The Mathematics and Statistics of Infectious Disease Outbreaks

Michael Höhle[1]

[1]Department of Mathematics
Stockholm University, Sweden

L09: Estimating Vaccine effectivness[1]

Stockholm University

[1]LaMo: 2022-04-25 @ 09:21:50

Outline

Overview

Outline

Vaccine efficacy (1)

- Vaccine efficacy is defined as the reduction in the attack rate attributable to the vaccine, i.e.

$$\mathrm{VE} = \left(\frac{p_u - p_v}{p_u} \right),$$

where p_u and p_v denote the risk of infection[2] in unvaccinated and vaccinated individuals over the observation period $[0, t]$, respectively,

- Alternatively, this is often written as

$$\mathrm{VE} = 1 - \mathrm{RR},$$

where RR is the relative risk of infection, i.e. p_v/p_u, also known as risk ratio.

[2] As described in Halloran et al. (2010, Tab 2.1) the VE for this endpoint is also denoted VE_S. Other endpoints, e.g disease, are also imaginable.

Vaccine efficacy (2)

- A hazard rate based definition of VE is in terms of the hazard ratio (HR) of infection in a survival context:

$$\mathrm{VE} = 1 - \mathrm{HR} = 1 - \frac{\lambda_v}{\lambda_u},$$

  where λ_u and λ_v are the constant hazard rates in the unvaccinated and vaccinated, respectively.
- If $\lambda_u t$ – and hence also $\lambda_v t$ – is small, the RR and HR based measure of VE do not differ much as seen from

$$p_u = F_u(t) = 1 - \exp(-\lambda_u t) \approx \lambda_u t \quad \text{and}$$
$$p_v = F_v(t) = 1 - \exp(-\lambda_v t) \approx \lambda_v t.$$

Vaccine efficacy (3)

- One often differs between *vaccine efficacy* for measures of the effect of vaccine in controlled clinial trials in the individual and *vaccine effectiveness* of the effect in the population under field conditions
- Vaccine effects in the population under field conditions are – in addition to biological factors – also influenced by
  - Vaccine storage
  - Vaccination schedules
  - Herd immunity
  - Human error

Estimating vaccine efficacy and effectiveness

- Vaccine efficacy and effectiveness can be estimated by the classical *cohort* and *case-control* designs known from epidemiology.
- Vaccination status of individuals is the exposure variable and outcome is whether individuals developed the disease.
- Furthermore, a regression contexts allows adjusment for possible confounders as in standard cohort or case-control analysis.
- In this course only the basic theory is shown, see Halloran et al. (2010) for a more thorough treatment of how to estimate analyse vaccine effects from study data.

Estimating vaccine efficacy from cohort studies

- In a cohort design, assume n_v and n_u vaccinated and unvaccinated individuals, respectively, are subject to follow-up without censoring. Suppose y_v and y_u individuals become cases during the observation period.
- An estimate of vaccine efficacy is then:

$$\widehat{\text{VE}} = 1 - \frac{y_v/n_v}{y_u/n_u}$$

- Confidence intervals for VE are derived from transforming confidence intervals for the RR.
- Poisson regression including person-years as offset can be used to adjust for additional confounders.

Salk polio vaccine field trial (1)

- 1954 randomized controlled trial to investigate polio vaccination in the USA (Francis, 1955)
- Incidence of polio at that time was about 50 per 100,000 children
- Vaccination of childrens in 1st to 3rd grade:

Vaccination \ Poliomyelitis	Yes	No
Yes	57	2.00745×10^5
No	142	2.01229×10^5

- One of the largest controlled trials ever
- Question: Does the vaccination work?

