PEER INFLUENCE ON SMOKING: CAUSATION OR CORRELATION?

Sophie Langenskiöld

AKADEMISK AVHANDLING

som för avläggande av ekonomie doktorsexamen
vid Handelshögskolan i Stockholm
framläggs för offentlig granskning
fredagen den 13 januari 2006, kl 15.15 i sal KAW,
Handelshögskolan, Sveavägen 65,
Stockholm

STOCKHOLM SCHOOL
OF ECONOMICS
HANDELSHÖGSKOLAN I STOCKHOLM

EFI, The Economic Research Institute

EFI Mission

EFI, the Economic Research Institute at the Stockholm School of Economics, is a scientific institution which works independently of economic, political and sectional interests. It conducts theoretical and empirical research in the management and economic sciences, including selected related disciplines. The Institute encourages and assists in the publication and distribution of its research findings and is also involved in the doctoral education at the Stockholm School of Economics. At EFI, the researchers select their projects based on the need for theoretical or practical development of a research domain, on their methodological interests, and on the generality of a problem.

Research Organization

The research activities at the Institute are organized in 23 Research Centres.
Centre Directors are professors at the Stockholm School of Economics.

EFI Research Centre:	*Centre Director:*
Management and Organisation (A)	Sven-Erik Sjöstrand
Centre for Ethics and Economics (CEE)	Hans de Geer
Centre for Entrepreneurship and Business Creation (E)	Carin Holmquist
Public Management (F)	Nils Brunsson
Information Management (I)	Mats Lundeberg
Centre for People and Organization (PMO)	Andreas Werr (acting)
Centre for Innovation and Operations Management (T)	Christer Karlsson
Centre for Risk Research (CFR)	Lennart Sjöberg
Economic Psychology (P)	Guje Sevón
Centre for Consumer Marketing (CCM)	Magnus Söderlund
Centre for Information and Communication Research (CIC)	Per Andersson (acting)
Marketing, Distribution and Industrial Dynamics (D)	Björn Axelsson
Centre for Strategy and Competitiveness (CSC)	Örjan Sölvell
Centre for Business and Economic History (BEH)	Håkan Lindgren
Accounting and Managerial Finance (B)	Johnny Lind
Centre for Financial Analysis and Managerial Economics in Accounting (BFAC)	Kenth Skogsvik
Finance (FI)	Clas Bergström
Centre for Health Economics (CHE)	Bengt Jönsson
International Economics and Geography (IEG)	Mats Lundahl
Economics (S)	Lars Bergman
Economic Statistics (ES)	Anders Westlund
Law (RV)	Erik Nerep
Centre for Tax Law (SR)	Bertil Wiman

Chair of the Board: Professor Carin Holmquist
Director: Associate Professor Filip Wijkström

Address

EFI, Box 6501, SE-113 83 Stockholm, Sweden • Homepage: www.hhs.se/efi/
Telephone: +46(0)8-736 90 00 • Fax: +46(0)8-31 62 70 • E-mail efi@hhs.se

Peer Influence on Smoking: Causation or Correlation?

Sophie Langenskiöld

STOCKHOLM SCHOOL
OF ECONOMICS
HANDELSHÖGSKOLAN I STOCKHOLM

EFI, The Economic Research Institute

Dissertation for the Degree of Doctor of Economics
Stockholm School of Economics

ABSTRACT: In this thesis, we explore two different approaches to causal inferences. The traditional approach models the theoretical relationship between the outcome variables and their explanatory variables, i.e., the science, at the same time as the systematic differences between treated and control subjects are modeled, i.e., the assignment mechanism. The alternative approach, based on Rubin's Causal Model (RCM), makes it possible to model the science and the assignment mechanism separately in a two-step procedure. In the first step, no outcome variables are used when the assignment mechanism are modeled, the treated students are matched with similar control students using this mechanism, and the models for the science are determined. Outcome variables are only used in the second step when these pre-specified models for the science are fitted.

In the first paper, we use the traditional approach to evaluate whether a husband is more prone to quit smoking when his wife quits smoking than he would have been had his wife not quit. We find evidence that this is the case, but that our analysis must rely on restrictive assumptions. In the subsequent two papers, we use the alternative RCM approach to evaluate if a Harvard freshman who does not smoke (observed potential outcome) is more prone to start smoking when he shares a suite with at least one smoker, than he would have been had he shared a suite with only smokers (missing potential outcomes). We do not find evidence that this is the case, and the small and insignificant treatment effect is robust against various assumptions that we make regarding covariate adjustments and missing potential outcomes. In contrast, we do find such evidence when we use the traditional approach previously used in the literature to evaluate peer effects relating to smoking, but the treatment effect is not robust against the assumptions that we make regarding covariate adjustments.

These contrasting results in the two latter papers allow us to conclude that there are a number of advantages with the alternative RCM approach over the traditional approaches previously used to evaluate peer effects relating to smoking. Because the RCM does not use the outcome variables when the assignment mechanism is modeled, it can be re-fit repeatedly without biasing the models for the science. The assignment mechanism can then often be modeled to fit the data better and, because the models for the science can consequently better control for the assignment mechanism, they can be fit with less restrictive assumptions. Moreover, because the RCM models two distinct processes separately, the implications of the assumptions that are made on these processes become more transparent. Finally, the RCM can derive the two potential outcomes needed for drawing causal inferences explicitly, which enhances the transparency of the assumptions made with regard to the missing potential outcomes.

KEYWORDS: Peer effects, smoking, Harvard freshmen, quasi-experimental study, Rubin's Causal Model, imputation, propensity score matching, Mahalanobis-metric matching, Bonferroni-adjusted p-levels.

ISBN NR 91-7258-692-3

PRINTED BY:
Elanders Gotab, Stockholm 2005

DISTRIBUTED BY:
EFI, The Economic Research Institute
Stockholm School of Economics
P O Box 6501, SE-113 83 Stockholm
www.hhs.se/efi

Contents

Acknowledgements

Before I began my graduate studies, I knew that the experience would be a challenge. But I never quite understood how daunting a challenge it would be. In fact, the first semester was one of the hardest times in my life. Because I started my graduate studies five years after I finished my undergraduate studies, I had forgotten much of the basic requirements for the first two courses in mathematics and microeconomics. As a consequence, I sat through lectures that were completely incomprehensible to me from beginning to end. But, I got through these two courses, the rest of the coursework, and my thesis work. I would never have seen it through, however, without the support that I was fortunate enough to have from the people around me.

I have been fortunate to have Professor Magnus Johannesson at Stockholm School of Economics as my advisor. Magnus has always been very dedicated to my work and available to discuss research and to comment on my written material on short notice. His support has been vital to me for finishing my dissertation. Magnus also helped me get the most out of my graduate studies by encouraging my wish to spend nearly four years at Harvard University. Without his willingness to direct my research over the phone and his involvement in my applications for funding and other critical factors, I would not have had this opportunity.

