Time Trends of Non-Hodgkin’s Lymphoma: Are They Real? What Do They Mean? 1

Theodore R. Holford, 2 Tongzhang Zheng, Susan T. Mayne, and Lisa A. McKay

Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, Connecticut 06510

Abstract

Factors that need to be considered in the analysis of time trends in disease incidence are age, year of diagnosis, and birth cohort. When these are included in a log-linear model, a nonidentifiability problem arises from the linear dependence among these three time factors so that only specified functions of the parameters can be unambiguously determined. One of these invariant functions is the drift or the sum of the period and cohort trend. Non-Hodgkin’s lymphoma incidence rates from Connecticut for the period 1935-1989 were analyzed for males and females. In addition to an age effect, both period and cohort significantly improved the fit of the model. The estimated drift shows that there has been a 10.3% increase in risk every 5 years since 1965 for females and 9.2% for males. It is unlikely that a trend of this magnitude can be attributed entirely to data artifact.

Introduction

The analysis of time trends for disease incidence is an important first step in understanding the etiology of a disease. While such analyses have proven useful, there are also underlying limitations to their interpretation so that some of the knowledge we seek to gain from an analysis of trends in rates will inevitably elude us. Three fundamental time effects that should be considered when investigating trends in disease rates are (a) age at diagnosis, (b) period or year of diagnosis, and (c) cohort or the year in which an individual was born (1).

There are many reasons for studying time trends in disease incidence, aside from the information they provide for assessing the health care needs of the population. Time trends have sometimes furnished valuable clues to disease etiology, leading to more definitive studies of the causes of disease. In addition, trends in population-based rates are the ultimate result of population changes in risk factor exposure so that understanding these interrelationships in a specified population can assist in the development of strategies for disease control.

The fact that risk for many cancers is associated with the aging process is well recognized; therefore, it almost goes without saying that age should be part of any descriptive model of cancer trends.

Period, or year of diagnosis, on the other hand, can be affected by any factor that increases the risk of cancer diagnosis for all age groups. This might be due to exposure of all ages to an etiological agent, such as the exposure that one might expect to come from something in the air or water. However, it might also be induced by artifactual changes in the methods of diagnosis. For example, this may be the result of changes in medical technology and/or the extent to which more sensitive methods of diagnosis are used in clinical practice. In addition, one cannot ignore the possibility that there are occasionally fads in selecting a particular diagnosis. These artifactual changes do not represent real changes in the health of the public, but they are only changes in the way disease is classified.

Exposure to a causal factor at birth might be manifested as a cohort effect. However, these are really more general than factors related to birth in that they represent a generation factor; therefore, if the risk of disease varies by generation, then we might expect to see a cohort effect. Changes in risk factor exposure that only affect a particular age group induce an increased risk that becomes associated with the generation that is at the appropriate age when exposure changes. For instance, the strong cohort trends for lung cancer have been attributed to cigarette smoking, because the habit tends to be started in a fairly narrow age range, the late teens to early 20s, and it is the leading cause of the disease (2, 3). While data artifact caused by change in diagnostic patterns would seem to have the greatest influence on the period effects, we cannot rule out the possibility that cohort could be affected by data artifact as well. Some generations might be less tolerant of uncomfortable medical procedures required for a definitive diagnosis, which could, in turn, induce an artifactual cohort trend. In addition, subgroups such as the elderly may be utilizing a component of the health care system that is not using the latest diagnostic procedures so that they have a different probability of receiving a particular diagnosis.

Materials and Methods

For many years, epidemiologists have recognized some of the important etiological implications that may arise from these three time factors. However, it was not until researchers tried to model these factors in a more formal way that the fundamental inferential limitations were more fully appreciated. This arises from the linear dependence among these factors (4), in that the year of birth can be directly calculated by taking the difference between the period and the age.

Cohort = Period - Age

To understand the associations in a regression analysis, it is generally recommended that linearly dependent variables not be included in the analysis. However, the implications of this dependence are really much deeper than just the problem of fitting a regression model. It really involves any attempt to interpret time trends, including simple graphs of age-specific rates. Therefore, one really needs to grapple with, rather than ignore, this problem.

When one tries to represent the overall trend or slope for a particular time effect, then the linear dependence implies that each slope can only be determined up to an unidentifiable constant. If \( \beta \) is the expected value of the slope for one of the time factors that arises from a particular method of estimation, and \( \beta^* \) is the true slope, then there is an indeterminate constant that affects all three slopes for the time factors. This can be represented as:

\[
\begin{align*}
\beta_a &= \beta^*_a + \nu \\
\beta_p &= \beta^*_p - \nu \\
\beta_c &= \beta^*_c + \nu
\end{align*}
\] (A)

where \( \beta_a \) is the age slope, \( \beta_p \) is the period, \( \beta_c \) is the cohort, and the indeterminate constant is represented by \( \nu \) (5).

