

...Introduction

Risk ratios (RR) and odds ratios (OR) are both used to analyze factors associated with increased risk of disease. They are different in that a risk ratio is the ratio of two risks and an odds ratio is the ratio of two odds. Risks and odds are calculated differently. For example, with one toss of a fair die, the risk of rolling a 1 is $1/6 = 0.167$. The odds of rolling a 1 are $1/5 = 0.200$. In textbook terms, risk is successes divided by total trials and odds are successes divided by failures. In this example rolling a 1 is a success and rolling 2 through 5 is a failure.

In R, glm (generalized linear model) is used to compute adjusted risk ratios (ARR's) and adjusted odds ratios (AOR's). Adjusting eliminates the influence of correlations between factors on the ratios. In statistical terms, adjusting controls for confounding. In addition to confounding, the RR's and OR's can be influenced by interaction which is also called effect modification. This occurs when the ratio for one group is different from the ratio for another group. For example, non-smokers who work with asbestos might have 4.0 times the risk of getting lung cancer compared to non-smokers who do not work with asbestos. This is for non-smokers. For smokers, this risk ratio might be much larger at 8.0, meaning smokers who work with asbestos have 8.0 times the risk of getting lung cancer compared to smokers who do not work with asbestos. In this case there is interaction between smoking and asbestos exposure because the RR for non-smokers is 4.0 and the RR for smokers is 8.0. Alternatively, it could be said that smoking modifies the effect of the asbestos exposure on lung cancer.

For a given data set, the AOR's and ARR's will almost always agree in terms of statistical significance. The AOR's and ARR's will be different but the p values will be close. This is not true for interaction terms. Often, using the same data, an analysis of interaction using OR's will indicate significant interaction while the RR's will indicate no interaction. This phenomenon is explained very well in the article "Distributional interaction: Interpretational problems when using incidence odds ratios to assess interaction" by Ulka B Campbell, Nicolle M Gatto and Sharon Schwartz. (Epidemiologic Perspectives & Innovations 2005, 2:1 doi:10.1186/1742-5573-2-1) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1079911/>.

The authors described the problem for RR's and OR's and included mathematical explanations. The purpose of this article is to show how this problem occurs when using adjusted RR's and adjusted OR's with the R glm function to assess interaction.

Creating a test data set

The first step is to create a test data set. The R code below produces a data set of 1,000,000 lines with simulated variables for two risk factors, a grouping variable and an outcome variable. This is designed to simulate a situation where there are two groups of people that have different risks of disease. There are also two risk factors and these are designed to have the same risk ratios for the disease for simulated people in both groups.

```
>
> rm(list=ls(all=TRUE))
>
> nx=1000000
>
> set.seed(523)
>
> x1=runif(nx)
> x2=runif(nx)
> grpx=runif(nx)
> dep1=runif(nx)
>
> rf1=ifelse(x1 < 0.20,1,0)
> rf2=ifelse(x2 < 0.30,1,0)
> grp=ifelse(grpx < 0.5,1,0)
>
```

```
> d1=as.data.frame(cbind(rf1,rf2,grp,depl))
```

The R output above creates the two risk factor variables (rf1 and rf2) and the group variable (grp). The proportions are 0.20, for rf1, 0.30 for rf2 and 0.50 for grp 0 and grp 1. This simulates a scenario where the risk factors are not too rare and the two groups are similar to a male/female grouping. The proportions will vary slightly due to randomizing with the runif R function.

First Simulation

```
> #Sim 1
>
> attach(d1,warn.conflict=F)
>
> xg12f=(grp*100)+(rf1*10)+rf2
>
> disprop=0.01
>
> disease=ifelse(xg12f == 0 & depl < disprop,1,0)
> disease=ifelse(xg12f == 1 & depl < (1.5*disprop),1,disease)
> disease=ifelse(xg12f == 10 & depl < (1.7*disprop),1,disease)
> disease=ifelse(xg12f == 11 & depl < (1.5*1.7*disprop),1,disease)
>
> disease=ifelse(xg12f == 100 & depl < (2*disprop),1,disease)
> disease=ifelse(xg12f == 101 & depl < (3*disprop),1,disease)
> disease=ifelse(xg12f == 110 & depl < (3.4*disprop),1,disease)
> disease=ifelse(xg12f == 111 & depl < (5.1*disprop),1,disease)
>
> sum(disease)/nrow(d1)
[1] 0.019679
>
```

