Sickness Absence and the Effects of Having a Spouse - Can twins reveal the selection effect?

by William Nilsson[1]
Department of Economics, Umeå University, Sweden and
Centre for Research on Welfare Economics (CREB), Barcelona, Spain

Abstract

Individuals that are married are often found to be healthier than singles. A crucial issue is to distinguish if this is due to a selection effect or due to a true protective effect of partnership. The purpose of this study is to distinguish these effects as explanations for a lower risk of having long-term sickness among individuals with a spouse. In this study an innovative method based on information on twins is developed to reveal the selection effect into partnership that provides a lower risk for long-term sickness absence. Important selections are found for both male and female samples.

JEL: C35, I12, J12
Keywords: sickness absence, spouse, heterogeneity, selection, twins

Umeå Economic Studies 686, 2006

[1] I wish to thank workshop participants at the Department of Economics, Umeå University, Sweden. Financial support from the Swedish Council for Working Life and Social Research is gratefully acknowledged. (Grant number: 2005-1330). All data, except for the identification of the twin sample come from Statistics Sweden (SCB). The twin information comes from the Swedish Twin Registry (STR). The Swedish Twin Registry is supported by grants from the Swedish Research Council. To contact the author, write to william.nilsson@econ.umu.se.

1 Introduction

A growing empirical literature studies the effects, usually benefits, of marriage. For example, married men are often found to be wealthier, healthier, happier and live longer than singles. As Lundberg (2005) notes, it is still common to include "controls" for marital status in applied labor economics. Including marital status as an exogenous variable can, however, be a misspecification since being married is in fact a choice variable. With the growing literature on the benefits of marriage, focus on this issue has increased, as any estimate of the benefits of marriage could be largely misleading if individuals are selected into partnership in a non-random way (Lillard & Panis, 1996, Brockmann & Klein, 2004, Lundberg, 2005). It is possible that the married men tested would have been wealthy, healthy, and live longer even without being married. The crucial issue is to separate the causal effects of marriage from selection effects.

The purpose of this study is to investigate the effects of partnership on the chance to not needing long-term sickness benefits. A particular concern in the study is to model selection into partnership and to distinguish a selection effect from a causal effect. An additional purpose is to measure the degree of intergenerational transmission on the chance to not need long-term sickness benefits. That is, how important is the social background, including the genes, for sickness absence?

An important article concerning the marital status and selection effects is Lillard & Panis (1996). They suggest that if a protective effect actually exists, this will provide an incentive for the less healthy to actually enter into marriage to be able to benefit from the effect. They use a simultaneous equation framework to handle the endogeneity of marital status. Lillard & Panis (1996) found evidence of both positive and negative (adverse) selection into marriage and that the positive selection dominates. Further, when marital status is modeled as endogenous, they found no significant difference between never married and married men. They did, however, find a positive health effect of remarriage for divorced men.

Gardner & Oswald (2004) studied the effects of marriage on mortality for British data. They found large benefits of marriage, but interestingly the results do not suggest that the effect works through reduced stress.

Empirical studies that intend to estimate the benefits of marriage have to take into

account that marital status is endogenous. Instrumental variable methods rely on the availability of suitable instruments. In addition, these methods can fail to identify the causal effect if the effects of marriage are different for different individuals, and if individuals take these differences into consideration in the decision to marry (Heckman, 1997). Using longitudinal data to estimate fixed-effects model can also provide unreliable estimates. The method does not handle reverse causality and it is also possible that the selection effect, in fact, is a time variant unobserved factor.

This study uses an alternative and innovative method to identify the selection effect. The method takes advantage of twin data to identify the selection effect that has its origin in the social background and genes. The idea is to estimate the aggregate effect for one twin's, (twin A's), partnership on the chance of not needing the benefit of long-term sickness absence. This effect can include both a protective effect and a selection effect. The next step is to estimate the effect of whether his twin sibling, (twin B), has a partner or not, on his (i.e. twin A's) chance to not be in need of long-term sickness absence. Under the assumption that a partner only provides protection to her/his partner and not to the partner's twin sibling, any measured effect will be based on selection. Comparing the aggregate effect and the effect from the twin sibling's partner will give an upper bound for the true causal effect of being in a partnership.

In this study a multivariate probit model is estimated for male and female twins separately. The model includes partnership status and sickness absence status for both twins. The study contributes to the literature in the following ways: First, a new method based on twins is used to distinguish selection from protection effects of partnership on the chance to not need sickness absence. Second, the method also provides estimates for the overall effect of the common background, including genes, on sickness absence. This is important as unobserved heterogeneity is suggested to be important in the literature on health and socioeconomic status (Fuchs, 2004, Lundberg, 2005, Zimmerman & Katon, 2005). The results indicate both important effects of the common background and an important selection effect into partnership on the chance to not need long-term sickness absence.

The remainder of the paper is structured in the following way; Section 2 presents the model with theoretical foundation, econometrical specification and the method to

distinguish selection effects and protection effects with twin data. Section 3 explains the data and section 4 the results. Finally, concluding remarks can be found in section 5.

2 Model

In this section the model is described. First, the theoretical motives to observe a causal effect on health from marriage, as well as the selection effect, are explained. The theoretical foundation also includes an overview of how data on twins can be used to separate these effects. The econometric model is also explained in this section. Based on the econometric model, the method of using twins to separate selection effects from causal effects is finally detailed.

2.1 Theoretical foundation

A common observation is that married men tend to be healthier and live longer when compared to unmarried men. An important question is whether this observation is due to a protective effect from having a partner or if the observed relationship is due to a selection into partnership. This theoretical overview summarizes the motives for a protection effect as well as a selection effect.

First, marriage can protect against sickness through behavioral changes. It is, for example, possible that a partner provide both a control against an unhealthy risky lifestyle and support during stressful events. A caring partner can also be a partial substitute for professional health care. These explanations have, in the literature, been summarized as the "guardian role" in which the partner can provide a protection effect for better health. Interestingly, Gardner & Oswald (2004) did not found that the protection effect works through reduced levels of stress.