Salk polio vaccine field trial (2)

- The function `Epi::twoby2` can be used to compute the relative risk and associated CIs from simple 2×2 tables:

```
salk <- t(matrix(c(57,200745,142,201229),2,2))
dimnames(salk) <- list(c("vacc","non-vacc"),c("case","nocase"))
require("Epi")
## Indlæser krævet pakke:  Epi
tab <- twoby2(salk,F.lim=1e9,print=FALSE)
(RR <- tab$measures["             Relative Risk:",])
##           95% conf.  interval
## 0.4025459 0.2960563 0.5473391
(VE <- 1-RR[c(1,3,2)])
##            interval 95% conf.
## 0.5974541 0.4526609 0.7039437
```

COVID-19 Example

- Phase III placebo controlled trial for the BNT162b2 mRNA COVID-19 Vaccine (Polack et al., 2020):

```
biontech <- t(matrix(c(8,21720,162,21728),2,2))
dimnames(biontech) <- list(c("vacc","non-vacc"),c("case","nocase"))
tab <- Epi::twoby2(biontech,F.lim=1e9,print=FALSE)
RR <- tab$measures["             Relative Risk:",]
(VE <- 1-RR[c(1,3,2)])
##             interval 95% conf.
## 0.9502491 0.8988468 0.9755307
```

Estimating vaccine efficacy from case-control studies

- One can also use case-control studies to estimate VE, i.e. let n_{cases} and $n_{controls}$ be the number of cases and controls, respectively, and let p_{cases} and $p_{controls}$ be the proportions vaccinated in the two groups.
- Here, one uses the odds ratio of vaccination in cases and controls, which is equal to the odds ratio of infection in vaccinated and unvaccinated
- Provided that the attack rate is low (i.e. $RR \approx OR$) one obtains

$$\widehat{\mathrm{VE}} \approx 1 - \frac{p_{cases}/(1-p_{cases})}{p_{controls}/(1-p_{controls})}$$

- Logistic regression can be used to adjust the estimates for additional confounders.

Critical Vaccination Fraction (1)

- Transmission models can be used to determine vaccine effectiveness in the population (Halloran et al., 2010, Ch. 5).
- Let $R_f(t)$ be the reproductive number at time t when the fraction f of the population is vaccinated
- We can measure the effect of a vaccination strategy by comparing the relative reduction in the reproduction number:

$$\mathrm{VE}_{R,f}(t) = 1 - \frac{R_f(t)}{R_0}.$$

- Let f^* denote the minimum proportion to be vaccinated for eradication if vaccination induces complete protection, i.e. $VE_S = 1$.

Critical Vaccination Fraction (2)

- In a homogeneous model and completely susceptible population:

$$R_f = R_0(1-f) < 1 \Rightarrow f^* > 1 - \frac{1}{R_0}$$

- Example: With $R_0 = 3$ means that 67% of the population needs to be vaccinated.
- If the vaccine instead is *leaky* then everyone can become infected, but the probability of a contact being infectious is θp instead of p with $0 < \theta < 1$ (i.e. $\mathrm{VE}_S = 1 - \theta$)
- In this case $R_1 = c\theta pd = \theta R_0$ and, hence, $R_f = fR_1 + (1-f)R_0$. Thus, if V_S is high