The block of output above creates a new variable, xg12f, that combines the three variables, grp, rf1 and rf2. This makes it easier to create the outcome variable named disease. Each value of xg12f defines a unique combination of the variables grp, rf1 and rf2. The disease proportion (disprop) is initially set to 0.01 and then the series of ifelse statements increases the disease proportion by multiples of disprop based on the risk factors and group. This simulates a situation where the disease is rare for people with no risk factors (xg12f = 0) and then increases with various combinations of risk factors and group. For example, a simulated person in group 0 with risk factor 1 but not risk factor 2 would have a value for xg12f of 10. The third ifelse statement sets the disease proportion for simulated persons in this category to 1.7 times the disprop value. In general, simulated persons with risk factors have higher disease proportions and simulated persons in group 1 have higher disease proportions. Simulated persons in group 1 with both risk factors (xg12f = 111) have the highest multiple at 5.1 times the proportion of disease for simulated persons in group 0 with no risk factors (xg12f = 0). As shown above, the overall proportion of simulated persons with the disease is 0.019679.

```
>
> r1=glm(disease~rf1+rf2+grp+rf1*rf2+rf1*grp+rf2*grp,
+ family=binomial(link=log))
> summary(r1)Call:
glm(formula = disease ~ rf1 + rf2 + grp + rf1 * rf2 + rf1 * grp +
    rf2 * grp, family = binomial(link = log))Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.3265  -0.2001  -0.2001  -0.1422   3.0328 Coefficients:
              Estimate Std. Error z value Pr(>|z|)
```

```

(Intercept) -4.598839    0.018000 -255.492    <2e-16 ***
rf1          0.500622    0.029614   16.905    <2e-16 ***
rf2          0.401599    0.026798   14.986    <2e-16 ***
grp          0.677464    0.021601   31.362    <2e-16 ***
rf1:rf2      -0.004331    0.031359   -0.138     0.890
rf1:grp       0.032458    0.032800    0.990     0.322
rf2:grp       0.032862    0.030670    1.071     0.284
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 (Dispersion
parameter for binomial family taken to be 1)    Null deviance: 193574  on
999999  degrees of freedom
Residual deviance: 189361  on 999993  degrees of freedom
AIC: 189375 Number of Fisher Scoring iterations: 7> x1=exp(confint.default(r1))
> rr=exp(coef(r1))
> round(cbind(rr,x1),digits=2)
              rr 2.5 % 97.5 %
(Intercept) 0.01  0.01  0.01
rf1          1.65  1.56  1.75
rf2          1.49  1.42  1.57
grp          1.97  1.89  2.05
rf1:rf2      1.00  0.94  1.06
rf1:grp      1.03  0.97  1.10
rf2:grp      1.03  0.97  1.10
>

```

The R output above uses the R function glm to produce ARR's with 95% confidence intervals and p values. In this glm the family is binomial with link = log. This produces coefficients that are the logs of the ARR's and these are used to produce the table of ARR's and 95% CI that appears below the glm output. It should be noted that the intercept is the log of the disease proportion for simulated persons that have no risk factors. In the table of ARR's the data for the intercept is not the ARR but rather the proportion of disease for simulated persons with no risk factors, 0.01.

As expected, the ARR's are statistically significant for rf1, rf2 and grp. Also as expected, the interaction terms rf1:rf2, rf1:grp and rf2:grp, are not statistically significant.

Next are the AOR's.

```

>
> r1=glm(disease~rf1+rf2+grp+rf1*rf2+rf1*grp+rf2*grp,
+ family=binomial(link=logit))
> summary(r1)Call:
glm(formula = disease ~ rf1 + rf2 + grp + rf1 * rf2 + rf1 * grp +
    rf2 * grp, family = binomial(link = logit))Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.3265  -0.2000  -0.2000  -0.1423   3.0326 Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.588289    0.018205 -252.037    <2e-16 ***
rf1          0.505839    0.030181  16.760    <2e-16 ***
rf2          0.405558    0.027242  14.887    <2e-16 ***
grp          0.686728    0.021929  31.316    <2e-16 ***
rf1:rf2      0.002334    0.032418    0.072     0.943
rf1:grp      0.042153    0.033593    1.255     0.210
rf2:grp      0.040418    0.031327    1.290     0.197
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 (Dispersion

