METHODOLOGY

Why and how to control for age in occupational epidemiology

Dario Consonni, Pier Alberto Bertazzi, Carlo Zocchetti

Abstract

In occupational epidemiology, the need to consider the age factor properly influences the choice of study design and analytical techniques. In most studies, age is viewed as a potential confounder. Age is strongly associated with end points of interest in occupational epidemiology (diseases, physiological characteristics, doses of xenobiotics, etc), but to measure age as a confounder it must be associated with the exposure under study. When the exposure of interest is time related—for example, duration of employment, time since first exposure, cumulative exposure—a strong intrinsic association with age can be anticipated, and age will behave as a (usually strong) confounder. When occupational exposures without a direct relation with age—for example, job, department, type of exposure—are evaluated, the degree and direction of confounding bias cannot be anticipated. Control of the confounding effect of age can be accomplished in the design phase of a study by way of randomisation, restriction, and matching. Randomisation is seldom viable in occupational settings. Restriction is rarely used in the case of age. Matching is often used in a case-control study as a method to increase the study efficiency, but it must be followed by proper matched or stratified analysis. Options for age adjustment in the analysis phase involve stratification and regression methods. In longitudinal studies the modified life table analysis is used to take into account the fact that subjects cross categories of age as the study proceeds. Stability of relative measures of effect over age strata favoured the greater use of relative risks than risk differences. In the presence of effect modification the influence of age should not be eliminated; its interaction with exposure should be explicitly considered.

(*Occup Environ Med* 1997;54:772–776)

Keywords: age; confounding bias; occupational epidemiology

Istituti Clinici di Perfezionamento, Clinica del Lavoro "L Devoto", Milan, Italy
D Consonni
C Zocchetti

EPOCA Epidemiology Research Centre, Institute of Occupational Health, University of Milan, Italy
P A Bertazzi

Correspondence to:
Dr Pier Alberto Bertazzi, Institute of Occupational Health, University of Milan, Via San Barnaba 8, 20122 Milano, Italy.
Tel: 0039 2 5450812; Fax 0039 2 5512344; Email pieralberto.bertazzi@unimi.it

Accepted 11 August 1997

Age is among the most relevant variables affecting the health experience of working populations.[1] When the possible association between a given exposure and a specified health outcome is evaluated, age may in various ways affect this evaluation. We consider here, in particular, the issue of age as a potential confounding factor—that is, an extraneous factor capable of distorting the estimates of the effect of the exposure of interest—the condition most commonly encountered in occupational epidemiology.

Age as a confounding factor

For a variable to be a confounding factor in an epidemiological study, three criteria must simultaneously be satisfied:[2–5] (*a*) it must be a determinant (risk factor) of the disease, even in the absence of the exposure of interest; (*b*) it must be associated with the exposure of interest in the study base, from which the cases derive; (*c*) it should not be an intermediate step in the pathway from cause to effect.

The criterion (*c*) has, in principle, no relevance in the case of age.

The criterion (*a*), on the contrary, is usually satisfied because age is a recognised determinant for many disorders or conditions of interest in occupational epidemiology. The incidence of many chronic diseases—for example, neoplastic, cardiovascular—strongly increases with age. Physiological characteristics—for example, pulmonary function, blood pressure—often investigated in occupational epidemiology—also recognise age as a determinant. An association with age is often present even when the outcome of interest is the internal dose of a xenobiotic, particularly in the case of substances which accumulate in the body.

The criterion (*b*) could be satisfied or not in occupational epidemiology. In practice, two situations are encountered, one which involves time related variables as disease determinants, the other in which the relevant variables bear no direct relation with time. Many time related variables are of key interest in occupational epidemiology.[5 6] For example, analysis of disease frequency by time since first employment (or since first exposure to a specific agent) is performed for chronic disorders that need long latency periods to become apparent; duration of employment (or of exposure to a specific

Table 1 All cause mortality, 1949–91, by duration of employment and time since first employment in a cohort of workers of an Italian oil refinery

	Years of employment			*Years since first employment*			
	<5	*5–14*	*≥15*	*<9*	*10–9*	*20–9*	*≥30*
Proportion of person-years <50 (%)	86	70	33	93	73	43	1
Crude RR*	1.0†	2.7	3.8	1.0†	3.3	6.4	12.3
Age adjusted RR*	1.0†	1.5	1.1	1.0†	1.5	1.2	1.1

*RRs calculated with Poisson regression.
†Reference category.

