HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

Everything keeps changing! What COVID-19 has taught us about pandemic surveillance

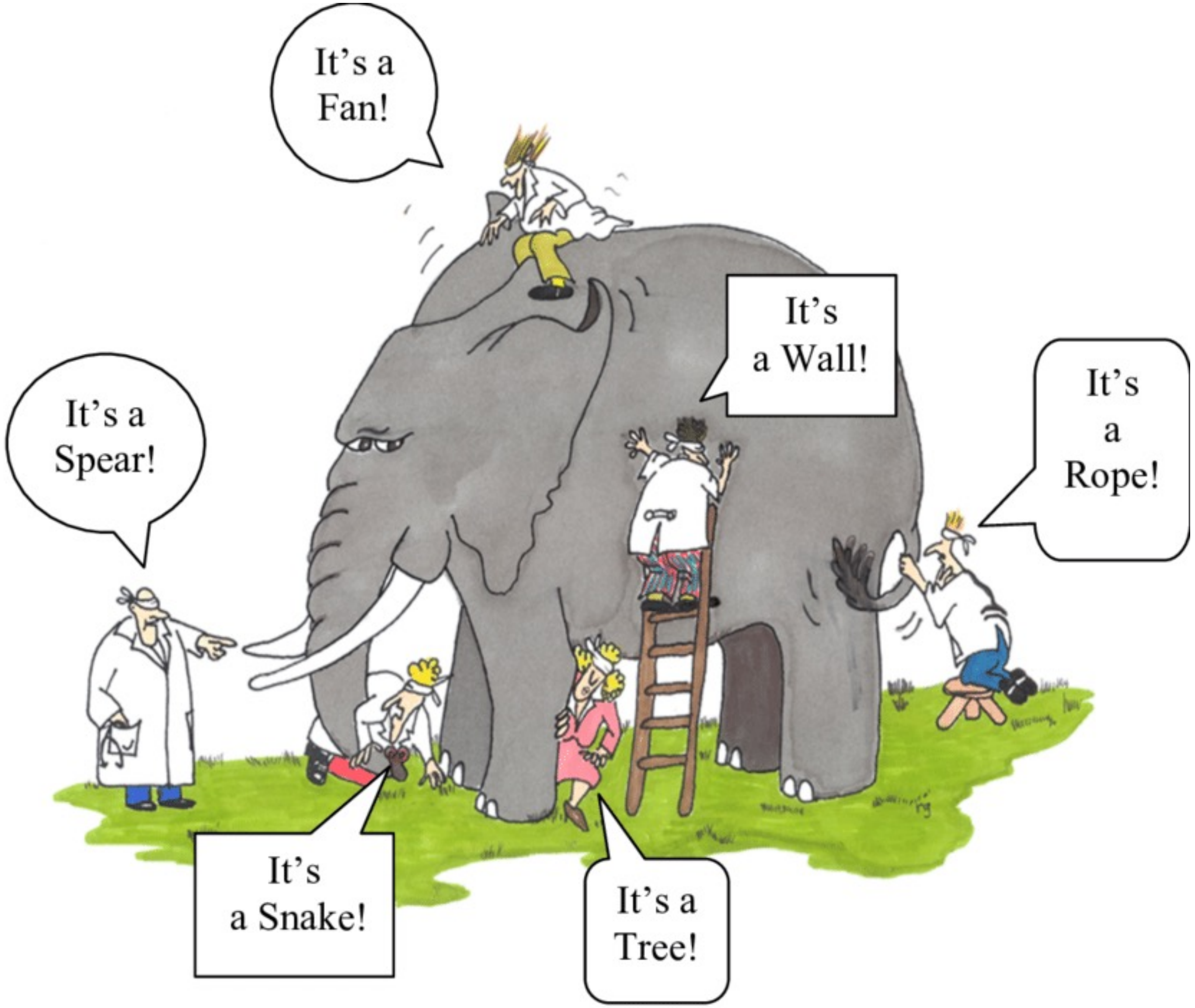
Marc Lipsitch

Fields Institute MfPH seminar

January 24, 2023

CENTER *for* COMMUNICABLE DISEASE DYNAMICS

Surveillance means many different things to different people



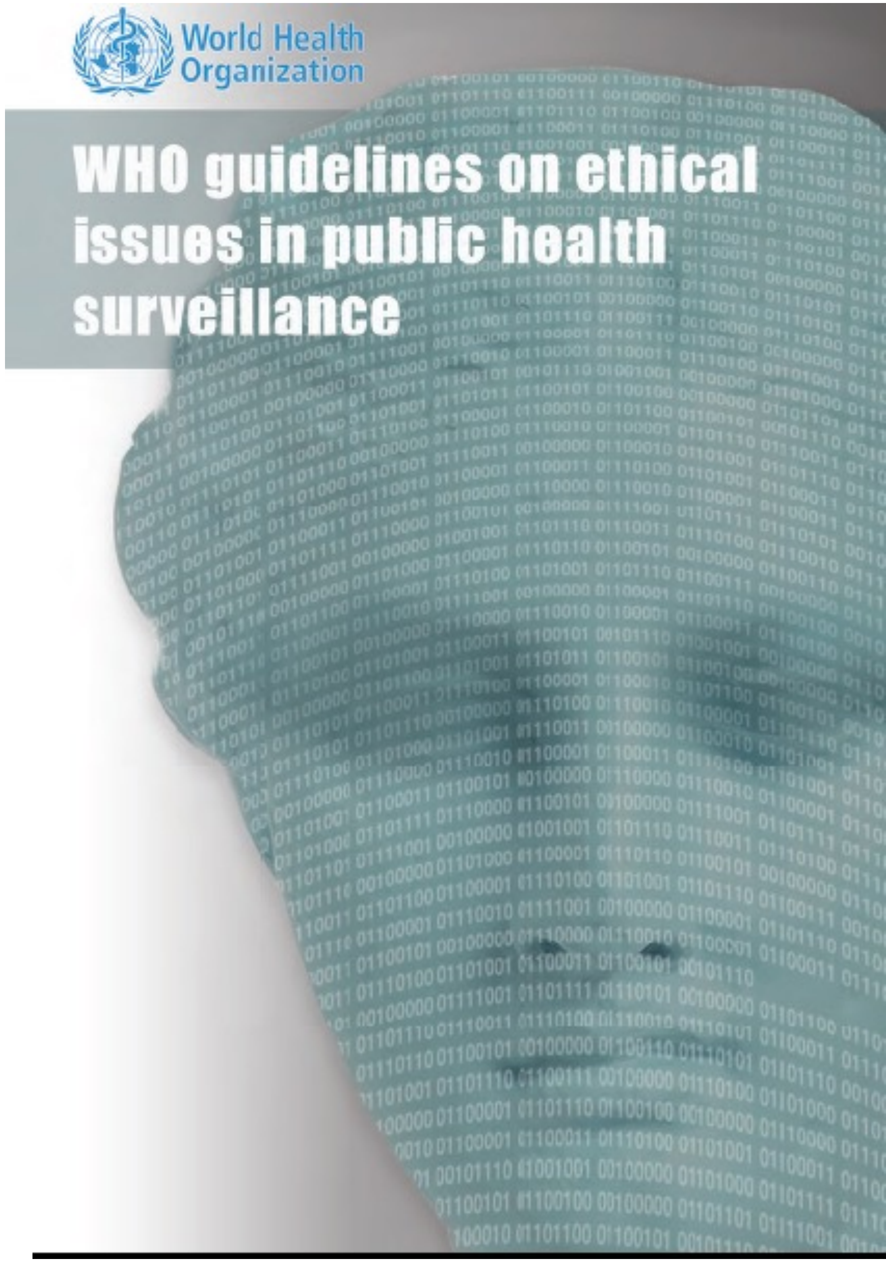
Each view is different, and strongly held

An ethical imperative for surveillance

Guideline 1. Countries have an obligation to develop appropriate, feasible, sustainable public health surveillance systems. Surveillance systems should have a clear purpose and a plan for data collection, analysis, use and dissemination based on relevant public health priorities.

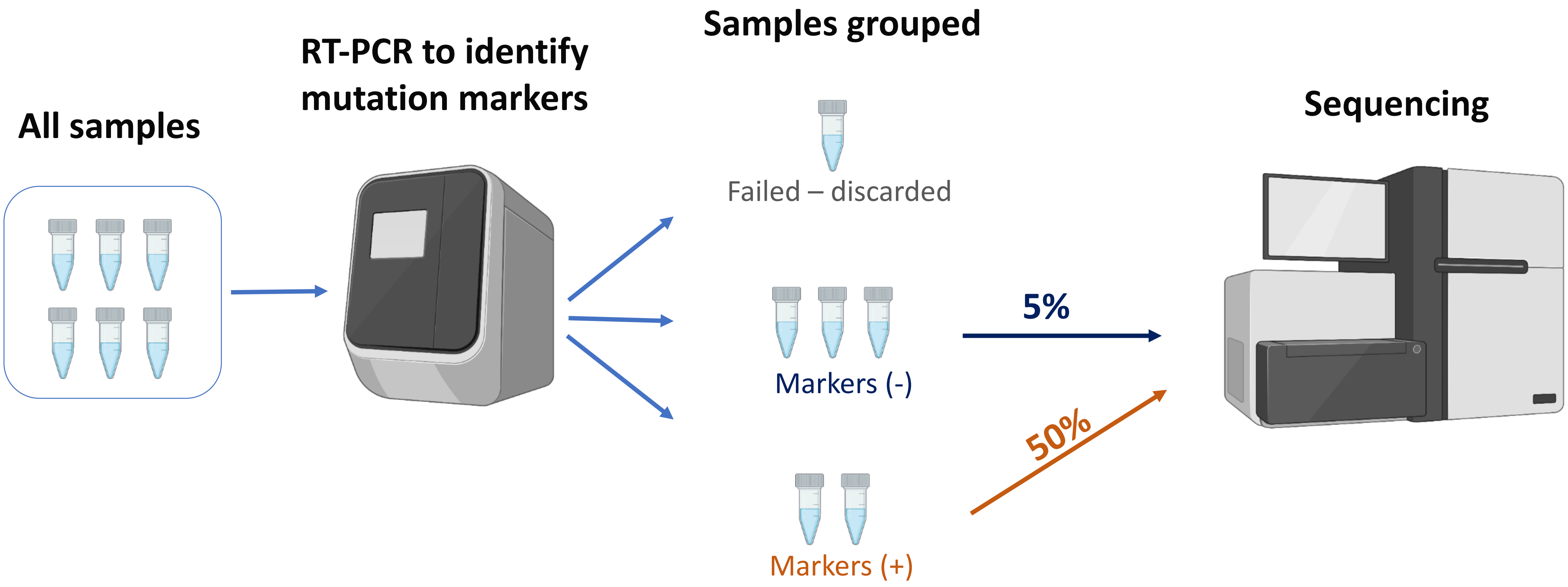
Guideline 4. Countries have an obligation to ensure that the data collected are of sufficient quality, including being timely, reliable and valid, to achieve public health goals.

