Reanalysis of the occurrence of back pain among construction workers: modelling for the interdependent effects of heavy physical work, earlier back accidents, and aging

Markku Nurminen

Abstract
Objectives—To re-examine the relation between heavy physical work and the occurrence of sciatic pain among construction workers reported previously to be absent in an epidemiological study. Methods—Poisson log linear regression was used to model the frequency of sciatic pain among concrete reinforcement workers and maintenance house painters with adjustment for the interactive effects of earlier back accidents and aging that modified the relation. Results—Concrete reinforcement work not only had a direct effect on the frequency of sciatic pain, but it also contributed significantly to the risk indirectly through earlier back accidents. The risk of sciatic pain increased from age 25 to 54 in a different manner for a worker depending on his occupational group and record of back accidents. Conclusions—Epidemiological studies on low back pain need to be analysed with sound methodology. This is important in view of future meta-analyses that will be performed for the purpose of providing guidelines on the prevention of back disorders in heavy physical work.

Keywords: back pain research; heavy physical work; epidemiological methods

Previously, Riihimäki1 reported an investigation of the effect of heavy physical work on back pain in a cohort of concrete reinforcement workers and a control group of maintenance house painters. Subjects were interviewed about back symptoms during the 12 months preceding the examination and data on occupational history and earlier back accidents were recorded with a questionnaire. In a logistic regression analysis, the occurrence of sciatic pain was assessed in relation to current occupation, earlier back accidents, and the age of the workers. Among the main conclusions of the study were the following:1 “According to this analysis, age and back accidents had an independent effect on the occurrence of sciatic pain. The rate of sciatic pain was not affected by occupation after age and back accidents were accounted for.” Other studies have suggested that occupational physical stress is linked with low back syndromes,2,3 but the causal connection hypothesis is still debatable because of conflicting results.4 As it is, the role of factors related to work both in the development of pathological changes and in the production of pain remains poorly understood.5 The analysis of epidemiological data demands sound methodology, and the conclusions reached from the analysis must be based on impeccable interpretation of results. This is important both from the occupational and public health points of view because available scientific evidence is increasingly being reviewed for the purpose of providing guidelines on the prevention of back disorders and the management of low back pain. Accordingly, a reanalysis was carried out to test the proposition—the null hypothesis accepted by Riihimäki—that heavy physical work has no effect on the occurrence of sciatic pain when the effects of age and back accidents have been adjusted for.

Materials and methods
The groups that participated in the study comprised 217 occupationally active male concrete reinforcement workers aged 25-54 and a control group of 202 maintenance house painters. The cohorts were selected from the register of the Helsinki regional section of the Construction Workers Union. To control for possible confounding effects by differences in the age distributions of the compared groups, frequency matching was performed by five-year age strata. By design, the study was retrospective: data on occupational history, occurrence of back accidents, or recurring pain before 1977, etc, were obtained by questionnaire. The answers were checked by a physiotherapist before conducting a standardised interview on back symptoms.

The reanalysis presented in this paper differs from Riihimäki’s1 analysis in four respects. Firstly, I used a Poisson log linear model for frequencies in a contingency table, instead of a logistic regression used by Riihimäki for proportions of workers with sciatic pain. Secondly, I represented the categorical variable age in the log linear model in terms of second order orthonormal polynomials that are scaled by their norms. This procedure resulted in a more parsimonious model than a model in which age is treated as a set of age category
indicators. The orthogonality of the polynomials also reduced the collinearity between age and its interaction with factor variates. Thus this technique enhanced the lack of bias of the estimates of the coefficients and the precision of their SEMs. Thirdly, I added a three factor interaction term involving occupation, age, and sciatic pain to the model to represent possible causal connections between these factors. Fourthly, I fitted separate regressions of sciatic pain on age within the classes defined by back accidents with two intercepts and two slope parameters.

The definition and estimation of interactions are problematic in epidemiological studies. A "Statistical interaction" is defined as the coefficients of product terms in the model. This means that the presence or absence of interaction is entirely dependent on the algebraic form of the statistical model that one chooses. For the same data interaction may seem to be present in a multiplicative model but absent in an additive model. Interactions can take the form of change in shape of the relation—for example, linear for some subgroups of workers, quadratic for others—so the product terms that I considered in the search for the best fitting model were not simple ones (see later). With multiple related product terms, I used a pooled $\chi^2$ test for interaction for the assessment of effect modification. Finally, the term interaction can also mean causal interaction or interdependence of coaction, but I do not make the distinction here.

