Original Article

Effect of Comparison Group on Inference about Effect Modification by Demographic Factors in Cohort Risk Regression

John Cologne*1, Shizue Izumi*1, Yukiko Shimizu*2, and Dale Preston*1
Departments of *1Statistics, and *2Epidemiology, Radiation Effects Research Foundation
5-2 Hijiyama Park, Minami-ku, Hiroshima 732-0815
E-mail: cologne@rerf.or.jp

Epidemiologic cohort studies frequently make use of a comparison group to infer what background rate of death or disease might have occurred in the exposed cohort in the absence of exposure. Unlike cohort analyses utilizing the standardized risk ratio, regression analyses of risk can often be performed without the need for a comparison group, avoiding possible bias in the risk estimate. Demographic factors related to the background rate may also modify the risk (effect modification). Including the comparison group can improve the precision of effect-modification parameter estimates, but if there is inadequate adjustment for heterogeneity between exposed and comparison groups in the background effects of these factors, the effect-modification parameter estimates can be biased. We studied this bias and the precision of effect-modification parameter estimates using theory and simulation. The problem is illustrated using data from studies of radiation exposure of atomic-bomb survivors that include a comparison group selected from distal geographic areas having different gender-specific rates of death. We conclude that, for studies of effect modification in cohorts covering a wide range of exposures including doses close to zero, there may be no advantage to including a comparison group, as long as internal standardization is feasible.

Key words: standardized rate ratio; internal standardization; bias; background rate

1. Introduction

Epidemiologic cohort studies conducted to assess the risk of an exposure are often based on comparing mortality or disease incidence rates in the exposed persons with those from an unexposed comparison group or from a reference population (Breslow et al., 1983). A point of concern is whether the comparison and exposed groups are comparable in terms of the background rate (Yanagawa and Hoel, 1993). When exposure is measured on a continuous scale and covers a wide range of doses, including doses close to zero, it may be possible to estimate the dose response using internal standardization without a zero-dose comparison group (Cologne and

Preston, 2001). In this case, the background (zero-dose) disease or mortality rate is estimated by the intercept of the dose response.

The exposed sub-cohort may be small, or studies may be conducted on small subsets of the exposed cohort, such as in second-stage studies (Zhao and Lipsitz, 1992). In such studies it may be useful to utilize the comparison group for precise estimation of demographic effects in the background rates because imprecisely estimated background rates can affect the power and precision of the risk estimate. When effect modification is not considered, unadjusted heterogeneity in the effects of demographic factors on the background rate should not impact relative risk estimation if the overall background mortality rate estimate is unbiased. For example, Cologne and Preston (2001) found no bias in an internally standardized relative-risk regression estimate when only a standardized rate ratio (SRR) parameter (Breslow et al., 1983) was used to adjust for heterogeneity between comparison-group and exposed-cohort background mortality rates. However, the SRR parameter alone does not adequately describe the background-rate heterogeneity in terms of demographic factors. Thus, even though the SRR parameter removes overall intercept-related bias in the dose-response, there may remain unadjusted demographic factor-specific differences between the true background rate for the exposed cohort and the fitted rate based on including the comparison group. Failure to correctly adjust for these differences might bias inference about effect modification by these same factors, but adding further parameters to the background part of the model to adjust such differences may decrease precision of inference about exposure risk and its effect modification.

There seems, therefore, to be the potential for a bias-variance trade-off when considering whether to include a comparison group in an analysis of effect modification. We consider the impact on inference about effect modification of including the comparison group with adjustment for background-rate heterogeneity between the comparison group and exposed cohort. We begin by describing the problem from a simple theoretical point of view. We then present real-data analyses based on the study of radiation risk and mortality in atomic-bomb survivors. Finally, we investigate the practical impact of sample size via simulation.

2. Theoretical considerations

Our goal is to make inference about effect modification by examining statistical interaction in a risk-regression analysis. Interaction means modification of the effect measure, not necessarily of the effect itself (Rothman, 2002, pages 169–176). In the case of radiation risk assessment, it is often not known whether the effect of exposure acts additively or multiplicatively on background rates of disease or death when risk estimates are transported to other populations. Therefore, it is necessary to assess possible interaction on both scales even if there is no evidence of biological interaction.

