

Frailty Modeling of the Bimodal Age–Incidence of Hodgkin Lymphoma in the Nordic Countries

Tom Grotmol¹, Freddie Bray^{1,2,5}, Harald Holte⁴, Marion Haugen², Lauren Kunz², Steinar Tretli¹, Odd O. Aalen², and Tron A. Moger^{2,3}

Abstract

Background: The bimodality of the age–incidence curve of Hodgkin lymphoma (HL) has been ascribed to the existence of subgroups with distinct etiologies. Frailty models can be usefully applied to age–incidence curves of cancer to aid the understanding of biological phenomena in these instances. The models imply that for a given disease, a minority of individuals are at high risk, compared with the low-risk majority.

Methods: Frailty modeling is applied to interpret HL incidence on the basis of population-based cancer registry data from the five Nordic countries for the period 1993 to 2007. There were a total of 8,045 incident cases and 362,843,875 person-years at risk in the study period.

Results: A bimodal frailty analysis provides a reasonable fit to the age–incidence curves, employing 2 prototype models, which differ by having the sex covariate included in the frailty component (model 1) or in the baseline Weibull hazard (model 2). Model 2 seemed to fit better with our current understanding of HL than model 1 for the male-to-female ratio, number of rate-limiting steps in the carcinogenic process, and proportion of susceptibles; whereas model 1 performed better related to the heterogeneity in HL among elderly males.

Conclusion: The present analysis shows that HL age–incidence data are consistent with a bimodal frailty model, indicating that heterogeneity in cancer susceptibility may give rise to bimodality at the population level, although the individual risk remains simple and monotonically increasing by age.

Impact: Frailty modeling adds to the existing body of knowledge on the heterogeneity in risk of acquiring HL. *Cancer Epidemiol Biomarkers Prev*; 20(7); 1350–7. ©2011 AACR.

Introduction

Lymphomas are grouped into 2 broad but distinct categories, Hodgkin lymphoma (HL) and non-HL that together comprise about 4% of all malignancies in high-resource areas of the world. The age–incidence curve of non-HL commonly exhibit monotonically increasing patterns typical for most epithelial cancers, whereas HL curves tend to exhibit 2 peaks, with the respective ages at peak incidence at about 20 and 75 years. The bimodality property of the HL age–incidence curve was first described in the late 1950s by MacMahon (1), who ascribed the heterogeneity in incidence by age to subgroups each with distinct etiologies (2). The infectious

etiology of HL has been postulated for decades, based initially on clinical and histopathologic presentation, and subsequently on epidemiologic evidence (3, 4).

Although there were indications of Epstein–Barr virus (EBV) infection being involved in the etiology of HL, it was not until 1987 that EBV DNA was detected in the malignant Hodgkin/Reed–Sternberg cells of the tumor (5). EBV infection is clearly implicated in the etiology of a proportion of HL, but it is not established which factors determine the development of EBV-related HL, although immunologic control of the infection is likely to be involved (6). The first peak of the age–incidence curve is dominated by HL among young adults and has been suggested to be caused partly by a primary infection (largely non-EBV related; ref. 7). The first mode also contains a small group of HL among children that is most likely attributed to primary EBV infection. The second peak in older adults is thought to be due to loss of immunologic control of latent infection (mostly EBV; ref. 6).

One possible explanation for the bimodal age–incidence curve is that it reflects risk at the individual level, that is, a typical person has a risk that changes over his or her lifetime with 2 peaks. Although possible, this does not fit with carcinogenesis models which seem to predict monotonically increasing risk (8, 9). Another explanation

Authors' Affiliations: ¹Cancer Registry of Norway, Institute of Population-Based Cancer Research; ²Department of Biostatistics, Institute of Basic Medical Sciences; ³Department of Health Management and Health Economics, University of Oslo; ⁴Department of Oncology, Oslo University Hospital; and ⁵Cancer Information Section, International Agency for Research on Cancer, Lyon, France

Corresponding Author: Tom Grotmol, Cancer Registry of Norway, Institute of Population-Based Cancer Research, PO Box 5313 Majorstuen, NO-0304 Oslo, Norway. Phone: 47-22-45-13-00; Fax: 47-22-45-13-70; E-mail: tom.grotmol@krefregisteret.no

doi: 10.1158/1055-9965.EPI-10-1014

©2011 American Association for Cancer Research.