Secondly, having a partner usually means shared expenses and as a consequence a higher real income per partner. If income matters for health this could be a motive for a protection effect from partnership. It should, however, be noted that it is not necessary that income, *per se*, matters. This is, in fact, an empirical question that has to be modeled with care. The reason is the risk for mistakenly concluding a causal effect from income to health when a "third variable" or reverse causality could, in fact, be the reasons for a correlation

between income and health.

Apart from these motives for a protection effect from marriage, it is important to remember that entering into a marriage or a partnership depends both on an individual choice and the availability of a possible partner. If health is an important asset on the marriage market, healthy individuals are expected to more easily find a possible partner, as well as a higher chance to actually stay in the partnership. Lillard & Panis (1996) also note that if marriage actually protects against sickness, unhealthy individuals will have an incentive to find, and stay with a partner to take advantage of this effect. It is, accordingly, possible to observe an adverse selection into marriage. These two selection effects work in the opposite direction and refer to issues concerning demand and supply on the marriage market.

To be able to confirm a true protective effect from marriage, it is important to control for any possible selection effect into partnership. It is important to deal with unobserved heterogeneity as biased estimates otherwise can give misleading conclusions. It is not hard to suggest possible omitted variables that can explain both marital and health status. McGue and Lykken (1992) found, for example, that divorce risk is, "to a substantial degree genetically mediated". Johanson *et al*. (2004) found that "genetic influences on personality contribute to the propensity to marry". If lifestyle and risk-behavior are suggested as possible motives for a protection effect, it is of course also possible that genetics that affect personality also can influence behaviors related to health.

Wilson & Oswald (2005) explain in a survey, a common approach to deal with these problems in the literature for marital status and mental health. A measure for mental health before marriage has often been included as an additional control besides explanatory variables. With this method it is, however, important to include all relevant explanatory variables. There is also an important risk that the health measure observed before marriage is not sufficiently detailed. Both these problems could imply omitted variable bias and a failure to identify a true causal effect from marriage.

In the literature of marital status and mortality the selection effect has, apart from the method mentioned above, also been modeled separately. Lillard & Panis (1996) is an important example of the method. They specify equations for the risks of marriage formation and dissolution, apart from the main interest of health and mortality. All

equations are estimated simultaneously. Lillard & Panis (1996) found both positive and adverse health selection into marriage. They did not find any positive net effect of entry into first marriage, while remarriage of divorced men significantly increased self-reported health.

Kohler *et al*. (2005) study happiness, partnership and fertility status. They use variation between twins to control for unobserved heterogeneity. In this way they intend to reveal a causal contribution of marriage. They found that once current partnership is controlled for, the history of partnerships and separations is not found to be important. Kohler *et al*. (2005) also found that OLS overestimate the effect of current partnership for a female sample, while the opposite is the case for a male sample. It is, however, important to remember that this method relies on the fact that which twin actually is married is assumed to have been assigned randomly. If initial subjective well-being differs between twins, the twin method does not necessarily mean a smaller bias than OLS. If within twin difference in initial subjective well-being is an important determinant of whom among the twins will be married, the method can actually give more biased estimates than OLS. This criticism, or at least caution, is also present in the literature that uses twins to handle ability bias in estimation of the returns to schooling (Neumark, 1999). As in that literature, it is unfortunately difficult to actually test if this is the case, since ability, as well as initial subjective well-being, is unobserved, and in fact the motive to use twins in the first place! In the next section another way to use twin data, with a different method to identify the causal effect, is explained.

2.2 Econometric model

The purpose of this study is to distinguish selection effects from a true causal effect of partnership on a reduced risk of long-term sickness. To do this, data on twins is used. In this study, a four-variate probit model is estimated where the risk for sickness absence and the chance to have a partner is estimated for both twins simultaneously. In this way the estimates are conditioned on unobserved heterogeneity that is correlated between the twins. Apart from this way of controlling for unobserved heterogeneity, it is possible to identify the selection effect based on a common background and genes. If the probability for one twin to have sickness absence is less, given that his twin sibling had a spouse, this has to be

due to selection, as a protection effect is highly unlikely to be present.

In addition to information on twin siblings, the data also includes information on sickness absence early in life, i.e. observations for one year when the twins were between 23 and 25 years old. A measure, although far from perfect, for early sickness absence can, accordingly, be included in the model. No information is, however, available for actual health status or measures for lifestyle. All estimations are accordingly interpreted as reduced form (Contoyannis *et al.*, 2004*)*. The strength of the estimation is, however, the possibility to see the relative importance of common background factors, i.e. the selection effect. As mentioned, the model consists of equations for both sickness absence and partner status. The latent propensity to have an identified spouse, M^*_{sit-1}, is assumed to be determined by;

$$M^*_{1it-1} = \boldsymbol{\gamma}'_1 \mathbf{m}_{1it-1} + \eta_{1i} + \xi_{1it-1} \quad (1)$$

$$M^*_{2it-1} = \boldsymbol{\gamma}'_2 \mathbf{m}_{2it-1} + \eta_{2i} + \xi_{2it-1} \quad (2)$$

$$M_{sit-1} = I(M^*_{sit-1} > 0)$$
$$s = 1, 2,\ i = 1,..., N,\ \text{and}\ t = 2,..., T$$

Explanatory variables are included in $\mathbf{m}_{sit-1}$ and $\boldsymbol{\gamma}_s$ are vectors of parameters to be estimated. Subindex $s = 1$ indicates the first group of twins, and $s = 2$ indicates the second group of twins, consisting of the sibling twins of $s = 1$. Subindex *i* refers to the pair of twins $i = 1,\ldots,N$ and $t = 2,\ldots,T$ refers to different periods. Individuals are either observed to have a spouse ($M_{sit-1} = 1$), or not ($M_{sit-1} = 0$). $I(M^*_{sit-1} > 0)$ is an indicator function which takes the value 1 if the inequality is satisfied, and zero otherwise. The error term, v_{sit-1}, includes an individual specific effect, η_{si}, and an orthogonal white noise, ξ_{sit-1}. The error term is assumed $v_{sit-1} \sim N(0,1)$.