$$R_f < 1 \Rightarrow f^* > \frac{1}{1-\theta}\left(1 - \frac{1}{R_0}\right)$$

- Example: For $R_0 = 3$ and $\theta = 0.1$ we get $f^* > 74\%$

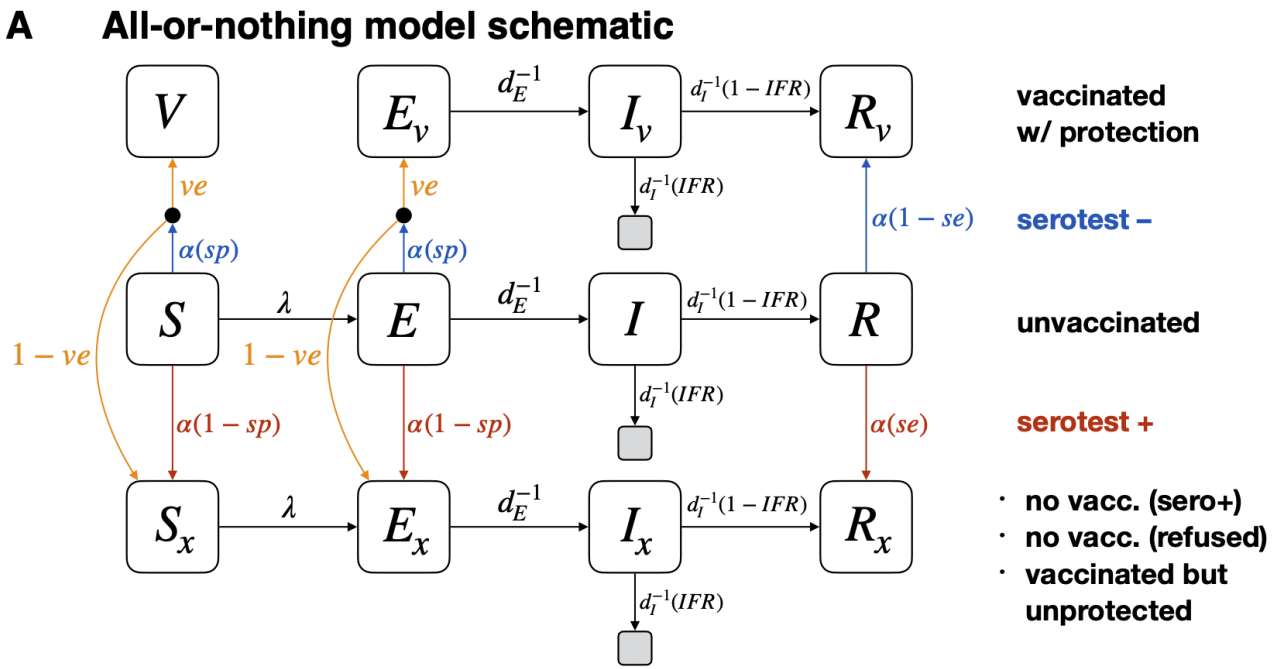
Critical Vaccination Fraction (3)

- For an *all-or-nothing* vaccine, where a proportion $0 < \alpha < 1$ is completely protected and a proportion is not protected at all the equivalent expression is $f^* > \frac{1}{\alpha}\left(1 - \frac{1}{R_0}\right)$.
- Example: For $R_0 = 3$ and $\alpha = 0.9$ we again get $f^* > 74\%$.
- Note: Both the leaky and all-or-nothing critical fractions are of the type

$$f^* > \frac{1}{\mathrm{VE}_S}\left(1 - \frac{1}{R_0}\right).$$

Transmission models and vaccination

- In SIR models, vaccination is typically reflected by introducing an additional 'vaccinated' (V) state
- COVID-19 example from Bubar et al. (2021):



Outline

Screening method (1)

- Farrington (1993) describes a retrospective method to assess the effectiveness of a vaccine under field conditions.
- This *screening method* requires only data on the cases – together with an estimate of vaccine coverage from sources external to the study (e.g. vaccination register data).
- Since the method is fitted within a GLM framework control for confounders (e.g. age, geographic location) is immediate by adding them as covariates.

Screening Method (2) – Notation

Random sample of disease cases in a given period from a defined population.

- Use sample to estimate the proportion cases vaccinated (PCV)
- Use data on the proportion of the population vaccinated (PPV)
- Vaccine effectiveness (VE) is given by

$$\text{VE} = 1 - \frac{\text{PCV}}{1 - \text{PCV}} \cdot \frac{1 - \text{PPV}}{\text{PPV}}$$

- Note that $1 - \text{VE}$ corresponds to the odds ratio of vaccination in cases and the population.

