Estimating Lifetime Risk of Developing High Serum Total Cholesterol: Adjustment for Baseline Prevalence and Single-Occasion Measurements

Michael J. Pencina¹, Ralph B. D’Agostino¹, Alexa S. Beiser², Mark R. Cobain³, and Ramachandran S. Vasan⁴

¹ Department of Mathematics and Statistics, Boston University, Boston, MA.
² School of Public Health, Boston University, Boston, MA.
³ Unilever Research, Sharnbrook, Bedfordshire, United Kingdom.
⁴ School of Medicine, Boston University, Framingham, MA.

Received for publication April 13, 2006; accepted for publication July 18, 2006.

The lifetime risk statistic is a powerful tool in epidemiology. It has been successfully applied to estimate and highlight the risks of numerous diseases, including breast cancer, Alzheimer’s disease, stroke, and coronary heart disease and some of its risk factors. Application of this method to health-related conditions that may have an onset early in young adulthood or to measurements that can fluctuate over time introduces problems of under- or over-estimation of risk. To correctly quantify the long-term risk of developing high serum total cholesterol (≥240 mg/dl or use of lipid-lowering medication), the authors propose a key modification of the lifetime risk statistic: adjustment for baseline prevalence. It accounts for the fact that many people already have the condition at a young age (an age often chosen as baseline). The authors derive point estimators and confidence intervals and supply a SAS macro (SAS Institute, Inc., Cary, North Carolina). For assessment of the risk inflation due to single-occasion measurement, the authors suggest two diagnostic tools, one requiring the condition to be present on two consecutive occasions and the other taking into account intrasubject variability. As an illustration, the authors calculate risk estimates for US Caucasians based on hypercholesterolemia incidence (1971–early 2001) from the Framingham Heart Study and prevalence data from the 1999–2000 National Health and Nutrition Examination Survey.

cholesterol; disease-free survival; hypercholesterolemia; incidence; prevalence; risk adjustment; risk assessment; survival rate

Cumulative incidence is a common way of quantifying the risk of developing an adverse health condition. Presented as a long-term or lifetime risk statistic, cumulative incidence offers important additional information that is not captured by cross-sectional prevalence estimates. For example, the burden posed by breast cancer or coronary heart disease might seem relatively low if prevalence estimates are considered, but the true public health burden of these conditions is better understood when the lifetime risk statistics are taken into account (1, 2). Gaynor et al. (1) modified the traditional survival analysis techniques used to estimate cumulative incidence (for example, the Kaplan-Meier estimator (3)) by adjusting for the competing risk of death (or other competing causes). The practical application of such an approach was facilitated by the development of the practical incidence estimators (PIE) macro by Beiser et al. (4). This macro has been used successfully to estimate long-term and lifetime risks of coronary

Abbreviations: NHANES, National Health and Nutrition Examination Survey; PIE, practical incidence estimators; PIPE, practical incidence prevalence-adjusted estimators.

Correspondence to Dr. Michael Pencina, Department of Mathematics and Statistics, Boston University, 111 Cummington Street, Boston, MA 02215 (e-mail: mpencina@bu.edu).
In the present article, we focus on quantifying the long-term and lifetime risks of developing a high serum total cholesterol level (or hypercholesterolemia, defined as serum total cholesterol ≥240 mg/dl or use of lipid-lowering medication) in 30-, 40-, and 50-year-old men and women enrolled in the offspring cohort of the Framingham Heart Study (9). Using National Health and Nutrition Examination Survey (NHANES) 1999–2000 data (10), we note that the prevalence of high serum total cholesterol in the age groups of interest ranges from 14 percent to 38 percent (see table 1). Such a high prevalence of hypercholesterolemia suggests that the life-span lifetime or long-term risk would be underestimated if the baseline prevalence were ignored.

Similarly to obesity, hypercholesterolemia is defined with a threshold that a continuous variable and can be considered reversible. It is not uncommon that people are classified as experiencing an event on the basis of a single measurement taken on a single occasion (such as a single Framingham Heart Study examination), which might unnecessarily inflate the risk estimates. Thus, it is important to assess the amount of risk inflation introduced by single-occasion measurement.

