Anders Stockmarr

anst@dtu.dk

DTU Section for

Statistics and Data Analysis

Gladsaxe, April 3 2020

Group sizes for pooled tests under imperfect testing

Assume that a test is being conducted for a disease with prevalence p. The test has a sensitivity of λ, ie.

$$P(Test\ positive|Disease\ positive) = P(T+|D+) = \lambda$$

Perfect specificity is assumed, ie.

$$P(Test\ negative|Disease\ negative) = P(T-|D-) = 1$$

I want to investigate the situation where subjects are pooled into groups ("pools"), such that a test for that a pool member is positive is carried out. In case of a positive result, all pool members are tested individually, while no individuals from a test negative group is tested. With a pool size of k, the number of tests carried out per subject is thus

- $1/k$ if the pooled test is negative;
- $1 + 1/k$ if the pooled test is positive.

The following assumptions are made:

1. Test sensitivity for pooled samples is the same as the test sensitivity for individual samples.
2. Individuals are assumed independent, thus ignoring effects of living together in case of a contagious disease etc.
3. All tests are considered independent, thus ignoring that the same sample can be part of an individual and a pooled test.

With the assumptions above, it is immediate that the probability of a disease positive pool/group is

$$q = 1 - (1-p)^k$$

This means that the total number of tests carried out per subject is

$$\frac{1}{k}P(T-) + \left(1+\frac{1}{k}\right)P(T+) =$$
$$\frac{1}{k}\big(P(D-) + (1-\lambda)P(D+)\big) + \left(1+\frac{1}{k}\right)\lambda P(D+) =$$
$$\frac{1}{k}(1-q+(1-\lambda)q) + \left(1+\frac{1}{k}\right)\lambda q =$$
$$\lambda q + \frac{1}{k} = \lambda\big(1-(1-p)^k\big) + \frac{1}{k}$$

For given λ and p, I would like to select the group size k to minimize the number of tests carried out per subject. For this, we consider k is a continuous index, and choose the vertex in the expression above which we find by differentiating. Since the value of the above expression for $k = 2$ is $\frac{1}{2} + \lambda p(2-p)$, and for $k = \infty$ is λ, any lower and lowest value among stationary points indicates a global minimum.

The expression

$$\lambda\big(1-(1-p)^k\big) + \frac{1}{k} \tag{1}$$

differentiated after k, is found as

$$\frac{\partial}{\partial k} = -\lambda\log(1-p)\,(1-p)^k - \frac{1}{k^2} = 0 \iff$$
$$-k^2\lambda\log(1-p)\,(1-p)^k = 1 \iff$$
$$ke^{\frac{k\log(1-p)}{2}} = \frac{1}{\sqrt{-\lambda\log(1-p)}} \iff$$
$$(\log(1-p)/2)ke^{k\log(1-p)/2} = \frac{\log(1-p)}{2\sqrt{-\lambda\log(1-p)}}$$

Since the function $f\colon x \mapsto xe^x$ is inverted by the two branches of the Lambert function W_0 and W_{-1}, it follows from the above that k can be found as

$$k = 2argmin\left(W_0\left(-\sqrt{-\frac{\log(1-p)}{4\lambda}}\right), W_{-1}\left(-\sqrt{-\frac{\log(1-p)}{4\lambda}}\right)\right)/\log(1-p)$$

Where 'argmin' refers to the one that results in the lowest value when inserted in the formula for the effort, formula (1). In all cases considered in this note, the primary branch W_0 resulted in the lowest effort.

Below results are tabled for a series of values of λ and p, rounded to integers. The effort value, the expected number of tests carried out per subject. For large values of k, one should be aware of that any dilution effect may compromise the assumption of similar sensitivity of individual and pooled tests.

	$p = 0.1$		$p = 0.01$		$p = 0.001$		$p = 0.0001$		$p = 0.00001$	
	k	Effort	k	Effort	k	Effort	k	Effort	k	Effort
$\lambda = 1$	4	0.59	11	0.20	32	0.06	101	0.02	317	0.006
$\lambda = 0.9$	4	0.56	11	0.19	34	0.06	106	0.02	334	0.006
$\lambda = 0.8$	4	0.53	12	0.17	36	0.06	112	0.02	354	0.006
$\lambda = 0.7$	5	0.49	13	0.16	39	0.05	120	0.02	379	0.005

Table 1: *Sizes of pools and effort values.*

A consequence of using pooled tests is that the risk of non-detection (false negative) is increased, as there are two tests that both needs to be positive, rather than one test when individual subjects are tested. For individual testing, the risk of non-detection is $1 - \lambda$. For the pooled test, the risk of non-detection is $(1 - \lambda)(1 + \lambda)$, and thus the increased relative risk (if $\lambda < 1$) is exactly λ. If, say, $\lambda = 0.8$, the amount of non-detections will increase by 80%. Below is a list of the expected number of non-detections per 10.000 individuals. As above, one should be aware that for high values of k, the assumption of similar sensitivity of pooled and individual tests may be compromised.

	$p = 0.1$		$p = 0.01$		$p = 0.001$		$p = 0.0001$		$p = 0.00001$	
	$indiv.$	$pooled$	$indiv.$	$pooled$	$indiv.$	$pooled$	$indiv.$	$pooled$	$indiv.$	$pooled$
$\lambda = 1$	0	0	0	0	0	0	0	0	0	0
$\lambda = 0.9$	100	190	10	19	1	2	0	0	0	0
$\lambda = 0.8$	200	360	20	36	2	4	0	0	0	0
$\lambda = 0.7$	300	510	30	51	3	5	0	1	0	0

Table 2: *Expected non-detections per 10.000 subjects.*

It is obvious that an imperfect test results in non-detections. Table 2 indicates that for the chosen values of sensitivity and prevalence, the number of non-detections will increase with between 50% and 90%.

Gladsaxe, April 3, 2020

Ander Stockmarr