








Science in the Age of Covid19 Part 2

Kaiser Fung

May 2021

Kaiser Fung @JunkCharts

The Research Zoo

Lab Experiments	Statistical Experiments	Observational Studies	Models / Simulation	Exploratory Analyses	Surveys
Vaccine antibodies	Vaccine Trials	Real-world studies of vaccines	Forecasts	Vaccine dose intervals	Vaccine hesitancy
Masks	Therapeutics Trials	Sewage	R0	Vaccine efficacy for subgroups	Attitudes & behavior
Variants		Parking lots	Interventions e.g. campus re-opening, J&J pause	Vaccine efficacy for deaths	
Prevalence			Herd immunity		
			Predicting infections		

Kaiser Fung @JunkCharts

Statistical Experiments

aka Randomized Clinical Trials (RCTs)

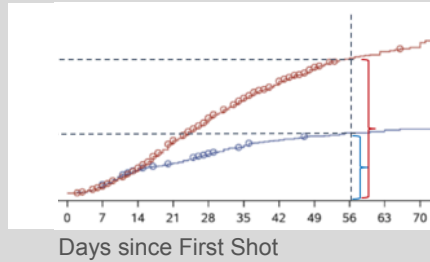


Covid-19 Vaccine Trials

Random Sample	~ 30-40,000 volunteers; due to low case rate, ~150 total cases for preliminary analysis Sample not truly representative of the population
Control	Yes
Randomization	Yes
Placebo effect	Yes (saline)
Double-blind	No
Pre-specified	Yes - trial protocols made public

Vaccine Trials Pre-specify Outcome Metric

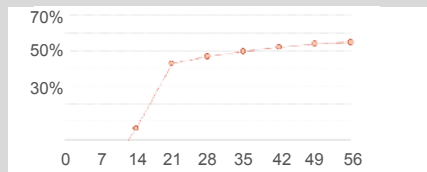
Case Rate
(cumulative)



Placebo

Vaccine

Vaccine
Efficacy
(VE)



VE = 1 - Relative ratio of case rates

Typically, follow-up runs for 1-2 years.
Emergency use: < 2 months.

Source: Johnson & Johnson vaccine FDA Briefing Document, 2021; my analysis



The Gold Standard for Causal Inference

Randomization



Co-variate
Balance

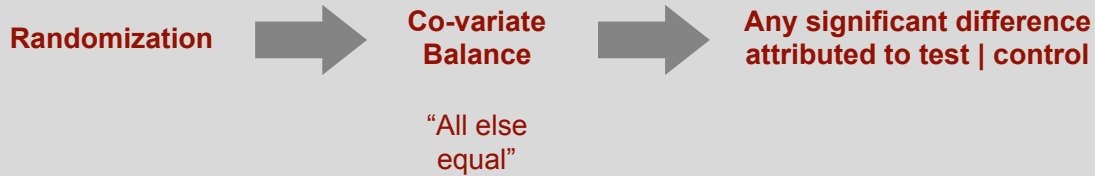
“All else
equal”

Subgroup	Ad26.COV2.S	Placebo
Per-protocol set	19630	19691
Age group (years)		
18-59	12830 (65.4%)	12881 (65.4%)
≥60	6800 (34.6%)	6810 (34.6%)
≥65	3984 (20.3%)	4018 (20.4%)
≥75	755 (3.8%)	693 (3.5%)
Sex		
Female	8702 (44.3%)	8777 (44.6%)
Male	10924 (55.6%)	10910 (55.4%)
Undifferentiated	2 (<0.1%)	4 (<0.1%)
Unknown	2 (<0.1%)	0
Race		
American Indian or Alaska Native	1643 (8.4%)	1628 (8.3%)
Asian	720 (3.7%)	663 (3.4%)
Black or African American	3374 (17.2%)	3390 (17.2%)
Native Hawaiian or other Pacific Islander	54 (0.3%)	45 (0.2%)
White	12200 (62.1%)	12216 (62.0%)
Multiple	1036 (5.3%)	1087 (5.5%)
Unknown	603 (3.1%)	662 (3.4%)

Source: Johnson & Johnson vaccine FDA Briefing Document, 2021



The Gold Standard for Causal Inference



Subgroup	Ad26.COVS.S	Placebo
Per-protocol set	19630	19691
Age group (years)		
18-59	12830 (65.4%)	12881 (65.4%)
≥60	6800 (34.6%)	6810 (34.6%)
≥65	3984 (20.3%)	4018 (20.4%)
≥75	755 (3.8%)	693 (3.5%)
Sex		
Female	8702 (44.3%)	8777 (44.6%)
Male	10924 (55.6%)	10910 (55.4%)
Undifferentiated	2 (<0.1%)	4 (<0.1%)
Unknown	2 (<0.1%)	0