Table 1. The number of HL cases, age-standardized (World) incidence rates, and person-years in the Nordic countries by sex, for the period (1993–2007)

Country	Males			Females			Total	
	Cases	ASR	Person-years	Cases	ASR	Person-years	Cases	Person-years
Denmark	1,094	2.5	39,514,034	784	1.8	40,437,285	1,878	79,951,319
Finland	1,113	2.7	37,902,359	878	2.2	39,750,803	1,991	77,653,162
Iceland	59	2.6	2,126,533	42	1.7	2,108,671	101	4,235,204
Norway	849	2.3	33,376,330	603	1.7	34,002,616	1,452	67,378,946
Sweden	1,441	1.9	66,121,311	1,182	1.6	67,503,933	2,623	133,625,244
Total	4,556	2.3	179,040,567	3,489	1.6	183,803,308	8,045	362,843,875

Abbreviation: Age-standardized incidence rates, ASR.

is that individuals have increasing risk functions but that heterogeneity between individuals creates the 2-peaked incidence. A peak followed by a subsequent decline in age–incidence curves, as seen for HL, may be interpreted as a frailty phenomenon. The risk at the population level declines once those susceptible, or frail, have acquired the disease. Once they are removed, they leave a pool of low-risk individuals, and a decline in the age–incidence curves is observed after a certain age. Statistical fitting of frailty models to these curves provides an estimate of the variation in risk between individuals at the population level, the fractions of those susceptible in the total population, as well as the number of genetic or epigenetic events required for a cell to reach malignancy. Such frailty analyses have been successfully applied to cancers of the testis (10, 11), colorectum (12), and nasopharynx (13, 14), to add to the existing body of knowledge.

The frailty model is here presented as a hypothesis-driven exploration of the observed peaks in incidence. It obviously represents a simplification, but one would expect that some individual heterogeneity must necessarily be present. Our aim is to see to what extent frailty modeling can explain the observed pattern. An alternative explanation for the first peak might be that an EBV or other infection could by pure chance initiate cancer development in a small minority of unfortunate individuals. Our hypothesis is that this is not a chance phenomenon but also relies on an inborn vulnerability of the individual.

The aim of this study is to advance the understanding of the etiology of HL by applying frailty modeling on HL incidence data from population-based cancer registries. We used data from the Nordic countries to assess the adequacy of various candidate models and interpret the model parameters in light of our current understanding of the biology and etiology of this disease.

Materials and Methods

HL incidence data for the 5 Nordic countries (Denmark, Finland, Iceland, Norway, and Sweden) were

obtained by 5-year age group (0–4, 5–9, ..., 80–84, 85, and over) and sex for the period 1993 to 2007 from NORDCAN, an online database (15) that includes detailed information on cancer incidence, mortality, and prevalence from the respective national population-based registries for more than 5 decades (16). A total of 8,045 incident cases were observed in the study period and 362,843,875 person-years at risk were accumulated (Table 1).

Because of the inherent random variation in the national age-specific rates and small numbers of cases within each age group for each country, aggregation across countries was necessary. Figure 1 shows that the incidence of HL was homogenous across the Nordic countries, justifying a Nordic-wide aggregation before fitting the frailty model.

Modeling

Frailty modeling is an extension of the Cox proportional hazards model in which a random effect (the frailty) has a multiplicative effect on the hazard function. Hence, the hazard rate for an individual is the unobserved frailty, Z , multiplied by the baseline hazard rate for all individuals, $\lambda(t)$, that is, $h(t|Z) = Z\lambda(t)$. Individuals with a high value of Z have a high risk of developing the disease, whereas individuals with a low value of Z are at low risk. The frailty follows a specified distribution, commonly gamma or compound Poisson. The latter distribution was used in this study as it allowed for a nonsusceptible fraction of the population (for which $Z = 0$), given by the parameters (ν and η) of the distribution, as justified below.

The mathematical details of the model are given in the Appendix. In brief, let the subscripts 1 and 2 of i denote parameters for subgroups of the incident cases characterized by the first and second peaks of a bimodal age–incidence curve, respectively. Following a validation of the model as applied to nasopharyngeal cancer by Haugen and colleagues (12), we included 2 independent frailties to allow for the 2-peak age–incidence of HL, with Z_1 representing the risk of developing HL in early

