Chapter 12

Does moderate alcohol intake reduce mortality?

Hans Olav Melberg

Introduction

Several studies have shown that those consuming moderate amounts of alcohol – less than a glass of wine, or a beer a day – have lower all-cause-mortality than abstainers and heavy consumers (Doll et al. 2005, Gronbaek et al. 1994, Ellison 2002, Corrao et al. 2000). This has been interpreted to imply that moderate alcohol consumption causes a reduction in mortality and it is sometimes labelled the J-curve hypothesis. In 1995, Ole-Jørgen Skog warned against this inference when he wrote: "I belong to the subspecies of stubborn alcohol researchers who still do not feel entirely comfortable with mainstream interpretations of the J-curve. The reason for my discomfort is 2-fold – one pertaining to the causal nature of the curve, the other being related to the public health implications of a J-shaped curve" (Skog 1995).

In this paper I shall follow up on some of the ideas presented in Skog's comment from 1995 and his article on the J-curve from 1996. More specifically, I shall explore a new approach to the problem of isolating causal relationships in non-experimental data. This approach will be used to test the effects of moderate drinking on mortality, with special attention to Skog's argument about the importance of controlling for social isolation and other confounders.

I shall argue – as Skog also did in 1995 and 1996 – that despite the numerous studies demonstrating a correlation, it is still too early to conclude that moderate consumption of alcohol causally reduces mortality. First of all, one should be cautious because there is good reason to suspect that the relationship is vulnerable to confounders. The more of these we control for, the smaller the difference between non-drinkers and drinkers of alcohol become

(Fillmore et al. 2006). In the tests carried out in this paper the relationship disappears after controlling for confounders. Second, I shall argue that there are still interesting statistical problems that need to be solved before more reliable tests can be carried out.

Background

The relationship between alcohol and mortality has been studied for at least one hundred years (For a survey of the historical literature, see Klatsky 2002). Initially some argued that alcohol consumption increased mortality by weakening the heart muscle. Later, the interest focused on the association between alcohol and vitamin deficiency and how this affected the functioning of the heart. In this early literature there was an implicit assumption of monotonicity. Since heavy alcohol consumption was bad, moderate alcohol consumption was assumed to be unhealthy

The fact that heavy consumption leads to higher mortality is not very strong evidence against moderate use of alcohol. This was noted quite early by Pearl (1926) who presented observational data indicting that moderate alcohol consumers had lower mortality rates than both abstainers and heavy consumers. The causal nature of this constant conjunction is open to at least two kinds of criticism. First of all, there could be confounding variables. Second, Pearl did not specify a causal mechanism. The following literature can be viewed as trying to close this gap by testing out various confounders, applying more sophisticated statistical methods, and trying to uncover the potential causal mechanisms.

As the literature grew, it became clear that the association between moderate alcohol consumption and reduced mortality was robust to several possible confounders. In a major article on the topic right before Skog's article, Rehm and Sempos (1995) concluded that neither age, nor gender reversed the relationship. Admittedly, the beneficial effects of alcohol seemed to be largest for middle-aged males and small for young females, but the relationship was not reversed. The relationship also held after controlling for smoking, previous alcohol consumption, and some dietary habits. More controversially, they also concluded that the J-curve was robust with respect to the so called "sick-quitter-hypothesis". In several articles, Shaper and Wanamethee (1988; 1989) had suggested that the J-curve could be explained by the fact that the abstainers included a significant group of individuals with diseases and that few previous studies had accounted for this. However, Rehm and Sempos noted that several articles had found that the ben-

eficial effects of moderate alcohol consumption remained even after controlling for diseases and other confounding factors.

In addition to the finding that the correlation is robust to several potential confounders, experimental studies have identified at least two possible causal mechanisms (Standridge, Zylstra and Adams 2004). First of all alcohol reduces the risk of heart problems because it raises the level of high density lipoprotein which, in turn, removes fat in blood vessels. Second, alcohol reduces coagulation, which in some cases may be beneficial for people with heart problems. At high consumption levels these beneficial effects are outweighed by an increase in other risks associated with alcohol such as cancer, liver problems and other alcohol-related diseases (Room, Babor and Rehm 2005), but at moderate levels there was experimental evidence for several causal mechanism implying that moderate amounts of alcohol could have positive health effects.