The latent sickness absence propensity, H^*_{sit}, is assumed to be determined by;

$$H^*_{1it} = \boldsymbol{\beta}'_1 \mathbf{x}_{1it-1} + \mu_{1i} + \varepsilon_{1it} \quad (3)$$

$$H^*_{2it} = \boldsymbol{\beta}'_2 \mathbf{x}_{2it-1} + \mu_{2i} + \varepsilon_{2it} \quad (4)$$

$$H_{sit} = I(H^*_{sit} > 0)$$
$$s = 1,2,\ i = 1,...,N,\ \text{and}\ t = 2,...,T$$

Individuals are observed to receive sickness absence during the year, $H_{sit} = 1$, or not, $H_{sit} = 0$. $I(H^*_{sit} > 0)$ is an indicator function which takes the value 1 if the inequality is satisfied, and zero otherwise. Explanatory variables that are expected to affect the probability to have long-term sickness absence are included in $\mathbf{x}_{sit-1}$, and $\boldsymbol{\beta}_s$ are vectors of the parameters to be estimated. One of the explanatory variables is M_{sit-1}. In the estimation, there is no reason to believe that $\boldsymbol{\beta}_1$ should differ systematically from $\boldsymbol{\beta}_2$. The same concerns γ_1 and γ_2 and it is, accordingly, possible to constrain $\boldsymbol{\beta}_1$ to be equal to $\boldsymbol{\beta}_2$ and γ_1 to be equal to γ_2. u_{sit} is an error term which includes μ_{si}, an individual specific effect, and ε_{sit}, an orthogonal white noise error. The error term is assumed to follow a standard normal distribution, $u_{sit} \sim N(0,1)$.

Many of the variables included in $\mathbf{m}_{sit-1}$ are also used $\mathbf{x}_{sit-1}$. It is however appropriate to include variables in the equations for the spousal status (1 and 2) that can be excluded from equation 3 and 4. Otherwise identification would only rely on the presence of a nonlinear functional form. Finding instruments is, however, difficult, keeping in mind that variables for parents' socioeconomic situation both can affect health and the probability of having a spouse. In this study, early marital status is used with the assumption that this does not affect the risk of sickness absence necessity when current spousal status is taken into account. In the empirical section the validity of the suggested instruments are tested given the presence of a nonlinear functional form.

The joint distribution of the error terms, $v_{1it-1}, v_{2it-1}, u_{1it-1}, u_{2it-1}$ is assumed to have the correlation matrix Σ. The correlations between the error terms for the different

equations are labeled according to;

$$Corr\begin{pmatrix} v_{1it-1},\ v_{1it-1} & & & \\ v_{2it-1},\ v_{1it-1} & v_{2it-1},\ v_{2it-1} & & \\ u_{1it},\ v_{1it-1} & u_{1it},\ v_{2it-1} & u_{1it},\ u_{1it} & \\ u_{2it},\ v_{1it-1} & v_{2it-1},\ v_{2it-1} & u_{2it},\ u_{1it} & u_{2it},\ u_{2it} \end{pmatrix} \equiv$$

$$\begin{pmatrix} 1 & & & \\ \rho_{21} & 1 & & \\ \rho_{31} & \rho_{32} & 1 & \\ \rho_{41} & \rho_{42} & \rho_{43} & 1 \end{pmatrix}$$

The subindeces indicate between which two equations the correlation refers. The equations are estimated simultaneously, and the correlations between the error terms are also parameters to estimate. The log-likelihood function for each pair of twins, $i = 1,\ldots,N$ and $t = 2,\ldots,T$, is;

$$\log L_i = \log \Phi_4(\mu_i;\Omega) \tag{5}$$

where $\Phi_4(\mu_i;\Omega)$ is a standard four-variate normal cdf, with

$$\mu_i = \{K_{i1}\boldsymbol{\gamma}_1'\mathbf{m}_{i1t-1},\ K_{i2}\boldsymbol{\gamma}_2'\mathbf{m}_{i2t-1},\ K_{i3}\boldsymbol{\beta}_1'\mathbf{x}_{i1t-1},\ K_{i4}\boldsymbol{\beta}_2'\mathbf{x}_{i2t-1},\}$$

where $K_{i1} = 2M_{i1t-1} - 1$, $K_{i2} = 2M_{i2t-1} - 1$, $K_{i3} = 2H_{i1t} - 1$, and $K_{i4} = 2H_{i2t} - 1$.

The matrix, Ω is symmetric;

$$\Omega = \begin{pmatrix} 1 & K_{i2}K_{i1}\rho_{21} & K_{i3}K_{i1}\rho_{31} & K_{i4}K_{i1}\rho_{41} \\ K_{i2}K_{i1}\rho_{21} & 1 & K_{i3}K_{i2}\rho_{32} & K_{i4}K_{i2}\rho_{42} \\ K_{i3}K_{i1}\rho_{31} & K_{i3}K_{i2}\rho_{32} & 1 & K_{i4}K_{i3}\rho_{43} \\ K_{i4}K_{i1}\rho_{41} & K_{i4}K_{i2}\rho_{42} & K_{i4}K_{i3}\rho_{43} & 1 \end{pmatrix}$$

A simulation method based on the GHK (Geweke-Hajivassiliou-Keane) simulator is used to evaluate the multivariate standard normal distribution function. The multivariate probit model is estimated with a Stata program written by Cappellari & Jenkins (2003). (Stata users can obtain the program by typing -findit mvprobit- at the Stata prompt). Before turning to the estimations it is, however, important to explain how the model can be used to distinguish protection effects from selection effects.

