Diagnostic test interpretation using epiR

Mark Stevenson and Evan Sergeant

12 Nov 2026

Contents

1 Diagnostic test performance measures . . . 2
1.1 Sensitivity and specificity . . . 2
1.2 Positive and negative predictive values . . . 3
2 Interpretation of diagnostic tests in series and parallel . . . 4
2.1 Parallel interpretation . . . 4
2.2 Series interpretation . . . 4
2.3 Increasing the predictive value of a positive test . . . 4
3 Estimation of the true prevalence of disease based on imperfect test results . . . 6
4 Estimate the number of false positive and false negative test results . . . 9
5 Estimate the characteristics of diagnostic tests applied at the herd (group) level . . . 10
6 Estimate herd test characteristics when samples are pooled . . . 11
7 Compute the post-test probability of disease given characteristics of a diagnostic test . . . 12
References . . . 13

A test is defined as any process or device designed to detect (or quantify) a sign, substance, tissue change, or body response in a patient. Tests include:

- Routine physical examinations.
- Questions posed to a patient (or the owner of a patient) during history taking.
- Symptoms or clinical signs.
- Laboratory findings — haematology, serology, biochemistry, histopathology.
- Post mortem findings.

If tests are to be used for decision making context, test selection should be based on its ability to alter your assessment of the probability that a disease does or does not exist. Diagnostic tests can be categorised by what they measure or detect:

- A physiological or production parameter (e.g., white blood cell count or daily milk volume).
- The infectious agent, its antigens or nucleic acid.
- The subject's immune response to the infectious agent (e.g., antibody response).

Diagnostic tests can also be classified by how their results are expressed:

- As a binary outcome (i.e., positive/negative, diseases/not diseased, pregnant/not pregnant).
- As a continuous measure (e.g., blood pressure, blood chemistry, serum enzyme concentration, serology).

For tests where the outcome is expressed as a continuous measure we often need to set a cut-off value that distinguishes positives from negatives (or normal from abnormal). This is a critical stage in the development and validation of a new diagnostic test and most, if not all, tests that produce binary outcomes (positive/negative) have, at some stage, been developed and validated by a researcher who has defined a

cut-off value to dichotomise test results that were originally continuous.

1 Diagnostic test performance measures

The two key requirements of a diagnostic test are: (1) to identify diseased individuals correctly; and (2) to identify non-diseased individuals correctly. To work out how well a diagnostic test performs, we need to compare it with a 'gold standard.' A gold standard is a test or procedure that is perfectly accurate. It diagnoses all diseased individuals that are tested and misdiagnoses none.

Sensitivity and specificity

Once samples are tested using a gold standard and the test to be evaluated, a 2 $\times$ 2 table can be constructed, allowing test performance to be quantified. The usual format is shown in Table 1.

Table 1: Table 1: A 2 × 2 diagnostic test contingency table.

	Disease+	Disease-	Total
Test+	a	b	a + b
Test-	c	d	c + d
Total	a + c	b + d	a + b + c + d

The **sensitivity** of a test is defined as the proportion of individuals with disease that test positive $p_{[T+|D+]}$. A highly sensitive test will rarely misclassify individuals with disease. Using the notation from Table 1, the formula to calculate diagnostic sensitivity is:

$$\text{Sensitivity} = \frac{a}{(a+c)}$$

The **specificity** of a test is defined as the proportion of individuals without disease that test negative $p_{[T-|D-]}$. A highly specific test will rarely misclassify individuals that are not diseased. Using the notation from Table 1, the formula to calculate diagnostic specificity is:

$$\text{Specificity} = \frac{d}{(b+d)}$$

Sensitivity and specificity are inversely related and, in the case of test results measured on a continuous scale, they can be varied by changing the cut-off value. In doing so, an increase in sensitivity will often result in a decrease in specificity, and *vice versa*.

The optimum cut-off level depends on what you're trying to achieve. If the primary objective is to find diseased individuals (i.e., to minimise the number of false negatives and accept a limited number of false positives) a test with a high sensitivity is required. If the objective is to make sure that every test positive is 'truly' diseased (i.e., minimise the number of false positives and accept a limited number of false negatives) the diagnostic test should have a high specificity.