In the exposed cohort, there will be a background rate of death or disease, $B$, which may
depend on many demographic factors (age, calendar time, gender, geographical location, etc.). For simplicity, let there be one demographic factor, gender (coded $s$ = 1 if female, 0 if male) and a dichotomous exposure (coded $x$ = 1 if exposed, 0 if not). We begin by considering three common classes of risk regression model: excess relative risk (ERR), relative risk (RR), and excess absolute rate (EAR) models:

- **ERR model**
  \[
  \lambda(E, x) = B e^{\alpha_E s} \left( 1 + E x e^{\gamma_E s} \right)
  \]  
  \[(1a)\]

- **RR model**
  \[
  \lambda_R(s, x) = B e^{\alpha_R s + x \log(R) + \gamma_R x}
  \]  
  \[(1b)\]

- **EAR model**
  \[
  \lambda_A(s, x) = B e^{\alpha_A s} + A x e^{\gamma_A s}
  \]  
  \[(1c)\]

where $B$ is the background rate, $E$ is the excess relative risk, $R = E + 1$ is the relative risk, and $A = B \times E$ is the excess rate, all quantities pertinent to males. Assume that the observed number of deaths $D_{jk}$ for gender $j$ ($j = 1, 2$) and exposure group $k$ ($k = 1, 2$) follows a Poisson distribution having mean $N_{jk} \lambda(s_j, x_k)$ with $N_{jk}$ person years of observation. To focus on the parameters for gender, assume that $B$, $R$, $E$, and $A$ are all known.

2.1 **Bias in effect modification**

The interest here is in how a non-comparable comparison group affects inference about the effect-modification parameter, $\gamma$. Assume that the effect of gender on rate of death or disease differs between the comparison group and the exposed cohort, so that adding the comparison group to the analysis without adjusting for the heterogeneity in rate leads to a biased estimate of the background gender effect; call the biased parameter $\alpha^*$ and call the resulting effect-modification parameter $\gamma^*$. Bias in the excess relative-risk model can be examined by writing (1a) as

\[
\lambda_E(s, x) = B e^{\alpha_E^* s} \left( 1 + E x e^{\gamma_E^* s} \right),
\]

where there is bias $\zeta_E = \alpha_E^* - \alpha_E$ in the background gender effect and it is desired to determine the resulting relationship between $\gamma_E^*$ and the true effect-modification parameter $\gamma_E$. If $\alpha_E$ and $\alpha_E^*$ are known,

\[
\left. \frac{\partial \ell(\alpha_E, \gamma_E)}{\partial \gamma_E} \right|_{\gamma_E^*} = \left. \frac{\partial \ell(\alpha_E^*, \gamma_E^*)}{\partial \gamma_E^*} \right|_{\gamma_E^*} = 0,
\]  
\[(2)\]

where $\ell(\alpha_E, \gamma_E) = \sum_{j,k} \left( D_{jk} \log \left( N_{jk} \lambda(s_j, x_k) \right) - N_{jk} \lambda(s_j, x_k) \right)$ is the log likelihood and $\gamma_E$ and $\gamma_E^*$ are the maximum likelihood estimates for the corresponding models. The log likelihood has derivatives

\[
\left( \frac{\partial \ell}{\partial \alpha_E} \right) = \sum_j s_j \left[ D_{j+} - N_{j+} B e^{\gamma_E s_j} \left( 1 + R \frac{N_{j2}}{N_{j+}} e^{\gamma_E s_j} \right) \right],
\]

\[(3)\]

\[
\left( \frac{\partial \ell}{\partial \gamma_E} \right) = R \sum_j s_j e^{\gamma_E s_j} \left[ \frac{D_{j2}}{1 + Re^{\gamma_E s_j}} - N_{j2} B e^{\gamma_E s_j} \right]
\]
where the subscript "+" denotes summation over the corresponding index. By substituting the derivatives with respect to \( \gamma_E \) into (2) and substituting the expectation of \( D \) with respect to the true parameters \((\alpha_E, \gamma_E)\), we obtain

\[
\frac{1 + E \gamma_E^\alpha}{1 + E \gamma_E^\gamma} = \frac{e^{\alpha \gamma_E^\gamma}}{e^{\alpha E}},
\]

which shows that the ratio of biased to true risk functions is inversely equal to the corresponding ratio in the background rates. If \( E \gamma_E \) and \( \gamma_E \) are both close to zero, \( \gamma_E^\gamma - \gamma_E \approx -(\alpha_E - \alpha E)/E \)

is an approximate measure of the bias in \( \gamma_E^\gamma \). The bias in the effect modification parameter is therefore approximately equal to the opposite of the bias in the background parameter weighted by the magnitude of the excess relative risk \( E \). In other words, if the excess relative risk \( E \) is small, bias resulting from unadjusted heterogeneity between the comparison group and exposed cohort can induce bias of much larger magnitude in the effect-modification parameter estimate.