Table 2 All cause mortality, 1949–91, by department in a cohort of workers of an Italian oil refinery

	Department			
	A	*B*	*C*	*D*
Proportion of person-years <50 (%)	63	65	71	46
Crude RR	1.0†	1.0	0.5	1.4
Age adjusted RR	1.0†	0.9	0.7	0.9

*RRs calculated with Poisson regression.
†Reference category.

substance) is a crude, but usually accurately measured, index of exposure; cumulative exposure combines intensity and duration of exposure, and the duration component often predominates.[7] Age is strongly associated with all of these variables. Hence, when workers are categorised and compared on the basis of the value of time related covariates such as duration of employment, latency, or cumulative exposure, it would be expected that the resulting groups also differ in terms of age composition. Workers with long duration, latency, and higher cumulative exposure will generally be older than those with lower values of these variables, and will have higher morbidity or mortality for the only reason of being older. In these situations, in which age has a strong association with disease, and a strong association with the exposure, it is initially expected that age is a confounder that will bias the effect estimates.

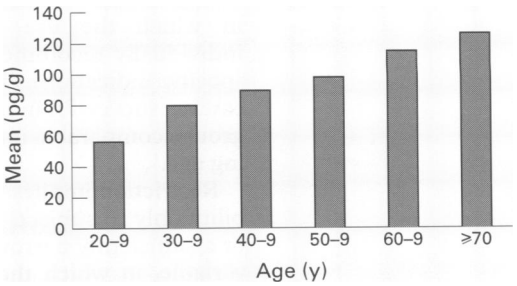
Both previous knowledge and study data can help evaluation if a variable can act as a confounder in a particular study. In practice, although some authors argue that the assessment of confounding should not be based exclusively on the data in hand,[8] the presence and the degree of confounding in a study is usually evaluated by contrasting crude and adjusted estimates of the effect measures.[3–5]

Table 1 shows the degree of confounding bias resulting in analysing time related variables without taking age into account, the crude and age adjusted relative risks of overall mortality by duration of employment, and latency in an Italian cohort of oil refinery workers,[9] followed up over 40 years. The proportion of person-years with age<50 is used as a convenient, although rough, indicator of the different age distribution of workers in different categories. Although no clear patterns of risk are found when age is taken into account, strong, yet spurious increasing trends are found in the unadjusted analyses, the reasons being that the workers in the longer exposure categories are necessarily, on the average, older.

In other situations the association of age with the exposure of interest can only be established retrospectively. For example, when examining the frequency of disease in groups categorised by non-time related variables—for example, factory, department, job, type of exposure—a confounding effect of age may or may not be present, depending on the age composition of the study and control groups. In this case the direction of the bias is unpredictable: the crude effect estimate for the exposure could be unbiased, if the age distribution is similar in the groups, or biased in any direction with respect to the null (no effect) value. As an example consider again the study of oil refinery workers. The cohort was subdivided in exposure categories corresponding to departments, according to the longest job held. Table 2 shows the results for some of these departments and shows the different directions of the confounding bias. The age composition of departments A and B is similar, and the crude mortalities are 9.8 and 9.4 $\times 10^{-3}$ person-years, respectively; people in department C are moderately younger (mortality 5.2 $\times 10^{-3}$); whereas workers in department D are substantially older (mortality 14.1 $\times 10^{-3}$). Taking, arbitrarily, the department A as reference, the crude and adjusted relative risks for B workers remain almost the same; the crude relative risk for C workers seems to be biased downward (the adjusted estimate is greater than the crude estimate); for workers in department D, the crude estimate was obviously biased upward and the adjusted relative risk was actually found to be slightly negative.

The striking difference of the magnitude of bias involved in analysing time related and non-time related variables can be appreciated from the comparison of tables 1 and 2. The crude RR is about 3.5 times the adjusted RR in the highest category of duration of employment, and over 10 times in the longest latency category. In table 2 the differences, although important, are definitely smaller.

The figure shows an example of age as a potential confounder in studying the dose of xenobiotics, in which the mean plasma concentration (pg/g blood lipids) of one PCB congener measured in a sample of the population living in an industrialised area in Italy is stratified by age category (MT Landi, L Needham, personal communication). It is obvious how the dose varies with age. The comparison of these levels with those of another population requires similarity of age distributions.



Plasma concentrations in blood (PCB) of people (n=121) living in an industrial area near Milan, Italy.

Table 3 Lung cancer risk and exposure to pesticides, 1965–81, in a cohort of United States licensed pest control workers

	Exposure			
	A	*B*	*C*	*D*
Crude OR*	3.3	2.4	3.0	1.5
Age adjusted OR†	3.2	2.2	3.0	1.4

*Exposed *v* non-exposed.
†ORs calculated with Mantel-Haenszel method.
Dead controls.

