PRACTICE OF EPIDEMIOLOGY

A Methodological Issue in the Analysis of Second-Primary Cancer Incidence in Long-Term Survivors of Childhood Cancers

Yutaka Yasui1, Yan Liu1, Joseph P. Neglia2, Debra L. Friedman1,3, Smita Bhatia4, Anna T. Meadows5, Lisa R. Diller6, Ann C. Mertens2, John Whitton1, and Leslie L. Robison2

1 Cancer Research Program, Fred Hutchinson Cancer Research Center, Seattle, WA.
2 Department of Pediatrics, University of Minnesota School of Medicine, Minneapolis, MN.
3 Department of Pediatrics, University of Washington School of Medicine, Seattle, WA.
4 Division of Pediatric Oncology and Bone Marrow Transplantation, City of Hope National Medical Center, Duarte, CA.
5 Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA.
6 Department of Pediatrics, Harvard Medical School, Boston, MA.

Received for publication March 4, 2003; accepted for publication June 18, 2003.

Survival of childhood cancer patients has increased remarkably in the last several decades due to therapeutic improvements. Associated with this progress is the emerging need to accurately assess/minimize late effects of cancer therapy in long-term survivors. This paper considers a methodological issue in assessing the risk of second-primary malignant neoplasms, a major late effect of concern, using second-primary female breast cancer as an example. In the assessment of second-primary malignant neoplasm risk, attained age is a critical factor that must be taken into account. Even with follow-up of decades, childhood-cancer survivors are still at relatively young ages for developing adult-onset diseases. Attained ages at follow-up, however, modify cancer risk considerably; for example, in the general population, women aged 40 years have about fivefold increased breast cancer risk compared with women aged 30 years. A failure to account for the natural age-associated increase of risk could alter, or even reverse, analytical conclusions. This problem was studied empirically by both descriptive and regression analyses of two major studies of long-term childhood-cancer survivors, the Childhood Cancer Survivor Study (1975–1999) and the Late Effects Study Group (1955–1994). These showed appreciable differences in the analytical results by not accounting for the natural age-associated increase of risk, illustrating a significant impact of this methodological issue on study conclusions.

As a result of remarkable improvements in the therapies for childhood cancers in the last several decades, the number of long-term childhood cancer survivors is growing (1). Second-primary malignant neoplasms (SMNs) are a major late effect of concern for long-term survivors of childhood cancers, their families, and physicians. Understanding SMN risk based on patients’ primary cancer types, therapeutic exposures, and other characteristics is critical in order for clinicians to provide appropriate follow-up care for survivors and to modify treatment protocols to avoid this serious late effect. The assessment of SMN risk among survivors is, therefore, an emerging issue of major clinical importance.

Abbreviations: CCSS, Childhood Cancer Survivor Study; CI, confidence interval; LESG, Late Effects Study Group; SEER, Surveillance, Epidemiology, and End Results; SMN, second-primary malignant neoplasm.
One of the primary research goals in large long-term follow-up studies of childhood-cancer survivors, such as the Childhood Cancer Survivor Study (CCSS) (2) and the Late Effects Study Group (LESG) (3), is to assess the risk of developing an SMN after childhood cancer. In the analysis of SMN data from these studies, an important methodological issue has emerged, relating to age at exposure and time since exposure, that has not previously been addressed. To illustrate, we will use as an example the posited association between risk of second-primary breast cancer among female survivors and age at diagnosis/treatment for childhood cancer.

This association may be hypothesized based on the concept that exposures to therapeutic doses of radiation to the chest during puberty, when cells in the breast tissue are in a high mitotic state, could increase relevant mutation rates, thereby increasing subsequent breast cancer risk. Using a popular approach to rate ratio evaluations in follow-up studies, namely, Cox regression, a report from the LESG showed a significant increase in breast SMN risk in the group diagnosed/treated between 10 and 16 years of age compared with the group diagnosed/treated between 0 and 9 years of age, after adjusting for dose of radiotherapy. This finding in the large high-profile study, which was consistent with the above hypothesis (3), has provided a basis for age-gender-specific clinical recommendations regarding radiation therapy for Hodgkin’s disease patients (4). Given the importance of the clinical implications from these study findings, the goal of this paper is to describe a methodological issue underlying the inconsistent results from the two large studies of SMN risk in long-term childhood cancer survivors, explaining why the two reports came to different conclusions and discussing appropriate methods that address the methodological issue.

