Selection Bias in Educational Research

A.C. TUIJNMAN
OCDE France

The problem of methodology:
quantitative vs. qualitative approaches,
internal vs. external evaluation

In this entry the conceptual and methodological problems arising from the use in educational research of samples drawn from selected or self-selected populations are discussed. The existence of selection bias greatly limits the validity of generalizations that can be drawn from the analysis of data associated with samples where selection has taken place. Moreover, it commonly distorts the findings of educational research, where no allowance has been made for its effects.

The entry provides examples of research studies in which the presence of selection bias can harm the accuracy of the parameter estimates and impair the validity of statistical inference. Methods for detecting and reducing selection bias are also reviewed.

1. Definitions

Wei and Cowan (1988 p. 332) define selection bias as "the bias that occurs in a clinical trial, experiment or sample survey because of the effects of the mechanism used to select individuals or units for inclusion in the experiment or survey". In sample surveys selection bias can influence the estimates of certain parameters because the selection probabilities for inclusion in a survey can be related to the variables being studied. This may be the case if there are unobserved determinants of the probability of inclusion that correlate significantly with unobserved determinants of the criterion variables.

At the core of selection bias is a restriction of range. The children enrolled in the Headstart Program, for example, because of the way they were chosen for inclusion in the program were likely to have restricted ranges on home background measures, scholastic achievement, and many other variables. Much of the research on selection bias has originated from work in the area of social and educational program evaluation. The evaluation of Headstart and public training programs for the unemployed are two examples. The harmful effects of selection bias on the accuracy of mean scores and regression estimates are now widely recognized and documented.

In the case of regression analysis for example, it is normally assumed that there is no range restriction on the dependent variables. Selection bias in the regressors distorts regression estimates and other statistics only in certain specifiable ways. According to Darlington (1990 p. 198) statistics such as the regression slope b_j, the intercept *a*, the predicted variance of the dependent variable $\hat{Y}_i$, the residual variance e_i, and the mean square error, MSE, are *not* distorted by a restriction of range. However, statistics such as the multiple correlation *R*, the variance of the dependent variable *Y*, all values of the correlation r_{Yj}, and the values of the partial correlation pr_j for the restricted variables are affected. In general, however, the statistics undistorted by range restriction are those used in causal analysis, whereas the distorted statistics are used in prediction. Therefore, prediction studies are seriously distorted by range restriction more often than regression studies involving causal analysis.

Different concepts exist for describing the phenomenon highlighted above. Statisticians seem to be neutral; they tend to use the practical term selection bias. For partly theoretical reasons *self*-selection bias has much currency in the economics of education. By contrast, educational sociologists tend in certain situations to use concepts such as 'retentivity' and 'selectivity'. These are considered to capture the idea that individuals often have no control over the selection mechanism, which may be random or nonrandom. The addition 'self' can be taken as suggesting that participation in an educational program is based on rational choice, and that the people who are *not* selected for enrollment in a program carry responsibility for themselves. Since many people seem to have no such choice; the term self-selection bias can be considered misleading.

2. Selection Bias in Sample Surveys

In survey research methodology a distinction is often made between the desired target population and the defined target population, from which a sample can be drawn. Differences normally occur between the ideal and the achievable, for example, because it may not be possible to obtain observations for all of the desired units. The accuracy of a sample statistic, $\bar{x}$, based on a probability sample is generally assessed in terms of the variance of $\bar{x}$, denoted Var($\bar{x}$), which quantifies the sampling stability of the values of $\bar{x}$ around the expected value E($\bar{x}$) (Ross, 1987). Sampling error, which is a property of the entire sampling distribution, belongs to both the selection and estimation procedures that are employed in obtaining sampled data. Selection bias is therefore a very special case of sampling bias.

The information obtained from a sample is normally used to advance useful generalizations about the defined population. The usefulness of statistical inference depends, among other factors, on how closely the achieved sample reproduces the characteristics of the defined population. Because the exact population parameters are unknown, the influence of sampling error is usually assessed on the basis of the internal evidence of a single sample of data. But sampling error, conventionally defined, does not necessarily take account of the possible differences between the desired target and the defined target population. The ideal is a population free of any selection mechanism that is related to the variables being studied and that influences the selection probabilities for inclusion in a sample. As the examples given below suggest, this ideal is often unattainable. Hence systematic differences often arise between the desired and defined target populations. Such differences may introduce selection bias in sample surveys. Even if the sampling errors are very small, which is usually the case in well designed studies, then the presence of selection bias in the data can seriously harm the accuracy of the parameter estimates and impair the validity of statistical inference. (See **Sampling*; **Survey Research Methods*).

