Time Dependence of Pfizer-BioNTech Vaccine Efficiency

Samuel D. McDermott

April 16, 2021

How does the efficacy of the Pfizer-BioNTech Covid-19 vaccine evolve in time? Using data from Fig. 3 of [1], reproduced in my Fig. 3 and with data as presented in Tab. 1, I propose two simple models to determine the efficacy of the vaccine as a function of d, defined to be the day after the first vaccine dose.

The first model of expected number of confirmed cases per day is

$$\mathcal{N}_3(d|e_i, e_f, d_i, \mathcal{N}_p) = \begin{cases} (1-e_i)\mathcal{N}_p & d \leq d_i \\ (1-e_f)\mathcal{N}_p & d > d_i \end{cases}, \tag{1}$$

where $\mathcal{N}_p$ is the number of cases per day recorded amongst the placebo group. I assume this to be a fixed number for this simple analysis. Because 82 cases were contracted amongst the placebo group in the first 21 days of the trial, which is to say before the second dose (which is when we are most interested in the efficacy), I take $\mathcal{N}_p = 82/21 \simeq 3.9$, but I confirm that the results are not sensitive to this parameter for choices $3.5 < \mathcal{N}_p < 4$. The parameters of interest are: e_i, the efficiency immediately after the first dose; e_f, the "final" efficiency before the second dose, which occurs $\sim \mathcal{O}(10)$ days after the first dose; and d_i, the day on which the transition happens (more specifically: the final day characterized by efficiency e_i). I constrain $0 \leq e_i < 0.5 \leq e_f < 1$ and $3 \leq d_i \leq 10$ with $d_i \in \mathbb{Z}$.

The second model is

$$\mathcal{N}_5(d|e_i, e_t, e_f, d_i, d_t, \mathcal{N}_p) = \begin{cases} (1-e_i)\mathcal{N}_p & d \leq d_i \\ (1-e_t)\mathcal{N}_p & d_i < d \leq d_t \\ (1-e_f)\mathcal{N}_p & d > d_t \end{cases}. \tag{2}$$

Again I take $\mathcal{N}_p = 82/21 \simeq 3.9$. The parameters of interest are: e_i, the efficiency immediately after the first dose; e_t, the "transitional" efficiency $\sim \mathcal{O}(10)$ days after the first dose; e_f, the "final" efficiency before the second dose; d_i, the day on which the first transition happens (more specifically: the final day characterized by efficiency e_i); and d_t, the day on which the second transition happens (more specifically: the final day characterized by efficiency e_t). I constrain $0 \leq e_i < 0.3 \leq e_t < 0.7 \leq e_f < 1$ and $3 \leq d_i < d_t \leq 10$ with $d_i, d_t \in \mathbb{Z}$.

I take the log-likelihoods to be Poissonian (note: this is the *positive* log-likelihood, which we want to *maximize* – it is *not* the *negative* log-likelihood that we often use, which we want to *minimize*),

Table 1: Number of cases per day on days with nonzero cases, from [1] shown in Fig. 3.

d	2	3	4	5	6	7	8	9	10	11	13	14	17	18	25	26	35	43	54	57	61	64	69	97
$N(d)$	6	2	1	3	6	3	6	3	3	2	1	1	1	1	1	1	1	1	1	1	1	1	1	2

such that

$$\begin{aligned}\mathcal{L}_3(e_i, e_f, d_i, \mathcal{N}_p) &= -2\sum_{a=1}^{f}\left[\mathcal{N}_3(d_a) - N(d_a) + N(d_a)\ln(N(d_a)/\mathcal{N}_3(d_a))\right] \\ \mathcal{L}_5(e_i, e_t, e_f, d_i, d_t, \mathcal{N}_p) &= -2\sum_{a=1}^{f}\left[\mathcal{N}_5(d_a) - N(d_a) + N(d_a)\ln(N(d_a)/\mathcal{N}_5(d_a))\right],\end{aligned} \tag{3}$$