2.3 Distinguishing true marriage effect from selection

Two important probabilities are used to measure the aggregate effect of having a spouse present on the risk of needing sickness absence. First, the probability of not having sickness absence, given that the individual had a spouse present the previous period is;

$$\Pr(H_{sit} = 0 \mid M_{sit-1} = 1) = \frac{\Phi_2(-\boldsymbol{\beta}'_{s1}\mathbf{x}_{sit-1}, \boldsymbol{\gamma}'_s\mathbf{m}_{sit-1}; \rho)}{\Phi(\boldsymbol{\gamma}'_s\mathbf{m}_{sit-1})} \tag{6}$$

where $\rho = -\rho_{31}$ for $s = 1$ and $\rho = -\rho_{42}$ for $s = 2$. $\Phi(.)$ and $\Phi_2(.)$ are the cumulative density functions of univariate and bivariate standard normal distributions. This probability can be compared to the probability of not having sickness absence, given that the individual was not found to have a spouse;

$$\Pr(H_{sit} = 0 \mid M_{sit-1} = 0) = \frac{\Phi_2(-\boldsymbol{\beta}'_{s2}\mathbf{x}_{sit-1}, -\boldsymbol{\gamma}'_s\mathbf{m}_{sit-1}; \rho)}{\Phi(-\boldsymbol{\gamma}'_s\mathbf{m}_{sit-1})} \tag{7}$$

where $\rho = \rho_{31}$ for $s = 1$ and $\rho = \rho_{42}$ for $s = 2$.

Using these probabilities an Aggregate Marriage Effect, AME, is defined as;

$$AME_s = \left(\frac{\sum_{i \in (M_{sit-1}=1)} Pr(H_{sit}=0 | M_{sit-1}=1)}{\sum_i M_{sit-1}} \right) - \left(\frac{\sum_{i \in (M_{sit-1}=0)} Pr(H_{sit}=0 | M_{sit-1}=0)}{\sum_i (1-M_{sit-1})} \right)$$

AME measures the difference in probability to not need sickness absence depending on if a spouse was present or not. If a spouse reduces the risk of receiving sickness absence this could either be due to a True Marriage Effect (TME), i.e. a protective causal effect, or due to a selection effect. AME does not distinguish between these effects, but information on twins can be used to separate the effects. An important assumption is that if a twin has a spouse present this does not causally affects the risk for his/her twin sibling to receive sickness absence the following year. If this assumption is valid, any correlation that may be found is, accordingly, due to a similar family background that affects both the risk for sickness absence and the probability of having a spouse.

To identify the selection effect based on family background two more probabilities are necessary. The probability for the *first* twin to be without sickness absence, given that the twin sibling had a spouse present the previous year is,

$$\Pr(H_{1it} = 0 \mid M_{2it-1} = 1) = \frac{\Phi_2(-\boldsymbol{\beta}_1' \mathbf{x}_{1it-1}, \boldsymbol{\gamma}_2' \mathbf{m}_{2it-1}; -\rho_{32})}{\Phi(\boldsymbol{\gamma}_2' \mathbf{m}_{2it-1})} \tag{8}$$

If the interest is to calculate the probability that the *second* twin is without sickness absence, conditional on whether the *first* twin had a spouse, changes in explanatory variables and parameters are obvious. Note that ρ_{41} would be used instead of ρ_{32}. The probability of being without sickness absence, conditional on whether the twin sibling did not have a spouse present the previous year can be calculated as;

$$\Pr(H_{1it} = 0 \mid M_{2it-1} = 0) = \frac{\Phi_2(-\boldsymbol{\beta}_1' \mathbf{x}_{1it-1}, -\boldsymbol{\gamma}_2' \mathbf{m}_{2it-1}; \rho_{32})}{\Phi(-\boldsymbol{\gamma}_2' \mathbf{m}_{2it-1})} \tag{9}$$

Similar to the Aggregate Marriage Effect these two probabilities can be compared for two different groups.

$$M\_twin_1 = \left(\frac{\sum_{i\in(M_{2it-1}=1)} Pr(H_{1it}=0|M_{2it-1}=1)}{\sum_i M_{2it-1}} \right) - \left(\frac{\sum_{i\in(M_{2it-1}=0)} Pr(H_{1it}=0|M_{2it-1}=0)}{\sum_i (1-M_{2it-1})} \right)$$

M_twin measures the difference between the average probability of not having sickness absence for those who had a twin sibling with a spouse the previous year and the average probability of not having sickness absence for those who had a twin sibling without a spouse the previous year. Since it is not expected to find a protection effect that goes from the partner of one twin to his twin sibling, any difference between these averages should be a selection effect. This selection effect is positive if the common family background influences the chance of having a present spouse. To identify the upper bound of the True Marriage Effect it is possible to compare the Aggregate Marriage Effect and the lower bound for the selection effect identified with the twins.

$$AME_1 = TME_1 + M\_Twin_1 \tag{10}$$

If the common background, i.e. selection, is important, M_Twin will be close to AME and the True Marriage Effect would, as a consequence, be very small. If, on the other hand, M_Twin is small relative to AME, this would indicate a large causal effect of having a spouse. Of course, this method can only capture selection based on common characteristics and shared experiences. The experiences that differ even for identical twins are not included, and the selection effect that is identified has to be interpreted as a lower bound.

3 Data

In this study a sample of twins born between 1949 and 1958 is used. The information for the identification of the twin sample and information of whether the twins are monozygotic (identical) or not, came from the Swedish Twin Registry (STR). Statistics Sweden (SCB) has identified a possible spouse for each of the years 1994 to 1999. A spouse is identified if the twin is either married or living with a spouse with whom the twin has at least one common child. Biological parents and siblings are attached to the data, as well as social parents and social siblings that are found in the nationwide Population and Housing Census, for the years 1960, 1965, 1970 and 1975. Explanatory variables found in the longitudinal database LOUISE are attached to all included individuals for the period 1994-1999. In addition, several variables from the Income and Wealth Register are included for the years 1968, 71, 74, 77, 80 and 83. These variables are included both for the twins and their parents. For this study, the data is separated into subsamples of monozygotic and dizygotic twins with the same sex. The main focus is of course on monozygotic twins, but the method is also used on dizygotic twins. Different samples are also used for male and female twins. Summary statistics are included in Table 1.

