### APPENDIX: DATA STRUCTURE AND PROGRAM USING

#### A. Data structure

The raw data for the HIV/AIDS cases in the five provinces of Southern Thailand in 1991-1996 are provided by the 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:

<table>
<thead>
<tr>
<th>Code</th>
<th>Diagnosis</th>
<th>Sex</th>
<th>Age</th>
<th>Date</th>
<th>Status</th>
<th>Behaviour</th>
<th>Method of Transmission</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1 24</td>
<td>1</td>
<td>1</td>
<td>43</td>
<td>90</td>
<td>10/05/91</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>2 1 48</td>
<td>1</td>
<td>1</td>
<td>61</td>
<td>90</td>
<td>14/06/93</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1 1 32</td>
<td>1</td>
<td>1</td>
<td>91</td>
<td>90</td>
<td>12/10/92</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1 1 18</td>
<td>4</td>
<td>1</td>
<td>91</td>
<td>90</td>
<td>02/07/94</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1 2 2 5</td>
<td>1</td>
<td>1</td>
<td>61</td>
<td>90</td>
<td>01/05/93</td>
<td>1</td>
<td>9</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Code</th>
<th>Diagnosis</th>
<th>Sex</th>
<th>Age</th>
<th>Date</th>
<th>Status</th>
<th>Behaviour</th>
<th>Method of Transmission</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 2 3</td>
<td>1</td>
<td>1</td>
<td>61</td>
<td>91</td>
<td>01/04/92</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1 1 33</td>
<td>2</td>
<td>1</td>
<td>45</td>
<td>91</td>
<td>10/04/93</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2 1 2</td>
<td>1</td>
<td>1</td>
<td>61</td>
<td>91</td>
<td>06/09/93</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>1 1 2 6</td>
<td>1</td>
<td>1</td>
<td>61</td>
<td>91</td>
<td>19/09/94</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1 1 4 0</td>
<td>1</td>
<td>1</td>
<td>43</td>
<td>91</td>
<td>29/07/94</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Code</th>
<th>Diagnosis</th>
<th>Sex</th>
<th>Age</th>
<th>Date</th>
<th>Status</th>
<th>Behaviour</th>
<th>Method of Transmission</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1 3 2</td>
<td>1</td>
<td>1</td>
<td>61</td>
<td>91</td>
<td>01/05/96</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>2 1 2 3</td>
<td>1</td>
<td>1</td>
<td>61</td>
<td>94</td>
<td>01/04/94</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1 1 2 4</td>
<td>2</td>
<td>1</td>
<td>43</td>
<td>94</td>
<td>10/04/95</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2 1 2 9</td>
<td>1</td>
<td>1</td>
<td>61</td>
<td>94</td>
<td>06/09/93</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>1 1 4 4</td>
