Confounding in logistic regression - Practical

Phillip Ayieko

We simulate a data set that follows this scenario in R as follows:

Confounding

```
# Create any first independent variable (round to one decimal place)
x1 <- round(rnorm(400, mean=0, sd=1), 1)
# Create any second independent variable (round to one decimal place)
x2 <- round(rnorm(400, mean = 4, sd=2), 1)
# Now create a third independent variable that is related by not a direct function of the first two var
x3 <- round(3*x1 + 2 *x2 + rnorm(400, mean = 0, sd=5), 1)
# Create a binary outcome variable that depends on all three variables
# Note that the probability of the binomial is an inv.logit function
# We will use smaller effects this time as well, more realistic.
# Note that a coefficient of 0.2 has an OR of
# exp(0.2) = 1.22 / one unit change
y <- rbinom(400, 1, exp(.2*x1 + .3*x2 -.3 * x3)/(1+ exp(.2*x1 + 2*x2 -3 * x3)))
# Put all variables into a data frame
confounding.dat <- data.frame(x1=x1, x2=x2, x3=x3, y=y)
# If looked at pairwise, the very strong confounding is not obvious
# because it arises from three variables working together
pairs(confounding.dat)
```

Now to analyze the data, comparing univariate to multivariate model outputs

```
# First univariate logistic regressions for each of the three variables
output <- glm(y ~ x1, data = confounding.dat, family = binomial)
logistic.regression.or.ci(output)
output <- glm(y ~ x2, data = confounding.dat, family = binomial)
logistic.regression.or.ci(output)
output <- glm(y ~ x3, data = confounding.dat, family = binomial)
logistic.regression.or.ci(output)
# Now let's run a logistic regression with all three variables included:
output <- glm(y ~ x1 + x2 + x3, data = confounding.dat, family = binomial)
logistic.regression.or.ci(output)
```

To investigate the above results for confounding, let's form a comparative table:- how different the results are