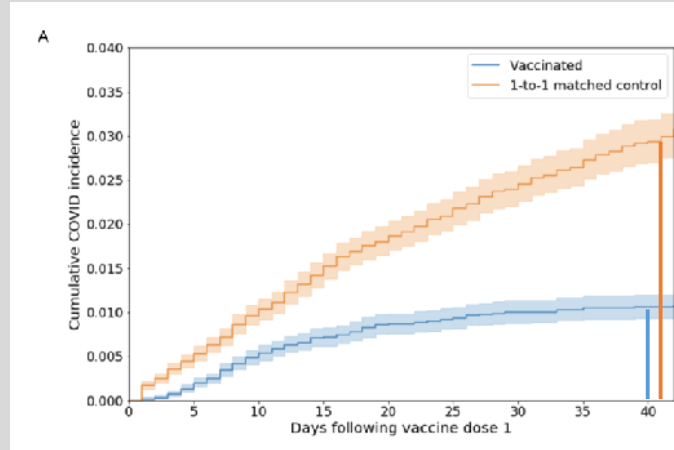
Real-world Study of Vaccine Effectiveness

An observational study, not a statistical experiment

Random Sample	Self-selected sample (“observed”), NOT random Conditioned by vaccination strategy, access, etc.
Control	“Unvaccinated”, also self-selected
Randomization	No
Placebo effect	No
Double-blind	No
Pre-specified	Possible, usually not

Real-world Studies Produce a Similar Outcome Metric

But is it the same metric as from a clinical trial?



Unvaccinated 

Vaccinated 

Source: Mayo Clinic/inference study. Pawlowski, et. al. 2021.

Observational Data Lack Covariate Balance

Not "all else equal"

TABLE 1. Characteristics of health care personnel, first responders, and other essential and frontline workers with reverse transcription-polymerase chain reaction (RT-PCR)-confirmed SARS-CoV-2 infections and percentage receiving one or more doses of a messenger RNA (mRNA) COVID-19 vaccine — eight U.S. locations, December 14, 2020–March 13, 2021

Characteristic	No. (column %)	SARS-CoV-2 infection		Vaccinated with at least one*	
		No. (row %)	p-value**	No. (row %)	p-value**
Total	3,950 (100)	205 (5.2)	—	989 (25.0)	—
Cohort location					
Phoenix, Arizona	555 (14.1)	39 (7.0)	<0.001	147 (26.3)	408 (73.6)
Tucson, Arizona	1,199 (30.4)	79 (6.6)		325 (27.1)	874 (72.9)
Other Arizona	320 (8.1)	16 (5.0)		69 (21.5)	251 (78.5)
Miami, Florida	221 (5.6)	19 (8.6)		119 (53.4)	103 (46.6)
Duluth, Minnesota	445 (11.3)	12 (2.7)		47 (10.6)	400 (90.5)
Portland, Oregon	469 (11.8)	4 (0.8)		61 (13.2)	407 (86.7)
Tempe, Texas	289 (7.3)	18 (6.2)		71 (24.5)	218 (75.4)
Salt Lake City, Utah	450 (11.4)	18 (4.0)		132 (29.3)	318 (70.7)
Sex					
Female**	2,654 (67.2)	164 (6.2)	0.009	529 (21.5)	3,124 (78.4)
Male	1,497 (37.8)	95 (6.4)		460 (30.7)	1,037 (69.3)
Age group, yrs					
18–49	2,859 (72.4)	146 (5.1)	0.55	755 (26.4)	2,104 (73.6)
≥50	1,111 (28.1)	59 (5.3)		234 (21.2)	857 (77.1)
Race					
White	3,408 (86.3)	178 (5.2)	0.02	616 (23.2)	2,794 (76.8)
Other	542 (13.7)	27 (5.0)		175 (32.3)	367 (67.7)
Ethnicity					
Hispanic/Latino	674 (17.1)	37 (5.5)	<0.001	235 (35.2)	438 (64.8)
Other	3,276 (82.9)	144 (4.4)		754 (23.0)	2,522 (77.0)
Occupation**					
Primary health care personnel	835 (21.1)	16 (1.9)	<0.001	55 (7.3)	779 (92.7)
Other health care personnel	3,385 (85.9)	47 (1.4)		242 (7.1)	1,999 (59.1)
First responder	852 (21.6)	73 (8.6)		369 (43.3)	483 (56.7)
Other essential and frontline worker	926 (23.5)	47 (5.1)		374 (40.3)	552 (59.7)
Chronic condition					
None††	2,723 (68.9)	141 (5.2)	0.02	711 (26.1)	2,012 (73.9)
≥1	1,227 (31.1)	64 (5.2)		278 (22.7)	949 (77.3)

Source: MMWR, CDC, April 2, 2021.