I have also had the opportunity to have Professor Donald Rubin at Harvard University as my advisor during the last couple of years. Donald has been very involved in the two last papers in this thesis and he has been tremendously generous, not only with his time, but also with his ideas and feedback. I have learned enormously from the many discussions that we have had, as well as the detailed feedback that I received from him on written work. Not only have I gained a deeper knowledge about the field that Donald has been instrumental in developing, but from him I also learned about the art of writing a paper and the responsibility that comes along with being a researcher. Donald's intuition for statistics and his intuitive interpretation of data have also been great sources of inspiration to me.

Moreover, my dissertation has benefited from the support that I received from Professor Dale Jorgenson at Harvard University, Björn Persson, Jim Greiner, and Gopi Goswami. Professor Jorgenson directed my research especially during my first two years at Harvard, and the enthusiasm he showed for my work was both encouraging and inspiring. He also introduced me to the Harvard community and helped me make the most of my stay there. Björn has been an enormous help throughout my graduate studies not only by supporting me as a friend, but also by always being available to discuss various problems I encountered both during my course work and my thesis work. I am deeply grateful to Jim and Gopi for the opportunity to ask for their help on a number of statistical and coding problems.

Also, the study that this dissertation is mainly based on, i.e., the freshman study, could not have been carried out without the advice I got from Donald and Magnus and the funding that I received through Magnus. The contributions by Justin Ashmall and Peter Brown were also key to its realization. Justin made it, for example, possible to collect all the data on-line through a secure website and Peter stored all the data on a Harvard server. Jane Calhoun (Committee on the Use of Human Subjects), Charles Ducey (Committee on Student Research Participation), and Elizabeth Nathans (Freshman Dean) were also a great support when the study was reviewed for approval. The enrollment of nearly 600 freshmen was facilitated by the cookies that the students received in compensation for their participation and that Stephen Weinberg helped me bake.

Many others have supported my work by commenting on my thesis, by giving me excellent advice concerning choices that I faced during my graduate studies, or even by funding my research. I would especially like to express my gratitude to Professors Jörgen Weibull, Tore Ellingsen, and Bengt Jönsson. The assisance that I have received from Carin Blanksvärd, Pirjo Furtenbach, Ritva Kiviharju, and Dale Rinkel is also gratefully acknowledged. In addition, I would like to thank the Marcus Wallenberg Foundation for Advanced Education in International Industrial Entrepreneurship, Apoteksbolagets Fond för Forskning och Studier i Hälsoekonomi och Socialfarmaci, and Carl Silfvéns Stipendiefond för Studier Utomlands for their financial contributions. Without their contribution, I would never have been able to finance my stay at Harvard University.

Last but not least, I am grateful to my family. My mother Bodil, my father Göran,

and my sister Louise have never been afraid of working hard to reach their goals. Giving up is not even part of their vocabulary. Their attitude towards work has been a great source of inspiration and their encouragement has meant a great deal to me all throughout my graduate studies.

Stockholm in December 2005
Sophie Langenskiöld

Part 1

Introduction

Description of thesis

The causal effect of treatment represents a comparison of the potential outcomes an agent experiences when assigned treatment (observed potential outcome) and the potential outcome the same agent would have experienced when assigned no treatment (missing potential outcome). This thesis explores two different approaches of such causal inference. In the procedure traditionally followed by researchers, the model for the theoretical relationship between the outcome variable and its explanatory variables, i.e., the science, are fit at the same time as they control for the systematic differences that the assignment mechanism created between the treated and the control students. The comparison of the potential outcomes are only implicit, because the missing potential outcomes are not explicitly derived. In an alternative approach based on Rubin's Causal Model (RCM), we can first fit the model for the assignment mechanism. Using this model, we can then design a study where the treated students are similar to, i.e., balance, the control students without using outcome variables. The model for the science can subsequently be fit on this well-balanced designed study. Moreover, the comparison of the potential outcomes need no longer be implicit. It can be explicit, because the missing potential outcomes can now be explicitly derived. The RCM perspective was developed by Rubin in a series of articles (1974, 1975, 1976, 1977, 1978, 1979, 1980).

In the traditional economic approach, causal inference sometimes relies on restrictive assumptions that are not transparent. Consider, for example, the argument between Professors Hoxby (Harvard) and Rothstein (Princeton) concerning the effect of school competition on students' test scores (The Wall Street Journal). The effect is hard to isolate because competition in a school district may not only reflect the competition itself but also other factors that may have an independent effect on students' test scores. The effect of the competition, i.e., causation, must then be separated from the effect of the other factors, i.e., correlation. Hoxby assumes, however, that the number of rivers, especially small rivers, in a school district only influences test scores through the effect it has on school competition. Given that this restrictive assumption holds, the number of rivers creates a variation in school competition that can be used to evaluate the

effect of competition on students' test scores. Using this identification strategy, Hoxby counts the number of rivers in different school districts and finds a large and significant effect of the school competition on students' test score (2000). However, Rothstein also counts this number for different school districts but arrives at different number for some district because, as Rothstein argues, Hoxby's algorithm is not transparent. For that reason, Rothstein cannot replicate Hoxby's results (2005).

In the alternative approach based on RCM, causal inference often relies on less restrictive and more transparent assumptions. Because the RCM does not use the outcome variable when the assignment mechanism is modeled, this model can be re-fit repetitively without biasing the model for the science, the model that is used to evaluate the effect of treatment on outcomes. The model for the assignment mechanism can then often be modeled to fit the data better and, because the model for the science can consequently better control for the assignment mechanism, the model for the science can be fit with less restrictive assumptions. Moreover, because the RCM models two distinct processes separately, i.e., the assignment mechanism and the science, the implications of the assumptions that are made on these processes can be more easily understood, i.e., the assumptions become more transparent. Finally, the RCM can derive the two potential outcomes needed for drawing causal inferences explicitly, and this enhance the transparency of the assumptions made with regard to the missing potential outcomes.

In order to highlight the advantage of the alternative approach of the traditional one, we choose a research question for which causal inference is based on especially restrictive and often not so transparent assumptions. Peer effects on smoking is such a question, because the agent's habits may not only be influenced by his peers' same habits but also by other factors that are correlated with both the agent's and his peers' habits. The former causal effects must then be separated from the latter correlated ones. The papers in this thesis use different agents. In the first paper, we study how a husband's decision to quit smoking is influenced by his wife's smoking habits. In the latter two papers, we study how a Harvard freshman, who does not smoke, is influenced by his roommates' smoking habits.

These questions are interesting not only from a methodological perspective, but also from a policy perspective. Peer effects evaluated in this thesis would be interesting from a policy perspective, because they had enhanced or challenged the effectiveness of various policies. In the first paper, peer effects would mean that a policy maker

needed to consider not only the direct effects of a policy on a husband's smoking but also the possible indirect effects resulting from its direct effects on his wife's smoking. Otherwise, the effects of the policy would be underestimated. In the latter two papers, peer effects would mean that the policy to assign Harvard freshmen into suites without considering their smoking habits may be unethical. Although the freshmen might not be influenced by their roommates' smoking habits in their suites, because smoking was not allowed there, they might be more prone to become friends with their roommates and consequently be influenced by their habits in other environments where smoking was allowed.

