# Methodology

# Methodology

Population statistics inevitably involve comparisons of sub-populations, regions, and time periods. To many, such comparisons are often too complex so that interpretation becomes a formidable task. However, those comparisons are necessary in order to understand the health status of specific populations within British Columbia. The text that accompanies the tables and figures in this report explains the basic meaning of the comparisons but, for some, a more in-depth explanation is useful and necessary.

This section provides the reader with computational examples of how various measures are calculated. All data shown in the examples are hypothetical. These routines are referenced in the discussion accompanying specific tables and figures where they are used and are arranged alphabetically. In some cases, a test of statistical significance is noted in the discussion and those routines will be found at the end of this part of the report.

Examples of the these statistical computations follow:

**RATES**
- Age Standardized Mortality Rate (ASMR)
- Fertility Rates
  - Total Fertility Rate (TFR)
  - Age Specific Fertility Rates (ASFRs)
- Potentional Years of Life Lost (PYLL) and Standardized Rate (PYLLSR)

**RATIOS**
- Observed versus Expected Ratios
  - Low Birth Weight (LBW) Live Births
  - Potential Years of Life Lost Index (PYLLI)
  - Standardized Mortality Ratio (SMR)

**ESTIMATION OF SMOKING ATTRIBUTABLE MORTALITY (SAM)**

**STATISTICAL TESTS OF SIGNIFICANCE**
- Chi-Square
- Confidence Intervals
- P-Value

**RATES**
- Age Standardized Mortality Rate (ASMR)

Although a hypothetical LHA is used in the example cited here, the ASMR was also calculated for yearly death data, for example Figure 16, and specific cause groups, for example Table 21, to permit comparisons between items in those tables or figures. The example shown below can be applied to those measures as well. The test of statistical significance is described under Rates in Statistical Tests of Significance at the end of this Appendix.

| | | LHA | | |
|---|---|---|---|---|
| Age Group (i) | Standard Population ($\pi_i$) | Estimated Population ($p_i$) | Death Rate/10,000 ($m_i$) | Observed Deaths ($d_i$) |
| < 1 | 403,061 | 1,339 | 22.4 | 3 |
| 1 – 4 | 1,550,285 | 5,483 | 1.8 | 1 |
| . | . | . | . | . |
| . | . | . | . | . |
| . | . | . | . | . |
| 80 – 84 | 382,303 | 1,198 | 701.2 | 84 |
| 85 + | 287,877 | 908 | 1596.9 | 145 |
| TOTAL | 28,120,065 | 81,016 | | 561 |

***For the Local Health Area:***

$$ASMR = \frac{\sum m_i \times \pi_i}{\prod} = \frac{22.4 \times 403{,}061 + \ldots + 1{,}596.9 \times 287{,}877}{28{,}120{,}065} = 46.2$$

Where: $p_i$ =area population in age group i;
$\pi_i$ =standard population in age group i;
$\prod$ $= \sum \pi_i$ = total standard population;
$d_i$ =deaths in LHA population in age group i; and
$m_i$ $= d_i / p_i \times 10{,}000$ = mortality rate per 10,000 LHA population in age group i.

e.g., $m_i = \frac{3 \times 10{,}000}{1{,}339} = 22.4$, for age group 1.

- Fertility Rates

Fertility Rates include the Total Fertility Rate (TFR) and Age Specific Fertility Rates (ASFRs). Although the TFR is calculated for a hypothetical LHA in the example cited here, the calculation method was applied to each year in Table 3 and Figure 5 and to each of the LHAs in Table 10. The teenage fertility rates shown in Table 10 and Figure 29 are the teenage-specific fertility rates, that is the ASFRs for 15-19 year olds, exemplified below.

| Age Group (i) | LHA | | |
|---|---|---|---|
| | Live Births ($b_i$) | Female Population ($w_i$) | Age Specific Fertility Rate ($ASFR_i$) |
| 15 – 19 | 19 | 598 | 31.8 |
| 20 – 24 | 46 | 440 | 104.5 |
| 25 – 29 | 74 | 498 | 148.6 |
| 30 – 34 | 51 | 745 | 68.5 |
| 35 – 39 | 12 | 690 | 17.4 |
| 40 – 44 | 2 | 581 | 3.4 |
| TOTAL | 204 | 3,552 | 374.2 |

***For the Local Health Area:***

1) the age specific fertility rate (ASFR) for age group 15–19 years is:

$$ASFR_i = \frac{b_i}{w_i} \times 1{,}000 = \frac{19}{598} \times 1{,}000 = 31.8$$

Where: $b_i$ =number of live births for age group i; and
$w_i$ =number of female population for age group i.

