Statistical Methods in Epidemiology
Lab 9 - Solutions.
Case-Control Studies

I. Matched Case-Control Studies: The Salmonella Typhimurium dataset

1.
```
. clogit case pork, group(set) nolog
note: 6 groups (12 obs) dropped due to all positive or all negative
outcomes.

Conditional (fixed-effects) logistic regression   Number of obs  =      117
                                                  LR chi2(1)     =     0.35
                                                  Prob > chi2    =   0.5538
Log likelihood = -42.435054                      Pseudo R2      =   0.0041

------------------------------------------------------------------------------
     case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------+-------------------------------------------------------------------
     pork |   .2656662   .4539626     0.59   0.558    -.6240842    1.155416
------------------------------------------------------------------------------
```

Pork consumption is associated with an average increase of the odds of salmonella by exp(.2656662)-1 = 1.3043-1 = 30.43%.

2.
```
. clogit case beef, group(set) nolog
note: 5 groups (8 obs) dropped due to all positive or all negative outcomes.

Conditional (fixed-effects) logistic regression   Number of obs  =     122
                                                  LR chi2(1)     =    0.08
                                                  Prob > chi2    =  0.7784
Log likelihood = -44.480256                       Pseudo R2      =  0.0009

------------------------------------------------------------------------------
     case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------+-------------------------------------------------------------------
     beef |  -.1188162   .4223471    -0.28   0.778    -.9466012    .7089689
------------------------------------------------------------------------------

. clogit case poultry, group(set) nolog
note: 7 groups (13 obs) dropped due to all positive or all negative
outcomes.

Conditional (fixed-effects) logistic regression   Number of obs  =    116
                                                  LR chi2(1)     =   0.11
                                                  Prob > chi2    = 0.7440
Log likelihood = -42.269322                       Pseudo R2      = 0.0013

------------------------------------------------------------------------------
     case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------+-------------------------------------------------------------------
  poultry |    .127572   .3896615     0.33   0.743    -.6361506    .8912945
------------------------------------------------------------------------------

. clogit case liverp, group(set) nolog
note: 1 group (1 obs) dropped due to all positive or all negative outcomes.

Conditional (fixed-effects) logistic regression   Number of obs  =    133
                                                  LR chi2(1)     =   2.90
                                                  Prob > chi2    = 0.0884
Log likelihood = -47.056878                       Pseudo R2      = 0.0299
```

```
---------------------------------------------------------------------------
     case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------+----------------------------------------------------------------
   liverp |  -1.791759   1.154701    -1.55   0.121    -4.054931     .471412
---------------------------------------------------------------------------

. clogit case meat, group(set) nolog
note: 2 groups (3 obs) dropped due to all positive or all negative outcomes.
Conditional (fixed-effects) logistic regression   Number of obs   =    130
                                                  LR chi2(1)      =   1.88
                                                  Prob > chi2     = 0.1699
Log likelihood = -46.468407                       Pseudo R2       = 0.0199

---------------------------------------------------------------------------
     case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------+----------------------------------------------------------------
     meat |   .6132185    .463145     1.32   0.185     -.294529    1.520966
---------------------------------------------------------------------------

. clogit case veg, group(set) nolog
note: 2 groups (3 obs) dropped due to all positive or all negative outcomes.

Conditional (fixed-effects) logistic regression   Number of obs   =    129
                                                  LR chi2(1)      =   0.84
                                                  Prob > chi2     = 0.3582
Log likelihood = -46.582661                       Pseudo R2       = 0.0090

---------------------------------------------------------------------------
     case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------+----------------------------------------------------------------
      veg |  -.6300026   .6871156    -0.92   0.359    -1.976724    .7167193
---------------------------------------------------------------------------

. clogit case fruit, group(set) nolog
note: 3 groups (5 obs) dropped due to all positive or all negative outcomes.

Conditional (fixed-effects) logistic regression   Number of obs   =    127
                                                  LR chi2(1)      =   9.47
                                                  Prob > chi2     = 0.0021
Log likelihood =  -41.57635                       Pseudo R2       = 0.1022

---------------------------------------------------------------------------
     case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------+----------------------------------------------------------------
    fruit |  -1.811938   .6591034    -2.75   0.006    -3.103757   -.5201195
---------------------------------------------------------------------------

. clogit case egg, group(set) nolog
note: 7 groups (13 obs) dropped due to all positive or all negative
outcomes.

Conditional (fixed-effects) logistic regression   Number of obs   =    114
                                                  LR chi2(1)      =  -0.00
                                                  Prob > chi2     = 1.0000
Log likelihood = -41.511701                       Pseudo R2       =-0.0000

---------------------------------------------------------------------------
     case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------+----------------------------------------------------------------
      egg |  -1.42e-10   .5303301    -0.00   1.000    -1.039428    1.039428
---------------------------------------------------------------------------

. clogit case plant7, group(set) nolog
note: 10 groups (16 obs) dropped due to all positive or all negative
outcomes.

Conditional (fixed-effects) logistic regression   Number of obs   =    107
                                                  LR chi2(1)      =  10.09
                                                  Prob > chi2     = 0.0015
Log likelihood = -33.982814                       Pseudo R2       = 0.1292

---------------------------------------------------------------------------
     case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------+----------------------------------------------------------------
   plant7 |   1.497087   .5193115     2.88   0.004     .4792554    2.514919
---------------------------------------------------------------------------
```