Another important issue that arises in this thesis is the importance of data quality. In the first paper that analyzes husbands and wives, we rely on data from the Panel Survey of Income Dynamics (PSID). These data are not ideal for studying peer effects and, consequently, we need to rely on unnecessary restrictive assumptions. For that reason, we decide to collect our own data so that we can relax some of these assumptions. The data is tailored to study how Harvard freshmen are influenced by their roommates' smoking habits or more generally any of their roommates' other health habits. The data collection involves a prospective on-line study that enrolls 600 Harvard freshmen during their first two weeks of classes. They answer the first interview when they enroll in the study and the subsequent two interviews after another five weeks and six months, respectively. The study was approved by the Harvard Institutional Review Board and by the Committee on Student Research Participation at Harvard, and we was granted a Certificate of Confidentiality from the U.S. government.

In the first paper, we use the traditional approach to evaluate if a husband's decision to quit smoking is influenced by his wife's smoking habits. The statistical model for the assignment mechanism is consequently fit at the same time as the model for the science. We find that a husband's probability to quit smoking increases by 24 percentage units if his wife quits. This result is, however, subject to restrictive and non-transparent assumptions and should be interpreted cautiously.

In the latter two papers, we use the alternative approach based on RCM to evaluate if a Harvard freshman who does not smoke is influenced by his roommates' smoking habits. In the first of these papers, outcome variables are not used when the statistical model for the assignment mechanism is fitted and is used to match treated students with similar control students, and when the statistical models for the science are determined. By so doing, we design a study that can be used to evaluate peer effects without

biasing our result by accidently capitalizing on the particular random variation in the outcome variables in our sample. We find that the treated students were more similar, i.e., balanced, to the control students than they would have been in a randomized experiment. In the last of these paper, the pre-specified statistical models for the science are fit and the variation in our estimates of the treatment effects across the different models help us evaluate the robustness of our result to various assumptions made with regard to the covariate adjustments and the missing potential outcomes. We do not find that Harvard freshmen are more prone to begin smoking when they share a suite with at least one smoker than they would have been when they had shared a suite with only non-smokers. The estimates of the treatment effects are close to zero and they are far from significant. Moreover, they are robust to the various assumptions that we make with regard to the covariate adjustments and the missing potential outcomes.

Interestingly, we find peer effects when we instead use the traditional approach that has been used previously to evaluate peer effects relating to smoking. This results is, however, not robust to the assumptions that we make with regard to the covariate adjustments. Furthermore, it is unclear if the result is robust to the assumptions that we make with regard to the missing potential outcomes, because the missing potential outcomes are only implicitly defined in the traditional approach. In other words, we find that it is hard to replicate the results of a well-balanced study using a traditional approach. This result is reminiscent of the classic results of LaLonde (1986). Because the results differ so much depending on the approach, and because the RCM approach leads to more robust results, these findings cast some doubt on previous peer effect analyses on smoking. Are the results in these studies capturing causation or just correlation?

Summary of papers

In this thesis, we explore some of the methods that can be used to evaluate causal effects of peer smoking, and their respective advantages. In the first paper, we start by evaluating peer effects using a traditional approach. In the subsequent two papers, we continue this evaluation using a "new" approach that is based on Rubin's Causal Model (RCM). The RCM perspective was developed by Rubin in a series of articles (1974, 1975, 1976, 1977, 1978, 1979, 1980).

The first paper considers peer effects that might arise between husbands and wives. The aim is to better understand what causes a husband to quit smoking. With deeper knowledge of these causes, we can more effectively address them with policy interventions.

The subsequent two papers, we consider peer effects that might arise among Harvard freshmen in their suites. The purpose is to evaluate the current praxis of not considering the freshmen's smoking habits when they are grouped into suites. If freshmen who do not smoke are encouraged to smoke by their roommates who do smoke, this praxis could be regarded as unethical[1].

Causal inference according to the traditional approach
Peer effects and smoking among husbands and wives

If we observe a simultaneous movement of an agent and his reflection in the mirror, we would not know if the agent caused the reflection to move or the reflection caused the agent to move unless we understand optics (Manski 1993). For the same reason, we cannot know from a joint distribution of an agent's and his peers' smoking habits what caused the observed habits, i.e., the agent's or the peers' habits, unless we understand social interactions. In this paper, the social interaction that occurs between an agent and his peers is assumed to be captured in a linear model. In addition to the endogenous influence between the agent and his peers, the agent and his peers are

[1] The freshmen are not allowed to smoke in their suites, but they are still assumed to influence one another's habits by spending a disproportionate amount of time together.

both assumed to be influenced by observed or unobserved exogenous characteristics, i.e., contextual and correlated effects, respectively.

Manski (1993) demonstrated that the prospects of identifying peer effects in such models using cross-sectional data are poor. If contextual effects are present, our hypothesis of endogenous effects cannot be tested, because endogenous effects cannot be separated from contextual effects. Moreover, if the attributes that directly influence the agent's habits are functions of the attributes that determine his reference group, our hypothesis that the agent's habits reflect the mean peer group habits are always consistent with observed data. Unless we know the agent's peer group, we can thus change the definition of the agent's peer group until the attributes that directly influence the agent's habits becomes a function of the attributes that determine the peer groups so that we can show peer effects.

In this paper, we weaken these assumptions using data from the Panel Survey of Income Dynamics (PSID). We assume that the exogenous characteristics influencing the agents and his peers' smoking habits are constant over time. Then we can produce estimates of peer effects that are not biased by those characteristics, by only using the variations that deviate from last year's variations (difference-in-difference estimators). Furthermore, we choose an agent whose peer group is known – a husband and his wife, whom may be considered his most important peer – and use the agent's description of his peer group's behavior. By so doing, we are less likely to have identified peer effects by manipulating the composition of the agent's peer group, or the agent's perception of the peer group behaviors.

As in other studies, however, first we need to establish if we can use all the observed variations in peer habits to identify peer effects, or if we can use only the part of the observed variations that we consider exogenous to the variation in the husbands' habits. The part of the observed variation in wives' habits that is created by the wives' pregnancies, their babies' birth weights, and their previous smoking habits is argued to be exogenous in this paper. We can then test the hypothesis that the observed variation is exogenous. Because we cannot reject this hypothesis, we use the observed variation to identify peer effects.

We find that the probability that a husband quits smoking increases by 24 percentage units if his wife quits. This means for example that a husband who must quit smoking, perhaps for medical reasons, may have a greater chance of success in a smoking

cessation program if both he and his wife are enrolled in it. Also, it means that a policy maker needs to consider not only the direct effects of a policy on a husband's smoking but also the possible indirect effects resulting from its direct effects on his wife's smoking. Otherwise, the effects of the policy will be underestimated.