Positive and negative predictive values

The **positive predictive value** is the proportion of individuals with a positive test that actually have the disease $p_{[D+|T+]}$. Using the notation from Table 1, the formula to calculate positive predictive value is:

$$\text{Positive predictive value} = \frac{a}{(a+b)}$$

The **negative predictive value** is the proportion of individuals with a negative test that do not have the disease $p_{[D-|T-]}$. Using the notation from Table 1, the formula to calculate negative predictive value is:

$$\text{Negative predictive value} = \frac{d}{(c+d)}$$

EXAMPLE 1

A new diagnostic test was trialed on 1586 patients. Of 744 patients that were disease positive, 670 were test positive. Of 842 patients that were disease negative, 640 were test negative. What is the diagnostic sensitivity and specificity of the new test?

```
library(epiR)

# If there are 744 disease positive individuals and 670 of them are test
# positive that means 744 - 670 = 74 are test negative. Similarly, if there
# are 842 disease negative individuals and 640 of them are test negative then
# 842 - 640 = 202 are test positive. Enter the subject counts for the 2 by 2
# table directly into R as a vector.

dat.v01 <- c(670,202,74,640)
rval.tes01 <- epi.tests(dat.v01, method = "exact", digits = 2, conf.level = 0.95)
print(rval.tes01)
#>           Outcome +    Outcome -      Total
#> Test +          670          202        872
#> Test -           74          640        714
#> Total           744          842       1586
#>
#> Point estimates and 95% CIs:
#> -------------------------------------------------------------
#> Apparent prevalence *                  0.55 (0.52, 0.57)
#> True prevalence *                      0.47 (0.44, 0.49)
#> Sensitivity *                          0.90 (0.88, 0.92)
#> Specificity *                          0.76 (0.73, 0.79)
#> Positive predictive value *            0.77 (0.74, 0.80)
#> Negative predictive value *            0.90 (0.87, 0.92)
#> Positive likelihood ratio              3.75 (3.32, 4.24)
#> Negative likelihood ratio              0.13 (0.11, 0.16)
#> False T+ proportion for true D- *      0.24 (0.21, 0.27)
#> False T- proportion for true D+ *      0.10 (0.08, 0.12)
#> False T+ proportion for T+ *           0.23 (0.20, 0.26)
#> False T- proportion for T- *           0.10 (0.08, 0.13)
#> Correctly classified proportion *      0.83 (0.81, 0.84)
```

```
#> ------------------------------------------------------------
#> * Exact CIs
```

Test sensitivity is 0.90 (95% CI 0.88 to 0.92). Test specificity is 0.76 (95% CI 0.73 to 0.79). The argument `method = exact` in `epi.tests` returns exact binomial confidence limits (Collett 1999) for sensitivity, specificity and each of the the predictive value outcomes.

2 Interpretation of diagnostic tests in series and parallel

Clinicians often request multiple tests to increase their confidence that a patient has a particular diagnosis. When multiple tests are performed and all are positive, the interpretation is straightforward: the probability of disease being present is relatively high. It is far more likely however, that some of the tests return a positive result and others will be negative. We can deal with this problem by interpreting test results in either parallel or series.

Parallel interpretation

Parallel interpretation of diagnostic test results means that when multiple tests are performed an individual is declared positive if *at least one* of the tests returns a positive result. Interpreting test results in parallel increases diagnostic sensitivity — your criteria to declare a patient as disease positive is low which means the suite of tests will detect a large proportion of individuals with disease. The trade off is that specificity and positive predictive value will be lowered.

Series interpretation

Series interpretation of diagnostic test results means that when multiple tests are performed an individual is declared positive if *all* tests return a positive result. Interpreting test results in series increases diagnostic specificity. Your criteria to declare a patient disease positive is high which means that there's a high probability that a test positive individual will be disease positive (i.e., the positive predictive value will be high). The trade off is that sensitivity is reduced.

Increasing the predictive value of a positive test

One way to increase the predictive value of a positive test is to use the test in a population where the prevalence of disease is relatively high. So, in a screening program designed to identify disease positive individuals we might target testing efforts towards those that are likely to have the disease in question (e.g., individuals of a certain age).