<td>2</td>
<td>1</td>
<td>43</td>
<td>94</td>
<td>10/04/95</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Year</th>
<th>Sex</th>
<th>Age Group</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>98</td>
<td>f</td>
<td>0-4</td>
<td>13300</td>
</tr>
<tr>
<td>90</td>
<td>f</td>
<td>5-9</td>
<td>14170</td>
</tr>
<tr>
<td>90</td>
<td>f</td>
<td>10-14</td>
<td>12293</td>
</tr>
<tr>
<td>96</td>
<td>m</td>
<td>0-4</td>
<td>8082</td>
</tr>
<tr>
<td>90</td>
<td>m</td>
<td>5-9</td>
<td>10033</td>
</tr>
<tr>
<td>90</td>
<td>m</td>
<td>10-14</td>
<td>10020</td>
</tr>
<tr>
<td>Year</td>
<td>Sex</td>
<td>Age</td>
<td>Value</td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
<td>-----</td>
<td>-------</td>
</tr>
<tr>
<td>91</td>
<td>f</td>
<td>0-4</td>
<td>6648</td>
</tr>
<tr>
<td>91</td>
<td>f</td>
<td>5-9</td>
<td>4891</td>
</tr>
<tr>
<td>91</td>
<td>f</td>
<td>10-14</td>
<td>8197</td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
<td>-----</td>
<td>-------</td>
</tr>
<tr>
<td>91</td>
<td>m</td>
<td>0-4</td>
<td>4186</td>
</tr>
<tr>
<td>91</td>
<td>m</td>
<td>5-9</td>
<td>3171</td>
</tr>
<tr>
<td>91</td>
<td>m</td>
<td>10-14</td>
<td>11207</td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
<td>-----</td>
<td>-------</td>
</tr>
<tr>
<td>94</td>
<td>f</td>
<td>0-4</td>
<td>9702</td>
</tr>
<tr>
<td>94</td>
<td>f</td>
<td>5-9</td>
<td>22293</td>
</tr>
<tr>
<td>94</td>
<td>f</td>
<td>10-14</td>
<td>10082</td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
<td>-----</td>
<td>-------</td>
</tr>
<tr>
<td>94</td>
<td>m</td>
<td>0-4</td>
<td>10039</td>
</tr>
<tr>
<td>94</td>
<td>m</td>
<td>5-9</td>
<td>8820</td>
</tr>
<tr>
<td>94</td>
<td>m</td>
<td>10-14</td>
<td>12293</td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
<td>-----</td>
<td>-------</td>
</tr>
<tr>
<td>95</td>
<td>f</td>
<td>0-4</td>
<td>8882</td>
</tr>
<tr>
<td>95</td>
<td>f</td>
<td>5-9</td>
<td>7833</td>
</tr>
<tr>
<td>95</td>
<td>f</td>
<td>10-14</td>
<td>10120</td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
<td>-----</td>
<td>-------</td>
</tr>
<tr>
<td>95</td>
<td>m</td>
<td>0-4</td>
<td>8882</td>
</tr>
<tr>
<td>95</td>
<td>m</td>
<td>5-9</td>
<td>7833</td>
</tr>
<tr>
<td>95</td>
<td>m</td>
<td>10-14</td>
<td>10120</td>
</tr>
<tr>
<td>Year</td>
<td>Gender</td>
<td>Age</td>
<td>Country</td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>-----</td>
<td>-------------</td>
</tr>
<tr>
<td>96</td>
<td>f</td>
<td>3-4</td>
<td>7482</td>
</tr>
<tr>
<td>96</td>
<td>f</td>
<td>5-9</td>
<td>7633</td>
</tr>
<tr>
<td>96</td>
<td>f</td>
<td>10-14</td>
<td>10140</td>
</tr>
</tbody>
</table>