[Table 1, about here]

Note that many variables measuring characteristics during childhood or early adulthood are measured at a particular age, rather than year of observation. Since data does not cover all years it is not possible to observe the variables at the exact same age for the different twins. If the variable is indicated for ages 12-16 this means that twins born in 1949 and 1954 were 16 years old. Twins born 1950 and 1955 were 15 years old, and so on. The censuses for 1965 and 1970 were, accordingly, used to construct these variables. The same method is used for variables included in the Income and Wealth Register, although the variation in age is smaller since more years are available. The income variables are relative income compared to the average in the municipality at the time. These variables are averaged over all available years where the income was observed. For example, a father that died and had a missing value in one occasion has an average over the two remaining observations.

The main interest in this study is the probability that an individual received sickness benefits, including benefits due to work related injuries, benefits due to rehabilitation or early pension. All of these welfare systems are only granted after medical confirmation of the individual's health status. Only sickness absence for more than 14 consecutive days is observed, i.e. the measure captures long-term sickness. Early pension is only granted if the reduced working ability is permanent. For the male sample about 12 percent received sickness related benefits. For the female sample about 18 percent had long-term sickness benefits. About 67 percent of the male sample had a spouse present according to the definition above. The corresponding number for the female sample is 69 percent.

Several variables capture whether the twins come from a home with divorced or unavailable parents. Information for the relative income of the biological father is also included to capture his socioeconomic status. The income is compared to the average income among individuals aged 30-60 in the municipality. The variable is an average over one to three observation when the twins were 17-25 years old. A dummy variable is also available that identifies whether the twins' oldest social sibling was a sister, when the twins were between 12 and 16 years old. Only cases where the sister was older than the twins are included. The idea is that an older sister could be an extra support for the parents in caring for the younger twins. Apart from these explanatory variables related to the family situation when the twins were young, a set of other variables are available for early labor market situation for the twins.

Income is available when the twins were between 17 and 25 years old. An average of relative income is constructed in the same way as for biological fathers. In addition, the amount of sickness benefits and Cash Unemployment Allowance, KAS, is available when the twins were between 23 and 25 years old. Unfortunately these different benefits are aggregated in the raw data from the Income and Wealth Register for the years 1974 and 1977. It is, accordingly, not possible to construct a variable that only includes sickness benefits. KAS, was for example applicable for individuals that were not a member of an unemployment insurance association, and did not qualify for unemployment insurance. The data for 1980 and 1983, when KAS and sickness benefits are not aggregated, reveal that KAS is both uncommon and small relatively to sickness benefits. Hereafter this variable will, accordingly, be interpreted as sickness benefits. Two dummy variables are

constructed to capture different degrees of sickness benefits. The first is whether the twin had a positive amount of sickness benefits, while the amount did not exceed 5000 SEK measured in the price level of 2001. The second dummy variable includes those who had a larger amount than 5000 SEK during the year. Another dummy variable is also available that indicates whether the twin received unemployment insurance.

In addition to these variables two different dummy variables are available for the level of education. These variables are measured between 1994 and 1999. The reference case is compulsory schooling. All of the above mentioned explanatory variables can explain both the risk for sickness absence and the chance to have a present spouse in adulthood. As mentioned, variables for early marital status are included to affect the probability to have a spouse in 1995 - 1998, while assumed to not affect the probability of sickness absence in adulthood when spousal status is included in the model. The first dummy variable captures those twins that were married between 20 and 22 years old. The second dummy variable includes those who were not married when observed between age 20 and 22, but were found to be married when observed between 23 and 25. As expected, a higher percentage was married at these ages for the female sample. These two variables are used to explain spousal status, while excluded from the equations explaining the probability to have sickness benefits.

4 Results

The results from the multivariate probit models are included in Table 2. Note that coefficients are presented and the magnitude of the effects is left aside in this study. Instead it will be noted if variables have significant negative or positive coefficients. This overview of the results does not discuss all coefficients since the main interest actually is on the estimates for the Aggregate Marriage Effect, and the True Marriage Effect.

[Table 2, about here]

An interesting result is that in the male sample, if the oldest sibling was a female sister, the probability to have a spouse is increased and the risk for being in need of sickness absence is reduced. It seems as if an older female sister can provide an

educational and protective effect on childhood. These effects are not present for the female sample. In fact, having a sister as the oldest sibling seems to increase the risk of having sickness absence in adulthood for the female sample. Having post secondary school or post graduate education, and/or a high relative averaged income when the twin was between 17 and 25 also provide a protective effect later in life. As expected these variables also increase the chance to have a spouse.

Interestingly, to have sickness absence of 1-5000 SEK during a year when the twin was between 23 and 25 increases the chance to have a spouse for the male sample. A higher level of sickness absence does not increase the probability. While it should be underlined that the coefficient is only significantly different from zero at the 10% level, this result indicates an adverse selection into partnership. Those who have a spouse and had a small amount of sickness benefits does also have a higher probability to have sickness absence in adulthood. This is not the case for those who did not have a spouse. A Wald test confirms that the coefficients are significantly different from each other at the 5 percent significance level. $(chi^2(1) = 4.64 > 3.84)$. This is another indication for an adverse selection, as it looks like those who had health problems that will affect future sickness absence, are more likely to be in partnership. On the other hand, for those, who do not have a partner, small health problems do not seem to affect the probability of sickness absence in adulthood. These differences are not present for those who received more than 5000 SEK as sickness benefits during a year. These results suggest that it is important in empirical studies to have a variable that can capture different degrees of health status early in life. These indications for adverse selection are not found for the female sample.

As expected, being married at age 20-22 or 23-25, increase the probability of having a spouse later in life. Likelihood-ratio tests confirm that these variables can be excluded for the sickness absence equation when a dummy for present spouse is included. Note that the coefficient for having a present spouse is not significantly different from zero in the equation for sickness absence, when the possible endogeneity is taken into account. This indicates that the causal effect is rather small.

It should be underlined that the coefficients are not interpreted as causal effects. No variables concerning health or health behavior are available and some variables are in fact based on individual choices. Education has, for example, been treated as endogenous when

estimating its effect on health in Auld & Sidhu (2005). In that study, the focus is to control for cognitive ability as it is expected to be correlated with both education and health, and, accordingly, could induce an omitted variable bias when the effect of education is estimated. While this could be a problem in this application, it should be noted that the model condition on correlated unobserved variation between the twins. Since monozygotic twins share the genes and the social background, the risk for ability bias, is less severe in this case. The correlations between the error terms are included in Table A1. The correlation between twins' status of sickness absence and also the correlation between the partnership status are positive and significantly different from zero. The correlations between the error terms for sickness absence and partnership for the same twin are negative, but never significantly different from zero.

