Meta-analytical approaches for pooling time to event individual data: an application to non Hodgkin lymphoma survival studies

Giulia Carreras, Anna Maria Pizzo

Affiliation: ISPO Cancer Prevention and Research Institute,
via di S.Salvi12, 50135, Florence, ITALY

Corresponding Author:
ISPO Cancer Prevention and Research Institute,
via di S.Salvi12, 50135, Florence, ITALY
Tel +390556268375, Fax +39055679954
Email: g.carreras@ispo.toscana.it

Summary

Introduction. In combining data from several studies, meta-analytical methods must be adopted. Meta-analysis involves combining summary information or pooling individual data from different studies to integrate findings. Meta-analyses are performed by using fixed or random effects models, the latter adopted for analysing heterogeneity.

Objectives. This work aims to examine and compare meta-analytical techniques applied to survival analysis for individual data.

Methods. In survival data, analyzed by Cox regression models, heterogeneity may be attributed to differences in baseline functions or to the investigated effect. Three different random effect models were identified depending on variables that model heterogeneity: the baseline hazard, the covariate effect or both. A fixed effect model was also analyzed. Data from four studies on survival from Non-Hodgkin lymphoma were gathered. The pooled effect of smoking on overall survival was evaluated by applying the four models.

Results. Results showed a better performance of random effect models respect to the fixed effect one. The baseline hazard appeared to better capture heterogeneity, suggesting that heterogeneity was mainly due to differences among study-populations.

Conclusions. Random effect models resulted easy to be implemented in meta-analysis of individual survival data, necessary for studying heterogeneity and in our application better performing respect to fixed ones.

KEY WORDS: *Cox model regression, individual patient data, meta-analysis, random effects, shared frailty model*

Introduction

Meta-analysis is frequently used as a statistical method for combining information (results or data) from several studies comparable in outcome and exposure with the aim of increasing the study-power and integrating the findings.

Many meta-analytical works use aggregate data (AD) and focus on the combination of published summary statistics usually in the form of weighted averages (1-5). Alternatively, individual patient data (IPD) can be employed and their use is now considered the gold standard in meta-analytical works (1, 6-7). IPD meta-analyses include more detailed data, allowing accurate classification of patients based on individual characteristics. Meta-analyses of IPD also offer investigators the opportunity to examine uncommon exposures, rare diseases, and variation in associations among population subgroups with greater statistical power than in individual studies (8). IPD are however costly to be obtained.

In performing a meta-analysis, using both AD or IPD, source of variation is a crucial point to consider. This can be due at least to three different aspects: sampling error, study-level characteristics

(within-study variability) and between-study variability or heterogeneity (8-9).
The sampling error measures the precision of the study-specific estimate and derives from the use of a sample instead of the all population. Uncertainty due to study-level characteristics is the variability in the patient responses within the primary studies. Variation due to factors that vary across the pooled studies results in heterogeneity. Differing design features, inclusion criteria, methodological quality, clinical procedures, patient population or characteristics and therapeutic schemes or practices are all factors that can contribute to such variability and should be investigated (9).
First of all, in performing a meta-analysis, exposure, confounding, and outcome variables must be standardized to remove potential sources of heterogeneity across studies. Afterwards, meta-analysis can be undertaken by the use of fixed or random effects models.
Fixed effects models deal with sampling error and uncertainty on study-level characteristics, assuming that each study is measuring the same underlying parameter, indeed assume no between-study variability. Random effects models take into account heterogeneity and, conversely to the fixed effects models, assume that each study is associated with a different but related parameter (9-10).
IPD meta-analyses can be tackled with two approaches: one and two-stage. The one-stage is a fixed effect approach that considers the IPD pooled as one dataset and performs a single analysis (1). Heterogeneity cannot be handled because the investigated effect is assumed to be identical in each study. On the other hand, in the two-stage approach studies are analyzed separately and then summary statistics are combined using standard meta-analytical random effect methods (1), as the one developed by Der Simmonian and Laird (3).
A review of 44 IPD meta-analyses published during the years 1999-2001 showed that 81% of the meta-analyses for time to event data adopted fixed effect models (1). Recently, some authors (6, 11-12) compared fixed and random effects models for time to event IPD, anyway random effect models remain not largely diffused (13).
This work examined and compared meta-analytical techniques applied to time to event IPD from a pooled dataset. Data from four different studies on survival from Non-Hodgkin Lymphoma (NHL) were used and the effect of tobacco smoking exposure on overall survival was investigated.

Material And Methods

A general framework for conducting IPD meta-analyses entails identifying all studies meeting identical inclusion criteria, obtaining each study's primary data, creating a standardized database, estimating study-specific exposure-disease associations and examining whether the study-specific results are heterogeneous (8). Even if all these points were faced, this work focused on comparing fixed and random effect models for time to event individual data, in investigating heterogeneity and in calculating pooled estimates.