- HIGHER RISK**
 - Men**
 - Hispanic**
 - Non-white**
 - > 50**
 - Chronic condition**
 - First responder, Frontline workers**
 - Florida, Arizona, Texas**
- LOWER RISK**
 - Women**
 - Non-Hispanic**
 - White**
 - 18 - 49**
 - None**
 - Healthcare workers**
 - Minnesota, Oregon**


Observational Data Lack Covariate Balance

Not "all else equal"

TABLE 1. Characteristics of health care personnel, first responders, and other essential and frontline workers with reverse transcription-polymerase chain reaction (RT-PCR)-confirmed SARS-CoV-2 infections and percentage receiving one or more doses of a messenger RNA (mRNA) COVID-19 vaccine — eight U.S. locations, December 14, 2020–March 13, 2021

Characteristic	No. (column %) of participants	SARS-CoV-2 infection		Unvaccinated		Vaccinated with ≥1 dose*	
		No. (row %)	p-value†	No. (row %)	No. (row %)	p-value†	
Total	3,950 (100)	205 (5.2)	—	989 (25.0)	2,961 (75.0)	—	
Cohort location							
Phoenix, Arizona	555 (14.1)	39 (7.0)	<0.001	147 (26.3)	408 (73.7)	<0.001	
Tucson, Arizona	1,199 (30.4)	79 (6.6)		525 (43.7)	674 (56.3)		
Other Arizona	320 (8.1)	16 (5.0)		69 (21.5)	251 (78.5)		
Miami, Florida	221 (5.6)	19 (8.6)		119 (53.4)	102 (46.6)		
Duluth, Minnesota	445 (11.3)	12 (2.7)		47 (10.5)	400 (90.5)		
Portland, Oregon	469 (11.8)	4 (0.8)		61 (13.2)	407 (86.7)		
Tempe, Texas	289 (7.3)	18 (6.2)		71 (24.5)	218 (75.4)		
Salt Lake City, Utah	490 (11.4)	18 (4.0)		132 (26.9)	358 (73.0)		
Sex							
Female**	2,954 (74.8)	164 (4.4)	0.08†	529 (17.9)	2,424 (82.0)	<0.001	
Male	1,497 (37.9)	96 (6.4)		460 (30.7)	1,037 (69.3)		
Age group, yrs							
18–49	2,859 (72.4)	146 (5.1)	0.55	755 (26.4)	2,104 (73.6)	0.48	
≥50	1,111 (28.1)	59 (5.3)		234 (21.1)	877 (78.9)		
Race							
White	3,408 (86.3)	178 (5.2)	0.22	814 (23.9)	2,594 (76.1)	<0.001	
Other	542 (13.7)	27 (5.0)		175 (32.3)	367 (67.7)		
Ethnicity							
Hispanic/Latino	674 (17.1)	57 (8.5)	<0.001	235 (34.9)	438 (65.1)	<0.001	
Other	3,276 (82.9)	148 (4.5)		754 (23.0)	2,522 (77.0)		
Occupation**							
Primary health care personnel	832 (21.1)	16 (1.9)	<0.001	55 (7.3)	777 (92.7)	<0.001	
Other health care personnel	3,335 (83.9)	47 (1.4)		242 (7.2)	3,093 (92.8)		
First responder	852 (21.6)	73 (8.6)		369 (43.3)	483 (56.7)		
Other essential and frontline worker	926 (23.5)	47 (5.1)		324 (35.0)	599 (65.0)		
Chronic condition							
None††	2,723 (68.9)	141 (5.2)	0.92	711 (26.1)	2,012 (73.9)	0.11	
≥1	1,227 (31.1)	64 (5.2)		278 (22.7)	949 (77.3)		

Source: MMWR, CDC, April 2, 2021.

- HIGHER RISK**
- LOWER RISK**
- Men
- Women 
- Hispanic
- Non-Hispanic
- Non-white
- White
- > 50
- 18 - 49
- Chronic condition
- None
- First responder, Frontline workers
- Healthcare workers
- Florida, Arizona, Texas
- Minnesota, Oregon





Observational Data Lack Covariate Balance

Not "all else equal"

TABLE 1. Characteristics of health care personnel, first responders, and other essential and frontline workers with reverse transcription-polymerase chain reaction (RT-PCR)-confirmed SARS-CoV-2 infections and percentage receiving one or more doses of a messenger RNA (mRNA) COVID-19 vaccine — eight U.S. locations, December 14, 2020–March 13, 2021