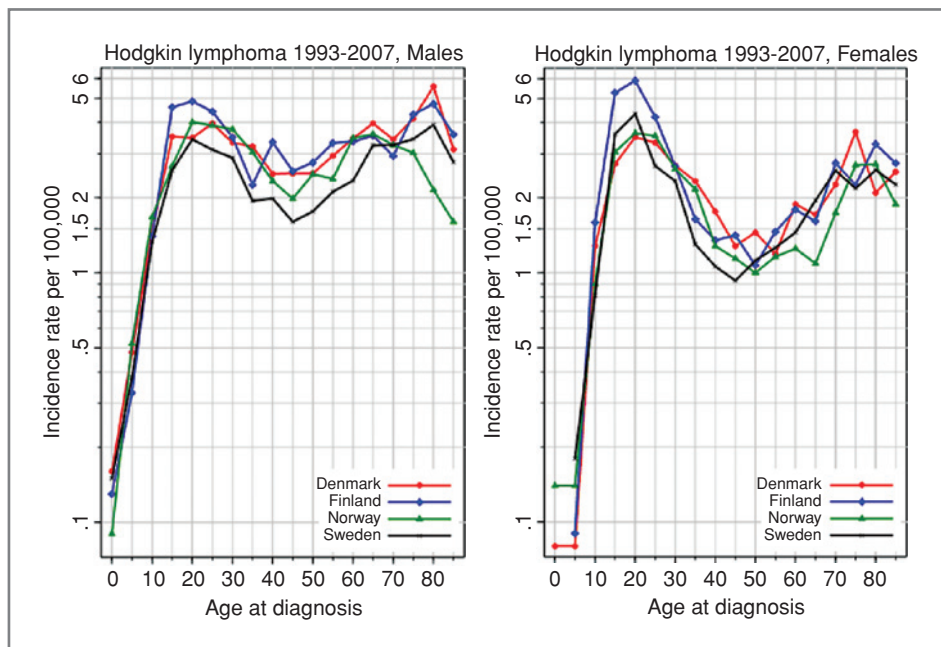


Figure 1. Observed age-specific incidence rates of HL in Denmark, Finland, Sweden, and Norway by sex for the period 1993 to 2007.

adulthood, and Z_2 representing the risk of HL incidence at later ages. The frailty model for the hazard at age t for each individual can then be formulated as:

$$h(t|Z_1, Z_2) = Z_1\lambda_1(t) + Z_2\lambda_2(t) \quad (1)$$

Here Z_1 and Z_2 are frailty variables assumed to follow compound Poisson distributions having 2 parameters η_i and v_i , with $i = 1, 2$. The compound Poisson distribution consists of a positive probability of having a zero frailty combined with a distribution over positive frailty values. The zero frailty corresponds to a group with zero susceptibility (in practice the susceptibility may not be exactly zero, but only very small). The compound Poisson distribution is mathematically a very attractive choice when there is a group with very low susceptibility (17). Hence, the quantity of interest is the proportion of individuals susceptible to HL in each peak, specified as:

$$P(Z_i > 0) = 1 - \exp(-v_i/\eta_i).$$

The baseline hazards $\lambda_1(t)$ and $\lambda_2(t)$ were assumed to follow Weibull distributions for all individuals. The use of Weibull distributions for modeling carcinogenesis dates back to the classical Armitage–Doll model (8). Many cancers show age–incidence rates that can be well approximated by Weibull distributions. These distributions have a scale parameter a_i , modeling the level of incidence of the disease, and a shape parameter k_i . The shape parameter is often interpreted biologically as the number of (genetic or epigenetic) events required on average for a cell to become malignant (8), and therefore should be constant across sex, birth cohort, diagnosis period, and country. As the Weibull shape parameters are allowed to vary for the 2 peaks in the model, this

corresponds to letting HL in the first and second peak to be characterized by a different number of events leading to malignancy, a result that can be interpreted in terms of possible differences in disease etiology.

Explanatory factors can be included either in the frailty distribution or the baseline hazard, leading to models with different interpretations. If for example, sex is included in the frailty distribution (i.e., allowing either η_i and v_i or both to vary by sex), it would yield a model with different proportions of susceptibles for males and females. This approach will fit with biological knowledge indicating that there should be a difference in risk between males and females very early in life. It will also correspond to an interaction between sex and the frailty. If, on the other hand, sex is included in the baseline hazard (i.e., allowing a_i to vary by sex), the proportion of susceptibles is equal for all. However, the underlying disease process as modeled by the baseline hazard may be at a higher level by age for one sex compared with the other, corresponding to for example, different exposure to environmental factors by age. Accordingly, we wish to present 2 contrasting models, one for each of the 2 alternatives given above, and then discuss which model seems most reasonable given the data and the current biological knowledge of HL. Comparing 2 prototype models yields insight, and it is not *a priori* clear which of them would be most correct.

The data were analyzed by using the software packages, Stata (18) and R (19), with the R function *nlminb* used to maximize the log-likelihood function (see Appendix), and standard errors calculated from the Hessian matrix by using the R function *optim*. Model selection was ascertained on comparing log-likelihood values and on the basis of Akaike’s information criterion (AIC; ref. 20,

designed to protect against overfitting and for which the model fit improves as the AIC value decreases) for different models, in addition to visual inspection of the fit of the model, and allowing the parameters to differ by sex at peak incidence. The 95% CIs for the proportions of susceptibles are based on the bootstrap percentile method, using 2.5 and 97.5 percentiles of estimates from 1,000 bootstrap samples of the data, sorted in ascending order.

Results

The 2 models both have 10 parameters. Model 1 allows the frailty parameters η_1 and ν_1 for the first peak to vary by sex (Log likelihood = 36,959, AIC = -73,899). On the basis of AIC and visual inspection, it was unnecessary to let them vary by sex for the second peak. As mentioned in the modeling section, this corresponds to saying that there should be a difference between males and females very early in life, causing them to have different proportion of susceptibles in the first peak. The fit of model 1 to the observed incidence rates for males and females are given in Figure 2A and B, respectively. In model 2, the Weibull scale parameters a_1 and a_2 vary by sex (Log likelihood = 36,952, AIC = -73,884). Hence, in this model, the proportion of susceptibles varies by peak only. The fit of model 2 to the observed incidence rates for males and females are given in Figure 2C and D, respectively.