Models for meta-analysis of primary data

Usually, models for meta-analysis on time to event IPD refers to the standard Cox proportional hazards regression model (model 0) (6, 11-13). The hazard rate for individual *i* in the *j*th study at time *t* is given by:

$$\text{model 0} \qquad \lambda_{ij}(t) = \lambda_0(t)e^{\beta \cdot x_{ij}},$$

where: the covariates x_{ij} are the exposure or the confounding variables under study for individual *i* and study *j;* the term $\lambda_0(t)$ is the baseline hazard function for time since entry into the study *t*, that is the time-dependent hazard when all the covariates are equal to zero; the coefficient β estimates the log relative risk for a one-unite increase in the covariates and it is therefore a measure of the covariate effect. Model 0 assumes a multiplicative relationship between the hazard function and the log-linear function of the covariates (the proportionality of hazards assumption).
Here, four meta-analytical models on time to event IPD were identified. Model 1 is a fixed effect model, while models 2, 3 and 4 are random effect models.
In the fixed-effect Cox regression model stratified by study (model 1) the baseline hazards $\lambda_0(t)$ are allowed to be different in each stratum *j* (study),

but the covariate effect β is assumed to be the same in all the strata (6, 12):

$$\text{model 1} \qquad \lambda_{ij}(t) = \lambda_{0j}(t)e^{\beta x_{ij}}.$$

The disadvantage of this model is that we do not get a simple summary estimate of the effect of the stratifying variable.

Random effect Cox models can model the heterogeneity both in the baseline hazard and/or in the covariate effect (8). Three different random effect models were then identified depending on the variables that model heterogeneity: the baseline hazard, the covariate effect or both.

In the shared frailty model (model 2) heterogeneity in the baseline hazard is taken into account (13). In this case, a group-level frailty α_j, that is an unobservable positive latent random effect, enters multiplicatively on the hazard function or additively as intercept (β_{1j}):

$$\text{model 2} \qquad \begin{aligned} &\lambda_{ij}(t) = \lambda_0(t)\alpha_j e^{\beta x_{1ij}} = \lambda_0(t)e^{\beta x_{1ij}+\beta_{1j}} \\ &\beta_{1j} = \ln(\alpha_j) \sim N(0,\tau^2). \end{aligned}$$

The frailty effect α_j measures the variation in study-specific hazard after controlling for observed variables. The frailty α_j natural logarithm transformation is assumed to follow a Normal distribution with 0 mean and variance τ^2, where the τ^2 represents the between-study variability on the baseline hazard. A larger variance τ^2 implies greater heterogeneity across studies; a null variance τ^2 indicates no heterogeneity over studies. Many frailty models have been considered depending on the frailty distribution. Commonly applied frailty distributions are the gamma distribution (14), the positive stable distribution (15), a three-parameter distribution (16), the compound Poisson distribution (17) and the log-normal distribution (18). In this work, the log-normal distribution was used for consistency with the other following models.

In the covariate random effect model (model 3) heterogeneity in covariates is taken into account by assuming that β varies randomly among studies (j) (6):

$$\text{model 3} \qquad \begin{aligned} &\lambda_{ij}(t) = \lambda_{0j}(t)e^{\beta_j x_{ij}} \\ &\beta_j = \beta + b_j,\ b_j \sim N(0,\tau_1^2), \end{aligned}$$

where t_1^2 represents the between studies variability of the effects.

Heterogeneity both in the baseline and in the covariate effect was also modelled (model 4) combining the shared frailty and the covariate random effect models (6, 11, 19):

$$\text{model 4} \qquad \begin{aligned} &\lambda_{ij}(t) = \lambda_0(t)e^{\beta_j x_{1ij}+\beta_{1j}} \\ &\beta_j = \beta + b_j,\ b_j \sim N(0,\tau_1^2) \\ &\beta_{1j} = \beta_1 + b_{1j},\ b_{1j} \sim N(0,\tau^2). \end{aligned}$$

The terms τ^2 and τ_1^2 represent the between-study variability on the baseline hazard and the heterogeneity of the effects over studies respectively.

All the models were sex and age-adjusted, considering these variables as confounders in each cancer survival analysis.

The partial penalized likelihood approach was used to estimate parameters in the Cox models that include random effects (19).

Performance of all models was tested using a log-likelihood ratio test (log-LRT) comparing the models log-likelihoods L_1 with null model one L_0 by the use of the statistic $2(L_1-L_0)$ distributed as a chi-square with degree of freedom equal to the number of covariates (20).

Common methods to assess heterogeneity are the Q-test (3) and the LRT (6). In this work the LRT was used comparing the performance of the random effect models with the corresponding fixed effect model i.e. the same model with null between-study variability: random effect model log-likelihood L_R was compared with the corresponding fixed effect model one (L_F) using the statistic $2(L_R-L_F)$ distributed as a chi-square with degrees of freedom equal to the number of added random terms. The R packages coxme and survival were used respectively for the random effect meta-analysis and for the other analyses (19-21).

Application to NHL data

To examine meta-analytical techniques applied to time to event IPD using the defined fixed and random effects models, NHL survival data were used. The NHL data refers to a subset including only incident cases and ever smokers, selected from case-control studies part of the InterLymph Consortium (information about it are available at http://epi.grants.cancer.gov/InterLymph/). Data from three Italian and one USA's InterLymph case-control studies (n=2618) were available. Table 1

shows study-specific characteristics including total number of cases and number of the main frequent NHL histotype (diffuse Large B cell Lymphoma, DLBCL) (n=728). In each of the individual studies, data were collected using trained interviewers, who administered standardized, structured questionnaires to study subjects. Consent was obtained from all participant subjects in the different studies. Individual-level information was collected.