Otherwise, the age difference could bias the estimate of the exposure difference.

If age is associated with the disease but not simultaneously with the exposure as well, there is no need, in principle, to adjust for it. In a study of a cohort of licensed pest control operators in Florida, for example, a nested case-control design was implemented to thoroughly investigate individual exposure to pesticides and to other factors (other occupational exposures, tobacco smoking, etc).[10] Lung cancer deaths were selected as cases. Two groups of controls (one of living and one of deceased cohort members) were matched to cases for a few variables, including age. Table 3 shows the crude and age adjusted odds ratios (OR) of lung cancer for different exposure variables. It is obvious that the two sets of estimates are not substantially different. Thus, age, which is a definite risk factor for lung cancer, was not associated with exposure in this study base, and hence it was not acting as a confounder in this particular study.

How to control confounding

Adequate control for age as a confounding factor can be accomplished at the design or the analysis phase of a study.

DESIGN OPTIONS

Options for the control of age at the design phase of a study include randomisation, restriction, and matching.[3 11]

Randomisation involves the random assignment of subjects to exposure groups; the main advantage of the method is that it is capable (in sufficiently large studies) of ensuring comparability of exposure groups for several potential confounders (known and unknown). Randomisation is seldom viable in occupational studies, one possible application is in the evaluation of preventive measures.[12] For example, in a study of the efficacy of barrier creams in preventing skin complaints, workers at a fabric dyeing and printing factory were randomly allocated to two treatment groups (one in which the use of barrier creams was individually recommended, the other with no specific indication). The randomisation process was successful in making the two treatment groups comparable for several variables including age.[13]

Restriction implies the limitation of admissibility only to subjects falling within a category or a sufficiently narrow range of the values of a variable, in which the frequency of a disease could be considered constant. Although often used for other variables—for example, race, sex—its use for controlling the effect of age would involve a substantial reduction of the population available for the study and consequently an unacceptable loss of information in most situations.

Matching can be viewed as a particular form of restriction in which, after selecting the study group, restriction is imposed on the subjects in the comparison group. It can be performed on an individual basis (individual matching), or at a group level (frequency matching). In follow up studies it involves admitting to the non-exposed category only subjects similar in age to the exposed. In this way, comparable age distributions between exposure groups are obtained, and the confounding effect of age is eliminated. Matching is rarely used in follow up studies, and control for age in the analysis phase is usually preferred. In case-control studies, matching implies selecting controls with the same or similar age as the cases. However, differently from follow up studies, matching in these studies can introduce rather than remove confounding, if a matched (or stratified) analysis of the data is not performed.[3 14] Matching in case-control studies is intended primarily to provide a balanced distribution of subjects across age strata, thus permitting a more efficient (in a statistical sense) analysis. For variables like age, even after individual matching, subjects can be grouped in a few categories and a stratified analysis is usually sufficient and even preferable. Sometimes, in fact, the case or the control may be missing in some matching sets (due to unsuccessful matching, non-response, etc): such sets would be unavailable for a matched analysis.[5]

ANALYSIS OPTIONS

Adjustment for the effect of age in the analysis phase is generally accomplished by treating age as a stratification variable and by calculating stratum-specific (unconfounded) estimates of effect. Several stratification methods, each with its own advantages and limitations, are then available for obtaining summary measures of effect in cohort, case-control, and cross sectional studies. Further in depth analysis can be accomplished through regression modelling (Poisson, Cox's, and logistic regression methods), which has the merit of simultaneously taking into account several variables, including those measured on a continuous scale.[3-5 14-17]

In occupational cohort studies, the traditional method of adjusting for age and other variables is the so called "indirect standardisation". By this method control of age is obtained by multiplying a set of national or local reference rates (specific for sex, race, age, and calendar year) by the stratum specific person-years of the study cohort to obtain expected deaths (or cases) for a given disease. The ratio of the observed to the expected deaths is called standardised (and not standard!) mortality ratio (SMR); for incidence data, the term standardised incidence ratio (SIR) is used. Notwithstanding its good statistical properties, the method has an important drawback for the non-comparability of SMRs computed for different categories of exposure.[18] For this reason

many authors hold that comparison of standardised rates is valid (free of bias) only when a "direct" method of standardisation is used, in which age specific rates of the exposure groups are multiplied by the age stratum specific person-years deriving from a population used as standard. This method can be viewed as a process in which a weighted summary of stratum specific rates is calculated with a common set of weights.[3 14]

Notwithstanding the limitations of the SMR, this method continues to be used in most occupational epidemiology studies. In a survey of statistical methods used in the analysis of occupational cohort studies[19] it was found than in the years 1990–1, the standardised mortality ratio (SMR) was used as the main analysis in 75% of the studies, and was the only method in about 50%.