This result may be sensitive to the restrictive assumptions we make. For that reason, we should replicate this evaluation in an approach using less restrictive assumptions before we establish that a husband's decision to quit smoking is influenced by his wife's smoking habits. The following two papers present such an approach.

Designing a study using Rubin's Causal Model (Part I) Peer effects and smoking among roommates at Harvard College

Subjects' treatment effects are defined as the difference between the potential outcome they would experience in the world with treatment (treated condition) and the potential outcome they would experience in the world without treatment (control condition). Because only the subjects' experienced potential outcomes can be observed, their missing potential outcomes need to be defined (either explicitly or implicitly) in order to evaluate the treatment effect.

The traditional approach involves models that evaluate such treatment effects at the same time as they control for systematic differences between treated and control subjects. These models often fail to provide adequate fits to the observed data when they are first tried and consequently must be re-adjusted. When re-fitting the model, it may be difficult not to be influenced by what was learned about the relationship between the outcome variables and the treatment. Unless this influence is controlled, the traditional approach could be criticized for capitalizing on random variation in a particular sample to demonstrate results that are either publishable or favorable to some a priori viewpoint.

The traditional approach can also be criticized for defining the missing potential outcomes only implicitly, thereby making an evaluation of the assumptions regarding these outcomes difficult. Also, it models simultaneously processes that are in fact distinct, i.e., the process that creates the differences between the treated and the control students and the process that explains the outcomes. This practice makes it, again, difficult to evaluate how different assumptions relate to these separate processes.

We suggest an alternative approach based on Rubin's Causal Model (RCM) in which

causal inference is less subject to these criticisms. The approach is illustrated by assessing whether Harvard freshmen who do not smoke are more prone to begin smoking when sharing a suite with at least one roommate who smokes (treated condition) than they would have been sharing a suite with only non-smokers (control condition). The traditional approach is split into two steps.

In the first step, conducted in this paper, the outcome variables are left aside, and a study that will be used to evaluate these peer effects is designed. We model the assignment mechanism that creates the systematic differences between the treated and the control students, and by using it to match our treated students with similar control students, we control for these differences. Also, we determine the models for the science that we will use to evaluate the effects themselves and their sensitivity regarding covariate adjustments. These models preferably derive the missing potential outcomes explicitly, but the outcomes can also be defined implicitly in order to evaluate the sensitivity of the results to assumptions regarding them. In the second step, conducted in the next paper, the pre-specified models for the science are fit.

We find that the treated students that we choose to study and their matched control students are more similar than we would have expected in a random experiment. For that reason, covariate adjustments in the final analyses are not needed for other reasons than to improve the precision of the estimates. We therefore plan not to include any covariate adjustments in our primary analyses. Moreover, we decide to estimate the treatment effects when the missing potential outcomes are derived both explicitly and implicitly, in order to evaluate the sensitivity of our estimates to various assumptions we make with regard to the missing potential outcomes.

In the secondary analyses, however, we plan to include various covariate adjustments because we want to see how sensitive the estimates are to such adjustments. We determine the pools of possible covariates from which we pick the covariates, and also the rule with which the covariates are choosen from these pools to be included as covariate adjustments. Again, the missing potential outcomes are derived both explicitly and implicitly.

With this specification of the primary and secondary analyses, we have completed the first step in our approach. We are now prepared to reveal the outcome variables and conduct these primary and secondary analyses in the second step, the topic of the next paper.

Analyzing the results using Rubin's Causal Model (Part II) Peer effects and smoking among roommates at Harvard College

We can now reveal the outcome variables because, irrespective of what we learn about the outcomes, this knowledge cannot influence the final results. The models that determine the final results cannot be changed and consequently nor can the results. Before these models are fit, we need to account for the fact that some of the outcomes may be missing.

Outcomes could be missing, for example, because only the students' observed potential outcomes are revealed, not their missing potential outcomes. These missing outcomes can, however, be explicitly derived in the Bayesian approach by drawing them from their posterior predictive distribution. They can also be implicitly defined in the frequentist's approach by the hypotheses chosen to be tested, e.g., that the treatment effects for all the units, or for the average unit, in the sample are zero.

We choose the Bayesian approach to derive explicitly our missing outcomes, because the assumptions we make with regard to the missing potential outcomes then become transparent and because the resulting complete distributions of the missing potential outcomes can be used to define any estimator of our preference. Also, we prefer this approach because it then becomes clear how we should treat outcomes that are missing for other reasons, i.e., to draw them from their posterior predictive distribution. Here, we are concerned with outcomes that are missing because students drop out from the study or choose not to respond to some questions.

We also use the frequentist's approach to define the missing potential outcomes, however, because we want to evaluate how robust our results are to the assumptions we make with regard to the missing potential outcomes. The missing potential outcomes are implicitly defined by assuming that the average treatment effect is zero, i.e., the Neyman approach (Rubin 1990).

On our first attempt, we fit the statistical models for our primary and secondary analyses. The results are consistent. Students cannot be shown to be more prone to begin smoking when sharing a suite with at least one smoker than they would have been when sharing a suite with only non-smokers. This result is robust to the assumptions we make with regard to the covariate adjustments and the missing potential outcomes in our well-balanced sample.

These results are quite different from those previous studies have found. One reason for the difference could be the more rigorous requirements we impose on our analysis. For example, none of the previous studies are based on a plausible template experiment, a proper design of the study before the analysis is conducted, or adequate attention to missing information. Given these results, we test if our results would change if we used a model previously used in the literature.

One such statistical model that we fit at our first attempt demonstrates a significant treatment effect of 0.08 units. Other statistical models that we fit thereafter demonstrate an even larger and more significant treatment effect. In fact, a range of possible significant treatment effects can be demonstrated by varying the number and the kind of covariates.

We find that if we evaluate peer effects on a sample that is well-balanced in observed covariates, we obtain estimates of peer effects that are robust to the assumptions that we make with regard to covariate adjustments and missing potential outcomes. When we evaluate peer effects at the same time as we control for imbalances in observed covariates, we can only evaluate the sensitivity of our results to covariate adjustments. We find the results not to be robust against such adjustments. This finding is reminiscent of the classic results of LaLonde (1986).