In this work, two survival analyses were carried out: on all NHL cases and on a subgroup of NHL cases classified as DLBCL histotype. A histotype analysis is a main epidemiological issue in NHL survival because the survival pattern considerably differs in respect to the histotype. For some aggressive NHL histotypes the survival curves decline rapidly in the early months following diagnosis, but eventually levels off over time; this contrasts dramatically with that of some indolent lymphomas, where a gradual steady decline was observed over the entire period of follow-up.

The analyses were carried out on a selected group of NHL cases, i.e. ever smokers (current and former smokers) (Table 1) studying the effect of tobacco smoking on survival. The selection of ever smokers arose from the results of a previous study on a subset of cases from Italian multicentre case control study. This study indicated that heavy smokers had a worse survival compared with those with a lower cumulative exposure to tobacco smoking (22).

In this work the tobacco smoking exposure was referred to prior diagnosis and the lifetime tobacco consumption was measured with the pack-year variable, computed as number of cigarettes smoked per day multiplied by number of years smoked divided for 20 (1 pack has 20 cigarettes). In the analyses, the pack-year variable was multiplied by 10, to better illustrate differences on hazard for increasing levels of smoking consumption. The pack-year is the covariate assumed to vary between studies in the random effect models.

As explorative analysis, the differences among studies on survival by smoking attitude were studied with univaried Kaplan Meier analysis. Two smoking categories were defined on the basis of the cut-off in pack-year category that was significantly associated to the risk of disease occurrence in an InterLymph aetiological pooled analysis (23): those who smoked less than 20 pack-years and those who smoked 20 or more pack-years. Moreover the cumulative exposure of 20 pack-years corresponded to the median of the distribution of our data. Kaplan Meier curves were calculated for both groups and Peto-Wilcoxon tests were applied to test differences among study-specific survival curves.

Results

Results of the preliminary explorative analysis to investigate differences on survival among studies are in Figure 1. The Kaplan-Meier five-years survival curves by study both for all histotypes and

Table 1. Study characteristics.

Study name (abbreviation)	Year(s) of diagnosis	NHL Cases (smokers)	DLBCL histotype cases (smokers)
Aviano/Aviano-Milan (Aviano)	1983–1992 / 1999–2002	353 (203)	114 (61)
Italian multicentre case-control study (Italy)	1991-1993	1624 (959)	434 (258)
Nebraska NHL Study (Nebraska)	1983-1986	383 (195)	91 (52)
EPILYMPH – Sardinia (Sardinia)	1999-2004	258 (167)	89 (63)
Total		2618 (1524)	728 (434)