The data were grouped using the SPIDA statistical package (Gebski et al. 1992), to create records matching and joined to the population data, creating the dataset `caseprep.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=marrired, 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, 92= pattani, 93= yala, 96= Narathiwat)
  8: date of diagnosis (dd/mm/yy)
  9: status (1=alive, 3=dead)
  10: behaviour (1=homosexual, 2=hetero, 3=besexual, 9=unknown)
  11: transmission (1=ivio, 2=sex, 3=donor, 4=mother, 9=unknown)
```

Recod date: into years (1991 - 0, 1992 - 1, etc.).

```
Sdays := case[8] - "31/12/90"
```
$jan1 := 1.368;731;1097;1442;1807;2172
$dec31 := 365;730;1096;1441;1806;2171;2437
$year := rec($days:=($jan1,$dec31,(0;6)))
$y := case[(1;7),5year,case[(9;11)]

! recode ages into age groups (0-4=0,5-9=1,10-14=2,...,75+=15)
$Sage := $y[3]
$lowerage := 5*(0;15)
$upperage := 5*(1;15;20):1
$ag := rec($Sage,$g=(lowerage,upperage,(0;15)))
$y := $y[(1;2),$Sage,$y[(4;11)]

! recode province (90=0,91=1,94=4,95=5,96=6)
$y := rec($y,col=7,=((90;91;94;95;96),(0;1;4;5;6)))

! create index for combination (province+year+gender+agegroup)
$index := 10000*$y[(7;3)]+1000*$y[[(8;2);$y[(3)]
$y := $y,$index
$ylab := case.nam;="index"
%label Sy $ylab
desc($y)

! separate aids cases and hiv cases
$sids := $y[(1)="1" ?: $y
$shiv := $y[(1)="2" ?: $y

! separate ivdu cases and sex cases
$sivdu := $y[(11)="1" ?: $y
$sex := $y[(11)="2" ?: $y

! stratify cases by index
$z := condense($y,y=12,x=7;8;2;3,typ=3,data=1)
$mr := rdim($z)
$z := $z[(2,$mr-1)[(1;6)
$szids := condense($szids,y=12,x=7;8;2;3,typ=3,data=1)
$sr <- rlm($szaids)
$szaids <- szaids[2,]$sr-1][1:6]
$zhiv <- condense($zhiv,y=12,x=7;8;2;3,typ=3,date=1)
$sr <- rlm($zhiv)
$zhiv <- $zhiv[2,]$sr-1][1:6]
$Sivdu <- condense($Sivdu,y=12,x=7;8;2;3,typ=3,date=1)
$sr <- rlm($Sivdu)
$Sivdu <- $Sivdu[2,]$sr-1][1:6]
$sex <- condense($sex,y=12,x=7;8;2;3,typ=3,date=1)
$sr <- rlm($sex)
$sex <- $sex[2,]$sr-1][1:6]

! Create stratified population counts.
! from data stored in province files (Sengkha.dat, syn.dat)
! as province, gender, rum, year, agegroup(0-4, 5-9, ..., 75+), count
$y = songpot syn
$y <- rec($y,col=1,=((90;94;95;96),(0;1;4;5;6)))
$y <- rec($y,col=2,="m","f"),(1,2))
$y <- rec($y,col=3,="(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,=($ag,0,15))
$y <- $index,by=[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,$szaids,typ=3)
$y := y[0]$
$y := join($y, z[1]$, 'type' = 7)
$y := y[0]$
$y := join($y, z[2]$)
$y := y[0]$
$y := join($y, z[3]$, 'type' = 3)
$y := y[0]$
$y := join($y, z[4]$, 'type' = 3)
$y := y[0]$
$ylab := ylab,"total","aids","hiv","ivdu","sex"
label y ylab
desc(y)
$totinc := round(100000*y[7]/y[6])$
$sum := round($sum, 16, 7$)
$sf := round($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
  101  0  0  1  1  63377  0  0  0  0
  102  0  0  1  2  61603  0  0  0  0
  103  0  0  1  3  61765  0  0  0  0
  104  0  0  1  4  56384  1  0  1  1
  105  1  0  1  5  5393  3  1  2  1  1
  106  0  0  1  9  44078  0  0  0  0
  107  0  0  1  7  38737  1  0  0  0
  108  0  0  1  8  33124  1  1  0  1
  109  0  0  1  9  25412  0  0  0  0
  110  0  0  1  2  24579  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, pputnum, pfun, pack, describe, savevar, track) in the ASP (McNeil et al 1997) package, the following program is used.

```matlab
system dependant(14,'on');
getfile caseexpnum
y=getnum;
y(:,6)=y(:,6)/1000;
putnum(y)
f = getfn;
f = str2num(f(1:5,:)/pop(1000,:),f(7:11,:));
putfun(f)
%describe hist=1 col=2,11
for j=7:11
y(:,j)=y(:,j)/y(:,6);
```
end
putin(ein)
s0 = ( y(1,2) = 0 | y(1,2) = 1 | y(1,2) = 4 | y(1,2) = 5 | y(1,2) = 6 ) & ( y(1,3) = 5 )
s0 = y(s0,);
pnum(s0)
paddx(HIV/AIDS incidence/1000 in 5 provinces in 1994)
describe hist=1 col=2:11
female = s0(1,4) = 0;
s0(female,7:11) = s0(female,7:11)-0.01;
male = s0(1,4) = 1;
s0(male,7:11) = s0(male,7:11)+0.01;
pnum(s0)
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 is 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 an additional column identifying no cases as 0s and cases as 1s is added 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 62377
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 8 61151
0 0 0 1 3 51423
0 0 0 1 4 41259
0 0 0 1 5 21435
0 0 0 1 2 25239
0 0 0 1 6 55078
0 0 0 1 4 39737
0 0 0 1 2 61151
```