There is no evidence of a statistically significance association between nutrition and salmonella occurrence except from fruit consumption which have a protective effect. More specifically, the odds of salmonella is 1 – exp(-1.81) = 84% lower for people who eat fruits compared with people who do not include fruits in their daily nutrition.

Plant 7 should also be thought as a significant risk factor.

3. Accounting for the effects of both fruit consumption and plant7 (as a meat origin) on salmonella onset, we fit firstly a model including only main effects of these univariatly significant factors:

```
. clogit case fruit plant7 , group(set) nolog
note: 11 groups (18 obs) dropped due to all positive or all negative
outcomes.

Conditional (fixed-effects) logistic regression   Number of obs  =    103
                                                  LR chi2(2)     =  14.15
                                                  Prob > chi2    = 0.0008
Log likelihood = -30.448621                       Pseudo R2      = 0.1885
---------------------------------------------------------------------------
     case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------+----------------------------------------------------------------
    fruit |  -1.418644   .7525455    -1.89   0.059    -2.893606    .0563179
   plant7 |   1.496342   .5859389     2.55   0.011     .3479225    2.644761
---------------------------------------------------------------------------
```

Adjusting for plant7, the protective effect of fruit consumption decreases both in strength and in significance.

In order to test whether plant7 modifies the effect of fruit consumption, we have to include an interaction term in our model:

```
 . xi: clogit case i.fruit*i.plant7, group(set) nolog
i.fruit           _Ifruit_0-1         (naturally coded; _Ifruit_0 omitted)
i.plant7          _Iplant7_0-1        (naturally coded; _Iplant7_0 omitted)
i.fruit*i.pla~7   _IfruXpla_#_#       (coded as above)
note: 11 groups (18 obs) dropped due to all positive or all negative
outcomes.

Conditional (fixed-effects) logistic regression   Number of obs  =    103
                                                  LR chi2(3)     =  14.28
                                                  Prob > chi2    = 0.0025
Log likelihood = -30.382707                       Pseudo R2      = 0.1903
---------------------------------------------------------------------------
     case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------+----------------------------------------------------------------
_Ifruit_1 |  -1.118084    1.11626    -1.00   0.317    -3.305913    1.069745
_Iplant7_1|   2.011522   1.558159     1.29   0.197    -1.042414    5.065458
_IfrXpl_~1|  -.5877987   1.632351    -0.36   0.719    -3.787147     2.61155
---------------------------------------------------------------------------
```

log(OR)	fruit	
plant7	0	1
0	Nuisance parameter – not reported	-1.118084
1	2.011522	-1.118084+2.011522-.5877987

The interaction term is not statistically significant implying that there is no evidence for effect modification. Here the reference category includes persons who consume neither fruits nor meat from plant7. The fact that fruits have a protective effect while plant7 acts as a risk factor makes interpretation meaningless. It would be more appropriate to change our reference category so as to comprise people who eat meat from plant7 and are not fruit consumers:

```
. char plant7[omit] 1

. xi: clogit case i.fruit*i.plant7, group(set) nolog
i.fruit           _Ifruit_0-1         (naturally coded; _Ifruit_0 omitted)
i.plant7          _Iplant7_0-1        (naturally coded; _Iplant7_1 omitted)
i.fruit*i.pla~7   _IfruXpla_#_#       (coded as above)
note: 11 groups (18 obs) dropped due to all positive or all negative
outcomes.

Conditional (fixed-effects) logistic regression   Number of obs  =     103
                                                  LR chi2(3)     =   14.28
                                                  Prob > chi2    =  0.0025
Log likelihood = -30.382707                       Pseudo R2      =  0.1903

------------------------------------------------------------------------------
     case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------+-------------------------------------------------------------------
_Ifruit_1 |  -1.705883   1.137987    -1.50   0.134    -3.936295      .52453
_Iplant7_0|  -2.011522   1.558159    -1.29   0.197    -5.065458    1.042414
_IfrXpl_~0|   .5877987   1.632351     0.36   0.719     -2.61155    3.787147
------------------------------------------------------------------------------
```

Parameterizing our model in this way, one could e.g. conclude that the odds of salmonella is on average 1- exp(-1.706-2.012+.588) = 96% lower for fruit consumers who do not eat meat originated from plant7 in comparison to non-fruit consumers who eat meat from plant 7. Be careful that such an interpretation would be convenient if our analysis had produced significant results.