2) the total fertility rate (TFR) is:

$$TFR = a \times \sum ASFR_i = 5 \times (31.8 + \ldots + 3.4) = 1{,}871$$

Where: $ASFR_i$ =age specific fertility rate for age group i; and
$a$ =number of years in each age group i.

- Potential Years of Life Lost (PYLL) and Standardized Rate (PYLLSR)

The Potential Years of Life Lost (PYLL) measures presented in this report are based on the number of years of life lost when a person dies before the age of 75 years. Infant deaths (age less than one year old) are included.

| Age Group (i) | Age Factor ($75-Y_i$) | Standard Population ($\pi_i$) | LHA Estimated Population ($p_i$) | LHA Death Rate/1,000 ($m_i$) | LHA Observed Deaths ($d_i$) | LHA Observed PYLL ($d_i(75-Y_i)$) |
|---|---|---|---|---|---|---|
| < 1 | 74.5 | 403,061 | 1,339 | 2.2 | 3 | 223.5 |
| 1 – 4 | 72.0 | 1,550,285 | 5,483 | 0.2 | 1 | 72.0 |
| 5 – 9 | 67.5 | 1,953,045 | 6,553 | 0.2 | 1 | 67.5 |
| . | . | . | . | . | . | . |
| . | . | . | . | . | . | . |
| . | . | . | . | . | . | . |
| 65 – 69 | 7.5 | 1,084,588 | 3,538 | 18.7 | 66 | 495.0 |
| 70 – 74 | 2.5 | 834,024 | 2,779 | 28.8 | 80 | 200.0 |
| TOTAL | | 28,120,065 | 79,140 | | 239 | 3,183.0 |

***For the Local Health Area:***

$$\text{PYLL} = \sum d_i \times (75 - Y_i)$$

Where: $d_i$ =number of deaths in age group i;
$Y_i$ =age at midpoint of age group i; and
$\sum$ =summation.

$$PYLLSR = \frac{\sum m_i \times \pi_i \times (75 - Y_i)}{\prod} = \frac{2.2 \times 403{,}061 \times 74.5 + \ldots + 28.8 \times 834{,}024 \times 2.5}{28{,}120{,}065} = 37.0$$

Where: $p_i$ =LHA population in age group i;
$\pi_i$ =standard population in age group i;
$\prod = \sum \pi_i$ = total standard population;
$d_i$ =deaths in LHA population in age group i;
$Y_i$ =age at midpoint of age group i; and
$m_i = (d_i/p_i) \times 1{,}000$ = mortality rate per 1,000 LHA population in age group i.

**RATIOS**

- Observed versus Expected Ratios

The following are hypothetical examples that apply to the vital event ratios shown in this report. The first example shows low birth weight (LBW) live births (less than 2,500 grams), but other live birth ratios, such as cesarean deliveries or live births with maternal or perinatal complications, as well as infant deaths ratios can be substituted. Tables 12, 16, 18, 20, and 26 and Figures 30, 32, 33, 34, and 36 present these ratios. Ratios for live births to teenage mothers, elderly gravida live births, pre-term live births, or live births by cesarean, although not shown in this report, would also be calculated the same way as the low birth weight ratios. These ratios based on live births should not be confused with observed versus expected ratios that involve age and gender standardization, such as Standardized Mortality Ratio (SMR) and Potential Years of Life Lost Index (PYLLI). The test of statistical significance is described under Ratios in Statistical Tests of Significance at the end of this Appendix.