The estimates for some of the components of models 1 and 2, together with corresponding 95% CIs, are given in Table 2. In model 1, the Weibull shape parameter for the first peak was estimated to be 3.97, indicating that the number of events required for a cell to reach malignancy among the cases diagnosed at ages associated with this peak, is around 4. For the second peak, the corresponding number of events is between 6 and 7. In model 2, the Weibull shape parameter is similar for both peaks, with estimates from 3.5 to 4.0, indicating that 3 to 4 events (rate-limiting steps) are necessary to reach malignancy in both peaks.

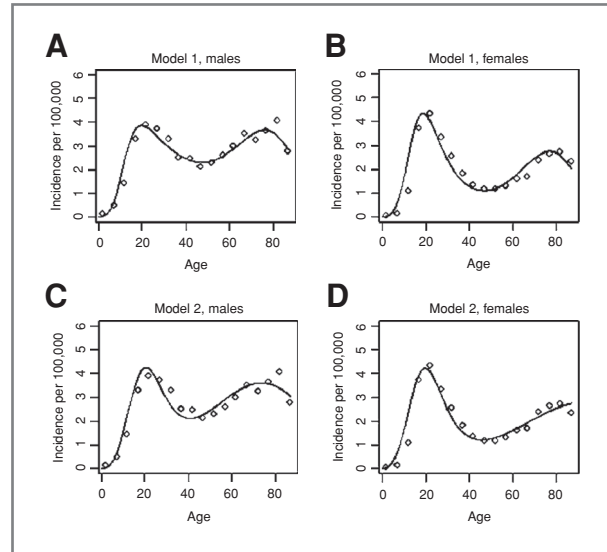


Figure 2. Fit of models 1 and 2 to the observed data by sex. Model 1: Sex difference in frailty distribution (A: Males, B: Females). Model 2: Sex difference in baseline hazard (C: Males, D: Females).

In model 1, the proportion of susceptibles for HL in the first peak is 0.52% for the male population, and 0.11% for the female population, with very wide CI for males. For the second peak, the proportion of susceptibles is 0.09% for both sexes. This results in a higher proportion of susceptibles in both peaks for males, corresponding to the higher incidence of HL observed in males compared with females. In model 2, on the other hand, the proportion of susceptibles is 0.09% for the first peak and 0.20% for the second, thus yielding a proportion of susceptibles of 0.29% for both males and females. This is a dramatic difference compared with model 1, but is compensated by the larger Weibull scale parameter values in model 2. Hence, the level of risk in the baseline hazard is higher (corresponding to greater exposure to environmental factors in this model), resulting in a similar fit to data for both models. As

Table 2. Estimates for certain components in the 2 models

	Model 1 Estimate (95% CI)	Model 2 Estimate (95% CI)
Susceptible males, 1st peak	0.52% (0.29–2.02)	0.09% (0.08–0.10)
Susceptible females, 1st peak	0.11% (0.10–0.12)	
Susceptibles, 2nd peak	0.09% (0.08–0.14)	0.20% (0.16–0.25)
Susceptible males, both peaks	0.61% (0.40–2.12)	0.29% (0.26–0.53)
Susceptible females, both peaks	0.20% (0.19–0.24)	
Weibull shape, 1st peak	3.97 (3.67–4.28)	3.55 (3.30–3.81)
Weibull shape, 2nd peak	6.33 (5.09–7.58)	3.75 (3.17–4.33)
Weibull scale, 1st peak males	5.0×10^{-9} ($2.2 \cdot 10^{-9}$ – $6.8 \cdot 10^{-9}$)	1.3×10^{-8} (6.5×10^{-9} – 2.5×10^{-8})
Weibull scale, 1st peak females		1.4×10^{-8} (7.0×10^{-9} – 2.8×10^{-8})
Weibull scale, 2nd peak males	8.1×10^{-16} ($3.8 \cdot 10^{-18}$ – $2.8 \cdot 10^{-15}$)	1.4×10^{-10} (1.2×10^{-11} – 1.8×10^{-9})
Weibull scale, 2nd peak females		5.9×10^{-11} (4.3×10^{-12} – 7.9×10^{-10})

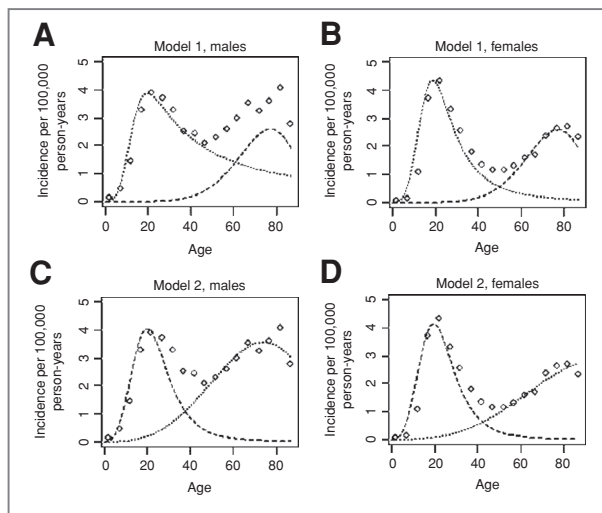


Figure 3. The contribution of each part of the models to the estimated curves. Adding the 2 curves from each part in the panels, gives the estimated age–incidence curves in Figure 2. Model 1: Sex difference in frailty distribution (A: Males, B: Females). Model 2: Sex difference in baseline hazard (C: Males, D: Females).