Guideline 6. The global community has an obligation to support countries that lack adequate resources to undertake surveillance.



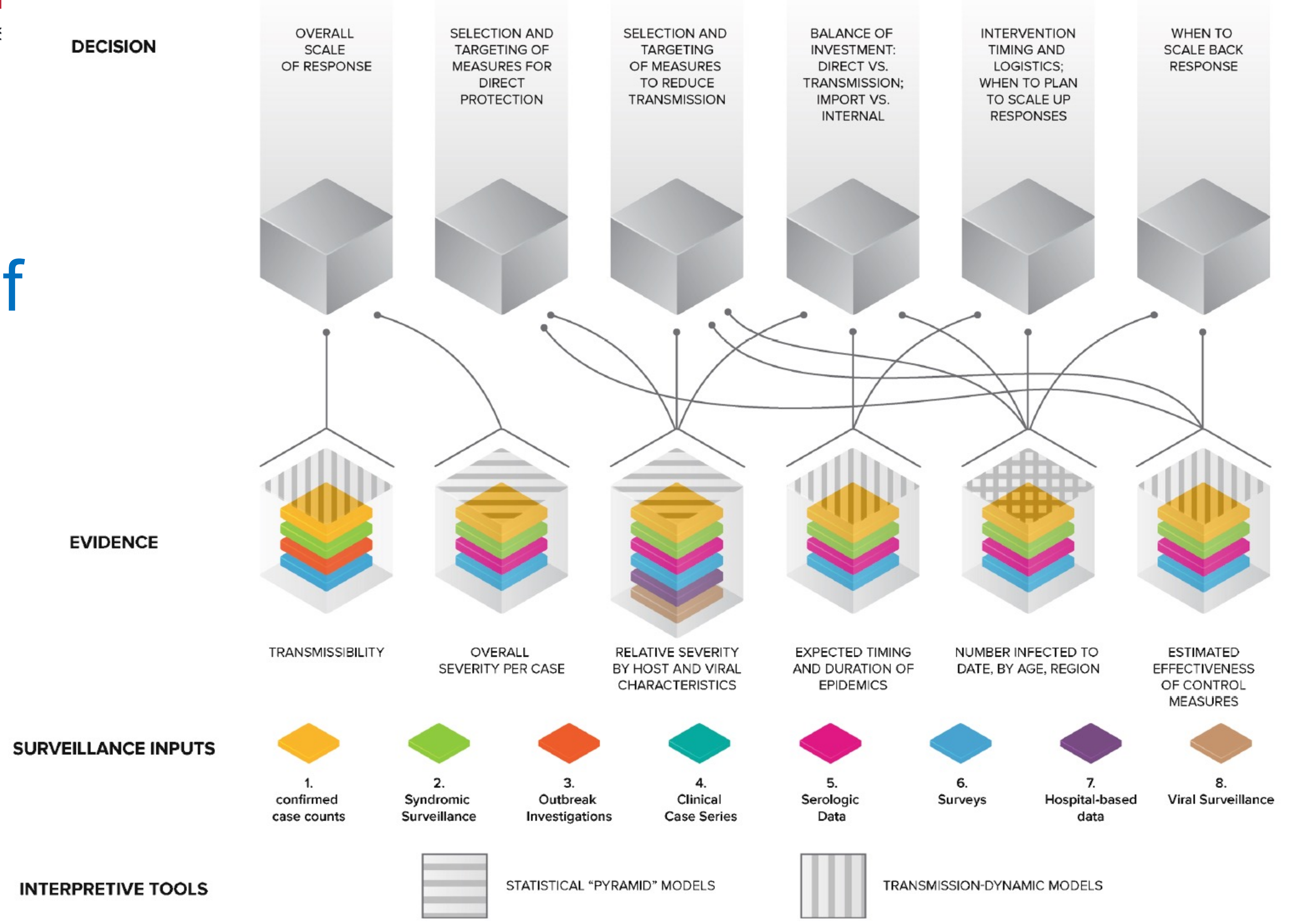
A data science view of surveillance

Laboratory-centric view of surveillance (it's still around!): enrich for the unusual/interesting



Xueting Qiu

A public health view of surveillance



Lipsitch & Santillana 2019 Curr Topics Micro Immuno

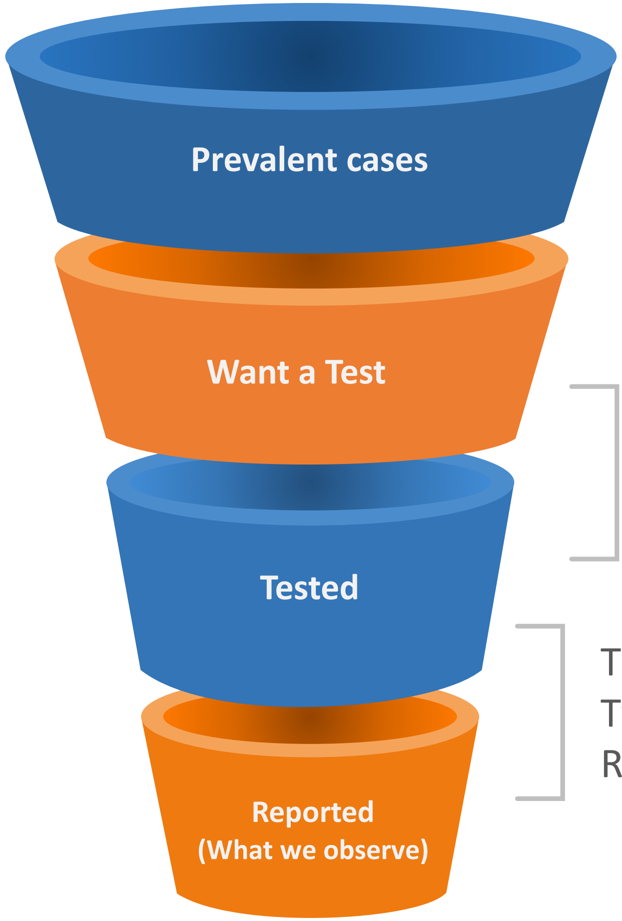
Lipsitch et al. Biosec Bioterror 2011

COVID-19 case surveillance measures no “natural” quantity

Determinants

Incidence, duration

Travel
Work/school screenings
Known exposure
Symptoms



Barriers

Fear of losing work (+/- sick leave)
Attitudes toward COVID-19

Test/reagent/swab shortages
Distance/cost/time loss for testing

Test sensitivity
Type of test (at home not reportable)
Reporting fail/delay

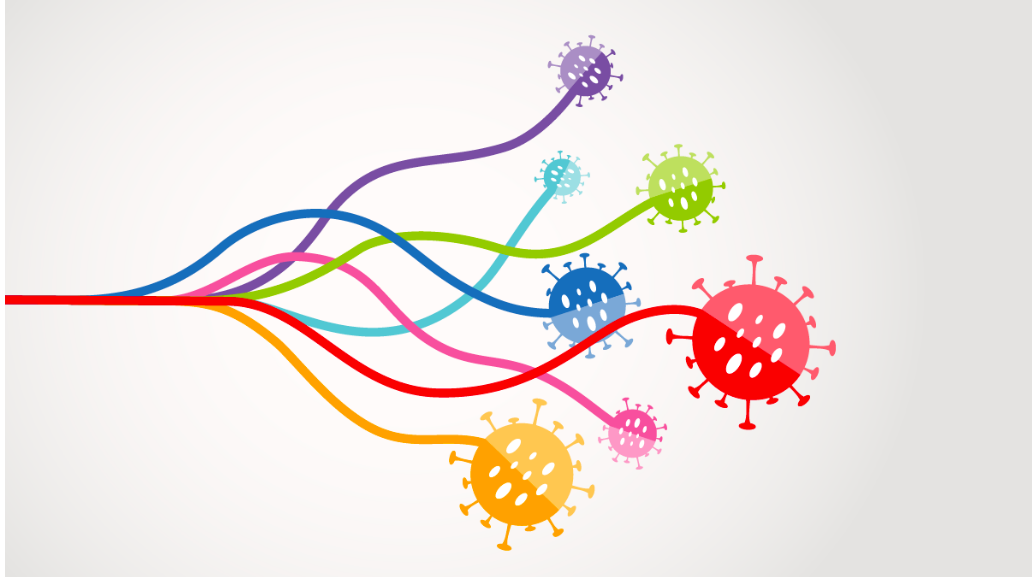
Surveillance as a Swiss Army Knife

A portfolio of specialized instruments, designed for different purposes

- Detection
- Individual assessment/treatment
- Characterizing severity, countermeasure effects
- Burden

Detection: Why?