...
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.

```plaintext
system dependent('on')
getfile mhail.num
y=gennum;
fp90m = ((y(:,2)==0) & (y(:,4)==1)); %males in Songkhla
y=y(fp90m,:);
y(:,3) = 5-y(:,2);
[lab, colid] = getlab;
putlab(lab,colid)
putnum(y)
puslin('HIV/AIDS incidence of males in Songkhla province')
setvar y=3 x=1 z=6
opplot log=1
```
getfile mhall.mun
y=getnum;
f90m = (y(.,2)==1) & (y(.,4)==1)); % males in Satun
y=y(f90m.);
y(.,3) = 5-y(.,3);
[lab, colid] = getlab;
putlab(lab, colid)
putnam(y)
p udało('HIV/AIDS incidence of males in Satun province ')
setvar y=3 x=1 z=6
orplot log=1
getfile mhall.mun
y=getnum;
f90m = (y(.,2)==4) & (y(.,4)==1)); % males in Pattani
y=y(f90m.);
y(.,3) = 5-y(.,3);
[lab, colid] = getlab;
putlab(lab, colid)
putnum(y)
p удалось('HIV/AIDS incidence of males in Pattani province ')
setvar y=2 x=1 z=6
orplot log=1
getfile mhall.mun
y=getnum;
f90m = (y(.,2)==5) & (y(.,4)==1)); % males in Yala
y=y(f90m.);
y(.,2) = 5-y(.,2);
[lab, colid] = getlab;}
The crude odds ratios for females are those plotted using a straightforward modification of this program. A further slight modification of the program, program the Mantel-Haenszel-adjusted odds ratios. The following program is used.

```
system dependent(14,'on')
getfile mhail.num
y=getnum;
p90m = ((y(:,2)==6) & (y(:,4)==1)), %males in Narathiwat
y=y(p90m,);
y(:,3) = 5*y(:,3);
[lab, colid] = getlab;
pubbh(lab,colid)
pudfn("HIV/AIDS incidence of males in Yala province")
setvar y=3 x=1 z=6
orect log=1
getfile mhall.num
y=getnum;
p90m = ((y(:,2)==6) & (y(:,4)==1)), %males in Narathiwat
y=y(p90m,);
y(:,3) = 5*y(:,3);
[lab, colid] = getlab;
pubbh(lab,colid)
pudfn("HIV/AIDS incidence of males in Narathiwat province")
setvar y=3 x=1 z=6
orect log=1
```
pus(t) lab(colid)
pus(t) nnum(y)
pus(t) nnum(HIV/AIDS incidence of males in Songkhla province)
sset var y=3 x=1 x=6
or plot log=1
gset(f) title nhall num
y=geta num,
f90m = (y(.,2)==1) & (y(.,4)==1)); %males in Satan
y=y(f90m,);
ys(.,3) = 5-y(.,3);
[lab, colid] = gset lab;
pus(t) lab(colid)
pus(t) nnum(y)
pus(t) nnum(HIV/AIDS incidence of males in Songkhla province)
sset var y=3 x=1 x=6
or plot log=1
gset(f) title nhall num
y=geta num;
f90m = ((y(.,2)==4) & (y(.,4)==1)); %males in Pattani
y=y(f90m,);
y(.,3) = 5-y(.,3);
[lab, colid] = gset lab;
pus(t) lab(colid)
pus(t) nnum(y)
pus(t) nnum(HIV/AIDS incidence of males in Pattani province)
sset var y=3 x=1 x=6
or plot log=1
getfile nhall.men
y=getnum;
fp90m = (y(,2)==5) & (y(,4)==1); % males in Yala
y=y(fp90m,1);
y(,3) = 5-y(,3);
[lab, colid] = getlab;
putnum(y)
putdn(HIV/AIDS incidence of males in Yala province )
sset var y=3 x=1 z=6
orplot log=1
getfile nhall.men
y=getnum;
fp90m = (y(,2)==6) & (y(,4)==1); % males in Narathiwat
y=y(fp90m,1);
y(,3) = 5-y(,3);
[lab, colid] = getlab;
putnum(y)
putdn(HIV/AIDS incidence of males in Narathiwat province )
sset var 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 mhaul.num
y = getnum;
n = size(y, 1);
m = n/2;
y0 = y(1:m, :);
y1 = y(n+1:m, :);
y = [y0(:, 6), y1(:, 2:6)];
males = (y(:, 4) == 1);
ym = y(males, :);
ym(:, 3) = 5 - ym(:, 3);
[lab, colID] = getlab;
putilab(lab, colID)
pyint = ym(:, 2)*10 + ym(:, 3);
yu = [ym pyint];
putilnu(yu);
fn = getfa;
fn = str2mat('population', [n, 2:6, 'provyear']);
purt(f)
putnu('males hiv/aids incidence 1991-96')
svar x:1 y:2 3 5 y:6
lreg prop=1 ncat=16 show=3 font=6 new=6
svar x:1 y: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