simple simulation and power analysis for prs testing

Joshua Schmidt

09/02/2022

GRADE cohort simulation and power tests.

Some simple functions with usage examples to derive power calculations of PRS vs outcome, and pr affected across PRS quantiles.

Example: generate a GRADE like data set.

This study has recruited ~ 1000 individuals, with ~ population affected rates 6-10%.

Defining deciles from PRS, the odds ratio for glaucoma in the top 10% versus bottom 10% is 14.9, or 25 for HTG. We can use the base rate, quantile definiton and OR to simulate a dataset.

A simulation

```
pacman::p_load('data.table','kableExtra','pwr','ggplot2','tidyverse',install=T,update = F)
source('R/power_functions.R')
```

```
n_cohort <- 1000
n_quantile <- 100
quantiles <- c(0.1,0.9)
OR <- 14.9
glaucoma_rate <- 6/100
set.seed(42)
data <- simulate_GRADE(n_ind = n_cohort, n_group = n_quantile,
                       OR = OR,base_prevalence = glaucoma_rate,
                       cuts = quantiles)
```

Table 1: Example GRADE cohort simulation

prs	glaucoma	label
1.6895	1	top
1.4665	0	top
1.8664	0	top
-0.2486	0	middle
-0.2485	0	middle
-1.6626	0	bottom
-1.6025	0	bottom

We can then check prevalence in cohort, which thankfully we find is 5.9% in this simulation. This is natural sampling variance around the 6% prevalence we required.

Table 2: Prevalence of glaucoma cases in simulation

glaucoma	N
0	941
1	59

Simulated association between prs and glaucoma

We can also check the association between PRS and glaucoma (NB: exp(coefficient) is NOT comparable to the quantile OR, by definition!).

Table 3: glaucoma prs regression, full sample

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.9066	0.1508	-19.2721	0
prs	0.5929	0.1379	4.2994	0

The OR is not comparable, because the model prs coefficient is per unit of prs, which being defined as a standard normal is one standard deviation. In contrast, defining odds ratios between the top and bottom deciles is equivalent to 3.509 sd units. i.e. the mean prs of the top minus the mean prs of the bottom. Therefore, the relationship between model coefficient and decile OR is:

$$OR = exp(coefficent) \times (\mu_{top} - \mu_{bottom})$$

Using these gives an estimate of the top bottom OR of 6.35.
This is an underestimate, likely due to the small number of samples.

For completeness, regression only on the subsets (100 from top, bottom, and middle of PRS distribution), also finds a significant association (at least in this simulation).

Table 4: glaucoma prs regression, 100 per group

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.9790	0.3038	-9.8062	0.0000
prs	0.5721	0.1870	3.0597	0.0022

Within study Odds Ratios

Finally, lets look at some realised odds ratios. These are to ORs for this within sample comparison:

Table 5: realised OR

comparison	OR
topVsBottom	4.4091
topVsMiddle	2.3430

The top vs. bottom OR is likely very inaccurate in GRADE, because we would expect very few cases in the bottom decile.

For giggles, lets simulate 1 million participants to check this intuition!

Table 6: realised OR, cohort of one million

comparison	OR
topVsBottom	14.9867
topVsMiddle	3.5911

Huzzah, we were right, and we nicely recapture our simulated OR! (Of course this is tautological, it rather shows that the simulation is just doing what it is supposed to!!)

Power analysis

Now that we have functions that can simulate a realistic data set, we can use them to test the power of different statistical tests within GRADE.

From the protocol, it seems there are distinct tests that need to be assessed.

regression

The protocol mentions in the section "Statistical analyses":

```
For association analysis, logistic or linear regression will be used, including covariates
to account for confounding variables as clinically and statistically appropriate.
```

This could either be a regression of glaucoma case/suspect status on prs, or glaucoma case/suspect status on label ("top","bottom","middle").

