WORK IN PROGRESS

PLEASE DO NOT CITE

METHODOLOGICAL COMPARISON OF DIFFERENT APPROACHES USED TO CONTROL FOR SELECTION BIAS IN NON-RANDOMISED STUDIES

Author and contact details

Hema Mistry[1]* and Stephen Morris[1]

[1]Health Economics Research Group, Brunel University, Uxbridge, Middlesex, UB8 3PH

*Correspondence to H Mistry: email: hema.mistry@brunel.ac.uk , telephone: 01895 265447

Abstract

The randomised controlled trial is often seen as the gold standard in health care research, because it provides the strongest research design and has the lowest threat of bias. Despite the growing use of non-randomised studies to evaluate health care technologies, there is currently no 'gold standard' approach to control for selection bias in non-randomised studies. Whilst a number of approaches have been used in order to obtain reliable estimates of costs and effects, no formal comparisons of the different methods have been made. This paper aims to explore and compare alternative methods used to control for selection bias in non-randomised studies.

Firstly, different definitions of selection bias are covered. The paper then reviews the literature on methods used to control for selection bias in non-randomised studies, such as the more traditional method (regression analysis) and more recent methods including propensity score analysis and instrumental variables analysis. Strengths and limitations of the different approaches will be highlighted.

The paper attempts to set out criteria and provide recommendations on which methods may be more suitable to obtain reliable cost and effect estimates to apply to economic evaluations of health care technologies, where data are not randomised and numbers are small.

Introduction

The randomised controlled trial (RCT) is often seen as the gold standard in health care research, because it provides the strongest research design and has the lowest threat of bias. However, in some situations, RCTs may be unnecessary, inappropriate, impossible or inadequate [Black, 1996]. For example, in some situations, treatment on humans may be unethical. In these circumstances, non-randomised studies (or observational studies) may be the only way to assess the costs and effectiveness of treatments. However, as participants in observational studies are not randomised there is a potential for bias. It is the potential for selection bias that most clearly differentiates randomised from non-randomised studies. One of the main reasons that RCTs are said to be superior to other types of studies is its ability to deal with selection bias. In another words, in a sufficiently large trial, randomisation eliminates selection bias. Whilst a number of approaches have been used in order to obtain reliable estimates of costs and effects, no formal comparisons of the different methods have been made. Only limited comparisons between some of the different methods have been made [Deeks et al, 2003; Crown, 2001; Earle et al, 2001; Shah et al, 2005; Zanutto, 2006; Posner et al, 2001; Morris, 2006].

First, the paper looks at different definitions of selection bias. Then, the paper explores and compares alternative methods used to control for selection bias in non-randomised studies which were identified from the literature review. Strengths and limitations of the different methods will be highlighted. The paper also attempts to set out criteria and provide recommendations on which methods may be more suitable to obtain reliable cost and effect estimates to apply to economic evaluations of health care technologies, where data are not randomised and numbers are small.

Methods

Categorisation of biases and selection bias

Selection bias refers to an absence of comparability between the groups being studied, and can occur during any stage of the research. In other words, the selection of study participants for the intervention group have different characteristics from those allocated to the control group. The choice of intervention in these circumstances is usually due to clinician's preference, but may also be due to the patient's preference, patient's characteristics and also clinical history [Deeks et al, 2003]. On average, selection bias tends to make treatment effects appear larger than they are and the size of these distortions can be as large as or larger than the size of effects that are being measured [Kunz and Oxman, 1995].

Several classifications of biases exist in clinical research. Sackett (1979) compiled a list of 35 biases that can arise in sampling and measurement of data; 22 of these biases identified occur in specifying and selecting the study sample. Delgado-Rodriguez and Llorca (2004) in their paper describe 34 biases which go under the subgroup of selection bias. On the other hand, Feinstein (1985) consolidated biases into four categories that can arise during research, one of these being susceptibility bias which refers to differences in baseline characteristics. The Cochrane Collaboration handbook (2005) also highlighted four main sources of systematic bias and all these biases can affect non-randomised studies. One of the biases being selection bias which refers to systematic differences in comparison groups. The latter two classifications of selection or susceptibility bias are similar.