Characteristic	No. (column %) of participants	SARS-CoV-2 infection		Unvaccinated		Vaccinated with ≥1 dose*	
		No. (row %)	p-value†	No. (row %)	No. (row %)	p-value†	
Total	3,950 (100)	205 (5.2)	—	989 (25.0)	2,961 (75.0)	—	
Cohort location							
Phoenix, Arizona	555 (14.1)	39 (7.0)	<0.001	147 (26.3)	408 (73.7)	<0.001	
Tucson, Arizona	1,199 (30.4)	79 (6.6)		525 (43.7)	674 (56.3)		
Other Arizona	320 (8.1)	16 (5.0)		69 (21.5)	251 (78.5)		
Miami, Florida	221 (5.6)	19 (8.6)		119 (53.4)	102 (46.6)		
Duluth, Minnesota	445 (11.3)	12 (2.7)		47 (10.5)	400 (90.5)		
Portland, Oregon	469 (11.8)	4 (0.8)		61 (13.2)	407 (86.7)		
Tempe, Texas	289 (7.3)	18 (6.2)		71 (24.5)	218 (75.4)		
Salt Lake City, Utah	490 (11.4)	18 (4.0)		132 (26.9)	358 (73.0)		
Sex							
Female**	2,954 (74.8)	164 (4.4)	0.08†	529 (17.9)	2,424 (82.0)	<0.001	
Male	1,497 (37.9)	96 (6.4)		460 (30.7)	1,037 (69.3)		
Age group, yrs							
18–49	2,859 (72.4)	146 (5.1)	0.55	755 (26.4)	2,104 (73.6)	0.48	
≥50	1,111 (28.1)	59 (5.3)		234 (21.1)	877 (78.9)		
Race							
White	3,408 (86.3)	178 (5.2)	0.22	814 (23.9)	2,594 (76.1)	<0.001	
Other	542 (13.7)	27 (5.0)		175 (32.3)	367 (67.7)		
Ethnicity							
Hispanic/Latino	674 (17.1)	57 (8.5)	<0.001	235 (34.9)	438 (65.1)	<0.001	
Other	3,276 (82.9)	148 (4.5)		754 (23.0)	2,522 (77.0)		
Occupation**							
Primary health care personnel	832 (21.1)	16 (1.9)	<0.001	55 (7.3)	777 (92.7)	<0.001	
Other health care personnel	3,335 (83.9)	47 (1.4)		242 (7.2)	3,093 (92.8)		
First responder	852 (21.6)	73 (8.6)		369 (43.3)	483 (56.7)		
Other essential and frontline worker	926 (23.5)	47 (5.1)		324 (35.0)	599 (65.0)		
Chronic condition							
None††	2,723 (68.9)	141 (5.2)	0.92	711 (26.1)	2,012 (73.9)	0.11	
≥1	1,227 (31.1)	64 (5.2)		278 (22.7)	949 (77.3)		

Source: MMWR, CDC, April 2, 2021.

- HIGHER RISK**
- LOWER RISK**
- Men
- Women 
- Hispanic
- Non-Hispanic 
- Non-white
- White
- > 50
- 18 - 49
- Chronic condition
- None
- First responder, Frontline workers
- Healthcare workers
- Florida, Arizona, Texas
- Minnesota, Oregon



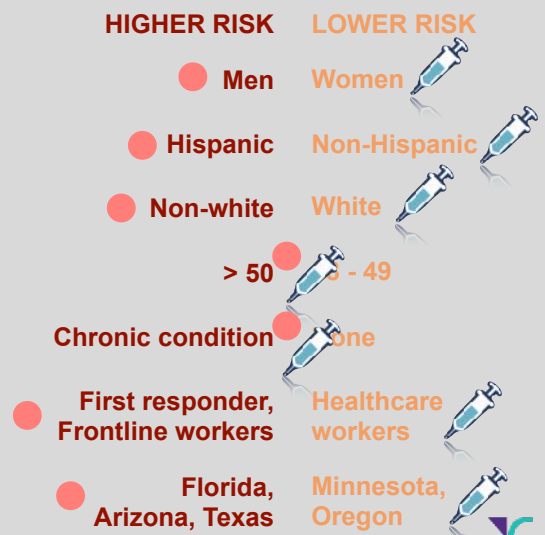
Observational Data Lack Covariate Balance

Not “all else equal”

TABLE 1. Characteristics of health care personnel, first responders, and other essential and frontline workers with reverse transcription-polymerase chain reaction (RT-PCR)-confirmed SARS-CoV-2 infections and percentage receiving one or more doses of a messenger RNA (mRNA) COVID-19 vaccine — eight U.S. locations, December 14, 2020–March 13, 2021