3. Examples of Selection Bias and Adjustment Methods

Selection bias is usually not a problem in educational research as long as the defined target population from which a sample is drawn comprises all students in an age group attending compulsory school. Of course, the noncoverage and nonresponse of certain categories of school-aged children may introduce threats to accuracy and validity that are, conceptually at least, very similar to problems associated with selection bias. But serious selection bias can occur if the enrollment status of students in post-compulsory programs is used as a criterion for sampling, since the mechanisms that impinge on individual choice and the decision to enroll in a program of post-compulsory education may well determine the characteristics of the target population from which a sample can be drawn. Because the participants in such programs may differ markedly and systematically from those who are not enrolled, studies involving samples drawn from heterogeneous target populations associated with certain post-compulsory school programs can be highly vulnerable to selection bias and other sampling effects. Consider the following examples:

IEA Study of Science Achievement

In a study of science education in 23 countries conducted in 1983-84 by the International Association for the Evaluation of Educational Achievement (IEA), three target populations were selected in each school system (Postlethwaite and Wiley 1991, Keeves 1992): (1) Population 1 -- all students aged 10:0 to 10:11 on the specified date of testing or all students in the grade where most 10-year-olds were to be found; (2) Population 2 -- all students aged 14:0 to 14:11 on the specified date of testing or all students in the grade where most 14-year-olds were to be found; (3) Population 3 -- all students in the final grade of full-time secondary education in programs leading to entry into higher education.

Two developing countries excepted, close to 100 percent of an age group was in school at the Population 1 level. Also at the Population 2 level nearly 100 percent was in school in the systems of the industrialized countries participating in the survey. However, as can be seen from Table 1, the percentage of an age group attending full-time education at Population 3 level varied substantially across systems. The Table also shows that the mean scores on comparable tests of student achievement in chemistry and physics differed across systems.

Insert Table 1 Here

Inferences based on IEA data, such as that included in Table 1, often involve the making of comparisons across educational systems at the same time; for example: "in 1983/84 the physics yield of Japanese schools at the Population 3 level was substantially higher than that of comparable Australian schools". This particular example may either be true, since in both systems 11 percent of an age group take physics, or it may not, since the standard deviations are different. Which inferences can safely be drawn from the data presented in this Table remains an open question, but one aspect stands out clearly: there is unwanted variation in variables such as age and percent in school which can influence the dependent variable being measured -- the system mean score on the IEA physics and chemistry tests.

The problem of accounting for undesired but natural variation in age and school enrollment in selective educational systems was clearly recognized at the time the first IEA Mathematics project was carried out in the early 1960s. Walker in Husén (1967) developed a mathematical model to represent the effects of selection, which at that time was known as retentivity bias. The basic idea underlying the model is that "each country has the same distribution of mathematical ability in the complete age group and that the differences in means and variances found at the preuniversity mathematics stage are a result of selection procedures" (op. cit. p. 135). One can of course question the validity of this assumption, but it serves well in demonstrating how the adjustment model works using, in this case, the formulae derived for the truncated normal distribution:

$$E(\bar{x}) = y/q; \ \text{Var} = 1-(y/q)\{(y/q)-k\} \quad (1)$$

where $E(\bar{x})$ = expected or adjusted mean score

q = percentage of students selected

y = ordinate of normal curve at point of cut off

k = point of cut off

Walker used the formula in Eqn (1) to calculate the expected mean scores for data of the type shown in Table 1. The adjusted mean scores are said to be *unbiased* estimators of some population parameter. After comparing the expected with the observed scores, he concludes (op. cit. p. 136): "for such sweeping assumptions as have been made, the agreement between theory and results is moderately good". If one were willing to accept Walker's basic assumption concerning the equal and normal distribution of mathematical ability across countries, then one could use the formula in Eqn (1) to calculate expected mean scores for the chemistry and physics test results shown in Table 1. Walker (op. cit. p. 136) also shows how the model can be adapted to take account of other assumptions or observations, for example; a correlation r between the variable (or possibly variables) representing the selection mechanism and the outcome:

$$E(\bar{x}) = r\,(y/q); \ \text{Var} = r^2\,(y/q)\{(y/q)-k\} \quad (2)$$

(See **Models and Model Building*)

