IN: *CHANCE*, 2025, 38:1, 25-30.

.

DOES A DAILY GLASS OF WINE LENGTHEN LIFE? INSIGHT FROM A SECOND CONTROL GROUP

PAUL R. ROSENBAUM
DEPARTMENT OF STATISTICS AND DATA SCIENCE
UNIVERSITY OF PENNSYLVANIA, PHILADELPHIA, PA, U.S.

Abstract. Does a daily glass of wine lengthen or shorten life? Despite extensive research, the question is still debated. In this context, a simple example is presented in which a second control group provides insight into unmeasured biases in observational studies of light daily alcohol consumption and longevity.

1. Does a Daily Glass of Wine Lengthen Life?

A popular notion is that light daily consumption of alcohol lengthens life by reducing cardiovascular mortality, perhaps in part by increasing the level of high-density lipoprotein (HDL) cholesterol, the so-called "good cholesterol." Is this true?

In chapter nine of the book, *Eat, Drink and Be Healthy: The Harvard Guide to Healthy Eating*, the distinguished nutritional epidemiologist, Walter Willett, and his colleagues review some of the relevant literature, both pro and con. One of many articles lending empirical support to the claim of cardiac benefits from light alcohol consumption was written by Il Suh and colleagues and was published in the *Annals of Internal Medicine* in 1992.

Both the American Heart Association and the American Society of Clinical Oncology have published position papers critical of the evidence underlying this popular notion. Understandably, the oncologists' position paper, authored by Noelle LoCante and colleagues, emphasizes the established fact that alcohol is a carcinogen.

What concerns have been raised about the evidence? One concern is the focus on mortality from cardiovascular diseases rather than overall mortality, including mortality from cancer, liver diseases, accidents and violence. A second concern is that a newer tactic, known as "Mendelian randomization," seems to suggest that alcohol increases, rather than decreases, cardiovascular risk, as discussed by Michael Holmes and colleagues; so, different methods and study designs seem to yield different answers. At the most basic level, there is the inescapable concern that the decision to drink alcohol is not governed by random assignment, as it would be in a clinical

trial; so, people in different treatment groups who look comparable may not actually be comparable.

Support for this most basic concern was demonstrated by Bo Petersson and colleagues in a study of alcohol consumption by men in Sweden. They found elevated total morality among men who abstained from alcohol, but concluded that this was likely not an effect caused by alcohol, writing:

> Most of these men, however, had chronic disease as the reason for abstention, or even a past history of alcoholism. Increased mortality in non-drinkers may create a false impression of a preventive effect of any versus no daily drinking.

Briefly, they are suggesting that some people abstain from alcohol because they are ill, rather than being ill because they abstain from alcohol.

2. Insight From Two Control Groups

As a statistician, I am interested, not only in this specific scientific question but also in how questions of this sort should be studied. Here, I will illustrate the insight that is sometimes provided by incorporating a carefully chosen second control group in an observational or nonrandomized study. To choose carefully means that one pattern of responses is expected in the treated group and the two control groups if the treatment caused its ostensible effect, and a different pattern is expected if there is no treatment effect, and the ostensible effect instead reflects a specific, plausible bias in who is exposed to treatment and who is spared exposure. There are limits to what a second control group can do, but perhaps with care it can distinguish these two quite distinct situations.

Building upon an idea of Morton Bitterman, Donald Campbell suggested that two control groups should be selected by the principle of "control by systematic variation." If there is concern about a particular unmeasured source of bias, then control groups should be selected so this potential bias is present to a much greater extent in one group than in the other. If the control groups exhibit similar outcomes despite being very different in terms of a particular source of bias, then that is not incompatible with a minor or negligible role of this source of bias. Conversely, very different outcomes in the two control groups heightens concern about this potential source of unmeasured bias. "Control by systematic variation" has been formalized in terms of its statistical properties, such as power to detect bias, or the ability to offer complementary information or "evidence factors;" see the Further Reading.

