Stata Demo on Prevalence: Part I

1. **Objectives**
   a. Calculate the prevalence of smoking in the Framingham Data Set and interpret the results
   b. Restrict an analysis to non-missing data
   c. Create a 2x2 table to examine changes in self-reported smoking status between visit 1 and visit 2

2. **Calculate the proportion of people at each visit that report current smoking.**
   In this data set, current smoking status us coded as 0=not current smoker, 1=current smoker
   a. Dropdown:
      i. Statistics→ Summaries, tables and tests→Tables→ Multiple One Way Tables
      ii. cursmoke1 cursmoke2 cursmoke3 or cursmoke*
      iii. Check box for "treat missing values like other values"
      iv. Submit
   b. Command Window Syntax: `tab1 cursmoke1 cursmoke2 cursmoke3, missing`

3. **Calculate the proportion of people at each visit that report current smoking among those with data on smoking status at that visit.**
   a. Dropdown:
      i. Statistics→ Summaries, tables and tests→Tables→ Multiple One Way Tables
      ii. cursmoke1 cursmoke2 cursmoke3 or cursmoke*
      iii. Submit
   b. Command Window Syntax: `tab1 cursmoke1 cursmoke2 cursmoke3`

4. **Calculate the proportion of people at each visit that report current smoking among those with data on smoking status at all 3 visits.**
   A missing datum counts as infinity when making comparisons. Therefore, if the value is not missing, it is considered less than infinity.

   We can create a variable for no missing smoking status at any examination cycle:
   `gen cursmokenotmiss = cursmoke1<. & cursmoke2<. & cursmoke3<.`

   If all 3 cursmoke variables are less than infinity (i.e. not missing), then cursmokenotmiss equals true which is recorded as 1, but otherwise cursmokenotmiss equals false which is recorded as 0.

   a. Dropdown:
      i. Statistics→ Summaries, tables and tests→Tables→ Multiple One Way Tables
      ii. Categorical variables: `cursmoke1 cursmoke2 cursmoke3` or `cursmoke*`
      iii. Check box for "treat missing values like other values"
      iv. Under tab for by/if/in Restrict observations: `cursmokenotmiss==1`
      v. Submit
   b. Command Window Syntax: `tab1 cursmoke1 cursmoke2 cursmoke3 if cursmokenotmiss==1, missing`

5. **What could explain the declining prevalence of smoking?**
    a. Over time, the prevalence of smoking is declining in the population
    b. Current smokers have a shorter life
    c. Several smokers choose not to participate in the 2nd and 3rd visits

6. **Calculate the change in smoking prevalence between the 1st and 2nd visit.**
    a. Dropdown:
        i. Tables→ Two Way Tables with Measures of Association
        ii. Row variable: cursmoke1
        iii. Column variable: cursmoke2
        iv. Check boxes for
            1. "treat missing values like other values"
            2. Cell contents→"Within-column relative frequencies"
            3. Cell contents→"Within-row relative frequencies"
            4. Cell contents→"Relative frequencies"
        v. Submit
    b. Command Window Syntax: `tabulate cursmoke1 cursmoke2, cell column miss row`

7. **Calculate the change in smoking prevalence between the 1st and 2nd visit among those with data on smoking status at both visits.**
    a. Dropdown:
        i. Tables→ Two Way Tables with Measures of Association
        ii. Row variable: cursmoke1
        iii. Column variable: cursmoke2
        iv. Check boxes for
            1. Do not check box for "treat missing values like other values"
            2. Cell contents→"Within-column relative frequencies"
            3. Cell contents→"Within-row relative frequencies"
            4. Cell contents→"Relative frequencies"
        v. Submit
    b. Command Window Syntax: `tabulate cursmoke1 cursmoke2, cell column row`

8. **Conclusions**
    a. Smoking prevalence declined over time
        i. Smokers are quitting
        ii. Smokers have a shorter life
        iii. Smokers are less likely to participate
    b. Stata can be used to
        i. Restrict an analysis to non-missing data
        ii. Create a 2 x 2 table to cross-classify two nominal variables

Stata Demo on Prevalence: Part 2

1. **Objectives**
   a. Create an ordinal variable from continuous data
   b. Calculate the prevalence of CHD for different levels of smoking at visit 1

