VE estimates with administrative data

Arseniy Khvorov

October 2019

Simulation repository
https://github.com/khvorov45/ve-admin

Contents

1 Methods

1.1 Core simulation

Starting population was the general population. Its size was set to 500,000. Every individual had their attributes randomly allocated in the order shown in Figure 2 and with probabilities shown in tables below.

First allocated attribute was true vaccination status. Measurement of this status for the purposes of the simulated study was allowed to be inaccurate in order to simulate exposure miscalssification. True vaccination status was used to determine which probability to use to allocate individuals to the flu-infected category. Both vaccinated and unvaccinated subjects were allocated to the non-flu infected category with the same probability. The remaining subjects ended up as part of the non-ARI category.

Everyone with an ARI (either infected with flu or a non-flu pathogen) was assigned to either the symptomatic or the asymptomatic group. Those who were symptomatic were assigned to the clinically assessed or unassessed groups. Being clinically assessed in this context means that they presented to a clinic with ARI illness and they were classified as an ARI case. These clinically assessed ARI cases got one probability of being tested, everyone else got another. Tests were allowed to be imperfect to simulate outcome misclassification.

1.2 Population summary

Each population was collapsed down to summary results - each individual was considered to be part of one of eight categories: administrative/surveillance vaccinated/unvaccinated case/control as shown in Figure 1.

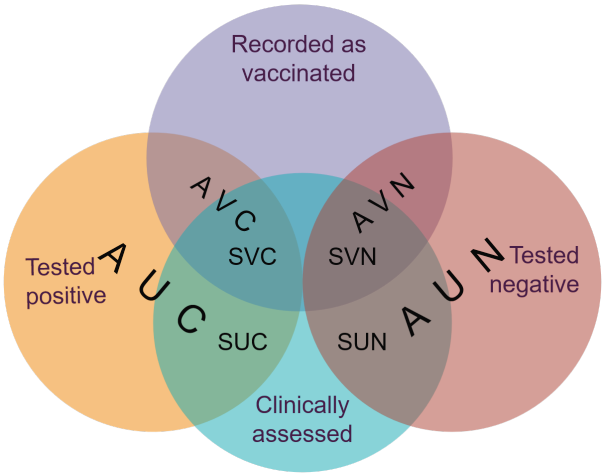

Figure 1: Assignment of an individual to appropriate categories. A - administrative, S - surveillance, V - vaccinated, U - unvaccinated, C - case, N - non-case (control).

Individuals in each of the categories were counted. These counts were representative of those that could have been obtained if a test-negative study was done on that population either using administrative or surveillance data. VE estimates could then be calculated as 1 - OR where $\text{OR} = \frac{\text{Odds in vaccinated}}{\text{Odds in unvaccinated}}$ where $\text{Odds} = \frac{\text{Count of cases}}{\text{Count of controls}}$.

