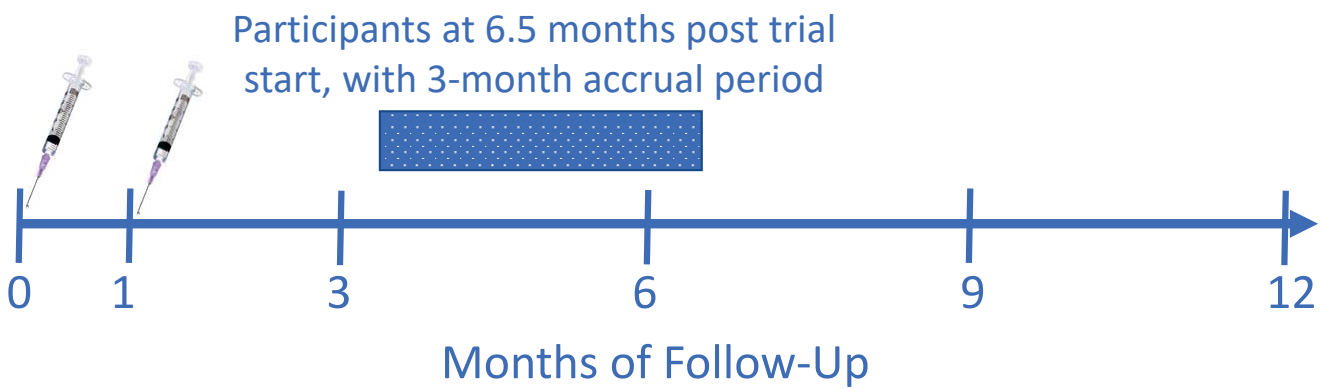
- Test statistic:

$$T = sup_{t \in [6,9]} \frac{\left[log(1 - 0.30) - log(1 - \widehat{VE}(t)) \right]}{\widehat{\sigma}(t)}$$

where $\widehat{\sigma}(t)$ is the square-root of a estimate of the variance of $log(1 - \widehat{VE}(t))$
(hypothesis test suggested by Alex)

Simpler Option

- Plan the final analysis for 6-month vaccine efficacy when BOTH reach the needed number of events AND $\approx$ 1000 participants reach the month 6 visit, and use doubly-robust covariate-adjusted analysis of Cm. VE(6)
  - Final analysis 2.5 months earlier; may need to increase total sample size by 20% to achieve same number of events (average follow-up 5 months vs. 6 months)

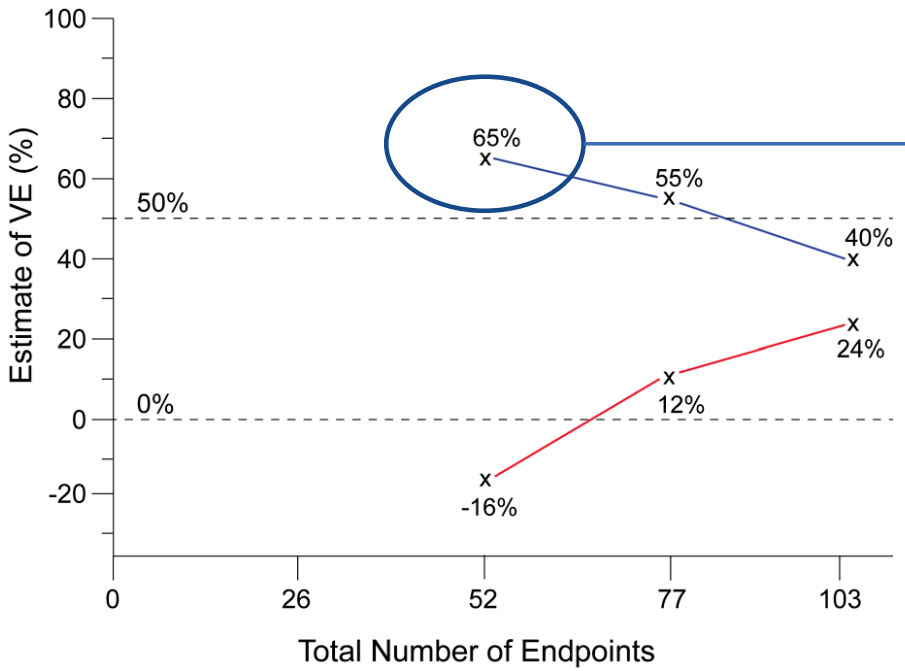


Primary Approach to Assess Vaccine Efficacy Against COVID-19?