To address the challenges described above, we present an extension of the method used in the construction of the PIE macro that introduces an additional adjustment for the baseline prevalence of the condition of interest and derives confidence intervals for the estimated rates. We apply this new method to estimate the life-span lifetime and long-term risks of hypercholesterolemia in the United States based on Framingham Heart Study incidence and NHANES 1999–2000 prevalence. Moreover, we propose two methods for assessing the amount of risk inflation due to single-occasion

TABLE 1. Characteristics of participants in study samples used to estimate long-term and lifetime risks of developing a high* serum total cholesterol level

<table>
<thead>
<tr>
<th>Component of analysis</th>
<th>Age 30 years</th>
<th>Age 40 years</th>
<th>Age 50 years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample size (no.)</td>
<td>Prevalence of high STC (%)</td>
<td>Sample size (no.)</td>
</tr>
<tr>
<td>Prevalence††</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>175</td>
<td>14.2</td>
<td>126</td>
</tr>
<tr>
<td>Men</td>
<td>137</td>
<td>14.2</td>
<td>118</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P-Y of follow-up</td>
</tr>
<tr>
<td>Incidence§§</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>760</td>
<td>14,560</td>
<td>1,258</td>
</tr>
<tr>
<td>Men</td>
<td>621</td>
<td>10,777</td>
<td>998</td>
</tr>
</tbody>
</table>

* Defined as serum total cholesterol ≥240 mg/dl or use of lipid-lowering medication.
† STC, serum total cholesterol; P-Y, person-years.
§§ Based on data from the Framingham Heart Study (Framingham, Massachusetts, 1971–early 2001) (9).
measurement. The first approach estimates the risk of developing sustained high serum total cholesterol, which requires the condition to be present at two consecutive examinations. The second approach takes into account the intraindividual variability in serum total cholesterol and factors it into the definition of hypercholesterolemia. We apply both approaches to assess the risk inflation in the Framingham example.

MATERIALS AND METHODS

Lifetime risk models

Cumulative incidence adjusted for the competing risk of death. For easy reference, we adopt notation paralleling that used by Beiser et al. (4). Let \( A_1 < A_2 < \ldots < A_j < \ldots < A_J \), \( J \leq N \), represent the ordered event ages among \( N \) subjects. Denote by \( r_{A_j} \) the number of persons at risk at age \( A_j \) (their failure or censoring age is greater than or equal to \( A_j \)), by \( e_{A_j} \) the number of people who develop the event of interest at age \( A_j \), and by \( c_{A_j} \), the number of people who develop the event of interest or die at age \( A_j \), out of \( r_{A_j} \) people at risk. The hazard of failing from the event of interest at age \( A_j \) is \( h_{A_j} = e_{A_j}/r_{A_j} \), and the hazard of failing from either the event of interest or death can be calculated as \( c_{A_j}/r_{A_j} \). The estimated survival probability (event-free and alive) is (see the paper by Beiser et al. (4))

\[
\hat{U}_{A_k} = 1 - \sum_{j=1}^{k} \frac{c_{A_j}}{A_j} \times \hat{U}_{A_{j-1}},
\]

or equivalently (induction with \( \hat{U}_{A_0} = 1 \))

\[
\hat{U}_{A_k} = \prod_{j=1}^{k} \left( 1 - \frac{c_{A_j}}{A_j} \right).
\]

The cumulative incidence adjusted for the competing risk of death can be expressed as the sum of the unconditional probabilities of failure at age \( A_j \), \( \hat{F}_{A_k} = h_{A_k} \hat{U}_{A_{k-1}} \):

\[
\hat{F}_{A_k} = \sum_{j=1}^{k} \hat{F}_{A_j} = \sum_{j=1}^{k} h_{A_j} \hat{U}_{A_{j-1}}.
\]

The standard error of the adjusted cumulative incidence \( \text{SE}(\hat{F}_{A_k}) \) can be estimated using a Taylor series linear expansion, described elsewhere (1, 4, 11, 12).

