APPENDIX

APPENDIX: DATA STRUCTURE AND PROGRAM USING

A. Data structure

Tha raw data for the HIV/AIDS cases in the five provinces of Southern Thailand in 1991-1996 are provided by Division of Epidemiology, Ministry of Public Health. This file is called *case.dat* and consists of 11 columns (diagnosis, sex, age, marital, race, occupation, province, date of diagnosis, status, behaviour and method of transmission), as follows

```
1 1 24 1 1 43 90 10/05/91 3 9 1
2 1 48 1 1 61 90 14/06/93 1 2 1
1 1 32 1 1 91 90 13/10/92 1 2 9
1 1 18 4 1 91 90 02/07/94 1 2 9
1 2 25 1 1 61 90 01/05/93 1 9 9
.....
.....
2 2 31 1 1 61 91 01/04/92 1 2 1
1 1 33 2 1 43 91 10/04/93 1 2 2
2 1 22 1 1 61 91 06/09/93 1 9 1
1 1 26 1 1 61 91 19/09/94 1 2 1
1 1 40 1 1 43 91 29/07/93 1 2 2
.....
.....
1 1 32 1 1 61 94 01/05/96 1 9 9
2 1 23 1 1 61 94 01/04/94 1 2 1
1 1 24 2 1 43 94 10/04/95 1 2 2
2 1 29 1 1 61 94 06/09/93 1 9 1
1 1 44 2 1 43 94 10/04/95 1 2 2
.....
.....
```

1 2 28 1 1 61 95 19/09/93 1 2 1

2 1 32 1 1 61 95 01/05/95 1 9 9

1 2 28 1 1 61 95 19/09/93 1 2 1

1 1 32 1 1 61 95 01/05/96 1 9 9

2 1 44 1 1 61 95 01/04/94 1 2 1

.....

.....

1 1 33 2 1 43 96 10/04/95 1 2 2

2 1 25 1 1 61 96 06/09/93 1 9 1

1 1 26 1 1 61 96 19/09/93 1 2 1

2 1 25 1 1 61 96 06/09/93 1 9 1

1 1 26 1 1 61 96 19/09/93 1 2 1

.....

.....

The population data information is obtained from the Human Resources Planning Division National Economic and Social Development Board. These data were stored in two files, called *Songkhla.dat* (Songkhla and Pattani data) and *syn.dat* (for the other three provinces), each comprising 5 columns (province, sex, year, age group and count) as follows.

90	f	93	0-4	13309
90	f	93	5-9	14170
90	f	93	10-14	12293

.....

.....

90	m	91	0-4	8082
90	m	91	5-9	10033
90	m	91	10-14	10020

.....

.....

91	f	93	0-4	6648
91	f	93	5-9	4891
91	f	93	10-14	8197
.....				
.....				
91	m	91	0-4	4186
91	m	91	5-9	3171
91	m	91	10-14	11207
.....				
.....				
94	f	93	0-4	9702
94	f	93	5-9	12293
94	f	93	10-14	10082
.....				
.....				
94	m	91	0-4	10033
94	m	91	5-9	8820
94	m	91	10-14	12293
.....				
.....				
95	f	93	0-4	8882
95	f	93	5-9	7833
95	f	93	10-14	10120
.....				
.....				
95	m	91	0-4	8882
95	m	91	5-9	7833
95	m	91	10-14	10120
.....				
.....				

96	f	93	0-4	7482
96	f	93	5-9	7633
96	f	93	10-14	10140
.....				
.....				
96	m	91	0-4	7882
96	m	91	5-9	5233
96	m	91	10-14	10110
.....				
.....				