A second way to increase the positive predictive value is to use a more specific test or change the cutpoint of the current test to increase specificity. As specificity increases, positive predictive value increases.

A third, and very common, way to increase positive predictive value is to use more than one test and interpret the results in series. Series interpretation: (1) increases the specificity of the test procedure (reducing the risk of false positives); and (2) reduces diagnostic sensitivity (there will be more cases of disease missed).

If tests are going to be applied in series, it makes sense to first test all individuals with the test that is less expensive and/or more rapid. Once results for the first test are returned all of the test positives receive the second test. This method is called sequential testing. It provides the same results as simultaneous testing, but at a lower cost because only those that are positive to the first test are followed-up with the second test.

EXAMPLE 2

An immunofluorescent antibody (IFAT) and polymerase chain reaction (PCR) test are to be used to diagnose infectious salmon anaemia. Counts of IFA positive salmon that tested positive to the IFAT and PCR and counts of IFA negative salmon that tested positive to the IFAT and PCR are shown in Tables 2 and 3.

Table 2: Table 2: IFAT and PCR test results for 176 salmon known to be infectious salmon anaemia positive.

	IFAT+	IFAT-	Total
PCR+	134	29	163
PCR-	4	9	13
Total	138	38	176

Table 3: Table 3: IFAT and PCR test results for 176 salmon known to be infectious salmon anaemia negative.

	IFAT+	IFAT-	Total
PCR+	0	12	12
PCR-	28	534	562
Total	28	546	574

Calculate the sensitivity and specificity of the two tests and a 95% confidence interval for sensitivity and specificity of the two tests using the exact method:

```
test <- rep(c("ifat","pcr"), each = 2)
perf <- rep(c("se","sp"), times = 2)
num <- c(138,546,163,562)
den <- c(176,574,176,574)
dat.df <- data.frame(test, perf, num, den)

tmp <- epi.conf(dat = as.matrix(dat.df[,3:4]), ctype = "prevalence",
   method = "exact", N = 1000, design = 1, conf.level = 0.95)
dat.df <- cbind(dat.df, tmp); dat.df
#>   test perf num den       est     lower     upper
#> 1 ifat   se 138 176 0.7840909 0.7159082 0.8424496
#> 2 ifat   sp 546 574 0.9512195 0.9302677 0.9673444
#> 3  pcr   se 163 176 0.9261364 0.8770029 0.9600863
#> 4  pcr   sp 562 574 0.9790941 0.9637665 0.9891521

# The diagnostic sensitivity and specificity of the IFAT is 0.784 (95% CI 0.716
# to 0.842) and 0.951 (95% CI 0.930 to 0.967), respectively.

# The diagnostic sensitivity and specificity of the PCR is 0.926 (95% CI 0.877
# to 0.960) and 0.979 (95% CI 0.964 to 0.989), respectively.
```

It is known that the two tests are dependent, with positive and negative covariances of 0.036 and -0.001, respectively. What is the expected sensitivity and specificity if the two test results are interpreted in parallel?

```
# Create a matrix listing the point estimate, lower 95% confidence limit and
# upper 95% confidence limit (as columns) for the diagnostic sensitivity of
# each test (as rows):

se <- matrix(c(0.784,0.716,0.842,0.926,0.877,0.960), ncol = 3,
   byrow = TRUE); se
#>       [,1]  [,2]  [,3]
#> [1,] 0.784 0.716 0.842
```

```
#> [2,] 0.926 0.877 0.960

# Do the same for diagnostic specificity:

sp <- matrix(c(0.951,0.930,0.967,0.979,0.964,0.989), ncol = 3,
   byrow = TRUE); sp
#>       [,1]  [,2]  [,3]
#> [1,] 0.951 0.930 0.967
#> [2,] 0.979 0.964 0.989

# Diagnostic sensitivity and specificity if the tests are interpreted in parallel:
rsu.dxtest(se = se, sp = sp, covar.pos = 0.035, covar.neg = -0.001,
   tconf.int = 0.95, method = "exact", interpretation = "parallel",
   conf.int = 0.95, nsim = 999)
#> $independent
#>   test       est       low       upp
#> 1   se 0.9843199 0.9721600 0.9920185
#> 2   sp 0.9304412 0.9045442 0.9491383
#>
#> $dependent
#>   test       est       low       upp
#> 1   se 0.9493199 0.9371600 0.9570185
#> 2   sp 0.9294412 0.9035442 0.9481383
#>
#> $covar.pos
#>      test   est
#> 1 1 and 2 0.035
#>
#> $covar.neg
#>      test    est
#> 1 1 and 2 -0.001
```