```

```

parameter for binomial family taken to be 1)      Null deviance: 193574  on
999999  degrees of freedom
Residual deviance: 189361  on 999993  degrees of freedom
AIC: 189375Number of Fisher Scoring iterations: 7> x1=exp(confint.default(r1))
> or=exp(coef(r1))
> round(cbind(or,x1),digits=2)
              or 2.5 % 97.5 %
(Intercept) 0.01  0.01  0.01
rf1          1.66  1.56  1.76
rf2          1.50  1.42  1.58
grp          1.99  1.90  2.07
rf1:rf2       1.00  0.94  1.07
rf1:grp       1.04  0.98  1.11
rf2:grp       1.04  0.98  1.11
>

```

The R output above is similar to the previous glm output but in this glm the family is binomial with link = logit instead of log. This produces coefficients that are the logs of the AOR's and these are used to produce the table of AOR's and 95% CI. The line for the intercept in the table is not AOR's but is the odds of disease for simulated persons with no risk factors. As with ARR's, the AOR's are statistically significant for rf1, rf2 and grp. The interaction terms rf1:rf2, rf1:grp and rf2:grp, are not statistically significant which also agrees with the glm output for the ARR's.

Second Simulation

```

> #sim 2
>
> disprop=0.05
>
> disease=ifelse(xg12f == 0 & dep1 < disprop,1,0)
> disease=ifelse(xg12f == 1 & dep1 < (1.5*disprop),1,disease)
> disease=ifelse(xg12f == 10 & dep1 < (1.7*disprop),1,disease)
> disease=ifelse(xg12f == 11 & dep1 < (1.5*1.7*disprop),1,disease)
>
> disease=ifelse(xg12f == 100 & dep1 < (2*disprop),1,disease)
> disease=ifelse(xg12f == 101 & dep1 < (3*disprop),1,disease)
> disease=ifelse(xg12f == 110 & dep1 < (3.4*disprop),1,disease)
> disease=ifelse(xg12f == 111 & dep1 < (5.1*disprop),1,disease)
>
> sum(disease)/nrow(d1)
[1] 0.09847
>

```

The R output above uses the same data file but the disprop variable is increased from 0.01 to 0.05. This reflects a situation where the disease proportion is higher. As shown above, the overall proportion of simulated persons with the disease is 0.09847.

```

>
> r1=glm(disease~rf1+rf2+grp+rf1*rf2+rf1*grp+rf2*grp,
+ family=binomial(link=log))
> summary(r1)Call:
glm(formula = disease ~ rf1 + rf2 + grp + rf1 * rf2 + rf1 * grp +
      rf2 * grp, family = binomial(link = log))Deviance Residuals:
      Min       1Q   Median       3Q      Max
-0.7718  -0.4588  -0.4168  -0.3207   2.4468 Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.9935084   0.0078756 -380.099  <2e-16 ***
rf1          0.5071062   0.0127189  39.870  <2e-16 ***

```

```

rf2          0.4029193  0.0115901  34.764  <2e-16 ***
grp          0.6901061  0.0093582  73.743  <2e-16 ***
rf1:rf2      0.0007967  0.0130359   0.061   0.951
rf1:grp      0.0155518  0.0139306   1.116   0.264
rf2:grp      0.0206406  0.0131138   1.574   0.115
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 (Dispersion
parameter for binomial family taken to be 1)    Null deviance: 643416  on
999999  degrees of freedom
Residual deviance: 620356  on 999993  degrees of freedom
AIC: 620370Number of Fisher Scoring iterations: 6> x1=exp(confint.default(r1))
> rr=exp(coef(r1))
> round(cbind(rr,x1),digits=2)
              rr 2.5 % 97.5 %
(Intercept) 0.05  0.05   0.05
rf1          1.66  1.62   1.70
rf2          1.50  1.46   1.53
grp          1.99  1.96   2.03
rf1:rf2      1.00  0.98   1.03
rf1:grp      1.02  0.99   1.04
rf2:grp      1.02  0.99   1.05
>