**SPECIAL FEATURES OF LONG-TERM FOLLOW-UP STUDIES OF CHILDHOOD CANCER SURVIVORS**

Long-term follow-up studies of children differ from those of adults with respect to study features that are particularly relevant to assessing risk of adult-onset disease during follow-up. First, even after a follow-up of decades, childhood cancer survivors are still at relatively young ages for developing adult-onset diseases such as breast cancer. In a study with follow-up of 20 years, for example, females diagnosed with childhood cancers at ages 5–20 years would be only 25–40 years of age at the end of follow-up, still relatively young ages for developing breast cancer. Second, when considered in a relative scale (e.g., risk ratio), these attained ages at follow-up modify adult-onset disease risk appreciably. The following example illustrates the importance of these points.

Figure 1 presents breast cancer incidence rates reported in the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute (5). The SEER registry reports cancer diagnoses in geographically defined US populations, covering approximately 10 percent of the total US population. The part of the age-incidence curve particularly relevant to long-term follow-up studies of children is expanded in the right panel. Note that, in the general
population, women aged 40 years, although at substantially lower risk than older women, have about a fivefold increased risk of breast cancer relative to women aged 30 years. Translating this into the setting of a follow-up study of children, consider a group of 5,000 girls diagnosed/treated for childhood cancer at age 5 years and followed for 20 years. We would expect, based solely on the SEER incidence rates shown in figure 1, 3.5 cases of breast cancer in the 20-year follow-up of this group (table 1). Note that the expected number is based on a hypothetical assumption that the group of female childhood cancer survivors has the same incidence rates of breast cancer as the general population. If the girls’ age at childhood cancer diagnosis is 10 years, instead of 5 years, the expected number of breast cancers in the 20-year follow-up is 25.5. If the age at diagnosis is 15 years, the expected number of breast cancers in the 20-year follow-up is 92.0–26.3 times higher than the case where the age at diagnosis of childhood cancer is 5 years.

It is, therefore, critical to consider the natural rise of breast cancer risk with age when assessing the potential association between risk of breast SMN and age at diagnosis/treatment for childhood cancer, or any other factors that may be associated with age at diagnosis/treatment. Specifically, observing a higher rate of breast SMN during a particular length of follow-up in girls treated for a childhood cancer at older ages is not surprising, based on the natural rise of breast cancer risk with age seen in the general population. Such an observation is by no means an indication of an increased elevation of risk among girls treated at older ages.

### DESCRIPTIVE ANALYSIS: INCIDENCE RATES VERSUS STANDARDIZED INCIDENCE RATIOS

To elucidate how important it is to account for this natural rise of breast cancer risk with age when assessing the potential association between risk of breast SMN and age at diagnosis/treatment for childhood cancer, or any other factors that may be associated with age at diagnosis/treatment. Specifically, observing a higher rate of breast SMN during a particular length of follow-up in girls treated for a childhood cancer at older ages is not surprising, based on the natural rise of breast cancer risk with age seen in the general population. Such an observation is by no means an indication of an increased elevation of risk among girls treated at older ages.

### TABLE 1. Expected numbers of breast cancer in a hypothetical cohort of 5,000 female childhood cancer survivors followed for 20 years, by age at treatment for childhood cancers

<table>
<thead>
<tr>
<th>Start age (years)</th>
<th>Follow-up</th>
<th>Expected breast cancer cases among the 5,000 survivors</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>25</td>
<td>3.5</td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>25.5</td>
</tr>
<tr>
<td>15</td>
<td>35</td>
<td>92.0</td>
</tr>
</tbody>
</table>

### FIGURE 2. Incidence rates of breast second-primary malignant neoplasms by age at diagnosis/treatment of childhood cancers: A, Childhood Cancer Survivor Study (1975–1999); B, Late Effects Study Group (1955–1994). The rates among female survivors of Hodgkin’s disease are shown in black bars, while those among female survivors of all childhood cancers are shown in white bars.

substantially with age at diagnosis/treatment of childhood cancer.

To take the natural rise of breast cancer risk with age into account, standardized incidence ratios can be used (6, 7). A standardized incidence ratio of breast SMNs for a cohort of survivors is an estimate of the rate ratio for the cohort relative to the general population, given by the ratio of the observed to expected numbers of breast SMNs, that is, $\frac{\sum_i D_i}{\sum_i E_i}$, with $E_i$ being the expected number of breast SMNs for survivor $i$. A set of age-specific reference rates yields the expected number for each survivor by $E_i = \sum_j R_j \times T_{ij}$, where $R_j$ and $T_j$ are the age-specific incidence rate in the general population and the person-years at risk for survivor $i$, respectively, in age stratum $j$ (see Yasui and Whitton (8) for an improved standardized incidence ratio calculation in the setting of long-term follow-up of children). The standardized incidence ratios of breast SMNs in the CCSS are shown in figure 3, part A, and in the LESG, in figure 3, part B. Neither
long-term survivors of childhood cancers, regression analyses of clinical and demographic factors as potential modifiers of SMN risk have often failed to account for this phenomenon. However, the same problem seen in the descriptive analysis above (i.e., rates vs. standardized incidence ratios) applies to regression analyses, although the issue is not as transparent as in the case of descriptive analyses.