2. **Calculate the prevalence of coronary heart disease (CHD) at visit 1 by categories of cigarettes per day**

   PREVCHD is defined as pre-existing angina pectoris, myocardial infarction (hospitalized, silent or unrecognized), or coronary insufficiency (unstable angina)
   0=Free of disease, 1=Prevalent disease

   Create 4 categories of cigarette packs per day (0,1-20,21-40,≥41). Since the values reflect a particular ordering, it is an ordinal variable.

```
gen packs1=.
replace packs1=0 if (cigpday1==0)
replace packs1=1 if (cigpday1>=1 & cigpday1 <= 20)
replace packs1=2 if (cigpday1>=21 & cigpday1 <= 40)
replace packs1=3 if (cigpday1>=41 & cigpday1<.)
```

   a. Dropdown:
      i. Tables→ Two Way Tables with Measures of Association
      ii. Row variable: packs1
      iii. Column variable: prevchd1
      iv. Check boxes for
         1. Do not check box for “treat missing values like other values”
         2. Cell contents→”Within-column relative frequencies”
         3. Cell contents→”Within-row relative frequencies”
         4. Cell contents→”Relative frequencies”
   b. Command Window Syntax: `tabulate packs1 prevchd1, cell column row`

3. **What could explain the higher prevalence of CHD among non-smokers compared to those who smoke 1 or more cigarettes per day?**
   a. High incidence, Long duration
   b. Cross-sectional data is susceptible to reverse causation
   c. Other common suspects
      i. Bias
      ii. Confounding
      iii. Chance

4. **Conclusions**
   a. Stata can be used to create an ordinal variable based on continuous data.
   b. CHD prevalence was lower among people with higher levels of smoking.
   c. Prevalence is a function of incidence and duration.
   d. In addition to a causal effect of exposure on disease risk, there are several alternative explanations for observing an association between two factors of interest.

Stata Demo on Cumulative Incidence and Incidence Rate of Death

Objectives

1. Calculate the cumulative incidence of death among smokers and nonsmokers in Stata
2. Calculate the incidence rate of death among smokers and nonsmokers in Stata

1. Calculate the cumulative incidence of death among smokers and nonsmokers in Stata
    a. Dropdown
        i. Statistics → Epidemiology and related → Tables for epidemiologists → Cohort study risk-ratio etc.
        ii. Under "Case variable" select "death"; under "exposed variable" select "cursmoke1"
    b. Command window type: `cs death cursmoke1`

The value in the risk row is Cumulative Incidence

2. Calculate the incidence rate of death among smokers and nonsmokers in Stata
    a. Dropdowns:
        i. Statistics → Epidemiology and related → Tables for epidemiologists → Incidence rate-ratio
        ii. Under "Case variable" select "death"; under "exposed variable" select "cursmoke1"; under "person-time variable" select "timedth"
    b. Command window type: `ir death cursmoke1 timedth`

Stata Demo on Incidence Rate of CHD

Objectives

Calculate the incidence rate of CHD among smokers and nonsmokers in Stata

1. Calculate the incidence rate of CHD among smokers and nonsmokers in Stata
   a. Dropdowns:
      i. Statistics → Epidemiology and related → Tables for epidemiologists → Incidence rate-ratio
      ii. Under "Case variable" select "anychd"; under "exposed variable" select "cursmoke1"; under "person-time variable" select "timechd"
   b. Command window type: `ir anychd cursmoke1 timechd`

To find incidence rate of CHD, the time variable will be timechd !
Exposed Variable = cursmoke1
Case Variable = anychd

Stata Demo on Measures of Association

1. Objectives

a. Examine the association between smoking and death
   - i. risk difference
   - ii. risk ratio
   - iii. attributable fraction among the exposed
   - iv. attributable fraction among the total population
   - v. odds ratio
   - vi. rate difference
   - vii. rate ratio

b. Examine the association between smoking and coronary heart disease (CHD)
   - i. rate difference
   - ii. rate ratio

2. Calculate the association between smoking status at visit 1 (cursmoke1) and the 24-year <u>risk</u> and <u>odds</u> of death (death).