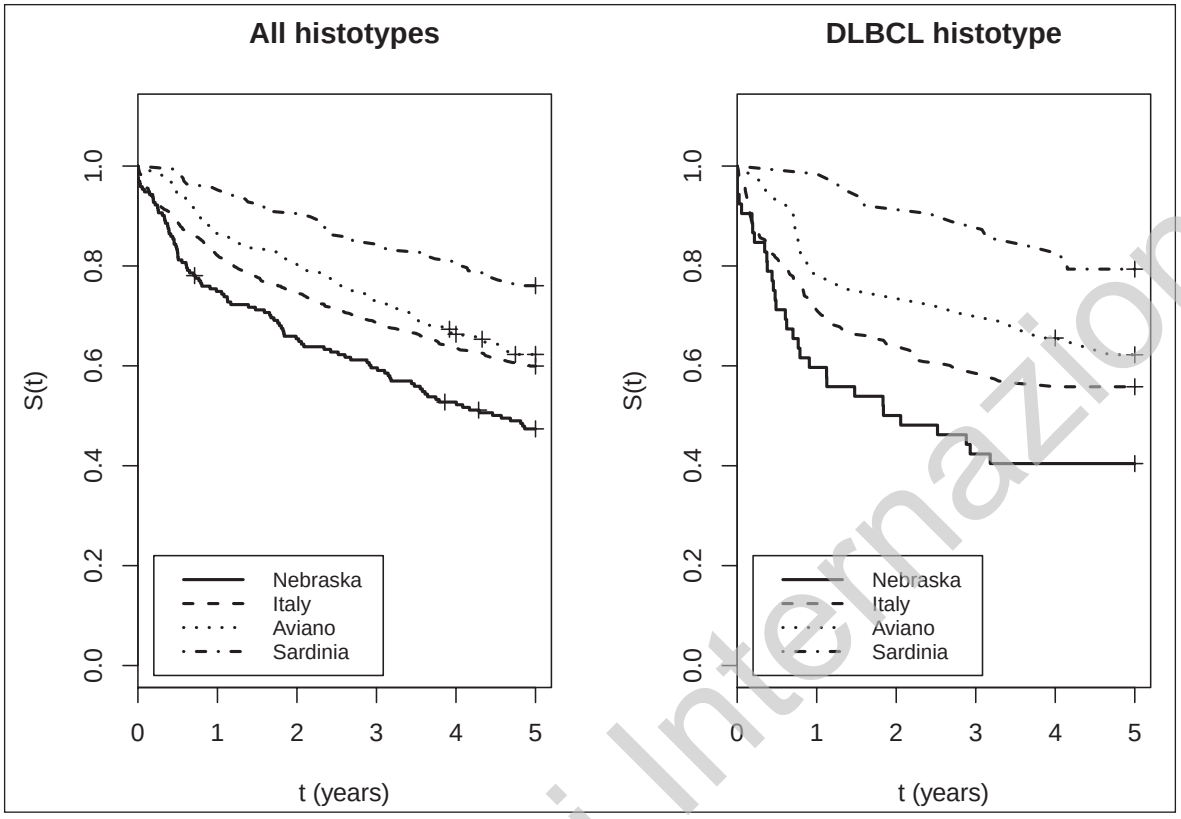

Figure 1. Kaplan-Meier five-years survival curves both for all histotypes and only for DLBCL histotype by study.

for DLBCL histotype resulted different among studies (Peto-Wilcoxon p<0.0001, for all histotypes and for DLBCL histotype).

In Figure 2 are the Kaplan-Meier five-years survival curves both for all histotypes and for DLBCL by smoking group. There were significant differences in survivals among studies in both smoking categories for all histotypes (Peto-Wilcoxon test p= 0.0108, for 20 or less pack-years and p=0.0195 for more than 20 pack-years smoked), as well as for DLBCL histotype analysis (Peto-Wilcoxon p= 0.0166, for 20 or less pack-years and 0.0389 for more than 20 pack-years smoked).

We applied the four models to study the tobacco smoking effect on NHL survival. A model with the study x_{1ij} as a covariate was initially considered, however that model did not verified the proportional hazard assumption and it was excluded from these analyses. Model performances were tested with the log-LRTs (Table 2). All the investigated models significantly better fit the data respect to the null one both in all histotypes and in DLBCL histotype analysis. All the random effect models (model 2, 3, 4) better performed respect to the fixed effect one (model 1): there is an improvement in likelihood going from the Cox stratified model to all the random effect models. Considering the performance of the random effect models, the model 2 and the model 4 showed a very similar LR hence no improvement was associated to the tobacco smoking random effect. This was confirmed by a lower LR value for model 3 respect to model 2 and 4.

The log-LRTs were applied also to investigate heterogeneity: models with random effect terms were compared with the corresponding fixed effect models (Table 3).

Both for all NHL histotypes and for DLBCL histotype, models 2 and 4 showed a statistically significant improvement in LR respect to the corresponding fixed effect models: the random effect terms were significantly different from zero indicating the existence of between study variability.

Model 3 had a different behaviour between the all histotypes and the DLBCL analyses. In the all histotypes analysis model 3 did not show an improvement in LR respect to the corresponding fi-