Activate countermeasure development

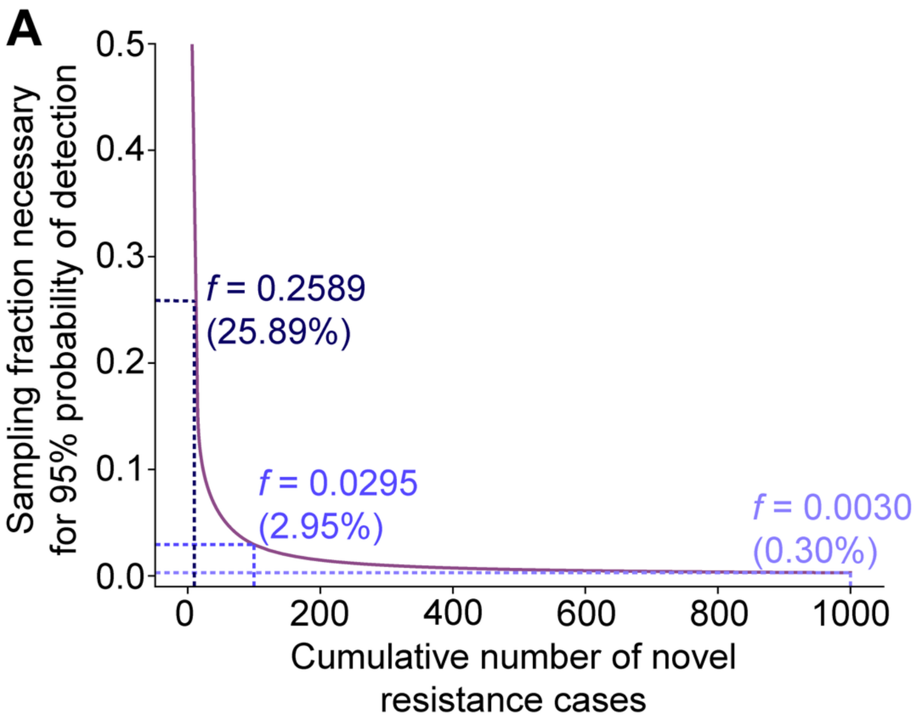


Prepare for variant spread

Detection: How?

Geographic scale

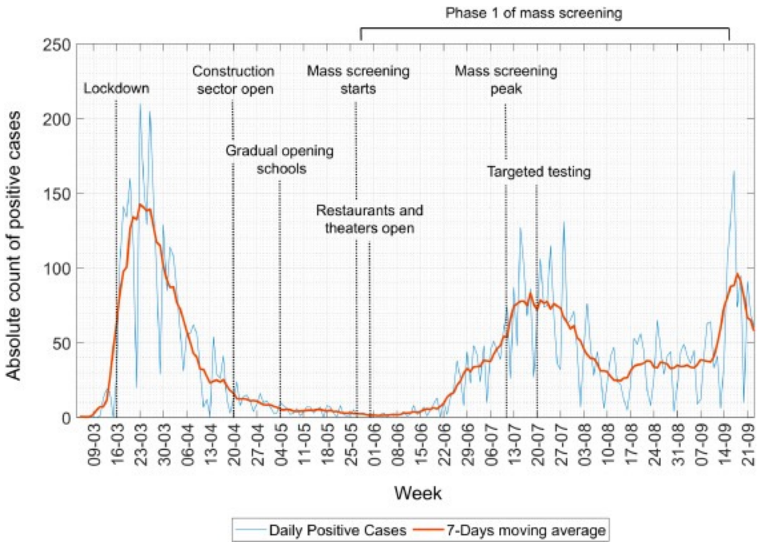
Size: for a new pathogen, bigger is better;
not that big for variants
(rule of 3)



Hicks et al. *PLoS Biology* 2019

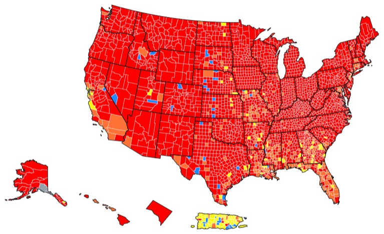
Quantifying burden: why?

- Situational awareness/ trends
- Population immunity from natural infection
- Evaluating control measures
- Identify hotspots/needs
- Identify inequities



Wilmes et al. *Lancet RH Eur* 2021

Level of Community Transmission of All Counties in US

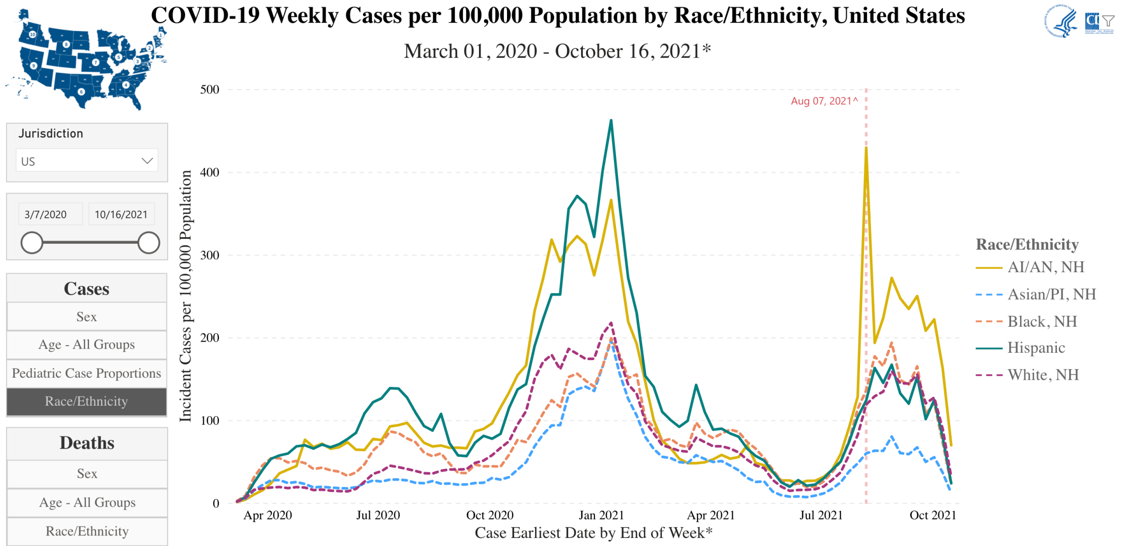


Community Transmission in US by County

	Total	Percent	% Change
High	2726	84.61%	-5.31%
Substantial	319	9.9%	3.76%
Moderate	122	3.79%	0.99%
Low	49	1.52%	0.5%

How is community transmission calculated?

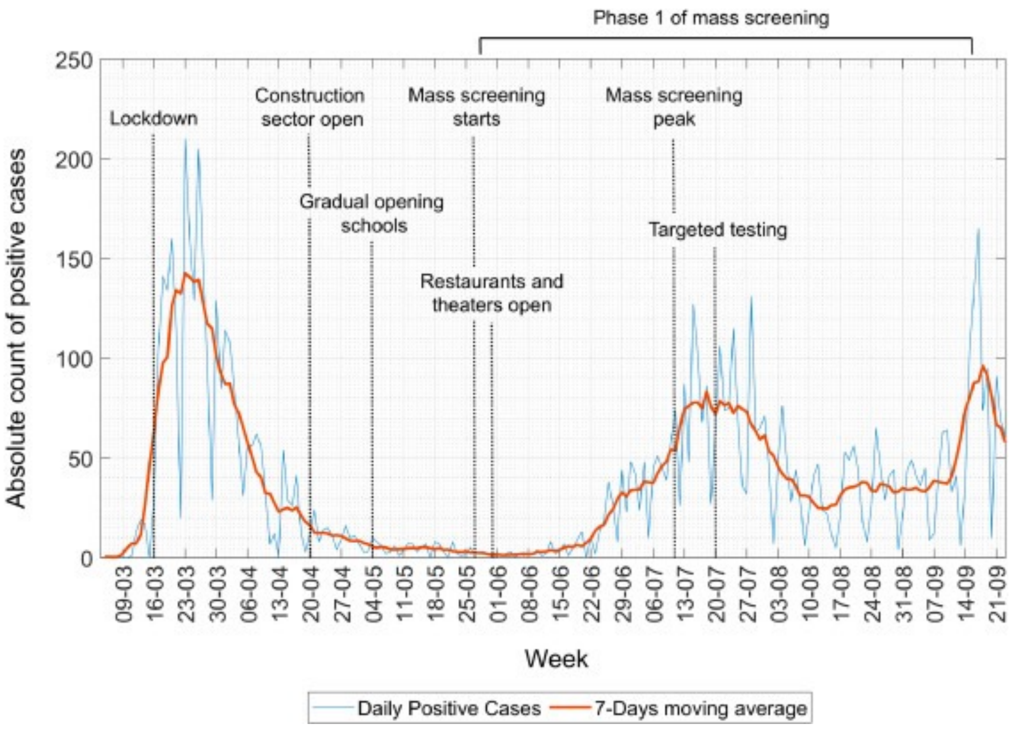
High Substantial Moderate Low No Data



Quantifying burden: how?