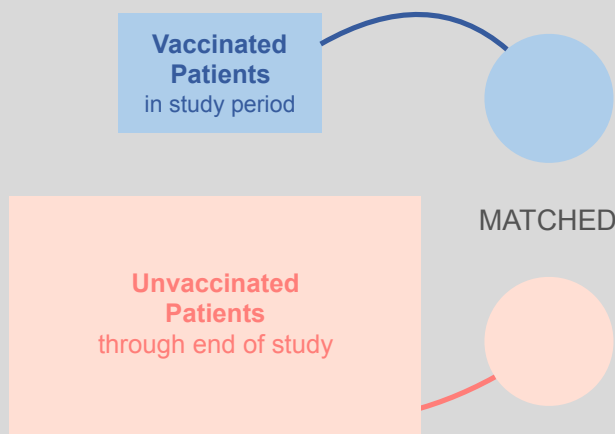
Characteristic	No. (column %) of participants	SARS-CoV-2 Infection		Unvaccinated		Vaccinated with ≥1 dose*	
		No. (row %)	p-value**	No. (row %)	No. (row %)	p-value**	
Total	3,950 (100)	205 (5.2)	—	989 (25.0)	2,961 (75.0)	—	
Cohort location							
Phoenix, Arizona	555 (14.1)	39 (7.0)	<.001	147 (26.3)	408 (73.6)	<.001	
Tucson, Arizona	1,199 (30.4)	79 (6.6)		525 (43.7)	674 (56.3)		
Other Arizona	320 (8.1)	16 (5.0)		65 (20.3)	255 (79.7)		
Miami, Florida	221 (5.6)	19 (8.6)		119 (53.4)	102 (46.6)		
Duluth, Minnesota	445 (11.3)	12 (2.7)		47 (10.6)	400 (90.4)		
Portland, Oregon	469 (11.8)	4 (0.9)		61 (13.0)	407 (87.0)		
Tempe, Texas	289 (7.3)	18 (6.2)		71 (24.6)	218 (75.4)		
Salt Lake City, Utah	450 (11.4)	18 (4.0)		132 (29.3)	318 (70.7)		
Sex							
Female**	2,951 (74.7)	164 (4.4)	0.087	529 (17.9)	2,422 (82.1)	<.001	
Male	1,497 (37.9)	95 (6.4)		460 (30.7)	1,037 (69.3)		
Age group, yrs							
15–49	2,859 (72.4)	146 (5.1)	0.85	755 (26.4)	2,104 (73.6)	0.48	
≥50	1,111 (28.1)	59 (5.3)		234 (21.1)	877 (78.9)		
Race							
White	3,408 (86.3)	178 (5.2)	0.22	816 (23.9)	2,592 (76.1)	<.001	
Other	542 (13.7)	27 (5.0)		173 (31.7)	367 (68.3)		
Ethnicity							
Hispanic/Latino	674 (17.1)	57 (8.5)	<.001	235 (34.9)	439 (65.1)	<.001	
Other	3,276 (82.9)	148 (4.5)		754 (23.0)	2,522 (77.0)		
Occupation**							
Primary health care personnel	832 (21.1)	16 (1.9)	<.001	55 (7.3)	777 (92.7)	<.001	
Other health care personnel	1,385 (35.0)	47 (3.4)		242 (17.5)	1,143 (82.5)		
First responder	852 (21.6)	75 (8.8)		369 (43.3)	483 (56.7)		
Other essential and frontline worker	925 (23.5)	47 (5.1)		324 (35.1)	599 (64.9)		
Chronic condition							
None††	2,723 (68.9)	141 (5.2)	0.82	711 (26.1)	2,012 (73.9)	0.11	
≥1	1,227 (31.1)	64 (5.2)		278 (22.7)	949 (77.3)		

Source: MMWR, CDC, April 2, 2021.



Fixing Covariate Imbalance: Matching

Matching produces a synthetic control group



Study of Mayo Clinic patients

Use demographics to predict probability of getting vaccinated (aka propensity score)

For each vaccinated patient, find an unvaccinated patient with similar propensity score

Matched groups have covariate balance

Fixing Covariate Imbalance: Regression Adjustment

“Removes” effects of other influential factors

Effect of Calendar Time Adjustment In Danish Study

	Unadjusted VE ²		Adjusted VE ³	
	VE	95% CI	VE	95% CI
Nursing home residents				
Unvaccinated	-	-	-	-
0-14 days after 1 st dose	-0.19	-0.44;0.01	-0.40	-0.62;-0.02
> 14 days after 1 st dose until 2 nd dose	0.60	0.46;0.71	0.21	-0.11;0.44
0-7 days after 2 nd dose	0.80	0.70;0.88	0.52	0.27;0.69
> 7 days after 2 nd dose	0.96	0.91;0.98	0.64	0.14;0.84

Source: Danish nursing home study, Moustsen-Helms, et. al., 2021.