References

Hilsenrath, Jon E (2005) "Novel Way to Assess School Competition Stirs Academic Row" *The Wall Street Journal*, October, 24

Hoxby, Caroline M (2000) "Does Competition among Public Schools Benefit Students and Taxpayers? *The American Economic Review*, 90(5):1209-1238

LaLonde, Robert J (1986) "Evaluating the Econometric Evaluations of Training Programs with Experimental Data" *The American Economic Review*, 76(4):604-620

Rothstein, Jesse (2005) "Does Competition among Public Schools Benefit Students and Taxpayers? A Comment on Hoxby (2000)" *NBER-working paper nr 11215*, March

Rubin, Donald B (1990) "[On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.] Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies" *Statistical Science*, 5(4):472-480

Rubin, Donald B (1979) "Discussion of 'Conditional independence in statistical theory,' by A.P. Dawid" *Journal of the Royal Statistical Society*, B41:27-28

Rubin, Donald B (1978) "Bayesian inference for causal effects: the role of randomization" *Annals of Statistics*, 6:34-58

Rubin, Donald B (1977) "Assignment of treatment group on the basis of a covariate" *Journal of Educational Statistics*, 2:1-26

Rubin, Donald B (1976) "Inference and missing data" *Biometrika*, 63:581-592

Rubin, Donald B (1975) "Bayesian inference for causality: the importance of randomization" *Proceedings of the Social Statistics Section of the American Statistical Association*, 233-239

Rubin, Donald B (1974) "Estimating causal effects of treatments in randomized and nonrandomized studies" *Journal of Educational Psychology* 66:688-701

Part 2

Papers

Causal inference according to the traditional approach

Peer effects and smoking among husbands and wives

ABSTRACT. We contribute to the literature on peer effects of smoking by studying a husband's decision to quit smoking as a function of his wife's smoking. We relax some of the strong assumptions made in previous studies. For example, we study a known peer group and allow for observed or unobserved time invariant exogenous characteristics that may influence both the agent's and his peers' smoking habits. Using data from the Panel Survey of Income Dynamics (PSID), we find a significant peer group effect. The point estimate suggests that the probability that a husband quits smoking increases by 24 percentage units if his wife quits smoking. Because we still make some strong assumptions, this estimate should be interpreted with caution.

1. Introduction

In this paper, we study how an agent's decision to quit smoking is influenced by one of his peers' smoking habits. The main objective of this exercise is to more fully explore the causes behind an agent's decision to quit smoking. Once we understand these causes better, we can more easily consider how to shape policy interventions. This paper also tries to increase our understanding of how an agent's decision to quit smoking is affected by smoking taxation and regulation. In the presence of peer effects, an agent would not only be directly influenced by these smoking policies, but also indirectly influenced through the effects these policies had on his peers. That is, the effects of smoking policies would be multiplied.

To our knowledge, no previous study has analyzed the relationship between an agent's decision to quit smoking and his peers' smoking behavior. However, some previous studies have analyzed the relationship between an agent's and his peers' smoking. These studies either identify peer effects by fitting a reduced form of the statistical model and claiming to use only exogenous variations (Norton et al. 1998; Gaviria et

[0] I am grateful to my advisor Magnus Johannesson who has been crucial to the progress of this paper. Also, I am grateful to Dale Jorgenson who directed my research with great enthusiasm. Sune Karlsson and Björn Persson contributed helpful comments. Financial support from the Marcus Wallenberg Foundation for Advanced Education in International Industrial Entrepreneurship and the Apoteksbolagets Fond för Forskning och Studier i Hälsoekonomi och Socialfarmaci is gratefully acknowledged.

al. 2001; Powell et al. 2003) or by fitting the full statistical model using both exogenous and endogenous variations (Kraut 2002; Kooreman et al 2002; Tamer 2003; Harris 2004; Kraut 2004; Nakajima 2004). We fit a reduced form of the statistical model in this paper.

In previous reduced-form studies, it is not clear whom an agent perceived as his peers and how an agent weighted his peers' respective behaviors. Consequently, the peers could have been identified and their behaviors might have been weighted in such a way that peer effects were demonstrated (Manski 1993). In this paper, we avoid this problem by choosing an agent whose peer group is known – a husband and his wife, whom he may consider to be his most important peer – and use the agent description of his peer group's behavior. We will then achieve more robust estimates of peer effects.

Also in previous reduced-form studies, peer effects were identified by assuming that the agent's and his peers' smoking habits were not influenced by any exogenous characteristics. Instead, potentially exogenous characteristics were used to instrument for peer behavior (Norton et al 1998; Gaviria et al 2001). In this paper, we allow for such observed as well as unobserved exogenous characteristics, as long as they are constant over time, by only using the part of the variations that deviate from last year's variations (a difference-in-difference estimator). Thereby, we have a good chance of separating the endogenous effects from the effects due to exogenous characteristics. However, we still need to instrument for peer behaviors. We use a wife's prior smoking habits, her pregnancies, and her babies' birth weight as instruments.

Section 2 discusses the difficulties of attempts to identify peer effects of this kind. In Section 3, we describe our strategy for overcoming these difficulties and other details relating to its implementation. The data we use to identify peer effects, i.e., the Panel Survey of Income Dynamics (PSID), are described in Section 4. Section 5 presents the results, and we conclude in Section 6.

2. Theory

There are many reasons why we expect peers to have similar smoking habits. It is possible that peers mutually influence their respective smoking habits (endogenous effects), or that their smoking habits are similarly affected by some exogenous characteristics defining them as a group or their environments that, to the researcher, are either known (contextual effects) or unknown (correlated effects). Manski (1993) demonstrated in his seminal paper that the prospects of separately identifying these effects are poor.

To illustrate the difficulty, we follow Manski's demonstration and assume the following representation for an agent's smoking habits y:

$$y = \alpha + \beta E(y|x) + E(z|x)'\gamma + z'\eta + u. \tag{2.1}$$

Here, $E(y|x)$ and $E(z|x)$ are the mean smoking habits and the mean observed exogenous characteristics for the agent's peers who are identified by the attributes x; z and u are the observed and unobserved attributes that directly influence an agent's habits. When $E(u|x,z) = x'\delta$, the linear form of the mean regression of y on x, z is:

$$E(y|x,z) = \alpha + \beta E(y|x) + E(z|x)'\gamma + z'\eta + x'\delta. \tag{2.2}$$

Given that $E(y|x,z)$, $E(y|x)$, and $E(z|x)$ are estimable, integrating the left and the right hand side of this equation with respect to $z|x$ and solving for $E(y|x)$ results in the social equilibrium equation:

$$E(y|x) = \frac{\alpha}{1-\beta} + E(z|x)'\frac{\gamma+\beta\eta}{1-\beta} + z'\eta + x'\frac{\delta}{1-\beta}. \tag{2.3}$$

The parameters in this social equilibrium equation cannot be identified, because there are five parameters and only four estimated coefficients. The parameters that can be identified, however, can be understood by replacing $E(y|x)$ in the mean regression of y on x, z with this social equilibrium equation, i.e., by replacing $E(y|x)$ in Equation 2.2 with Equation 2.3:

$$E(y|x,z) = \frac{\alpha}{1-\beta} + E(z|x)'\frac{\beta\eta+\gamma}{1-\beta} + z'\eta + x'\frac{\delta}{1-\beta}. \tag{2.4}$$

Although the five parameters are still unidentified, we can determine the presence of at least some peer effects when the regressors $[1, E(x|z), z, x]$ are linearly independent. A non-zero coefficient of $E(z|x)$ then indicates that at least one of the endogenous peer effects ($\beta\eta \neq 0$) or contextual peer effects ($\gamma \neq 0$) is present.