a. Dropdown:
   - i. Statistics→ Epidemiology and Related→Tables for Epidemiologists→Cohort study risk-ratio etc.
   - ii. Case variable: death
   - iii. Exposed variable: cursmoke1
   - iv. On the options tab, check box for “Report odds ratio” Add "or" to get the odds ratio
   - v. Submit

b. Command Window Syntax: `cs death cursmoke1,or`

c. Calculation of Results

risk difference	$Risk_{E+} - Risk_{E-} = 0.36 - 0.34 = 0.023$
risk ratio	$\frac{Risk_{E+}}{Risk_{E-}} = \frac{0.36}{0.34} = 1.07$
attributable fraction among the exposed	$\frac{RR-1}{RR} = \frac{1.07-1}{1.07} = 0.065$
attributable fraction among the total population	p=prevalence of exposure=2181/4434=0.49 $\frac{p(RR-1)}{1+p(RR-1)} = \frac{.49(1.07-1)}{1+.49(1.07-1)} = 0.033$
disease odds ratio	$\frac{Odds_{D+\|E+}}{Odds_{D+\|E-}} = \frac{R_{D+\|E+}/1-R_{D+\|E+}}{R_{D+\|E-}/1-R_{D+\|E-}} = \frac{\frac{788/2181}{1393/2181}}{\frac{762/2253}{1491/2253}} = \frac{788/1393}{762/1491} = \frac{0.566}{0.511} = 1.11$
exposure odds ratio	$\frac{Odds_{E+\|D+}}{Odds_{E+\|D-}} = \frac{R_{E+\|D+}/1-R_{E+\|D+}}{R_{E+\|D-}/1-R_{E+\|D-}} = \frac{\frac{788/1550}{762/1550}}{\frac{1393/2884}{1491/2884}} = \frac{788/762}{1393/1491} = \frac{1.03}{0.93} = 1.11$

3. Calculate the association between smoking status at visit 1 (cursmoke1) and the 24-year rate of death (death) over follow-up (timedth).

- a. Dropdown:
  - i. Statistics→ Epidemiology and Related→Tables for Epidemiologists→Incidence rate-ratio etc.
  - ii. Case variable: death
  - iii. Exposed variable: cursmoke1
  - iv. Person-time variable: timedth
  - v. Submit
- b. Command Window Syntax: `ir death cursmoke1 timedth`
- c. Calculation of Results

rate difference	$\text{Rate}_{E+} - \text{Rate}_{E-}$ = 0.0177-0.0163 = 0.0014 cases / person-year
rate ratio	$\frac{\text{Rate}_{E+}}{\text{Rate}_{E-}} = \frac{0.0177}{0.0163} = 1.09$

4. Calculate the association between smoking status at visit 1 (cursmoke1) and the 24-year rate of coronary heart disease (anychd) over follow-up (timechd).

- a. Dropdown:
  - i. Statistics→ Epidemiology and Related→Tables for Epidemiologists→Incidence rate-ratio etc.
  - ii. Case variable: anychd
  - iii. Exposed variable: cursmoke1
  - iv. Person-time variable: timechd
  - v. Submit
- b. Command Window Syntax: `ir anychd cursmoke1 timechd`
- c. Calculation of Results

rate difference	$\text{Rate}_{E+} - \text{Rate}_{E-}$ = 0.016-0.015 = 0.00048 cases / person-year
rate ratio	$\frac{\text{Rate}_{E+}}{\text{Rate}_{E-}} = \frac{0.016}{0.015} = 1.03$

5. Conclusions for Module 4.1

- a. Measures of disease frequency
  - i. risks
  - ii. odds
  - iii. rates
- b. Measures of association
  - i. difference measures
  - ii. ratio measures
  - iii. attributable fractions
- c. In this study,
  - i. positive association between smoking at visit 1 and risk/odds/rate of death
  - ii. positive association between smoking at visit 1 and rate of CHD

Objectives for Module 9.1 – Crude and Age-Adjusted Risk Ratio for Death

1. **Calculate the crude risk ratio of death for smokers compared to nonsmokers in Stata**
2. **Calculate the age-adjusted risk ratio of death for smokers compared to nonsmokers in Stata**

I. Calculate the crude risk ratio of death for smokers compared to nonsmokers in Stata

   a. Dropdown

      i. Statistics → Epidemiology and related → Tables for epidemiologists → Cohort study risk-ratio etc.