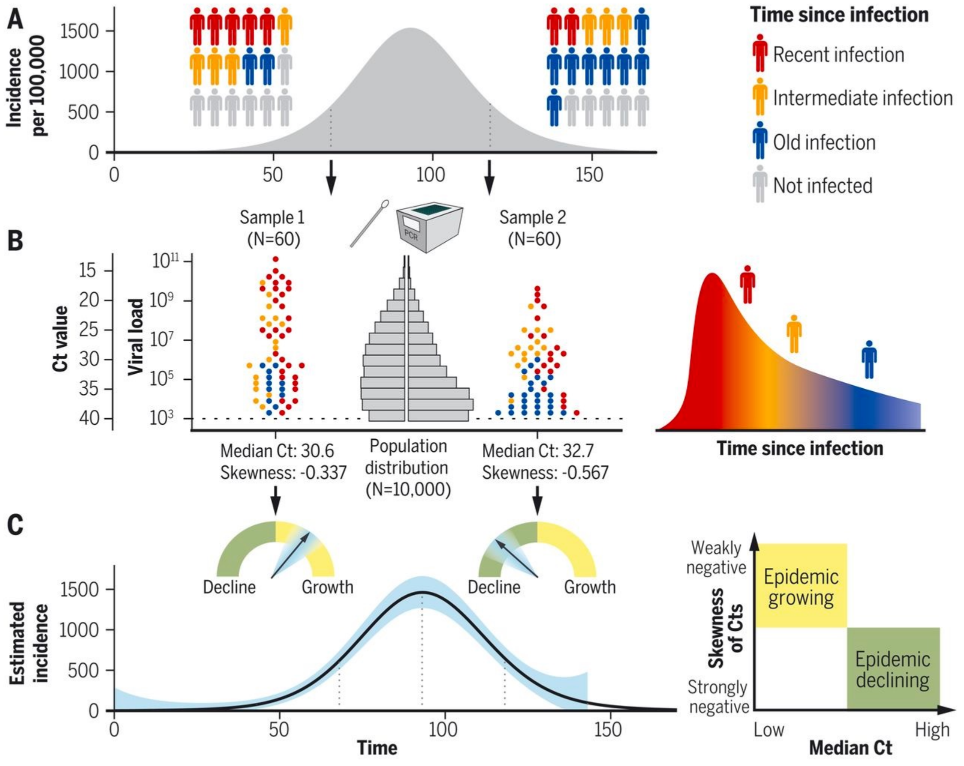
Representativeness!!!

Data completeness



Wilmes et al. *Lancet RH Eur* 2021

Viral load can help identify trend



Hay et al. *Science* 2021

Random sampling: the gold standard for quantifying disease and a lot better than alternatives

A personal take on science and society

World view

By Natalie Dean

Tracking COVID-19 infections: time for change

To manage the pandemic effectively, channel the power of random sampling.

One of the best ways the world has to get a clear view of COVID-19 is going underused. It's time to exploit the power of random sampling.

Last September, the US Centers for Disease Control and Prevention estimated that only one in four SARS-CoV-2 infections in the United States had been reported. Across Africa, the average is closer to one in seven. Why? Many people who are quite ill, or worried

> "Without random sampling, there's a vicious cycle of guesswork."

blood tests monthly from around 150,000 people. In late January, one in 20 people tested positive for current infection. But age really mattered: one in 10 of the youngest children tested positive, as did one in 15 of the older children. The results signalled an enormous pool of infections, and were quickly made available to guide policy and family decisions.

Forecasting the course of the pandemic demands reliable estimates of current infection levels. Without accurate knowledge of these levels, epidemiologists must make many assumptions (on the likelihood that, for example, infected people will develop symptoms, or be tested). That guesswork informs mathematical models and, con-

UK random samples (REACT and ONS CIS) provided timely, variant-specific estimates of

- Prevalence of infection (and thus estimates of incidence)
- Cumulative incidence of infection
- Detection and growth rate of new variants
- Vaccine effectiveness dependent on vaccine, number of doses, time since last dose, and variant
- Symptom profile and its evolution through time
- Evidence of reinfection and degree of protection
- Equity measures – comparisons across groups
- Infection-fatality rate

Alternative: testing on hospital admission for non-covid, followed by reweighting to adjust demographics to the full population

Routine Hospital-based SARS-CoV-2 Testing Outperforms State-based Data in Predicting Clinical Burden

Leonard Covello,[a] Andrew Gelman,[b] Yajuan Si,[c] and Siquan Wang[d]

Epidemiology • Volume 32, Number 6, November 2021

Target product profiles: The right tool(s) for the job(s)

	Detect	Assess Individuals	Characterize severity/ countremeasures	Quantify Burden
Size	+++/++	+++	++	+
Geographic Coverage	+++	+++	+	+++
Sensitivity	++	++ (but specific meaning)	++	+
Specificity	++	+++	++	+
Frequency	+	+++	++	+
Data completeness	+	+	+++	+++
Precision (VL)	+	++	-	++
Representativeness	+/-	-	+/-	+++
		Test/trace Clinical	Cohorts/ Payer-provider data	Random sample

Before COVID

Parameters – estimate once

- Virus sequence
- Severity
- Viral load kinetics
- Transmissibility
- +- Vaccine effectiveness
- Natural history
- Sequelae

Surveillance targets – changing daily

- Cases
- Hospitalizations
- Deaths
- Etc.

COVID

Surveillance targets – changing weekly-monthly due to virus and host changes

- Virus sequence
- Severity
- Viral load kinetics
- Transmissibility
- Vaccine effectiveness
- Natural history
- Sequelae

Surveillance targets – changing daily

- Cases
- Hospitalizations
- Deaths
- Etc.

Value of integrated health systems

Conclusions (Part I)

- Surveillance is not one general activity, but many specific ones
- Scientific questions and decisions should motivate requirements for each surveillance activity
- This imposes constraints and resource needs, but can also be freeing: only certain requirements for each activity
- In particular:
  - Surveillance is about more than just detection
  - For certain purposes, quality of data is more important than representativeness or scale
  - Well-designed epidemiologic studies, especially longitudinal ones, are an integral part of surveillance
  - For other purposes, approximating a random sample is critical
- COVID has expanded the remit of surveillance to include activities to monitor changing quantities that used to be called research to measure fixed quantities. Part II to explore

Part II: Integrated health systems: potential role in pandemic surveillance

- Longitudinal data on individuals permits:
  - Detailed matching of cohorts for prospective studies to control confounding
  - Subgroup analysis
  - Detailed follow-up of individuals in transition from outpatient to inpatient
  - ?Automated merging of vaccine data with outcomes
  - No need for subtraction as has been done in many MOH studies: direct data on the unvaccinated
- Multiple uses for these platforms:
  - Vaccine effectiveness and waning
  - Vaccine safety
  - Sequelae
  - Severity
  - Resource use/ length-of-stay

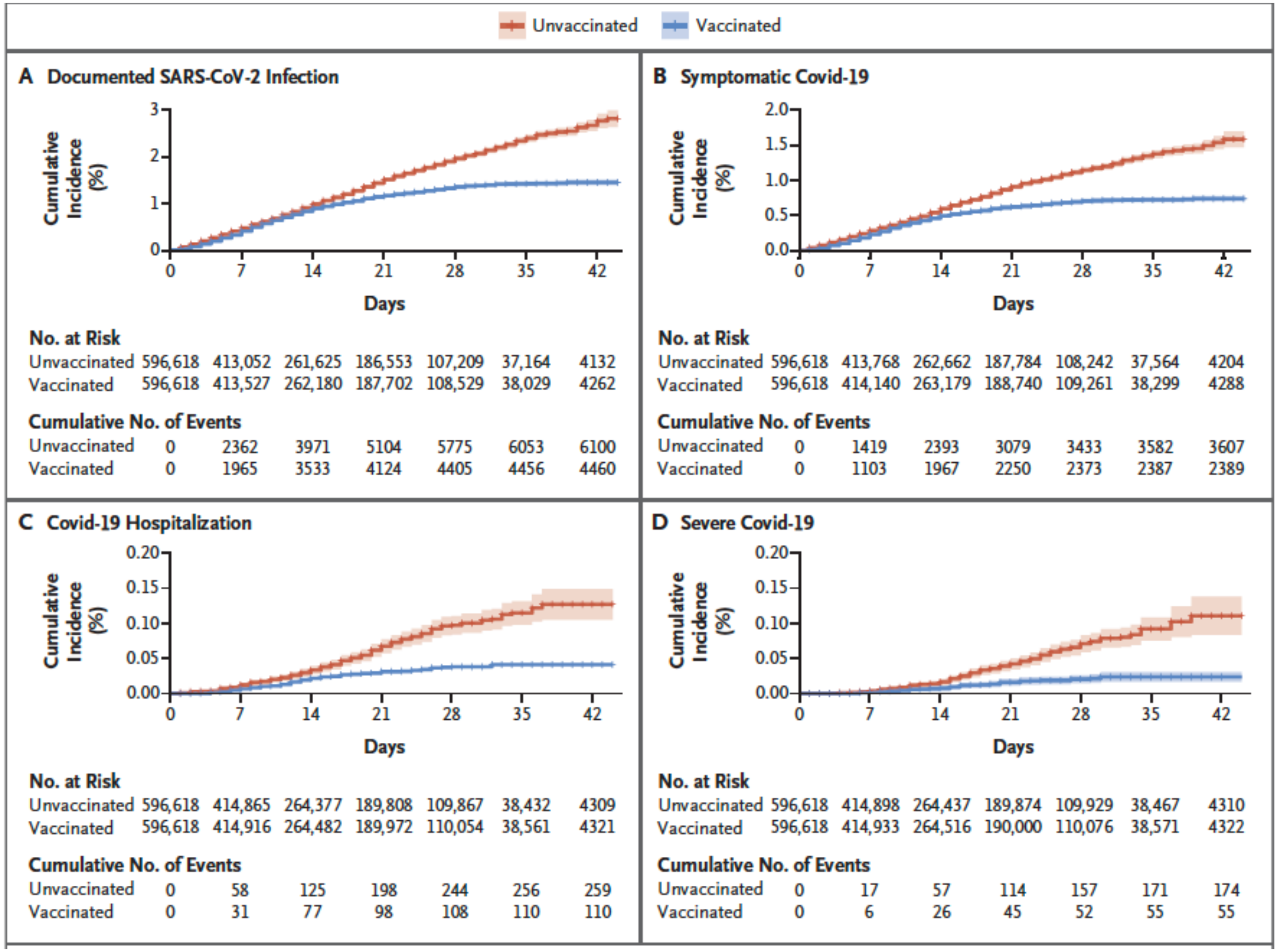
1. VE with Clalit Research Institute, Israel

