Alternative analysis of mpox vaccine effectiveness data from Israel

Christopher B. Boyer, PhD and Marc Lipsitch, DPhil*
Center for Communicable Disease Dynamics,
Harvard T.H. Chan School of Public Health,
677 Huntington Avenue, Boston, MA 02115 USA

*corresponding author

Text including introductory/abstract paragraph: 1,098 words
2 figures
1 supplementary text
1 supplementary table

This work was supported by the Morris-Singer Fund's gift to the Center for Communicable Disease Dynamics.

Data availability statement: all data are obtained from the original published paper on which this comments.

Code availability: R code for all results is available at https://github.com/boyercb/nature-med-mpox-response.

Wolff Sagy et al. report that a single dose of MVA-BN mpox vaccine was 86% effective in preventing incident mpox. In their cohort study, all eligible men were enrolled at the start of a vaccination campaign on July 31, 2022 and followed until they tested positive for mpox or December 25, 2022, whichever was earlier. To avoid the potential for immortal time bias, participants vaccinated after July 31 were considered unvaccinated from July 31 through their date of vaccination and then considered vaccinated from their date of vaccination, which was set as their new Day 0. Incidence of mpox was compared using time-dependent Cox regression on this time scale, with typically different calendar days being Day 0 for vaccinated and unvaccinated persons. In this communication, we show that while their method eliminates immortal time, it does not account for confounding by calendar time which is particularly important during an outbreak where incidence rates vary drastically over time. We evaluate the potential magnitude of this bias through simulation and a re-analysis of initial data reported by Wolff Sagy et al. and conclude with a discussion of how it can be resolved in future studies.

Randomized trials of vaccines are not always possible due to logistical or ethical constraints. Even when trials are possible, they cannot estimate vaccine effectiveness (VE) under all circumstances and they may be invalidated as the disease environment changes, such as when new variants emerge. When trials aren't possible, so-called "real world" estimates of VE use observational data to emulate, either implicitly or explicitly, the randomized trial which cannot be conducted[1,2]. However, emulating a trial from observational data requires careful consideration of both the design and analysis to avoid bias and ensure comparability.

In their article, Wolff Sagy et al. estimate effectiveness of the MVA-BN mpox vaccine using electronic medical records data from a large healthcare system in Israel[3]. In the methods, they identify the potential for *immortal time bias*[4] as the vaccinated must, by definition, survive long enough to receive a vaccine while the unvaccinated are followed from the first calendar date. They then describe a common analytical approach for eliminating immortal time: splitting follow-up for the vaccinated, assigning the person-time prior to vaccination as unvaccinated time, assigning person-time after vaccination as vaccinated time under a reset time zero, and using a time-dependent Cox regression to estimate VE.

However, in this commentary, we aim to show:

- This setup induces a form of confounding by calendar time that can readily arise in infectious disease outbreaks when incidence varies markedly over time.
- The bias from this can be substantial and may have led to overestimates of VE in the study by Wolff Sagy et al. as incidence was decreasing over time as the mpox outbreak subsided.
- Both immortal time bias and confounding by calendar time can be accommodated via methods that explicitly emulate the target trial of interest.

To demonstrate the potential for confounding by calendar time, we simulated data from vaccine campaign during an outbreak in which the vaccine provided no protection (i.e. VE = 0%, the setup is described in more detail in the Supplementary Materials). Figure 1 shows simulated

outbreaks under three hypothetical rollouts of the vaccine campaign (Panels A and B): one in which vaccinations are done all at once, another with a constant rate of vaccinations, and a third with an accelerating vaccination rate. As in the Wolff Sagy et al. study, in all cases, the start of the campaign coincides with a period of declining incidence in infections. We then implemented the time-dependent Cox regression approach described by Wolff Sagy et al. as well as possible alternatives to estimate VE using the simulated data. The bottom panel of Figure 1 shows the average VE estimate produced by the time-dependent Cox model with reset time zero improves over a naive approach but is still biased upwards when vaccines are rolled out over time. This bias is caused by the fact that vaccinated person-time is later, on average, than unvaccinated person-time, so vaccinated persons experience lower incidence, due to declining case counts, irrespective of any vaccine effect.