When Skog wrote his article, the evidence seemed to indicate that there really was a causal relationship, and therefore, it was quite a bold move to argue that he was not comfortable with mainstream interpretations. Still, the article had a major impact because it demonstrated both the underlying problems that led one to suspect the existence of further confounders, as well as bringing up a previously ignored issue: How research about the optimal level of alcohol consumption at the individual level could not simply be translated into recommendations about alcohol consumption at the aggregate level. This last contribution is important and it is frequently quoted by others. It implies that the optimal average consumption for a country will most likely be different from the optimal level of consumption for an individual. The reason is simply that heavy consumers are unlikely to cut their consumption down to the optimal level, and if the abstainers start to adjust their consumption upwards to the optimal level this will create a "wet culture" that, in turn, will lead some moderate consumers to become heavy consumers. Hence, to prevent too many heavy consumers, the optimal average consumption of alcohol in a country is likely to be lower than the optimal amount for an average individual. This is an interesting and important argument, but the focus in this paper is on the first of Skog's arguments – the problem of confounding.

Social integration as a confounder

People who do not drink alcohol typically do so for a reason. This makes it difficult to interpret observed correlations as causal relationships because the reasons for not drinking could be correlated with mortality. This is obvious

in the case of certain diseases. Having a disease is a reason why some people do not drink alcohol, as well as a cause of higher mortality. Less obviously, Skog argued that social factors also could create a spurious correlation between alcohol consumption and mortality. Both abstainers and heavy drinkers tend to have weak social networks. If social isolation increases mortality, the higher mortality among abstainers should not be attributed to the lack of alcohol consumption but to the effects of social isolation.

Skog did not only suggest the possibility of a confounding factor. Building on his previous work on social networks and alcohol consumption, he formally demonstrated that social factors could create the J-curve pattern (Skog 1980). To support his claim he also presented a numerical example which demonstrated the potential importance of these factors. However, he did not have the data to test the empirical significance of social integration. This was also a problem for others in the field who acknowledged the importance of his point, but admitted that the only variable related to social integration they could control for was martial status (Rehm and Sempos 1995). New datasets have changed this and combined with a new methodology, it is now possible to follow up on Skog's suggestion and empirically investigate the importance of social integration.

The dataset that will be used is from the Health and Lifestyle Survey (HALS) organized by the University of Cambridge Clinical School. The original dataset consists of a large representative sample of 9 003 adult individuals living in the UK who were interviewed in 1984/85. Based on the respondents own answers, 11.5% were abstainers, 19.5% used alcohol on special occasions (Christmas, weddings), 44.4% were occasional drinkers and 24.6% were regular drinkers. The last two categories (occasional and regular drinkers) are classified as "drinkers" in this paper. The survey also included a question about previous drinking habits. Hence, to isolate the effects of moderate alcohol consumption, former drinkers were excluded from the non-drinking group (6.2% of the sample) and both current and former heavy drinkers were excluded in the analysis (2.3% of the sample). In addition to the interview, nurses collected physiological measures like blood pressure and respiratory functions from 82% of the sample. Finally, 88.6% of the respondents filled in a self-report booklet with questions used to assess personality and psychiatric status. After the exclusions and non-responses, the analysis was carried out on a sample of between 4 079 and 5 512 individuals, the difference depending mainly on whether questions related to physical measurements collected by a nurse were included in the analysis.

Information about the date and cause of death on these individuals is recorded in a separate data file and the last update is from May 2003. By that time 24.1% of the individuals had passed away. The data are publicly available and more information about the selection process and representativeness can be found in the Users' manual to the survey and the data.

In order to analyse this data, I will use a method developed by Rubin and Rosenbaum (1983) which explicitly focuses on the problem of how to isolate causality when faced with non-experimental data. The method is called propensity score analysis, but to explain what it does, it is necessary to take a short detour into the history of statistics.

Statistics and the measurement of causal effects

Statistical views on causation and the measurement of causation in observational data have been surveyed in two published debates from 1986 and 1999 (Holland 1986, Rosenbaum 1999). In this literature causal effects are defined counterfactually as the difference between the outcome when the individual receives a treatment and when she does not receive a treatment. In the current context, treatment is defined as "moderate drinking of alcohol" and the outcome is the risk of dying before May 2003. Formally, the causal effect of alcohol for an individual (i) could then be measured by:

$$\Delta(i) = Y_1(i) - Y_0(i) \qquad (1)$$

Where Y_0 is the risk of dying if you do not use alcohol and Y_1 is the risk of dying if you consume moderate amounts of alcohol.

The problem is that it is impossible to observe both these values. A person cannot drink alcohol and abstain at the same time, so one piece of information is always missing. This is sometimes called the fundamental problem of causal inference and in order to get around this problem it is necessary to impose some assumptions on the problem. One such approach is to restrict our attention to the overall expected effect of treatment within a population. For instance, one might want to know the overall expected effect of drinking alcohol:

$$E(\Delta) = E(Y_1) - E(Y_0) \qquad (2)$$

Although this approach avoids the problem of reference to individuals who receive both treatments at the same time, it is not enough since the expression above does not correspond to the information we find in the data. All we observe is the death rate in the two groups, not what it would have been if the groups had been different. Let $D=1$ indicate that the individual receives

the treatment (i.e., moderate alcohol consumption) and $D=0$ indicate abstention. Formally, from observed variables we can estimate the outcome for those who drink alcohol and those who abstain:

$$E(Y_1 \mid D=1)$$
$$E(Y_0 \mid D=0) \qquad (3)$$

The question is then if and how we can use the information on observed mortality from (3) to estimate unobserved components in (2) that are needed to find the effect of moderate alcohol consumption.