The possibility to identify the parameters improves if the assumption of no contextual effects ($\gamma = 0$) and no correlated effects ($\delta = 0$) can be made. In that case, the endogenous effect can be identified if $\eta \neq 0$ and $E(z|x)$ is not a linear function of $[1, z]$:

$$E(y|x,z) = \frac{\alpha}{1-\beta} + E(z|x)'\frac{\beta\eta}{1-\beta} + z'\eta. \tag{2.5}$$

Manski also explained that if x and z are functionally dependent, the statistical model (Equation 2.2) holds tautologically. An agent's behavior would, for example, be a perfect reflection of the mean of his peers' behaviors when z is a function of x (this implies that $\beta = 1$ and $\alpha = \gamma = \delta = \eta = 0$). This means that if we do not know how an agent forms his peers and weights his peers' respective behaviors, we could manipulate

with the composition of the peers and the weighting of their behaviors until z becomes a function of x and peer effects can be established.

3. Empirical strategy

In this paper, we use a panel to identify peer effects, in order to improve our chances of separating the endogenous effects from the contextual or correlated effects. Furthermore, we have an agent whose peer choice and perception of peer behaviors we know. Consequently, the possibility of manipulating the results diminish.

3.1. The valid causal question. We ask how a husband's decision to quit smoking is influenced by his wife's smoking habits. Although a husband may form many peer groups, the peer group represented by his wife probably constitutes the most important one. Moreover, a husband's own perception of his wife's behavior can be found by using data on a family unit level.

However, we need to ascertain that the question we address is truly of causal nature. The causality is not solved unless we can conclude that a husband is more likely to quit smoking if his wife also quits, than he would have been if his wife had instead continued to smoke, i.e., we need to compare the potential outcomes in the two possible scenarios (Rubin 1974). If we can visualize an experiment in which husbands are stochastically assigned an incentive to quit smoking and then manage to replicate this experiment, we have found a way to address this question causally.

Such an experiment could, for example, randomly assign the smoking wives of smoking husbands to participate (treated condition) or not to participate (control condition) in a smoking cessation program. In the presence of peer effects, the encouragement that the wives are exposed to would propagate into an encouragement for their husbands to quit smoking. Consequently, their husbands can be said to be randomly exposed to an encouragement to quit smoking. The causal effect on the husbands' decision to quit smoking of their wives' same decision could then be analyzed for those husbands who comply with their assignment (Hirano et al. 2000).

Unfortunately, we only have a small sample of husbands and wives who both smoke. For that reason, we decide to study all husbands who smoke regardless of whether their wives do. We make the assumption that the husbands' decisions to quit smoking are encouraged or discouraged by the same percentage units when their wives quit or begin smoking. Consequently, we imagine an experiment that randomly assigns

husbands who smoke either to participate in a program that encourages or discourages their smoking (treated condition) or not to participate in such a program (control condition).

3.2. The endogenous separated from the contextual and correlated effects. We do not have such an encouragement-designed experiment at our disposal. Instead, we need to find shocks that assign the husbands to the treated and the control condition respectively, i.e., instruments that have no direct influence on the husbands' decision to quit smoking but that influence them indirectly by influencing their wives' decision to quit or to begin smoking. Furthermore, we need to control for the contextual and correlated effects that cannot be assumed to be the same in the treated and the control condition in such a non-random experiment.

We assume that husbands' smoking habits are only affected by their wives' pregnancies through the effects that the pregnancies have on the wives' smoking decisions[1]. Therefore, we plan to instrument for the wives' smoking habits with indicators for the wives' pregnancy status during the interview, the previous interview, or the interview two years ago. Furthermore, we also assume that husbands are only affected by their wives' prior smoking habits through the effect that these prior habits have on their wives' current ones. Consequently, the wives' smoking habits can also be instrumented with their previous habits (smoking habits two and three years before the interview) or with indicators for their previous habits such as their babies' weights at birth (weights of babies born during the year of the interview, the previous interview, or the interview two years ago). Finally, we need to assume that these instruments similarly affect all wives and influence the wives' decisions to quit as well as to begin smoking.

Furthermore, it is hard to control for the exogenous and the correlated effects unless the same husbands and wives are followed over time. If these effects are assumed to be constant over time, then they can be controlled by using only the part of the variation that deviates from the previous period's variation (difference-in-difference model). We believe that it is reasonable to assume that these effects are constant because we are studying stable marriages, i.e., marriages that not only lasted during the study period in question, i.e., between 1985 and 1993, but even until 2001 or later.

[1] This is a reasonable assumption if we study husbands and wives before 1993, because it was not until 1993 that the hazard of second-hand smoke can be assumed to have reached the general public by means of a report published by the Environmental Protection Agency (EPA). In this report, the EPA stated that the bulk of the scientific evidence demonstrated that second-hand smoke causes lung cancer and other significant health threats to children and adults.

3.3. The linear model as a first attempt. There are different approaches that we can follow to identify how a binary outcome is influenced by an endogenous dummy covariate. Following the traditional approach (Angrist 2001), we would have used a structural model where the 0/1 values of the outcome would have been explained by the influence of an instrument on an unobserved latent variable. From a policy perspective, however, the causal effect on an unobserved latent variable is not important. Only the causal effect on an observed variable is policy relevant. For that reason, the causal effect on an unobserved variable would have needed to be translated to the causal effect on an observed variable. As Angrist argues, however, this translation is unnecessary, because we could have identified the causal effect on an observed variable directly by following the new econometric approaches. In these new approaches, the causal effect of treatment is estimated for those subjects whose treatment status is affected by the instrument.

Among these new approaches, we could have used 2SLS regardless of the support of the outcome if there were no covariates except the treatment indicator, or if the covariates were discrete and sparse. The statistical model for the outcome would then be linear. We could also have used 2SLS when the support of the outcome is continuous and when the treatment effect is assumed constant, an assumption that may not be realistic but that often leads to similar average effects as more general estimation strategies. Also in that case, the model for the outcome would be linear. In both of these cases, the model for the first-stage would best be treated as linear because, as Angrist argues, 2SLS with a linear first-stage achieves consistent estimates even though the first-stage is non-linear, whereas other two-step procedures with a non-linear first-stage only achieve consistent estimates when the first-stage is correctly estimated.

However, we use 2SLS even though the support of the outcome is binary and the treatment effect may not be constant. Therefore, we can only regard our results as first approximations of the causal effects relating a husband's decision to quit smoking to his wife's smoking habits. Any results that we find must consequently be re-estimated in future works using any of the other new econometric methods that, according to Angrist, allow for binary outcomes and non-constant treatment effects, i.e., using a non-linear model (Mullahy 1997) or a linear approximation of a non-linear model (Abadie 1999) to estimate these results. Angrist found, however, that 2SLS differed little from these two latter methods in an empirical exercise that he used to exemplify these other new econometric methods. Consequently, there is at least some support that the 2SLS may not necessarily differ much from these latter methods.