      ii. Under "Case variable" select "death"; under "exposed variable" select "cursmoke1"

   b. Command window type: `cs death cursmoke1`

II. Calculate the age-adjusted risk ratio of death for smokers compared to nonsmokers in Stata.

First, create 4 categories for age using the following code:

```
gen age4cat=.
replace age4cat=0 if (age1<=40)
replace age4cat=1 if (age1>40 & age1 <= 50)
replace age4cat=2 if (age1>50 & age1 <= 60)
replace age4cat=3 if (age1>60 & age1<.)
```

   a. Dropdown

      i. Statistics → Epidemiology and related → Tables for epidemiologists → Cohort study risk-ratio etc.

      ii. Under "Case variable" select "death"; under "exposed variable" select "cursmoke1"

      iii. Go to the "Options" tab; click the box next to "stratify on variables"; use the dropdown menu to select "age4cat"
      Note: Under "Within-stratum weights" the button next to "Use Mantel-Haenszel" should be automatically selected

   b. Command window type: `cs death cursmoke1, by(age4cat)`

Objectives for Module 9.2 – Crude and Age-Adjusted Incidence Rate Ratio for CHD

1. **Calculate the crude incidence ratio of CHD for smokers compared to nonsmokers in Stata**
2. **Calculate the age-adjusted incidence ratio of CHD for smokers compared to nonsmokers in Stata**

I. Calculate the crude risk ratio of CHD for smokers compared to nonsmokers in Stata
   a. Dropdowns:
      i. Statistics → Epidemiology and related → Tables for epidemiologists → Incidence rate-ratio
      ii. Under “Case variable” select “anychd”; under “exposed variable” select “cursmoke1”; under “person-time variable” select “timechd”
   b. Command window type: `ir anychd cursmoke1 timechd`

II. Calculate the age-adjusted risk ratio of death for smokers compared to nonsmokers in Stata.

First, create 4 categories for age using the following code:

```
gen age4cat=.
replace age4cat=0 if (age1<=40)
replace age4cat=1 if (age1>40 & age1 <= 50)
replace age4cat=2 if (age1>50 & age1 <= 60)
replace age4cat=3 if (age1>60 & age1<.)
```

   a. Dropdown
      i. Statistics → Epidemiology and related → Tables for epidemiologists → Incidence rate-ratio
      ii. Under “Case variable” select “anychd”; under “exposed variable” select “cursmoke1”; under “person-time variable” select “timechd”
      iii. Go to the “Options” tab; click the box next to “stratify on variables”; use the dropdown menu to select “age4cat”
      Note: Under “Within-stratum weights” the button next to “Use Mantel-Haenszel” should be automatically selected
   b. Command window type: `ir anychd cursmoke1 timechd, by(age4cat)`

Stata Demo on Effect Measure Modification and Standardization

1. Objectives

- a. Review concepts of
  - i. Confounding vs. effect measure modification
  - ii. Pooling vs. standardization
- b. Conduct a test of homogeneity to determine whether age modifies the association between smoking at visit 1 and the risk of cardiovascular disease.
- c. Calculate and interpret a summary estimate using standardization
  - i. Total population
  - ii. Unexposed individuals
  - iii. Exposed individuals

2. Confounding vs. Effect Measure Modification

- a. *Confounding*
  - i. Incorrect estimates due to the impact of a third factor that is associated with the exposure and a risk factor for the outcome independent of exposure; arises due to non-exchangeability (noncomparability) between the exposed and unexposed
  - ii. Results in an invalid estimate; we would like to remove confounding
  - iii. Not scale-dependent- if the ratio measure (e.g. rate ratio, risk ratio) is confounded, so is a difference measure (e.g. rate difference, risk difference)
  - iv. We can present stratum-specific estimates, standardized estimates or pooled estimate
- b. *Effect measure modification*
  - i. A third factor that modifies the strength of the association between the exposure-outcome association
  - ii. Provides useful information that we would like to highlight and describe in our findings
  - iii. Scale dependent: if the strength of the association between exposure and outcome varies for different subgroups (effect measure modification), it can be seen on one scale or on both scales, e.g. the rate ratios for two subgroups may be different but the rate differences for the two subgroups may be similar.
  - iv. We can present stratum-specific estimates or standardized estimates. A pooled summary estimate (e.g. Mantel-Haenszel adjusted estimate) is not appropriate.