```

The R output above is glm output with family = binomial and link = log. As before, the coefficients are the logs of ARR's and are used to produce the table of ARR's. Also as before the line for the intercept in the table is not ARR's but is the prevalence of disease for simulated persons with no risk factors. In this case it is 0.05 in contrast to the earlier glm result with the previous data file which was 0.01. As expected, this output shows statistically significant ARR's for the risk factors and no statistically significant results for the interaction terms.

```

>
> r1=glm(disease~rf1+rf2+grp+rf1*rf2+rf1*grp+rf2*grp,
+ family=binomial(link=logit))
> summary(r1)Call:
glm(formula = disease ~ rf1 + rf2 + grp + rf1 * rf2 + rf1 * grp +
    rf2 * grp, family = binomial(link = logit))Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7701  -0.4586  -0.4157  -0.3210   2.4459
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.939696   0.008344 -352.305 < 2e-16 ***
rf1          0.534243   0.014054  38.012 < 2e-16 ***
rf2          0.423218   0.012628  33.515 < 2e-16 ***
grp          0.740304   0.010119  73.158 < 2e-16 ***
rf1:rf2      0.042163   0.015615   2.700  0.00693 **
rf1:grp      0.072082   0.015824   4.555  5.23e-06 ***
rf2:grp      0.063938   0.014672   4.358  1.31e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 (Dispersion
parameter for binomial family taken to be 1)    Null deviance: 643416  on
999999  degrees of freedom
Residual deviance: 620356  on 999993  degrees of freedom
AIC: 620370Number of Fisher Scoring iterations: 5> x1=exp(confint.default(r1))
> or=exp(coef(r1))
> round(cbind(or,x1),digits=2)
              or 2.5 % 97.5 %
(Intercept) 0.05  0.05   0.05

```

```

rf1      1.71  1.66  1.75
rf2      1.53  1.49  1.57
grp      2.10  2.06  2.14
rf1:rf2   1.04  1.01  1.08
rf1:grp   1.07  1.04  1.11
rf2:grp   1.07  1.04  1.10
>

```

The R output above is glm output with family = binomial and link = logit. The coefficients are the logs of AOR's and are used to produce the table of AOR's. The line for the intercept in the table is not AOR's but is the odds of disease for simulated persons with no risk factors. As expected, this output shows statistically significant AOR's for the risk factors. However, in contrast to the results with the ARR's, this glm output shows statistically significant results for the interaction terms. In short, with this data set, the interaction terms are not statistically significant when using ARR's but they are statistically significant when using AOR's.

Lastly, the value of the disprop variable is increased again, this time to 0.10. The output below shows the creation of the new disease variable and shows the overall proportion of simulated persons with the disease is increased to 0.196368.

The output below is the, by now familiar, glm output for ARR's and AOR's. With the increased proportion of disease, the ARR's still show the interaction terms as not statistically significant. But the AOR's show the interaction terms as having p values less than 2×10^{-16} which is pretty low. Once again, the ARR's show no interaction while the AOR's indicate strong interaction.

Third Simulation

```

>
> #sim 3
>
> disprop=0.10
>
> disease=ifelse(xg12f == 0 & dep1 < disprop,1,0)
> disease=ifelse(xg12f == 1 & dep1 < (1.5*disprop),1,disease)
> disease=ifelse(xg12f == 10 & dep1 < (1.7*disprop),1,disease)
> disease=ifelse(xg12f == 11 & dep1 < (1.5*1.7*disprop),1,disease)
>
> disease=ifelse(xg12f == 100 & dep1 < (2*disprop),1,disease)
> disease=ifelse(xg12f == 101 & dep1 < (3*disprop),1,disease)
> disease=ifelse(xg12f == 110 & dep1 < (3.4*disprop),1,disease)
> disease=ifelse(xg12f == 111 & dep1 < (5.1*disprop),1,disease)
>
> sum(disease)/nrow(d1)
[1] 0.196368
>
> r1=glm(disease~rf1+rf2+grp+rf1*rf2+rf1*grp+rf2*grp,
+ family=binomial(link=log))
> summary(r1)Call:
glm(formula = disease ~ rf1 + rf2 + grp + rf1 * rf2 + rf1 * grp +
    rf2 * grp, family = binomial(link = log))Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.2029  -0.6670  -0.5688  -0.4587   2.1465
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.303778    0.005401 -426.540  <2e-16 ***
rf1          0.512466    0.008491  60.356  <2e-16 ***
rf2          0.402541    0.007816  51.502  <2e-16 ***
grp          0.691635    0.006344 109.025  <2e-16 ***