We can derive two important facts from the relationship among the slopes shown in Equation A. First, none of the slopes can be determined, and each can, in principle, take any value (6), suggested here by...
Only the slopes are not determined, and these are interrelated by a years. Denominators for the rates were obtained by cumulating the effect is not related to the rate. This analysis is based on all non-Hodgkin’s lymphoma cases re apparent using the additive scale. A similar graph can be con- but the suggested lack of parallelism in these data is just as scale. This graph shows incidence rates plotted on the log scale, factors affecting the age-specific rates were age and period, then the age-specific rates. Fig. 2 shows the age-specific rates plotted against period of diagnosis for females. If the only two lines should be parallel on either the log or arithmetic trend, or zero period slope, and that assumption results in a able because of the non identifiability problem alluded to earlier. One might be tempted to think that the implications of some of these results are avoided by statistical methods that do not use models, such as direct adjusted rates. This is not true, because these adjustments generally ignore the effect of cohort, which is only reasonable when the effect is not related to the rate.

This analysis is based on all non-Hodgkin’s lymphoma cases reported to the Connecticut Tumor Registry between 1935 and 1988. Data were first tabulated, using 5-year age groups (20 through 84 years) and 5-year periods, with the exception that the last period is only 4 years. Denominators for the rates were obtained by cumulating the mid-year population estimates obtained by the Connecticut State Health Department, which provides an estimate of person-years experience.

To estimate the effect of the underlying time factors on non-Hodgkin’s lymphoma incidence, we fitted the log-linear age-period-cohort model, shown in Equation B to the Connecticut incidence data. It was assumed that the number of cases had a Poisson distribution, and maximum likelihood estimates of the parameters were obtained by using the regression package, GLIM (generalized linear interactive modeling) (10).

Results

We first consider the trends for the age-specific rates. Fig. 1A shows the age-specific rates plotted against period of diagnosis for females. If the only two factors affecting the age-specific rates were age and period, then these lines should be parallel on either the log or arithmetic scale. This graph shows incidence rates plotted on the log scale, but the suggested lack of parallelism in these data is just as apparent using the additive scale. A similar graph can be con-
Table 1 Significance tests for goodness of fit and time effects for the age-period-cohort model

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goodness of fit</td>
<td>99</td>
<td>86.80*</td>
<td>90.84*</td>
</tr>
<tr>
<td>Cohort curvature</td>
<td>21</td>
<td>100.70*</td>
<td>68.19*</td>
</tr>
<tr>
<td>Period curvature</td>
<td>9</td>
<td>27.57*</td>
<td>35.46*</td>
</tr>
</tbody>
</table>

* P < 0.001.

To determine the extent to which the observed trends are real, we must specify how much is likely to be due to artifact, or at least we should identify periods when diagnostic practice may have had a substantial influence on the trends. A crude indication that it has been a factor in the past is indicated by the percentage reported to the tumor registry with only death certificate information, shown in Fig. 5. It is apparent that, in the early years of the Connecticut Tumor Registry, a substantial portion of the cases had only death certificate information, but since 1965, that percentage has been very small. Another indication that the quality of the diagnostic information has improved is suggested by the percentage of the diagnoses that are microscopically confirmed, shown in Fig. 6. Again, the data suggest that the quality of the registry information has been quite good since 1965.

If we assume that the incidence data collected since 1965 have not been influenced by diagnostic artifact, then the period-cohort drift since 1965, *i.e.*, the sum of the slopes, will suggest the real increase for non-Hodgkin’s lymphoma in recent years. We estimate the drift by using the period slope for the five intervals since 1965, as well as the slope for the nine most recent cohorts. The estimate of drift is 0.0980 (SE 0.0219) log units/5 years for females, which corresponds to a relative risk of 1.103/5-year period. Similarly, for males the trend is 0.0880 (SE 0.0192) log units/5 years, for a relative risk of 1.092.

Fig. 3. Age, period, and cohort effects for females constrained so that the sum is 0 (βp = 0 for solid line, βp = 0.1 for dashed line, βp = 0.2 for dotted line, and βp = 0.3 for dot-dashed line).

Fig. 4. Age, period, and cohort effects for males constrained so that the sum is 0 (βp = 0 for solid line, βp = 0.1 for dashed line, βp = 0.2 for dotted line, and βp = 0.3 for dot-dashed line).

Results for males are shown in Fig. 4, and the conclusions are very similar to those reached for females.

(continued)
justed tests include the other interactions, and the unadjusted
curved or shape of the trends for females and males. Table 2
cohorts is zero. While we cannot obtain an estimate of the slope
effects. Only interactions with period appear to be marginally
for each group by assuming that the slope for the last nine
females are compared in Fig. 7. We obtained a unique trend line
other two.