- Low Birth Weight Live Births

| Year (i) | LHA: Low Birth Weight Live Births Observed ($O_i$) | LHA: Low Birth Weight Live Births Expected ($E_i$) | LHA: Total Live Births ($L_i$) | British Columbia: Low Birth Weight Live Births Observed ($b_i$) | British Columbia: Total Live Births ($B_i$) |
|---|---|---|---|---|---|
| 1995 | 92 | 82.9 | 1,701 | 2,096 | 42,989 |
| 1996 | 69 | 74.6 | 1,588 | 1,965 | 41,846 |
| 1997 | 102 | 80.2 | 1,582 | 2,113 | 41,655 |
| 1998 | 85 | 74.7 | 1,495 | 2,145 | 42,913 |
| 1999 | 91 | 78.1 | 1,501 | 2,267 | 43,586 |
| TOTAL | 439 | 390.6 | 7,867 | 10,586 | 212,989 |

***For the Local Health Area:***

1) the expected low birth weight live births for year i = 1995 were:

$$E_i = \frac{b_i}{B_i} \times L_i = \frac{2{,}096}{42{,}989} \times 1{,}701 = 82.9$$

Where: $b_i$ =number of LBW live births for the province in year i;
$B_i$ =number of live births for the province in year i; and
$L_i$ =number of live births for the LHA.

2) the ratio of observed over the expected LBW live births for the five-year period was:

$$\text{Ratio} = \frac{\sum O_i}{\sum E_i} = \frac{92 + \ldots + 91}{82.9 + \ldots + 78.1} = \frac{439}{390.6} = 1.1$$

Where: $O_i$ =observed LBW live births for year i; and
$E_i$ =expected LBW live births for year i.

- Potential Years of Life Lost Index (PYLLI)
***Note that this method is both age and gender standardized.***

| Age Group (i) | Gender (j) | Age Factor ($75-Y_{ij}$) | LHA: Estimated Population ($p_{ij}$) | LHA: Death Rate/1,000 ($m_{ij}$) | LHA: Observed Deaths ($d_{ij}$) | LHA: Observed PYLL ($d_{ij}(75-Y_{ij})$) | LHA: Expected PYLL ($e_{ij}(75-Y_{ij})$) | British Columbia: Estimated Population ($P_{ij}$) | British Columbia: Death Rate/1,000 ($D_{ij}/P_{ij} \times 1{,}000$) | British Columbia: Observed Deaths ($D_{ij}$) | British Columbia: Observed PYLL ($D_{ij}(75-Y_{ij})$) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| <1 | M | 74.5 | 1,339 | 2.2 | 3 | 223.5 | 766.3 | 42,700 | 7.7 | 328 | 24,436.0 |
| <1 | F | 74.5 | 1,301 | 1.8 | 2 | 177.3 | 620.8 | 40,600 | 6.4 | 260 | 19,380.3 |
| . | . | . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . | . | . |
| 70–74 | M | 2.5 | 1,587 | 71.3 | 113 | 282.8 | 233.2 | 65,500 | 58.8 | 3,969 | 9,921.4 |
| 70–74 | F | 2.5 | 2,779 | 28.8 | 80 | 200.0 | 182.3 | 107,000 | 26.2 | 2,807 | 7,017.5 |
| TOTAL | | | 79,140 | | 239 | 3,183.0 | 5,100.0 | 2,966,500 | | 11,068 | 200,265.5 |

***For the Local Health Area:***

$$PYLLI = \frac{O}{E} = \frac{\sum d_{ij} \times (75 - Y_{ij})}{\sum e_{ij} \times (75 - Y_{ij})} = \frac{223.5 + 177.3 + \ldots + 282.8 + 200.0}{766.3 + 620.8 + \ldots + 233.2 + 182.3} = \frac{3{,}183}{5{,}100} = 0.6$$

Where: O =observed PYLL;
E =expected PYLL;
$d_{ij}$ =observed deaths in age group i and gender j;
$e_{ij}$ =expected deaths in age group i and gender j;
$Y_{ij}$ =age at midpoint of age group i and gender j;
$p_{ij}$ =LHA population for age group i and gender j;
$P_{ij}$ =provincial population for age group i and gender j;
$D_{ij}$ =provincial deaths for age group i and gender j.