One possible answer can be found in basic statistical theory which tells us that if the outcome (Y) is independent of the selection into treatments (D) then

$$E(Y_1 \mid D=1) = E(Y_1)$$
$$E(Y_0 \mid D=0) = E(Y_0) \qquad (3)$$

That is, the observed average outcome for the selected group is an unbiased estimator of the outcome for the whole group as long as the method of selecting members to the groups is independent of the mortality rates. As long as this is true we can write the expected effect of alcohol on mortality in terms of variables that can be estimated from the observed data:

$$E(\Delta) = E(Y_1 \mid D=1) - E(Y_0 \mid D=0) \quad (4)$$

In summation, it is possible get an estimate of causal effects by restricting attention to overall net average treatment results and by making sure that the way individuals are selected into the different treatments is independent of the outcome of the treatment (the assumption of independence).

To focus only on the overall net effect is a restrictive assumption (Heckman and Vytlacil 2001). When investigating the effects of a treatment one might be interested in many other variables. For instance, assume that the treatment - drinking alcohol - has strong beneficial effects for a small minority while the majority experience no change or even a slight increase in mortality. In this case one might want to know about the probability of the treatment to *increase* the client's problems, how many this applies to, and how strong the various effects are for the sub-groups. These questions cannot be answered by examining the overall expected effect. What we ideally would like to have is knowledge about the whole distribution of the parameter we are interested in. However, unable to find all the information,

we settle for the second best solution of trying to at least identify the overall expected effect, and the key to this seems to be independence between the selection mechanisms and the outcome.

How do we make sure that the selection into treatment is independent of the outcome and is there any reason to expect that they are not independent in observational studies? In an experiment the independence between the outcome and the selection method is ensured by using a random selection mechanism. As long as the selection is done in this way, the mechanism does not (on average) generate any systematic differences between those receiving the treatment and the control group. In observational studies the situation is different. If there is external selection, then the rules often distinguish between cases in a way that implies sending hard cases to one treatment while easier cases remain in the control group. If there is self-selection, it is plausible that people tend to select the treatment that they expect to work best for them. In both cases there is a relationship between the mechanism of selection and the final outcome and it would be wrong to use the observed difference between the groups as an unbiased estimate of the causal effect.

From the arguments above one might tend to think that conclusions based on observational data are of little value since they are likely to be biased because of selection effects. Conclusions from randomized experiments, on the other hand, provide convincing evidence on the effects of different treatments. In the medical community some even argue that "science recognize today that randomization is the only valid method of comparing the effects of different treatments" (Barer 1994). With this starting point little can be said about the effects of alcohol since it is difficult to imagine a large study in which people are randomized into a life of abstention or moderate drinking.

The claim that knowledge can only be found by randomized experiments ignores the fact that statistical methods might be used to adjust for selection effects in observational data. This has been an active field of research during the past 30 years, and the many methods trying to deal with selection bias in observational studies can be divided into two main groups. First, there are some methods that rely on regression techniques that are modified to take selection into account (the regression approach). An example of this is estimation using the method of two-stage least squares (Heckman 1979). The second general approach tries to extract and compare subgroups of similar people from each treatment (matching). For instance, if it is claimed that the overall result is biased because one groups consists of individuals with higher than average mortality, one might exclude these cases and compare only the results for the individuals who are similar except for their drinking habits.

The regression approach has been criticized for relying on extreme extrapolations and inference into areas in which there are few or no observations (Rosenbaum and Rubin 1983). Moreover, it is claimed, these extrapolations are done without due warning to the researcher. A regression equation produces a result without telling the reader about the number of observations in the neighborhood in which the inference is made. The regression approach also usually assumes a specific functional form, often linearity. Some also complain about the distributional assumptions needed to justify a regression - for instance that the errors are independent and normally distributed. Finally, sometimes the result of a regression is difficult to interpret, especially for non-experts.