[Table 3, about here]

Table 3 includes measures for the Aggregate Marriage Effect and the twin measures for selection. The aggregate marriage effect is found to be between 0.07 and 0.11 for the male sample. The probability of being without sickness absence is, accordingly, between 0.07 and 0.11 higher if the twin had a spouse present. These measures do not distinguish if this is due to selection or a true causal effect. The twin measure for selection is estimated to 0.05 for the male sample. Accordingly, between 46 and 75 percent of the higher probability of not needing sickness absence is found to be selection based on the common background, including genes. This is a lower bound for the selection effect, because any selection that is not common amongst the twins is not included. The causal effect is, accordingly, smaller than 24 to 54 percent. For the female sample, the lower bound for the selection effect is found to represent about 65 and 74 percent, and, thus, the upper bound for the causal effect is between 26 and 35 percent.

Included in Table 3 are also estimates on the aggregate background effect. This is estimated similar to M_twin but in this case the probabilities are conditioned on the sickness absence status of the twin sibling. The aggregate background effect measures the difference between the average probability of not having sickness absence for those who had a twin sibling without sickness absence and the average probability of not having

sickness absence for those who had a twin sibling with sickness absence. The more important the common background is for the risk of having sickness absence, the larger the difference between the averages. This measure is similar to measures of sibling correlation used in the literature for intergenerational income mobility (Solon, 1999). The estimates in Table 3 indicate an important influence of the background on the probability of not being in need of sickness absence. The measure is about 0.16 and 0.19 for the male sample and about 0.22 for the female sample.

In the lower part of Table 3 the corresponding measures are included for dizygotic sample. The lower bound for the selection effect is found to be between 0 and 12.8 percent for the male sample. The female sample has a lower bound of about 20 and 40 percent. The aggregate background effect is estimated to be between 0.09-0.14 for both the male sample and the female sample. These estimates are lower than for the monozygotic twins. It seems to be important to analyze monozygotic twins where genes are identical. The differences between the monozygotic and dizygotic sample suggest that genes matter both for the identification of the selection effect of partnership and the risk of sickness absence. Of course, differences between the samples can also occur if monozygotic twins are treated differently compared to dizygotic twins. It is, for example, possible that parents to monozygotic twins induce an extra effort in treating the twins the same way. For this reason it is difficult to confirm if it actually is the genes that explain all the difference between monozygotic and dizygotic samples. Remember, however, that dizygotic twins also share genes, which contribute to the estimates. The results do, accordingly, suggest an important role for genetics.

5 Concluding remarks

The purpose of this study is to distinguish selection effects and causal effects as explanations to why, in general, individuals with a partner have lower risk of long-term sickness. Entering into partnership is a decision based on individual characteristics and preferences. Further, the decision depends, crucially, on the availability of a possible partner. The probability to have a partner can be affected both positively and negatively by initial health status. The reason is that if having a partner in fact can improve the health status, individuals with health problems, can take advantage of this to a greater extent. At

the same time, being unhealthy or engaging in unhealthy activities, can be unattractive characteristics for possible partners, and can, accordingly, decrease the probability to find and/or keep a spouse.

In this study a multivariate probit model with equations for partnership and sickness absence status is estimated. Equations are included for both twin siblings to control for unobserved characteristics that are correlated for twins. In addition, using data on twins makes a new method available to reveal the selection effect based on the common social background and genes that work through partnership. It is unlikely that a protecting effect of one twin's partner is also present for his/her twin sibling. Accordingly, if the chance to not be in need of sickness benefits increases conditionally on whether the twin sibling had a spouse, this is due to a selection effect based on the common background and genes. The selection effect can, for example, capture personality traits that are an advantage for both staying healthy and to finding, and staying, with a partner. The selection effect that is identified is of course a lower bound for the selection effect, as, even for identical twins, differences can exist that induce a selection.

The results in this study indicate some adverse selection into partnership for the male sample. No such adverse selection is found for the female sample. The lower bound for the overall selection effect is found to explain between 46 and 75 percent of the lower risk of sickness absence for individuals with a partner for the male sample. The corresponding measures for the female sample are between 64 and 75 percent. The results, accordingly, indicate a large positive selection based on the social background into partnership. The causal effect from having a partner is, as a consequence, less than a simple overview of the data would indicate.

The empirical model is also suitable for calculating an aggregate measure of the importance of the common background for the chance to be without need of long-term sickness absence. The measure compares twins that have a twin sibling without sickness absence with the group of twins that have twin siblings with sickness absence. The difference between the probabilities to not be in need of sickness absence is about 0.16-0.19 for the male sample, and about 0.22 for the female sample. These measures indicate intergenerational transmission, or heritability, of the need of sickness absence. For dizygotic samples these measures are lower and this indicates that genetics matter for the

probability to not be in need of sickness benefits. Of course, it is difficult to be sure that all the difference is due to the genes as it is possible that monozygotic twins are in general treated more alike compared to dizygotic twins. The results, do, however, suggest that any study of individual health or sickness absence should be aware that genetics could be an important omitted variable that can cause biased estimates.

It is of course important to remember that the results in this study are found for a particular sample of twins. It is, for example, possible that different results would be found for different countries or even different age groups. The possibility to generalize the results depends, for this reason, on further empirical studies. In particular it would be interesting to see the twin method applied with different health indicators as dependent variable.

References

Auld. CM & Sidhu N. (2005) Schooling, cognitive ability and health. *Health Economics* **14**: 1019-1034.

Brockman H, Klein T. (2004) Love and Death in Germany: The Marital Biography and Its Effect on Mortality. *Journal of Marriage and Family* **66**: 567-581.

Contoyannis P. Jones AM. Rice N. (2004) The Dynamics of Health in the British Household Panel Survey. *Journal of Applied Econometrics* **19**: 473-503.