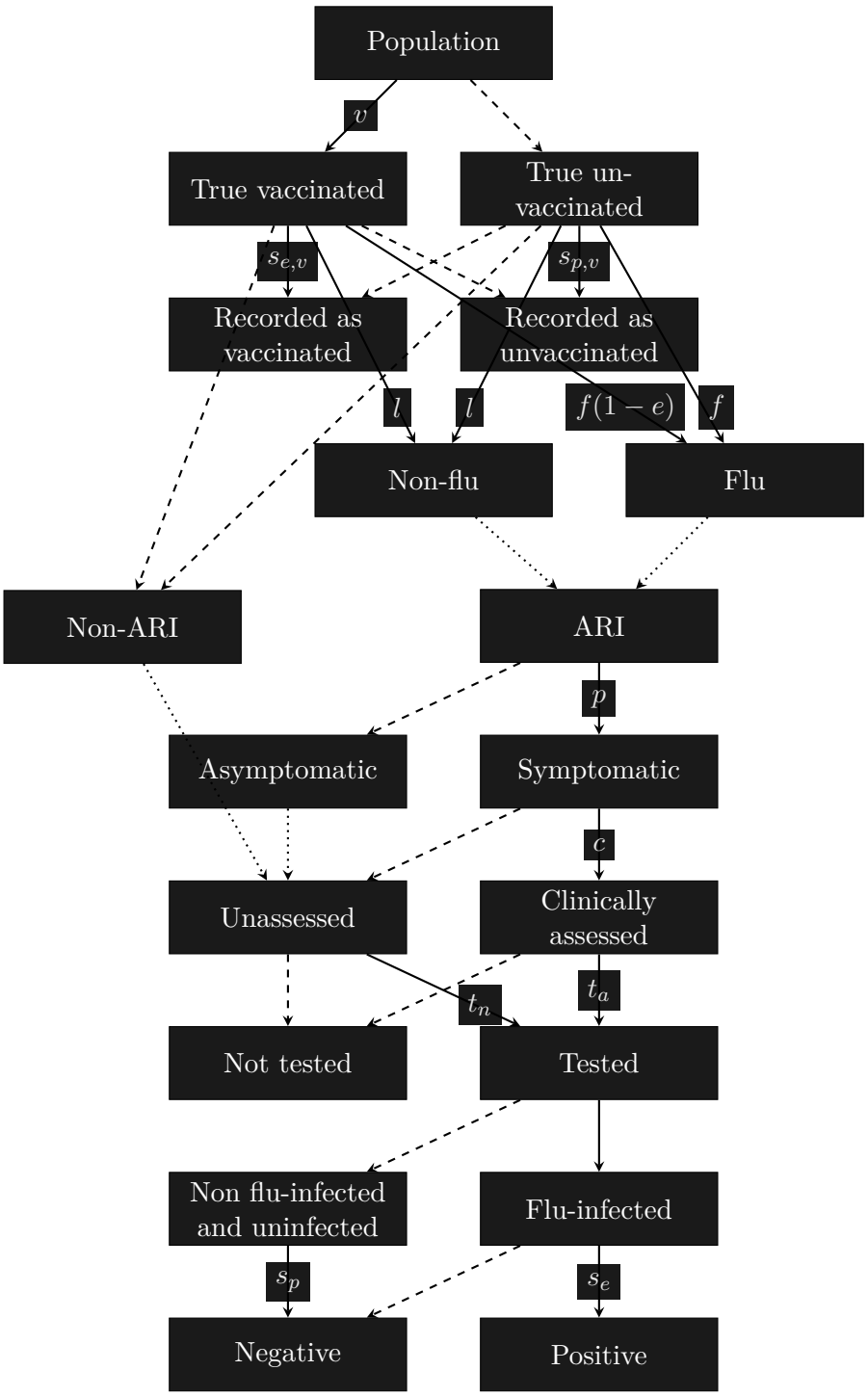

Figure 2: Simulation decision tree. Parameter key is in Table 1. Solid lines mean allocation with probabilities represented by the indicated parameters. Dashed line probabilities are complements of corresponding solid line probabilities. Dotted lines represent full-group allocation. Probability of being flu-infected for those who are tested isn't indicated because it wasn't necessary for the purposes of simulations.

1.3 Parameter Variation

Every parameter in the simulation was set at a prespecified value. Some of the parameters were set to vary (their prespecified value would have been ignored then). Setting a parameter to vary meant meant that the parameter was assigned a small set of values, set to the first of those values, a set amount of populations were simulated using that value, then it was set to the next value and so on until the simulation went through the entire set.

If multiple parameters were set to vary the simulation would have gone through all possible combinations of all parameters values.

1.4 Mixed-group simulations

If a simulation required a population to be composed of multiple groups, each group was simulated as if it were a separate population. The only mixed combination used was children/adults/elderly. The total size of each group was obtained by multiplying the total requested sample size (usually 200,000) by the specified proportion of each group in the population (this is the w parameter which would have been set to 1 if there was only one group in the population). Amounts of cases and controls were counted in each of the groups and added together to represent the counts obtained from the full population.

1.5 Additional simulations

To determine the effects of individual parameters, an additional set of simulations was performed with parameters fixed to values shown in Table 3. This set of parameter values produced unbiased VE estimates in surveillance data. Using this as a baseline, required parameters could be varied (e.g. s_p can be set below 1) to observe their effect in absence of other sources of bias. Additional simulations were also used to observe the effect some parameters have on others (e.g. how s_p set below 1 affects variation of VE estimates at different values of t_n).

1.6 Parameter estimates used

Table 1: Parameter names, meanings and values used in simulations. "Range used" shows the range of values used for variation in individual age group simulations. Tables 2 contains values and patterns used for variation in mixed group simulations. Every parameter except c represents an absolute probability (some of them only apply to subsets of the population). Only relative probability estimates could be obtained for c (by comparing presentation counts found in ASPREN data [1] to expected underlying population size derived from other parameter estimates), its values were set to 1 in individual group simulations unless it is the parameter varied. Parameter w was only relevant if the population was requested to be composed of multiple groups. Shown values are the ones used in children/adults/elderly mixed simulation. In individual simulations, w was set to 1. Estimates of v and e were derived from provided data.