The standard approach to the regression analysis of SMN risk in long-term childhood cancer survivors has been Cox regression, taking time since study entry as the time scale of the model. This approach models modifications of the hazard rate \( h(t) \) of breast SMNs at time \( t \) since study entry by clinical and demographic characteristics \( X(t) \) that are possibly changeable over time: \( h(t) = h_0(t) \exp\{X(t)\beta\} \), where \( h_0 \) is an unspecified baseline hazard function, and \( \beta \) is a vector of log rate ratio parameters corresponding to the covariate vector \( X(t) \). The earlier report from the LESG (3) used this standard approach and found the significant increase in breast SMN risk among female survivors of Hodgkin’s disease diagnosed/treated at ages 10–16 years, compared with those diagnosed/treated at ages 0–9 years, adjusting for radiotherapy dosage. In fact, in most uses of the Cox regression method, it is the time since study entry that is taken as the time scale of the model (9, p. 75). The great majority of SMN risk analyses in long-term follow-up studies of childhood cancer survivors indeed used this approach.

The standard approach, however, does not take into account the natural rise of breast cancer risk with age. One may consider including age as a time-dependent covariate to account for its effects: Age has to be treated as a time-dependent covariate because it changes during follow-up. A critical problem in including age as a covariate in our example is the need to specify the functional form of age effects on the hazard rate, which appear to be strong based on figure 1. The specification, regardless of the particular functional form being chosen (e.g., log linear, step function, etc.), makes an additional assumption in the analysis, and its appropriateness for the specific cohort of interest is difficult to confirm.

A tenable approach to this regression problem is to use age, instead of time since study entry, as the time scale of the Cox model (see Andersen et al. (10) and Therneau and Grambsch (9) for discussions and examples of time scales as an alternative to time since study entry). Given the sharp increase of breast cancer incidence rates in the relevant age range seen in figure 1 (e.g., an approximately fivefold increase between ages 30 and 40 years), it seems appropriate to use age as the time scale of the Cox model and to introduce time since study entry as a time-dependent covariate as necessary. This approach models the hazard rate \( h(t) \) of breast SMNs at age \( t \) with clinical and demographic characteristics \( X(t) \): \( h(t) = h_0(t) \exp\{X(t)\beta\} \), where \( h_0 \) is an unspecified baseline hazard function. To fit this form of Cox models, the data set must be formulated into the so-called “counting process form” where each survivor’s at-risk period during follow-up is represented by a set of age intervals each of which spans between a starting age and an ending age at risk, and covariate values are constant in each interval.
natural rise of breast cancer risk with age. The standard Cox regression approach or its modification with age as the time scale of the regression data. This formulation has a number of practical advantages including capabilities for handling left-truncation (i.e., entry to the study at age \( t > 0 \)), time scales alternative to time since study entry, and time-dependent covariates (see Therneau and Grambsch (9, chapter 3.7) for details). Commonly used statistical software, such as STATA (Stata Corporation, College Station, Texas), SAS (SAS Institute, Inc., Cary, North Carolina), and S-Plus (Insightful Corporation, Seattle, Washington), can handle the counting process form of survival data.

Another workable approach is to model modifications of standardized incidence ratios (SIR) with covariates using Poisson regression models (6, 7): 
\[
\log \text{SIR}(t) = X(t) \beta
\]
This approach is quite different from the aforementioned standard Cox regression approach or its modification with age as the time scale: It uses external reference age-specific rates, for example, the SEER incidence rates, to account for the natural rise of breast cancer risk with age. The standard Cox regression approach and its modification, on the other hand, control for age effects internally within the study data set without using external reference rates. Since the calculation of standardized incidence ratios itself takes into account the natural rise of breast cancer risk with age, age effects are controlled for without being introduced in the regression model as a covariate. This advantage comes from making the assumptions that all covariates modify the external reference age-specific rates multiplicatively and that the multiplicative modifications are uniform over age (i.e., no effect modification of covariate effects by age), the necessary assumption in comparing standardized incidence ratios (7, chapter 15.8).