Results

Before embarking on modelling, I first performed stratified analyses to validate the model choices. Table 1 shows the reported frequencies of sciatic pain when the data were also cross classified by back accidents and age. To study whether occupation bears an independent relation to sciatic pain without the influence of back accidents and aging, I carried out a stratified analysis by occurrence of accidents and age group. The risk of sciatic pain among reinforcement workers was 20% higher than the risk among painters (risk ratio 1.2, and its 95% confidence interval (95% CI) 0.9 to 1.6, estimated by the method of Miettinen and Nurminen), but the increase was not significant ($P=0.20$). At any rate, a regression model with the basic variates that define the contingency table should roughly reproduce the result of the stratified analysis.

The application of the logistic regression (with the GLIM program) for the proportion of sciatic pain (based on table 1), yielded the following analytical result: The two term additive model including factors AGE (age) and BACC (back accidents) gave a good fit on the data; the inclusion of factor OCC (occupation) into this model did not improve the fit significantly. The product terms (interaction terms) of the factors also proved to be non-significant. I assessed the probability of significance associated with the occupation factor by comparing the two models that differed from each other only by the OCC term; the deviance $\chi^2$ test gave $P=0.17$. Thus the modelling approach and stratified analysis yielded qualitatively the same result. I next considered whether it was possible to improve the model for an in depth analysis of the occurrence relation.

The logistic parameterisation of the adopted model for the effects of the risk factors age and back accidents was

$$
\log(OR) = \beta_1 (AGE) + \beta_2 (BACC) (model 1)
$$

where OR is the odds ratio—that is, the ratio of the odds of sciatic pain among reinforcement workers versus the odds of pain among painters, and the $\beta$s are the regression coefficients to be estimated. If the proportion is low (say, $<10\%$) at all factor levels, the logistic linear model for OR approximates well to the log linear model or the equivalent multiplicative model for the risk ratio (RR):

$$
RR = \exp(\beta_1 (AGE)) \times \exp(\beta_2 (BACC)) (model 2)
$$

In models 1 and 2, the effect of aging—that is, a change from one age category to another—is measured by the amount that such a change multiplies the OR or RR. If this multiplicative scale is used to measure effects, then independence of the factors AGE and BACC implies that the effects of one factor are homogeneous across the categories of the other. However, when changing, for example, from the youngest age group (25–29) to the next one the magnitude of RR was multiplied by 2.9 among workers with no back accidents, whereas among workers with back accidents the respective change was 6.3-fold. From age 30–54 the risk of sciatic pain increased significantly among workers with no back accidents (Cochran-Armitage test for linear trend, $P=0.018$), whereas among workers with back accidents the risk stayed at the same (higher) level. These inconsistencies suggest that the effects of the joint action of age and back accidents were interdependent, not independent. If the effects of the two factors were independent (in the sense that the factors did not alter each other's biological action on the risk of sciatic pain), an additional third parameter would be needed to represent this discrepancy from the logistic model 1.

Riihimäki used the term additive somewhat confusingly. In the theory of general linear models (GLM), linearity is assessed with
Table 2  Analysis of the frequency data of table 1 assuming a Poisson log linear model