The case by case data were grouped using the SPIDA statistical package (Gebski et al 1992), to created data records matching and joined to the population data, creating the datafile *casepop.dat*. This program listing is as follows.

```
!column   1: diagnosis (1=aids,2=hiv)
!         2:  gender (1=male,2=female)
!         3:  age in years
!         4:  marital status (1=single,2=married,3=separated,4=widowed,
                              5=divorced,6=under 15,9=unknown)
!         5:  race (1=thai,2=chinese,3=other,9=unknown)
!         6:  occupation
!         7:  province (90=Songkhla,91=satul,94=pattani,95=yala,96=Narathiwat)
!         8: date of diagnosis (dd/mm/yy)
!         9: status (1=alive,3=dead)
!        10: behaviour (1=homosexual,2=hetero,3=bisexual,9=unknown)
!        11: transmission (1=ivdu,2=sex,3=bdonor,4 mother,9=unknown)
!
! recode dates into years (1991=0, 1992=1, etc.)
$days := case[;8]-"31/12/90"
```

```
$jan1 := 1;366;731;1097;1442;1807;2172
$dec31 := 365;730;1096;1441;1806;2171;2437
$year := rec($days,g=($jan1,$dec31,(0;;6)))
$y := casc[;1,,7],$year,case[;9,,11]
!
! recode ages into age groups (0-4=0,5-9=1,10-14=2,...,75+=15)
$age := $y[;3]
$lowerage := 5*(0;;15)
$upperage := 5*(1;;15;20)-1
$ag := rec($age,g=($lowerage,$upperage,(0;;15)))
$y := $y[;1,2],$ag,$y[;4,,11]
! recode province (90=0,91=1,94=4,95=5,96=6)
$y := rec($y,col=7,t=((90;91;94;95;96),(0;1;4;5;6)))
! create index for combination (province+year+gender+agegroup)
$index := 10000*$y[;7]+1000*$y[;8]+100*$y[;2]+$y[;3]
$y := $y,$index
$ylab := case.nam;"index"
%label $y $ylab
desc($y)
! separate aids cases and hiv cases
$aids := $y[;1]==1 ?: $y
$hiv := $y[;1]==2 ?: $y
! separate ivdu cases and sex cases
$ivdu := $y[;11]==1 ?: $y
$scx := $y[;11]==2 ?: $y
! stratify cases by index
$z := condense($y,y=12,x=7;8;2;3,typ=3,data=1)
$nr := rdim($z)
$z := $z[2,,$nr-1][;1,6]
$zaids := condense($aids,y=12,x=7;8;2;3,typ=3,data=1)
```

```
$nr := rdim($zaids)
$zaids := $zaids[2,,$nr-1][;1,6]
$zhiv := condense($hiv,y=12,x=7;8;2;3,typ=3,data=1)
$nr := rdim($zhiv)
$zhiv := $zhiv[2,,$nr-1][;1,6]
$zivdu := condense($ivdu,y=12,x=7;8;2;3,typ=3,data=1)
$nr := rdim($zivdu)
$zivdu := $zivdu[2,,$nr-1][;1,6]
$zsex := condense($scx,y=12,x=7;8;2;3,typ=3,data=1)
$nr := rdim($zsex)
$zsex := $zsex[2,,$nr-1][;1,6]
! create stratified population counts
! from data stored in province files (Songkhla.dat, syn.dat)
! as province,gender(m,f),year,agegroup(0-4,5-9,..,75+),count
$y := songpat;syn
$y := rec($y,col=1,t=((90;91;94;95;96),(0;1;4;5;6)))
$y := rec($y,col=2,t=(("m";"f"),(1;2)))
$y := rec($y,col=3,t=((91;;97),(0;;6)))
$ag := "0-4";"5-9";"10-14";"15-19";"20-24";"25-29";"30-34";"35-39"
$ag := $ag;"40-44";"45-49";"50-54";"55-59";"60-64";"65-69";"70-74";"75+"
$y := rec($y,col=4,t=($ag,(0;;15)))
$index := 10000*$y[;1]+1000*$y[;3]+100*$y[;2]+$y[;4]
$y := $index,$y[;1,3,2,4,5]
$ylab := "index";"province";"year";"gender";"agegroup";"population"
%label $y $ylab
desc($y)
! join to population counts
$y := join($y,$z,typ=3)
$y := $y|0
$y := join($y,$zaids,typ=3)
```

```
$y := $y|0
$y := join($y,$zhiv,typ=3)
$y := $y|0
$y := join($y,$zivdu,typ=3)
$y := $y|0
$y := join($y,$zsex,typ=3)
$y := $y|0
$ylab := $ylab;"total";"aids";"hiv";"ivdu";"sex"
%label $y $ylab
desc($y)
$totinc := round(100000*$y[;7]/$y[;6])
$sm := $y[;2]==0 & $y[;4]==1 ?: $totinc
$sm := resh($sm,16;7)
$sf := $y[;2]==0 & $y[;4]==2 ?: $totinc
$sf := resh($sf,16;7)
casepop := $y
```