Fuchs, VR. (2004) Reflections on the Socio-economic Correlates of Health. *Journal of Health Economics* **23**: 653-661.

Gardner J. Oswald A. (2004) How is mortality affected by money, marriage, and stress? *Journal of Health Economics* **23**: 1181-1207.

Heckman JJ. (1997) Instrumental Variables: A Study of Implicit Behavioral Assumptions in One Widely Used Estimator, *Journal of Human Resources* **32**(3), 441-462.

Jockin V, McGue M, Lykken DT. (1996) Personality and Divorce: A Genetic Analysis. *Journal of Personality and Social Psychology* **71**(2): 288-299.

Johnson W, McGue M, Krueger RF., Bouchard Jr. TJ. (2004) Marriage and Personality: A Genetic Analysis. *Journal of Personality and Social Psychology* **86**(2): 285-294.

Kohler H-P, Behrman JR., Skytthe, A. (2005) Partner + Children = Happiness? The Effects of Partnerships and Fertility on Well-Being. *Population and Development Review* **31**(3): 407-445.

Lillard LA, Panis CWA. (1996) Marital Status and Mortality: The Role of Health. *Demography* **33**(3): 313-327.

Lindahl M. (2005) Estimating the Effect of Income on Health and Mortality Using Lottery Prizes as Exogenous Source of Variation in Income. *Journal of Human Resources* **40**(1): 144-168.

Lundberg S. (2005) Men and islands: Dealing with the family in empirical labor economics. *Labour Economics* **12**: 591-612.

Neumark D. (1999) Biases in twin estimates of the return to schooling. *Economics of Education Review* **18**: 143-148.

Solon G. (1999) Intergenerational Mobility in the Labor Market. In *Handbook of Labor Economics*, volume 3A, Ashenfelter O, Card D (eds.). Elsevier Science: Amsterdam, 1761-1800.

Wilson CM, Oswald AJ. (2005) How Does Marriage Affect Physical and Psychological Health? A Survey of the Longitudinal Evidence, *IZA Discussion Paper*: No. 1619.

Zimmerman FJ, Katon W. (2005) Socioeconomic Status, depression disparities, and financial strain: what lies behind the income-depression relationship? *Health Economics* **14**: 1197-1215.

Table 1 – Summary Statistics

Monozygotic sample	Male sample		Female sample	
Variable	Mean	Std.dev.	Mean	Std.dev.
Received long-term sickness related benefits	0.1215	0.3267	0.1792	0.3835
Spouse present	0.6716	0.4697	0.6920	0.4617
Social mother was biological mother, 12-16	0.9388	0.2397	0.9336	0.2491
Neither social father nor social mother were identified, 12-16	0.0931	0.2906	0.1018	0.3025
Biological mother was not in labor market, 12-16	0.5374	0.4987	0.5176	0.4997
Biological father or mother was divorced, 20-22	0.0849	0.2788	0.1020	0.3027
Biological father not identified or dead before twin turned 25	0.0755	0.2642	0.0775	0.2674
Biological mother not identified or dead before twin turned 25	0.0251	0.1564	0.0261	0.1596
Relative averaged income for biological father when twin was 17-25 [a]	1.4054	0.8350	1.3495	0.8400
Twin had older social sister, when 12-16 years old	0.2163	0.4118	0.1852	0.3885
Twin 41-45 years old [b]	0.4057	0.4911	0.4146	0.4927
Twin 46-50 years old [b]	0.3421	0.4745	0.3353	0.4721
Education, B (1=upper secondary school) [c]	0.4643	0.4988	0.4882	0.4999
Education, C (1=post secondary school and post graduate education) [c]	0.3025	0.4594	0.3469	0.4760
Relative averaged income for twin when twin was 17-25 years old [a]	0.6381	0.2885	0.5173	0.2474
Sickness benefits and Cash Unemployment Benefits (KAS) < 5000 SEK (and > 0), twin between 23 and 25	0.3762	0.4845	0.3767	0.4846
Sickness benefits and Cash Unemployment Benefits (KAS) >= 5000 SEK, twin between 23 and 25,	0.2968	0.4569	0.3180	0.4658
Received unemployment insurance, twin between 23 and 25	0.0615	0.2402	0.0738	0.2614
Twin was married when aged 20-22	0.0205	0.1416	0.0781	0.2682
Twin was not married at age 20-22, but was married at age 23-25	0.0891	0.2849	0.1786	0.3831

Notes: a) The incomes are relative to the average income among 30-60 years old in municipality. b) The reference case for the twins' age is 36-40 years old. c) The reference case for education is compulsory school.

Table 2 – Estimates from four variate probit models.