There are various forms of selection bias that have been named. For example, Sackett (1979) and Delgado-Rodriguez and Llorca (2004) agree on two types of selection bias: Berkson bias (also known as admission rate bias) which can result from differential rates for hospital admission for cases and controls; and Neyman bias (also known as prevalence-incidence bias) which arises when a time gap occurs between the exposure and the selection of study participants. For example, this bias occurs in studies of diseases that are quickly fatal and can create a study group which is not representative of the general population. Other types of selection bias include unmasking (detection signal) bias, non-respondent bias and membership bias. Unmasking bias refers to an innocent exposure which may become suspect if, rather than causing the disease, it causes a sign or symptom which precipitates the search for the disease; whereas non-respondent bias refers to non-respondents (or 'late-comers') from a specified sample may exhibit exposures or outcomes which differ from those of respondents (or 'early comers'); and membership bias refers to membership in a group i.e. people who attend a swimming club, may imply a degree of health which differs systematically from that of the general population.

Selection bias in studies may occur not only due to observed covariates, but may also be due to unobserved covariates. Both the observed and unobserved covariates can be either known or unknown to the clinicians, as well as recorded or not recorded during the study.

Literature review

The aim of the literature review was to gather evidence on methods identified to control for selection bias in non-randomised studies. The following databases were searched for papers: MEDLINE, Ingenta, BIDS IBSS and EconLit from the earliest possible date up to April 2006. The search terms used for the literature search are shown in Appendix 1.

Results

Literature review results

The literature search identified 245 abstracts (after removing duplicates). After reviewing all 245 titles and abstracts, 62 of these abstracts identified or indicated methods to control for selection bias. After obtaining and reviewing 60 articles (2 articles were only available in the form of abstracts), various methods were identified which are summarised in Table 1. Some articles identified more than one method to control for selection bias. These papers explicitly stated that they were dealing with selection bias using these methods. Both, theoretical and practical papers were identified, which were not just related to health care, but were also related to the labour market.

Overview of methods identified to control selection bias in non-randomised studies

1) Multiple regression analysis

The traditional approach to control for selection bias is to use regression models. Regression models estimate how much each independent variable such as risk factors relate to the dependent variable or the outcome. When the comparison between the groups is made, the adjustments are added or subtracted from the estimated treatment effect to account for the impact of differences in each of the baseline variables according to their estimated relationship with the dependent variable [Deeks et al, 2003]. Regression models can only deal with observed variables, and cannot adjust for unobserved variables.

2) Propensity Score

Propensity scores (PS) are estimated using a logit or probit regression model and are essentially probabilities which predict the likelihood that individuals are assigned to the treatment group, compared to the control group, conditional on the values of the observed covariates. The PS summarises all the background covariates for each patient into a single-index variable (the propensity score). The generation of one single characteristic allows the assessment of whether the treated and control groups overlap enough on background characteristics. When such overlap is present, the PS approach allows calculation of the estimated treatment versus control effects that reflect adjustment for differences in all observed background characteristics [Rubin, 1997]. In other words, selection bias is removed when comparisons are made between groups with similar propensity scores. Propensity scores have the advantage of reducing bias in observed differences, but remain subject to bias from unobserved differences [Rubin, 1997; Laudrum, 2001].

Once propensity scores have been obtained for each patient, there are three common ways to help control for selection bias: 1) matching - the idea behind PS matching method is to

match each patient who receives the treatment to a control patient with a similar PS. Matching on propensity scores is a way of matching on many variables indirectly, instead of matching directly on many variables, which becomes increasingly more difficult with more variables. The most common method of matching is nearest neighbour matching. This is when patients in the intervention group are matched to patients in the control group who have the closest PS; 2) stratification (subclassification) – the idea is to create groups (or strata) that are similar in terms of background characteristics. Once the strata are defined, treated and control patients who are in the same group or stratum are compared directly [D'Agostino, 1998]; this is done for each stratum and usually the mean propensity scores for each group are compared; and finally 3) regression adjustment[1] - where PS are added as an independent variable into the regression model. Inclusion of PS as a covariate in the regression model takes into account the likelihood for treatments and thereby controlling for selection bias.

3) Instrumental Variables

Instrumental variables (IV) are another method to control for selection bias and accounts for selection bias in both observed and unobserved covariates. In order to conduct an IV analysis you have to find one or more variables (instruments) that have two essential properties. First, they should be related to the choice of treatment; the higher the correlation between the instrument and the treatment variable, the better the instrument. Second, the instrument should be uncorrelated or have no direct effect on the outcome. If the instrument satisfies both properties then the instruments are said to be good and consistent. The IV analysis is done in two stages. In the first stage, a regression model is conducted and in the second stage, the IV regression is carried out using the predicted probabilities obtained from the first regression instead of the observed values. This method estimates an 'incremental' or 'marginal' effect of treatment only over a range of variation in the treatment across the IV groups.