The *casepop.dat* file these comprises 11 columns (index, province, year, gender, agegroup, population, total, aids, hiv, ivdu, sex) and 960 records as follows.

100	0	0	1	0	61253	0	0	0	0	0
101	0	0	1	1	63377	0	0	0	0	0
102	0	0	1	2	61603	0	0	0	0	0
103	0	0	1	3	61765	0	0	0	0	0
104	0	0	1	4	56584	1	0	1	0	1
105	0	0	1	5	51393	3	1	2	1	1
106	0	0	1	6	44078	0	0	0	0	0
107	0	0	1	7	38737	1	0	1	0	0
108	0	0	1	8	33124	1	1	0	0	1
109	0	0	1	9	25412	0	0	0	0	0
110	0	0	1	2	24579	0	0	0	0	0

111	0	0	1	1	63377	0	0	0	0	0
112	0	0	1	2	61603	0	0	0	0	0
113	0	0	1	3	61765	0	0	0	0	0
114	0	0	1	4	56584	1	0	1	0	1
115	0	0	1	5	51393	3	1	2	1	1
116	0	0	1	6	44078	0	0	0	0	0
117	0	0	1	7	38737	1	0	1	0	0
118	0	0	1	8	30325	1	1	0	0	1
119	0	0	1	9	25642	0	0	0	0	0

.....

.....

.....

B. Graphing Incidence Rates

To plot incidence rates of HIV/AIDS cases stratified by age group for each gender, province, year and the method of transmission using matlab together with functions (getfile, getnum, putnum, putfn, putdn, describe,setvar, track) in the ASP (McNeil et al 1997) package, the following program is used.

```
system_dependent(14,'on')
getfile casepop.num
y=getnum;
y(:,6) = y(:,6)./1000;
putnum(y)
fn = getfn;
fn = str2mat(fn(1:5,:),'pop(1000s)',fn(7:11,:));
putfn(fn)
%describe hist=1 col=2:11
for j=7:11
y(:,j) = y(:,j)./y(:,6);
```

```
end
putfn(fn)
s0 =( (y(:,2)==0 | y(:,2)==1 | y(:,2)==4 | y(:,2)==5 | y(:,2)==6) & (y(:,3)==3) );
ss0 = y(s0,:);
putnum(ss0)
putdn('HIV/AIDS incidence/1000 in 5 provinces in 1994')
describe hist=1 col=2:11
female = ss0(:,4)==2;
ss0(female,7:11) = ss0(female,7:11)-0.01;
male = ss0(:,4)==1;
ss0(male,7:11) = ss0(male,7:11)+0.01;
putnum(ss0)
setvar y=7 'x=5 2 4'
track size=9 ymax=1.5
setvar y=8 'x=5 2 4'
track size=9 ymax=1.5
setvar y=9 'x=5 2 4'
track size=9 ymax=1.5
setvar y=10 'x=5 2 4'
track size=9 ymax=1.5
setvar y=11 'x=5 2 4'
track size=9 ymax=1.5
```