As noted above, some people abstain from alcohol as a matter of scruple, perhaps for religious reasons, perhaps because of a history of alcoholism, perhaps due to illness, or perhaps based on concern about interactions between alcohol and needed medications. Other people consume very little alcohol in a year, but do not think

twice about raising a glass at a wedding, a wake or a holiday gathering. These are two different attitudes about consuming alcohol — perhaps scruple and casual indifference — yet the biochemical effects of no alcohol in a year and a few drinks in a year are likely to be similar, particularly when compared to daily drinking. So, this is a pair of control groups that systematically varies an attitude about alcohol, thereby perhaps varying certain unmeasured biases.

For instance, a brochure, *Harmful Interactions: Mixing Alcohol with Medicines*, published on-line by the US National Institute on Alcohol Abuse and Alcoholism, contains six pages of single-line warnings about harmful interactions between particular medications and alcohol. The interactions include complications like internal bleeding, heart problems, and difficulty breathing. They also include warnings that alcohol may render the medication less effective or ineffective. Quite possibly, scrupulous abstention from alcohol is more common among people who take these medications, and consequently who have the health problems that necessitate taking these medications.

3. The Data and the Comparisons

The data are from six of the US National Health and Nutrition Examination Surveys (NHANES), from 2005 to 2016, together with linked death certificate records from the National Death Index. The surveys for these years offer two advantages: (i) the NHANES questions about alcohol and smoking are pretty much the same over this period of time, and (ii) there is a reasonable period of follow-up for mortality beginning at the date of an individual's NHANES clinical examination.

In NHANES, some questions characterize excessive or binge drinking as: (i) four or more alcoholic drinks in a day for a woman, or (ii) five or more alcoholic drinks in a day for a man, while others refer to five or more drinks without reference to sex. These definitions of binge drinking cannot be modified given the way NHANES questions were asked.

Light daily drinking was defined as drinking on at least $260 = 5 \times 52$ days in the past year, drinking between 1 and 3 drinks on typical drinking days, with at most 12 days of binge drinking during the year. This group — briefly, the "treated group" — contained 1130 people, and these individuals reported drinking, on average, 533 alcoholic drinks in the year prior to their NHANES survey. Abstention — the first type of control — was defined as zero drinking days and zero days of binge drinking during the year; so, by definition, they had zero drinks in the past year. To save limited space in plots, Abstention is "None," meaning no alcohol. Rare drinking — the second type of control — was defined as: (i) not qualifying as abstaining, (ii) drinking at most 3 drinks on at most $24 = 2 \times 12$ days, with (iii) at most 12 days of binge drinking during the year. In the matched comparison, the "rare drinking"

TABLE 1. Covariate balance in $I = 1130$ matched blocks of size $J = 5$, with two controls from each of two control groups. "Ever-binged" refers to a past period of drinking at least 4 or 5 drinks on most days.

	Daily	Rare	None
Sample size	1130	2260	2260
Female, %	36	36	36
Age, mean	62	61	62
Education, mean	3.95	3.90	3.95
EverBinged, %	12	12	12
Never Smoked, %	36	36	36
No Longer Smoke, %	49	49	49
Smoke Some Days, %	3	3	3
Smoke Everyday, %	12	12	12
NHANES0506, %	16	15	16
NHANES0708, %	18	18	18
NHANES0910, %	23	21	21
NHANES1112, %	16	15	17
NHANES1314, %	15	17	16
NHANES1516, %	11	14	12

group reported between 1 and 72 drinks, with a median of 10 drinks and a mean of 14.4 drinks in the past year. In terms of averages, the daily drinkers of alcohol drank $533/14.4 = 37$ drinks for every drink consumed by the rare drinkers.

The block design consists of $I = 1130$ blocks, each block containing $J = 5$ adults, aged at least 20, one daily drinker, two abstainers and two rare drinkers. The blocks were matched for age, sex, five categories of education, four categories of smoking, a binary indicator of binge drinking in the past, and the years of the six NHANES surveys. Because abstainers and rare drinkers were matched to all available daily drinkers, the distributions of these covariates in every matched group resembles the distribution among the daily drinkers. So, the design looks at the possible effects of daily drinking on the type of person who typically engages in daily drinking.