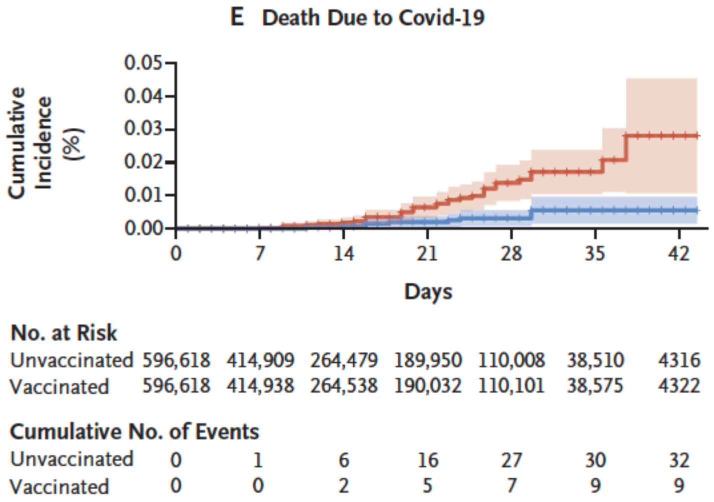
Ran Balicer
Noa Dagan
Noam Barda
Ben Reis
Miguel Hernán

Sonia Hernández-Diaz
Michael Leschinsky
Eldad Kepten

Tal Biron-Shental
Maya Makov-Assif
Galanit Key
Isaac Kohane
Yatir Ben-Shlomo
Doron Netzer
Jacob Waxman

Vaccine effectiveness in Israel





N Dagan, N Barda et al. *NEJM* 2021

Approach

- Matched cohort VE design emulating an RCT, 596,000+ individuals per arm
- Matching on age, sex, social sector, neighborhood, history of flu vaccine, pregnancy, number of coexisting conditions
- Negative control outcome – protection during d0-12 post-dose-1. Used to check adequacy of matching
- Individuals who had been unexposed could become exposed when vaccinated and rematch to a new unexposed

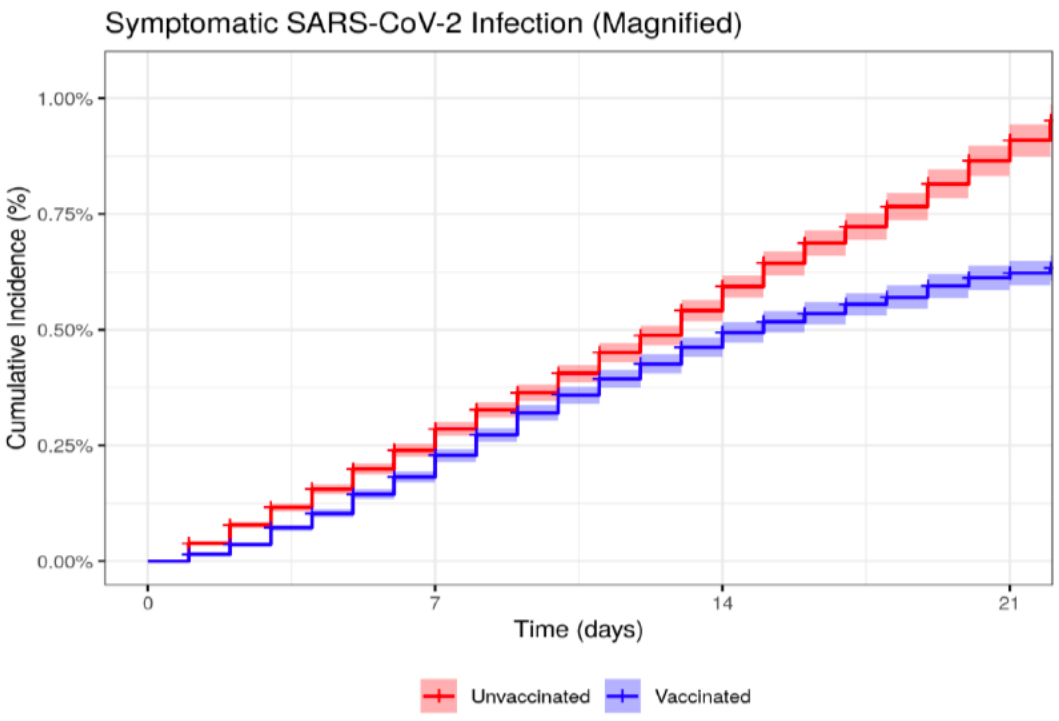
Vaccine effectiveness by subgroup

Table 3. Estimated Vaccine Effectiveness against Covid-19 Outcomes in Subpopulations According to Characteristics at Baseline.*

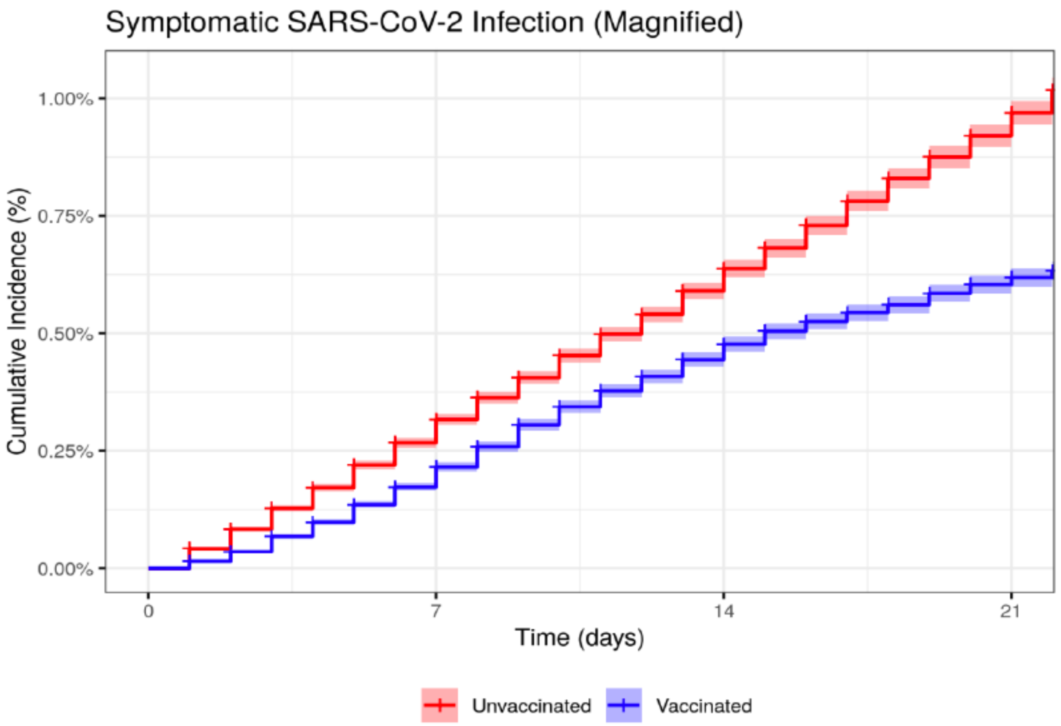
Characteristic and Period	Documented Infection		Symptomatic Illness	
	1–RR	Risk Difference	1–RR	Risk Difference
	% (95% CI)	*no./1000 persons (95% CI)*	*% (95% CI)*	*no./1000 persons (95% CI)*
Age, ≥70 yr				
14 to 20 days after first dose	22 (−9 to 44)	0.81 (−0.28 to 1.89)	44 (19 to 64)	1.36 (0.48 to 2.36)
21 to 27 days after first dose	50 (19 to 72)	1.40 (0.42 to 2.35)	64 (37 to 83)	1.35 (0.62 to 2.22)
7 days after second dose to end of follow-up	95 (87 to 100)	6.10 (3.43 to 9.61)	98 (90 to 100)	4.77 (2.14 to 7.70)
No coexisting conditions				
14 to 20 days after first dose	49 (42 to 56)	2.13 (1.69 to 2.59)	55 (45 to 63)	1.32 (0.98 to 1.67)
21 to 27 days after first dose	66 (58 to 73)	2.49 (1.99 to 2.98)	73 (62 to 82)	1.27 (0.92 to 1.64)
7 days after second dose to end of follow-up	91 (83 to 96)	7.67 (4.90 to 11.07)	93 (78 to 100)	3.54 (1.79 to 5.90)
One or two coexisting conditions				
14 to 20 days after first dose	43 (32 to 53)	2.05 (1.41 to 2.73)	57 (45 to 66)	1.74 (1.25 to 2.24)
21 to 27 days after first dose	56 (45 to 65)	2.43 (1.77 to 3.16)	62 (47 to 73)	1.56 (1.05 to 2.06)
7 days after second dose to end of follow-up	95 (88 to 98)	10.53 (6.73 to 14.40)	95 (88 to 100)	6.21 (3.82 to 8.95)
Three or more coexisting conditions				
14 to 20 days after first dose	37 (12 to 55)	1.60 (0.43 to 2.76)	62 (43 to 77)	2.19 (1.20 to 3.18)
21 to 27 days after first dose	37 (−1 to 62)	1.03 (−0.03 to 2.02)	47 (11 to 73)	0.97 (0.16 to 1.86)
7 days after second dose to end of follow-up	86 (72 to 95)	5.83 (3.16 to 9.03)	89 (68 to 98)	3.97 (1.41 to 6.68)

N Dagan, N Barda et al. *NEJM* 2021

Data depth allowed detailed monitoring of confounding (enabled by RCT)