**1) Observed PYLL (O)**

The number of potential years of life lost (PYLL) based on the number and age at death of deaths that occurred in the LHA. For example, for age group under one year of age and gender j, the observed PYLL are:

Observed PYLL = deaths x age factor = $d_{ij}$ (75-$Y_{ij}$) = 3 x 74.5 = 223.5

**2) Expected PYLL (E)**

The number of potential years of life lost (PYLL) expected for residents of the LHA based on the PYLL from the expected deaths in the age group. For example, for age group under one year of age and gender j, the expected PYLL are:

Expected PYLL = expected deaths x age factor = $e_{ij}$ (75 – $Y_{ij}$) = $\frac{D_{ij}}{P_{ij}} \times p_{ij} \times (75 - Y_{ij})$

$$= \frac{328}{42{,}700} \times 1{,}339 \times 74.5 = 766.3$$

- Standardized Mortality Ratio (SMR)

***Note that this method is both age and gender standardized.***

| Age Group (i) | Gender (j) | LHA Estimated Population ($p_{ij}$) | LHA Death Rate/1,000 ($m_{ij}$) | LHA Observed Deaths ($d_{ij}$) | LHA Expected Deaths ($e_{ij}$) | British Columbia Estimated Population ($P_{ij}$) | British Columbia Death Rate/1,000 ($M_{ij}$) | British Columbia Observed Deaths ($D_{ij}$) |
|---|---|---|---|---|---|---|---|---|
| < 1 | M | 1,339 | 2.2 | 3 | 10.3 | 42,700 | 7.7 | 328 |
| < 1 | F | 1,301 | 1.8 | 2 | 8.3 | 40,600 | 6.4 | 260 |
| . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . |
| 85 + | M | 1,198 | 70.1 | 84 | 87.2 | 48,100 | 72.8 | 3,502 |
| 85 + | F | 908 | 159.7 | 145 | 138.8 | 34,500 | 152.8 | 5,272 |
| TOTAL | | 81,016 | | 561 | 595.1 | 3,131,700 | | 23,389 |

***For the Local Health Area:***

$$SMR = \frac{\sum d_{ij}}{\sum e_{ii}} = \frac{3 + 2 + \ldots + 110 + 145}{10.3 + 8.3 + \ldots + 92.6 + 138.8} = \frac{561}{595.1} = 0.9$$

Where: $d_{ij}$ =observed deaths in age group i and gender j; and
$e_{ij}$ =expected deaths in age group i and gender j.

**1) Observed Deaths (d)**

The actual number of deaths that occurred in the LHA. For example, for age group under one year of age and gender j, the observed deaths are three.

**2) Expected Deaths (e)**

The number of deaths expected for residents of the LHA based on the age specific mortality rates for the province as a whole and the population age structure of the LHA. For age group under one year and gender j, the expected deaths are:

$$e_{ij} = \frac{D_{ij}}{P_{ij}} \times p_{ij} = \frac{328}{42{,}700} \times 1{,}339 = 10.3$$

Where: $p_{ij}$ =LHA population for age group i and gender j;
$D_{ij}$ =provincial deaths for age group i and gender j; and
$P_{ij}$ =provincial population for age group i and gender j.

- Estimation of Smoking Attributable Mortality (SAM)

This report uses an estimation method to approximate the extent of smoking-attributable deaths based on the concept of attributable risk. To define attributable risk mathematically, consider $d_0$ and $d_1$ respectively to represent the death rates, in a given time period, in two cohorts from a population — those not exposed and those exposed to a given risk factor. The attributable risk of this factor, $AR_1$ , would then be:

$$AR_1 = \frac{d_1 - d_0}{d_1} = \frac{r_1 - 1}{r_1}$$

Where: $r_1 = d_1/d_0$ is the relative risk of the exposed cohort.

The relative risk of the unexposed cohort is $r_0 = 1$; the attributable risk of this cohort is $AR_0 = 0$.

The attributable risk (AR) for the population as a whole (exposed plus unexposed cohorts) is given by:

$$AR = \frac{p_1 (r_1 - 1)}{p_1 (r_1) + (1 - p_1) (r_0)} = \frac{(p_1) (r_1 - 1)}{(p_1) (r_1 - 1) + 1}$$

Where: $p_1$ = the proportion or fraction of the population exposed to the risk factor; and
$1-p_1$ = the proportion or fraction of the population not exposed to the risk factor.