Study of Danish nursing home residents

Unadjusted VE = 96%

Adjusted VE = 64%

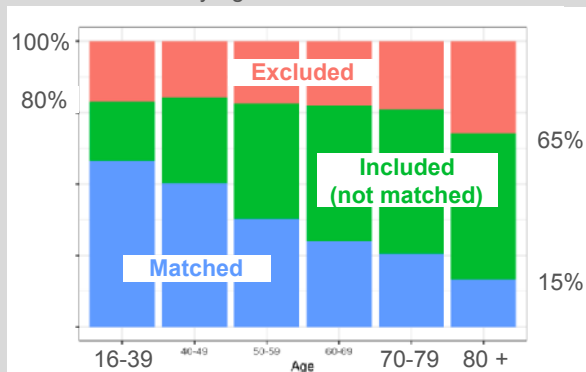
Adjustment to remove effect of rapidly declining infection rate over course of study



Incurable Imbalance

Look out for match rate

Mix of Clalit (insurer) Patients By Age & Match Status



Source: Israel Clalit study, NEJM, Dagan, et. al., 2021.

Study by Israel’s largest health insurer

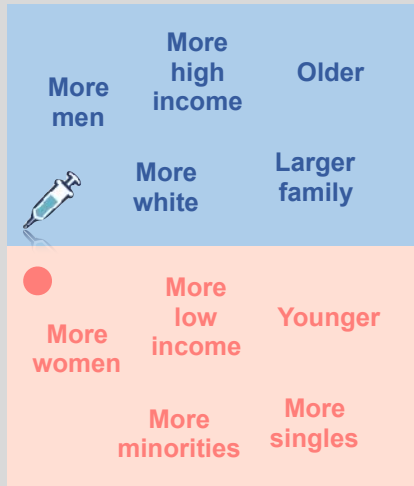
Older patients vaccinated first, and have only 25% match rate

Exact matching is tough

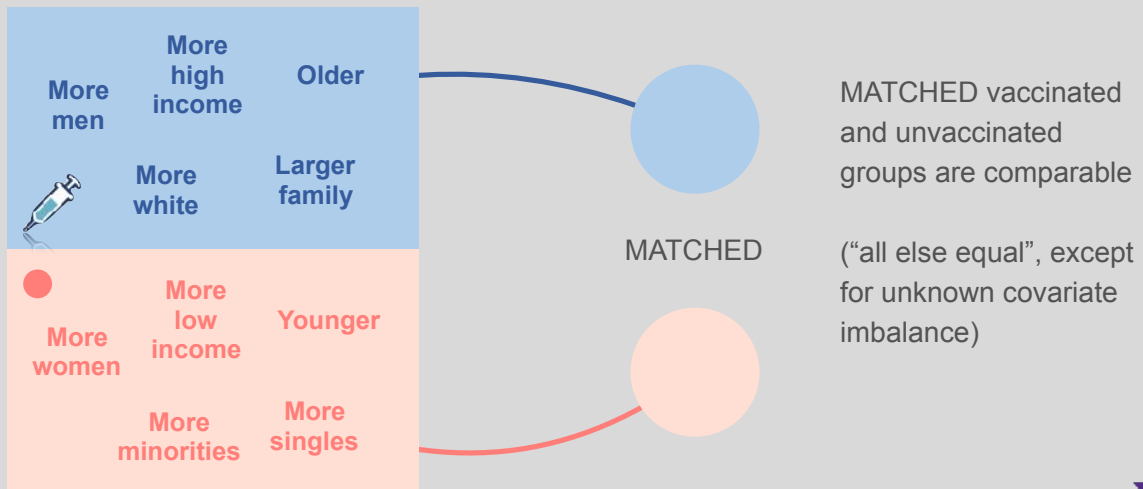
Approximate matching methods (trade-off)