Interpreting test results in parallel and accounting for the lack of test indepdendence returns a diagnostic sensitivity of 0.949 (95% CI 0.938 to 0.957) and diagnostic specificity of 0.929 (95% CI 0.906 to 0.947).

3 Estimation of the true prevalence of disease based on imperfect test results

EXAMPLE 3

Assume that from a very large herd of dairy cows, 200 animals are randomly sampled and 26 animals test positive for *Mycobacterium avium* subspecies *paratuberculosis* (MAP) infection using an ELISA, yielding an apparent prevalence of 0.13. What is the 95% confidence interval for the apparent prevalence estimate?

```
dat.m01 <- as.matrix(cbind(26,200))
epi.conf(dat.m01, ctype = "prevalence", method = "wilson", N = 1000,
   design = 1, conf.level = 0.95)
#>    est        se      lower     upper
#> 1 0.13 0.1558198 0.09028202 0.1836635
```

The apparent prevalence of MAP infection in this herd is 13 (95% CI 9 to 18) cases per 100 cows at risk. What is the true prevalence of MAP infection in this herd? Assume the sensitivity and specificity of the MAP ELISA is 0.30 and 0.96, respectively.

```
epi.prev(pos = 26, tested = 200, se = 0.30, sp = 0.96, method = "wilson",
   units = 1, conf.level = 0.95)
#> $ap
```

```
#>    est      lower     upper
#> 1 0.13 0.09028202 0.1836635
#>
#> $tp
#>         est     lower    upper
#> 1 0.3461538 0.1933924 0.552552
#>
#> $test.positive
#>   est lower upper
#> 1  26    17    36
#>
#> $true.positive
#>   est lower upper
#> 1  21    13    30
#>
#> $false.positive
#>   est lower upper
#> 1   5     1    10
#>
#> $test.negative
#>   est lower upper
#> 1 174   164   183
#>
#> $true.negative
#>   est lower upper
#> 1 126   112   139
#>
#> $false.negative
#>   est lower upper
#> 1  48    37    61
```

The estimated true prevalence of MAP infection in this herd is 35 (95% CI 19 to 55) cases per 100 animals at risk — substantially higher than the apparent prevalence estimatefsu. Importantly, the analyses show that of the 26 MAP ELISA test positive cows 5 (95% CI 1 to 10) were expected to be false positives and of the 174 MAP ELISA negative cows 48 (95% CI 37 to 61) were expected to be false negatives.

EXAMPLE 4

Cysticercus bovis is the intermediate (larval) stage of *Taenia saginata*, the human beef tapeworm. Cattle become infected with *Cysticercus bovis* by ingesting materials contaminated with tapeworm eggs originating from human faeces. Humans, the definitive host, become infected via consumption of raw or undercooked beef. In humans, adult tapeworms range from 5 mm to 15 mm in length and whilst infection may occasionally be associated with diarrhoea or abdominal pain it is usually asymptomatic.

A new PCR for *C. bovis* has been developed and you have been asked to use the new test to estimate the prevalence of infection in a herd that has had a history of *C. bovis* detections at the abbatoir.