Scholastic Aptitude Testing

The selection problems besetting IEA studies that involve sampled data from students in post-compulsory grades of the educational system apply of course equally well to other studies in which the mean scores of selected samples of students are compared across countries and regions. Several examples are known from the United States. A much publicized case concerns the ranking of states by the mean scores obtained by students taking the College Board's Scholastic Aptitude Test (SAT) or the American College Testing program (ACT). Since the proportion of students taking such tests differs from state to state, and also because certain variables such as college requirements and state funding policies influence the selection probabilities for inclusion in a SAT sample, meaningful statistical inferences cannot be made simply on the basis of unadjusted SAT mean scores. The question that has occupied many statisticians for decades is how such an adjustment can be made.

The simplest approach is to adjust for the selection ratio, by covarying out the participation rate. Powell and Steelman (1984 p. 400) conclude on the basis of an analysis that sought to adjust for the selection ratio that "about 82 percent of the variation in states' performance can be attributed to this variable" and that "at best, only 18 percent can be considered 'real' variation". This approach to adjustment is questioned on grounds that "percentage taking the SAT" can be considered as a posttreatment concomitant variable that itself is influenced by the selection process, the properties of which are unknown (see Wainer 1986 p. 11-13). Other approaches have been proposed, namely regression adjustment for relevant background variables, and enlarging the SAT pool by means of imputation and other methods. These approaches and methods are discussed in a section below.

Training Programs for the Unemployed

A third group of researchers developed an interest in the phenomenon of selection bias in connection with the evaluation of public employment training programs. Many governments assigned high priority to the reduction of unemployment, especially in the latter half of the 1970s and early 1980s. Skills training was considered important, and publicly supported training programs were either set up or, in countries where they already existed, extended.
Especially in Sweden, the United States and the United Kingdom much effort was devoted to evaluating the effectiveness of public training programs in terms of the postprogram employment status and earnings of the trainees (for an overview, see Barnow 1987). However, the findings of this expanding body of research were inconclusive in that the expected benefits could often not be demonstrated. The absence of the expected outcomes led some econometricians to suggest that program effectiveness could not be determined because the participants enrolled in employment training programs were negatively self-selected with respect to the important predictors of program success; hence the doubtful or even negative relationships between program treatments and outcomes. This example has become a classic case in the writings on self-selection bias, not so much because a definite methodological breakthrough was reached but rather as a consequence of the energy which economists devoted to the problem. (See **Evaluation of Public Training Programs*)

Higher and Continuing Education

Even with the recent expansion of intake, enrollment in higher and continuing education remains selective on a number of accounts. Compared with all eligible students,,university entrants are likely to be more homogeneous with respect to variables such as cognitive ability, achievement motivation and potential, and expected life career. Both groups are also likely to be more homogeneous than the general population. Selection bias may thus be present in studies in higher education that make use of samples in which students are selected for inclusion on the basis of their enrollment status. (see *Validity)

During the 1980s the research on the effects of education was increasingly focused on the post-initial sector. Because of the magnitude of the social and economic stakes involved, funds for conducting large-scale evaluation studies of continuing education, and especially job training, became accessible in many countries, and relevant data sets were assembled as a result.

Two conclusions soon emerged from this line of work. First, the designs for research have to be developed so that one can distinguish between effects due to initial education and those due to continuing education. The failure to account for previous experience of learning will otherwise result in an upward bias in the parameter estimated for the effect of continuing education. Second, any group of women and men who take part in programs of adult education, public employment training, and human resource development in firms is atypical in some sense. Because of the variety of post-initial educational programs and the heterogeneity of their clientele the results obtained in evaluation studies of continuing education are seldom comparable and generalizable. Since selection bias obviously operates, doubt is cast on the validity of any statistical inference made.

Tuijnman and Fägerlind (1989) report a longitudinal study of differences between participants and nonparticipants in continuing education at various stages in the lifespan (see *Longitudinal Research Methods*). Table 2 shows that there are marked differences between the two groups on a number of indicators and at different age levels. The sheer size of the differences makes it clear that evaluation studies involving data sampled only from program participants are likely to suffer from a very large selection bias.