A simple and efficient method for stratified analysis is the Mantel-Haenszel method,[20] which can also be used for the analysis of matched data. Its properties favoured the large use of the method in case-control studies. Formulae based on a similar principle have been developed that can also be used in other study contexts.[3 5]

When the outcome under study is measured on a quantitative scale—for example, physiological variables, measures of internal dose of a substance—many parametric and non-parametric techniques are available for the analyses. In this context, multiple regression modelling is a powerful tool because it allows age to be taken into account together with other quantitative or qualitative variables.[21]

Age and effect modification

Aging modifies the way in which the human subject interacts with occupational exposures of any type (chemical, psychosocial, etc). Few epidemiological studies have been performed to specifically consider the issue of age as a modifier in occupational settings, even because the extreme age values are not represented in these settings. An example comes from a recently completed follow up of a Finnish population of middle aged workers in various municipal occupations.[22] After 11 years of aging, the workers considered that their work (which had not changed) had become heavier both mentally and physically, and reported an increase of the work related physical and mental strain. Changes in lifestyle—for example, physical activity—in the frequency of physical symptoms or diseases (especially musculoskeletal and cardiorespiratory)—and mental symptoms—for example, avoidance reactions—were found. The ability to work had declined significantly. In this investigation, aging clearly stood out as a major factor affecting the way in which the work environment affected exposed people; and in fact, consistently, age was found to be associated in "risk occupations" with early mortality, and retirement because of disability.

In the case of effect modification, the influence of age on the effect measure should not be eliminated; rather, its possible interaction with exposure should be explicitly considered. Age at first exposure is a possible example of age as a modifier in occupational epidemiology, due to the fact that the susceptibility to several disease agents—for example, carcinogens—may be different at different ages.[23]

Therefore, when considering age in occupational cancer studies, one is not only concerned with controlling its possible confounding effects, but also with examining the possibly changing pattern of the exposure related disease occurrence in different age strata. An example is taken from an environmental epidemiology study examining the long term health effects of exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin in the population of Seveso, Italy. The decision was made to examine separately the cancer occurrence in the age category <20 years given the probable special susceptibility to the exposure of interest and its possible carcinogenic action at younger ages.[24] As another example, a greater susceptibility of children and younger women to the effects of radiation has been suggested.[15]

Age and person-time at risk

In longitudinal studies with follow up extending over many years, aging implies the passage of a subject across different categories of age and other time related variables. The need for a proper consideration of this temporal aspect led to the development of analytical methods capable of incorporating the time dimension in assessing the frequency of diseases The modified life table method[25 26] involves the splitting of subjects' observation time in different categories of time varying determinants (age, etc). The fundamental unit of observation is not the subject, but the duration of time for which the subject is observed in a specified time or age band.[17] This is the most appropriate method to examine the individual risk experience when the periods of observation vary greatly among subjects, the usual case when studying long latency diseases.[4]

Age and relative measures of effect

Most epidemiological studies make use of relative (rate, risk, and odds ratios) rather than absolute measures of effect (rate, risk differences, or excesses). In this way, an implicit reference is made to a statistical model called multiplicative (as opposed to additive). Some authors[15 17] noted that the empirical stability of relative risks over age strata—which implies that age is not a modifier of relative measures—has been an important motivation for the use of the relative risk. In fact, a fundamental property of an effect measure should be its capability to express with a single summary measure the effect of a given exposure, without detailing the effects in individual strata of covariates (age, time, sex, etc). Relative measures do possess this attribute in numerous circumstances,[15 27] and are thus often preferable to risk difference. Non-uniformity of relative risks over age strata may signify that: (*a*) a different measure—for example, risk difference—could be a better descriptor of the exposure-disease relation;[2 3 15 17] and (*b*) disease susceptibility, do

vary with age.[15] In 1990–1, only one out of 200 occupational cohort studies adopted an additive measure of effect.[19]

Conclusions

In conclusion, although age is rarely the specific focus of occupational epidemiology studies, none the less the criteria to enrol subjects, the methods to examine their risk experience over time, the choice of the effect measure, and the type of analysis do, to some extent, depend on the need to properly take age into consideration. In most situations age is important as a potential confounder, and should be controlled to avoid a distortion in the effect estimate. Age, in other instances, can affect the mode of action of a given exposure, and in these instances it should not be treated as extraneous, rather its interaction with exposure should be explicitly considered in analysis and interpretation of results. In studies lasting for long periods, analytical techniques should make allowance for aging, which is in itself an indicator of the time related changing risk of many diseases. Finally, age is one of the reasons why relative risks are more commonly used as effect measures than risk differences: the relative risks have greater stability over age strata.