Stratification by GLM (1)

- We now consider n strata with θ_i and π_i denoting PCV and PPV in the cohort, respectively.
- Let $R_i = 1 - \mathrm{VE}_i$ be the relative risk of disease in vaccinated relative to unvaccinated individuals in each stratum.
- Assume k covariates are also available in each stratum, i.e. let $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ik})'$.
- Log-linear model to control for confounding by stratification

$$\log(R_i) = \alpha + \boldsymbol{x}_i'\boldsymbol{\beta}, \quad i = 1, \ldots, n$$

Stratification by GLM (2)

- In case of a rare disease $R_i = \text{OR}_i$. Hence,

$$\log(R_i) \approx \log\left(\frac{\theta_i}{1-\theta_i}\frac{1-\pi_i}{\pi_i}\right) = \alpha + \boldsymbol{x}_i'\boldsymbol{\beta}$$

- Re-arranging terms yields:

$$\log\left(\frac{\theta_i}{1-\theta_i}\right) = \underbrace{\log\left(\frac{\pi_i}{1-\pi_i}\right)}_{o_i} + \alpha + \boldsymbol{x}_i'\boldsymbol{\beta}$$

$$= o_i + \alpha + \boldsymbol{x}_i'\boldsymbol{\beta}$$

- This corresponds to a binary GLM with logit-link and offset o_i

$$\text{No. vaccinated cases} \sim \text{Bin}(\text{No. cases}, \theta_i).$$

Sample size (1)

- The user should specify: V_T anticipated true effectiveness, V_L the lower threshold effectiveness value as well as $PPV = \pi$, the power $1 - \beta$ and the significance level α
- Standard arguments for computing sample sizes for difference in proportion in $2\times$ 2 tables yield

$$N = \frac{(z_{\alpha/2} + z_{1-\beta})^2(1 - \pi V_L)^2(1 - V_T)}{\pi(1-\pi)(V_T - V_L)^2}, \qquad (1)$$

  where, e.g., $z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution.
- Alternatively, if $\beta = 0.5$ then $V_T - V_L$ can be thought of as anticipated lower half-width of the corresponding $(1 - \alpha/2) \cdot 100\%$ CI.

Sample size (2)

- R function implementing the sample size equation (1)

```
N <- function( VT, VL, pi, alpha=0.05, beta=0.2) {
  (qnorm(alpha/2) + qnorm(1-beta))^2 * (1-pi*VL)^2 * (1-VT) / ( pi*(1-pi)*(VT-VL)^2)
}
```

- Reproducing the first 3 rows of Table 2 in Farrington (1993):

```
VT<-seq(0.5,0.9,by=0.1)
t(sapply(c(0.05,0.1,0.15), function(delta) {
  structure(round(N( VT=VT,VL=VT-delta, pi=0.5,alpha=0.05,beta=0.5)),names=VT)
}))
##       0.5  0.6 0.7 0.8 0.9
## [1,] 1846 1292 840 480 203
## [2,]  492  346 226 130  55
## [3,]  232  164 108  62  27
```

Measles in Leeds 1980–1986

- Farrington (1993) illustrates the screening method using measles notifications from two districts in Leeds 1980–1986.

```
#Load data on measles notifications and vaccine coverage in Leeds
(measles <- read.csv2(file=file.path("Data", "farrington1993-table1.csv")))
##   Birth.cohort Cases Vaccinated Coverage
## 1         1980    82          5     70.0
## 2         1981    98          9     70.9
## 3         1982   180         28     76.0
## 4         1983   177         37     81.0
## 5         1984   112         22     83.7
## 6         1985   140         27     84.5
## 7         1986   151         27     83.1
#Create offset from PCV
measles$o <- with(measles, log(Coverage / (100-Coverage)))
#Create factor representing the birth cohort
measles$bc <- as.factor(measles$Birth.cohort)
m <- glm( cbind(Vaccinated, Cases-Vaccinated) ~ offset(o), family=binomial,data=measles)
m.bc <- glm( cbind(Vaccinated, Cases-Vaccinated) ~ offset(o) + bc, family=binomial,data=measles)
#Compare models
(p <- anova(m, m.bc,test="Chisq")$"Pr(>Chi)"[2])
## [1] 0.5758264
#Vaccine coverage
(VE <- (1-exp(as.numeric(coef(m)))))
## [1] 0.950991
#Confidence intervals
sort(1-exp(confint(m)))
## Waiting for profiling to be done...
##    97.5 %     2.5 %
## 0.9419066 0.9589255
```