To show that calendar time bias may have led to overestimates of VE in the study by Wolff Sagy et al., we estimate the number of cases that would have occurred in the vaccinated group under the null that VE = 0% and compare that to the number observed. We use data extracted from a Kaplan Meier table in their preliminary analysis[5] before anyone was censored for receiving a second dose. Our estimate is based on the insight that, under the null, the number of cases in the vaccinated should be the number vaccinated on each calendar date times the incidence rate of the unvaccinated. This method estimates 10.5 expected cases in the vaccinated group under the null hypothesis of no vaccine effect (Fig. 2). A crude estimate of vaccine effectiveness would compare the observed number of infections in the vaccinated group to the number expected if the vaccine had had no effect: this would be 100%-3/10.5=72%, compared to a crude estimate from the same preliminary data of 84%. While it is impossible from the summary data published to account for finer time scales or to fully adjust for additional covariates to control other sources of confounding, we think our work above suggests some caution is warranted.

What should the authors have done instead? Splitting time-dependent vaccination on calendar date rather than resetting follow up time in their analysis may have solved the issue (third approach, Fig. 1C). However, more explicit target trial emulation methods could have prevented both confounding by calendar time and immortal time bias[6]. This approach also would also force the authors to more clearly articulate the ideal trial that the analysis is attempting to emulate, which would aid both interpretation of the current analysis and provide a replicable benchmark for future studies, randomized or not. While many such emulation methods are possible, one approach would be to construct a sequence of nested trials starting on each calendar day after the start of the vaccination campaign in which those eligible and vaccinated on that day are compared to those eligible and unvaccinated on that day. Stratified estimates of VE in each nested trial could then be pooled. In Panel C of Figure 1, we show that unlike the method of Wolff Sagy et al. this approach correctly adjusts for the differences in incidence rate across calendar time.

In conclusion, we urge future investigators to carefully consider the role changing incidence during an outbreak can play in studies of vaccine effectiveness.

References:

1. Hernán, M. A. & Robins, J. M. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am. J. Epidemiol.* **183**, 758–764 (2016).
2. Hernán, M. A. *et al.* Observational Studies Analyzed Like Randomized Experiments: An Application to Postmenopausal Hormone Therapy and Coronary Heart Disease. *Epidemiology* **19**, 766–779 (2008).
3. Wolff Sagy, Y. *et al.* Real-world effectiveness of a single dose of mpox vaccine in males. *Nat. Med.* **29**, 748–752 (2023).
4. Suissa, S. Immortal Time Bias in Pharmacoepidemiology. *Am. J. Epidemiol.* **167**, 492–499 (2008).
5. Arbel, R. *et al.* Effectiveness of a single-dose Modified Vaccinia Ankara in Human Monkeypox: an observational study. Preprint at https://doi.org/[https://doi.org/10.21203/rs.3.rs-1976861/v2 (2022).
6. Hernán, M. A., Sauer, B. C., Hernández-Díaz, S., Platt, R. & Shrier, I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J. Clin. Epidemiol.* **79**, 70–75 (2016).