3. Pooling vs. Standardization

- a. *Pooling:* a method for adjusting for confounding when differences between strata are due to sampling variability
- b. *Standardization:* a method to compare two populations with different distributions of a stratification factor(s) that confounds and/or modifies an exposure-disease association

4. **Test for Effect Measure Modification (EMM)**
    a. Test of homogeneity
        i. H_0: stratum-specific estimates are homogenous (no EMM)
        ii. H_A: At least one stratum estimate is different from the others (EMM)
        iii. Degrees of freedom= # strata – 1
    b. Large p-value: Do not reject null
        i. Insufficient evidence of effect measure modification
        ii. Report stratum-specific estimates or calculate Mantel-Haenszel summary measure
    c. Small p-value: Reject null
        i. Effect modification is present
        ii. Report stratum-specific estimates or standardized measure

5. **Conduct a test of homogeneity to determine whether age (age4cat) modifies the association between smoking at visit 1 (cursmoke1) and the risk of death.**

First, create the 4 categories for age that we have used in previous sessions with the following code:

```
gen age4cat=.
replace age4cat=0 if (age1<=40)
replace age4cat=1 if (age1>40 & age1 <= 50)
replace age4cat=2 if (age1>50 & age1 <= 60)
replace age4cat=3 if (age1>60 & age1<.)
```

To see the 2x2 tables of smoking by death for each age category:
    a. Dropdown:
        i. Statistics→ Summaries, tables, and tests→ Two way tables with measures of association
        ii. Main tab
            1. Row variable: death
            2. Column variable: cursmoke1
        iii. By/if/in tab
            1. Stratify on variables: age4cat
        iv. Submit
    b. Command Window Syntax: `by age4cat, sort : tabulate death cursmoke1`

Age ≤ 40

		Death		
		Yes	No	Total
Smoker	Yes	67	385	452
	No	25	277	302
	Total	92	662	754

50 < Age < 60

		Death		
		Yes	No	Total
Smoker	Yes	286	281	567
	No	312	500	812
	Total	598	781	1379

40 < Age < 50

		Death		
		Yes	No	Total
Smoker	Yes	266	689	955
	No	110	574	684
	Total	376	1263	1639

Age > 60

		Death		
		Yes	No	Total
Smoker	Yes	169	38	207
	No	315	140	455
	Total	484	178	662

Now, we can look at the stratum specific estimates, crude overall estimate, Mantel-Haenszel estimate and test of homogeneity.

c. Dropdown:
   i. Statistics→ Epidemiology and Related→ Tables for Epidemiologists→ Cohort study risk-ratio etc.
   ii. Main tab
      1. Case variable: death
      2. Exposed variable: cursmoke1
   iii. Options tab
      1. Stratify on variable: age4cat
      2. Within-stratum weights: Use Mantel-Haenszel weights (default)
   iv. Submit

d. Command Window Syntax: `cs death cursmoke1, by(age4cat)`

```
          age4cat |       RR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
                0 |    1.790619      1.158273   2.768188     14.98674
                1 |    1.731975      1.418997   2.113985     64.09396
                2 |    1.312757      1.165099   1.479129     128.2843
                3 |    1.179281      1.078835   1.289079     98.49698
-----------------+-------------------------------------------------
            Crude |     1.06826      .9858212   1.157592
      M-H combined |    1.381036      1.279253   1.490917
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(3) =    19.107  Pr>chi2 = 0.0003
```

6. **Calculate the association between smoking at visit 1 (cursmoke1) and the risk of death after adjusting for age (age4cat) by standardizing to the total population under study.**

In order to standardize to the total population in Stata, we need to tell Stata that each category of age should be weighted equally, so we create a new variable equal to the proportion of the population in that age category:

```
table age4cat
```

```
----------------------
  age4cat |      Freq.
----------+-----------
        0 |        754
        1 |      1,639
        2 |      1,379
        3 |        662
----------------------
```

754/4434=	0.1700496
1639/4434=	0.3696437
1379/4434=	0.3110059
662/4434=	0.1493009

```
gen all=.

replace all= 754/4434 if (age4cat==0)
replace all= 1639/4434 if (age4cat==1)
replace all= 1379/4434 if (age4cat==2)
replace all= 662/4434 if (age4cat==3)
```

THIS SECTION HAS BEEN REVISED:

PLEASE USE THIS CODE AND NOT THE CODE PRESENTED IN THE VIDEO!