Bias in the relative-risk model (1b) can be derived by writing

\[
\lambda_R(s, x) = B e^{\alpha R^s + x \log(R) + \gamma_R^s x},
\]

from which we obtain

\[
\gamma_R^* - \gamma_R = - (\alpha_R^* - \alpha_R).
\]

Thus, the two biases are of the same magnitude, though opposite sign, regardless of the magnitude of the exposure effect \( R \). The difference between this and the excess relative-risk model is that, with the relative-risk model, the effect-modification term directly multiplies the background rate, so the effect-modification parameter can directly offset the background bias, whereas in the excess relative-risk model the effect-modification term only directly multiplies the excess relative risk, so the amount of change in the effect-modification parameter needed to overcome bias in the background term is proportional to the size of the risk and so can be large even if the background bias is small.

For the excess absolute rate model (1c), the bias is

\[
e^{\gamma_R^\alpha} - e^{\gamma_R^\alpha} = - \frac{B}{A} \left[ e^{\alpha_R^s} - e^{\alpha s} \right].
\]

In this case the effect modifier term is biased to the extent of the bias in the background rate divided by the excess rate—essentially the same result as with the excess relative-risk model, although with the different parameterization it applies to the difference between, rather than ratio of, biased and unbiased rates.

To verify the theoretical results, the bias was examined by simulation using S-plus (Mathsoft, Inc., Seattle, Washington) with 5,000 replications for each of seven values of background gender bias \( \gamma_E \) evenly spaced between \(-0.15 \) and \(0.15\). We used only the relative-risk model (1b) because it is the simplest to fit in S-plus and the relationship between \( \alpha^\gamma \) and \( \gamma^\gamma \) is conceptually the same.
with all three models. The results are shown in Fig. 1a and confirm that the effect modification parameter exactly offsets unadjusted bias in the background rate. The result did not change when the number of person years or the magnitude of exposure relative risk were varied (not shown).

2.2 Precision of effect modification

Regardless of the model, the asymptotic covariance matrix of the parameter estimates $\alpha$ and $\gamma$ is the inverse information matrix

$$i(\alpha, \gamma)^{-1} = \begin{pmatrix} i_{\alpha\alpha} & i_{\alpha\gamma} \\ i_{\gamma\alpha} & i_{\gamma\gamma} \end{pmatrix} = \begin{pmatrix} i^\alpha & i^\gamma \\ i^\gamma & i^{\gamma\gamma} \end{pmatrix}.$$  

The Fisher information for $\gamma$ is $(i^{\gamma\gamma})^{-1} = [i_{\gamma\gamma} - i_{\gamma\alpha}(i_{\alpha\alpha})^{-1}i_{\alpha\gamma}]$, not $i_{\gamma\gamma}$ [for example, McCullagh and Nelder (1989), page 472]. From (3) it is apparent that for the excess relative-risk model the value $i_{\gamma\gamma} = \text{E}((\theta \gamma)^2)$ does not depend on the comparison group data $\{D_{ij}, N_{ij}\}$ but $(i^{\gamma\gamma})^{-1}$ does. Therefore, the comparison group adds to the information about $\gamma_E$ in the excess relative-risk model, reducing its asymptotic variance. Similar results hold for the relative-risk and excess-rate models.

The standard error of the effect modification parameter in the relative-risk model (1b) was also examined via simulation with 5,000 replications. Fig. 1b shows the effect of reducing the number of person years contributed by the comparison subjects relative to a fixed number of exposed subjects' person years when there is no bias in $\alpha$. The simulations confirm the theoretical
result that the comparison subjects contribute to the precision of the effect-modification estimate.

In conclusion, a comparison group whose rate of death or disease is the same as the background rate in the exposed cohort can enhance precision of the effect-modification estimate, but if the comparison group is not directly comparable and the resulting background-rate heterogeneity is not adequately adjusted, there will be bias in the effect-modification parameter estimate.

3. Example from the atomic-bomb survivor cohort

The preceding theory suggests that there is a potential for bias-variance trade-off in effect-modification analyses when a comparison group is included. The bias can be controlled by adjusting for heterogeneity in background rates. If there is such heterogeneity, and if it is adjusted, requiring the introduction of further parameters, is there still any advantage in terms of precision? We investigated this using the following real-data example.