where $N(d)$ is the number of cases on day d, given in Tab. 1, and d_f is the final day in the sample. More properly, I should model the data and also the time evolution of the risk of infection. This would involve accounting for the delay time distribution between contracting SARS-CoV-2 and being diagnosed with Covid-19, which has a population average no shorter than 4 days [2, 3, 4, 5], and would also require accounting for environmental changes in the prevalence of the virus, which could be marginalized over by modeling the placebo group. Hopefully these are relatively small changes, so for the results shown here I am simply asserting that the likelihood of infection is constant, and I will account for the delay time distribution by simply "starting the clock" 4 days late, which means dropping the first 4 days of the data set (I confirm that results are not sensitive to this choice). Thus, I start on day 4. Accounting for this same delay, I take the final day as $d_f = 21 + 4 = 25$, since the second dose is received on day 21.

I determine the posteriors of each these parameters and the evidence for these models using the (dynamic) nested sampler `dynesty` [6, 7, 8, 9, 10]. I show corner plots in Fig. 1 and 2. I find that the initial efficacy is consistent with zero. **I find that the day before which we attain the "final" efficacy is day 7, and I find that that efficacy is $e_f \simeq 90^{+5}_{-7}\%$ in the 3-parameter model and $e_f \simeq 91^{+4}_{-6}\%$ in the 5-parameter model. Keep in mind that this is all before the second dose, and full vaccination is only attained one week after the second dose. At that time, efficacy is supposed to rise to 95% [1], which is compatible with my result.** I find that in the 5-parameter model, the "transitional" efficacy is prior-dependent. If I use the priors above I find that the data is not informative on e_t, whereas if I use $0 \le e_i < e_t < e_f < 1$ (with no further prior on the allowed ranges of each efficiency), it is bimodal with peaks at 30% and 70% and with mean $e_t \simeq 50\%$. However, it seems to robustly be true that $d_i = 6$ and $d_t = 7$ in this model. **I conclude that protection from SARS-CoV-2 at the 90% level starts on the 8^{th} day after the first dose.**

Comparison of the evidences E depends mildly on the priors. For the priors described above and using all days from $1-25$, I find $\Delta\ln(\mathsf{E}) = +0.82$ in favor of the 5-parameter model. If instead I use $0 \le e_i < e_t < e_f < 1$ (with no further prior on the allowed ranges of each efficiency) and all days from $1-25$ I find $\Delta\ln(\mathsf{E}) = +0.94$ in favor of the 5-parameter model. Similarly, if I take only days from $4-25$, the evidence becomes $\Delta\ln(\mathsf{E}) = +0.91$ in favor of the 5-parameter model. Thus the preference for the 5-parameter model seems robust, though it may not be worth the additional model complexity: I find the $\Delta\text{BIC} \simeq -4$ in favor of the 3-parameter model.