Full matching



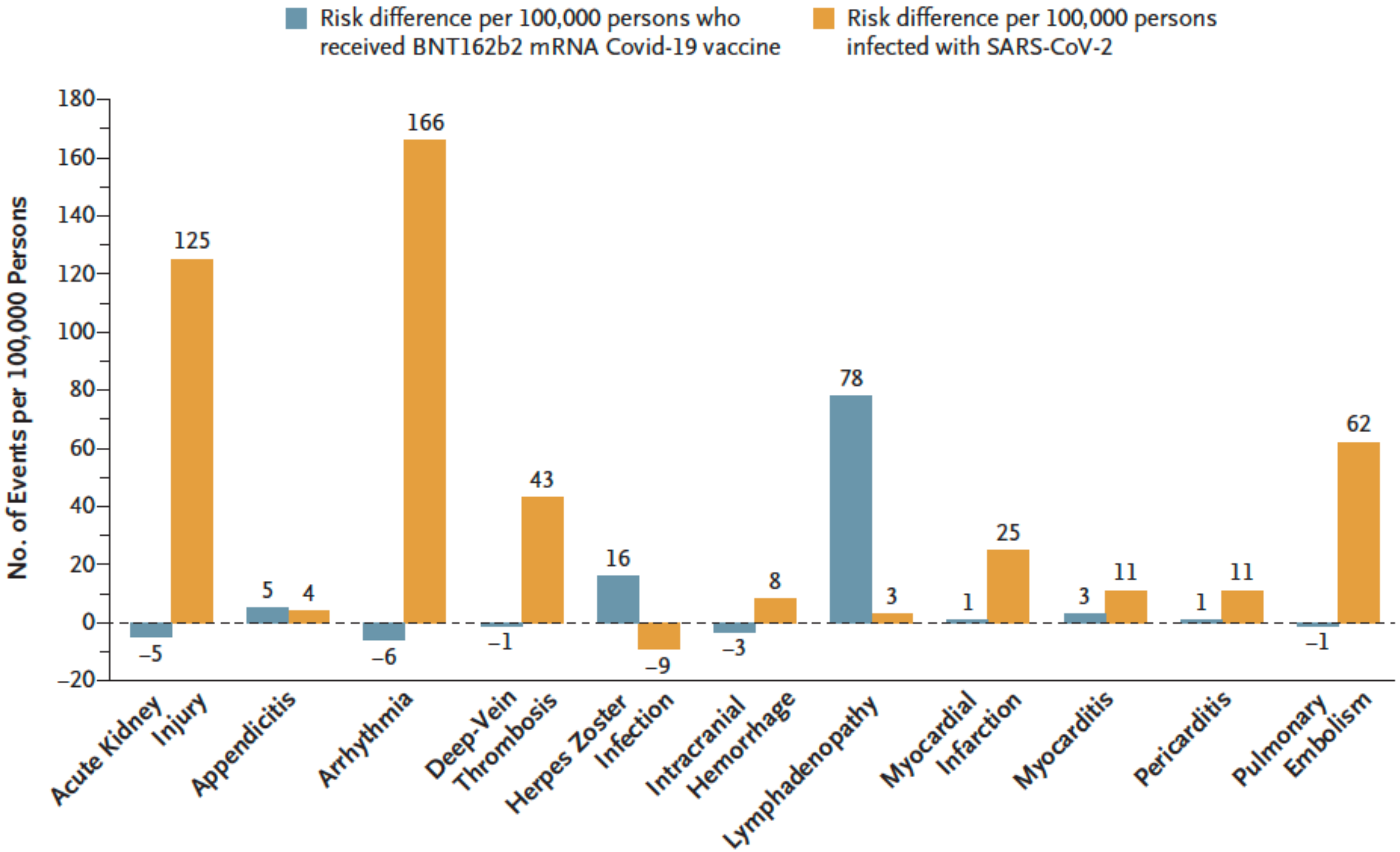
Matched only on age and sex

Vaccine safety and infection sequelae

- Matching as in VE study separately for two exposures: vaccination or documented infection
- Tracked outcomes following vaccination (vs. no vaccination) AND following documented COVID-19 (vs. no documented COVID-19)
- Not strictly comparable but illustrates the strength of the study platform and can give rough comparison