Insert Table 2 Here

The conclusion seems clear: because participants in post-initial education and training programs tend to differ systematically from nonparticipants in a number of respects, major problems arise in estimating the effectiveness of such programs. The consequences of selection bias can no longer be ignored. Priority may therefore have to be given to the definition of control groups and to the improvement of techniques for the sampling of populations representative of the nonparticipant groups in education.

4. Additional Methods

Because of the work carried out by statisticians, economists and researchers in education, each in his own field, understanding of selection bias has developed rapidly since the mid-1970s. Methods for detecting -- and possibly reducing -- selection bias have also been developed. The basic approaches were outlined above. Regression adjustment and imputation methods are discussed below.

Simple Allocation Designs

Allocation methods usually involve the making of assumptions about participation and adjusting for selection by 'partialling out' the nonparticipation effect from the data. The simplest approach was previously mentioned, namely to adjust for unequal selection by covarying out participation rates. Many variations on this theme can be found in the research literature.

Selection Bias as a Specification Error

Another approach is to employ at least one regressor variable measuring an important aspect of program exclusion restrictions or the decision to participate. If such information is available and specified in the equation determining enrollment, then selection bias can be reduced. Regression adjustment methods thus consider selectivity as a specification error (Heckman 1979). The key to this approach is to obtain measures on some or all of the characteristics that make comparisons unfair. By including these measures in the regression equation it is hoped that the biasing effects of selection can be removed from the variables of interest.

This approach is conceptually analogous to the method of handling unwanted design effects in stratified probability samples by carrying the stratification variables through all stages of the data analysis. The application of regression adjustment is straightforward in the case of controlling for design effects, because the stratification variables are known and data on them usually collected. The difficulty of using the method in order to adjust for selection effects is precisely that the variables associated with these effects are mostly unknown and data on them often not available. The method thus stands or falls with the quality of previously validated knowledge and data.

Selection Modeling

Selection modeling is a methodology proposed by Heckman and Robb (1986) to overcome some of the problems of treating selection bias as a specification error. It can be used to generate artificially the missing scores on the outcome variables for the (nonobserved) nonparticipants by means of a regression imputation technique. The methodology consists of three basic steps (Wainer 1986 p. 3): (1) *observe* the distribution of the outcome variables of interest; (2) *hypothesize* the functional relationship between the likelihood of participation and the variables observed; and (3) *calculate* the distribution of the outcome variables for both those who are observed and those who are not.

Not surprisingly, this method has met with strong criticism. The difficulty is of the same kind as that associated with the approach taken by Walker in the adjustment for selection effects in IEA data: the validity of the minimal identifying assumptions that guide the adjustment usually cannot be fully tested with data. Despite this criticism the method is not completely without merit, as Heckman and Robb (1986) demonstrate on the basis of experiments and simulations.

Imputation Adjustment

As in the example above, imputation adjustment also involves the making of assumptions about nonobservation. Imputation methods can be used to compensate for the effects of range restriction in ways analogous to the treatment of attrition and selective noncoverage and nonresponse in survey samples. (See **Missing Data and Nonresponse*).

Mixture Modeling

Mixture modeling also considers selection bias as a problem of missing data, e.g. nonresponse or nonobservation dependent on the outcome variable. The approach, as its name suggests, uses a combination of methods to predict the scores that are missing in range-restricted, dependent variables. The key assumption in mixture modeling is that the respondents and the nonrespondents form different populations, that a procedure can be designed for the multiple imputation of data values for nonrespondents, and that parameters for nonrespondents can be estimated using assumptions about the relationship between respondents and nonrespondents. In practice this means that the same person is simultaneously observed as a member of the program participant group and the excluded group. If both states can be observed then the selection effect can be identified from the treatment effect and, provided that the data are appropriately weighted, unbiased estimates of program impact may be obtained. Propensity score methodology (Rosenbaum and Rubin 1985) is a related approach.

5. Experimental and Quasi-Experimental Designs

It can be inferred from the diversity of available methods to adjust for selection bias that the search for alternatives to rigorous experimental controls in program evaluation has been a long one. Many statisticians and others contend that this effort has not yielded a method that can appropriately be used for inference from nonexperimental data derived from self-selected samples. In the absence of such valid methods, a strong case is made for program evaluation by experimental design. Experimental procedures commonly involve the use of multiple groups that, even though randomized, differ substantially with respect to their background characteristics and the level of program treatment. These groups are then compared in terms of pre- and post-program characteristics. But problems also beset this approach, for example, it may not be possible to employ strict controls in social settings.