In the Life Span Study cohort of atomic-bomb survivors (Pierce et al., 1996; Shimizu et al., 1999), all persons who received significant radiation exposures were within 3 km of the hypocenters at the time of the bombings, but the comparison group was constructed using persons who were up to 10 km from the hypocenters. Subsequent all-cause mortality in this cohort differed significantly between proximal (≤ 3 km) and distal (> 3 km) zero-dose comparison groups (Fig. 2). Because some studies of atomic-bomb survivors have used geographic location at the time of the bombing to define exposed and comparison groups (e.g., Mine et al., 1990), Cologne and Preston

![Graph showing excess relative mortality](image)

**Fig. 2.** Excess relative risk of death by radiation dose or distance from the hypocenter (adjusted for city, gender, age, age at start of follow-up, and calendar year). The baseline is the average of the two comparison groups (0–10 km combined). Adapted from Cologne and Preston (2000) and Cologne et al. (2001).
(2001) evaluated the effect of using either the proximal or distal zero-dose group alone as the comparison group in a regression analysis. They found that the choice of comparison group had a significant impact on the estimated slope and shape of the dose response.

The relative risk of mortality for radiation in the atomic-bomb survivors is modified by gender and age at exposure. Because age at exposure is equivalent to birth year (the entire cohort was exposed in 1945), which is equivalent to age and calendar year, various important radiation effect modifiers are equivalent to demographic factors used in modeling the background mortality rate. We therefore examined inference about effect modification using various degrees of comparison-group bias adjustment to determine what impact background-rate heterogeneity adjustment had on the bias and precision of the effect-modification parameter estimate. Risk models were fit by Poisson regression based on finely stratified tables of deaths and person years, using the Epicure software (Hirosoft International Corporation, Seattle, Washington). Details of the analyses are described briefly in the Appendix and in more detail elsewhere (Cologne and Preston, 2001). The three previously described types of model (excess relative risk, relative risk, and excess absolute rate) were considered.

Results for each of the three models are shown in Table 1 according to several degrees of adjustment for background-rate heterogeneity, which were used to assess the impact of heterogeneity adjustment on precision of the effect-modification parameter estimates. Complete adjustment includes SRR parameters and interactions between comparison-group indicator variables and all other background variables in a log-linear background model (including separate gender-by-age and city-by-year regression spline functions by group; see Appendix); in other words, the entire background-rate model was replicated with separate parameters for each group (for a richly parameterized background-rate model, this is almost equivalent to fitting the individual background person-year strata, tantamount to excluding the comparison group). SRR plus group interactions with main effects means that SRRs and interactions between the comparison-group indicators and the other background covariates city, gender, age, and year (but not the regression spline functions) were included in the model; adjustment for heterogeneity with this model is moderate but not complete. SRR only implies that only an indicator for each comparison group was included, resulting in a single SRR parameter for each comparison group; this does not adjust for heterogeneity in demographic effects on the background rate, but only corrects for bias in the intercept of the non-effect-modified dose response. Although not adjustments per se, two additional approaches were examined: either excluding the comparison group (exclusion), where the background rate was estimated internally using the exposed sub-cohort only, or including the comparison group with no adjustment (none), which assumes that the background rates for the exposed sub-cohort and the two comparison groups are the same.

For each degree of adjustment, the table shows the effect-modification parameter estimate and its standard error and mean square error, the likelihood-ratio test of effect modification by

Jpn J Biomet Vol. 23, No. 2, 2002
Table 1. Effect modification in the full cohort

<table>
<thead>
<tr>
<th>Degree of adjustment for background-rate heterogeneity (# of adjustment parameters)(^a)</th>
<th>Relative risk model</th>
<th>Excess relative risk model</th>
<th>Excess absolute rate model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate for gender(^b) (SE)</td>
<td>MSE (\times 10^{-4})</td>
<td>LR statistic</td>
</tr>
<tr>
<td>Exclusion of all comparison-group subjects-background rate estimated from exposed cohort (0)</td>
<td>.0593 (.0249)</td>
<td>6.19</td>
<td>29.02</td>
</tr>
<tr>
<td>Complete adjustment—a separate, full background-rate model for each group (40)</td>
<td>.0593 (.0249)</td>
<td>6.19</td>
<td>29.04</td>
</tr>
<tr>
<td>Moderate adjustment—a separate SRR for each group plus group interactions with main effects (18)</td>
<td>.0583 (.0248)</td>
<td>6.17</td>
<td>28.26</td>
</tr>
<tr>
<td>SRR only—an individual SRR only for each group; otherwise, a common background model (2)</td>
<td>.0704 (.0237)</td>
<td>6.83</td>
<td>32.65</td>
</tr>
<tr>
<td>None-comparison group included, but no adjustment for background-rate heterogeneity (0)</td>
<td>.0707 (.0237)</td>
<td>6.91</td>
<td>33.13</td>
</tr>
</tbody>
</table>