4) Sample selection models

Sample selection models (SSMs) such as the Heckman model and the Lee model control for selection bias due to both observed and unobserved factors associated with the treatment selection, which may also be correlated with the outcome variable of interest. In the first

[1] Inverse probability weighting or a marginal structural model is a variant of regression adjustment for propensity scores. This method is similar to the propensity score method, whereby the weighting is by the inverse probability of receiving treatment [Rosenbaum, 1987]. An individual is assigned a weight, equal to the inverse of the conditional probability of receiving his or her own treatment. However, it is simply not the inverse of the propensity score. Specifically, the weight is the inverse of the propensity score for treated subjects and it is the inverse of 1 minus the propensity score for untreated subjects.

stage, a probit model (for the Heckman model [Crown et al, 1998]) or a logit model (for the Lee model [Lee, 1983]) of treatment selection is estimated. Then the estimated probabilities from this probit or logit model, are used to calculate a new variable (an adjustment factor sometimes referred to as an inverse Mills ratio) for each patient which is the probability of not receiving the treatment given that the individual was 'at risk' of receiving the treatment. In the second stage, we want to predict the outcome of interest, and capture the factors which may influence the outcome. This adjustment factor serves as a direct test of whether selection bias is present and if so, what the direction of its impact is [Crown, 2001].

Table 2 highlights the key requirements for each method; strengths and weaknesses of each of the methods are identified; whether any of the methods are similar or different to other methods identified; and whether the method works best in small or large sample sizes. Briefly, starting with regression analysis: the method assumes that relationships are linear, whereas with the other methods you do not need to specify any functional forms; the method is easy to understand, explain and has more power than IVs and SSMs; and the method only controls for selection bias in observed variables and not in unobserved variables.

For the propensity score method: it summarises all background characteristics into one score, whereas the other methods uses all individual variables; requires some degree of "overlap" between the groups observations (i.e. 'balances out' the groups being compared), whereas for the other methods this is not necessary; PS models may have lower standard errors than other models; PS may not use all observations for the analysis, therefore leading to a smaller sample size; as the variables are combined into one variable i.e. the propensity score it may obscure important interactions, whereas with the other methods you can see which variables have an effect on treatment; rests on the assumption of strong ignorability whereas the other methods don't require strong assumptions (apart from SSMs); they are easier to understand and explain i.e. using histograms than IVs or SSMs; and once PS are obtained can be used in three ways - matching, stratification and in a regression model; and the method only deals with selection bias in observed variables and not unobserved variables.

For the IV method: the method requires a strong instrument, whereas the other methods don't need one; the instrument in the IV analysis is uncorrelated with the error terms or has no direct effect on outcome, whereas with PS method it has the opposite effect, the probability of treatment is highly correlated with the outcome term; SSMs with valid exclusion restrictions (variables that predict treatment but not the outcome) and IV models with valid

instruments (same thing) are closely related; and the method controls for selection bias in both observed and unobserved variables.

For SSMs: they require some strong joint distributional assumptions, whereas the other methods do not (apart from PS which requires the strong ignorability assumption); the method constructs an adjustment factor (lambda), the other methods don't compute such a value i.e. to tell you whether selection bias exists in the data; may not be as robust as PS; if the joint distributional assumption holds, SSMs are more efficient than IV; and the method deals with selection bias in observed and unobserved variables.

Discussion

To obtain unbiased estimates of cost and effect differences, we need large, adequately powered randomised controlled trials. However, this is not always possible and non-randomised studies are often used. Due to the lack of randomisation, there is a potential for selection bias. Also, non-randomised studies which are small often raise statistical issues concerning the robustness, and thus the usefulness of the data. This paper has tried to identify various methods that can be used to control for selection bias in non-randomised studies.

Some of the methods identified in the literature search are similar in the way they can be applied to the data. For each of these models (PS, IV, and SSMs), you first conduct a linear prediction model to obtain the predicted probabilities. It is in the second stage where the key differences in the models lie. For example, whether or not you include the predicted probabilities and how you might apply these predicted probabilities in the second stage.