Prevalence-adjusted adjusted cumulative incidence. The procedure described above works well if the event of interest does not occur before a certain age or if the sample is restricted to persons who are condition-free at a certain age. Then we can set \( \hat{U}_{A_0} = 1 \) in equation 1 as the starting point. This is a valid assumption for numerous conditions, including dementia, Alzheimer’s disease, and type II diabetes. However, it is not uncommon to encounter situations in which events of interest occur in young adulthood, such as in the case of hypercholesterolemia. If our objective is to estimate the long-term risk of developing hypercholesterolemia starting at baseline ages of 30, 40, and 50 years (see the paper by Pencina et al. (13)), we must consider the fraction of persons who may have already developed the condition at baseline. From the perspective of estimating the public health burden posed by hypercholesterolemia, it is important to determine the risk of developing a high serum total cholesterol level (the event) while taking into account the fact that the condition may have occurred prior to the baseline age. This will require an adjustment for the baseline prevalence of the condition.

Thus far in the methods outlined above, we were assuming \( \hat{U}_{A_0} = 1 \), where no one experienced an event before time \( A_1 \). Now, this is no longer true. We consider three “initial” ages: \( A_1 \) is the age at which we observe the first event in our sample; \( A_0 \) is the starting baseline age of our sample; and \( A_{-1} \) is an age prior to our baseline before which no events could have occurred. To introduce the adjustment for baseline prevalence, we can assume that all prior events took place right at \( A_0 \). Then we can set \( \hat{U}_{A_{-1}} = 1 \) and \( c_{A_0}/r_{A_0} = \delta \), where \( \delta \) is the estimated baseline prevalence of the condition of interest. We also have \( h_{A_0} = \delta \). Our estimation starts one age unit earlier than before—at age \( A_0 \). The event-free and alive survival of equation 2 now takes the form

\[
\tilde{V}_{A_k} = \prod_{j=0}^{k} \left( 1 - \frac{c_{A_j}}{r_{A_j}} \right) = (1 - \delta) \prod_{j=1}^{k} \left( 1 - \frac{c_{A_j}}{r_{A_j}} \right)
\]

for \( k \geq 0 \) and \( \tilde{V}_{A_{-1}} = \tilde{U}_{A_{-1}} = 1 \).

The adjusted cumulative incidence additionally adjusted for baseline prevalence can be expressed as

\[
\hat{G}_{A_k} = \sum_{j=0}^{k} h_{A_j} \tilde{V}_{A_{j-1}} = h_{A_0} \hat{V}_{A_{-1}} + \sum_{j=1}^{k} h_{A_j} (1 - \hat{\delta}) \hat{U}_{A_{j-1}}
\]

\[
= \hat{\delta} + (1 - \hat{\delta}) \sum_{j=1}^{k} h_{A_j} \hat{U}_{A_{j-1}} = \hat{\delta} + (1 - \hat{\delta}) \hat{F}_{A_k}.
\]

Since \( \hat{F}_{A_k} \) can be interpreted as the cumulative incidence conditional on event-free survival until baseline, equation 5 can be read as the sum of the conditional probabilities of having an event prior to or at age \( A_k \) given an event at baseline (equal to 1) times the probability of having an event at baseline \( (\hat{\delta}) \) plus the conditional probability of an event prior to or at \( A_k \) given no event at baseline \( (\hat{F}_{A_k}) \) times the probability of no event at baseline \( (1 - \hat{\delta}) \). Thus, \( \hat{G}_{A_k} \) represents the cumulative incidence up to age \( A_k \) without conditioning on event-free status at baseline.