a. Dropdown:
   i. Statistics→ Epidemiology and Related→ Tables for Epidemiologists→ Cohort study risk-ratio etc.
   ii. Main tab
      1. Case variable: death
      2. Exposed variable: cursmoke1
   iii. Options tab
      1. Stratify on variables: age4cat
      2. User-specified variable: all
   iv. Submit

b. Command Window Syntax: `cs death cursmoke1, by(age4cat) standard(all)`

```
         age4cat |       RR       [95% Conf. Interval]       Weight
-----------------+-------------------------------------------------
               0 |    1.790619     1.158273   2.768188     .1700496
               1 |    1.731975     1.418997   2.113985     .3696437
               2 |    1.312757     1.165099   1.479129     .3110059
               3 |    1.179281     1.078835   1.289079     .1493009
-----------------+-------------------------------------------------
           Crude |     1.06826     .9858212   1.157592
    Standardized |    1.372987     1.275679   1.477717
```

In a population with the age distribution of the **total population**, the risk of death is 1.37 times greater among smokers than among the nonsmokers.

7. Calculate the association between smoking at visit 1 (cursmoke1) the risk of death after adjusting for age (age4cat) by standardizing to the unexposed population (not current smokers at visit 1).

a. Dropdown:
   i. Statistics→ Epidemiology and Related→ Tables for Epidemiologists→ Cohort study risk-ratio etc.
   ii. Main tab
      1. Case variable: death
      2. Exposed variable: cursmoke1
   iii. Options tab
      1. Stratify on variables: age4cat
      2. Within-stratum weights: Use external
         estandard: external weights are the total number of unexposed
   iv. Submit

b. Command Window Syntax: `cs death cursmoke1, by(age4cat) estandard`

```
 age4cat |       RR       [95% Conf. Interval]       Weight
-----------------+-------------------------------------------------
               0 |    1.790619     1.158273   2.768188          302
               1 |    1.731975     1.418997   2.113985          684
               2 |    1.312757     1.165099   1.479129          812
               3 |    1.179281     1.078835   1.289079          455
-----------------+-------------------------------------------------
           Crude |     1.06826     .9858212   1.157592
 E. Standardized |    1.333775      1.24473    1.42919
```

In a population with the age distribution of the **non-smokers**, the risk of death is 1.33 times greater among smokers than among nonsmokers.

8. Calculate the association between smoking at visit 1 (cursmoke1) the risk of death after adjusting for age (age4cat) by standardizing to the exposed population (current smokers at visit 1).

a. Dropdown:
   i. Statistics→ Epidemiology and Related→ Tables for Epidemiologists→ Cohort study risk-ratio etc.
   ii. Main tab
      1. Case variable: death
      2. Exposed variable: cursmoke1
   iii. Options tab
      1. Stratify on variables: age4cat
      2. Within-stratum weights: Use internal
      istandard: internal weights are the total number of exposed
   iv. Submit

b. Command Window Syntax: `cs death cursmoke1, by(age4cat) istandard`

```
         age4cat |        RR       [95% Conf. Interval]       Weight
-----------------+-------------------------------------------------
               0 |    1.790619      1.158273   2.768188          452
               1 |    1.731975      1.418997   2.113985          955
               2 |    1.312757      1.165099   1.479129          567
               3 |    1.179281      1.078835   1.289079          207
-----------------+-------------------------------------------------
           Crude |     1.06826      .9858212   1.157592
 I. Standardized |      1.4271        1.3129   1.551233
```

In a population with the age distribution of the **smokers**, the risk of death among smokers is 1.43 times greater than among the nonsmokers.

9. Conclusions

a. Confounding and effect measure modification both involve a third factor, but they are separate concepts- a factor may or may not be a confounder and it may or may not modify the association between the exposure and outcome.
b. Both pooled estimates and standardized estimates adjust for confounding to the degree of stratification by that factor, but pooling is not appropriate in the presence of effect measure modification.
c. In the Framingham Heart study, self-reported smoking at visit 1 is associated with a higher risk of death, especially among younger participants.
d. The standardized estimates for smoking status at visit 1 and the risk of death are adjusted for differences in the age distribution between smokers and non-smokers. These standardized estimates and reflect the association between smoking and death in a population with the age distribution of the group to which we standardize.