4. In this case we do not care about those who eat meat from plant7. In terms of modelling this means that we would like to define our reference category so as to include persons that have not eaten meat from plant7 independently of whether they are fruit consumers or not. The model will include only one three-level variable. Of course this is not an interaction model.

```
. gen frpl1=0 if (fruit==0&plant7==0)|(fruit==1&plant7==0)
(75 missing values generated)

. replace frpl1=1 if (fruit==0&plant7==1)
```

```
(11 real changes made)

. replace frpl1=2 if (fruit==1&plant7==1)
(49 real changes made)



. xi:clogit case i.frpl1, group(set)  nolog
i.temp            _Itemp_0-2          (naturally coded; _Itemp_0 omitted)
note: 11 groups (18 obs) dropped due to all positive or all negative
outcomes.

Conditional (fixed-effects) logistic regression   Number of obs  =    103
                                                  LR chi2(2)     =  13.29
                                                  Prob > chi2    = 0.0013
Log likelihood = -30.879976                       Pseudo R2      = 0.1770

----------------------------------------------------------------------------
     case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------+-----------------------------------------------------------------
 _Ifrpl1_1 |   3.054601   1.169598     2.61   0.009     .7622302    5.346971
 _Ifrpl1_2 |   1.304073   .5928097     2.20   0.028     .1421873    2.465959
----------------------------------------------------------------------------
```

5.

log(OR)	fruit	
plant7	0	1
0	Nuisance parameter – not reported	
1	3.054601	1.304073

Those who eat meat from plant7 and are not fruit consumers have on average exp(3.054601)=21.21times higher odds of salmonella compared with people who have not eaten meat from plant7 (independently of whether they eat fruits or not). The aggravating effect of plant7 is remarkably limited for fruit consumers (Odds of disease = exp(1.304073)=3.68).

II. Choice of Controls in Case-Control Studies

1. Mahmood et al. (1989). Infant feeding and risk of severe diarrhea in Basrah city, Iraq: a case - control study. *Bulletin of the World Health Organization*, **67(6):** 701-706.

i) Source of controls: Health facility-based controls.

ii) Potential advantages & disadvantages of this approach.

Advantages:

1) Easier and less expensive than general population controls.

2) May be more aware of exposures and likely to cooperate than general population controls (healthier).

Disadvantages:

1) Controls are ill; distribution of the exposure may not reflect the distribution of exposure in the source population for cases. That's why controls should be limited to diagnoses for which there is no prior indication of a relation with exposure.

2) Subjects may have changed their exposure status as a result of being sick.

In this case controls are recruited from all seven MCHC in the city so as to limit the potential confounding effect of e.g. food, drinkable water, environmental factors etc. Moreover, the reason for controls visiting MCHC is either immunization or a routine check-up, which means that controls are in fact not ill.

iii) Sampling scheme for controls: Concurrent sampling, i.e. controls are selected concurrently from the population still at risk when a new case is diagnosed.

iv) Appropriate measure of relative incidence: Rate Ratio.

Justification: Diarrhea is a common disease. Individuals can experience more than one episode during follow-up and risk-based measures will tend to 1 as the period of follow-up increases. So what is of interest is how often an individual experience the disease.

v) Criterion(e) of exclusion of infants "3 months of age and older with no history of being taken to an MCHC for immunization" from the cases: The aim is to avoid selection bias. The underlying cohort should be uniform for both cases and controls, i.e. including all immunized infants with a potential of experiencing diarrhea during the study period. Generally speaking, in this way we ensure that the conditions from which the controls are suffering are not related to the exposure we are studying.

2. Mueller et al. (1987). Tonsillectomy and Hodkin's Disease: Results from Companion Population-Based Studies. *J Natl Cancer Inst*, **78:** 1-5.

Use of siblings of cases as controls.

Advantages:

1) Tend to be more cooperative than general population controls.

2) Similar to cases on factors such as socio-economic status, lifestyle, genetic characteristics and ethnic background.

Disadvantages:

1) The list of potential relative controls is often derived from the case; this dependence may add a potential source of bias.

2) Hence, relative controls may be too similar to cases regarding the exposure of interest.

3. Collins et al. (1999). Surgical treatment and risk of sporadic Creutzfeldt-Jakob disease: a case-control study. *Lancet*, **353:** 693-697.

Use of random telephone controls.

Reason:

To avoid the deficiencies of hospital-based controls, i.e. potential change in the exposure status of controls because of being ill and the fact that the distribution of the exposure may not reflect the distribution of exposure in the source population for cases. For the disease under study (CJD), the risk of bias due to the above reasons is high because the etiology is unknown.

Advantages:

1) May approximate random sampling from the source population.

2) Controls are often matched to cases on area code and prefix (i.e. SES matching).

Disadvantages:

1) Probability of contacting each eligible subject may differ due to time of day, number in household, answering machines, etc.

2) Lack of personal contact and limited time available for the interview may have as a result misleading or incorrect information.