Comparing sub-groups of similar individuals avoids some of the problems described above. It is easy to interpret for non-experts, the method explicitly focuses on the number of individuals in each group thus warning the researcher about the possibility of extreme extrapolation, and it does not rely on the same assumptions about linearity and error distribution implied in the regression models. There is, however, one problem that is almost fatal: When the number of variables grows, the number of groups we must compare grows exponentially. Distinguishing between males and females is easy (two groups), but if we add a second variable - say age - we must use four groups. Adding another variable - e.g. level of psychiatric problems - increases the number of groups to be compared to eight (e.g. young males with psychiatric problems vs. old females with few psychiatric problems). Adding yet another variable increases the number of groups to 16. In general, the problem is that if the number of variables we want to control for (n) increases the number of groups rapidly becomes very large (2^n) and we are unlikely to have enough observations in each group to draw any reliable conclusions.

Rosenbaum and Rubin's Propensity Scores

In an article from 1983 Rosenbaum and Rubin claim to have an answer to the "2^n-problem" or the curse of dimensionality, as it is sometimes labeled. The basic idea is very simple. In general we want to compare the outcomes for individuals who are as similar as possible but who have received different treatments. The question is then how we define similarity. The standard reply is to create groups that are similar in all dimensions (gender, age, previous diseases and so on). An alternative approach would be to aggregate all the supposed confounding variables in a single variable and then compare the individuals who have similar values on this variable. This is

essentially what Rosenbaum and Rubin do, but to understand exactly how, why and when this gives unbiased results more elaboration is needed (for a basic and applied introduction to propensity score analysis, see Dehejia and Wahba 2002, D'Agostino 1998, Rubin 1997).

Assume that we have a set of variables (denote the set X) - such as age, gender, previous use of alcohol, psychiatric problems–that we believe contaminate the overall result when we compare the mortality of two groups. Instead of trying to compare individuals who have similar values on all the confounding variables, we will use all the variables (X) to construct what is called the propensity score. This is the estimated probability that a person drinks alcohol given the values on all the confounding variables. In the dataset we observe who abstains and who drinks - and their characteristics - so in order to find the estimated probability of consuming alcohol (given the values on his or her X variables) we can use logistic regression of the confounding selection variables (X) on D (D is here 0 or 1 depending on whether the person drinks alcohol or not). This gives us an estimate of the probability that a person drinks alcohol given his gender, age, previous disease history, psychiatric profile and whatever else we believe might be different between drinkers and non-drinkers which also affect the mortality rate.

After working out the propensity score for each individual, the next step is to compare individuals with similar propensity scores to estimate what the result would have been without treatment. In terms of equation (1), we construct the counterfactual result $Y_0(i)$ by using the outcome from similar individuals, but who differ in terms of whether they drink or not. This can be done in many ways - comparing only pairs of individuals with similar propensity scores, comparing treated individuals to several of its closest non-treated neighbors, comparison with closest neighbors weighted by some function in which the most similar individuals get more weight and many other matching methods. In simulation studies these different methods tend to give the same results in large samples. The overall expected treatment effect can then be estimated by taking the average of all the individual differences. One should note that unlike randomized social experiments, this approach to measuring the causal effect gives information about the whole distribution of the parameter we are interested in. Each individual who drinks alcohol is assigned a counterfactual value and we can answer questions about how many actually benefit (or not) from the treatment and by how much.

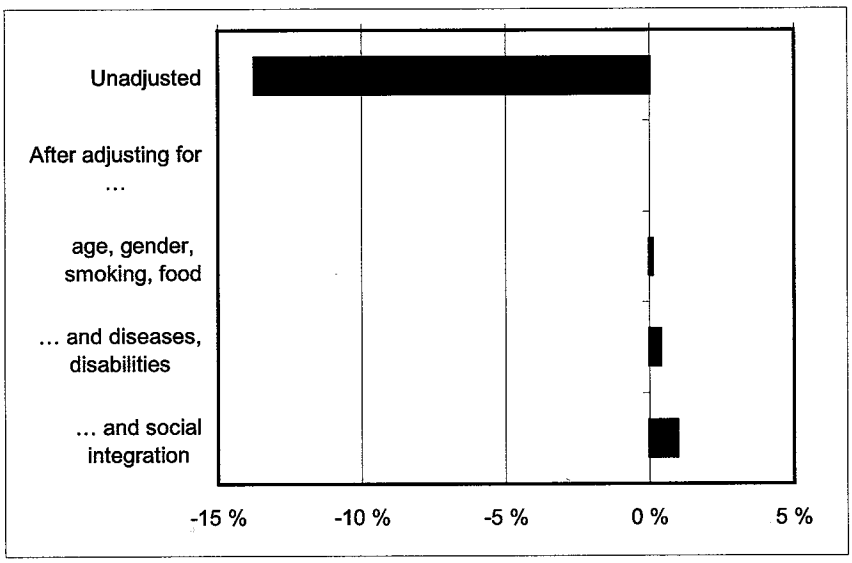
Why should we expect this to produce an unbiased estimate of the treatment effect? Based on the formal proof by Rubin and Rosenbaum, it is possible

to give an intuitive account of why the procedure works. In the data there are probably some individuals who have a high probability of drinking alcohol given their characteristics (gender, age and so on), but who in fact abstained. These people are similar to some who drink alcohol moderately in the sense that they have almost identical propensity scores. Assuming we have managed to identify all the important confounders, it is as if we have two similar individuals and used a random mechanism to assign them either to be in the group that abstains or the group that drinks alcohol. For this pair, all dependence between the selection mechanism and the outcome has been eliminated and – as with randomization – the estimated treatment effect is unbiased.