3.4. The adjustments for attrition. As we have mentioned, we use the potential of a panel to identify peer effects. The panel we use consists of all married couples who are present in the PSID between 1985 and 1993, and who not only are present in the 2001 panel but who also remain married at that time. Therefore, we need to control for the attrition caused by the couples who are present in the panel between 1985 and 1993, but who are not present in the 2001 panel or who do not remain married at that time. This attrition can be controlled by estimating the inverse Mills ratio (Heckman 1979), because this ratio is likely to be dependent on information that was collected for the couples when they were present in the panel. When using the inverse Mills ratio, we need to ensure that the standard errors are robust against the heteroskedasticity caused by modelling the attrition as a specification error. Moreover, we need to adjust the standard errors of our final regression if we can reject the hypothesis that the inverse Mills ratio is zero. The standard errors will otherwise be underestimated.

In this paper, we will estimate the inverse Mills ratio based on the propensity scores, i.e., the propensity to remain married in 2001 given a vector of covariates believed to influence this propensity. The propensity scores can help us understand the extent of differences in the multivariate distributions for the couples who can and cannot remain in our panel and, consequently, our potential to adjust for these differences by covariate adjustments. The means of these distributions are represented by the propensity score means for the couples who can and cannot remain in our panel, and their variances are represented by the variances of these propensity scores and of the part of covariates that are orthogonal to these propensity scores. If the standardized difference in propensity-score means between these distributions is less than 0.3, and the ratio of the variances along and orthogonal to the propensity-scores ranges between 4/5 and 5/4, the distributional differences can possibly be controlled by means of covariate adjustments and, thereby, the Mills ratio.

4. Data

In the 2001 wave, the Panel Survey of Income Dynamics (PSID) collected data about the smoking habits of heads of households and of their cohabiters. If the heads of households and/or their cohabiters said they smoked in the wave in question, additional information was collected about the years in which they had begun smoking. If the heads of households and/or their cohabiters said that they did not smoke, information was collected about their previous smoking habits, i.e., whether or not they had ever smoked and if they had, when they had begun and when they had quit. This information made it possible to infer the smoking/non-smoking status for heads of

households and their cohabiters during any year prior to the 2001 waves, even though the PSID waves for those previous years had not collected smoking information. However, these "smoking histories" do not take into account unsuccessful attempts to quit smoking.

The part of the PSID sample, originally an equal probability sample and drawn up by the Study Research Center, is used to derive all those male heads of households who were married to their cohabiters for one or more years between 1985 and 1993, and who were smokers given that they did not have missing values in any variables needed for the analyses (481 marriages/2684 observations). These husbands either remained married and were present in the PSID in 2001 (359 marriages/2112 observations) or were not married or present in the PSID in 2001 (122 marriages/572 observations). The smoking histories of wives are present for the former group of husbands but not for the latter. The analysis can therefore only be done using the former group.

The husbands took the interviews in 70% of the cases, whereas the wives took the interviews in less than 30% of the cases. If the husbands are chosen as the unit of analysis, the variable whose effect is of main interest — the effect of wives' smoking habits — will probably be measured with errors, but the other explanatory variables would not. If, on the other hand, the wives are chosen as the unit of analysis, the variable whose effect is of main interest will probably not be measured with errors, but the other explanatory variables probably will. Since the direction of the bias caused by measurement errors is easier to understand if fewer explanatory variables are measured with errors, the husbands are chosen as the unit of analysis.

The first time the husbands appear in the panel, they are all smokers. Among them, the average husband is 37 (std=9.5) years old and has 12.8 (std=2.2) years of education. The probability that he is employed is 95% (std=22%) and his yearly income is \$25,580 (std=\$16,340) for 2,229 (std=623) hours of work. His wife is 35 (std=9.2) years old and has an educational background similar to her husband, i.e., 12.8 (std=1.9) years. The probability that she is employed is 76% (std=43%), and she earns \$11,360 (std=\$10,200) and works 1467 (std=732) hours a year. Also, the probability that she smokes is 42% (std=49%).

The last time that an average husband appeared in the panel, the probability that he was still a smoker was 82% (std=38%) and the probability that his wife smoked

had decreased to 35% (std=48%)[2]. The probability of being employed had decreased to 91% (std=28%) for him and increased to 88% (std=33%) for her. The average husband also worked fewer hours, i.e., 82 hours, which was probably a consequence of the lower probability of the average man being employed; his wife, however, worked more hours, i.e., 108 hours. Finally, the couple's family income had increased to 58,200 (std=42,730).

Husbands who were married to non-smokers differed from husbands who were married to smokers. Husbands with nonsmoking wives were more likely to be black, had 0.4 years more education, and were 1.8 years older. Their wives had 0.7 years more education, were 1.4 years older, had $2,650 more in labor income in the beginning of the panel and $3,710 more at the end of the panel, probably as a consequence of working 181 and 123 more hours respectively; they were also more likely to be employed in the beginning of the panel and less likely to be unemployed at the end of the panel. These husbands and wives also had a $3,910 and a $6,340 higher family income in the beginning and at the end of the panel.

5. Empirical results

In this section, we find some evidence that the husbands are influenced by their wives' smoking habits. First, however, we derive the Mills ratio that controls for the attrition in our panel.

5.1. Using the Mills ratio to control for attrition. In our panel, we cannot include those smoking men who are married and present in the PSID in some or all of the years between 1985 to 1993, but who are either divorced in 2001 or who no longer are PSID study participants. We control for this attrition by means of the Mills ratio based on the linearized propensity scores. The Mills ratio must then be based on propensity scores that control for all such observed covariates that could influence the likelihood to divorce or to drop out from the study.

We assume that households can better economize on their expenditures and pool their risks if husbands and wives live together rather than separately. For these reasons, we expect to find a higher divorce rate for households with higher labor incomes or total incomes. Furthermore, we also assume that husbands and wives who are recently married, have few children, or have good chances of meeting new partners – perhaps because they are still young or live close to a large city – may be more prone to divorce. They are more independent, and divorce is thus simply less costly for them. We

[2] All comparisons are significant unless otherwise stated.

assume that external stress can also break up a household. For example, we expect long working hours to increase the risk for divorce. Finally, we assume the risk for divorce to be an increasing function of the number of years that remains until 2001. These variables that explain why couples may be divorced in 2001 are all included in the propensity score estimation.

In the propensity score estimation, we must also include the variables that explain why couples may have dropped out from the study before the 2001 interview. Unfortunately, we can only guess at these variables and we would, consequently, need to include many of them in order to believe that we had better controlled for the attrition. We believe that the loss we would incur by dropping the observations with missing values (on any of these variables) is likely to outweigh the gain we would achieve in including more variables in the propensity score estimation. For this reason, we do not add these variables in the propensity score estimation. In future analyses, however, it would be interesting to impute the missing values of the variables. We could then evaluate how robust our results are to the addition of more covariates in the propensity score estimation.