After identifying these methods, how best can they be applied to a non-randomised dataset; and what criteria is needed to judge which method is the most appropriate in controlling for selection bias? First, we need to look at the observed cost and effect results and then to apply each of these methods to the dataset, and then see which method provides the closest estimates to the observed (unadjusted) cost and effect data.

Some of the key criterion for judging which method might be the most appropriate are summarised below:

a) Need to identify three types of variables(s) that are required for such an analysis: 1) variables that only affect costs and/or effects but not the dependent variable; 2)

variables that only affect the dependent variable but not costs and/or effects; and 3) variables that affect costs and/or effects and the dependent variable.

b) What sample sizes do each of the methods work best with? Need some sort of cut-off point to define small, medium and large sample sizes.
c) Do the methods work best for observed or unobserved or for both types of variables?
d) What tests can be applied to see whether selection bias has been controlled for or reduced by each of these methods?

It is difficult to say which of these methods is more accurate in controlling for selection bias in non-randomised studies, here are some pointers. The multiple regression and propensity score method seem the easiest to apply, as it requires less assumptions to be made. However, both of these methods only control for observed variables and may leave out some key unobserved variables which may play a role in the selection problem. With propensity score matching, you are looking at a binary treatment (yes or no) and this may be better to use for healthcare studies than using regression analysis, as you are matching patients with similar scores, therefore you are not likely to be affected by outliers. Also, with propensity scores you are looking at 'individual' rather than 'average' level effects. Instrumental variables analysis seem a good method to apply because it controls for both observed and unobserved variables, but finding good instruments may be difficult. Sample selection models also control for both observed and unobserved variables, and can be applied to the same situations as propensity score, however, it requires variables not to be included in both equations to avoid problems of multicollinearity and requires some strong assumptions to be made concerning the error term.

The next stage is to compare these methods identified using non-randomised datasets, both small and large.

Issues for discussion

- Are there other types of selection bias?
- Are there any other methods for dealing with selection bias?
- Are there any other costs or benefits of the different methods discussed which have been missed?
- Any other criteria for judging which method is the most appropriate?
- What method(s) are most appropriate when sample sizes are small?

References

1. Black N (1996). "Why we need observational studies to evaluate the effectiveness of health care." British Medical Journal 312: 1215-1218.
2. Cochrane Collaboration Handbook (2005). 6. Assessment of study quality. Cochrane Handbook for Systematic Reviews of Interventions. Higgins JPT and Green S: 79-89.
3. Crown WH (2001). "Antidepressant selection and economic outcome: a review of methods and studies from clinical practice." The British Journal of Psychiatry 179(S2): S18-S22.
4. Crown WH, Obenchain RL, et al. (1998). "The application of sample selection models to outcomes research: the case of evaluating the effects of antidepressant therapy on resource utilization." Statistics in Medicine 17: 1943-1958.
5. D'Agostino RB (1998). "Propensity score methods for bias reduction in the comparison of a treatment to a non-randomised control group." Statistics in Medicine 17: 2265-2281.
6. Deeks JJ, Dinnes J, et al. (2003). "Evaluating non-randomised studies." Health Technology Assessment 7(27).
7. Delgado-Rodriguez M and Llorca J (2004). "Bias." Journal of Epidemiology and Community Health 58: 635-641.
8. Earle CC, Tsai JS, et al. (2001). "Effectiveness of chemotherapy for advanced lung cancer in the elderly: instrumental variable and propensity analysis." Journal of Clinical Oncology 19(4): 1064-1070.
9. Feinstein AR (1985). An outline of cause-effect evaluations. Clinical Epidemiology - Architecture of Clinical Research. W. S. Company. Philadelphia, WB Saunders Company: 39-52.
10. Kunz RA and Oxman AD (1995) Empirical evidence of selection bias in studies of the effects of health care: a systematic review. Presented at the Cochrane Colloquium, Oslo, 5-8 October 1995
11. Landrum MB and Ayanian JZ (2001). "Causal effect of ambulatory speciality care on mortality following myocardial infarction: a comparison of propensity score and instrumental variable analyses." Health Services and Outcomes Research Methodology 2(3-4): 221-245.
12. Lee LF (2003). “Generalized econometric models with selectivity.” Econometrica 51(2): 507-512.
13. Morris S (In press). “The impact of obesity on employment.” Labour Economics.
14. Posner MA, Ash AS, et al. (2001). "Comparing standard regression, propensity score matching, and instrumental variables methods for determining the influence of

mammography on stage of diagnosis." Health Services and Outcomes Research Methodology 2: 279-290.