Results

If one simply splits the end sample into current and former moderate drinkers (75% of the sample) and lifetime nondrinkers (abstainers/special occasion drinkers, 25%), there is a clear tendency for higher mortality among the abstainers. 34.7% of those identified as non-drinkers in 1984/5 had died 19 years later. In comparison, among those respondents classified as moderate drinkers, 20.1% were dead. Obviously, a large part of this difference can be explained by confounders such as age (old people drink less and have higher mortality rates) and other well known factors associated with mortality (Balia and Jones 2004). In the first step of the analysis the mortality was adjusted for some of these factors (age, gender, income, type and regularity of breakfasts, sleeping habits, consumption of nuts and juice).

After adjusting for several such factors believed to influence mortality rates using the propensity score method described, the difference between abstainers and moderate drinkers is reversed. The applied work was done in Stata using the psmatch2 command developed by Leuven and Sianesi 2003. The unadjusted difference between the groups was about 14% in favour of the drinkers, but after the first adjustment the difference is 0.1% in favour of the abstainers (see figure 1). The adjusted difference between drinkers and non-drinkers is too small to be statistically significant, but the change from the unadjusted figure is large and significant. This large change demonstrates the importance of adjusting important confounders in order to isolate the effects of alcohol.



* Percentage difference in deaths among drinkers compared to non-drinkers 19 years after the interview

Figure 1. Effects of moderate consumption of alcohol on mortality*

The second step of the analysis adjusted not only for age, gender and other characteristics, but also factors relating to the "sick quitter hypothesis" (disabilities, body mass index, blood preassure, lung capacity, pulse, past history of diseases like stroke, other heart problems, and diabetes). People with diseases or disabilities often drink very little and at the same time they have naturally high mortality rates. After including these variables in the analysis, the difference between the abstainers and drinkers became slightly larger, going from 0.1% to 0.3% in favour of abstainers. Finally, taking several indicators of social isolation into account produced an even larger change, making the difference in favour of the abstainers about 1%. In other words, when comparing like-with-like, the percentage of deaths among those drinking alcohol is estimated to go down by 1% from 20.8% to 19.8% in an alternative world with no drinking.

To interpret these numbers, note that there are 3 125 moderate drinkers in the final sample and 636 (20.8%) of these had died before May 2003. A reduction to from 20.8% to 19.8% implies that if none of the 3125 drinkers had used alcohol, there would have been 31 fewer deaths by May 2003. That is, a reduction in 5% in the number of deaths among moderate alcohol consumers - from 636 to 605.

Discussion

The results show that the impact of social isolation is in the expected direction, but the effect is modest compared to the change after adjusting for age, gender and other standard variables relating to mortality. Other studies trying to test Skog's suggestion on social isolation did not find statistically significant effects (Murray et al. 1999, Greenfield, Rehm and Rogers 2002). Moreover, unlike the current study, they found that moderate consumption of alcohol reduced mortality. Reflecting upon the possible reasons for this brings up an interesting statistical problem.

Recall Skog's argument that social isolation would be high among both non-drinkers and heavy drinkers. In other words, there is a non-linear relationship between mortality and social isolation – a pattern that looks like a J-shaped curve. The question is then whether the methods used are able to pick up or adjust for such relationships. For instance, in the propensity score approach it is common to use logistic regression to estimate the probability that a person drinks alcohol. Logistic regression assumes that the relationship between the dependent variable (mortality) and the variables believed to influence mortality – such as age and social isolation – looks like an extended "S". At low values changes in the variable (e.g. social isolation) will not produce large changes in mortality, in the middle small changes can have large effects, but towards the end the effect of changes in the variable is once again small. This seems to produce a problem: The theory – and empirical evidence – says that we have a J-shaped relationship, but in the statistical approach testing for this, we assume that it is S-shaped. How much of a problem is this? In the empirical application presented in this paper, the problem is reduced – but not eliminated - by excluding heavy drinkers from the analysis. This means that the upper part of the "S" is partly removed and the "S" looks more like a "J", but it is far from perfect.