mentioned in the modeling section, the discrepancies between the models are because of the differences in the parameterization of the models, giving alternative explanations as to what causes the 2 peaks and the observed differences in incidence by sex.

In Figure 3, the respective contribution of the component parts in models 1 and 2 to the observed age incidence is depicted. When the curves from each part are added, one gets the estimated age–incidence curves in Figure 2 for each sex and model. For model 1, the degree of overlap between the 2 parts of the model is particularly apparent for the observed second peak in males, where the large contribution from the curve estimated by the first part of the model, impacts on the observed level of incidence in the second peak. Hence, if there are 2 different etiologies underlying the 2 peaks, the cases in the second peak for males will be a mix of both. For model 2, the degree of overlap is fairly small for both sexes, and the 2 etiologies are more separated, except for a small period in mid age.

Discussion

Frailty models can be usefully applied to age–incidence curves to aid the understanding of biological phenomena in instances in which for a given disease, a minority of individuals are at high risk, compared with a low-risk or, in theory, nonsusceptible, majority. In this study, we have shown that a bimodal frailty model provides a reasonable fit to the age–incidence curves of HL, diagnosed between 1993 and 2007 in the 5 Nordic countries. It shows that interindividual variation in cancer susceptibility can give rise to complicated, for example, bimodal population incidence rates as a function of age, while at the

same time, the individual risk remains simple and monotonically increasing by age. The decline in incidence with age for the component peaks represents exhaustion of the susceptible population, rather than decreases in risk for any individual within the population, which is the basic idea of the frailty model. These findings provide additional insight into the bimodality of the age–incidence curve, the observed sex differences in incidence among older adults, and the variations in the susceptibility of acquiring the disease, each of which is addressed below.

It has been recognized for decades that a key epidemiologic feature of HL is the consistent observation of bimodality of the age–incidence curves and the hypothesized existence of at least 2 etiologically distinct subgroups characterized by their differing ages at peak incidence (1, 2). The current epidemiologic understanding of HL would suggest that the first peak (20 years) relates to specific factors affecting children (EBV-positive) and young adults (predominantly EBV-negative). The second peak (75 years) relates to the determinants of HL affecting older adults (largely resulting from a lack of immunologic control of latent EBV infection). The overall shape of the age–incidence curve is a function of the relative contribution of each of the underlying subgroups (6). The childhood group has been put forward as comprising an epidemiologically separate entity as part of a "three-disease model" (21). In this study, however, there was little sign of such a separate entity among children. Although the data may be too limited to consider a 3-hazard model in the early years, our results support the notion that the age–incidence curve of HL is adequately and robustly described by a bimodal distribution.

In the actual frailty analysis, 2 different modeling approaches have been employed. In model 1, the sex covariate was included in the frailty component, whereas in model 2, it was included in the baseline Weibull hazard. Even though the overall integrated result of both peaks is quite similar in both models, there are noticeable differences when comparing the estimates resulting from the 2 models. These differences include the male-to-female ratio of susceptibles, the k parameter (i.e., the number of rate-limiting steps in the carcinogenic process), and the proportion of susceptibles. The male-to-female ratio varies between 3:1 for model 1 and 1:1 for model 2, and k was estimated to 4 and 6 to 7 for peaks 1 and 2, respectively, in model 1, whereas in model 2, it was estimated to 3.5 to 4 in both peaks.

The subtypes of HL are the same regardless of whether the disease occurs as part of the first or the second peak. It is thus reasonable to assume that the disease process and thus also the number of rate-limiting steps, are similar. If biological data were available to support a similar process for both peaks, one would prefer model 2 over model 1, because the estimated k values of 3.55 and 3.75 for peaks 1 and 2, respectively, from model 2 are reasonable. There is also little biologic evidence supporting the notion that inherent susceptibility varies between males and females,

as implied by model 1. The initial step in the disease development is thought to comprise environmental factors acting in conjunction with inherited susceptibility genes, evenly distributed between males and females (22–28). The assumption of model 2 that susceptibility is similar in both sexes is thus in line with current understanding that the sex difference in HL incidence rates is explained by variation in exposures.