An alternative for controlled experimentation is the quasi-experimental or comparison group design. Quasi-experimental methods are often used in program evaluation. In its basic form it requires a preprogram measure, a postprogram measure that can reflect the effect of the program being studied, and a measure describing the status of the persons in the sample, usually whether they took part (treatment group) or did not (comparison group). In educational research the comparison group usually comprises nonparticipants who are assumed or shown to have many characteristics in common with the program treatment group. Bryant (1978) provides an example of how unbiased estimates of program effects might be obtained from a sample of participants and a comparison group drawn from a large household survey. Difficult problems concern both the definition and use of a comparison group, and whether characteristics of the groups influence treatment conditions. Range restriction may also beset quasi-experimental studies employing multiple samples. It may also be impractical or even impossible to identify an appropriate comparison group and collect data on the group members. Estimates of program impact are therefore likely to remain biased -- unless an appropriate control group can be identified and sampled, and characteristics of the groups have not influenced treatment conditions.

Unlike the conditions that are specific to an experimental design, the accuracy of parameter estimates in nonexperimental evaluations of program effectiveness depend on the way the program input, treatment and program outcome equations are specified. Some nonexperimental evaluations of post-initial education have made use of two-wave models or even multiple time-series designs. Although it is acknowledged that these and other adjustment methods are imperfect, nonexperimental evaluation cannot be ruled out, because many complex programs exist where one or more of the following conditions apply: (1) selection mechanisms or voluntary self-selection are a condition for inclusion; (2) it cannot be established with certainty whether and in what respects participants differ from nonparticipants; (3) the program treatments cannot be directly measured; (4) an adequate comparison group cannot be defined; and characteristics of the groups influence treatment conditions. (See **Experimental Studies*; **Quasi-experimentation*).

6. Conclusion

The only way to solve the selection bias problem is to draw a sample in such a way that the enrollment status of program participants becomes irrelevant as a criterion for sampling. An attractive means of overcoming selection and simultaneity problems is presented by longitudinal studies in which an age group is sampled before they leave compulsory school. Longitudinal data cannot be regarded as a panacea, however. Serious problems of data analysis remain even in cases where some of the data are collected before the treatment and the sample is taken from a population comprising both participants and nonparticipants. If unobserved variables and their errors correlate with the outcome variable, which is often the case, then specification errors may be present and unbiased estimators cannot be obtained. Hence, since their use may yield inconsistent estimates of program effects even in cases where selection bias has been ruled out, regression methods may have limited applicability in program evaluation research. (See **Longitudinal Research Methods*).

See also: *Regression Analysis; *Reliability; *Sampling Errors; *Validity

7. References

Barnow, B S 1987 The impact of CETA programs on earnings: A review of the literature. *Journal of Human Resources, 18,* 157-193.

Bryant, E C 1978 Survey statistics in social program evaluation. In David, H A (Ed.), *Contributions to survey sampling and applied statistics* (pp. 41-55). London.; Academic Press

Darlington, R B 1990 *Regression and linear models.* London: McGraw-Hill.

Heckman, J J 1979 Sample selection bias as a specification error. *Econometrica, 46,* 931-959.

Heckman, J J and Robb, R 1986 Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes. In H. Weiner (Ed.), *Drawing inferences from self-selected samples* (pp. 63-110). Berlin Springer-Verlag

Husén, T (Ed.), *International study of achievement in mathematics: A comparison of twelve countries* (Vol. 2, pp. 135-139) Stockholm Almqvist and Wiksell, 1967.

Keeves, J P 1992 *Changes in science education and achievement: 1970 to 1984.* Oxford Pergamon Press

Powell, B and Steelman, L C 1984 Variations in state SAT performance: Meaningful or misleading? *Harvard Educational Review, 54,* 389-412.

Postlethwaite, T N and Wiley, D E 1991 *Science achievement in twenty-three countries* Oxford Pergamon Press.

Rosenbaum, P and Rubin, D B 1985 Constructing a control group using multivariate sampling methods that incorporate the propensity score. *American Statistician,* 39, 33-38.

Ross, K N 1987 Sample design. *International Journal of Educational Research, 11* (1), 57-75.

Tuijnman, A C and Fägerlind, I 1989 Measuring and predicting participation in lifelong education using longitudinal data. *Scandinavian Journal of Educational Research, 33,* 47-66.