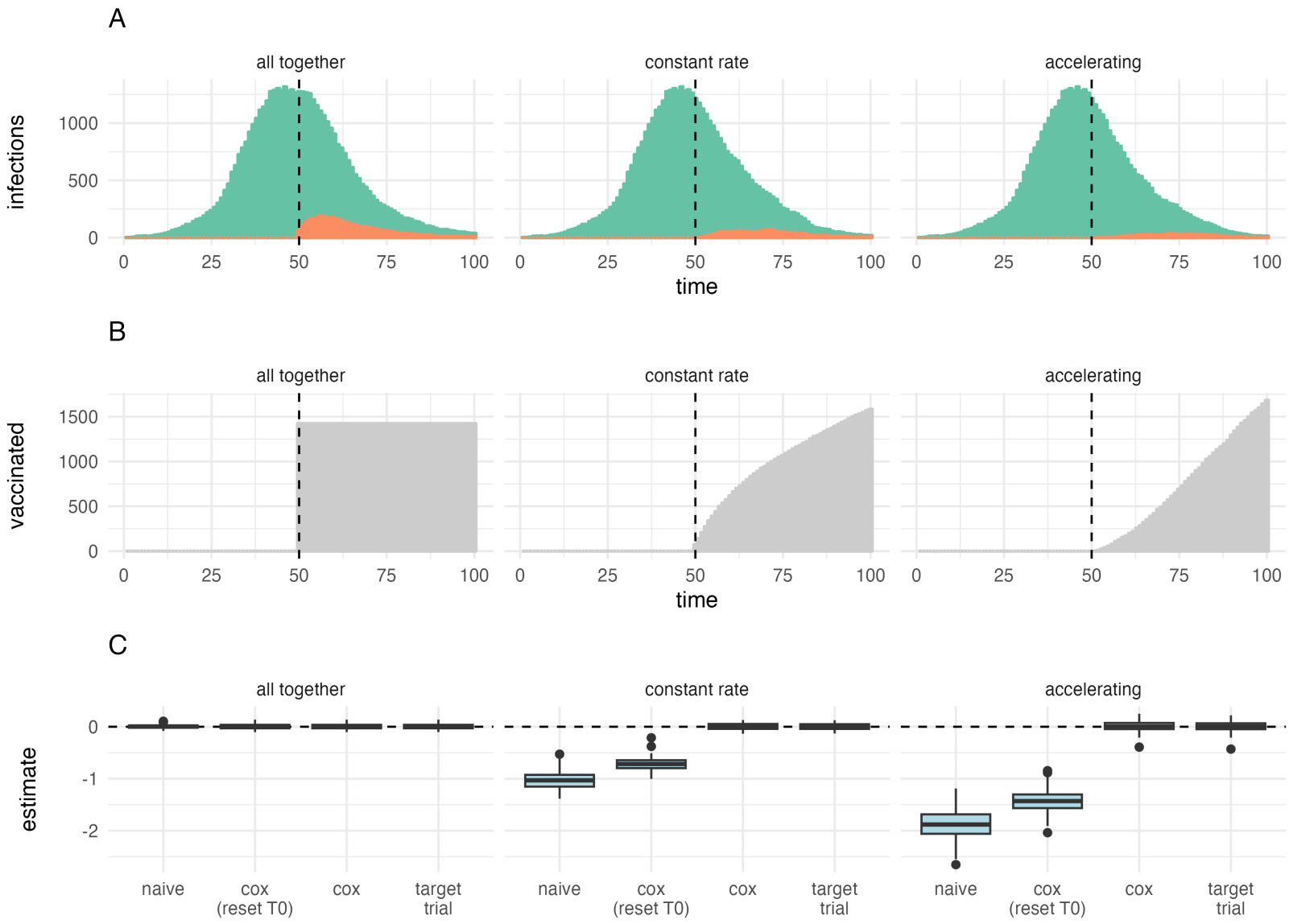
A
all together
constant rate
accelerating
infections
1000
500
0
0
25
50
75
100
time
B
all together
constant rate
accelerating
vaccinated
1500
1000
500
0
0
25
50
75
100
time
C
all together
constant rate
accelerating
estimate
0
-1
-2
naive
cox (reset T0)
cox
target trial

Figure 1: Simulated performance of different VE estimation strategies for a vaccination campaign during an outbreak. Panel A shows the number of infections among the unvaccinated (green) and vaccinated (orange) during the outbreak under three possible vaccination rollouts: every vaccinated at once, vaccinations occurring at a constant rate over follow up, or a slow start followed by an acceleration. Panel B shows the vaccines administered over time. Panel C shows the performance of the estimation strategies described in the Supplement across 500 simulated outbreaks.