Similar statistical problems also appear in the more standard approach used in the literature – survival analysis. In addition to all the standard assumptions used in survival analysis, statisticians have pointed out that a common method testing for the existence of a J-curve relationship is flawed (Goethghebeur and Pocock 1995). In a paper that is almost never cited in the substance abuse literature, they present two problems that lead us too easily to conclude that there is a J-shaped relationship. The first reason is simply that there are often very few observations in the tail. A high value might just be the result of random variations in a small group. Given the number of studies and observations, this does not seem to be likely in the case of mortality among abstainers.

The second reason is more interesting. Assume one tries to fit a function to some points that follow the general J-shaped pattern. It may well be the case that a quadratic function fits the points quite well, but the reason for this good fit may have more to do with the pattern on the right hand side (e.g. the increasing mortality associated with alcohol), than the fit to the left of the nadir. That is, we may end up selecting a function that confirms the J-curve because it gives the best fit for the many observations to the right of the turning-point. It would then be wrong to put much emphasis on the predictions to the left of the turning point.

In brief, there are still unsolved statistical problems here - how to quantify the problem of only using only approximately correct assumptions and how to implement approaches which allow more flexible functional forms between the variables. In addition to the other problems - such as trying to control for patterns of drinking as well as average consumption, and modeling more complex webs of causal interactions (Rehm 2000) - I believe this is one area that should be further explored in order to capture the true relationship between alcohol and mortality.

Limitations

Do the results above imply that there is no J-curve and that moderate consumption of alcohol increases mortality? Before making such a strong conclusion it is necessary to examine the results more closely.

First of all, the uncertainty associated with the results is quite large. Using bootstrapping techniques to estimate the confidence intervals, the 95% confidence interval for the effect of moderate alcohol consumption goes from a negative effect (1.8% fewer deaths) and up to a positive effect of 3.9% more deaths. This indicates that there is quite a lot of uncertainty associated with the results.

Second, it is necessary to analyse the extent to which the propensity score method manages to adjust for the confounders. Table 1 shows that the drinkers and non-drinkers were statistically significantly different on many variables before adjustment. For instance, before adjustment the average age among the drinkers was 45, while it was 54 among the non-drinkers. After creating groups of similar individuals, the age difference is eliminated. Overall the groups were different on 18 of 26 possible confounders before adjustment. After trying to create pairs of similar individuals, the two groups differed significantly on only 3 variables, and even for these three variables the difference between the drinkers and non-drinkers was reduced by at least

70%. In short, the table shows that the bias due to the observed confounding variables has been eliminated or greatly reduced after matching.

Table 1. Overview of important confounding variables, averages, and bias reduction

Variable	Sample	Average for drinkers	Average for non-drinkers	Percent difference	Reduction in bias (%)	p-value
Age	Unmatched	44,78	54,42	-56,5		0,00
	Matched	45,17	45,42	-1,4	97,4	0,21
Gender	Unmatched	0,53	0,25	60,0		0,00
	Matched	0,52	0,52	0,9	98,5	0,36
Income	Unmatched	5,58	4,45	45,4		0,00
	Matched	5,50	5,39	4,6	89,9	0,02
Smoking	Unmatched	0,35	0,27	18,4		0,00
	Matched	0,35	0,32	4,8	73,8	0,04
Disabilities	Unmatched	0,29	0,36	-13,0		0,00
	Matched	0,30	0,31	-3,4	73,5	0,23
Body Mass Index	Unmatched	24,50	25,08	-13,7		0,00
	Matched	24,54	24,52	0,6	95,3	0,89
Heart trouble	Unmatched	0,07	0,10	-10,6		0,00
	Matched	0,07	0,08	-3,0	72,2	0,24
Depression/ Nervous illness	Unmatched	0,18	0,22	-10,4		0,00
	Matched	0,18	0,17	1,8	82,7	0,67
High blood pressure	Unmatched	0,14	0,21	-18,7		0,00
	Matched	0,15	0,14	2,7	85,7	0,51
Index of social isolation (family)*	Unmatched	4,05	3,99	2,6		0,47
	Matched	4,07	4,13	-2,3	11,6	0,33
Index of social isolation (friends)*	Unmatched	3,50	3,01	21,3		0,00
	Matched	3,43	3,32	4,8	77,6	0,02
Quality of personal relationships (index)*	Unmatched	19,89	19,75	6,0		0,09
	Matched	19,88	19,91	-1,0	82,6	0,75
Feelings of loneliness	Unmatched	0,41	0,51	-13,8		0,00
	Matched	0,41	0,41	0,5	96,5	0,91

* These indexes were constructed using questions about the frequency of contacts with family, friends, and the degree to which the individuals felt these contacts were supportive and important

A third potential problem is data quality and validity. For instance, given that the information used is from one point in time, one might question

whether these answers are representative for the individual. A 20-year-old individual engaged in physical exercise could quit soon after the survey. This information would be relevant for estimating his mortality, but it is not available. More generally, it may not be valid to take the answers given in 1984/5 as indicators of the lifetime-characteristics we are interested in. This is an important criticism and there is no point in denying that one would have liked to have annual information on all the variables. However, one might argue that at least some of the variables are of a permanent nature - disabilities and permanent illnesses remain over time; a past history of heart problems does not go away the next year.