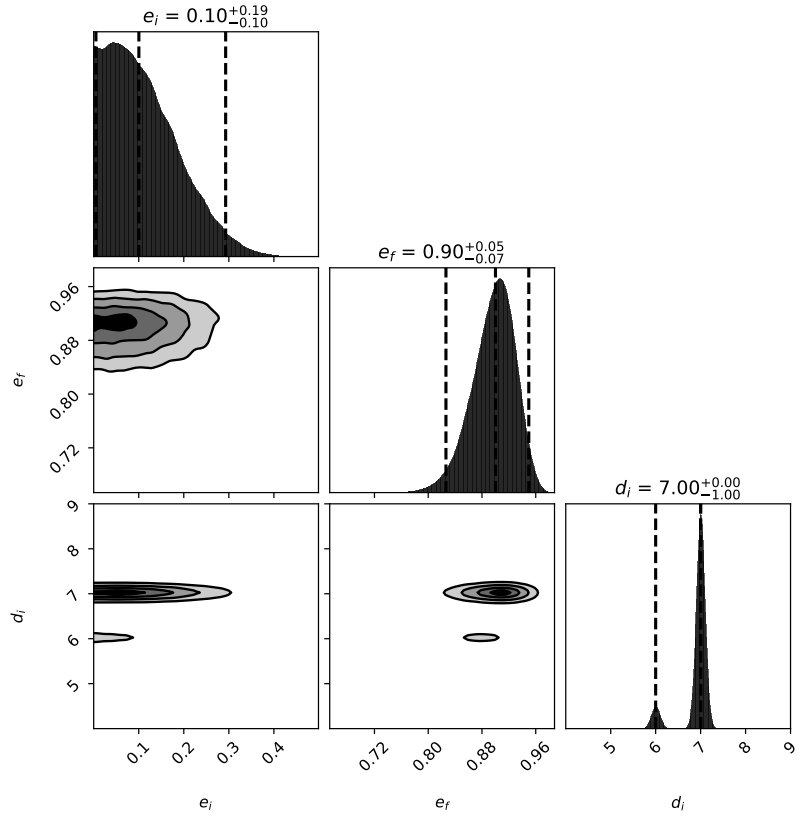


Figure 1: Corner plot from the 3-parameter model, Eq. (1). I take priors $0 \leq e_i < 0.5 \leq e_f < 1$ and $7 \leq d_i \leq 14$ with $d_i \in \mathbb{Z}$ and I use data from days $4 \leq d \leq 25$.

References

[1] F. P. Polack, S. J. Thomas, N. Kitchin, J. Absalon, A. Gurtman, S. Lockhart et al., *Safety and efficacy of the bnt162b2 mrna covid-19 vaccine*, *New England Journal of Medicine* **383** (2020) 2603–2615, [https://doi.org/10.1056/NEJMoa2034577].

[2] W. Tan, L. Y. Wong, Y. S. Leo and M. Toh, *Does incubation period of covid-19 vary with age? a study of epidemiologically linked cases in singapore*, *Epidemiology and infection* **148** (2020) .

[3] C. McAloon, Á. Collins, K. Hunt, A. Barber, A. W. Byrne, F. Butler et al., *Incubation period of covid-19: a rapid systematic review and meta-analysis of observational research*, *BMJ Open* **10** (2020) , [https://bmjopen.bmj.com/content/10/8/e039652.full.pdf].

[4] C. Daley, M. Fydenkevez and S. Ackerman-Morris, *A systematic review of the incubation period of sars-cov-2: The effects of age, biological sex, and location on incubation period*, *medRxiv* (2020) , [https://www.medrxiv.org/content/early/2020/12/24/2020.12.23.20248790.full.pdf].

[5] T.-k. Kong, *Longer incubation period of coronavirus disease 2019 (covid-19) in older adults*, *AGING MEDICINE* **3** (2020) 102–109, [https://onlinelibrary.wiley.com/doi/pdf/10.1002/agm2.12114].

[6] J. S. Speagle, *DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences*, *MNRAS* **493** (Apr., 2020) 3132–3158, [1904.02180].

[7] F. Feroz, M. P. Hobson and M. Bridges, *MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics*, *MNRAS* **398** (Oct., 2009) 1601–1614, [0809.3437].

[8] J. Skilling, *Nested Sampling*, in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering* (R. Fischer, R. Preuss and U. V. Toussaint, eds.), vol. 735 of *American Institute of Physics Conference Series*, pp. 395–405, Nov., 2004. DOI.

[9] J. Skilling, *Nested sampling for general Bayesian computation*, *Bayesian Analysis* **1** (2006) 833 – 859.

[10] E. Higson, W. Handley, M. Hobson and A. Lasenby, *Dynamic nested sampling: an improved algorithm for parameter estimation and evidence calculation*, *Statistics and Computing* **29** (2019) 891–913.