Monozygotic sample	Male sample		Female sample	
	$M_{st-1} = 1$	$H_{st} = 1$	$M_{st-1} = 1$	$H_{st} = 1$
Variable, (measured at *t*-1)	$(\gamma_1 = \gamma_2)^a$	$(\beta_1 = \beta_2)^a$	$(\gamma_1 = \gamma_2)^a$	$(\beta_1 = \beta_2)^a$
Social mother was biological mother, 12-16	0.1258	-0.5824***	0.0884	-0.2895**
	(0.71)	(-3.30)	(0.46)	(-2.02)
Neither social father nor social mother were identified, 12-16	-0.3086*	-0.0059	0.1213	-0.1867
	(-1.66)	(-0.03)	(0.73)	(-1.62)
Biological mother was not in labor market, 12-16	-0.1528**	0.1726**	-0.0398	0.0120
	(-2.00)	(2.17)	(-0.56)	(0.20)
Biological father or mother was divorced, 20-22	-0.0340	0.0660	-0.3328**	0.3143***
	(-0.23)	(0.43)	(-2.49)	(2.78)
Biological father not identified or dead before twin turned 25	0.2730	-0.0123	-0.1106	0.0547
	(1.41)	(-0.08)	(-0.73)	(0.43)
Biological mother not identified or dead before twin turned 25	-0.2778	0.4587***	-0.0494	-0.4682**
	(-0.92)	(2.92)	(-0.18)	(-2.01)
Relative averaged income for biological father when twin was 17-25	0.1719***	0.0042	0.0923*	-0.0716*
	(3.25)	(0.08)	(1.82)	(-1.80)
Twin had older social sister, when 12-16 years old	0.1884**	-0.2394***	-0.0290	0.1526**
	(2.08)	(-2.75)	(-0.33)	(2.13)
Twin 41-45 years old	0.0902	0.0266	-0.0092	0.1145**
	(1.50)	(0.43)	(-0.16)	(1.99)
Twin 46-50 years old	0.1601*	0.0249	-0.1092	0.2806***
	(1.79)	(0.29)	(-1.30)	(3.72)
Education, B (1=upper secondary school, less than three years)	0.0861	-0.3711***	0.0615	-0.1946**
	(0.93)	(-4.43)	(0.67)	(-2.44)
Education, C (1=post secondary school and post graduate education)	0.3289***	-0.8078***	0.1745*	-0.3297***
	(2.86)	(-6.81)	(1.70)	(-3.68)
Relative averaged income for twin when twin was 17-25 years old	0.3960***	-0.7296***	0.3157**	-0.4292***
	(2.73)	(-4.51)	(2.26)	(-3.68)
Sickness benefits and Cash Unemployment Benefits (KAS) < 5000 SEK, twin between 23 and 25	0.1633*		-0.0034	
	(1.95)		(-0.05)	
Sickness benefits and Cash Unemp. Benefits < 5000 SEK (and > 0), twin between 23 and 25, if $M_{1t-1} = 1$		0.2807***		-0.0380
		(3.01)		(-0.49)
Sickness benefits and Cash Unemp. Benefits < 5000 SEK (and > 0), twin between 23 and 25, if $M_{1t-1} = 0$		0.0220		0.0272
		(0.20)		(0.30)
Sickness benefits and Cash Unemployment Benefits (KAS) >= 5000 SEK, twin between 23 and 25,	0.0128		-0.0024	
	(0.14)		(-0.03)	
Sickness benefits and Cash Unemployment Benefits >= 5000 SEK, twin between 23 and 25, if $M_{1t-1} = 1$		0.3661***		0.2529***
		(3.45)		(2.67)
Sickness benefits and Cash Unemployment Benefits >= 5000 SEK, twin between 23 and 25, if $M_{1t-1} = 0$		0.3566***		0.2104***
		(3.41)		(2.61)
Received unemployment insurance, twin between 23 and 25	-0.0782	0.1570	0.1878	0.0123
	(-0.55)	(1.12)	(1.54)	(0.12)
Twin was married when aged 20-22	0.7937**		0.2954**	
	(2.26)		(2.38)	
Twin was not married at age 20-22, but was married at age 23-25	0.5949***		.3473***	
	(4.46)		(3.95)	
Spouse present ($M_{st-1} = 1$)		-0.1423		0.1344
		(-0.53)		(0.34)
Constant	-0.4373*	0.0009	0.0261	-0.4366
	(-1.78)	(0.00)	(0.11)	(-1.47)
Number of observations	3446		4215	
Log likelihood	-6330.44		-8745.39	

Notes: Estimated coefficients are found in the first line for each variable and second row includes *t*-ratios. The standard errors are corrected for repeated observations from the same twins over the years. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Likelihood ratio tests confirm that the instruments can be excluded from the equations for sickness absence. (χ^2(d.f.=2) = 2.43 < 5.99 and χ^2(d.f.=2) = 0.64 < 5.99). a) $\gamma_1 = \gamma_2$ and $\beta_1 = \beta_2$, i.e. the parameters are constrained to be the same for the spousal respective sickness absence equations.

Table 3 - Measures for Aggregate Marriage Effect and twin measure of selection

Measure	Male		Female	
	Twin 1	Twin 2	Twin 1	Twin 2
Panel A. Sample of monozygotic twins				
Aggregate Marriage Effect (AME)	0.1103	0.0717	0.1054	0.0845
Aggregate Background Effect on Sickness Absence	0.1577	0.1883	0.2199	0.2204
Twin measure of selection (*Twin_M*)	0.0512	0.0534	0.0781	0.0545
Share of AME due to selection (*Twin_M* / AME)*100;	46.4	74.6	74.1	64.6
Panel B. Sample of dizygotic twins				
	Twin 1	Twin 2	Twin 1	Twin 2
Aggregate Marriage Effect (AME)	0.1020	0.0613	0.0954	0.1085
Aggregate Background Effect on Sickness Absence	0.0937	0.1371	0.0878	0.1442
Twin measure of selection (*Twin_M*)	0.0130	-0.0139	0.0385	0.0223
Share of AME due to selection (*Twin_M* / AME)*100;	12.8	0 [a]	40.3	20.6

Note: a) The share of AME is set to zero as a consequence of the negative measure for *Twin_M*.

Table A1 - Correlations of error terms

Monozygotic sample		Male		Female	
		Estimate	t	Estimate	t
$M_{2t-1} = 1$ (twin 2), $M_{1t-1} = 1$ (twin 1)	ρ_{21}	0.3430***	6.61	0.4153***	8.85
$H_{1t} = 1$ (twin 1), $M_{1t-1} = 1$ (twin 1)	ρ_{31}	-0.2239	-1.34	-0.2842	-1.12
$H_{1t} = 1$ (twin 1), $M_{2t-1} = 1$ (twin 2)	ρ_{32}	-0.1046	-1.56	-0.1491*	-1.90
$H_{2t} = 1$ (twin 2), $M_{1t-1} = 1$ (twin 1)	ρ_{41}	-0.1123*	-1.78	-0.1197	-1.55
$H_{2t} = 1$ (twin 2), $M_{2t-1} = 1$ (twin 2)	ρ_{42}	-0.1126	-0.74	-0.2176	-0.95
$H_{2t} = 1$ (twin 2), $H_{1t} = 1$ (twin 1)	ρ_{43}	0.2669***	4.62	0.3505***	8.15

Note: Coefficients that are significantly different from zero at 10, 5 and 1 percent levels are marked with *, ** and ***.