C. Graphing Odds Ratios

To graph the odds ratios, the data used to plot the incidence rates need to be restructured to have separate records for cases and noncases. Only the total number of HIV/AIDS cases are included in these odds ratios calculations. Starting with the *casepop.dat* file, we create two files, one comprising columns 2-6 (for the noncases) and the other comprising columns 2-5 and 7 (for the cases). These files are then

stacked, and are additional column identifying noncases as 0s and cases as 1s is attacted as the first column. This gives rise to the following data file called *mhall.dat*

```
0 0 0 1  0 61253
0 0 0 1  1 63377
0 0 0 1  2 61603
0 0 0 1  3 61765
0 0 0 1  4 56584
0 0 0 1  5 51393
0 0 0 1  6 44078
0 0 0 1  7 38737
0 0 0 1  2 61151
0 0 0 1  3 51423
0 0 0 1  4 41250
0 0 0 1  5 21435
0 0 0 1  2 52393
0 0 0 1  6 45078
0 0 0 1  4 39737
0 0 0 1  2 61151
.....
.....
.....
1 0 1 1  0   0
1 0 1 1  1   0
1 0 1 1  2   0
1 0 1 1  3   0
1 0 1 1  4   1
1 0 1 1  5   3
1 0 1 1  6   0
1 0 1 1  7   1
```

```
1 0 1 1 2   0
1 0 1 1 3   0
1 0 1 1 4   1
1 0 1 1 5   3
1 0 1 1 2   0
1 0 1 1 3   0
1 0 1 1 4   1
1 0 1 1 5   3
.....
.....
.....
```

To plot crude odds ratios for males and females in each province comparing the pattern of incidence in different years, with year 1996 as baseline, and to plot crude odds ratios in each year comparing different provinces, with Songkhla as baseline, the following program is used.

```
system_dependent(14,'on')
getfile mhall.num
y=getnum;
fp90m = ((y(:,2)==0) & (y(:,4)==1)); %males in Songkhla
y=y(fp90m,:);
y(:,3) = 5-y(:,3);
[lab, colid] = getlab;
lab(3,1:43) = '3,0 1996,1 1995,2 1994,3 1993,4 1992,5 1991'
putlab(lab,colid)
putnum(y)
putdn('HIV/AIDS incidence of males in Songkhla province ')
setvar y=3 x=1 z=6
orplot log=1
```

```
getfile mhall.num
y=getnum;
fp90m = ((y(:,2)==1) & (y(:,4)==1)); %males in Satun
y=y(fp90m,:);
y(:,3) = 5-y(:,3);
[lab, colid] = getlab;
lab(3,1:43) = '3,0 1996,1 1995,2 1994,3 1993,4 1992,5 1991'
putlab(lab,colid)
putnum(y)
putdn('HIV/AIDS incidence of males in Satun province ')
setvar y=3 x=1 z=6
orplot log=1
getfile mhall.num
y=getnum;
fp90m = ((y(:,2)==4) & (y(:,4)==1)); %males in Pattani
y=y(fp90m,:);
y(:,3) = 5-y(:,3);
[lab, colid] = getlab;
lab(3,1:43) = '3,0 1996,1 1995,2 1994,3 1993,4 1992,5 1991'
putlab(lab,colid)
putnum(y)
putdn('HIV/AIDS incidence of males in Pattani province ')
setvar y=3 x=1 z=6
orplot log=1
getfile mhall.num
y=getnum;
fp90m = ((y(:,2)==5) & (y(:,4)==1)); %males in Yala
y=y(fp90m,:);
y(:,3) = 5-y(:,3);
[lab, colid] = getlab;
```

```
lab(3,1:43) = '3,0 1996,1 1995,2 1994,3 1993,4 1992,5 1991'
putlab(lab,colid)
putnum(y)
putdn('HIV/AIDS incidence of males in Yala province ')
setvar y=3 x=1 z=6
orplot log=1
getfile mhall.num
y=getnum;
fp90m = ((y(:,2)==6) & (y(:,4)==1)); %males in Narathiwat
y=y(fp90m,:);
y(:,3) = 5-y(:,3);
[lab, colid] = getlab;
lab(3,1:43) = '3,0 1996,1 1995,2 1994,3 1993,4 1992,5 1991'
putlab(lab,colid)
putnum(y)
putdn('HIV/AIDS incidence of males in Narathiwat province ')
setvar y=3 x=1 z=6
orplot log=1
```