Other methods to assess vaccine effectiveness (1)

- Vaccine effectiveness can also be quantified using time series models for disease incidence.
  - One option is to define periods before and after the introduction of vaccination and analyse the difference in model parameters for the two periods (Anderson et al., 1984; Antunes et al., 2007; Girard, 2000).
  - A second option is to include vaccination coverage as a (time-varying) covariate in the time series model (Herzog et al., 2011; Höhle et al., 2011).
  - Usually, coverage proportions are only available for certain age groups.
- Instead of a time series model one can also use a SIR transmission model with an extra "vaccinated" department.
- In either case it is important to realize that vaccine effects can be a combination of short term effects and long term effects.

Other methods to assess vaccine effectiveness (2)

- Time series regression model in Höhle et al. (2011) for the logarithm of the incidence containing intercept, trend and seasonal terms:

$$\log(y_t) = \beta_0 + \beta_1 \cdot t + f(t) + \epsilon_t, \quad t = 1, 2, \ldots, n,$$

$$\mu_t = E(y_t) = \exp\left(\beta_0 + \beta_1 \cdot t + f(t) + \frac{1}{2}\sigma^2\right),$$

  where $f(t)$ is a periodic function and ϵ_t is described by an ARMA(p, q) model with variance σ^2.

- Using $\beta_1 \cdot \text{coverage}_t$ instead of $\beta_1 \cdot t$ in the above model leads to

$$\text{VE} = 1 - \frac{\mu_{100\% \text{ coverage}}}{\mu_{0\% \text{ coverage}}} = 1 - \exp(\beta_1)^{100}.$$

Literature I


Anderson, R. M., B. T. Grenfell, and R. M. May (1984). "Oscillatory Fluctuations in the Incidence of Infectious Disease and the Impact of Vaccination: Time Series Analysis". In: *The Journal of Hygiene* 93.3, pp. 587–608.


Antunes, J. L., E. A. Waldman, C. Borrell, and T. M. Paiva (Dec. 2007). "Effectiveness of influenza vaccination and its impact on health inequalities". In: *Int J Epidemiol* 36, pp. 1319–1326.


Bubar, Kate M., Kyle Reinholt, Stephen M. Kissler, Marc Lipsitch, Sarah Cobey, Yonatan H. Grad, and Daniel B. Larremore (2021). "Model-informed COVID-19 vaccine prioritization strategies by age and serostatus". In: *Science* 371.6532, pp. 916–921. DOI: 10.1126/science.abe6959.


Farrington, C. P. (1993). "Estimation of Vaccine Effectiveness Using the Screening Method". In: *International Journal of Epidemiology* 22.4, pp. 742–746.

Literature II


Francis, T. (1955). "Evaluation of the 1954 field trial of poliomyelitis vaccine. Summary report". In: *Am J Public Health* 45.2 Pt. 2, pp. 1–51.


Girard, D. Z. (2000). "Intervention times series analysis of pertussis vaccination in England and Wales". In: *Health Policy* 54, pp. 13–25.


Halloran, M. E., I. M. Longini, and C. Struchiner (2010). *Design and Analysis of Vaccine Studies*. Springer-Verlag.


Herzog, S. A., M. Paul, and L. Held (2011). "Heterogeneity in vaccination coverage explains the size and occurrence of measles epidemics in German surveillance data". In: *Epidemiology and Infection* 139.4, pp. 505–515.

Literature III


Höhle, M., A. Siedler, M. Ludwig, H-M. Bader, U. Heininger, and R. von Kries (2011). "Assessment of varicella vaccine effectiveness in Germany: A time series approach". In: *Epidemiology and Infection* 139.11, pp. 1710–1719. DOI: 10.1017/S0950268810002815.


Polack, Fernando P. et al. (2020). "Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine". In: *New England Journal of Medicine* 383.27, pp. 2603–2615. DOI: 10.1056/NEJMoa2034577.