\(^a\) Number of adjustment parameters in addition to 25 background main-effect, interaction, and spline parameters.

\(^b\) Parameter estimates are shown only for the effect modifier gender; results for the age-at-exposure effect modifier were similar.

\(^c\) Calculated assuming that the parameter estimated with exclusion is unbiased.

\(^d\) For the test of joint effect modification by gender and age at exposure (2 degrees of freedom).

\(^e\) Degrees of freedom for the excess absolute rate model are one less because the excess-rate term includes age.
gender and age at exposure jointly, the goodness of fit, and the number of parameters required to make the adjustment. For simplicity, parameter estimates are shown only for the effect modifier gender; results for the age-at-exposure effect modifier were similar. Effect modification was insignificant on the absolute-rate scale and the relative-risk model did not fit as well as the excess relative-risk model. The results with complete adjustment for background-rate heterogeneity are nearly identical to those excluding both zero-dose comparison groups, as would be expected because the rich background-rate model should very nearly fit all of the comparison-group person-year strata exactly. Without complete adjustment, bias appeared in the effect-modification parameter estimates, but standard errors and likelihood-ratio tests gave the appearance of greater power with SRR only or no adjustment. The bias predominated when there was no adjustment, so that the MSE increased substantially. Therefore, in terms of making inference about effect modification there was no advantage to including the comparison group.

4. Simulations using randomly sampled sub-cohorts

Table 1 suggests that the excess relative risk model fits slightly better than the relative risk model, but a consideration of parsimony would lead to choosing the excess absolute rate model, which requires no effect modification. However, other considerations in risk assessment—such as transporting risk estimates to different populations—require that both the relative and absolute models be used, because it may not be known which is the appropriate scale of risk for a particular population. Therefore, it is important to study further the impact of the comparison group on inference about effect modification in the excess relative risk model, particularly in regards to the size of the cohort (both exposed and comparison subgroups).

The sub-cohort of atomic-bomb survivors who received significant radiation doses (whole-body doses of 0.005 Sievert or greater) is large enough to accurately estimate demographic effects on background mortality rates without using the comparison groups. However, in studies utilizing only subsets of the cohort—such as the Adult Health Study subset of the Life Span Study (Yamada et al., 1996)—or, more generally, in smaller cohort studies, the comparison group may be needed for estimating demographic patterns in the background rate even if all available exposed persons are included. We therefore evaluated the effect of cohort size by simulation using randomly selected samples of the atomic-bomb survivor exposed and comparison sub-cohorts.

We assessed the impact of comparison-group size on the bias and precision of effect modification parameter estimates for gender and age at exposure and the power of the likelihood ratio test for their joint significance. We compared these quantities with the comparison group excluded to those with randomly sampled comparison groups of various sizes: 12.5%, 25%, or 50%, or using all (total 34,064) of the comparison sub-cohort. Background-rate heterogeneity was adjusted to each of the degrees shown in Table 1 (complete adjustment, moderate adjustment using SRRs plus group interactions with main effects, or SRRs only). To assess the effect of exposed-cohort size, we randomly sampled various numbers of exposed persons: 1.0%, 2.5%, Jpn J Biomet. Vol. 23, No. 2, 2002
5%, 10% or 50% of the exposed sub-cohort (total 52,508) and compared the resulting bias and precision to that based on the full cohort. Only the excess relative-risk model (1a) was evaluated via simulation because, as seen in Table 1, there was no effect modification of excess rates and the log-linear relative risk model did not fit as well. In total 500 simulations were performed for each combination of exposed- and comparison-group size using the Epicure software.