The later seems most relevant. This can be accomplished by treating "top" as the reference level, and coding dummy variables indicating status in bottom or middle deciles: i.e. glaucoma ~ bottom + middle

Table 7: power to detect top vs bottom ot middle in GRADE

prevalence	OR	bottom	middle
6%	14.9	0.696	0.598
6%	21	0.61	0.757
10%	14.9	0.9	0.834
10%	21	0.874	0.932
30%	14.9	1	0.984
30%	21	0.999	0.995

This has moderate to high power. Note to that because of variance in case numbers between deciles, significance can be achieved for top vs bottom and middle, or only bottom or only middle - which explains why sometimes middle has greater power than the comparison to bottom. Note also that the p-values were corrected for 2 tests, so these results are the power after multiple testing correction.

Next using the prs itself as the predictor.

Table 8: power to detect glaucoma vs prs in GRADE

prevalence	OR	power
6%	14.9	0.992
6%	21.0	0.999
10%	14.9	1.000
10%	21.0	1.000
30%	14.9	1.000
30%	21.0	1.000

Thus, there is very high power to find this association.

Proportion of cases

The protocol also mentions, in "Study outcomes" that: `The primary outcome will be assessing the prevalence of glaucoma and AMD between the bottom decile, middle 80% and top decile of both respective PRS spectra.`

To me that seems to be a test of proportions. That could either be a test of the homogeneity of proportions, using a χ^2 test, or three Z score tests testing each of the pairwise differences in proportions.

As the Z score tests would be done with or without a χ^2 test, i focus on those. p-values are two-tailed and corrected for n = 3 tests.

Table 9: power to detect differences in pr of glaucoma cases across prs deciles in GRADE

prevalence	OR	top_bottom	top_middle	middle_bottom
6%	14.9	0.938	0.489	0
6%	21	0.97	0.618	0
10%	14.9	0.998	0.71	0.221
10%	21	0.999	0.834	0.271
30%	14.9	1	0.965	0.839
30%	21	1	0.993	0.915

These simulation results show that there is » 90% power to detect a difference in prevalence rates between top and bottom deciles. There is also moderate to high power to detect differences between the top and bottom, but is clearly contingent on assumptions of base rate of glaucoma and the phenotype tested, with higher power for HTG versus glaucoma.

We can also use simulations to get an estimate of the proportion of cases+suspects in each prs decile.
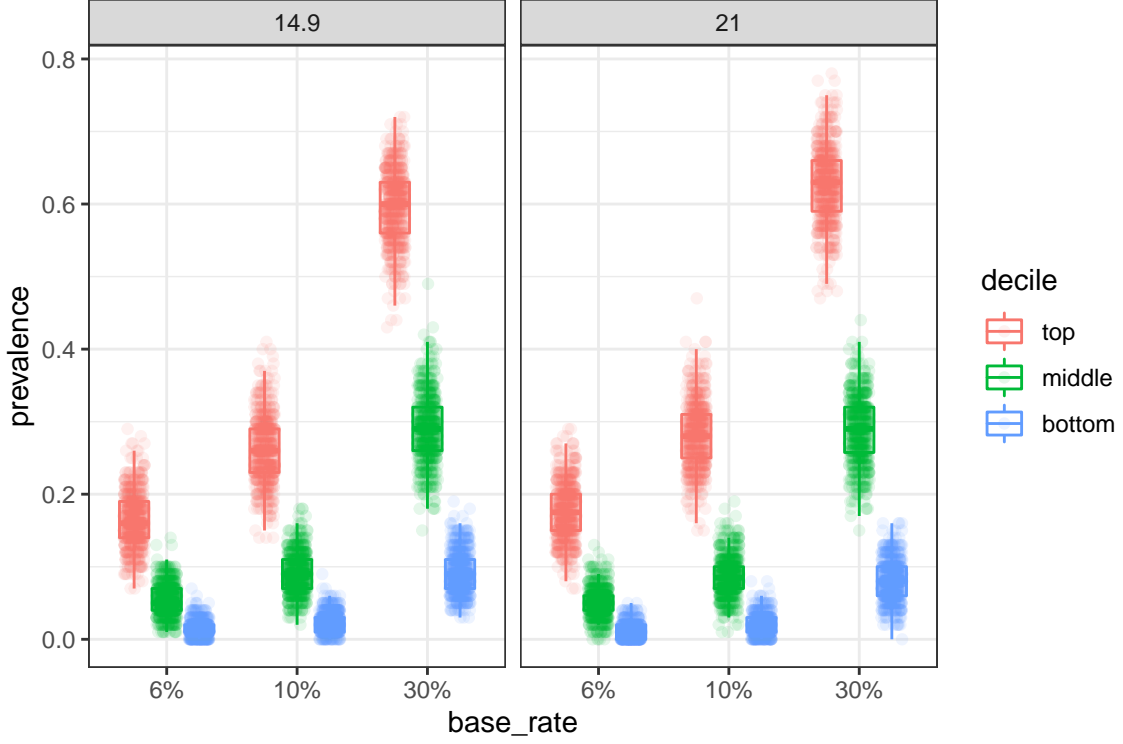
Figure 1: Distribution of prs decile glaucoma prevalance