The crude odds ratios for females are these plotted using a straightforward modification of this program. A further slight modification of the program, program the Mantel-Heanszel-adjusted odds ratios. The following program is used.

```
system_dependent(14,'on')
getfile mhall.num
y=getnum;
fp90m = ((y(:,2)==0) & (y(:,4)==1)); %males in Songkhla
y=y(fp90m,:);
y(:,3) = 5-y(:,3);
[lab, colid] = getlab;
```

```
lab(3,1:43) = '3,0 1996,1 1995,2 1994,3 1993,4 1992,5 1991'
putlab(lab,colid)
putnum(y)
putdn('HIV/AIDS incidence of males in Songkhla province ')
setvar y=3 x=1 z=6
orplot log=1
getfile mhall.num
y=getnum;
fp90m = ((y(:,2)==1) & (y(:,4)==1)); %males in Satun
y=y(fp90m,:);
y(:,3) = 5-y(:,3);
[lab, colid] = getlab;
lab(3,1:43) = '3,0 1996,1 1995,2 1994,3 1993,4 1992,5 1991'
putlab(lab,colid)
putnum(y)
putdn('HIV/AIDS incidence of males in Satun province ')
setvar y=3 x=1 z=6
orplot log=1
getfile mhall.num
y=getnum;
fp90m = ((y(:,2)==4) & (y(:,4)==1)); %males in Pattani
y=y(fp90m,:);
y(:,3) = 5-y(:,3);
[lab, colid] = getlab;
lab(3,1:43) = '3,0 1996,1 1995,2 1994,3 1993,4 1992,5 1991'
putlab(lab,colid)
putnum(y)
putdn('HIV/AIDS incidence of males in Pattani province ')
setvar y=3 x=1 z=6
orplot log=1
```

```
getfile mhall.num
y=getnum;
fp90m = ((y(:,2)==5) & (y(:,4)==1)); %males in Yala
y=y(fp90m,:);
y(:,3) = 5-y(:,3);
[lab, colid] = getlab;
lab(3,1:43) = '3,0 1996,1 1995,2 1994,3 1993,4 1992,5 1991'
putlab(lab,colid)
putnum(y)
putdn('HIV/AIDS incidence of males in Yala province ')
setvar y=3 x=1 z=6
orplot log=1
getfile mhall.num
y=getnum;
fp90m = ((y(:,2)==6) & (y(:,4)==1)); %males in Narathiwat
y=y(fp90m,:);
y(:,3) = 5-y(:,3);
[lab, colid] = getlab;
lab(3,1:43) = '3,0 1996,1 1995,2 1994,3 1993,4 1992,5 1991'
putlab(lab,colid)
putnum(y)
putdn('HIV/AIDS incidence of males in Narathiwat province ')
setvar y=3 'x=1 5' z=6
orplot log=1
```

Finally. The logistic regression model may be used to estimate the odds ratio for any combination of predictor variable compared to any other combination. the following program is used.

```
system_dependent(14,'on')
getfile mhall.num
y=getnum;
n=size(y,1);
m=n/2;
y0=y(1:m,:);
y1=y(m+1:n,:);
y=[y0(:,6) y1(:,2:6)];
male = (y(:,4)==1);
ym = y(male,:);
ym(:,3) = 5-ym(:,3);
[lab,collD] = getlab;
lab(3,3:43) = '0 1996,1 1995,2 1994,3 1993,4 1992,5 1991';
putlab(lab,collD)
pyint = ym(:,2)*10+ym(:,3);
ym = [ym pyint];
putnum(ym);
fn = getfn;
fn = str2mat('population',fn(2:6,:),'prov.year');
putfn(fn)
putdn('male hiv/aids incidence 1991-96')
setvar z=1 'x=2 3 5' y=6
lreg prop=1 ncat=16 show=3 font=6 new=0
setvar z=1 'x=7 5' y=6
lreg prop=1 ncat=31 font=7 show=1 out=1
predstres=getnum('res=1');
stres = [predstres(:,2) ym(:,2:6)];
stres(stres(:,1)>3,:)
lreg prop=1 ncat=31 font=7 show=2 out=1
```