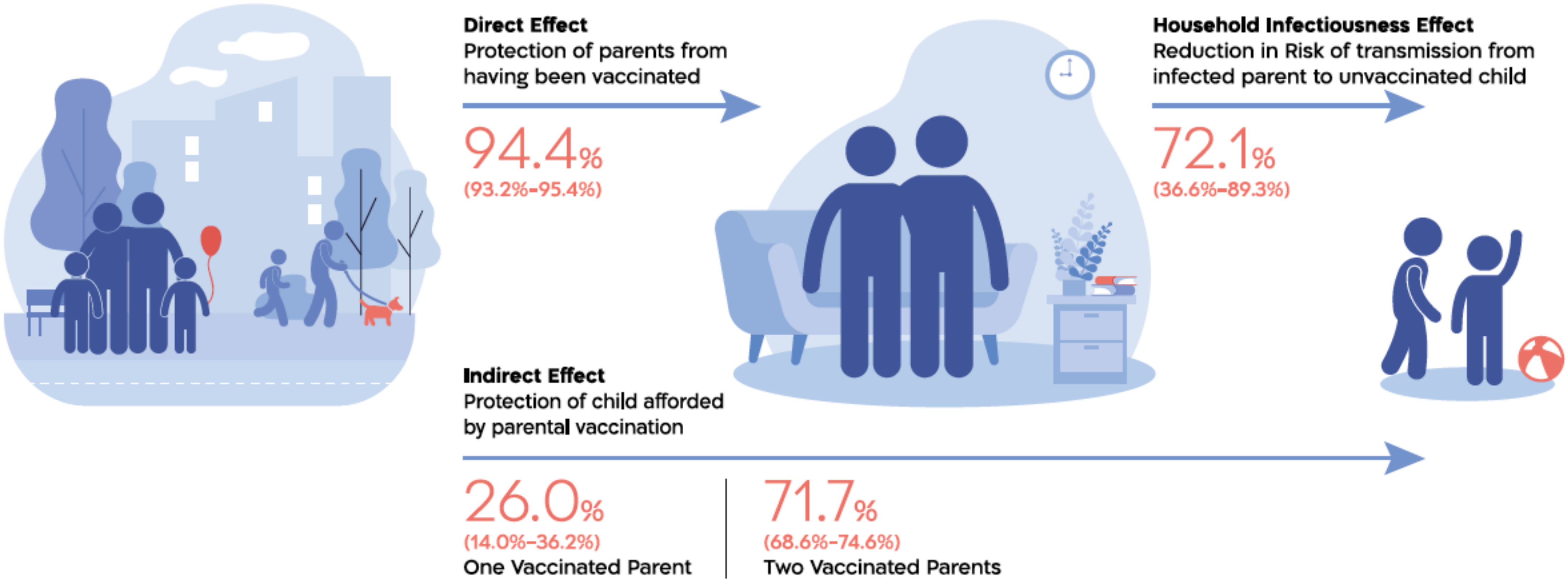
Barda et al. NEJM 2021

Vaccine safety and infection sequelae



Barda et al. NEJM 2021

Indirect effects: Parent vaccination effect on children



S Hayek et al. *Science* 2022

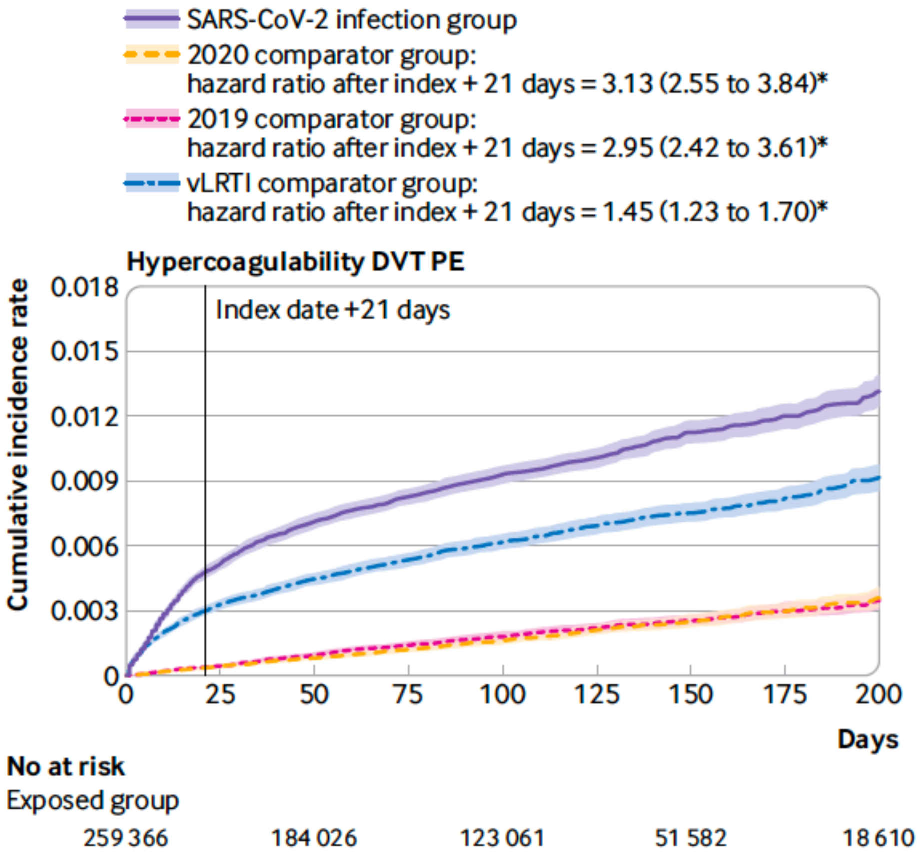
2. Sequelae with United Health

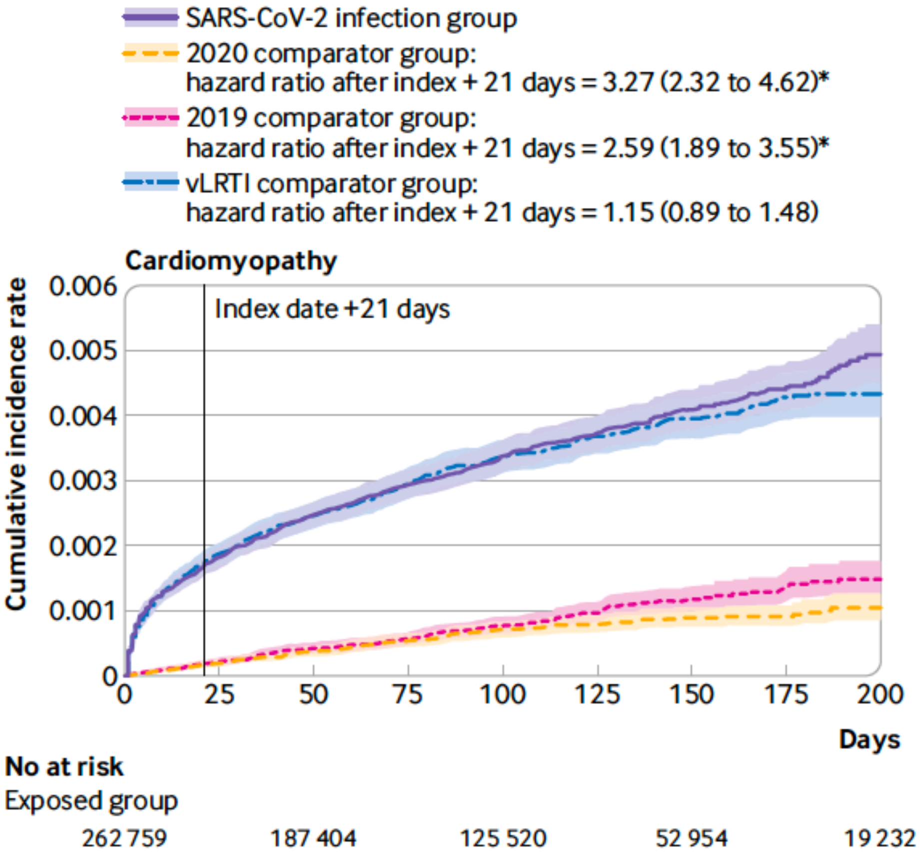
Sarah Daugherty
Ken Cohen
Kevin Heath
Yinglong Guo
Jirapat Samranvedhya
Karol Giuseppe Giubilo
Micah Desmariñas
Sheng Ren

Approach

- UHG Clinical Discovery Database: Claims, pharmacy, and hospitalization data
- Propensity-score matched analysis of individuals 18-65 diagnosed with COVID-19 1/1/2020 to 10/31/2020 vs
  - Contemporaneous comparators
  - Year-ago (2019) comparators
  - “Viral LRTI” comparators
- Counted new events for 120 days starting 21 days post-diagnosis
- Approximately 260K exposed individuals (different for each outcome)
- Kaplan-Meier risk difference and hazard ratio
- Atopic dermatitis as negative control

Different comparators



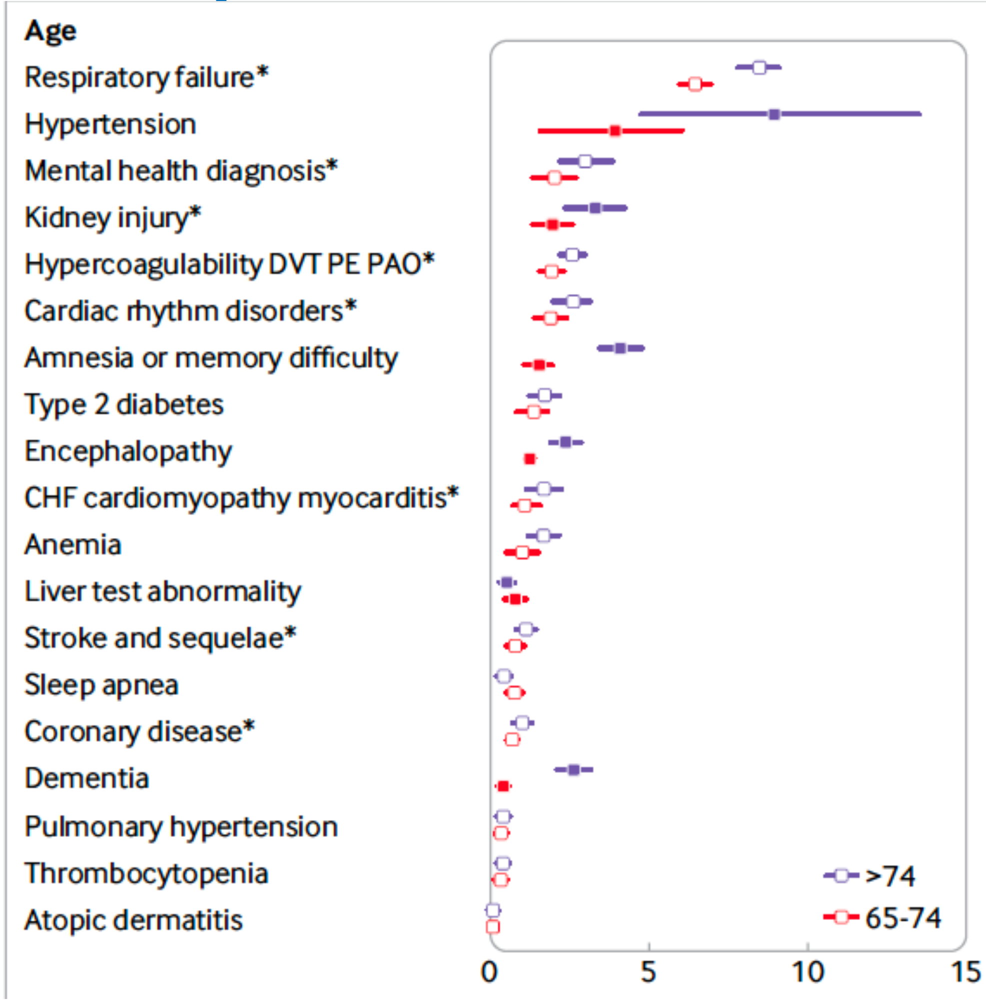


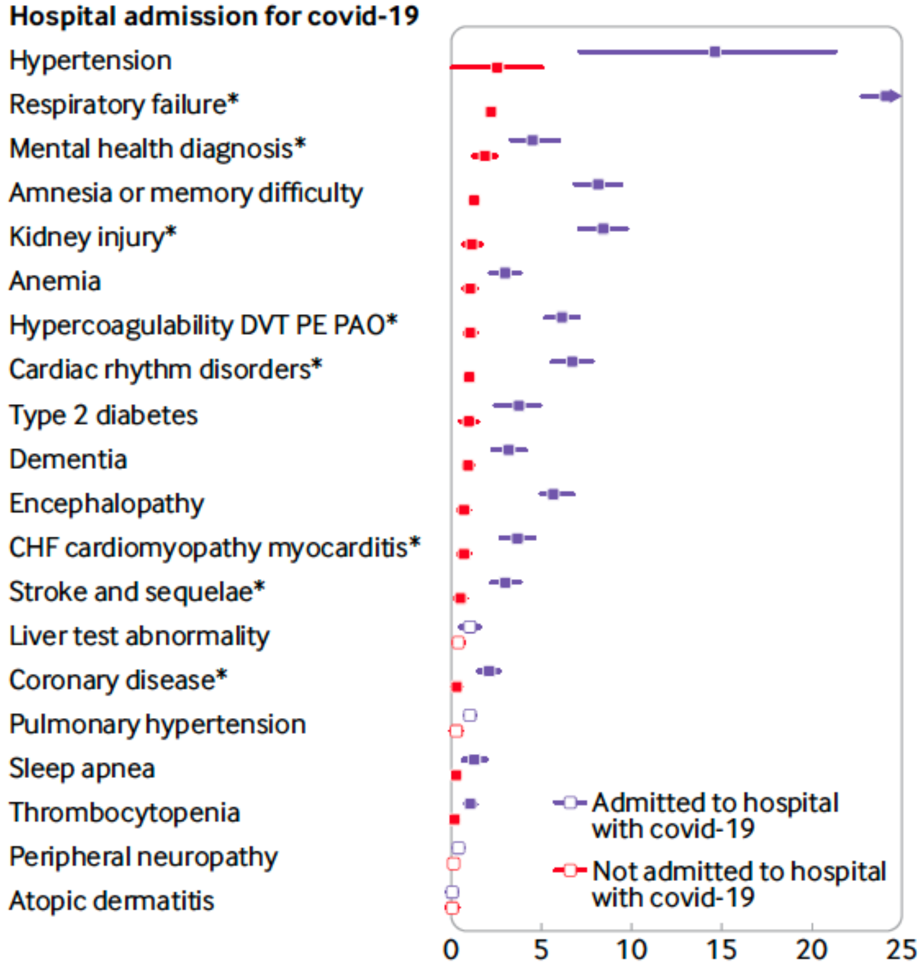
Medicare analysis

- Similar approach except counted *new and persistent postacute sequelae*: i.e. any diagnosis not present 14d prior to index date that occurred 21+ days after index date, regardless of whether it had appeared during the acute phase
- 132,847 individuals >65 with SARS-CoV-2 diagnosis during 2020 for primary (2020 unexposed) comparison

Cohen et al. *BMJ* 2022

Risk differences by age and hospitalization





K Cohen et al. *BMJ* 2022

2. Omicron vs. Delta Severity with KP of Southern CA

Joe Lewnard (UC Berkeley)
Sara Tartof (KPSC)
Manish Patel (CDC)
Rebecca Kahn (CDC)

Identified severity as a key unknown in late Nov.

Contacted Joe Lewnard 3 Dec.