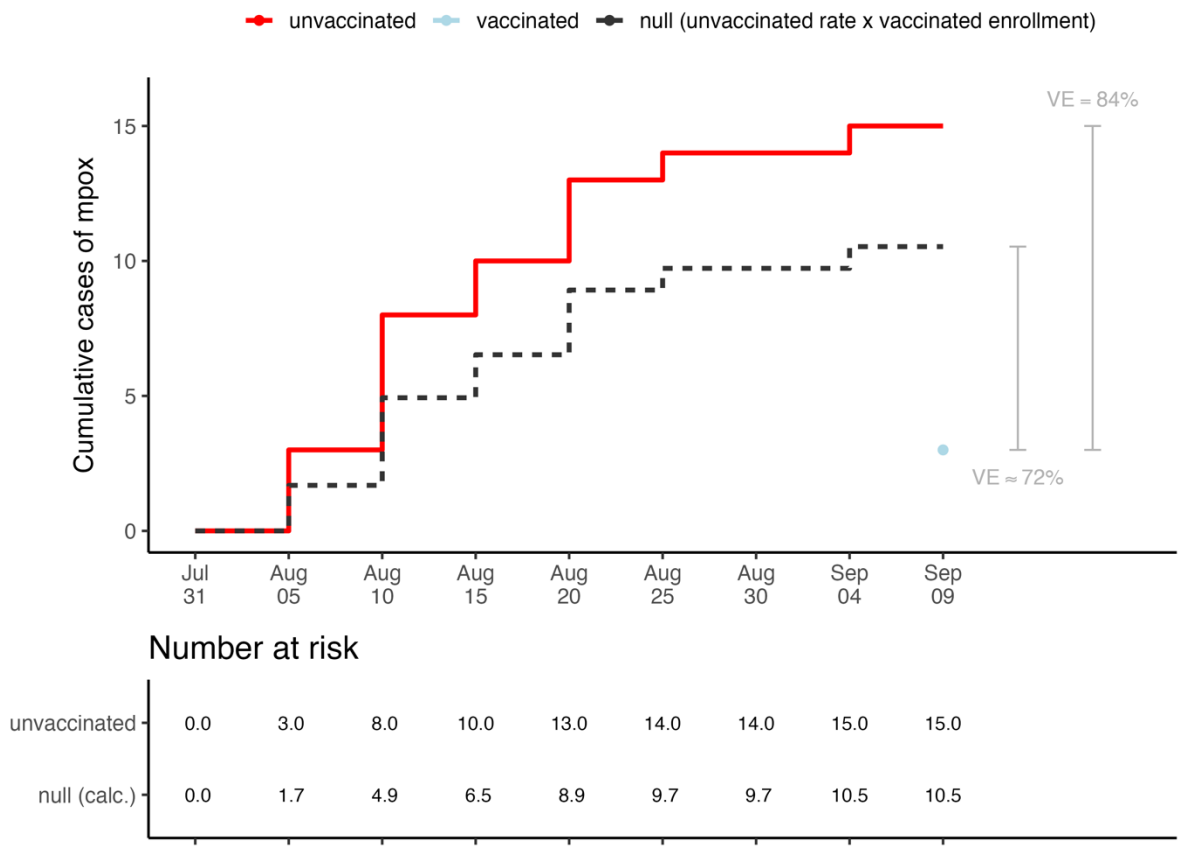

Figure 2: Cumulative mpox cases over follow up comparing vaccinated to unvaccinated and vaccinated if they had experienced the same incidence rate as unvaccinated.

Supplementary Materials:

Simulation details

To simulate an outbreak, we implemented a stochastic discrete-time individual-based SEIR model based on the population model:

$$S_{t+1} = \mathrm{S_t} - \beta I_t S_t$$
$$E_{t+1} = E_t + \beta I_t S_t - \delta E_t$$
$$I_{t+1} = I_t + \delta E_t - \gamma I_t$$
$$R_{t+1} = R_t + \gamma I_t$$

with compartments for susceptible (S), exposed (E), infectious (I), and recovered individuals with parameters $(\beta = 0.35, \delta = 0.2, \gamma = 0.2)$. To track transitions between states at the individual level, we define state indicator variables $\left(S_i^t, E_i^t, I_i^t, R_i^t\right)$ for individual i at time t. Then at each time step, we simulate transitions between states via:

$$\pi_{SE}^{t+1} \sim \mathrm{Bernoulli}\left(1 - exp\left(-\beta \sum_{i=1}^{n} I_i^t\right)\right)$$
$$\pi_{EI}^{t+1} \sim \mathrm{Bernoulli}(1 - exp(-\delta))$$
$$\pi_{IR}^{t+1} \sim \mathrm{Bernoulli}(1 - exp(-\gamma)),$$

and update state indicator variables to reflect these transitions:

$$S_i^{t+1} = S_i^t(1 - \pi_{SE}^{t+1})$$
$$E_i^{t+1} = S_i^t \pi_{SE}^{t+1} + E_i^t(1 - \pi_{EI}^{t+1})$$
$$I_i^{t+1} = E_i^t\, \pi_{EI}^{t+1} + I_i^t(1 - \pi_{IR}^{t+1})$$
$$R_i^{t+1} = I_i^t\, \pi_{IR}^{t+1}.$$

To simulate a vaccination campaign during the outbreak, we additionally define an indicator of vaccination status $\left(V_i^t\right)$ for individual i at time t. Prior to the start of the vaccination campaign, $V_i^t = 0$ for everyone. After the start of the campaign at $t^* = 50$, vaccination is assigned according to one of three different rollout strategies:

- *all together* – at the start of the campaign 30% of those who are not infectious or recovered are immediately vaccinated, i.e.
$$V_i^{t^*} \mid I_i^{t^*} = 0, R_i^{t^*} = 0 \sim \mathrm{Bernoulli}(0.3)$$
- *constant rate* – rollout of the vaccine occurs at a constant rate of 15 per thousand per day among those who are not infectious or recovered, i.e.
$$V_i^t \mid I_i^t = 0, R_i^t = 0, V_i^{t-1} = 0 \sim \mathrm{Bernoulli}(1 - exp(-0.015)) \text{ for } t \in (t^*, T)$$
- *accelerating rate* – the rate of vaccinations increases from zero at the start of the campaign to 40 per thousand per day at the end of the campaign among those who are not infectious or recovered, i.e.
$$V_i^t \mid I_i^t = 0, R_i^t = 0, V_i^{t-1} = 0 \sim \mathrm{Bernoulli}\left(1 - exp\left(\frac{-0.04}{T - t*}t + \frac{0.04}{T - t*}t^*\right)\right) \text{ for } t \in (t^*, T)$$

Because we are simulating under the null that VE = 0%, vaccination status does not affect transitions between compartments in our SEIR model nor is there need for a separate compartment for vaccination.

Finally, we estimate VE from the simulated outbreak data using the following four analytical approaches:

1. *naïve* – tally the number of infections among vaccinated and unvaccinated from the start of the campaign ($t = 50$) until the end ($t = 100$) and estimate VE as one minus the risk ratio.
2. *cox (reset T0)* – using the method of Wolff Sagy et al. assign person time after the start of the campaign until receipt of the vaccine as unvaccinated time and then vaccinated from their vaccination date after resetting time zero. Then use Cox regression with time-dependent covariate for vaccination and estimate VE as one minus the hazard ratio.
3. *cox* – as above assign person time after the start of the campaign until receipt of the vaccine as unvaccinated time and then vaccinated from their vaccination date but DO NOT reset time zero (instead use calendar time). Then use Cox regression with time-dependent covariate for vaccination and estimate VE as one minus the hazard ratio.
4. *target trial* – construct a sequence of nested trials starting on each calendar day after the start of the vaccination campaign in which those eligible and vaccinated on that day are compared to those eligible and unvaccinated on that day. Use Cox regression stratified by calendar day and estimate VE as one minus the pooled hazard ratio.

Simulations were repeated 500 times and results are shown in Figure 1.

Re-analysis details

We first estimated the number vaccinated each week as the difference between the number of vaccinated persons reported at risk in Figure 1 of ref. 5 in successive 5-day follow-up intervals, starting from the end. For example, 548 persons vaccinated in interval 1; 708-548 = 428 in interval 2; 773-708 = 113 in interval 3, and so forth. This estimation works because the number followed for the whole period should be the number vaccinated by interval 1; the number followed for all but one interval would be the number vaccinated by interval 2; and so forth. The use of 5-day intervals rather than days may introduce some error to this estimate. More than half incidence in the control group happened over the first 3 intervals (8 of 15 cases) (Ref 5, Fig. 1), yet 11% [(872-773)/872] of the vaccinated group did not experience this high-incidence period while vaccinated, violating the principle that the experience of the unexposed should reflect what would have happened to the exposed group, had they (counterfactually) been unexposed.

Using this estimate of the timing of vaccinations, in Table 1 we estimated the number of cases that would have occurred in the vaccinated group if the vaccine had been 0% effective; that is, we calculated for each week of follow up the product of the number of persons vaccinated by that week and the incidence rate in the unvaccinated group and summed this product over all weeks. For the number in the unvaccinated group each week, we used the number unvaccinated at risk in the first week in Figure 1 of Ref. 5 (1084), plus the number of individuals eventually vaccinated (871), minus the number whom we estimated above had been vaccinated by that calendar week.

This method estimates 10.5 expected cases in the vaccinated group under the null hypothesis of no vaccine effect (Figure 2 of this communication). A crude estimate of vaccine effectiveness would compare the observed number of infections in the vaccinated group to the number

expected if the vaccine had had no effect: this would be 100%-3/10.5 = 72%, compared to a crude estimate from the study data of 84%.

Table S1: Calculation of the number of cases among vaccinated that would have occurred if they experienced the same incidence rate as unvaccinated (i.e. VE = 0%)

Date	Unvaccinated (N)	Cumulative Cases (N)	Weekly Incidence rate (N per 1000)	Vaccinated (N)	Vaccinated x Incidence rate
Jul 31	1,423	0	0.0	548	0.0
Aug 05	1,259	3	2.4	708	1.7
Aug 10	1,191	8	4.2	773	3.2
Aug 15	1,091	10	1.8	869	1.6
Aug 20	1,088	13	2.8	869	2.4
Aug 25	1,084	14	0.9	872	0.8
Aug 30	1,084	14	0.0	872	0.0
Sep 04	1,083	15	0.9	872	0.8
Sep 09	1,084	15	0.0	871	0.0
TOTAL					**10.5**