Wainer, H 1986 The SAT as a social indicator: A pretty bad idea. In H. Weiner (Ed.), *Drawing inferences from self-selected samples* (pp. 8-21) Berlin: Springer-Verlag.

Wei, L J and Cowan, Ch D 1988 Selection bias. *Encyclopedia of Statistical Sciences, 8,* 332-334.

8. Suggestions for Further Reading

Ashenfelter, O 1987 The case for evaluating training programs with randomized trials. *Economics of Education Review, 6,* 333-338.

Behrman, J R 1987 Schooling and other human capital investments: Can the effects be identified? *Economics of Education Review, 6,* 301-305.

Fraker, T and Maynard, R 1987 The adequacy of comparison group designs for evaluations of employment-related programs. *Journal of Human Resources, 22,* 194-227.

Glynn, R J, Laird, N M and Rubin, D B 1986 Selection modeling versus mixture modeling with nonignorable nonresponse. In H. Weiner (Ed.), *Drawing inferences from self-selected samples* (pp. 116-142). Berlin Springer-Verlag.

Kiefer N 1979 Population heterogeneity and inference from panel data on the effects of vocational training. *Journal of Political Economy, 87,* S213-S226.

Rubin, D B 1987 *Multiple imputation for nonresponse in surveys* New York Wiley.

Willis, R J and Sherwin, R 1979 Education and self-selection. *Journal of Political Economy, 87* (5, Part 2), S7-S36.

Proposed Key Words

Selection Bias

- Definition
- Effects
- Detection
- Correction

Selection Bias Methods

- simple allocation designs
- imputation adjustment
- regression adjustment
- selection modeling
- mixture modeling.

Self-selection in education

Experimental design: selection bias

Quasi-experimental design: selection bias

Nonexperimental design: selection bias

Sampling: selection bias

Table 1. Percentage of an age group in school, mean ages in Population 3, grade level, means and standard deviations in Chemistry (25 items) and Physics (26 items)

				Chemistry			Physics		
Educational system [a]	grade level tested	% in school [b]	mean age [c]	% in school [d]	$\overline{X}$ test score	s.d. test score	% in school	$\overline{X}$ test score	s.d. test score
Australia	12	39	17:3	12	49.1	19.5	11	48.7	15.2
England	13	20	18:0	5	69.3	17.9	6	58.4	15.1
Finland	12	41	18:6	16	35.9	15.3	14	37.9	14.1
Hong Kong (F6)	12	27	18:3	20	68.2	17.7	20	61.2	14.5
Hungary	12	18	18:0	1	50.2	19.7	4	58.7	17.6
Japan	12	63	18:2	16	55.5	22.9	11	58.5	17.8
Korea	12	38	17:9	37	30.9	14.8	14	39.8	16.9
Norway	12	40	18:9	6	44.3	18.4	10	54.1	15.9
United States (1986)	12	83	17:7	2	37.7	18.2	1	45.3	15.9

[a] Data collected in 1983 and 1984; [b] Excluding students in vocational education; [c] Age given in years and months; [d] percent of an age group.

Source: Postlethwaite and Wiley (1991)

Table 2. Differences between participants and nonparticipants in continuing education at different ages with respect to selection-relevant characteristics (*Malmö Study* data [a]; effect size indices [b])

Variable/ characteristic	Participation age 30-35	Participation age 36-42	Participation age 43-56
Father's occupation	0.41	0.44	0.38
Cognitive ability measured at age 10	0.44	0.54	0.54
Perception of parental support for schooling	0.50	0.61	0.43
Formal educational attainment	0.63	0.70	0.58
Cognitive ability measured at age 20	0.54	0.54	0.65
Interest in continuing education [c]	0.98	0.83	0.61
Occupational level [c]	0.63	0.82	0.91
Earnings (logarithm) [c]	0.32	0.54	0.68
Job satisfaction [c]	0.04	0.23	0.37
N cases	716	703	671

[a] The *Malmö Study* is a longitudinal study of a group of 1432 Swedish men and women that was begun in 1938. The most recent data collection took place in the early 1990s (see **Longitudinal Research Methods*); [b] Effect size is the difference between two group means expressed in standard deviation units; [c] Measured, respectively, at age 35, 42 and 56 years.

Source: Tuijnman and Fägerlind (1989, p. 59)