Fourth, the method used to isolate causality could be criticized for not exploiting all the information that is in the data. Assume we wait 100 years and all the individuals in the sample have passed away. Then there will be no relative difference between drinkers and abstainers in terms of the share of deaths in the two groups. This illustrates that the method does not fully exploit information about the timing of the deaths. It could be adapted to take account of this by using "age at death" as the outcome variable, but as long as only a minority of the sample is dead, this is not a good option. Another option is exclude those who are so old that they were likely to pass away within the time period. It turns out that re-running the analysis after excluding those who were above 70 years old in 1984/85 produced almost the exact same difference: Moderate drinking increased the percentage of fatalities in the group by 1%. However, when running separate analyses for men and women (excluding those above 70), the males weakly benefited from alcohol consumption, while the females did not. None of the results were statistically significant. Note that in this case "not significant" does not imply "not interesting." Many studies claim that alcohol is a significant factor and a test showing the opposite is interesting in itself. However, while running the analysis on sub-groups of age and gender may give a better picture, it does not change the fact that the method throws away some information that standard survival analysis manages to include. On the other hand, survival analysis makes a number of other assumptions that can cause problems. Given that different methods makes different assumptions and have different strengths and weaknesses it makes sense to explore the issue using a variety of methods.

Fifth, and finally, there is the problem of intermediary variables. Assume we have the following causal relationship: alcohol → depression → mortality. Viewing depression as a separate causal variable means that we assign it some independent causal effect and eliminate this from the overall

relationship between alcohol and mortality. However, if depression itself is caused by alcohol, it seems wrong to eliminate it. We want to measure the gross effect of alcohol on mortality and if depression is caused by alcohol this effect should also be included, not eliminated. Hence, adjusting for variables that themselves are caused by alcohol use, leads to underestimation of the true effect of alcohol. Since some such intermediary variables have been included in the analysis, one may argue that the result is flawed. This is true, but as shown in figure 1 the reported results do not change much when some of the additional variables - some of which are intermediary like blood pressure and depression - are included. Second, the relationship between the intermediary variables and alcohol is not perfect in the sense that depression can be caused by many other factors than alcohol. Hence, to the extent that it has an independent causal effect it becomes important to control for it before measuring the effects of alcohol.

Conclusion

There are no magical bullets for isolating causal effects in observational data and in the conclusion of his article on the J-curve, Skog himself argued in favour of trying out several different methods. He also suggested that aggregate time-series analysis could be used to avoid some of the selection effects that create problems for the analysis of individual-level data. This suggestion has also been quite influential, inspiring a number of studies on the relationship between alcohol consumption and mortality (for a review, see Norström and Ramstedt 2005). In this paper a different approach has been explored in order to adjust for selection problems - propensity score analysis. Using this approach I did not find that moderate consumption of alcohol was causally important to reduce mortality. However, the paper also showed that the uncertainties are still large and that important methodological and data problems need to be solved before it is possible to draw more certain conclusions.

Acknowledgement

I am grateful for comments and suggestions from William Kerr, Sturla Nordlund, Thor Norström, Jürgen Rehm and the editors of this book.

References

Balia, S. and Jones, A.M. (2004), "Mortality, Lifestyle and Socio-Economic Status", *Working Paper*, Department of Economics and Related Studies, University of York.

Barer, D. (1994), "Patients' Preferences and Randomised Trials", *Lancet* 344 (8923): 688.

Corrao, G., Rubbiati, L., Bagnardi, V., Zambon, A. and Poikolainen, K. (2000), "Alcohol and Coronary Heart Disease: A Meta-Analysis", *Addiction* 95 (10): 1505-1523.

D'Agostino, R.B., Jr. (1998), "Propensity Score Methods for Bias Reduction in The Comparison Of A Treatment To A Non-Randomized Control Group", *Statistics in Medicine* 17 (19): 2265-2281.

Dehejia, R. and Wahba, S. (2002), "Propensity Score-Matching Methods for Non-Experimental Causal Studies", *The Review of Economics and Statistics* 84 (1): 151-161.