15. Rosenbaum PR (1987). "Model-based direct adjustment." Journal of the American Statistical Association 82(398): 387-394.
16. Rubin DB (1997). "Estimating causal effects from large data sets using propensity scores." Annals of Internal Medicine 127(8S): 757-763.
17. Sackett DL (1979). "Bias in analytic research." Journal of Chronic Diseases 32: 51-63.
18. Shah BR, Laupacis A, et al. (2005). "Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review." Journal of Clinical Epidemiology 58: 550-559.
19. Zanutto EL (2006). "A comparison of propensity score and linear regression analysis of complex survey data." Journal of Data Science 4: 67-91.

Appendix 1: Search terms

- Selection bias AND non-random$
- Selection bias AND non-experiment$
- Selection bias AND quasi-experiment$
- Selection bias AND observational
- Susceptibility bias AND non-random$
- Susceptibility bias AND non-experiment$
- Susceptibility bias AND quasi-experiment$
- Susceptibility bias AND observational
- Bias selection AND non-random$
- Bias selection AND non-experiment$
- Bias selection AND quasi-experiment$
- Bias selection AND observational

Table 1: Methods identified to control for selection bias in articles

Method to control for selection bias	**Frequency**
Multiple regression analysis	13
Propensity score:	
Matching	17
Stratification	7
Regression adjustment	3
Instrumental variables analysis	15
Sample selection models i.e. Heckman	22
Inverse probability weighting	3
Article not relevant (no methods were identified)	6

Table 2: Summary of methods identified to control for selection bias (SB)

Method	Main idea behind method	Does the method control for SB?	Key requirements	Why is this method different from the others?	Is this method similar to other methods
Regression analysis	Estimates how each independent variable relates to the dependent variable	Yes	- Independent variables included should not be highly correlated - Assumes linearity - Only works if relevant baseline factors have been measured	Regression model uses all observations/variables	Yes – PS regression adjustment if all the same variables are used, doesn't include the PS in regression equation, however, results should be similar
Propensity score matching	- Reduces all background characteristics into a single-index variable (the propensity score). Then subjects can be matched based on their propensity score. - There are various matching schemes	Balancing the covariates in the two groups	- Groups must overlap enough before matching takes place. - Rests on the assumption of "strong ignorability" - Used with dichotomous regressors	Creating a single-index variable makes it easier to match subjects on one variable, than matching across various variables.	- Similar to the IPW method in estimating treatment probabilities - Similar to SSMs
Propensity score - stratification method	- Reduces all background characteristics into a single-index variable (the propensity score). - Once propensity scores are obtained, strata or groups are formed, so subjects in each stratum can be compared directly.	Balancing the covariates in the two groups	- Groups must overlap enough before stratification takes place. - Rests on the assumption of "strong ignorability" - Rosenbaum and Rubin (1984) suggest that 5 strata (or groups) are sufficient to remove 90% of the bias. - Used with dichotomous regressors	- Creating a single-index variable makes it easier to compare and create strata. - Method can be used also for missing data imputation.	- Similar to the IPW method in estimating treatment probabilities - Similar to SSMs
Propensity score - regression adjustment	- Reduces all background characteristics into a single-index variable (the propensity score). - Then the propensity scores are added as an independent variable in the regression equation.	Balancing the covariates in the two groups	- Groups must overlap enough before regression is undertaken. - Rests on the assumption of "strong ignorability" - Used with dichotomous regressors	Creating a single-index variable makes it easier to add a single variable into the regression model than various other independent variables.	- Similar to the IPW method in estimating treatment probabilities - Similar to SSMs - Similar to regression analysis if all the same variables are used, and results should be similar
Instrumental variables analysis	In the first stage a logistic regression is conducted using both the independent variables and the instruments to obtain predicted	Yes	Instrument(s) must satisfy two criteria: 1) the instrument(s) should be highly correlated with treatment variable; 2) but	IV is a device that aims to achieve pseudo randomisation	Can be similar to SSMs if the instruments are valid.