One hundred and twenty five animals are tested using the new PCR and 50 return a positive test result. Assuming the diagnostic sensitivity of the test is 0.50 and the diagnostic specificity is 0.95, what is the apparent prevalence and true prevalence of *C. bovis* in this herd?

```
epi.prev(pos = 50, tested = 125, se = 0.50, sp = 0.95, method = "wilson",
   units = 1, conf.level = 0.95)
#> $ap
#>   est     lower     upper
#> 1 0.4 0.3183377 0.4876254
```

```
#>
#> $tp
#>         est     lower     upper
#> 1 0.7777778 0.5963059 0.9725009
#>
#> $test.positive
#>   est lower upper
#> 1  50    39    61
#>
#> $true.positive
#>   est lower upper
#> 1  49    38    59
#>
#> $false.positive
#>   est lower upper
#> 1   1     0     4
#>
#> $test.negative
#>   est lower upper
#> 1  75    64    86
#>
#> $true.negative
#>   est lower upper
#> 1  26    18    36
#>
#> $false.negative
#>   est lower upper
#> 1  49    38    59
```

The apparent prevalence of *C. bovis* in this herd is 40 (95% CI 32 to 49) cases per 100 cows at risk. The true prevalence of *C. bovis* is 78 (95% CI 60 to 97) cases per 100 animals at risk. The true prevalence is greater than the apparent prevalence because the diagnostic sensitivity of the test is relatively low.

Now consider a different herd. A total of 125 animals are tested using the new PCR and this time only 5 animals return a positive test result. What is the apparent prevalence and true prevalence of *C. bovis* in this herd?

```
epi.prev(pos = 10, tested = 125, se = 0.95, sp = 0.80,
   method = "wilson", tp.method = "rogan.gladen", bayes.variant = NULL,
   units = 100, conf.level = 0.95)
#> Warning in epi.prev(pos = 10, tested = 125, se = 0.95, sp = 0.8, method =
#> "wilson", : Apparent prevalence is less than (1 - Sp). Rogan Gladen estimate
#> of true prevalence invalid.
#> $ap
#>   est    lower    upper
#> 1   8 4.403303 14.10119
#>
#> $tp
#>   est    lower    upper
#> 1 -16 -20.7956 -7.86508
```

The true prevalence of *C. bovis* cannot be calculated using the Rogan Gladen formula (Rogan and Gladen 1978) because the apparent prevalence (0.04) is less than than (1 - diagnostic test specificity). The `epi.prev` function issues a warning to alert you to this fact. Use `tp.method = "simplified.bayes"` to estimate true prevalence using a simplied Bayesian approach (Kopacka and Fuchs 2026):

```
epi.prev(pos = 6, tested = 125, se = 0.95, sp = 0.95,
   method = "wilson", tp.method = "simplified.bayes",
   bayes.variant = "mode.hpd", units = 100, conf.level = 0.95)$tp
#>   est lower    upper
#> 1   0     0 5.490317
```

The true prevalence of *C. bovis* is estimated to be 0 (95% CI 0 to 5.5) cases per 100 cows at risk. This estimate can be improved using a full Bayesian approach. See Messam et al. (2008) for a concise introduction to analytical methods.

4 Estimate the number of false positive and false negative test results

EXAMPLE 5

On 10 July 2023 animal health officials from country A were informed by counterparts from country B that cattle from five consignments had tested PCR positive for lumpy skin disease (LSD) upon arrival in country B since mid-May 2023. In country A LSD is (thought to be) absent. In country B LSD is endemic.

On 28 July 2023, animal health officials from country B informed animal health officials from country A that: (1) 13 of 290 country A's cattle tested positive to LSD in country B post arrival; and (2) the cattle had been traced to four registered establishments in country A.

You are told that country B is sampling animals that are unvaccinated and are apparently free of LSD, coming into an area where LSD is present. Assume a diagnostic test sensitivity and specificity of 0.96 and 0.92, respectively for the LSD PCR.

The cattle that were shipped to country B were vaccinated on days 1 and 2 post arrival. One group tested positive when blood samples were tested by PCR on day 5 post arrival. The remainder were sampled and tested while they were on board the ship transporting them from country A to country B.

Assuming country A was free of LSD in July 2023, what are possible explanations for the positive test results?

The positive test results from the cattle that were vaccinated are likely to be vaccination induced. Laboratory contamination is a possible explanation for the cattle that were tested while still on board the ship (cross contamination always a possibility with PCR tests).