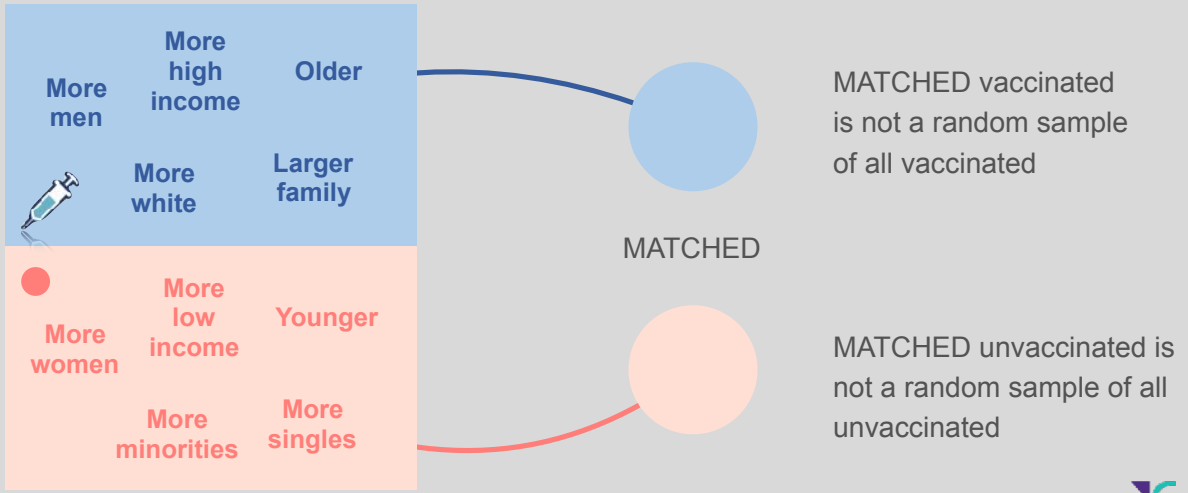
VE for Matched Population



VE for Matched Population



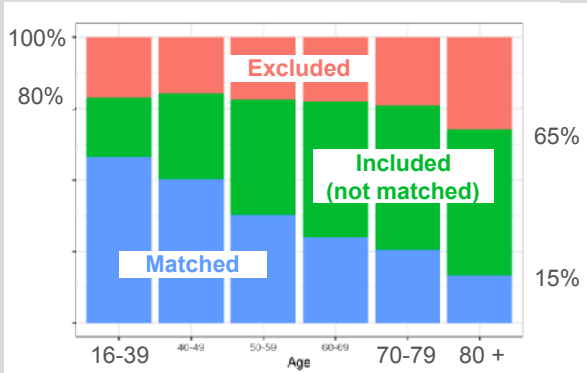
VE for Matched Population = VE for Research Population?



Matched Population Much Younger

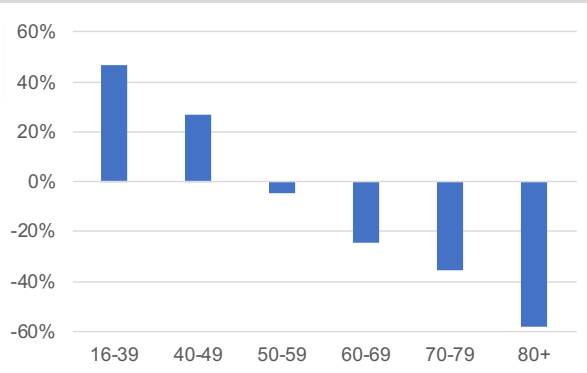
VE of matched population likely over-stated

Mix of Clalit (insurer) Patients
By Age & Match Status



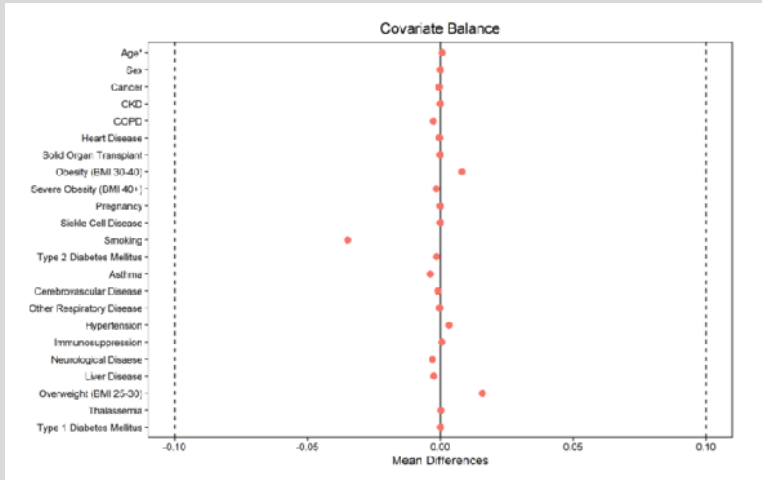
Source: Israel Clalit study, NEJM, Dagan, et. al., 2021.

Over/Under-representation by Age Group
Matched Vaccinated vs. All Vaccinated



Source: Israel Clalit study, NEJM, Dagan, et. al., 2021.
My analysis.

Covariate Balance is Insufficient



Source: Israel Clalit study, NEJM, Dagan, et. al., 2021.

Study by Israel's largest health insurer

Matched groups are comparable

But MATCHED vaccinated far younger than all vaccinated

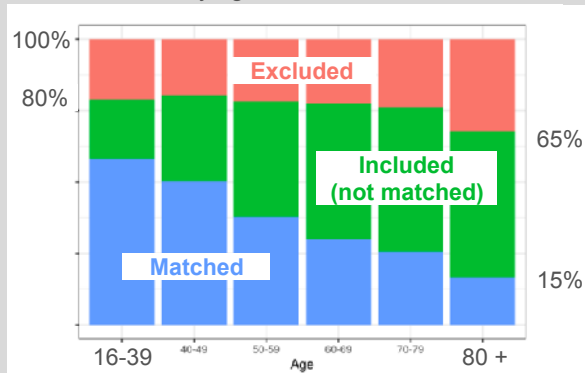
And more...