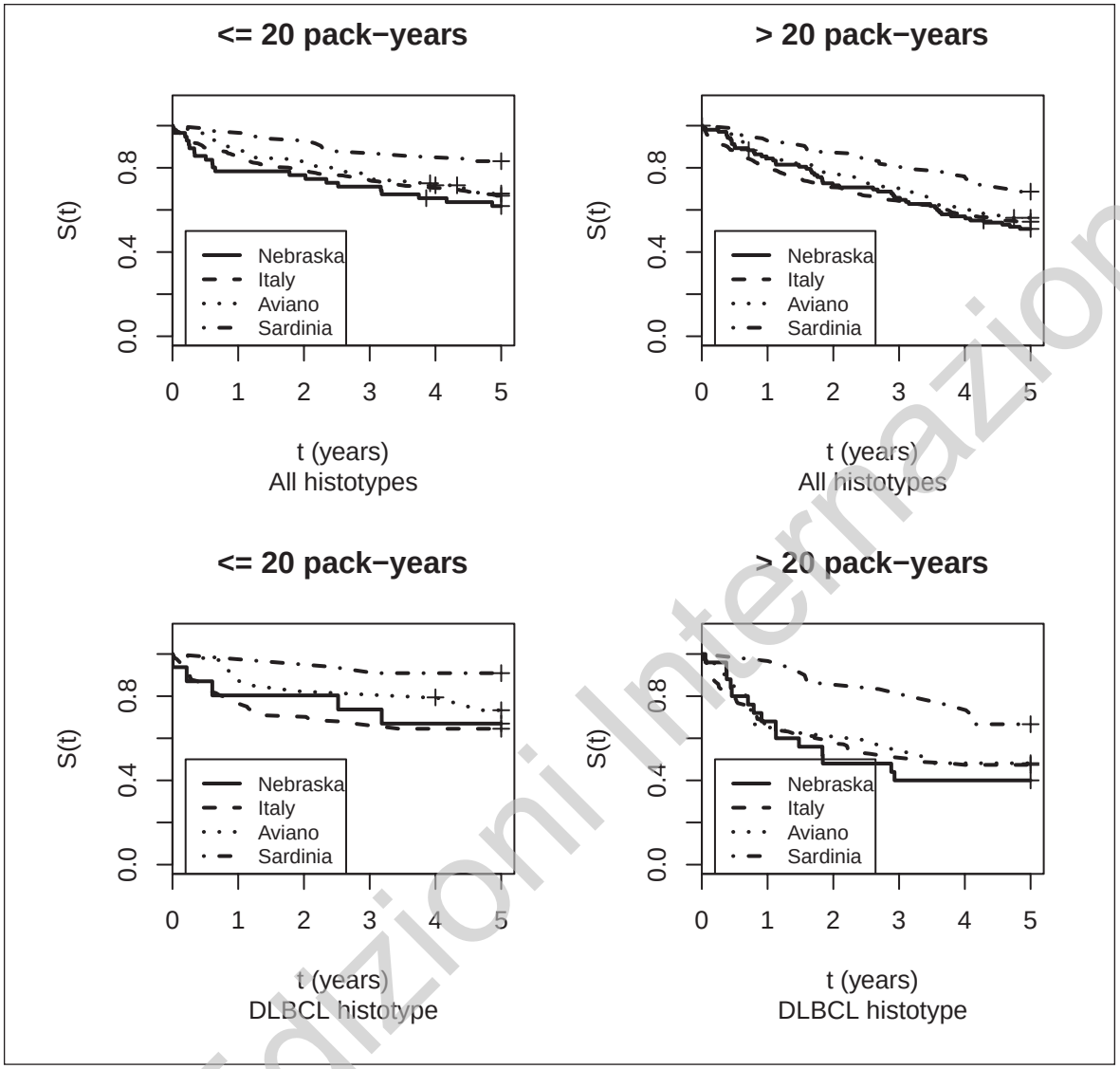

Figure 2. Kaplan-Meier five-years survival curves both for all histotypes and only for DLBCL histotype by study and smoking group.

xed effect model: the random component was not significantly different from 0. On the contrary, in the DLBCL analysis, a weak improvement was observed by adding the random effect, i.e. the random component of the tobacco smoking covariate was significantly different from 0.

Heterogeneity was also explored by analyzing the variances of the random terms (Table 4). Variance values were not negligible for the random intercept i.e. the effect associated to the baseline hazard. Variance values for the tobacco smoking random term were smaller than those for the intercept. The variance of the tobacco smoking random effect in the all subtypes analysis decreased from model 3 to model 4 i.e. by adding the random intercept parameter, while the opposite was observed in the DLBCL analysis. However these variances values were so small than their contrasting behaviours could be not meaningful.

The tobacco smoking hazard ratios (HRs) were also computed. Figure 3 shows the forest plot for the HRs estimates obtained by applying the Cox model (model 0) to each study and applying the four meta-analytical models to the pooled data. The reported HRs refers to all histotypes analysis since the same results were observed in the DLBCL analysis. When data were pooled, all the investigated models reported a statistically significant to-

Table 2. Log-likelihood ratios tests results: statistics (LR) with the corresponding p-values (p) and degrees of freedom (df).

	All histotypes		DLBCL histotype		
Model	**LR**	**P**	**LR**	**P**	**df**
model 1	79.7	<0.0001	29.80	<0.0001	3
model 2	93.42	<0.0001	40.66	<0.0001	4
model 3	84.40	<0.0001	37.56	<0.0001	4
model 4	93.42	<0.0001	40.67	<0.0001	5

Table 3. Log-likelihood ratios tests results: statistics (LR) with the corresponding p-values (p) and degrees of freedom (df).

	All histotypes		DLBCL histotype		
Model	**LR**	**P**	**LR**	**P**	**df**
model 2	9.19	0.0024	7.57	0.0059	1
model 3	0.17	0.6753	4.47	0.0344	1
model 4	9.19	0.0100	7.57	0.0226	2

Table 4. Between-study variability on the baseline hazard (τ^2) and on the tobacco smoking effect (pack-year) over studies (τ_1^2) in models 2, 3 and 4.

Model	**Random terms**	**All histotypes**	**DLBCL histotype**
model 2	τ^2 (intercept)	$7.37\ 10^{-2}$	$1.46\ 10^{-1}$
model 3	τ_1^2 (pack-year)	$1.52\ 10^{-14}$	$3.08\ 10^{-3}$
model 4	τ^2 (intercept)	$7.37\ 10^{-2}$	$1.47\ 10^{-1}$
	τ_1^2 (pack-year)	$5.21\ 10^{-10}$	$3.39\ 10^{-6}$

bacco smoking HR, confirming an increasing hazard risk with a ten-unit increasing of smoking cumulative exposure. The HR pooled estimates and their confidence intervals were very similar, differing only in decimals, whichever model was applied to the NHL data. As expected, the variability of the pooled HRs was reduced respect to the single studies ones.

Discussion

This work aimed to examine and compare fixed and random effect models for meta-analysis of survival IPD: a fixed effect model (model 1) and three different random effect models (model 2, 3, 4) were defined on the basis of the Cox regression model. By way of example, the models were applied to survival data of NHL cases, focusing on tobacco smoking exposure. Only tobacco smoking, as prognostic factor, and age and sex, as confounding variables, were studied because uniformly collected and measured in all the studies. This work is just a starting point for further analyses designed to examine the prognostic value of various factors on NHL survival making use of an extended database including more studies.