Assuming the LSD PCR has been used, how many false positives would you expect if 290 cattle have been tested assuming a design prevalence of 0.001?

```
epi.fpos(n = 290, pstar = 0.001, se.u = 0.96, sp.u = 0.92, conf.level = 0.95)
#> $test.pos
#>   est low upp
#> 1  23  15  33
#>
#> $true.pos
#>   est low upp
#> 1   0   0   2
#>
#> $false.pos
#>   est low upp
#> 1  23  15  33
#>
#> $test.neg
#>   est low upp
#> 1 267 257 275
#>
#> $true.neg
```

```
#>   est low upp
#> 1 267 257 275
#>
#> $false.neg
#>   est low upp
#> 1   0   0   0
```

Assuming a design prevalence of 0.001, if 290 cattle are tested using a test with sensitivity 0.96 and specificity 0.92 you can expect that 23 (95% CI 15 to 33) will return a positive test result and, of this group, all will be false positives. Assuming our estimates of sensitivity and specificity are correct and our assumption regarding design prevalence is appropriate, the finding that there were only 5 positive samples from 290 calls laboratory procedures in country B into question.

What would be your advice to country A officials based on the information provided above?

Test cattle in herds that are exporting to country B to confirm the absence of LSD. In the code that follows we assume there are 2500 cattle per herd and the within-herd LSD prevalence is in the order of 0.01: If LSD is present in a herd the within-herd prevalence should be reasonably high.

```
epi.ssdetect(N = 2500, pstar = 0.01, se = 0.96, sp = 0.92,
   interpretation = "series", covar = c(0,0), nfractional = FALSE,
   se.psu = 0.95, ss.se = 0.95)
#> $performance
#>     se   sp
#> 1 0.96 0.92
#>
#> $sample.size
#>   SUs total
#> 1 295   295
```

From a herd of 2500 cattle, 298 need to be sampled and tested to be 95% certain that, if all tests are negative, LSD is not present.

5 Estimate the characteristics of diagnostic tests applied at the herd (group) level

EXAMPLE 6

While pregnancy testing a group of heifers the attending herd manager asks you to examine a favourite cow because she believes the animal 'just isn't doing right'. You examine the cow to find, apart from a relatively low body condition score compared with other members of the herd, no clinically abnormal findings. On questioning the herd manager you learn that this cow was purchased three years ago. The herd manager cannot remember details of the herd from which the cow originated. You decide to take a blood sample for full blood count and biochemistry.

A few days later full blood count and biochemistry results have been returned from the lab. Even though Johne's disease was not mentioned in the clinical history, the laboratory elected to run a Johne's IDEXX ELISA on the blood sample. The Johne's IDEXX ELISA was positive. There is no previous history of Johne's disease in this herd.

As well as carrying out an additional round of testing on this cow the herd manager asks you to test every cow in the herd ($n = 584$) using the IDEXX ELISA. IDEXX report the sensitivity and specificity of their Johne's ELISA as 0.850 to 0.900 and 0.950 to 1.000, respectively. A publication by McCormick et al. (2010) reports the sensitivity and specificity of the same ELISA as 0.514 and 0.993, respectively.

Assuming the diagnostic sensitivity and specificity of the Johne's serum IDEXX ELISA is 0.514 and 0.993, respectively, what is the probability that you will detect Johne's disease in this herd if (say) 120 animals are tested, the design prevalence is 0.05 and you use $k = 1$ reactor to declare the herd positive?

```
rsu.sssep.rsfreecalc(N = 584, pstar = 0.05, mse.p = 0.95, msp.p = 0.95,
   se.u = 0.514, sp.u = 0.993, method = "hypergeometric",
   max.ss = 32000)$summary
#>     n   N c pstar         p1      se.p      sp.p
#> 1 306 584 5  0.05 0.04961991 0.9503801 0.9781908
```

If a random sample of 306 cows is taken from a herd of 584 and ≥ 5 reactors are found the probability that the herd is Johne's disease positive at a prevalence of 5% is 0.0496.

Your herd manager elects to sample and test every cow in the herd using the serum IDEXX ELISA. In total 584 cows are tested with 11 returning a positive result. What is the probability of there being 11 reactors in a population where the prevalence of Johne's is 5% or less?

```
rsu.sep.rsfreecalc(N = 584, n = 584, c = 11, pstar = 0.05, se.u = 0.514, sp.u = 0.993)
#> [1] 0.9953289
```

If 584 animals are tested from a population of 584 and a positive herd test result is defined as 11 or more individuals returning a positive test, the probability of detecting disease if the population is diseased at a prevalence of 5% is 0.995.