<table>
<thead>
<tr>
<th>Model term</th>
<th>β coefficient</th>
<th>SE</th>
<th>t value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.07</td>
<td>0.26</td>
<td>4.11</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>OCC</td>
<td>-0.01</td>
<td>0.30</td>
<td>-0.03</td>
<td>0.97</td>
</tr>
<tr>
<td>BACC</td>
<td>1.37</td>
<td>0.27</td>
<td>5.09</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>AGE4</td>
<td>9.48</td>
<td>4.03</td>
<td>2.35</td>
<td>0.027</td>
</tr>
<tr>
<td>AGE2</td>
<td>1.56</td>
<td>1.65</td>
<td>0.94</td>
<td>0.36</td>
</tr>
<tr>
<td>SCI</td>
<td>1.09</td>
<td>0.27</td>
<td>4.01</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>OCC x BACC</td>
<td>-0.92</td>
<td>0.22</td>
<td>-4.07</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>OCC x AGE</td>
<td>2.21</td>
<td>2.15</td>
<td>1.03</td>
<td>0.31</td>
</tr>
<tr>
<td>OCC x AGE2</td>
<td>-2.64</td>
<td>1.87</td>
<td>-1.41</td>
<td>0.17</td>
</tr>
<tr>
<td>BACC x AGE</td>
<td>0.17</td>
<td>2.53</td>
<td>0.07</td>
<td>0.95</td>
</tr>
<tr>
<td>BACC x AGE2</td>
<td>-2.74</td>
<td>1.48</td>
<td>-1.85</td>
<td>0.076</td>
</tr>
<tr>
<td>OCC x SCI</td>
<td>0.61</td>
<td>0.29</td>
<td>2.08</td>
<td>0.048</td>
</tr>
<tr>
<td>BACC x SCI</td>
<td>-1.14</td>
<td>0.26</td>
<td>-4.32</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>AGE x SCI</td>
<td>-11.1</td>
<td>3.94</td>
<td>-2.81</td>
<td>0.010</td>
</tr>
<tr>
<td>AGE2 x SCI</td>
<td>-2.85</td>
<td>1.49</td>
<td>-1.91</td>
<td>0.068</td>
</tr>
<tr>
<td>OCC x BACC x AGE</td>
<td>1.23</td>
<td>1.70</td>
<td>0.73</td>
<td>0.48</td>
</tr>
<tr>
<td>OCC x BACC x AGE2</td>
<td>-2.36</td>
<td>1.61</td>
<td>-1.46</td>
<td>0.16</td>
</tr>
<tr>
<td>OCC x AGE x SCI</td>
<td>-0.85</td>
<td>2.33</td>
<td>-0.36</td>
<td>0.72</td>
</tr>
<tr>
<td>OCC x AGE2 x SCI</td>
<td>4.99</td>
<td>2.05</td>
<td>2.43</td>
<td>0.023</td>
</tr>
<tr>
<td>BACC x AGE x SCI</td>
<td>-2.85</td>
<td>1.49</td>
<td>-1.91</td>
<td>0.068</td>
</tr>
<tr>
<td>Without back accidents (intercept)</td>
<td>9.44</td>
<td>4.90</td>
<td>-1.93</td>
<td>0.055</td>
</tr>
<tr>
<td>With back accidents (intercept)</td>
<td>8.01</td>
<td>4.20</td>
<td>-1.91</td>
<td>0.068</td>
</tr>
<tr>
<td>Without back accidents (slope)</td>
<td>7.29</td>
<td>4.79</td>
<td>1.52</td>
<td>0.14</td>
</tr>
<tr>
<td>With back accidents (slope)</td>
<td>11.3</td>
<td>4.64</td>
<td>2.43</td>
<td>0.023</td>
</tr>
</tbody>
</table>

SCI = Sciatic pain (no / yes); OCC = occupation (concrete reinforcement worker / maintenance house painter); BACC = back accidents (no / yes); AGE, AGE2 = age of worker, represented by orthogonal second-degree polynomials.

Coding of factors: each coefficient represents a comparison of that factor level with level 1 omitting level 1 itself.

Model given by the S-PLUS system code: GLM(formula = frequency ~ OCC + BACC + poly(AGE, 2) + OCC:BACC + BACC:poly(AGE, 2) + OCC:BACC:poly(AGE, 2) + OCC:poly(AGE, 2):SCI + BACC:poly(AGE, 2):SCI + poly(AGE, 2):SCI, family = Poisson, link = log, in which I is an indicator of the age range 30–54 (that is, I=1 in that age category and I=0 otherwise).†Pooled χ² test for interaction: 10.6 on 5 degrees of freedom, P=0.032; age range: 30–54.

regard to the regression parameters β, and additivity refers to the influence that the investigated factors have on the chosen effect parameter. A basic parameterisation of the additive model for the coaction of age and back accidents in terms of RR is 11:

\[
RR = 1 + \beta_1 (\text{AGE}) + \beta_2 (\text{BACC}) \quad (\text{model } 3)
\]

In this additive model, the effect of changing from one age category to another is measured by the amount such a change adds to the RR. Model 3 infers that the additive effect is the same for workers with and without back accidents.