This may be extended to account for multiple levels of exposure, as follows:

$$AR = \frac{\sum_{i=1}^{n} p_i (r_i - 1)}{\sum_{i=1}^{n} p_i (r_i - 1) + 1}$$

Where: $p_i$ = the proportion (prevelance) of the population in the ith level of exposure group;
$r_i$ = the relative risk at the ith level of exposure; and
i = the ith risk category.

When applied to smoking-attributable mortality (SAM), the attributable risk is often expressed as a percentage:
SAM (%) = AR x 100

Smoking-attributable deaths are derived by multiplying the smoking-attributable mortality percentage expressed as a decimal fraction by the number of deaths aged 35+ in each of 19 specified cause of death categories. These categories are comprised of selected malignant neoplasms, circulatory system diseases, and respiratory system diseases, and are listed in the Glossary.

Relative-risk data from the American Society's Cancer Prevention Study (CPS-II) 1982–1988[1] were selected for use, as they have been widely used for similar analyses. The data from CPS–II established the age groups and the classification of smokers (current, former, and never) for which smoking prevalence data were required. The relative risk age categories were for 35+, or 35-64 and 65+. BC prevalence rates for smoking were provided in the **Tobacco Use in B.C. (1997)** survey commissioned by the BC and Yukon Health and Stroke Foundation.[2]

**STATISTICAL TESTS OF SIGNIFICANCE**

- Chi Square

For ratios, such as SMRs, a Chi-square ($\chi^2$) test is applied to determine whether the observed number of cases is statistically significantly different from the expected number. For LHA ***l***:

$$\chi_l^2 = \frac{(O_l - E_l)^2}{E_l}$$

(with one degree of freedom).

Where: $O_l$ =Observed number for LHA ***l***; and
$E_l$ =Expected number for LHA ***l***.

If $\chi_l^2 > 3.84$, the ratio is statistically significant at 5% significance level.

For SMR values, the Chi-square statistic that is applied is:

$$X_l^2 = 9\hat{O}_l\,(1 - \frac{1}{9\hat{O}_l} - (\frac{E_l}{\hat{O}_l})^{1/3})^2$$

Where: $\hat{O}_l$ $= O_l$ *if* $O_l > E_l$; *otherwise*
$\hat{O}_l$ $= O_l + 1.$

- Confidence Intervals

For rates, such as ASMRs, the test employed to determine statistical significance is a confidence interval. The 95% confidence interval for the difference (D) between a LHA and a provincial rate is defined by the upper and lower limits of the interval as follows:

$$\text{Lower Limit} = D - 1.96\sqrt{\frac{R_l^2}{O_l} + \frac{R_p^2}{O_p}}$$

$$\text{Upper Limit} = D + 1.96\sqrt{\frac{R_l^2}{O_l} + \frac{R_p^2}{O_p}}$$

[1]Centres for Disease Control. (1990). Smoking and health: A national status report. (DHSS publication no. (CDC) 87-8396). 2nd Edition. Rockville, MD: U.S. Department of Health and Human Services.
[2]Tobacco Use in B.C., ANGUS REID GROUP survey results, September 1997.

Where: $R_l$ =Rate for LHA ***l***;
$R_p$ =Rate for the province;
$O_l$ =Observed number for LHA ***l***; and
$O_p$ =Observed number for the province.

If the Lower Limit > 0, then $R_l$ is statistically significantly higher than $R_p$;
if the Upper Limit < 0, then $R_l$ is statistically significantly lower than $R_p$; otherwise,
there is no statistically significant difference.

- P Value

The p-value is the probability of rejecting the null hypothesis when a specified test procedure is used on a given data set. This probability is the smallest level of significance at which the null hypothesis would be rejected. Once the p-value has been determined, the conclusion at any particular level α results from comparing the p-value to α (e.g., 0.05):

(a) p-value $\leq \alpha$ $\Rightarrow$ reject null hypothesis at level α,

(b) p-value $> \alpha$ $\Rightarrow$ do not reject the null hypothesis at level α,

and we call the data statistically significant when the null hypothesis is rejected and not significant otherwise.