You conclude that the Johne's disease positive and the within-herd prevalence is low (i.e., less than 5.0%).

6 Estimate herd test characteristics when samples are pooled

EXAMPLE 7

You want to test dairy herds for Johne's disease using faecal culture which has a sensitivity and specificity of 0.647 and 0.981, respectively.

You pool faecal samples from five cows together and collect six pooled samples per herd. Assuming the prevalence of Johne's disease in the herd is 0.10 and homogenous mixing, what is the herd level sensitivity and specificity? Assume that pooling of faeces results results in dilution of *Mycobacterium avium* subspecies *paratuberculosis* in the pooled sample.

```
epi.pooled(Se = 0.647, Sp = 0.981, P = 0.10, m = 5, r = 6, dilution = TRUE)
#> $Se
#> [1] 0.647
#>
#> $Sp
#> [1] 0.981
#>
#> $PlSe
#> [1] 0.6941491
#>
#> $PlSp
#> [1] 0.9085421
#>
#> $HSe
#> [1] 0.9160345
#>
#> $HSp
#> [1] 0.5624323
#>
#> $HAPneg
#> [1] 0.4375677
```

Herd level sensitivity is 0.916, herd level specificity is 0.562. Sensitivity at the herd level is increased using the pooled sampling approach. Herd level specificity is decreased.

EXAMPLE 8

To confirm your country's disease freedom status you intend to use a test applied at the herd level. The test is expensive so you decide to pool the samples taken from individual herds. How many pooled samples of size 5 are required to be 95% confident that you will have detected disease if 1% of herds are disease-positive? Assume a diagnostic sensitivity and specificity of 0.90 and 0.95 for the pooled testing regime.

```
rsu.sssep.rspool(k = 5, pstar = 0.01, pse = 0.90, psp = 0.95, se.p = 0.95)
#> [1] 32
```

A total of 32 pools (each comprised a samples from 5 herds) need to be tested.

If you decide to collect 60 pools, each comprised of samples from five herds what is the sensitivity of disease detection assuming a design prevalence of 0.01 and the sensitivity and specificity of the pooled test equals 1.0?

```
rsu.sep.rspool(r = 60, k = 5, pstar = 0.01, pse = 1, psp = 1)
#> $se.p
#> [1] 0.9509591
#>
#> $sp.p
#> [1] 1
```

This testing regime returns a population-level sensitivity of disease detection of 0.95. Repeat these calculations assuming the sensitivity of the pooled test equals 0.90.

```
rsu.sep.rspool(r = 60, k = 5, pstar = 0.01, pse = 0.90, psp = 1)
#> $se.p
#> [1] 0.933242
#>
#> $sp.p
#> [1] 1
```

If the sensitivity of the pooled test equals 0.90 the population-level sensitivity of disease detection is 0.93. How can we improve population-level sensitivity? Answer: Include more pools in the study.

```
rsu.sep.rspool(r = 70, k = 5, pstar = 0.01, pse = 0.90, psp = 1)
#> $se.p
#> [1] 0.9574806
#>
#> $sp.p
#> [1] 1
```

Testing 70 pools, each comprised of samples from 5 herds returns a population-level sensitivity of disease detection of 0.95.

7 Compute the post-test probability of disease given characteristics of a diagnostic test

EXAMPLE 9

You're working in a small animal practice in southern Australia. You're presented with a mature male desexed labrador recently acquired from a humane shelter with general debility, cachexia and a chronic cough. The dog's medical history is unknown but there was mention by humane shelter staff that the dog may have accompanied its previous owners who moved from the north of Australia the previous year.