Because effect modification as parameterized in the excess relative risk model requires that the sign of the dose response be the same for both genders, the fitted model failed to converge more frequently as the size of the sampled exposed sub-cohort got smaller because by chance the sign of the estimated dose-response parameter was opposite between the two genders and the overall dose-response parameter estimate was close to 0. The proportion of times the models failed to converge was 0.63 (1% of exposed sub-cohort), 0.43 (2.5%), 0.23 (5%), 0.09 (10%), and 0.002 (50%). This lack of convergence was not affected by sampling independently within dose strata because it was related to the chance of sampling few deaths at the rare high doses rather than under-sampling of high doses. Small samples were therefore associated on average with closer-to-zero estimates of the effect-modification parameters even when all comparison subjects were excluded, showing that the estimation of effect modification becomes increasingly more difficult the smaller the number of exposed individuals. Therefore the effect-modification parameter bias due to inclusion of the comparison group is presented relative to the parameter value obtained with no comparison group separately for each exposed-group size. Results were similar for the gender and age-at-exposure parameters, so we present results for gender only.

4.1 Bias

Fig. 3 shows the average bias in gender effect modification parameter estimates, relative to the unbiased (controls excluded) values. Overall amount of bias decreased with greater degree of adjustment for background-rate heterogeneity. With SRR adjustment only, inclusion of any proportion of the comparison group led to bias in the effect-modification parameter estimates. The bias was larger with larger numbers of comparison subjects and with smaller numbers of exposed persons. A similar pattern was observed with the SRR plus group main-effect interaction background-rate adjustment, but the level of bias was less than with SRRs only. With complete adjustment, the effect of which is close to that of excluding all controls, there was very little bias except with the smallest (1% and 2.5%) sizes of exposed sub-cohorts. Thus, with very small exposed cohorts (on the order of a thousand persons or less), there can be a pronounced impact of background-rate heterogeneity on bias in the effect-modification parameter estimate even if the heterogeneity is carefully adjusted.

4.2 Precision

Fig. 4 shows the average standard errors of the gender effect-modification parameter estimates. Precision was strongly related to size of the exposed group, as would be expected because information on effect modification derives predominantly from the exposed persons. The smaller
Bias in Effect Modifier Gender

**SRR-only Adjustment**

- Relative bias
- Size of comparison group:
  - 0 (0%)
  - 4,258 (12.5%)
  - 8,516 (25%)
  - 17,032 (50%)
  - 34,084 (100%)

**Main-effects Adjustment**

- Relative bias
- Size of comparison group:
  - 0 (0%)
  - 4,258 (12.5%)
  - 8,516 (25%)
  - 17,032 (50%)
  - 34,084 (100%)

**Complete Adjustment**

- Relative bias
- Size of comparison group:
  - 0 (0%)
  - 4,258 (12.5%)
  - 8,516 (25%)
  - 17,032 (50%)
  - 34,084 (100%)

*Fig. 3.* Bias in the effect-modification parameter estimate for gender, relative to that with all controls excluded, using the three background adjustment strategies shown in Table 1. Curves represent different sizes of randomly sampled exposed subcohorts: 52,508 (100%) [---]; 26,254 (50%) [★]; 5,251 (10%) [★]; 2,625 (5%) [★]; 1,313 (2.5%) [△]; and 525 (1%) [○]. The dotted line represents no relative bias. Each point represents the average from 500 simulations for each combination of exposed and comparison sub-cohort sizes.
Standard Error of Effect Modifier Gender

**SRR-Only Adjustment**

![SRR-Only Adjustment Graph]

**Main-effects Adjustment**

![Main-effects Adjustment Graph]

**Complete Adjustment**

![Complete Adjustment Graph]

**Fig. 4.** Standard error of the effect modifier parameter estimate for gender, using the three background adjustment strategies shown in Table 1 (see Fig. 3 legend for explanation of symbols).

the number of exposed subjects, the greater was the improvement in precision when comparison subjects were included and adjustment was by SRR only or by SRR plus group main-effect interactions, but the improvement was less noticeable with the greater adjustment. Furthermore,
improvement in precision occurred only up to 12.5% of the comparison group (about 4,000 persons) and did not improve with a greater number of comparison subjects regardless of how small the exposed group was. The improvement in precision with addition of a comparison group was less dramatic than in the case of the simple theoretical simulations (Fig. 1b), presumably because of the richer background-rate model and corresponding large number of parameters even without complete adjustment for heterogeneity. With complete adjustment, there was no improvement in standard error with the inclusion of comparison subjects.