The estimated proportions of susceptible to HL pertaining to both peaks combined varied between the models. For model 1, the estimates were 0.61% and 0.20% for males and females, respectively; and the corresponding values for model 2 were 0.29% for both males and females. The estimates deriving from model 2 are in better accordance with the lifetime (0–84 years) HL risk of 0.23% and 0.17% for males and females, respectively, based on data from NORDCAN (15) than the estimates from model 1. However, model 2 overestimated the proportion of susceptibles somewhat in both males and females. This discrepancy could possibly be due to factors related to the second peak for both men and women, where the estimated curve continues at a high level and would continue above 80 years and further on as well. This indicates that a number of susceptible people simply do not acquire the disease because they die from other causes. All in all, on the basis of the male-to-female ratio, the k parameter values, and the number of susceptible, model 2 seems to fit better with our current understanding of HL than model 1.

One difference between the 2 models is the degree of overlap between the 2 peaks in males. In model 1, the first peak in males is right-skewed, indicating that the second peak contains a mixture of HL with different etiology. In model 2, on the other hand, the 2 peaks in males are more separate, thus suggesting that the second peak contains HL with homogeneous etiology. One interpretation of this sex difference is less heterogeneity in HL in females than in males among elderly. There is some evidence for greater heterogeneity in HL among males compared with females in this age group, as there is a certain occurrence of nodular lymphocyte predominant subtype in elderly males in addition to the more common subtypes of nodular sclerosis and mixed cellularity/lymphocytes depleted (29). Accordingly, because the overlap between the 2 peaks in males is featured in model 1, but not in model 2, model 1 seems to fit better with our present knowledge of heterogeneity in HL in elderly males than model 2.

The idea underlying the frailty model is that a small fraction of individuals have a high risk of a particular disease, whereas the majority are associated with a small risk. Inherited susceptibility genes acting in conjunction with environmental factors are thought to comprise the initial steps in the disease development. The growing evidence of genetic predisposition playing a role in HL (22) might provide the biologic rationale for this model. It has been known for many years that genetically determined variation (polymorphisms) in the human leukocyte antigen (HLA) region is associated with sus-

ceptibility to HL (23). Recent studies have shown that HLA class I and III regions are associated with susceptibility to EBV-positive and EBV-negative HL, respectively (24). Furthermore, the etiological heterogeneity between these 2 types of HL has been shown to relate to HLA-A alleles and cytotoxic T-cell response to EBV infection (25). In addition, there is accumulating evidence suggesting that polymorphisms in other relevant genes are associated with the risk of HL, such as B-cell developmental genes (26), glutathione S-transferase genes (27), and DNA repair genes (28).

On the basis of a simulation study of genome-wide association studies, Dickson and colleagues (30) suggested that rare genetic variants with large effects on disease risk may be implicated in a large proportion of cases. Hence, risk might be concentrated on few individuals rather than being distributed more evenly over many people, which fits with our frailty thinking. Yet another mechanism possibly underlying frailty is a model in which random mutations occurring at a very young age produce a developmental disposition to cancer (31). This is not genetically determined but due to random and unpredictable events.

The uniform trends across countries provide a rationale for collapsing across them, thus ensuring a sufficient accrual of cases to adequately interpret the model estimates. From 1980 onwards, there was a decreasing trend of HL in European countries (32). This is likely the result of improved diagnostic tools introduced in these populations circa 1980, via enhancements in immunohistochemistry, for example, more specific B- and T-cell markers, in addition to Southern blot analysis of immunoglobulin and T-cell receptor gene rearrangements. An increasing proportion of patients previously diagnosed as mixed cellularity and lymphocyte-depleted HL were classified as non-HL (33). The incidence rates stabilized around 1990, and to avoid bias due to this artificial reduction in HL incidence rates, only data from more recent years of diagnosis, that is, 1993 to 2007, were used in the frailty analysis.

As pointed out most recently by Byrne (34), the Armitage–Doll model provides a reasonable fit for cancers of the colon, stomach, and pancreas (8) but is an inadequate representation for the majority of neoplasms. We show that the applicability of the Armitage–Doll model can be extended if one considers the heterogeneity between individuals, via frailty models, and even incidence curves with 2 peaks are consistent with the former model. The parameter k has traditionally been attributed as a proxy of the number of rate-limiting cellular changes needed for cancer progression. This view has been useful, but clearly represents an oversimplification of the complex mechanisms underlying cancer development (34). Still, the parameter is a natural model parameter, being part of the standard Weibull distribution, and remains of interest as a marker of important biological features. In the case of HL, the k parameter could be translated into

rate-limiting steps in the carcinogenic process leading to HL, such as acquisition of EBV infection, loss of immune control (pertaining to peak 2 of the age-incidence curve), and subsequent transitions toward malignancy.

In summary, the present study has shown that the age-incidence curves of HL, as diagnosed during 1993 to 2007 in the 5 Nordic countries, are well described by bimodal frailty models. The good fit is not in itself a proof of the validity of the model; however, it illustrates that heterogeneity in cancer susceptibility between individuals may give rise to bimodal age-incidence rates at the population level, although the individual risk remains simple and monotonically increasing by age. Frailty modeling of HL adds to the existing body of knowledge on the heterogeneity between individuals in terms of susceptibility pertaining to this cancer form. At present, there are more arguments in favor of model 2 than model 1. However, this may change if new information on the underlying mechanisms behind HL is acquired in the future. A follow-up modeling study encompassing the histologic subtypes of HL seems to be warranted to further explore the heterogeneity in risk of acquiring this disease.