1 Ilmarinen J. Aging and work: coping with strengths and weaknesses. *Scand J Work Environ Health* 1997;**23**(suppl1): 3–5.
2 Miettinen O. Confounding and effect-modification. *Am J Epidemiol* 1974;**100**:350–3.
3 Rothman KJ. *Modern epidemiology*. Boston: Little Brown, 1986.
4 Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic research. Principles and quantitative methods*. New York: Van Nostrand Reinhold, 1982.
5 Checkoway H, Pearce NE, Crawford-Brown DJ. *Research methods in occupational epidemiology*. New York: Oxford University Press, 1989.
6 Pearce N. Methodological problems of time-related variables in occupational cohort studies. *Rev Epidemiol Sante Publique* 1992;**40**:S43–54.
7 Finkelstein MM. Potential pitfall in using cumulative exposure in exposure-response relationships: demonstration and discussion. *Am J Ind Med* 1995;**28**:41–7.
8 Miettinen OS, Cook EF. Confounding: essence and detection. *Am J Epidemiol* 1981;**114**:593–603.
9 Bertazzi PA, Pesatori AC, Zocchetti C, Latocca R. Mortality study of cancer risk among oil refinery workers. *Int Arch Occup Environ Health* 1989;**61**:261–70.
10 Pesatori AC, Sontag JM, Lubin JH, Consonni D, Blair A. Cohort mortality and nested case-control study of lung cancer among structural pest control workers in Florida (United States). *Cancer Causes Control* 1994;**5**:310–8.
11 Hernberg S. *Introduction to occupational epidemiology*. Chelsea: Lewis Publishers, 1992.
12 Cherry N. Evaluation of preventive measures. In: McDonald JC, ed. *Epidemiology of work-related diseases*. London: BMJ, 1995:413–39.
13 Duca PG, Pelfini G, Ferguglia G, Settimi L, Peverelli C, Sevosi I, Terzaghi G. Efficacia dell'uso di creme-barriera nel prevenire affezioni dermatologiche in lavoratori delle tintostamperie: risultati di una sperimentazione randomizzata. *Med Lav* 1994;**85**:231–8.
14 Miettinen OS. *Theoretical epidemiology. Principle of occurrence research in medicine*. New York: John Wiley, 1985.
15 Breslow NE, Day NE. *Statistical methods in cancer research*. Vol 1. *The analysis of case-control studies*. Lyon: International Agency for Research on Cancer, 1980. (IARC Sci Publ No 32.)
16 Breslow NE, Day NE. *Statistical methods in cancer research*. Vol II. *The design and analysis of cohort studies*. Lyon: International Agency for Research on Cancer, 1987. (IARC Sci Publ No 82.)
17 Clayton D, Hills M. *Statistical models in epidemiology*. Oxford: Oxford University Press, 1993.
18 Miettinen OS. Standardization of risk ratios. *Am J Epidemiol* 1972;**96**:383–8.
19 Callas PW, Pastides H. Hosmer DW. Survey of methods and statistical models used in the analysis of occupational cohort studies. *Occup Environ Med* 1994;**51**:649–55.
20 Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959;**22**:719–48.
21 Kleinbaum DG, Kupper LL, Muller KE. *Applied regression analysis and other multivariable methods*. 2nd ed. Boston; PWS-KENT, 1988.
22 Tuomi K, Ilmarinen J, Seitsamo J, Huuhtanen P, Martikainen R, Nygard C-H, Klockars M. Summary of the Finnish research project (1981–92) to promote the health and work ability of aging workers. *Scand J Work Environ Health* 1997; **23**(suppl 1):66–71.
23 Leutzinger Y, Richie JP Jr. The effect of animal age on tumor induction. In: Arcos JC, Argus MF, Woo Y-T, eds. *Chemical induction of cancer. Modulation and combination effects*. Boston: Birkhauser, 1995: 373–95.
24 Pesatori AC, Consonni D, Tironi A, Zocchetti C, Fini A, Bertazzi PA. Cancer in a young population in a dioxin-contaminated area. *Int J Epidemiol* 1993;**22**:1010–3.
25 Case RAM, Lea AJ. Mustard gas poisoning, chronic bronchitis and lung cancer. *Br J Prev Soc Med* 1955;**9**:62–72.
26 Hill ID. Computing man-years at risk. *Br J Prev Soc Med* 1972;**26**:132–4.
27 International Agency for Research on Cancer. *Cancer: causes, occurrence and control*. Lyon: IARC, 1990. (IARC Sci Publ No 100.)