4.3 Power

Fig. 5 shows the proportion of times the two degree-of-freedom likelihood ratio test was rejected (i.e., the value of the LR test statistic was greater than 5.991). It is well known that power to detect interaction is generally lower than that for main effects, so it is not surprising to see that the power fell to below 0.5 with 10% or less of the exposed sub-group. This may be partly due to skewness in radiation doses of the cohort—there are very few persons with large doses. In actual studies based on samples from the cohort, it would be important to over-sample the persons with large doses through stratified sampling, matching or counter-matching, etc. However, our interest here is on the impact of the comparison group rather than size of the exposed group; we are interested in what effect the comparison group would have on a study with an exposed group of various sizes (i.e., assuming all available exposed persons are included). There was little affect of comparison group size, except for a slight increase in proportion of LR test rejections with the inclusion of comparison subjects when adjustment was by SRR only.

5. Conclusions

We can write the relative risk and excess relative risk models generally as

\[
\begin{pmatrix}
\text{total rate} \\
\text{rate}
\end{pmatrix} = \begin{pmatrix}
\text{background rate} \\
\text{rate}
\end{pmatrix} \times \begin{pmatrix}
\text{relative risk}
\end{pmatrix}
\]

(although this ignores the different ways in which they parameterize effect modification) and the excess absolute-rate model as

\[
\begin{pmatrix}
\text{total rate} \\
\text{rate}
\end{pmatrix} = \begin{pmatrix}
\text{background rate} \\
\text{rate}
\end{pmatrix} + \begin{pmatrix}
\text{excess rate}
\end{pmatrix}
\]

Because the total observed and expected numbers of events (deaths or cases of disease) are equal in a maximum likelihood analysis, bias in the background rate estimate will be compensated by an inverse and equal amount of bias in the exposure-risk portion of the model. Failure to adjust for heterogeneity in the effect of demographic factors causes bias in the background-rate estimate, which can lead to biased inference about effect modification—a spillover effect—by these factors in the risk part of the model, even if no such effect modification exists. This spillover effect is multiplicative with the excess relative-risk model \([B \times (1 + E)]\) or the relative-risk model.
Power of LR Test of Effect Modification (2 d.f.)

Fig. 5. Power of the two degree-of-freedom likelihood ratio test for effect modification by gender and age at exposure, using the three background adjustment strategies shown in Table 1 (see Fig. 3 legend for explanation of symbols).

$(B \times R)$, and it is additive with the excess rate model $(B + A)$. Even if overall heterogeneity in rates among groups is adjusted, such as with SRR parameters (e.g., Yanagawa and Hoel, 1993; Breslow et al., 1983), bias in effect modification can still occur. SRR adjustment alone should be
sufficient to prevent bias in the exposure dose response as long as demographic effect modifiers are not included in the analysis model (Cologne and Preston, 2001), but one must be careful to ensure that the background effects of individual demographic variables are properly adjusted before attempting to make inference about effect modification by those factors.

Theoretical results show that the comparison group contributes to the precision of effect-modification parameter estimates. This was also seen in the real-data simulations as precision and power improved with increasing number of controls using SRR adjustment only, but the effect was smaller with a greater degree of adjustment for background-rate heterogeneity. The lack of improvement in precision when background-rate heterogeneity was completely adjusted is presumably due to the large number of parameters needed for adjustment. On the other hand, the apparent better precision and power with SRR-only adjustment is achieved at the price of biased effect-modification parameter estimates. Thus, even with small numbers of exposed persons, the potential gain in precision of the effect-modification parameter estimates does not justify including the comparison group unless there is no heterogeneity. It is difficult to imagine, though, how one could be certain of no heterogeneity in small studies.

The previous argument leads us to conclude that the use of a comparison group is not recommended in cohort studies of effect modification of any size when internal standardization is possible. However, it leaves open the question of when internal standardization is feasible. Certainly the range and proximity to zero of doses is important, but that alone is insufficient because imprecisely estimated background rates will affect the power and precision of the risk estimate. Therefore, if it is not clear that the internally standardized analysis is more accurate, both analyses (with and without inclusion of the comparison group) should be performed and compared.

Generally, though, we believe that studies of large exposed cohorts should not be based on possibly inappropriate comparison groups when dose estimates are available and internal standardization can be performed. Greenland and Morgenstern wrote "It is well known that regional differences in background rates can produce large amounts of ecological bias ... The key to understanding the connection between ecological bias, confounding, and effect modification is to evaluate group as a confounder and an effect modifier at the individual level" (Greenland and Morgenstern, 1989). Thus, comparing the rates of different groups using geographical location as a surrogate for exposure should be avoided and studies based on such groups should be analyzed carefully to ensure proper adjustment of demographic factors related to heterogeneity in background rates.