Appendix

Let the subscripts 1, 2 indicate parameters for peaks 1 and 2, respectively. The compound Poisson distributions for the frailties Z_1 and Z_2 normally have 3 parameters ρ_i , η_i , v_i with expected value $E(Z_i) = \rho_i \eta_i / v_i$, $i = 1, 2$. The Weibull distributions have a scale parameter a_i and a shape parameter k_i with $\lambda_i(t) = a_i k_i t^{k_i - 1}$. Because both the compound Poisson distributions and the baseline hazards include scale parameters ρ_i and a_i , we fix the expected value of each frailty distribution to 1 by setting $\rho_i = v_i / \eta_i$ to secure identifiability of the model.

The individual survival function can be found by conditioning on the frailties Z_1 , Z_2 and integrating them out. The conditional survival function is

$$S(t|Z_1, Z_2) = \exp(-Z_1 \Lambda_1(t) - Z_2 \Lambda_2(t))$$

in which $\Lambda_i(t) = \int_0^t \lambda_i(s) ds = a_i t^{k_i}$ is the cumulative basic hazard rate. To find the unconditional or population

survival function, we calculate $E[S(t|Z_1, Z_2)]$ by integrating over the frailties (12). The result becomes:

$$S(t) = E[S(t|Z_1, Z_2)] = \exp\left(-v_1/\eta_1 \left[1 - \left(\frac{v_1}{v_1 + \Lambda_1(t)}\right)^{\eta_1}\right] - v_2/\eta_2 \left[1 - \left(\frac{v_2}{v_2 + \Lambda_2(t)}\right)^{\eta_2}\right]\right).$$

The proportion of susceptibles in each peak follows from the parameters of the compound Poisson frailty distributions as $P(Z_1 > 0) = 1 - \exp(-v_1/\eta_1)$ and $P(Z_2 > 0) = 1 - \exp(-v_2/\eta_2)$. Because Z_1 and Z_2 are independent, the total proportion of susceptibles in either first, second, or both peaks is calculated as

$$\begin{aligned} P(Z_1 > 0 \text{ or } Z_2 > 0) &= P(Z_1 > 0) + P(Z_2 > 0) \\ &\quad - P(Z_1 > 0) \cdot P(Z_2 > 0) \\ &= 1 - \exp(-v_1/\eta_1 - v_2/\eta_2) \end{aligned}$$

Assuming a Poisson model, the likelihood function is

$$L = \prod_{l=1}^{18} \prod_{m=1}^2 (\mu_{l,m})^{R_{l,m}} \times \exp(-\mu_{l,m})$$

in which $\mu_{l,m}$ is the expected number of cases in age interval l from 1 to 18 with sex m , and R is the observed number of cases. The expected number of cases is then

$$\mu_{l,m} = T_{l,m} [\ln(S(t_{l-1})) - \ln(S(t_l))]/5$$

in which $T_{l,m}$ is the number of person-years at risk for age interval l with sex m .

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Grant Support

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received September 24, 2010; revised April 24, 2011; accepted April 26, 2011; published OnlineFirst May 10, 2011.

References

- MacMahon B. Epidemiological evidence of the nature of Hodgkin's disease. *Cancer* 1957;10:1045-54.
- MacMahon B. Epidemiology of Hodgkin's disease. *Cancer Res* 1966;26:1189-201.
- Mueller N. Hodgkin's disease. In: Schottenfeld D, Fraumeni JF Jr, editors. *Cancer epidemiology and prevention* New York: Oxford University Press; 1996. p. 893-919.
- Biggar RJ, Jaffe ES, Goedert JJ, Chaturvedi A, Pfeiffer R, Engels EA. Hodgkin lymphoma and immunodeficiency in persons with HIV/AIDS. *Blood* 2006;108:3786-91.
- Weiss LM, Strickler JG, Warnke RA, Purtilo DT, Sklar J. Epstein-Barr viral DNA in tissues of Hodgkin's disease. *Am J Pathol* 1987;129:86-91.
- Hjalgrim H, Engels EA. Infectious aetiology of Hodgkin and non-Hodgkin lymphomas: a review of the epidemiological evidence. *J Intern Med* 2008;264:537-48.
- Jarrett RF. Viruses and lymphoma/leukaemia. *J Pathol* 2006;208:176-86.
- Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br J Cancer* 1954;8:1-12.
- Meza R, Jeon J, Moolgavkar SH, Luebeck EG. Age-specific incidence of cancer: Phases, transitions, and biological implications. *Proc Natl Acad Sci U S A* 2008;105:16284-9.
- Aalen OO, Tretli S. Analyzing incidence of testis cancer by means of a frailty model. *Cancer Causes Control* 1999;10:285-92.