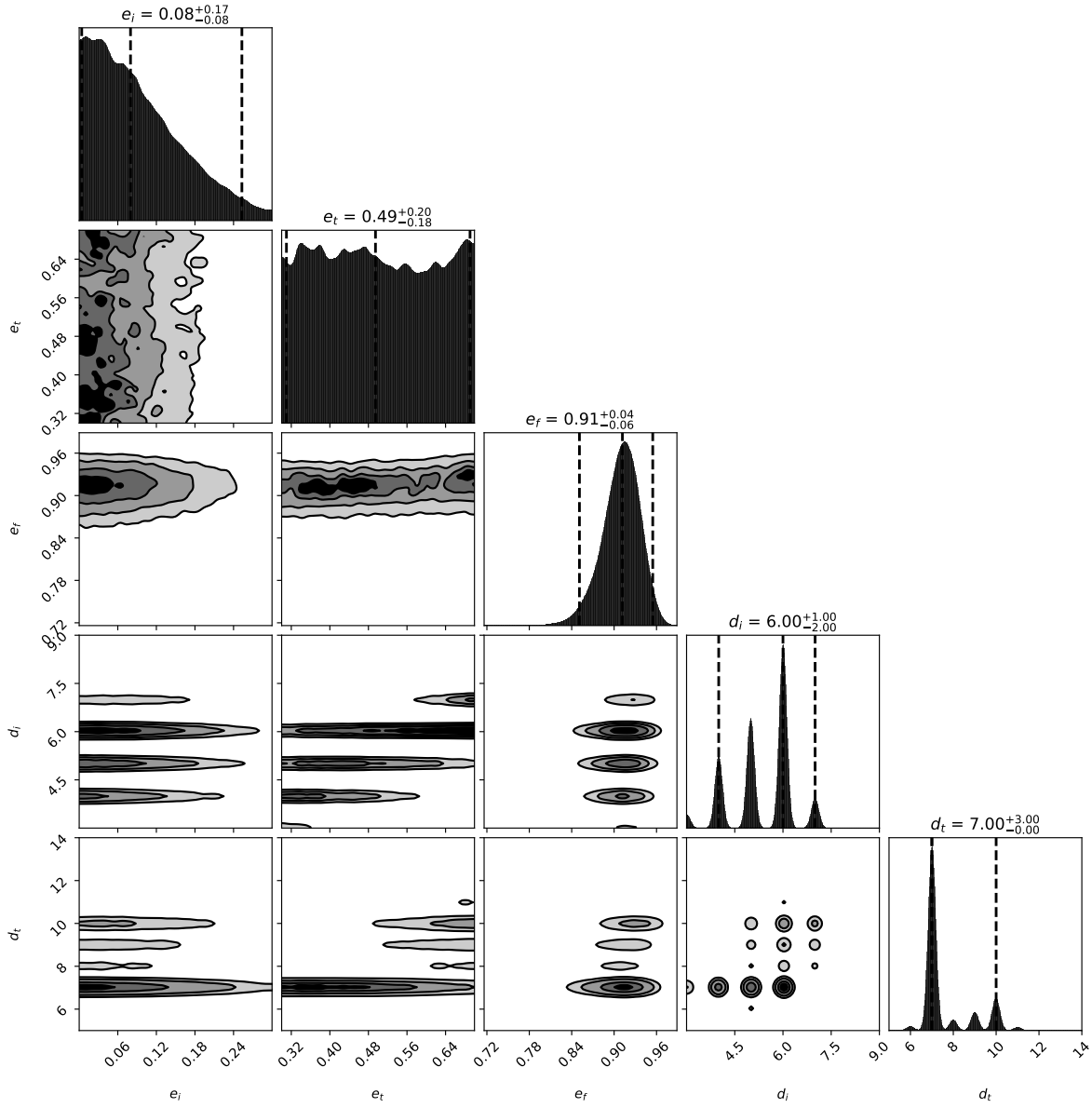
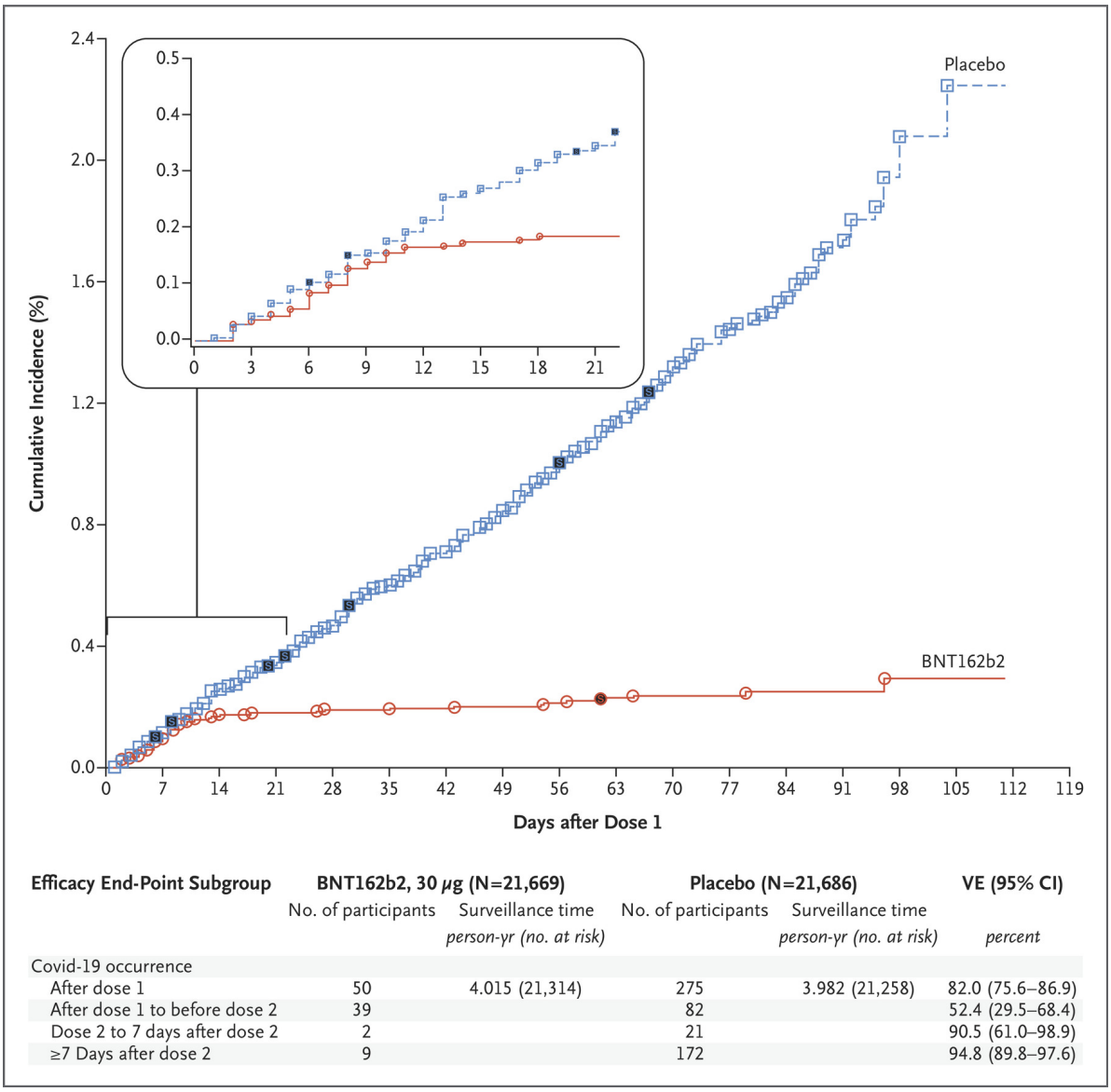


Figure 2: Corner plot from the 5-parameter model, Eq. (2). I take priors $0 \leq e_i < 0.3 \leq e_t < 0.7 \leq e_f < 1$ and $7 \leq d_i < d_t \leq 14$ with $d_i, d_t \in \mathbb{Z}$ and I use data from days $4 \leq d \leq 25$.



Efficacy End-Point Subgroup	BNT162b2, 30 μg (N=21,669)		Placebo (N=21,686)		VE (95% CI)
	No. of participants	Surveillance time *person-yr (no. at risk)*	No. of participants	Surveillance time *person-yr (no. at risk)*	*percent*
Covid-19 occurrence					
After dose 1	50	4.015 (21,314)	275	3.982 (21,258)	82.0 (75.6–86.9)
After dose 1 to before dose 2	39		82		52.4 (29.5–68.4)
Dose 2 to 7 days after dose 2	2		21		90.5 (61.0–98.9)
≥7 Days after dose 2	9		172		94.8 (89.8–97.6)

Figure 3: Number of Covid-19 cases in the placebo group versus the vaccine group versus date on which Covid-19 infection is confirmed. Data from [1].