The preliminary univariate analysis showed some differences in survival from NHL, observed both by study and by study and smoking group (figures 1 and 2). These differences could be due to heterogeneity among studies; such heterogeneity was investigated, measured and eventually explained by applying random effect models.

Variation in the investigated effect may be due to differences in baseline functions between studies i.e. at the population level, or may arise from the effect itself, i.e. the covariate may have a different prognostic effect in some studies than in others (12). There are at least two possible ways to capture this heterogeneity: by attributing it to the single covariates as well as to the baseline hazard. The

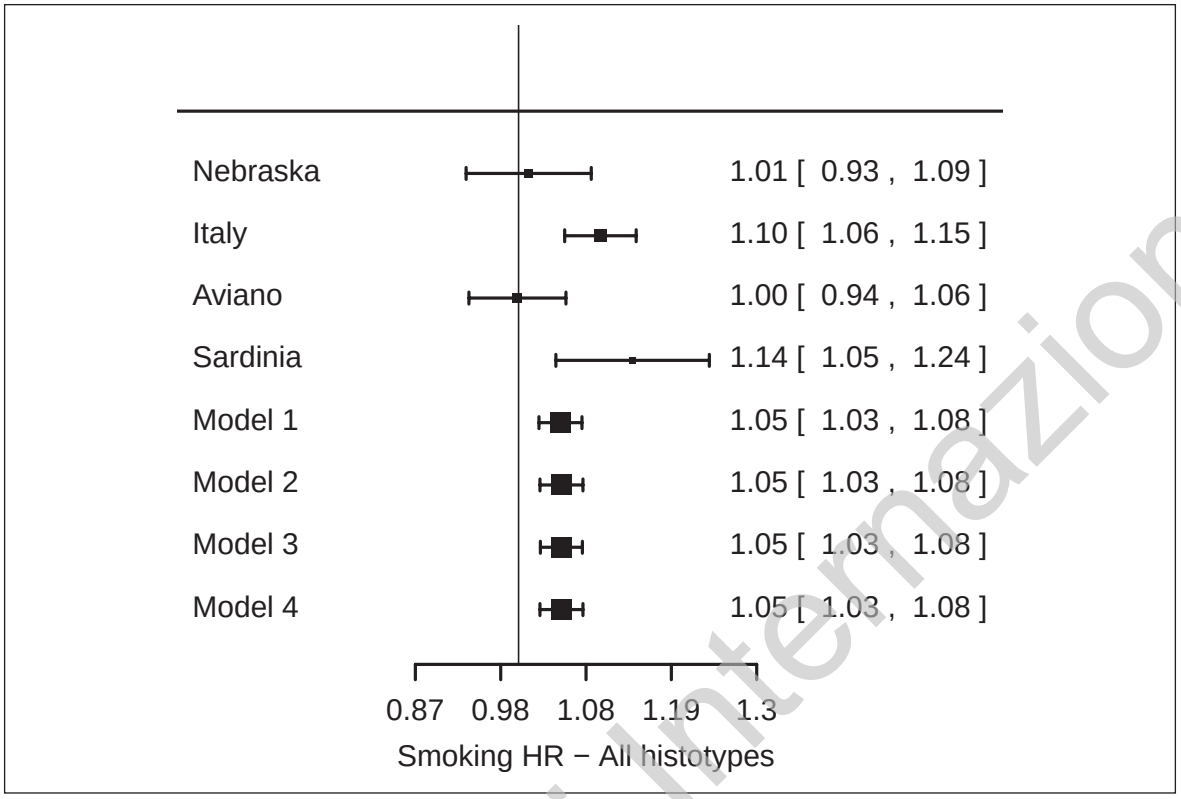

Figure 3. Forest plot for the smoking HR obtained by applying model (0) to each study and applying the four models to the pooled data.

three random models examined in this work differed in the choice of the parameters allowed to vary randomly with studies. In the shared frailty model (model 2) variability between studies is captured in the baseline hazard function by including in the model a random parameter explicitly accounting for the baseline between-study variability. In this case the baseline parameter varies randomly among studies but no variability is attributed to the covariates.

The covariate random effect Cox model (model 3) is adopted when the variability between studies arise from the investigated effect itself, i.e. the effect under study varies randomly among studies. In this model, the lone tobacco smoking effect was assumed to vary randomly, consequently all the heterogeneity was assumed to be held in the pack-year variable.

The model 4 is assumed when the variability between studies arises in both covariate and baseline parameters: this model should capture the great part of heterogeneity.

Apart from modelling techniques, the key issue of a meta-analytical analysis is to assess heterogeneity. This consists in identifying an eventual heterogeneity and, in the IPD analysis, in finding variables explaining such heterogeneity. Larger is the amount of the explored and explained heterogeneity and larger is the relevance and the scientific value of the meta-analytical analysis (11).

The stratified Cox model (fixed effect model 1) allows estimating the pooled covariate effect taking into account study-specific baseline hazards. The pooled tobacco smoking HR from model 1 is not dissimilar respect to the ones from the random effect models (figure 3). However, model 1 does not allow measuring the heterogeneity, which is a major limitation for meta-analysis. Fixed effect models may therefore be used once the absence of heterogeneity has been proved.