More importantly, the quoted results conflict with the stated conclusion of the study—namely, that “The accident rate of the reinforcement workers was higher than that of the painters, and this difference seemed to explain their higher rate of sciatic pain.”10 To show that the higher risk of sciatic pain among reinforcement workers was caused by their higher risk of back accidents, it would be necessary to verify by a statistical analysis that there is an association between (a) occupation and back accidents and (b) back accidents and sciatic pain. Although the data did not formally contradict the model reached by Riihimäki1 (a χ² goodness of fit statistic, deviance = 18.3, with the number of degrees of freedom, df = 17), model 1 embodies no terms to build such a causal connection.

A limitation of the binomial logistic regression is that it cannot be used to explicitly model for the association between the explanatory variates OCC and BACC that is possible within the log linear framework. Therefore, we used an alternative Poisson log linear model for the analysis of the relations in the multivariate contingency table 1. A graphical inspection of the data also showed that the frequency of sciatic pain depicted non-linear age relations in the occupational groups. I thus improved the model parsimony by the inclusion of two polynomial terms to account for the effect modification by age rather than having five categorical variates for the six age groups.

The final model comprising terms for the main effects and the interaction of all study factors (occupation, accidents, and age) as well as additional terms for their modifying effects on sciatic pain fitted the data of table 1 very well (deviance = 16.9, df = 24). The analysis of residuals indicated adequate compatibility with the model assumptions. Table 2 summarises the numerical results (obtained with the S-PLUS system11). The interpretations of this reanalysis are as follows: (a) to reflect the matched design, there was no significant association between occupation and age; (b) heavy physical work, earlier back accidents, and aging were significant risk factors of sciatic pain (P=0.048, P<0.001, and P=0.010, respectively), the main effect of back accidents being much stronger than that of heavy work; (c) occupation was strongly associated with back accidents (P<0.001) and thereby may also have caused more sciatic pain among reinforcement workers than among painters; (d) the risk of sciatic pain increased steadily with age among reinforcement workers throughout the age interval from 25 to 54 years, whereas among painters the increase was sharp from the youngest age group to the 30–34 year old category whereupon the risk stayed at the same level (P=0.023 for the non-linear modification of the age relation by occupation); and (e) there was a significant overall interaction between back accidents, aging, and sciatic pain (P=0.032), and specifically, from the age of 30 onwards the risk of sciatic pain for those workers with back accidents was consistently higher (slope significantly different from zero, P=0.023) than for those workers without back accidents. In short, heavy work, back accidents, and aging showed significant interdependent effects on the risk of sciatic pain.

Discussion

Although a logistic model is always equivalent to a hierarchical log linear model (that includes all main and interaction effects of the explanatory variates) in the sense that the parameter estimates of the logistic regression are exactly expressible by their log linear counterparts.13 However, the models have a different perspective for interpretation. A logistic model has fewer parameters than a corresponding log linear model, and therefore logistic regression is simpler to estimate, particularly in small samples. In principle, at least, it can be questioned whether it is appropriate to treat the 24 stratum total numbers (n TOT, table 4, Riihimäki) as fixed denominators of the proportions of workers with sciatic pain. A Poisson model for multinomial frequency data would seem to correspond better with the non-experimental epidemiological study design, in which only the two total numbers of the occupational groups were non-random. A distinct
difference between the two models is that the significance of, for example, the OCC $\times$ BACC term, can explicitly be determined from a log linear model but not from the logistic regression model. Riihimäki tested this association with a stratified analysis, because model I involves the term only implicitly.

Age is often controlled for in the analysis in too crude a fashion, leaving open the possibility of residual confounding. Despite the successfully age matched design, according to Riihimäki "Age was included in the analysis to control its confounding effect". Rather than to control for confounding bias, age was specified as a categorical variable in the logistic model to allow for possible effect modification. Categorical analysis for modelling exposure-disease relations has recently been criticised.