Table 2 Significance tests for the sex interaction with the curvature effects in the
age-period-cohort models for non-Hodgkin’s lymphoma incidence rates in

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>Adjusted*</th>
<th>P</th>
<th>Unadjusted</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age curvature</td>
<td>11</td>
<td>8.70</td>
<td>0.650</td>
<td>8.47</td>
<td>0.671</td>
</tr>
<tr>
<td>Period curvature</td>
<td>9</td>
<td>17.84</td>
<td>0.037</td>
<td>16.20</td>
<td>0.063</td>
</tr>
<tr>
<td>Cohort curvature</td>
<td>19</td>
<td>18.03</td>
<td>0.520</td>
<td>16.30</td>
<td>0.637</td>
</tr>
</tbody>
</table>

* Each interaction with a curvature effect is considered in the presence of the other two.

The trends in the age, period, and cohort effects for males and
females are compared in Fig. 7. We obtained a unique trend line
for each group by assuming that the slope for the last nine
cohorts is zero. While we cannot obtain an estimate of the slope
from the curve alone, these graphs are useful for comparing the
curvature or shape of the trends for females and males. Table 2
presents significance tests that compare the curvatures between
the sexes by considering sex-curvature interactions. The ad-
justed tests include the other interactions, and the unadjusted
rates only look at the interactions in the presence of main
effects. Only interactions with period appear to be marginally
significant. However, the graph shown in Fig. 7 suggests that
the magnitude of that difference is rather small, especially for
the more recent periods, suggesting that the pattern has been
rather similar for males and females.

Discussion

Our analysis of the effects of age, period, and cohort on the
trends in non-Hodgkin’s lymphoma incidence in Connecticut
suggests that both period and cohort effects are involved. Struc-
tural limitations of the fundamental time parameters in this
model limit the quantities that can be estimated; and we have
presented recent trends in terms of the drift or sum of period
and cohort slopes. This is a parameter that can be uniquely
estimated from the rates (7, 8), and it suggests that there is an
increase in the incidence rates that continues through the recent
time periods.

In order to determine whether the trends are real, one must
keep in mind a variety of artifactual factors that could be related
to an increase in reported incidence without changing the true
rate. Zheng et al. (11) reported that these factors include (a)
completeness of registration, (b) changes in disease classifica-
tion, (c) advances in diagnostic technology, and (d) changes in
related diseases such as AIDS.

It is difficult to document the completeness of registration by
the Connecticut Tumor Registry, but reporting is thought to be
essentially complete since 1965, and since 1971 there has been
a legal reporting requirement for all cancer cases occurring in
the state (11). Indirect evidence that the tumor registry has been
successful in registering cases since 1965 is apparent from the
small proportion reported only at death or without microscopic
confirmation of the diagnosis. Hence, it is unlikely that incom-
plete reporting would have very much impact on these estimates
of time trends.

Changes in the coding system used to classify the disease
resulted in some loss of specificity in cases reported prior to
1961, but the number of cases involved is very low. A more
likely source of an artifactual effect might have resulted from
advances in diagnostic technology that would have resulted in
the detection of more cases over time. Again, it is difficult to
quantify the magnitude of this effect. We have estimated a drift
of about 9–10%/5-year period, which translates into a 40%
increase over 20 years. It seems very unlikely that diagnostic
technology would have had this large an impact on the inci-
dence rates for non-Hodgkin’s lymphoma.

Finally, the ongoing AIDS epidemic in Connecticut would be
expected to have an impact on non-Hodgkin’s lymphoma inci-
dence. However, the trend up to this point appears to have
started much earlier than the AIDS epidemic. While we would
expect to see an effect of AIDS on future trends in non-
Hodgkin’s lymphoma incidence, it probably has not had much
of an effect on the trends reported here.

While data artifact has probably had some impact on the
increasing trends in non-Hodgkin’s lymphoma incidence, it
does not appear to be large enough to account for the magni-
tude of the effect seen in Connecticut. This leads us to the
conclusion that the trends are indeed real. In addition, the
significance of an effect due to birth cohort makes it plausible
that there have been changes in exposure to unidentified risk
factors in the Connecticut population.

References

1. Case, R. A. M. Cohort analysis of mortality rates as an historical or narrative
1308, 1981.
General, Washington, DC: U.S. Department of Health, Education and Wel-
4. Fienberg, S. E., and Mason, W. M. Identification and estimation of age-
period-cohort models in the analysis of discrete archival data. In: K. F.
Schuesler (ed.), Sociological Methodology 1979, pp. 1–67. San Francisco,
5. Holford, T. R. Understanding the effects of age, period, and cohort on inci-
9. Rogers, W. L. Estimable functions of age, period, and cohort effects. Am.
miology of Non-Hodgkin’s Lymphoma in Connecticut. Cancer (Phila.), in
Time Trends of Non-Hodgkin's Lymphoma: Are They Real? What Do They Mean?


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/52/19_Supplement/5443s

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link
http://cancerres.aacrjournals.org/content/52/19_Supplement/5443s.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.