Doll, R., Peto, R., Boreham, J. and Sutherland, I. (2005) "Mortality In Relation To Alcohol Consumption: A Prospective Study among Male British Doctors", *International Journal of Epidemiology* 34 (1): 199-204.

Ellison, R.C. (2002) "Balancing the Risks and Benefits of Moderate Drinking", *Annals of the New York Academy of Sciences* 957: 1-6.

Fillmore, K.M, Kerr, W.C., Stockwell, T., Chikritzhs, T. and Bostrom, A. (2006 forthcoming), "Moderate alcohol use and reduced mortality risk: Systematic error in prospective studies", *Addiction Research and Theory.*

Goethghebeur, Els J.T., and Pocock, Stuart J. (1995), "Detection and Estimation of J-Shaped Risk-Response Relationships", *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 158 (1): 107-121.

Greenfield, T. K., Rehm, J. and Rogers, J. D. (2002), "Effects of Depression and Social Integration on The Relationship Between Alcohol Consumption and All-Cause Mortality"m *Addiction* 97 (1): 29-38.

Gronbaek, M., Deis, A., Sorensen, T.I., Becker, U., Borch-Johnsen, K., Muller, C., Schnohr, P. and Jensen, G. (1994), "Influence of Sex, Age, Body Mass Index, and Smoking on Alcohol Intake and Mortality", *British Medical Journal* 308 (6924): 302-306.

Heckman, J.J. (1979), "Sample Selection Bias as a Specification Error", *Econometrica* 41 (1): 153-161.

Heckman, J.J., and Vytlacil, E.J. (2001), "Policy-Relevant Treatment Effects", *American Economic Review* 91 (2): 107-111.

Holland, P.W. (1986), "Statistics and Causal Inference", *Journal of the American Statistical Association* 81: 945-970.

Klatsky, A.L. (2002), "Alcohol and Cardiovascular Diseases: a Historical Overview", *Annals of the New York Academy of Sciences* 957: 7-15.

Leuven, E. and Sianesi, B. (2003) PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing. *http://ideas.repec.org/c/boc/bocode/s432001.html Version 2.0.8.*

Murray, R.P., Rehm, J., Shaten, J. and Connett, J.E. (1999) "Does Social Integration Confound The Relation Between Alcohol Consumption and Mortality in The Multiple Risk Factor Intervention Trial (MRFIT)?" *Journal of Studies on Alcohol* 60 (6): 740-745.

Norström T. and Ramstedt, M. (2005), "Mortality and Population Drinking: A Review Of The Literature", *Drug and Alcohol Review* 24: 537–547

Pearl, L.R. (1926), *Alcohol and Longevity.* New York: A.A. Knopf.

Rehm, J. (2000), "Alcohol consumption and mortality. What Do We Know and Where Should We Go?" *Addiction* 95 (7): 989-995.

Rehm, J., and Sempos, C.T. (1995), "Alcohol Consumption and All-Cause Mortality", *Addiction* 90 (4): 471-480.
(1995), "Alcohol Consumption and Mortality - Questions about Causality, Confounding and Methodology", *Addiction* 90, 493-498.

Room, R., Babor, T. and Rehm, J. (2005), "Alcohol and Public Health", *Lancet* 365 (9458): 519-530.

Rosenbaum, P.R. (1999), "Choice as An Alternative To Control In Observational Studies", *Statistical Science* 14 (3): 259-278.

Rosenbaum, P.R., and Rubin, D.B. (1983), "The central role of the propensity score in observational studies for causal effects", *Biometrika* 70 (1): 41-55.

Rubin, D.B. (1997), "Estimating causal effects from large data sets using propensity scores", *Annals of Internal Medicine* 127 (8): 757-763.

Shaper, A.G., Wannamethee, G. and Walker, M. (1988), "Alcohol and mortality in British men: explaining the U-shaped curve", *Lancet* 2 (8623): 1267-1273.
(1989) "Alcohol and the U-shaped curve", *Lancet* 1 (8633): 336.

Shaper, A.G., and Wannamethee, S.G. (2000), "Alcohol intake and mortality in middle aged men with diagnosed coronary heart disease", *Heart* 83 (4): 394-399.

Skog, O-J. (1995), "The J-curve, causality and public health", *Addiction* 90 (4): 490-492; discussion 493-498.

Skog, O-J. (1996), "Public health consequences of the J-curve hypothesis of alcohol problems", *Addiction* 91 (3): 325-337.

Skog, O-J. (1980), "Social interaction and the distribution of alcohol consumption", *Journal of Drug Issues* 10, 71-92.

Standridge, J.B., Zylstra, R.G. and Adams, S.M. (2004), "Alcohol consumption: an overview of benefits and risks", *Southern Medical Journal* 97 (7): 664-672.