	probabilities. In the second stage, another regression is conducted using the predicted probabilities instead of the actual values.		not correlated (or have direct effect) with outcome variable		
Sample selection models (SSM) i.e. Heckman	In the first stage, a probit model of treatment selection is estimated and the errors obtained are used to construct an adjustment factor. In the second stage, this factor is included as one of the variables in the outcome model.	Yes - The adjustment factor permits a direct test of whether selection bias is present and if so, what the direction of its impact is.	- If variables are different in the two stages, then SSMs will be more effective in controlling for SB - Need some valid exclusion restrictions (variables that predict treatment but not outcome) - Requires some joint distributional assumptions	The adjustment factor identifies the existence of selection bias by testing its individual significance. If there is selection bias, then the inclusion of the adjustment factor in the second stage controls for the selection bias.	-Are similar to IV analysis if it has valid exclusion restrictions. - Is similar to PS methods

Method	**What are the main weaknesses?**	**What are the main strengths?**	**Tests needed to check whether bias is reduced or eliminated**	**Does the method work well in large or small sample sizes or both?**	**How often is the method used*? In what situations is this method used?**
Regression analysis	- Can't control for unobserved variables, only controls for observed variables - Can't be sure that the covariates among the two groups are balanced (i.e. variables may not overlap enough) - The regression may not be as robust if variables are omitted (omitted variable bias) -If there is multicollinearity between independent variable, this may increase the standard error of the estimates.	- Can use more variables than matching or stratification methods - Can be used to predict values - Estimates a 'mean effect' for each variable	- RESET test can be conducted to check that the model is not mis-specified (or suffer from omitted variables bias) - or Hosmer-Lemeshow test can be done to check how well the model fits the data for logistic regressions.	Works in both situations; works much better with a larger sample size.	- Standard practice - Applied in various health care situations such as stroke and pneumonia
Propensity score matching	- Can't control for unobserved variables (hidden bias), only controls for observed variables. - Handling of weak covariates - Matching on a one-to-many or many-to-one basis may reduce sample size. - Depending on the criteria for matching some subjects may not be matched (matching range may be too narrow).	- Easy to implement, are efficient and more robust. - Less sensitive to non-linearities - Can control observed variables better than some methods - Requires few assumptions - Can calculate separate propensity scores for each pair of treatments - Uses as many variables as possible	- No specific tests. - Can compare estimates with the original data.	Works in both situations; works much better with a larger sample size and only when a few covariates	- Frequently - Applied in various health care situations such as breast cancer, asthma and joint replacement surgery.
Propensity score - stratification method	- Can't control for unobserved variables (hidden bias), only controls for observed variables. - Handling of weak covariates	-Easy to implement, are efficient and more robust. - Less sensitive to non-linearities - Can control observed variables better than some methods - Requires few assumptions - Can calculate separate propensity scores for each pair of treatments	- No specific tests. - Can compare estimates with the original data.	This method works better with a larger sample size.	-Frequently - Applied in various health care situations such as anti depressants.
Propensity score - regression adjustment	- Can't control for unobserved variables (hidden bias), only controls for observed variables.	-Easy to implement, are efficient and more robust. - Less sensitive to non-	- No specific tests. - Can compare estimates with the original data.	This method can work in both situations, works	-More often - Applied in various health care situations

	- Handling of weak covariates	linearities - Can control observed variables better than some methods - Requires few assumptions - Estimates a 'mean effect' for each variable		much better with a larger sample size.	such as renal failure.
Instrumental variables analysis	- May be really hard to find variables which can be classed as good instruments. - Instruments found may be not be highly correlated (strong assumption), therefore IV estimates may be more biased than OLS estimates - Has less statistical power than regression analyses	Controls for both observed and unobserved variables	Hausman test to see whether IV or OLS is more efficient	This method works better with a larger sample size.	- Infrequently - Mainly used in econometrics, but has been applied in health care studies such as acute MI
Sample selection models (SSM) i.e. Heckman	-Only controls for the unobserved variables in the treatment process -Problems with multicollinearity if observed variables in two stages are the same - Requires strong assumptions to estimate the model	Controls for both observed and unobserved variables	A likelihood ratio test can be conducted for the model to see whether the model fits the data well.	This method can work in both situations, works much better with a larger sample size	- More often - Mainly used in econometrics but applied in health care studies such as antidepressants

* Not used, infrequently used, more often, frequently used, standard practice