To illustrate the regression analysis of SMN risk in real data using the different approaches described above, we estimated the breast SMN rate ratio parameter for age at diagnosis/treatment of Hodgkin’s disease at 10–16 years versus 0–9 years using the LSEG data. The standard Cox regression approach taking time since study entry as the time scale gave a rate ratio estimate of 4.8 (95 percent confidence interval (CI): 1.1, 20.4) (table 2). Using age as the time scale of the Cox model reduced the rate ratio estimate to 1.9 (95 percent CI: 0.4, 7.9) and modified the corresponding \( p \) value from 0.03 to 0.40. Poisson regression of standardized incidence ratios produced a similar rate ratio estimate to the Cox regression with age as the time scale: 1.7 (95 percent CI: 0.4, 7.1) with a \( p \) value of 0.48. The two approaches that took age effects into account yielded similar results: a rate ratio estimate of less than two with a wide confidence interval indicating little statistical evidence for the association between age at diagnosis/treatment of Hodgkin’s disease and subsequent breast SMN risk.

### DISCUSSION

In the analysis of SMN risk among long-term survivors of childhood cancers, age at follow-up is a critical factor whose effects must be taken into account. This is not an evident matter since the risk for developing a specific type of malignancy of interest (e.g., breast cancer) may be quite low in the general population for the entire age range that survivors experience during follow-up. The key point is, however, that the SMN risk in the general population may be sharply increasing over the relevant age range in spite of its overall low level. It is this sharp increase of risk in the general population that necessitates the incorporation of follow-up age effects in the analysis. As illustrated in our example, a failure to consider the effects of follow-up age could alter significantly, or even reverse, conclusions of the analysis. Although we have focused on SMN risk in childhood cancer survivors in this paper, the same methodological issue applies to the risk of any adult-onset disease in the long-term follow-up of childhood disease patients.

Alternatively to the standard Cox regression of taking time since study entry as the time scale, we considered two approaches, Cox regression taking age as the time scale and Poisson regression of standardized incidence ratios. The former controls for follow-up age effects internally, while the latter uses a set of external age-specific rates. There are limitations in each of these approaches. Use of the Poisson standardized incidence ratio regression approach is limited to cases where appropriate external age-specific rates are available. An important assumption/limitation in the Poisson

### TABLE 2. Inference on rate ratios of breast second-primary malignant neoplasm by age at diagnosis/treatment of Hodgkin’s disease using Cox regression with two different time scales and Poisson regression of standardized incidence ratios, Late Effects Study Group, 1955–1994

<table>
<thead>
<tr>
<th></th>
<th>Rate ratio for age at diagnosis (10–16 vs. 0–9 years)</th>
<th>95% CI†</th>
<th>( p ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cox regression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time scale = time since study entry‡</td>
<td>4.8</td>
<td>1.1, 20.4</td>
<td>0.03</td>
</tr>
<tr>
<td>Time scale = age</td>
<td>1.9</td>
<td>0.4, 7.9</td>
<td>0.40</td>
</tr>
<tr>
<td>Poisson regression of SIRs‡</td>
<td>1.7</td>
<td>0.4, 7.1</td>
<td>0.48</td>
</tr>
</tbody>
</table>

† CI, confidence interval; SIRs, standardized incidence ratios.
‡ The common usage of Cox regression method.
standardized incidence ratio regression approach is that age should not be an effect modifier of any covariate effects in the model. With these limitations, however, the Poisson standardized incidence ratio regression can be applied for comparing subgroups whose age ranges during follow-up do not have a sufficient overlap. Suppose all survivors had the same length of follow-up, for example, 20 years. Then, subgroups defined by age at diagnosis/treatment of childhood cancers, for example, 10–16 years versus 0–9 years, cannot be compared using Cox regression with age as the time scale, because the age range of 30–36 years is experienced only by the older subgroup and not by the younger subgroup, making it impossible to internally control for follow-up age effects. This is a limitation in the Cox regression with age as the time scale. If the follow-up length of patients is relatively short compared with the age range of patients studied, this is an important issue to consider. The Poisson standardized incidence ratio regression, on the other hand, can be used in such situations under its assumptions/limitations and use of external age-specific rates.

Findings from epidemiologic analyses such as those discussed here are highly relevant to follow-up care of childhood cancer survivors and therapeutic improvements aimed at minimizing serious late effects. Although we kept our analytical illustration simple for clarity of the main point, excluding factors such as specific exposures associated with cancer therapy from the analysis, the analysis results of the LESG data using Poisson regression of standardized incidence ratios or Cox regression with age as time scale (table 2) are consistent with those of the CCSS data: Statistical evidence is weak for the hypothesized association between the risk of second-primary breast cancer among female survivors and age at diagnosis/treatment for childhood cancer. In view of the importance of clinical implications from these study findings, it is our hope that the methodological issue discussed here will be widely recognized and appropriately handled by researchers who are engaged in the evaluation of adult-onset disease risks in long-term follow-up studies of children.

ACKNOWLEDGMENTS
This work was supported by grant U24-CA-55727 of the National Institutes of Health.

REFERENCES