It is preferable that baseline prevalence, \( \hat{\delta} \) be estimated from a sample that is independent of the one used for calculating cumulative incidence using contemporary data. Since the cumulative incidence is estimated over a long period of time, baseline prevalence from the beginning of a study may be different from that calculated using current data. Use of contemporary estimates facilitates the applicability of the results. The variance of \( \hat{G}_{A_k} \) can be estimated using the delta method (14, 15). We have

\[
\var(\hat{G}_{A_k}) = (1 - \hat{F}_{A_k} - 1 - \hat{\delta}) \times \left( \begin{array}{c} \var(\hat{\delta}) \\ \cov(\hat{\delta}, \hat{F}_{A_k}) \\ \cov(\hat{\delta}, \hat{F}_{A_k}) \\ \var(\hat{F}_{A_k}) \\ \var(\hat{F}_{A_k}) \end{array} \right) \times \left( \begin{array}{c} 1 - \hat{F}_{A_k} \\ 1 - \hat{\delta} \end{array} \right).
\]
Since \( \hat{\delta} \) and \( \hat{F}_{Ak}^* \) are based on two different independent samples, we have \( \text{cov}(\hat{\delta}, \hat{F}_{Ak}^*) = 0 \), and equation 6 reduces to

\[
\text{var}(\hat{G}_{Ak}^*) = (1 - \hat{F}_{Ak}^*)^2 \text{var}(\hat{\delta}) + (1 - \hat{\delta})^2 \text{var}(\hat{F}_{Ak}^*).
\]  

The 95 percent confidence interval for the baseline prevalence-adjusted cumulative lifetime risk is given by

\[
\hat{G}_{Ak}^* \pm z_{1-\alpha/2} 2\text{SE}(\hat{G}_{Ak}^*) = \hat{\delta} + (1 - \hat{\delta}) \hat{F}_{Ak}^* \pm z_{1-\alpha/2} \times \left( (1 - \hat{F}_{Ak}^*)^2 \text{var}(\hat{\delta}) + (1 - \hat{\delta})^2 \text{var}(\hat{F}_{Ak}^*) \right)^{1/2},
\]  

where \( z_{1-\alpha/2} \) is the desired percentile of the standard normal distribution.

### Application: Life-span Lifetime Risk of Hypercholesterolemia

The Framingham Heart Study started in Framingham, Massachusetts, in 1948 (16, 17) with the enrollment of the “original” cohort of 5,209 persons. In 1971, the offspring of the original cohort and their spouses were enrolled in the Framingham Offspring Study. A total of 5,124 participants attended the first offspring examination, returned for a second examination every 3–4 years since then (9). All participants gave written informed consent, and the study protocol was approved by the institutional review board of the Boston Medical Center (Boston, Massachusetts). At each of the seven offspring examinations, blood samples were obtained from participants who had fasted for 12–24 hours. Plasma total cholesterol was measured using automated enzymatic assays (18), and plasma values were approximated to serum values by multiplying by 1.03 (19). Participants who had a serum total cholesterol level of 240 mg/dl or more or were using lipid-lowering medication were classified as having high serum total cholesterol (hypercholesterolemia), consistent with the guidelines of the National Cholesterol Education Program (20). The present analysis included participants who were free of hypercholesterolemia at their first eligible examination and free of myocardial infarction during the course of follow-up, which extended from 1971 to early 2001 (Framingham examinations 1–7).

Our objective was to determine the life-span lifetime and long-term risks of high serum total cholesterol for 30-, 40-, and 50-year-olds who were free of myocardial infarction. The traditional application of the PIE macro allows for the addition of persons older than the starting age to the original pool of subjects used for estimation. However, here we limited the original pool to subjects not older than 4 years above the index age (resulting in age groups of 30–34, 40–44, and 50–54 years). This was necessary to limit the possibility of including persons free of hypercholesterolemia at their first visit but with an unknown status before then. For example, a 40-year-old seen for the first time who was free of the condition would not be included in the “at-risk” pool of 30-year-olds, since we do not know that person’s status at age 30 years. This was not an issue in the traditional application of the PIE macro to irreversible conditions (e.g., dementia or Alzheimer’s disease): A 40-year-old entering the study free of Alzheimer’s disease would have been free of the disease at age 30 as well.