The use of random effect models is frequently aimed to the only heterogeneity analysis. This because a gain of efficiency of random effect models respect to the stratified models has been proved to be largest when moderate or large numbers of small groups of two or three patients, as in mul-

ticenter trials, were used in the analysis (24). In addition, in many applications the estimates of the effects under study seemed not to vary significantly by using fixed or random models.

Results from the application of the meta-analytical models to our NHL dataset showed a better performance of the random effect models respect to the fixed effect models (table 2). In addition the comparison of the random effect models with the corresponding fixed effect ones by log-LRTs (table 3) highlighted that both the models (models 2 and 4) with the random effect related to the intercept (the baseline) better performed than the corresponding fixed effect models. The model 2 (shared-frailty) was preferable since it gained in the number of degree of freedom respect to model 4. These findings suggested that definitely, heterogeneity occurred in our NHL data and that basically, this heterogeneity was captured when the baseline hazard function was allowed to vary randomly with studies. This heterogeneity, not explained by covariates, is named residual heterogeneity and should always be explored (9, 25).

The main interest when exploring heterogeneity in meta-analytical survival analyses is to assess heterogeneity in the investigated effects across studies, rather than just heterogeneity in baseline functions. Katsahian et al. (11) asserts that models that do not incorporate heterogeneity in the investigated effects should only be used in case of substantial evidence of homogeneity of the effects between studies. In fact, Katsahian et al. (11) demonstrated that models without random covariate effect perform poorly in case of heterogeneity in the investigated effects; on the other hand, models with random covariate effect, respect to model without, performed worse in case of absence of heterogeneity in the investigated effects. Our findings of a not considerable difference in the performance of the model 2 compared to the model 4, confirm Katsanian et al.'s (11) assertions.

Anyway a contrasting behavior in the two different histotype analyses was observed considering model 3. When the all histotypes analysis was performed, the model with the random tobacco smoking effect did not better perform than the corresponding fixed effect one, suggesting a homogeneous tobacco smoking effect on NHL survival among studies. When the only DLBCL cases were considered, model 3 showed a slightly better performance respect to the corresponding fixed effect model (table 3: LRT p=0.03). The same results were observed by examining the values of the tobacco smoking random effect variances (t_1^2) that measured the amount of investigated heterogeneity (table 4). Presumably, our results indicated that the investigated heterogeneity was not attributable to a different study-related effect of tobacco smoking on NHL survival; rather it seemed to be linkable to the differences among study-populations. This residual heterogeneity might be further explained by adding new factors to the survival analysis. Anyway, a minor amount of heterogeneity due to a study-different tobacco smoking prognostic value appeared when a subgroup of DLBCL cases was isolated. This could be attributed to a well-known homogeneity in survival trends for sub-groups of the same histological diagnosis. The different survival trends by histotypes could have influenced the heterogeneity assessment in the all histotype analysis.

In all the models of this work the age and sex covariates were always included. This is because a recent study suggested that individual patient covariates could influence overall results in case of not large datasets (12). The latter evaluated the effect of including patient-level covariates on heterogeneity in a cancer survival meta-analysis of about 11000 subjects. The authors found a limited influence on the estimates of the investigated effect but a reduction in the confidence intervals, explained by the large dataset leading to balanced covariates between studies.

Covariate-interactions were not considered in this work since we did not evidence heterogeneity in the covariate effect.

Finally, a concept to be mentioned is checking of model assumptions that is an important issue with the use of proportional hazard models (11). As an example, in this application an additional model with the study as a covariate was initially considered: since the proportional hazards assumption was not satisfied, this model was excluded from analyses. To be noticed that methods for checking PH assumption are not yet implemented for random effect models in all statistical packages. The NHL application afforded the original goal to investigate and apply alternative models to un-

dertake a meta-analysis of time to event IPD. Moreover the availability of patient level information provided a straightforward possibility for exploring heterogeneity. We assessed the effect of tobacco smoking on survival from NHL, however in this work we did not try to draw conclusions on survival hazards. The aim was not to investigate the effect of prognostic factors on NHL survival, neither to explain heterogeneity since no standardized information was available for other confounding variables or known prognostic factors. The availability of multiple covariates as well as the evaluation of covariate-interactions could partially explain heterogeneity. As Tudor Smith and Williamson (25) assert, one major advantage of IPD is the ability to investigate potential causes of heterogeneity by exploring covariate-interactions. When more studies will be available and more variable will be collected, a further analysis will be addressed to this topic.

In conclusion the comparison of meta-analytical models and the implementation of random effect models, that are not commonly applied to IPD turn out to be simply implemented, even if on a small number of pooled studies. Moreover the random effects models were found to perform better than fixed effect ones, and the component found to capture the larger amount of between-study variability in our application was the baseline hazard.