Acknowledgements

We thank Sachiko Teranishi for technical assistance. This publication is based partly on research performed at the Radiation Effects Research Foundation (RERF), Hiroshima and Nagasaki, Japan, a private nonprofit foundation funded equally by the Japanese Ministry of Health, Jpn J Biomet Vol. 23, No. 2, 2002
Labour and Welfare and the US Department of Energy through the National Academy of Sciences. The first two authors also acknowledge the support of Grant-in-Aid No. 14580356 from the Japanese Ministry of Education, Culture, Sports, Science, and Technology.

REFERENCES


Appendix

We fit the following three models—excess relative risk (ERR), relative risk (RR), and excess absolute rate (EAR)—to atomic-bomb survivor data on age and time of death using Poisson regression (Breslow and Day, 1987) with person-year tables stratified by city, gender, age, calendar year, and radiation dose:

\[
ERR \quad \lambda(c,s,a,y,f,d) = \lambda_0(c,s,a,y,f)(1 + \beta_E d e^{\gamma_1 s + \gamma_2 g}) \\
RR \quad \lambda(c,s,a,y,f,d) = \lambda_0(c,s,a,y,f)e^{\beta_R d + \gamma_1 s + \gamma_2 g} \\
EAR \quad \lambda(c,s,a,y,f,d) = \lambda_0(c,s,a,y,f) + \beta_A d \lambda_E(s,a,g),
\]

where the background mortality rate \( \lambda_0(\cdot) \) is a log-linear function of binary indicator variables representing city \( (c) \) and gender \( (s) \) and continuous variables representing age \( (a) \) and calendar year \( (y) \)—see equation (A2) below. Age at start of follow-up \( (f) \) was also included because entry into follow-up was staggered over several years. The parameter \( \beta \) is the risk per unit dose \( [d, \text{in Gray}], \) adjusted for measurement error (Pierce et al., 1992)—excess relative rate of mortality \( (\beta_E) \), logarithm of the relative rate of mortality \( (\beta_R) \), or absolute excess rate of mortality \( (\beta_A) \). The risk was allowed to include effect modification by gender and age at exposure \( (g = 1945 - [y - a]) \).

To obtain cohort average estimates of background mortality rates, the fixed, binary variables \( s \) and \( c \) were centered at their unweighted means (e.g., \( s = \{0 \text{ if male, 1 if female} \} \) — proportion of cohort that is female) and the continuous time variables were centered at integer values closest to their person-year weighted means. The effects of age and time were modeled using quadratic regression splines (Harrell et al., 1988) with a gender-specific spline in age \( \{\sigma_{\text{age}}(a,s)\} \) and a city-specific spline in year \( \{\sigma_{\text{year}}(y,c)\} \). The regression splines were of the form

\[
\sigma_{\text{time}}(t,v) = \phi_1 v t + \phi_2 v t^2 + \sum_{l=1}^L \eta_l v (t - \tau_l)_+^2
\]

Jpn J Biomet Vol. 23, No. 2, 2002
where “time” \( (t) \) is either age or year, \( v \) is either city or gender, \( L \) is the number of knots, \( \tau_i \) is the value of the \( i^{th} \) knot, and the function \( (z)_+ = \max\{z, 0\} \). The knots were \{40, 50, and 75\} for age and \{1910, 1920, and 1930\} for year.

SRR parameters \{\psi_i\} were estimated using indicators \{I_i\} of comparison group \((i = 1, 2)\) to allow for differences between either comparison group and the exposed group. Interactions between the comparison-group indicators and other background variables were included to allow a greater degree of adjustment for heterogeneity in the background rates. Thus, the completely adjusted background mortality-rate model was:

\[
\lambda_0(c, s, a, y, f, i) = \exp \left\{ \varphi_0 + \psi_i I_{\{\text{group}=i\}} + \alpha_0 s + \eta_0 c + \omega_0 f + (\alpha_i s + \eta_i c + \omega_i f) I_{\{\text{group}=i\}} + \sigma_{0,\text{age}}(a, s) + \sigma_{0,\text{year}}(y, c) + [\sigma_{i,\text{age}}(a, s) + \sigma_{i,\text{year}}(y, c)] I_{\{\text{group}=i\}} \right\},
\]

(A2)

so that \( \varphi_0 \) is the estimated cohort-average background rate for the exposed group.