We used NHANES 1999–2000 data (10) to calculate the baseline prevalence of high serum total cholesterol in Caucasian (for consistency with the predominantly Caucasian Framingham cohort) women and men aged 30–39, 40–49, and 50–59 years. Here we enlarged the age groups used for the incidence component to increase sample sizes and obtain more stable estimates. The numbers of women and men contributing to the incidence and prevalence portions of our estimation process are presented in table 1. Note that using a single, age-specific point prevalence for hypercholesterolemia introduces a strong birth cohort effect. While it adds to the contemporariness of our findings in this case, great caution should be used when selecting the prevalence sample in other situations. A long-term risk analysis (described above in “Prevalence-adjusted cumulative incidence”) was carried out using the practical incidence prevalence-adjusted estimators (PIPE) macro given in the Appendix.

### Assessment of Risk Inflation

Hypercholesterolemia, unlike dementia or Alzheimer’s disease, and similarly to obesity, is not an “absorbing state.” This means that serum total cholesterol can return to normal levels or fluctuate over time. Studying hypercholesterolemia prospectively is challenged by the possibility of misclassifying people as having events based on single-occasion measurements. This concern is valid in the context of the Framingham Heart Study, where the estimates are based on several single-occasion measurements (a maximum of seven examinations) taken approximately 4 years apart. This can lead to overestimation of the actual risk.

To assess the amount of potential risk inflation, we propose two diagnostic techniques. The first one can be labeled the clustering approach. We define a person as hypercholesterolemic if the condition is present at two consecutive Framingham examinations. Doing so, we no longer use single examinations as bases of event classification but consider two-examination clusters instead (“clustering” readings from two examinations together). Thus, in the Framingham example, the data are analyzed on the basis of event classifications (presence vs. absence of hypercholesterolemia) in the following six two-examination clusters: 1&2, 2&3, 3&4, 4&5, 5&6, and 6&7.

For illustration, consider a person who is 30 years old at examination 1 and has the following serum total cholesterol values (in mg/dl): 210 at examination 1; 242 at examination 2; 235 at examination 3; 230 at examination 4; 242 at examination 5; 248 at examination 6; and 250 at examination 7. As a convention, we take the age of the person at the first of the two examinations in each cluster as his or her age; hence, 30 years will be the person’s baseline age. The event status is “nonevent” in clusters 1&2, 2&3, 3&4, and 4&5 and changes to “event” in cluster 5&6. Note that even though serum total cholesterol exceeded 240 mg/dl at

examination 2, this was not classified as an event, since the levels at two “clustering” examinations were below 240 mg/dl. Additionally, the first cluster with an event is 5&6, even though serum total cholesterol already exceeded 240 mg/dl at examination 5; however, it did not do so at examination 4. Baseline event status is assessed consistently with the event definition of the clustering approach: The person in our example enters at examination cluster 1&2 as not having an event, even though his or her serum total cholesterol level exceeded 240 mg/dl at examination 2.

Note that this approach eliminates a single-examination occurrence of the condition from being classified as an event but does not miss any two consecutive occurrences. After the event status is defined with examination clusters as units, we can apply the PIE or PIPE macro to obtain the risk estimates. The adjustment for baseline prevalence could still be added; however, we believe that it is less of a concern here. The “clustering” approach is meant to serve more as an indicator of the robustness of the estimates obtained using the single examination event definitions rather than as an estimating procedure itself: The two-examination clusters are somewhat arbitrary and do not offer the simple interpretation that is possible with the traditional analysis.

A different approach to assessment of the amount of risk inflation takes into account the intrasubject variability in serum total cholesterol. This variability has been reported to range between 3.9 percent and 10.9 percent per year (21). We amend the definition of hypercholesterolemia with the additional requirement that the total cholesterol level must exceed a preselected cutoff based on the intrasubject variability. Here we define this cutoff as 10 percent above the baseline value. Hence, a person is considered hypercholesterolemic if his or her serum total cholesterol level equals or exceeds the maximum of 240 mg/dl and 1.1 times the baseline level or he/she undergoes a lipid-lowering treatment.