A few details follow. The indicator of past binge drinking is based on a question that asked whether you ever had a time your life when you drank 4 or 5 drinks almost everyday. A person who currently engages in such behavior would not qualify for the three study groups, daily, none, or rare drinker; however, some people in each group did behave in this way in the past. The five categories of education had 1 for "less than 9th grade," 3 for "high school degree or equivalent" and 5 for "at least a BA degree." To preserve confidentiality, NHANES caps reported age at 80 years. The matching was exact for sex, four smoking categories and the binary indicator of

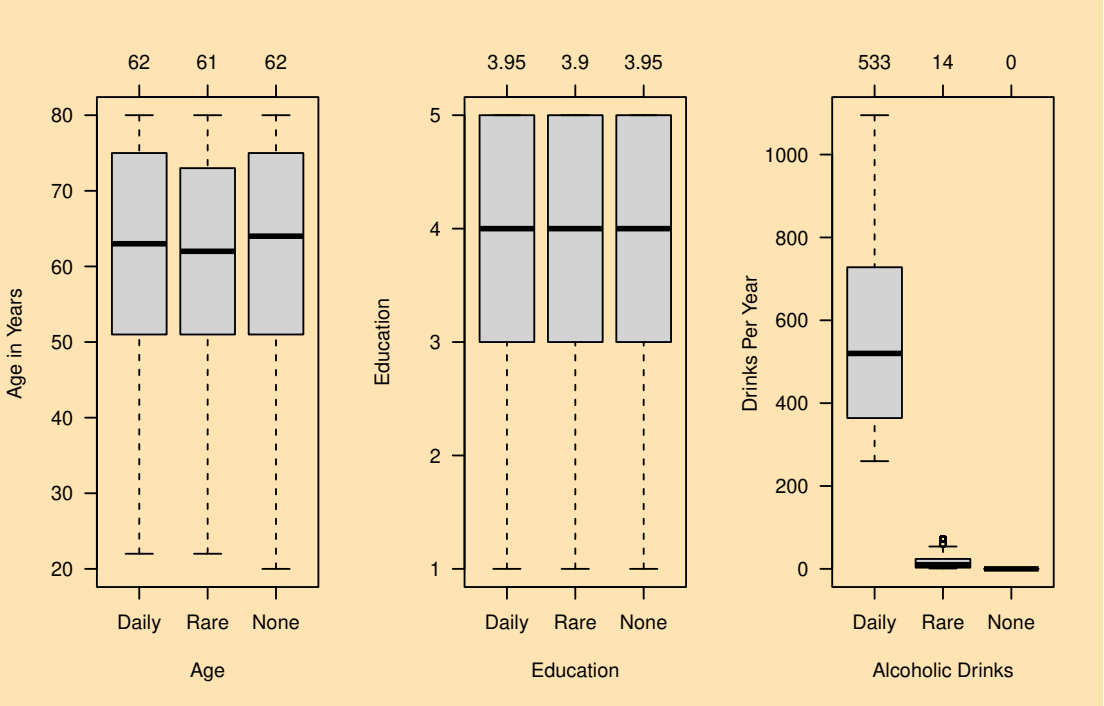

FIGURE 1. After matching, age and education, and alcoholic drinks per year in three groups, daily drinkers, individuals who drank no alcohol in the past year (None), and individuals who rarely drank alcohol in the past year (Rare). For education, 1 is "< 9th grade," 3 is "high school," and 5 is "≥ BA degree." Means are above boxplots.

past binge drinking; so, it is possible to look at these subgroups separately without breaking up any block. The matching used "two-criteria matching" as developed by Bo Zhang and colleagues.

4. Do the Groups Look Comparable?

Table 1 and Figure 1 show the block design, including covariate balance after matching, and the number of alcoholic drinks consumed per year in each alcohol group. After matching, the three alcohol groups look similar in terms of age, sex, smoking categories, past binge drinking, and the years of the NHANES. The year of the NHANES affects the duration of possible follow-up for mortality. Obviously, the groups might differ in terms of other covariates that are not measured.

Although daily drinkers consumed alcohol below the recommended limits at that time, in Figure 1 they drank much more alcohol than the negligible amounts consumed by the two control groups.