Table 10: mean expected prevelance per decile, by base rate of glaucoma and OR top vs. bottom

base_rate	decile	OR	prevalence
6%	top	14.9	16.49
6%	middle	14.9	5.54
6%	bottom	14.9	1.41
6%	top	21	17.65
6%	middle	21	5.06
6%	bottom	21	1.05
10%	top	14.9	26.20
10%	middle	14.9	9.02
10%	bottom	14.9	2.32
10%	top	21	28.22
10%	middle	21	8.65
10%	bottom	21	1.95
30%	top	14.9	59.65
30%	middle	14.9	29.02
30%	bottom	14.9	9.26
30%	top	21	62.62
30%	middle	21	28.72
30%	bottom	21	8.04

While powerful, these tests cannot correct for other predictors e.e. sex, age etc. Which should perhaps mean the generalised linear model framework is preferred.

Other tests

In the description of statistical power, it appears that a t-test will be used:

`Based on the combined estimated incidence of glaucoma plus glaucoma suspect cases in each group (i.e. 30% in the top decile vs 9% in the bottom decile), the current sample size will yield >80% power (alpha=0.05) to detect a significant difference between the top and bottom deciles of the PRS distribution (two-sided t-test).` From this description I am not quite sure of the power estimate.

Given the stated proportions of cases+suspects, sample sizes and alpha, I checked the power of a χ^2, t-test, and Z-test, though of course, χ^2 and Z-test are fundamentally the same test.

Table 11: power of Chi.sq, t- and z- tests, given proportions

method	power
chi.sq	0.9462
t-test	0.9608
z-test of proportions	0.9731

These three tests are essentially equally, highly powered.

Question

Note too that the quoted differences in prevalence between top and bottom deciles imply OR of 3.59-4.33:

$(0.1/0.9)/(0.03/0.97) = 3.59$ $(0.3/0.7)/(0.09/0.91) = 4.33$

These OR are similar to the one reported in Craig et. al. 2020 for top decile versus the rest. For example:

`While comparing the top and bottom deciles shows the dose-response across deciles, one can also consider the risk in the high-PRS individuals versus all others; when this is done in the ANZRAG cohort, the OR is 4.2 and 8.5 in the top 10 and 1%, respectively, of individuals versus all remaining individuals (Supplementary Table 9).`

Does this mean that 3% diagnosed and 9% diagnosed+suspects is for the middle 80%, rather than the bottom decile?

From the simulations, assuming a base rate of 6% and OR of 14.9 top vs bottom, the prevalences should be, top to bottom, 16.5%, 5.54% and 1% (realised OR ~ 14, 3.4). For 10% base rate this is 26.20%, 9.02% and 2.32% (realised OR ~ 14.9, 3.6). This later one is closest to the figures given in Georgie's GRADE paper.