**RESULTS**

**Lifetime risk of hypercholesterolemia**

The results for the long-term and lifetime risks of hypercholesterolemia are summarized in tables 2 and 3, with a graphical example presented in figure 1. Table 2 presents the 10-, 20-, and 30-year risks of developing high serum total cholesterol, adjusted for the competing risk of death, without the adjustment for baseline prevalence (i.e., cumulative incidences adjusted for the competing risk of death for persons free of the condition at baseline). The corresponding rates adjusted for baseline prevalence are displayed in table 3. For a 30-year-old person, these data can be interpreted as the risk of already having or developing hypercholesterolemia over time by ages 40, 50, and 60 years. Note that for 50-year-olds, the 30-year rates presented in tables 2 and 3 approximate their remaining and life-span lifetime risks, respectively, of developing a high serum total cholesterol level.

On the basis of table 2, we estimate that 1–2 of every 10 women who are free of high serum total cholesterol at age 40 or 50 years will develop the condition over the next 10 years. This 10-year risk increases to 3–5 in 10 when we account for baseline prevalence (table 3). The 30-year risk exceeds 5 in 10 for condition-free women and reaches 7 in 10 after accounting for baseline prevalence. The risks are fairly uniform among the three baseline age categories in men who are free of the condition at baseline; risks oscillate around 1 in 10 with the 10-year horizon and exceed 4 in 10 with 30 years of follow-up. Accounting for baseline prevalence, the rates increase to 2–4 in 10 and 5–7 in 10 for the 10- and 30-year periods, respectively.

**Risk inflation**

The results obtained using the two risk inflation assessment methods are presented in tables 4 and 5. Note that the
clustering approach is more conservative, leading to lower risk estimates. To assess the magnitude of risk inflation, we compare the estimates presented in tables 4 and 5 with those shown in table 2. The modified estimates are lower than their single-occasion measurement counterparts. For the clustering approach, the magnitude of this decrease amounts to about one person in 10. It is quite substantial for the 10-year estimates (decreasing the estimates from two people in 10 to one person in 10) but is less pronounced for the 20-year risk estimates—decreasing from four in 10 to three in 10. The estimates that take into account the intrasubject variability decrease by only, at most, five people in 100 across all age and time categories. Overall, we conclude that the longer-term risk of developing hypercholesterolemia remains substantial, even if we require that the condition be maintained at two consecutive examinations or that additional subject-specific thresholds be crossed.

**DISCUSSION**

In this paper, we have presented a modification of the PIE macro which allows for an adjustment to the long-term risk estimates for the baseline prevalence of the condition. Such a modification can be particularly useful in situations where the condition can occur in early adulthood, at or before the age of risk prediction; obesity and hypercholesterolemia in the Framingham Heart Study are good examples. The proposed method calculates the cumulative incidence, adjusted for baseline prevalence and the competing risk of death, as a function of cumulative incidence (adjusted for the competing risk of death) and the baseline prevalence of the condition. The resulting estimates can be interpreted as the probability of having or developing the condition of interest in a given time horizon. Confidence intervals around the adjusted incidence rates are obtained using the delta method. A SAS macro (SAS Institute, Inc., Cary, North Carolina) which incorporates these additions is given in the Appendix.

We have illustrated the application of this method to the estimation of long-term risk of hypercholesterolemia in the Framingham Heart Study. Baseline prevalence was obtained from the NHANES 1999–2000 survey. We conclude that long-term risk estimates obtained without the baseline prevalence adjustment may underestimate the future burden of the condition from a public health perspective. On the other hand, the clinical perspective (estimation of risk for a person who is condition-free at a baseline age) does not require adjustment for baseline prevalence.