A category indicator analysis of the data in table 1 in the log linear framework would result in an overfit model. The formulation of age as a second order polynomial resulted in a parsimonious model. I selected the polynomial form of the continuous variable to represent the effects of aging based on a visual inspection of the data. The orthogonal transformation of the age variable removed serious numerical problems that can arise in the design matrix. Graphical examination of the final model and the goodness of fit statistic showed that residual confounding by unaccounted covariates was unlikely. But, it should be borne in mind that the degree of confounding—and hence the strength of an effect—cannot be directly measured from the observed data. Confounding should always be evaluated against background risk of disease, knowledge of subject matter, logical argument, and evidence from previous studies.

Although log linear modelling for contingency tables with one dimension representing the outcome variable is a flexible tool, even simple data sets like the reanalysed one lead to quite complex problems of interpretation. The specification of a more complex model—for example, path analytical; one that would take into account the causal order of the studied variables—would be highly uncertain for these data. The reason for this is that the study was retrospective by design, and the questionnaire information was crude. For example, the exact time of the occurrence of back accidents earlier than the year immediately preceding the examination was not recorded or at least was not used in the original analysis. Moreover, no questions were asked on the type or severity of the accidents; this is an important omission as proneness to accidents is different at different ages, and age modified the risk relations. The outcome variable considered here was the occurrence of recurring sciatic pain during the previous year. However, the pain could have been the recurrence of back symptoms before entry into the present occupation. Fewer reinforcement workers than painters had back symptoms before entering their present profession. Physically heavy work can render the back liable to injury. Moreover, repeated injuries can lead to degenerative changes in the lumbar spine even in early adulthood and force a worker to change his trade. Thus it is possible for a health based selection bias both when entering and leaving reinforcement work. Together, the causal web that links degenerative back symptoms, back injuries, and occupational factors is very difficult to study. Nevertheless, the log-linear model explicitly embodied the significant terms OCC $\times$ BACC and BACC $\times$ SCI. But to prove that there is a causal linkage OCC $\rightarrow$ BACC $\rightarrow$ SCI requires arguments that are beyond statistical inference. For example, the link BACC $\rightarrow$ SCI is biologically credible because, as noted by Riihimäki, "There is a natural immediate association between back accidents and back symptoms; to be recalled, back accidents must have caused pain." Riihimäki discussed validity and issues of subject matter supporting the study hypothesis or detracting from it. My reanalysis intended merely to show that such a causal link is statistically nonexistent.

The associations between back pain, heavy physical work, and earlier back accidents may involve complex causal pathways where primary causes and secondary causes are difficult to separate in an epidemiological setting. The power to detect statistical interaction is typically an order of magnitude less than the power to detect main effects. Because of limitations of sample size and because of distortions produced by measurement error in questionnaire and interview data, it is necessary to be cautious about the potential of occupational epidemiology for assessment of interaction. A recent study alleviated some of these problems. The prospective design restricted the subjects to those without a history of recent or notable previous back problems, and thus simplified the analysis. The findings of this study that heavy occupational musculoskeletal loading and physical demands predicted future back pain agree with the results of this reanalysis.

Epilogue
My motive for undertaking this reappraisal of Riihimäki’s study was to clarify the role that heavy physical work has in the production of back pain. The previous interpretation of results destroyed the tentative evidence that heavy work has a direct effect on the risk of sciatic pain. The original analysis also failed to verify that work may contribute to the risk indirectly through the link with back accidents. Epidemiological data are laborious, tawdry, and expensive to come by, and they deserve a painstaking analysis and inference to render correct results for further evaluation and combination of data in a critical meta-analysis. Only valid study results provide a sound basis for developing recommendations on the prevention of back disorders and impaired performance in heavy physical work.

I thank the reviewers for their helpful comments on this paper.

Reanalysis of the occurrence of back pain among construction workers

17 Brenner H, Blettner M. Controlling for continuous confounders in epidemiologic research. Epidemiology 1997; 8:429-34.

Rejected manuscripts

From February 1994, authors whose submitted articles are rejected will be advised of the decision and one copy of the article, together with any reviewer’s comments, will be returned to them. The Journal will destroy remaining copies of the article but correspondence and reviewers’ comments will be kept.
Reanalysis of the occurrence of back pain among construction workers: modelling for the interdependent effects of heavy physical work, earlier back accidents, and aging.

M Nurminen

*Occup Environ Med* 1997 54: 807-811
doi: 10.1136/oem.54.11.807

Updated information and services can be found at:
http://oem.bmj.com/content/54/11/807

**Email alerting service**

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Notes**

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/