Acknowledgments

We would like to thank Luigino dal Maso (Centro di Riferimento Oncologico, Aviano, Italy), Pierluigi Cocco (Department of Public Health, Occupational Health Section University of Cagliari, Italy), Brian Chiu (Department of Preventive Medicine, Northwestern University, Feinberg School of Medicine, Chicago, Illinois, USA), Paolo Vineis (Division of Epidemiology, Public Health and Primary Care, Imperial College, London, UK) and Adele Seniori Costantini (Cancer Prevention and Research Institute – ISPO, Florence, Italy) for making their individual patient data available. We are grateful to Adele Seniori Costantini for their helpful comments on the earlier versions of the manuscript.

References

1. Simmonds MC, Higgins JPT, Stewart LA, Tierney JF, Clarke MJ, Thompson SJ. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. Clin Trials 2005; 2: 209-217.
2. Egger M, Davey Smith G, Schneider M, Minder C. Meta-analysis; principle and procedures. BMJ 1997; 315: 1533-1537.
3. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986; 7: 177-188.
4. DerSimonian R, Raghu K. Random-effects model for meta-analysis of clinical trials: an update. Contemp Clin Trials 2007; 28:105-114.
5. Whitehead A, Whitehead J. A general parametric approach to meta-analysis of clinical trials. Stat Med 1991; 10: 1664-1677.
6. Tudor Smith C, Williamson PR, Marson AG. Investigating heterogeneity in an individual patient data meta-analysis of time-to-event outcomes. Stat Med 2005; 24: 1307–1319.
7. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. Eval Health Prof 2002; 25: 76–97.
8. Smith-Warner SA, Spiegelman D, Ritz J, Albanes D, Beeson WL, Bernstein L, Berrino F, van den Brandt PA, Buring JE, Cho E, Colditz GA, Folsom AR, Freudenheim JL, Giovannucci E, Goldbohm RA, Graham S, Harnack L, Horn-Ross PL, Krogh V, Leitzmann MF, McCullough ML, Miller AB, Rodriguez C, Rohan TE, Schatzkin A, Shore R, Virtanen M, Willett WC, Wolk A, Zeleniuch-Jacquotte A, Zhang SM, Hunter DJAm J Epidemiol. Methods for pooling results of epidemiologic studies: the Pooling Project of Prospective Studies of Diet and Cancer. Am J Epidemiol 2006; 163: 1053-64.
9. Normand SLT. Tutorial in biostatistics meta-analysis: formulating, evaluating, combining and reporting. Stat Med 1999; 18: 321-359.
10. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of models. Stat Med 1999; 18: 2693-2708.
11. Katsahian S, Latouche A, Mary J, Chevret S, PorcherR, Practical methodology of meta-analysis of individual patient data using a survival outcome. Contemp Clin Trials 2008; 29: 220–230.
12. Michiels S, Baujat B, Mahé C, Sargent DJ, Pignon JP. Random effects survival models gave a better understanding of heterogeneity in individual patient data meta-analyses. J Clin Epidemiol 2005; 58: 238–45.
13. Sargent DJ. A General Framework for Random Effects Survival Analysis in the Cox Proportional Hazards Setting. Biometrics. 1998; 54(4):1486-1497
14. Clayton DG. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika. 1978; 65:141-151.
15. Hougaard P. A class of multivariate failure time distributions. Biometrika. 1986; 73: 671-678.

16. Hougaard P. Survival models for heterogeneous populations derived from stable distributions. Biometrika. 1986; 73:671-678.
17. Aalen OO. Modelling Heterogeneity in Survival Analysis by the Compound Poisson Distribution. Ann Appl Probab 1992; 4: 951-972.
18. McGilchrist CA, Aisbett CW. Regression with Frailty in Survival Analysis. Biometrics. 1991; 47: 461-466.
19. Therneau TM, Grambsch PM, Pankratz VS. Penalized Survival Models and Frailty. J Comput Graph Stat 2003; 12(1):156- 175
20. Casella G, Berger RL. Statistical Inference, Academic Internet Publishers Incorporated 2006, Second edition
21. Venables WN, Smith DM, the R Development Core Team. An introduction to R. 2009 http://cran.r-project.org/doc/manuals/R-intro.pdf
22. Battaglioli T, Gorini G, Costantini AS, Crosignani P, Miligi L, Nanni O, Stagnaro E, Tumino R, Vineis P. Cigarette smoking and alcohol consumption as determinants of survival in non-Hodgkin's lymphoma: a population-based study. Ann Oncol 2006; 17: 1283-1289.
23. Morton LM, Hartge P, Holford TR, Holly EA, Chiu BCH, Vineis P, Stagnaro W, Willett E, Franceschi S, La Vecchia C, Hughes AM, Cozen W, David S, Severson RK, Bernstein L, Mayne ST, Dee FR, Cerhan JR, Jheng T. Cigarette smoking and risk of non-Hodgkin lymphoma: a pooled analysis from the InterLymph consortium. Cancer Epidemiol Biomarkers Prev 2005; 14: 925–933.
24. O'Quigley J, Stare J. Proportional hazards models with frailties and random effects. Stat Med 2002; 21:3219–3233.
25. Tudor Smith C, Williamson PR. A comparison of methods for fixed effects meta-analysis of individual patient data with time to event outcomes. Clin Trials 2007; 4:621–630