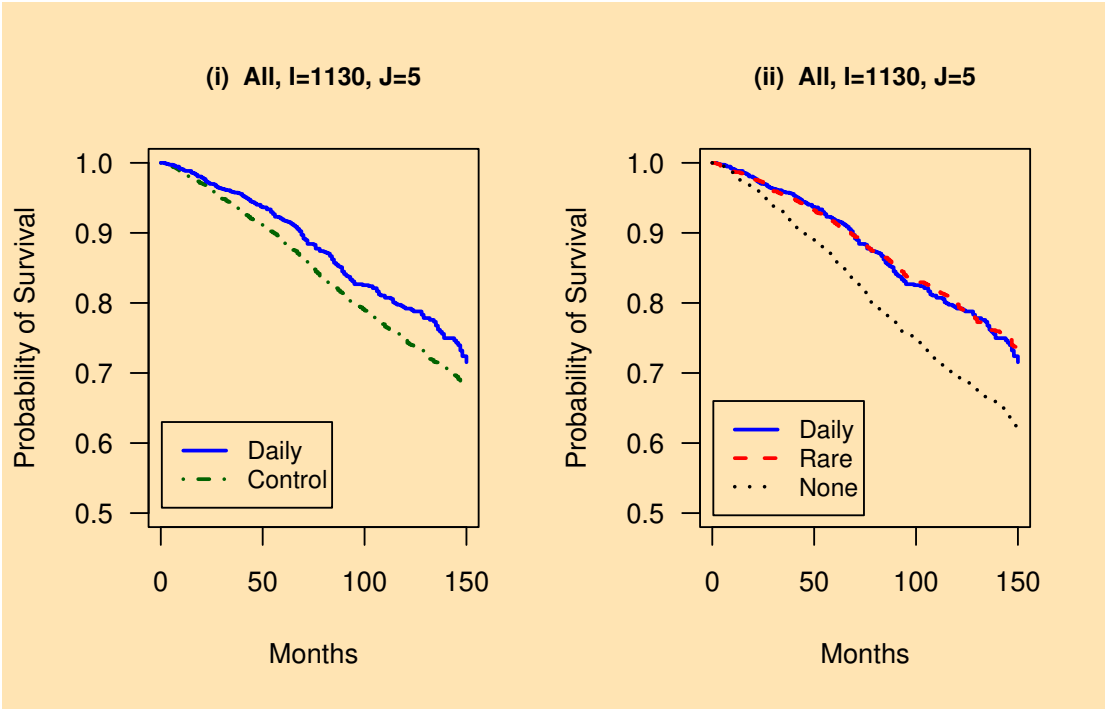

FIGURE 2. Estimated probability of survival to t months after NHANES clinical examination for all $5650 = 5 \times 1130$ matched individuals. The left panel (i) merges the two control groups, while the right panel (ii) separates the two control groups.

Because controls are matched to the daily drinkers, the distributions in Table 1 and Figure 1 reflect the distributions of these variables for daily drinkers. Compared to all adults in NHANES, daily drinkers are older, more often male, with somewhat more education.

5. Do Daily Drinkers Live Longer Than Controls?

Figure 2 compares the survival over 12.5 years (150 months) in the three alcohol groups, where Figure 2(i) merges the two control groups, and Figure 2(ii) separates them. The figure shows estimates of the probability of survival for t months plotted against t. The estimates are Kaplan-Meier survival curves. Notably, the two control groups have very different survival curves: abstainers are dying sooner than rare drinkers, despite only a small difference in alcohol consumption. In Figure 2(i), daily drinkers survive longer than controls, but in Figure 2(ii) this seems to reflect shorter survival among abstainers, with little difference between daily drinkers and rare drinkers, despite very different alcohol consumption in the Daily and Rare

groups in Figure 1. It is difficult to explain the pattern in Figure 2 as an effect actually caused by alcohol. The pattern in Figure 2 is not inconsistent with the possibility that the abstention group contains some individuals who abstain because of significant health problems. The pattern in Figure 2 is not easily attributed to chance.[1]

The two panels of Figure 3 look at subsets of the $I = 1130$ blocks, namely individuals who never smoked, and individuals who never had a lengthy period of binge drinking on most days. Although never smokers survive longer, the comparison of the three alcohol groups is similar to Figure 2(ii).