Par.	Description	Range Used	Children (<15)	Adults (15-65)	Elderly (65+)	Ref.
w	Proportion of the age groups in the general population		0.189	0.657	0.154	[2]
v	Probability of being vaccinated	0.05 - 0.5	0.1	0.25	0.66	
$s_{e,v}$	Sensitivity of exposure measurement	0.9 - 1	0.9	0.95	0.98	[3, 4, 5]
$s_{p,v}$	Specificity of exposure measurement	0.5 - 1	0.9	0.8	0.7	[3, 4, 5]
e	Vaccine effectiveness	0.1 - 0.9	0.6	0.5	0.4	
f	Influenza risk in unvaccinated	0.05 - 0.15	0.15	0.08	0.05	[6]
l	Non-influenza ARI risk in vaccinated and unvaccinated	0.1 - 0.3	0.3	0.15	0.1	[7, 8]
p	Probability of the ARI being symptomatic	0.1 - 0.9	0.84	0.84	0.84	[9]
c	Relative probability of being clinically assessed as having ARI when it is symptomatic	0.1 - 0.9	0.4	0.3	1	
t_a	Tested probability for clinically assessed ARI	0.1 - 0.9	0.17	0.35	0.22	[1]
t_n	Tested probability for everyone without clinically assessed ARI	0 - 0.3	0.15	0.15	0.15	
s_e	Sensitivity of influenza test	0.5 - 1	0.86	0.86	0.86	[10]
s_p	Specificity of influenza test	0.9 - 1	0.984	0.984	0.984	[10]

Table 2: Combinations (patterns of variation) and values used in fixed variation of parameters when multiple groups were present in the population. Combinations 1 and 3 were not used for w.

Combination	Children	Adults	Elderly
1	Low	Low	Low
2	Mid	Mid	Mid
3	High	High	High
4	High	Low	Low
5	Low	High	Low
6	Low	Low	High

Parameter	Low	Mid	High
w	0.15	0.33	0.7
v	0.05	0.3	0.5
$s_{e,v}$	0.9	0.95	1
$s_{p,v}$	0.5	0.75	1
e	0.1	0.5	0.9
f	0.05	0.1	0.15
l	0.1	0.15	0.3
p	0.1	0.5	0.9
c	0.1	0.5	0.9
t_a	0.1	0.5	0.9
t_n	0	0.15	0.3
s_e	0.5	0.75	1
s_p	0.9	0.95	1

Table 3: Parameter values used in the additional simulation set. Parameter w is missing because the additional simulations were always performed with only one parameter set in the population (equivalent to only having one age group).

Parameter	Value
v	0.5
$s_{e,v}$	1
$s_{p,v}$	1
e	0.5
f	0.3
l	0.3
p	1
c	1
t_a	1
t_n	1
s_e	1
s_p	1

2 Results and discussion

The following sections present and discuss results associated with every parameter whose variation within ranges defined in Table 1 had a perceivable effect on the bias of VE estimates.

2.1 Individual-group simulations

2.1.1 Effect of tested proportions — t_a and t_n

Changing t_a and t_n only affected the estimates of VE in administrative data as shown in Figures **??** and **??** respectively. The parameter that allowed the pattern seen in Figure **??** to be replicated in the additional simulation set was influenza test specificity s_p when set below 1. No additional simulation replicated the pattern seen in Figure **??** (most obvious in the elderly group).

Bibliography

[1] Australian Sentinel Practices Research Network;. Available from: https://aspren.dmac.adelaide.edu.au/.

[2] Australian Demographic Statistics. Australian Bureau of Statistics; 2018. Available from: https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3101.0Sep2018?OpenDocument.

[3] Irving SA, Donahue JG, Shay DK, Ellis-Coyle TL, Belongia EA. Evaluation of self-reported and registry-based influenza vaccination status in a Wisconsin cohort. Vaccine. 2009;27(47):6546–6549.

[4] Donald RM, Baken L, Nelson A, Nichol KL. Validation of self-report of influenza and pneumococcal vaccination status in elderly outpatients. American Journal of Preventive Medicine. 1999;16(3):173–177.

[5] Rolnick Sj, Parker Ed, Nordin Jd, Hedblom Bd, Wei F, Kerby T, et al. Self-report compared to electronic medical record across eight adult vaccines: Do results vary by demographic factors? Vaccine. 2013;31(37):3928–3935.

[6] Tokars JI, Olsen SJ, Reed C. Seasonal Incidence of Symptomatic Influenza in the United States. Clinical Infectious Diseases. 2017;66(10):1511–1518.

[7] Influeza Surveillance Report 2017. Australian Department of Health; 2017. Available from: https://www.health.gov.au/internet/main/publishing.nsf/Content/cda-ozflu-2017.htm.

[8] Influeza Surveillance Report 2018. Australian Department of Health; 2018. Available from: https://www.health.gov.au/internet/main/publishing.nsf/Content/ozflu-surveil-2018-final.htm.

[9] Leung NHL, Xu C, Ip DKM, Cowling BJ. The fraction of influenza virus infections that are asymptomatic: a systematic review and meta-analysis. Epidemiology. 2015;26(6):862–872.

[10] Druce J, Tran T, Kelly H, Kaye M, Chibo D, Kostecki R, et al. Laboratory diagnosis and surveillance of human respiratory viruses by PCR in Victoria, Australia, 2002–2003. Journal of Medical Virology. 2004 Nov;Available from: https://www.onlinelibrary.wiley.com/doi/10.1002/jmv.20246.