Additionally, we have assessed the potential risk inflation due to misclassification caused by ascertainment of event
status based on single-occasion measurements. We have introduced the clustering approach, which requires the condition to be present at two consecutive visits for a person to be classified as having experienced an event, and another diagnostic approach that amends the definition of hypercholesterolemia with an additional threshold based on intrasubject variability. We have observed that even though these modifications lower the risk estimates in the Framingham example, such estimates still remain quite high.

We believe that this article offers useful tools for investigators who want to estimate the long-term and lifetime risks of conditions that are already present at baseline and/or are classified on the basis of a single-occasion measurement. The combination of current baseline prevalence and cumulative incidence over a long period of time offers researchers the opportunity to enhance their understanding of future risk burden beyond that available from prevalence or incidence data alone.

**TABLE 4. Estimated long-term risk of a sustained* high† serum total cholesterol level, conditional on survival without high serum total cholesterol at baseline‡:**

<table>
<thead>
<tr>
<th>Gender and follow-up time (years)</th>
<th>Age 30 years</th>
<th>Age 40 years</th>
<th>Age 50 years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimated risk (%)§</td>
<td>95% CI¶</td>
<td>Estimated risk (%)§</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1.20</td>
<td>0.32, 2.08</td>
<td>6.60</td>
</tr>
<tr>
<td>20</td>
<td>12.00</td>
<td>8.92, 15.08</td>
<td>26.76</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3.50</td>
<td>1.87, 5.14</td>
<td>7.65</td>
</tr>
</tbody>
</table>

* Defined as high serum total cholesterol in two consecutive Framingham examinations.
† Defined as serum total cholesterol ≥240 mg/dl or use of lipid-lowering medication.
‡ Based on data from the Framingham Heart Study (Framingham, Massachusetts, 1971–early 2001) (9). Note that because of the design of this approach, 30-year risk estimates are not available.
§ Adjusted for competing risk of death.
¶ CI, confidence interval.

**TABLE 5. Estimated long-term risk of ever developing a high* serum total cholesterol level after accounting for intrasubject variability†, conditional on survival without high serum total cholesterol at baseline‡:**

<table>
<thead>
<tr>
<th>Gender and follow-up time (years)</th>
<th>Age 30 years</th>
<th>Age 40 years</th>
<th>Age 50 years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimated risk (%)§</td>
<td>95% CI ¶</td>
<td>Estimated risk (%)§</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1.89</td>
<td>0.91, 2.87</td>
<td>7.26</td>
</tr>
<tr>
<td>20</td>
<td>14.56</td>
<td>11.85, 17.28</td>
<td>38.64</td>
</tr>
<tr>
<td>30</td>
<td>44.60</td>
<td>39.23, 49.97</td>
<td>57.18</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>5.62</td>
<td>3.78, 7.45</td>
<td>10.17</td>
</tr>
<tr>
<td>20</td>
<td>25.95</td>
<td>22.20, 29.69</td>
<td>29.66</td>
</tr>
<tr>
<td>30</td>
<td>45.16</td>
<td>39.47, 50.84</td>
<td>44.34</td>
</tr>
</tbody>
</table>

* Defined as serum total cholesterol ≥240 mg/dl or use of lipid-lowering medication.
† Risk of developing a serum total cholesterol level at or above the maximum of 240 mg/dl and 1.1 times the baseline value or use of lipid-lowering medication.
‡ Based on data from the Framingham Heart Study (Framingham, Massachusetts, 1971–early 2001) (9).
§ Adjusted for competing risk of death.
¶ CI, confidence interval.
ACKNOWLEDGMENTS

This work was supported by the National Heart, Lung, and Blood Institute (contract N01-HC-25195).
Conflict of interest: none declared.

REFERENCES


APPENDIX

PIPE Macro

Here we present a simple modification of the practical incidence estimators (PIE) macro which enables the user to account for baseline prevalence (practical incidence prevalence-adjusted estimators (PIPE)). The number of variables specified in the PIE macro call is extended to include baseline prevalence (variable prev) and the sample size on which it is based (np) for each of the two groups considered. It takes the form

%macro PIPE(IDS, minage, maxage, agegrpw, group, level1, level2, agefree, o1, o2, np1, prev1, np2, prev2),

where all of the other inputs are described in the article by Beiser et al. (4). The PIE macro is composed of several submacros. Our adjustment alters only a portion of the %lr subroutine. Its call is extended to include

%lr (in, sendout, prev, np).