```

```

rf1:rf2      0.008824    0.008221    1.073    0.283
rf1:grp      0.013005    0.009147    1.422    0.155
rf2:grp      0.011577    0.008706    1.330    0.184
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 (Dispersion
parameter for binomial family taken to be 1)    Null deviance: 990652  on
999999  degrees of freedom
Residual deviance: 938046  on 999993  degrees of freedom
AIC: 938060Number of Fisher Scoring iterations: 6> x1=exp(confint.default(r1))
> rr=exp(coef(r1))
> round(cbind(rr,x1),digits=2)
           rr 2.5 % 97.5 %
(Intercept) 0.10  0.10  0.10
rf1          1.67  1.64  1.70
rf2          1.50  1.47  1.52
grp          2.00  1.97  2.02
rf1:rf2      1.01  0.99  1.03
rf1:grp      1.01  1.00  1.03
rf2:grp      1.01  0.99  1.03
>
> r1=glm(disease~rf1+rf2+grp+rf1*rf2+rf1*grp+rf2*grp,
+ family=binomial(link=logit))
> summary(r1)Call:
glm(formula = disease ~ rf1 + rf2 + grp + rf1 * rf2 + rf1 * grp +
      rf2 * grp, family = binomial(link = logit))Deviance Residuals:
      Min       1Q   Median       3Q      Max
-1.1920  -0.6657  -0.5655  -0.4604   2.1433 Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.190813    0.006088 -359.89  <2e-16 ***
rf1          0.561856    0.010543  53.29  <2e-16 ***
rf2          0.438542    0.009391  46.70  <2e-16 ***
grp          0.796766    0.007483 106.48  <2e-16 ***
rf1:rf2      0.134330    0.012407  10.83  <2e-16 ***
rf1:grp      0.169283    0.012129  13.96  <2e-16 ***
rf2:grp      0.124378    0.011124  11.18  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 (Dispersion
parameter for binomial family taken to be 1)    Null deviance: 990652  on
999999  degrees of freedom
Residual deviance: 938069  on 999993  degrees of freedom
AIC: 938083Number of Fisher Scoring iterations: 4> x1=exp(confint.default(r1))
> or=exp(coef(r1))
> round(cbind(or,x1),digits=2)
           or 2.5 % 97.5 %
(Intercept) 0.11  0.11  0.11
rf1          1.75  1.72  1.79
rf2          1.55  1.52  1.58
grp          2.22  2.19  2.25
rf1:rf2      1.14  1.12  1.17
rf1:grp      1.18  1.16  1.21
rf2:grp      1.13  1.11  1.16
>

```

Discussion

These simulations show that given the same data, different generalized linear models (glm) will sometimes lead to different conclusions about interaction between the dependent and independent variables. When the disease is rare, adjusted odds ratios and adjusted risk ratios tend to agree regarding the statistical significance of the interaction terms. However, as the prevalence of the disease increases, the adjusted odds ratios and adjusted risk ratios tend to produce conflicting results for the interaction terms. This is discussed and explained mathematically in the article referred to above: "Distributional interaction: Interpretational problems when using incidence odds ratios to assess interaction" by Ulka B Campbell, Nicolle M Gatto and Sharon Schwartz. (Epidemiologic Perspectives & Innovations 2005, 2:1 doi:10.1186/1742-5573-2-1) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1079911/>. The purpose of this paper was to illustrate how this works with adjusted risk ratios and adjusted odds ratios using the R glm function.

For practicing statisticians and epidemiologists, it might be good advice to use risk ratios whenever denominators are available; and use odds ratios only when denominators are not available, such as case-control studies.

For example, with case-control studies, data is collected for selected persons with a disease (cases) and also for a specific number of persons without the disease (controls). With this data there are no denominators for persons at risk of getting the disease and disease rates cannot be calculated. Without disease rates, rate ratios cannot be calculated, but case control data can be used to calculate odds ratios even though the rates are unavailable. For more on this see "Statistical Methods in Cancer Research Volume I: The Analysis of Case-Control Studies", IARC Scientific Publication No. 32, Authors: Breslow NE, Day NE at: <https://publications.iarc.fr/Book-And-Report-Series/Iarc-Scientific-Publications/Statistical-Methods-In-Cancer-Research-Volume-I-The-Analysis-Of-Case-Control-Studies-1980>