If a model predicted survival from an annual dose of alcohol, it would report what we see in Figure 2(i), and it would suggest that light daily alcohol lengthens survival, as is often claimed in the literature. If we distinguish abstention from rare drinking, as in Figure 2(ii), then we are left with the very different impression that daily drinkers and rare drinkers exhibit similar survival, while abstainers exhibit far worse survival. This impression from Figure 2(ii) does not favor daily drinking over rare drinking, and instead raises the possibility that some abstainers may be avoiding alcohol for reasons related their health, consistent with what Bo Petersson and colleagues claimed in their study of Swedish men. The important methodological point is that, with two control groups, we cannot miss the pattern in Figure 2(ii).

6. Take-Home Messages

The example of alcohol and longevity has illustrated and, I hope, supported some old but good advice about the design and analysis of observational studies for causal effects.

- William G. Cochran argued that an observational study should be designed to resemble a simple experiment, such as a block design formed by matching for covariates, with careful planning to support the step from association to causation.
- To this, Donald T. Campbell added that an observational study should incorporate what he called "quasi-experimental" devices intended to shed light on unmeasured biases from nonrandomized treatment assignment. A second control group is one simple quasi-experimental device.

[1]Formal analysis simply confirms the visual impression in Figure 2. In Figure 2(i), Cox's stratified proportional hazards model rejects the hypothesis of equal survival with a P-value of 0.000816, and hazard ratio of 0.76 for daily drinkers relative to all controls, with 95% confidence interval [0.66, 0.90], consistent with the literature suggesting light daily alcohol lengthens survival. When the two control groups are separated in Figure 2(ii), the hypothesis of equal survival for daily drinkers and rare drinkers is not rejected with P-value 0.850, while the hypothesis of equal survival for rare drinkers and abstainers has a P-value of 3.99×10^{-14} with abstainers having an estimated hazard ratio of 1.69 compared to rare drinkers, with 95% confidence interval [1.48, 1.94].