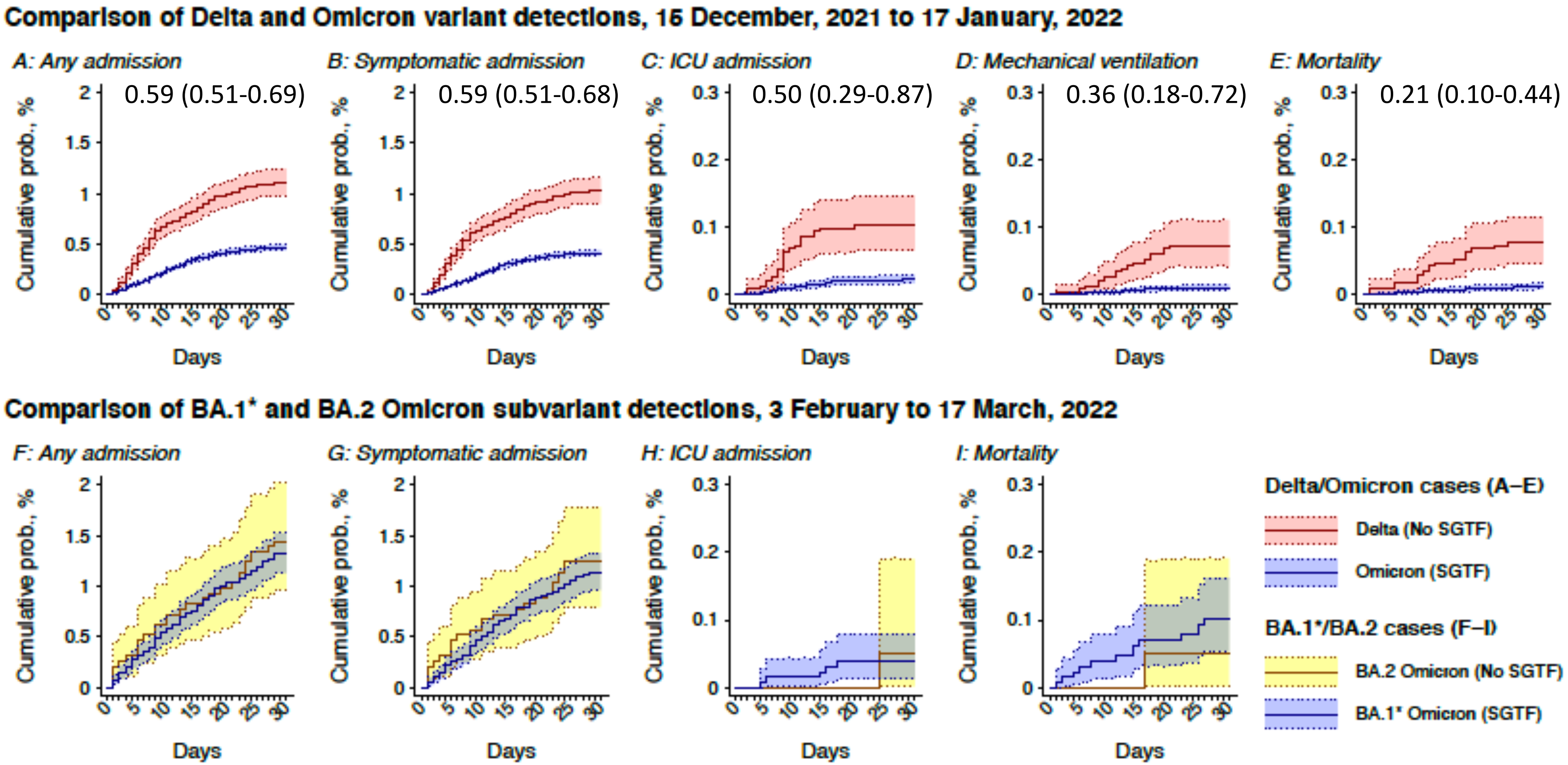
Designed analysis over the next 2 weeks.

First tables 3 Jan. Manuscript submission, White House press conference mention, and preprint 11 Jan.

Approach

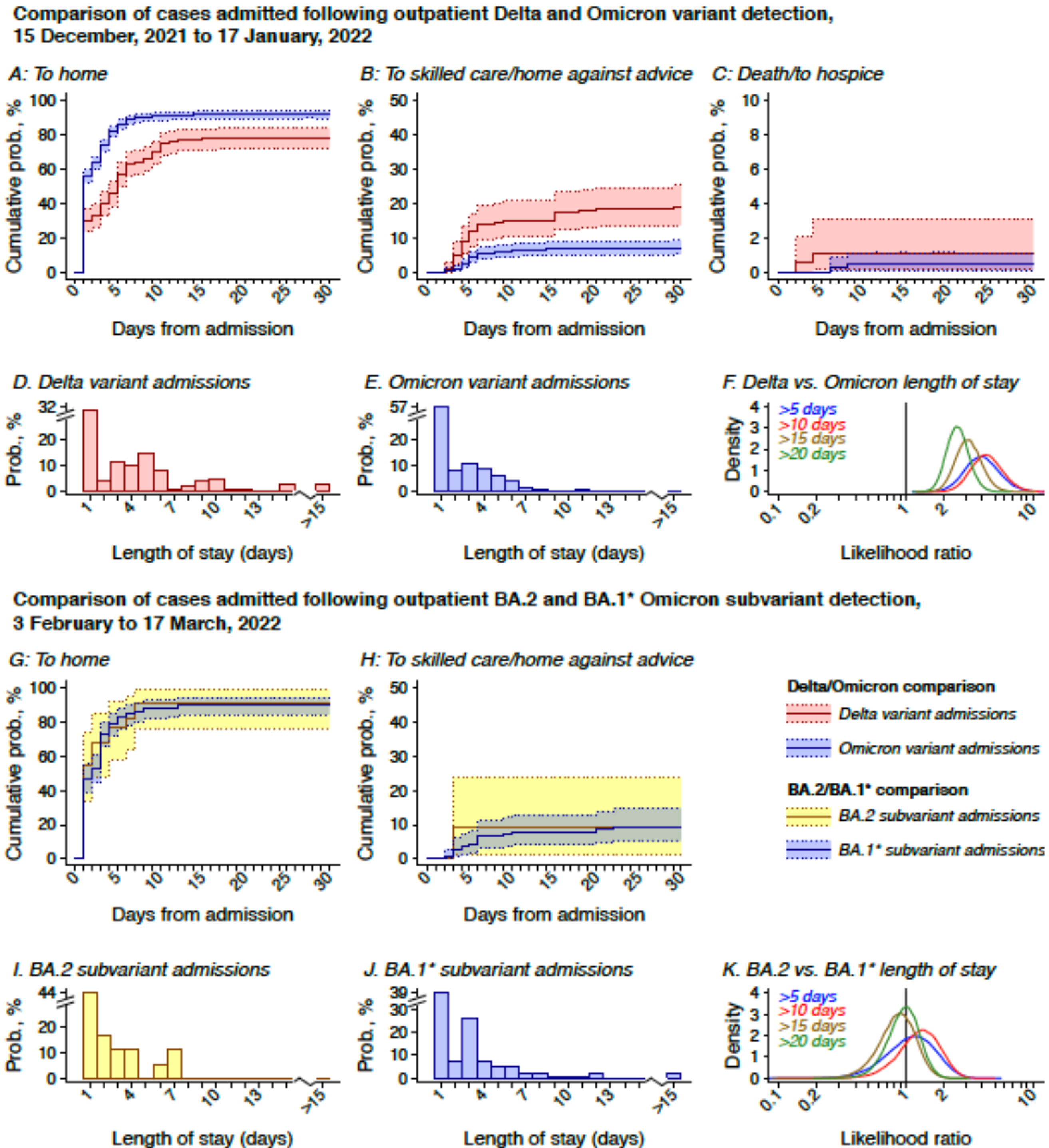
- Data comprise cases diagnosed based on tests processed using ThermoFisher TaqPath COVID-19 Combo Kit devices (identify S gene target failure)
- Compared time to severe clinical endpoints among patients **first ascertained via outpatient testing**
  - Delta vs. BA1/1.1 (SGT+ vs. SGTF). 15 Dec 2021- 17 Jan 2022
  - BA1/1.1 vs. BA.2 (SGTF vs. SGT+. 3 Feb-17 March 2022

BA.1 less severe than Delta; BA.2 similar to BA.1



J Lewnard et al. *Nat Med* 2022

Length of stay among outpt-diagnosed hospitalized cases



J Lewnard et al. *Nat Med* 2022

Serendipity

- Early BA.1 in Southern California
- Payer-provider with deep electronic health records and linkage to state vaccine registry
- Excellent analysts
- SGTF visible on their machine (TaqPath)
- SGTF distinguished BA.1 from Delta
- Existing CDC contract to build on

Serendipity is not a strategy. Need to harden these capacities, including integrating sequencing much more into clinical data, to prepare for future variants and future pandemics

Grad and Lipsitch *Genome Biology* 2014, **15**:538
http://genomebiology.com/2014/15/11/538

Genome Biology

REVIEW

Epidemiologic data and pathogen genome sequences: a powerful synergy for public health

Yonatan H Grad[1,2,3*] and Marc Lipsitch[1,2]

Conclusions part II

- Research data bases in insurers or payer-providers can provide near-real-time evidence on surveillance for severity, vaccine effectiveness, safety, sequelae, etc. Ability to emulate RCT's in observational data, perform subgroup analyses, and gain scale are all major advantages of these data sources
- COVID shows that these are really surveillance questions, and we should transition from a research approach (high activation energy) to a surveillance approach (constantly updating)
- The high quality of these data sets outweighs their moderate size and imperfect representativeness for many questions
- Need to link clinical data bases to fast sequencing

Further acknowledgments

Thought partners:
Dylan George
Rebecca Kahn
Dan Jernigan
Caitlin Rivers
Xueting Qiu
Jason Asher
Mauricio Santillana
Miguel Hernán

Funding:
Morris-Singer Fund
NCI Seronet