You decide to test the dog for heartworm (*Dirofilaria immitis*) using a commercially available ELISA antigen test. The test comes back positive. You review the literature on commercial antigen kits for the detection of heartworm in dogs (Henry et al. 2018) to find that the diagnostic sensitivity of the ELISA you're using is

0.9750 (95% CI 0.9426 to 0.9918) and the diagnostic specificity is 0.9400 (0.8345 to 0.9875). The prevalence of heartworm in dogs that are not on prophylactic treatment in northern Australia is thought to be in the order of 35%. What is the probability that this dog has heartworm given the positive test result?

```
epi.nomogram(pretest.ppos = 0.35, se = 0.9750, sp = 0.9400,
   tconf.int = 0.95, method = "exact", verbose = TRUE,
   cred.int = 0.95, nsim = 999)$postest.ppos
#>         est
#> 1 0.8974359
```

Given the positive test result the probability that this dog is heartworm positive is 90%.

Now assume the dog has always been in southern Australia where the prevalence of heartworm is much lower (assume 0.001). What is the probability that this dog has heartworm given the positive test result?

```
epi.nomogram(pretest.ppos = 0.001, se = 0.9750, sp = 0.9400,
   tconf.int = 0.95, method = "exact", verbose = TRUE,
   cred.int = 0.95, nsim = 999)$postest.ppos
#>          est
#> 1 0.01600591
```

Given the positive test result the probability that this dog is heartworm positive is only 2%. For this scenario you'd be wise to carry out additional testing before starting treatment.

To account for uncertainty in diagnostic sensitivity and specificity we provide `epi.nomogram` with the lower and upper confidence limits for sensitivity and specificity:

```
epi.nomogram(pretest.ppos = 0.001, se = c(0.9750,0.9426,0.9918),
   sp = c(0.9400,0.8345,0.9875), lratio.pos = NA, lratio.neg = NA,
   tconf.int = 0.95, method = "exact", verbose = FALSE, cred.int = 0.95,
   nsim = 999)
#> If the pre-test probability of being outcome positive is 0.0010 and the test is positive, the post-t
#> If the pre-test probability of being outcome positive is 0.0010 and the test is negative, the post-t
```

If the pre-test probability of being outcome positive is 0.001 and the test is positive, the post-test probability of being outcome positive is 0.02 (95% CrI 0.0059 to 0.05). Again, for this scenario you'd be wise to carry out additional testing before starting treatment.

References

Caraguel, CGB, and R Vanderstichel. 2013. "The two-step Fagan's nomogram: ad hoc interpretation of a diagnostic test result without calculation." *Evidence Based Medicine* 18: 125–28.

Christensen, J, and IA Gardner. 2000. "Herd-level interpretation of test results for epidemiologic studies of animal diseases." *Preventive Veterinary Medicine* 45: 83–106.

Collett, D. 1999. *Modelling Binary Data*. Boca Raton Florida: Chapman & Hall/CRC.

Haskins, R, PG Osmotherly, F Tuyl, and DA Rivett. 2014. "Uncertainty in clinical prediction rules: The value of credible intervals." *Journal of Orthopaedic and Sports Physical Therapy* 44: 85–91.

Henry, LG, KJ Brunson, HS Walden, N Wenzlow, SE Beachboard, L. Barr K, and MT Long. 2018. "Comparison of six commercial antigen kits for detection of *Dirofilaria immitis* infections in canines with necropsy-confirmed heartworm status." *Veterinary Parasitology* 254: 178–82.

Kopacka, I, and K Fuchs. 2026. "Overcoming limitations of the Rogan-Gladen correction: A closed-form solution to a simplified Bayesian method for true prevalence estimation." *Preventive Veterinary Medicine* 253: 253.

McCormick, J, S Chebanier, N Djuranovic, and L Estey. 2010. "Validation of a new IDEXX Johne's antibody ELISA for use on milk samples." In *43rd Annual Conference Proceedings American Association of Bovine Practitioners*, 43:241. Albuquerque, New Mexico: American Association of Bovine Practitioners.

Messam, LL, AJ Branscum, MT Collins, and IA Gardner. 2008. "Frequentist and Bayesian approaches to prevalence estimation using examples from Johne's disease." *Animal Health Research Reviews* 9: 1–23.

Mossman, D, and JO Berger. 2001. "Intervals for posttest probabilities: A comparison of 5 methods." *Medical Decision Making* 21: 498–507.

Rogan, WJ, and B Gladen. 1978. "Estimating prevalence from results of a screening test." *American Journal of Epidemiology* 107: 71–76.