Population Drift

Every study excludes people up-top

Mix of Clalit (insurer) Patients By Age & Match Status



Source: Israel Clalit study, NEJM, Dagan, et. al., 2021.

Excluded

People living in nursing homes

People confined to their homes

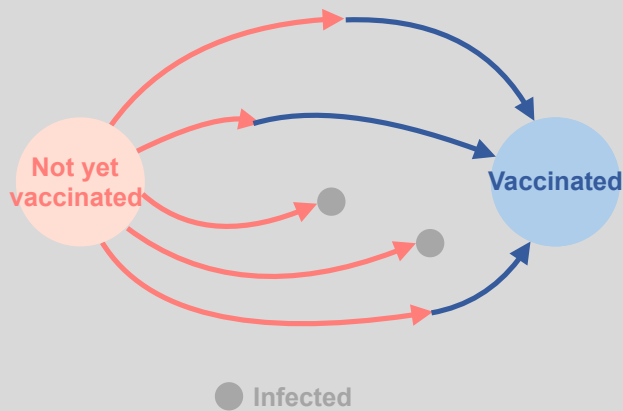
Healthcare workers

People who "had a health care interaction within 3 days before the vaccination date" (13%)



Immortal Time Bias

Biases pop up in complex ways



Each person starts as **unvaccinated**

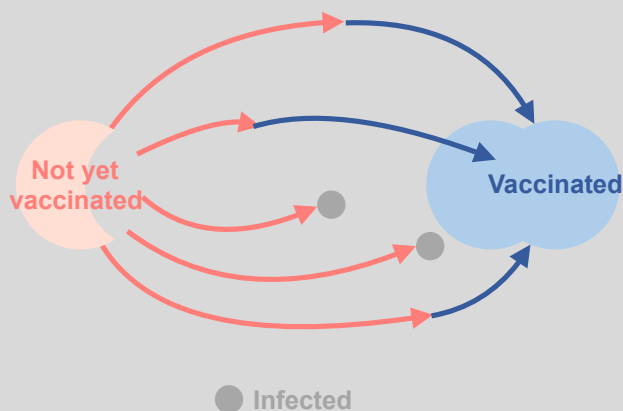
Over time, some/most people transition to **vaccinated**

Any vaccinated person cannot be infected (is “immortal”) prior to vaccination

Unadjusted analysis is biased against unvaccinated

Immortal Time Bias

Biases pop up in complex ways



Each person starts as **unvaccinated**

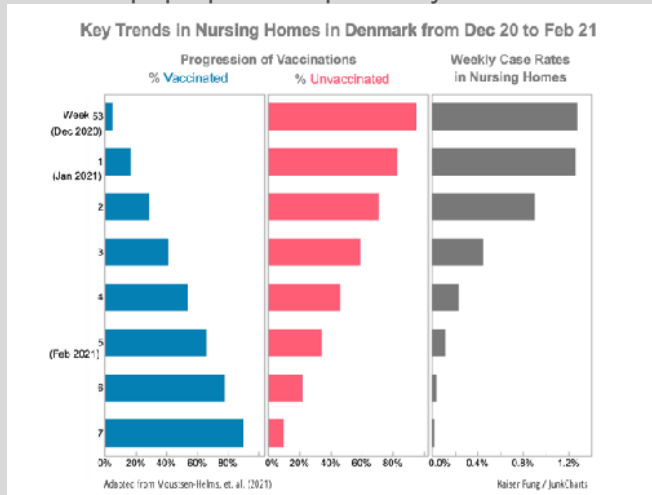
Over time, some/most people transition to **vaccinated**

Any vaccinated person cannot be infected (is “immortal”) prior to vaccination

Unadjusted analysis is biased against unvaccinated

Calendar Time Bias

Biases pop up in complex ways



The vaccination campaign coincides with a period of declining cases (reversal of holiday spike?)

At-risk time for **unvaccinated** is skewed toward the start of study period

Unadjusted analysis is biased against unvaccinated



Kaiser Fung @JunkCharts

Proper Usage of Real-world Studies

Using real-world studies to validate RCTs is like asking a C student to grade the work of an A student.



Kaiser Fung @JunkCharts

Proper Usage of Real-world Studies

Using real-world studies to validate RCTs is like asking a C student to grade the work of an A student.

Use real-world studies to answer open questions:

VE for subgroups (racial groups, age groups, pregnant women, etc.)

VE over a longer time

VE for hospitalizations, deaths



Kaiser Fung @JunkCharts

Kaiser Fung
@junkcharts

Email for consulting, coaching, collab
junkcharts@gmail.com



Kaiser Fung @JunkCharts