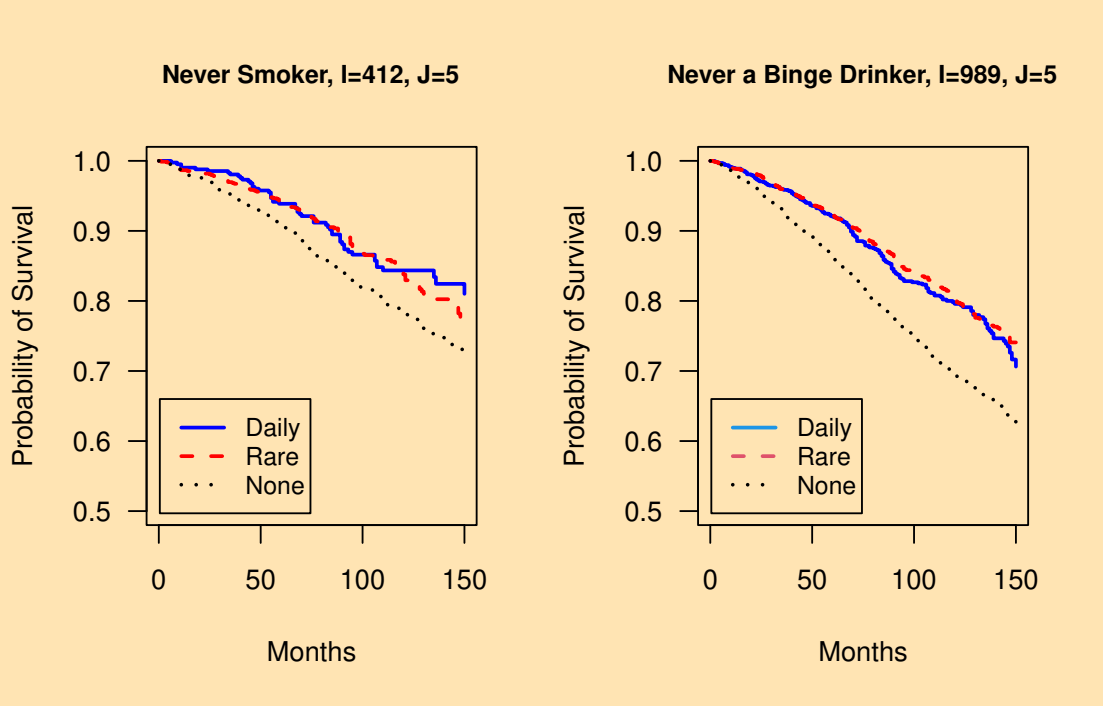

FIGURE 3. Estimated probability of survival, by alcohol group. The left panel describes all $2060 = 5 \times 412$ never smokers, and the right panel describes all $4945 = 5 \times 989$ individuals who never engaged in binge drinking on most days.

- Colin Mallows said: "The most robust statistical technique is to look at the data," but you cannot make sense of observational data until you have adjusted for measured covariates. Before matching, the pool of potential controls had a median age of 51, 59% were female, and 60% never smoked, all very different from the daily drinkers and their matched controls in Table 1. In a matched study, people can be plotted in simple ways that are adjusted for measured covariates, as in Figures 1-3, so matching permits use of Mallows' "most robust statistical technique." Henry James wrote: "No theory is kind to us that cheats us of seeing."

References

Campbell, D. T. (1988). Prospective: Artifact and control. In *Methodology and Epistemology for Social Science: Selected Papers*, pp. 167–190. Chicago: University of Chicago Press. Reprint of a paper from 1969.

Cochran, W. G. (1965). The planning of observational studies of human populations (with Discussion). *Journal of the Royal Statistical Society A 128*(2), 234–266.

Goldberg, I. J., L. Mosca, M. R. Piano, and E. A. Fisher (2001). Wine and your heart: A science advisory for healthcare professionals from the Nutrition Committee, Council on Epidemiology and Prevention, and Council on Cardiovascular Nursing of the American Heart Association. *Circulation 103*(3), 472–475.

Holmes, M. V., C. E. Dale, L. Zuccolo, R. J. Silverwood, Y. Guo, Z. Ye, D. Prieto-Merino, A. Dehghan, S. Trompet, A. Wong, et al. (2014). Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. *British Medical Journal 349*, g4164.

LoConte, N. K., A. M. Brewster, J. S. Kaur, J. K. Merrill, and A. J. Alberg (2018). Alcohol and cancer: A statement of the American Society of Clinical Oncology. *Journal of Clinical Oncology 36*(1), 83–93.

Petersson, B., E. Trell, and H. Kristenson (1982). Alcohol abstention and premature mortality in middle-aged men. *British Medical Journal 285*(6353), 1457–1459.

Rosenbaum, P. R. (1987). The role of a second control group in an observational study. *Statistical Science 2*(3), 292–306.

Rosenbaum, P. R. (2015). How to see more in observational studies: Some new quasi-experimental devices. *Annual Review of Statistics and Its Application 2*(1), 21–48.

Rosenbaum, P. R. (2023a). *Causal Inference*. Cambridge, MA: MIT Press. Chapter 9 reviews statistical issues concerning light alcohol and longevity.

Rosenbaum, P. R. (2023b). A second evidence factor for a second control group. *Biometrics 79*(4), 3968–3980.

Suh, I., B. J. Shaten, J. A. Cutler, and L. H. Kuller (1992). Alcohol use and mortality from coronary heart disease: The role of high-density lipoprotein cholesterol. *Annals of Internal Medicine 116*(11), 881–887.

Willett, W., P. J. Skerrett, and E. L. Giovannucci (2017). *Eat, Drink, and Be Healthy: The Harvard Medical School Guide to Healthy Eating*. New York: Simon and Schuster.

Zhang, B., D. S. Small, K. B. Lasater, M. McHugh, J. H. Silber, and P. R. Rosenbaum (2023). Matching one sample according to two criteria in observational studies. *Journal of the American Statistical Association 118*(542), 1140–1151.