11. Moger TA, Aalen OO, Halvorsen TO, Storm HH, Tretli S. Frailty modelling of testicular cancer incidence using Scandinavian data. *Biostatistics* 2004;5:1–14.
12. Svensson E, Moger TA, Tretli S, Aalen OO, Grotmol T. Frailty modelling of colorectal cancer incidence in Norway: indications that individual heterogeneity in risk is related to birth cohort. *Eur J Epidemiol* 2006;21:587–93.
13. Bray F, Haugen M, Moger TA, Tretli S, Aalen OO, Grotmol T. Age-incidence curves of nasopharyngeal carcinoma worldwide: bimodality in low-risk populations and aetiological implications. *Cancer Epidemiol Biomarkers Prev* 2008;17:2356–65.
14. Haugen M, Bray F, Grotmol T, Tretli S, Aalen OO, Moger TA. Frailty modeling of bimodal age-incidence curves of nasopharyngeal carcinoma in low-risk populations. *Biostatistics* 2009;10:501–14.
15. www.aacr.org
16. Engholm G, Ferlay J, Christensen N, Bray F, Gjerstorff ML, Klint A, et al. NORDCAN – a Nordic tool for cancer information, planning, quality control and research. *Acta Oncol* 2010;49:725–36.
17. Aalen OO, Borgan Ø, Gjessing HK. Unobserved heterogeneity: The odd effects of frailty. In: *Survival and event history analysis: a process point of view* New York: Springer-Verlag; 2008. p. 231–70.
18. Stata Statistical Software, College Station: StataCorp, editor, TX, StataCorp LP; 2009.
19. R: A language and environment for statistical computing R development Core Teameditor, Vienna, R Foundation for Statistical Computing; 2008.
20. Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974;19:716–23.
21. Armstrong AA, Alexander FE, Cartwright R, Angus B, Krajewski AS, Wright DH, et al. Epstein-Barr virus and Hodgkin's disease: further evidence for the three disease hypothesis. *Leukemia* 1998;12:1272–6.
22. Landgren O, Caporaso NE. New aspects in descriptive, etiologic, and molecular epidemiology of Hodgkin's lymphoma. *Hematol Oncol Clin North Am* 2007;21:825–40.
23. Hors J, Dausset J. HLA and susceptibility to Hodgkin's disease. *Immunol Rev* 1983;70:167–92.
24. Diepstra A, Niens M, Vellenga E, van Imhoff GW, Nolte IM, Schaapveld M, et al. Association with HLA class I in Epstein-Barr-virus-positive and with HLA class III in Epstein-Barr-virus-negative Hodgkin's lymphoma. *Lancet* 2005;365:2216–24.
25. Hjalgrim H, Rostgaard K, Johnson PC, Lake A, Shield L, Little AM, et al. HLA-A alleles and infectious mononucleosis suggest a critical role for cytotoxic T-cell response in EBV-related Hodgkin lymphoma. *Proc Natl Acad Sci U S A* 2010;107:6400–5.
26. Broderick P, Cunningham D, Vijayakrishnan J, Cooke R, Ashworth A, Swerdlow A, et al. IRF4 polymorphism rs872071 and risk of Hodgkin lymphoma. *Br J Haematol* 2010;148:413–5.
27. Lourenço GJ, Néri IA, Sforni VC, Kameo R, Lorand-Metze I, Lima CS. Polymorphisms of glutathione S-transferase Mu 1, glutathione S-transferase theta 1 and glutathione S-transferase Pi 1 genes in Hodgkin's lymphoma susceptibility and progression. *Leuk Lymphoma* 2009;50:1005–9.
28. El-Zein R, Monroy CM, Etzel CJ, Cortes AC, Xing Y, Collier AL, et al. Genetic polymorphisms in DNA repair genes as modulators of Hodgkin disease risk. *Cancer* 2009;115:1651–9.
29. Morton LM, Wang SS, Devesa SS, Hartge P, Weisenburger DD, Linet MS. Lymphoma incidence patterns by WHO subtype in the United States, 1992–2001. *Blood* 2006;107:265–76.
30. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. *PLoS Biol* 2010;8:e1000294.
31. Frank SA, Nowak MA. Developmental predisposition to cancer. *Nature* 2003;422:494.
32. Morgan G, Vornanen M, Puitinen J, Naukkarinen A, Brincker H, Olsen J, et al. Changing trends in the incidence of non-Hodgkin's lymphoma in Europe. Biomed Study Group. *Ann Oncol* 1997;8 Suppl 2:49–54.
33. Foss Abrahamsen A, Egeland T, Hansen S, Langholm R, Holte H, Kvaloy S. Hodgkin's disease in a national and hospital population: trends over 20 years. *Eur J Cancer* 1997;33:2380–3.
34. Byrne HM. Dissecting cancer through mathematics: from the cell to the animal model. *Nat Rev Cancer* 2010;10:221–30.