Then the following additions are made, starting after the statements within the data set “out” are completed:

data prev; yes=round(&prev*np); zno=&np-yes; run;
proc transpose data=prev out=tranprev; run;
proc freq data=tranprev noprint; tables _name_/binomial; weight coll;
output out=varprev bin; run;
data varprev; set varprev; varprev=e_bin*2; prev=bin_; keep prev varprev; run;

The above statements calculate the variance associated with the prevalence rate, taking into account the corresponding sample size. This variance and the prevalence rate are then added to the \textit{s1} data set and used to compute the baseline-adjusted adjusted cumulative incidence and its standard error (additions are shown in boldface type):

```plaintext
data s1; if _n_ eq 1 then set varprev; set out;
keep years cf&agefree secf&agefree cfstar&agefree secff&agefree
  lcl ucl lclstar uclstar adj Cfstar&agefree adj secff&agefree
  adj_lclstar adj_uclstar;
cf&agefree=cf&agefree*100;
secf&agefree=secf&agefree*100;
 cfstar&agefree=cfstar&agefree*100;
 secff&agefree=secff&agefree*100;
adj Cfstar&agefree=prev*100+(1-prev)*cfstar&agefree;
adj secff&agefree=
sqrt((100-cfstar&agefree)**2*varprev+(1-prev)**2*secff&agefree**2);
years=(age-\text{agefree}+1);
lcl=max(0,cf&agefree-1.96*secf&agefree);
ucl=min(100,cf&agefree+1.96*secf&agefree);
lclstar=max(0,cfstar&agefree-1.96*secff&agefree);
uclstar=min(100,cfstar&agefree+1.96*secff&agefree);
adj lclstar=max(0,adj Cfstar&agefree-1.96*adj secff&agefree);
adj uclstar=min(100,adj Cfstar&agefree+1.96*adj secff&agefree);
run;
data &sendout; set s1;
keep years cf&agefree secf&agefree
  cfstar&agefree secff&agefree
  adj Cfstar&agefree adj secff&agefree;
run;
data s2; set s1;
  if ((years/agegrp eq floor(years/agegrp)));
  format years yf.;
run;
title2 'Unadjusted Cumulative Incidence';
title3 'With 95\% Confidence Limits';
proc print;id years;var cf&agefree lcl ucl;run;
title2 'Cumulative Incidence, Adjusted for Competing Risk of Death';
title3 'With 95\% Confidence Limits';
proc print;id years;var cfstar&agefree lclstar uclstar;run;
title2 'Cumulative Incidence';
title3 'Adjusted for Competing Risk of Death and Baseline Prevalence';
title4 'With 95\% Confidence Limits';
proc print;id years;var adj Cfstar&agefree adj lclstar adj uclstar;run;
title;title2;title3;title4;
```

Note that when baseline prevalence is zero (\textit{&prev}=0), the baseline-adjusted adjusted cumulative incidence reduces to adjusted cumulative incidence.

The list of subroutines composing the outer shell of the \textit{PIPE} macro is identical to that of the original \textit{PIE} macro, with \textit{%lr} being invoked twice with an extended list of parameters:

\texttt{%lr(SDS1, &o1, &prev1, &np1) and %lr(SDS2, &o2, &prev2, &np2)}.

The macro call for an analysis of 30-year-olds takes the form

\texttt{%PIPE(IDS=over30, minage=30, maxage=70, agegrp=5, group=gender, level1=1, level2=2, agefree=30, o1=f30, o2=m30, prev1=0.142, np1=175, prev2=0.142, np2=137)},

where the first 10 entries are described in the article by Beiser et al. (4) and the last four are taken from table 1.