Selection of approach to primary analysis of vaccine efficacy needs to balance multiple issues

- Interpretability of answers
- Time-variations in VE
- Statistical robustness and efficiency
- Urgent need to approve a vaccine as soon as the evidence basis for meeting vaccine success criteria is sufficiently solid
- Also need to capture data on durability of VE (e.g., concern for "late vaccine-enhanced disease" as immune responses fall to low levels)
  - E.g., CYD-TDV dengue vaccine (Sridhar et al., 2018, *NEJM*)
- Communication with colleagues from diverse disciplines and with community advisory boards/public/press

Group Sequential Monitoring of Vaccine Efficacy



Estimated VE > 65% at 52 endpoints = Meet success criteria

Expect first analysis ~4.5 months after trial initiation
(3-month accrual, 90% baseline SARS-CoV-2 negative, 0.83% 6-month incidence in placebo arm)

*2:1 Vaccine:Placebo design with 2 interim analyses at 52 and 77 endpoints. O'Brien-Fleming monitoring boundaries to detect VE LB > 20% (benefit) and to detect VE UB < 50% (lack of benefit) based on 1-sided 0.025-level testing accounting for interim analyses.

14

Accounting for Covariate-Dependent Loss-to-follow-up

- Standard survival analyses (Kaplan-Meier, standard Cox PH) assume censoring is completely random
- Many methods exist that relax this assumption, e.g., allowing censoring to depend on covariates
  - E.g., TMLE methods (Moore and van der Laan, 2009; David B et al. papers) are doubly robust, providing consistent estimation of Cm. VE(t) if either (A) $F(t|X, A)$ is consistently estimated or (B) the conditional censoring distribution is consistently estimated
  - Thus, if retention is very high, or censoring can be approximately modeled correctly, then TMLE and related doubly-robust methods are robust to modeling of the outcome regression $F(t|X, A)$

Per-protocol (PP) VE*

- So far have focused on VE counting all events post enrollment (intention-to-treat analysis)
- Key supplementary analysis: Assess VE in the PP cohort, only counting events occurring more than 14 days post last vaccination [e.g., VE(1.5 – 6)]

  PP cohort = Endpoint free at 14 days post first vaccination AND received both inoculations AND did not have other specified protocol violations

  - Standard analyses are susceptible to post-randomization selection bias

*Horne, Lachenbruch, Goldenthal (2000, *Vaccine*)

Per-protocol (PP) VE: Two Approaches

1. **Average Treatment Effect (ATE)**: Assume $A = 1, PP = 1$ vs. $A = 0, PP = 1$ is randomized conditional on X and apply a method such as TMLE to estimate causal VE$(1.5 - 6)$
   - This randomization is not 100% conceivable, because $PP = 1$ includes being free of the endpoint by 14 days, which cannot be assigned
   - Only a minor concern given few endpoints by 14 days?

Per-protocol (PP) VE: Two Approaches

1. **Average Treatment Effect (ATE)**: Assume $A = 1, PP = 1$ vs. $A = 0, PP = 1$ is randomized conditional on X and apply a method such as TMLE to estimate causal VE$(1.5 - 6)$
   - This randomization is not 100% conceivable, because $PP = 1$ includes being free of the endpoint by 14 days, which cannot be assigned
   - Only a minor concern given few endpoints by 14 days?
2. **Principal stratification approach** (e.g., Gilbert, Shepherd, Hudgens, 2013, *JASA*): Assess VE$(1.5 - 6)$ in the subgroup who would be $PP = 1$ under both treatment assignments
   - A causal effect without an inconceivable assignment
   - Identifiability challenges due to missing counterfactual data
   - Standard analysis valid if no-vaccine-harm and include all covariates X that dually predict PP and the endpoint; otherwise may give biased results

Per-protocol (PP) VE: Two Approaches

1. **Average Treatment Effect (ATE)**: Assume $A = 1, PP = 1$ vs. $A = 0, PP = 1$ is randomized conditional on X and apply a method such as TMLE to estimate causal VE$(1.5 - 6)$
   - This randomization is not 100% conceivable, because $PP = 1$ includes being free of the endpoint by 14 days, which cannot be assigned
   - Only a minor concern given few endpoints by 14 days?
2. **Principal stratification approach** (e.g., Gilbert, Shepherd, Hudgens, 2013, *JASA*): Assess VE$(1.5 - 6)$ in the subgroup who would be $PP = 1$ under both treatment assignments
   - A causal effect without an inconceivable assignment
   - Identifiability challenges due to missing counterfactual data
   - Standard analysis valid if no-vaccine-harm and include all covariates X that dually predict PP and the endpoint; otherwise may give biased results

Covariates X play an important role, and given it is not possible to fully verify assumptions, sensitivity analysis is warranted

Assumption of Independent Units/No Interference

- All of the methods discussed assume that one participant's treatment assignment does not affect the outcome of any other participants
- May be violated, e.g.
  - Enroll multiple members of a household
  - Enroll co-workers who work next to each other
  - Enroll multiple individuals from the same institution (e.g., nursing home staff)
- To what extent should design recruitment strategy to minimize interference?
- How much does the interference invalidate estimation and inference, and what methods to use? (Gary Chan)

A Lot of Work to Do

"We will have the first readouts on COVID-19 vaccine efficacy in January of 2021"

- Anthony Fauci, May 2020

Next week:
Part II Alex Luedtke "Comparison of Statistical Methods for Assessing Vaccine Efficacy in COVID-19 Vaccine Efficacy Trials"