Here is the R program used in this analysis. It should run when copied and pasted into the R window.

```
rm(list=ls(all=TRUE))

nx=1000000

set.seed(523)

x1=runif(nx)
x2=runif(nx)
grpx=runif(nx)
dep1=runif(nx)

rf1=ifelse(x1 < 0.20,1,0)
rf2=ifelse(x2 < 0.30,1,0)
grp=ifelse(grpx < 0.5,1,0)

d1=as.data.frame(cbind(rf1,rf2,grp,dep1))

#Sim 1

attach(d1, warn.conflict=F)

xg12f=(grp*100)+(rf1*10)+rf2

disprop=0.01

disease=ifelse(xg12f == 0 & dep1 < disprop,1,0)
disease=ifelse(xg12f == 1 & dep1 < (1.5*disprop),1,disease)
disease=ifelse(xg12f == 10 & dep1 < (1.7*disprop),1,disease)
disease=ifelse(xg12f == 11 & dep1 < (1.51*disprop),1,disease)

disease=ifelse(xg12f == 100 & dep1 < (2*disprop),1,disease)
disease=ifelse(xg12f == 101 & dep1 < (3*disprop),1,disease)
disease=ifelse(xg12f == 110 & dep1 < (3.4*disprop),1,disease)
```



```

disease=ifelse(xg12f == 111 & dep1 < (5.1disprop),1,disease)

sum(disease)/nrow(d1)

r1=glm(disease~rf1+rf2+grp+rf1rf2+rf1grp+rf2grp,
family=binomial(link=log))
summary(r1)
x1=exp(confint.default(r1))
rr=exp(coef(r1))
round(cbind(rr,x1),digits=2)

r1=glm(disease~rf1+rf2+grp+rf1rf2+rf1grp+rf2grp,
family=binomial(link=logit))
summary(r1)
x1=exp(confint.default(r1))
or=exp(coef(r1))
round(cbind(or,x1),digits=2)

#sim 2

disprop=0.05

disease=ifelse(xg12f == 0 & dep1 < disprop,1,0)
disease=ifelse(xg12f == 1 & dep1 < (1.5disprop),1,disease)
disease=ifelse(xg12f == 10 & dep1 < (1.7disprop),1,disease)
disease=ifelse(xg12f == 11 & dep1 < (1.51.7disprop),1,disease)

disease=ifelse(xg12f == 100 & dep1 < (2disprop),1,disease)
disease=ifelse(xg12f == 101 & dep1 < (3disprop),1,disease)
disease=ifelse(xg12f == 110 & dep1 < (3.4disprop),1,disease)
disease=ifelse(xg12f == 111 & dep1 < (5.1disprop),1,disease)

sum(disease)/nrow(d1)

r1=glm(disease~rf1+rf2+grp+rf1rf2+rf1grp+rf2grp,
family=binomial(link=log))
summary(r1)
x1=exp(confint.default(r1))
rr=exp(coef(r1))
round(cbind(rr,x1),digits=2)

r1=glm(disease~rf1+rf2+grp+rf1rf2+rf1grp+rf2grp,
family=binomial(link=logit))
summary(r1)
x1=exp(confint.default(r1))
or=exp(coef(r1))
round(cbind(or,x1),digits=2)

#sim 3

disprop=0.10

disease=ifelse(xg12f == 0 & dep1 < disprop,1,0)
disease=ifelse(xg12f == 1 & dep1 < (1.5disprop),1,disease)
disease=ifelse(xg12f == 10 & dep1 < (1.7disprop),1,disease)
disease=ifelse(xg12f == 11 & dep1 < (1.51.7disprop),1,disease)

disease=ifelse(xg12f == 100 & dep1 < (2disprop),1,disease)
disease=ifelse(xg12f == 101 & dep1 < (3disprop),1,disease)
disease=ifelse(xg12f == 110 & dep1 < (3.4disprop),1,disease)
disease=ifelse(xg12f == 111 & dep1 < (5.1disprop),1,disease)

```

```
sum(disease)/nrow(d1)
```

```
r1=glm(disease~rf1+rf2+grp+rf1rf2+rf1grp+rf2grp,  
family=binomial(link=log))  
summary(r1)  
x1=exp(confint.default(r1))  
rr=exp(coef(r1))  
round(cbind(rr,x1),digits=2)
```

```
r1=glm(disease~rf1+rf2+grp+rf1rf2+rf1grp+rf2grp,  
family=binomial(link=logit))  
summary(r1)  
x1=exp(confint.default(r1))  
or=exp(coef(r1))  
round(cbind(or,x1),digits=2)
```