Once we have estimated the propensity scores, we drop those husbands and wives who can be included in our panel, but who have a higher propensity-score than the highest propensity score for those husbands and wives who cannot be included (42). Similarly, we drop those husbands and wives who cannot be included in our panel and who have a lower propensity score than the lowest propensity score for those who can be included (4). Then, we compare the remaining husbands and wives who can and cannot be included in the panel. We find that the standardized mean bias of the propensity scores for these respective couples is 0.44. The ratio of the variances along the propensity scores is 1.35 and the ratios of the variances for the part of these covariates that are orthogonal to the propensity scores range between 4/5 and 5/4 for 67% of the covariates (Figure 1).

We can, therefore, conclude that the propensity score means and variances differ somewhat more than we hoped for between the two groups. The importance of these imbalances can be evaluated by the magnitude and the significance of the Mills ratio in the final regression.

5.2. Estimating the presence of peer effects. Because we intend to identify peer effects by only using the part of agents' variations that deviates from their last year's variations, we need only to control for covariates that change over time. One of

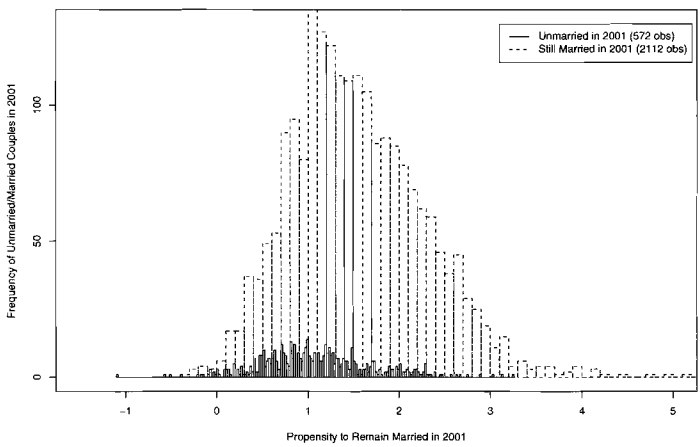

FIGURE 1. Adjusting for propensity score (PS) (Without refinements)
$Bias = 0.44$ and $\sigma^2_M/\sigma^2_{UM} = 1.35$

the most important covariates to control for is price. In addition to price, we also add price interacted with an indicator for those husbands who are 25 years old or younger, because there is evidence that younger people are more price sensitive than adults[3]. Reasons for such increased price sensitivity among younger people could be their lower level of addiction or their greater responsiveness not only to price changes but also to the effect of price changes on their peers (Lewit 1981). Younger people may spend a larger fraction of their disposable income on cigarettes, or may be more present-oriented than adults, which may also make them more price sensitive (Grossman and Chaloupka 1997).

We can come close to isolating the causal effect of price. The price is, for example, not likely to be endogenous because we use individual level data, not aggregated data[4]. Also, the effect of price can be separated from the effect of those attitudes towards smoking that are constant over time, because we use only variations that deviate from last year's variation to identify peer effects, not variations that are constant over time.

[3] Previous papers that have studied the difference in price responsiveness for adults and young adults have tended to use 24 or 25 as the age that separates the former from the latter (Chaloupka 2000).

[4] An individual can hardly be expected to influence any of those prices, whereas an aggregation of individuals may have that influence.

Other covariates that are important to control for are family income and education. Family income and the husband's education should have opposing effects on the husband's decision to quit smoking. A husband's incentive to quit smoking could be expected to decrease as his income increases, because he becomes less cash constrained. Furthermore, a husband's incentive to quit smoking could be expected to increase as his level of education improves because, with education, he becomes more informed about the risks of smoking. We include total family income and income squared, because we are interested not only in the marginal effect of income but also the change in the marginal effect of income as income changes. We also add wage and wage squared, because these covariates can proxy for the marginal effect of education and the change in the marginal effect of education as education changes. Wage and wage squared are good proxy covariates, because the change in an individual's wage over time, i.e., the identifying variation we use, can be assumed to represent the individual's change in education or experience over time.

An individual's tolerance to smoking can also influence a husband's and his wife's smoking habits and, therefore, has the potential to bias the estimates. Race and faith interacted with time dummy variables are added to control for the changes in an individual's tolerance over time. To control for the public attitudes towards smoking that are not constant over time, we add dummy variables for the region and for the size of the greatest city in the county where the husbands live.

Socioeconomic status has been shown to be correlated with different smoking habits. Thus far, we have controlled for income and education, but not for occupational status. We add controls for whether or not the husbands are temporarily unemployed, permanently unemployed, retired, disabled, or for other reasons not working.

We assume that all these covariates are exogenous except the peer effects themselves. Peer effects can, however, be treated as exogenous if the Hausman regression based test for exogeneity cannot be rejected and if this test can be assumed to be valid. In our case, the Hausman test cannot reject exogeneity (F(1,1194)=0.49, p-value=0.48). Moreover the instruments are correlated with the endogenous variable in the first-stage regression ($\chi^2(6) = 143$, p-value< 0.001) and the instruments that are not needed for identification cannot be shown to be correlated with the first differenced errors ($\chi^2 = 0$)[5] (Wooldrigdge 2002). For these reasons, the Hausman test can be assumed

[5] These tests are robust against heteroscedasticity, because heteroscedasticity becomes a problem once a non-linear relationship is approximated with a linear relationship.

to be valid, and peer effects can best be treated as exogeonous. The results of both models are presented in Table 2.

In the OLS regression for first differences, a husband's decision to quit smoking is significantly affected by his wife's smoking habits. The probability that a husband quits smoking increases by 24 percentage units if his wife quits smoking and, consequently, it decreases by 24 percentage units if his wife instead begins smoking (p-value=0.003)[6].

Wage, as a proxy for education, and income have the expected effects. If the husband's hourly wage increases by one USD, the probability that he quits smoking increases by 0.4 percentage units (p-value=0.02), and if the husband's annual income increases by 1,000 USD, the probability that he quits smoking decreases by 0.2 percentage units (p-value=0.07). The marginal effect of wage does not seem to decrease, whereas the marginal effect of income increases (p-value<0.001).

A husband's decision to quit smoking is not significantly affected by cigarette prices. However, a husband who is 25 or younger is 3 percentage units (p-value<0.001) more likely to quit smoking following a one USD price increase than is a husband who is older.

Finally, changes in the effects of race and faith over time seem to matter ($F(6,1637)=4.18$, p-value<0.001 and $F(6,1637)=5.17$, p-value<0.001, respectively). A black husband's probability to quit smoking is 4.5 percentage units lower than a husband of another race in 1987, and this difference increases over time until 1990. In 1992, an atheist is 16 percentage units less likely to quit smoking than someone who does not claim to be an atheist.

The rest of the covariates are insignificant. One insignificant result is worth discussing here. The Mills ratio is not significant in the regression (p-value=0.16). Therefore, the imbalances resulting from the couples who could and could not be included in the panel may not matter and, consequently, the somewhat larger imbalances than adjustable by means of the inverse Mills ratio may not be important. An insignificant Mills ratio also suggests that the standard errors of the regressions do not need to be adjusted. In addition, a husband's decision to quit smoking is not significantly influenced by his occupational status ($F(5,1637)=0.90$, p-value=0.48) or changes in his educational level over time ($F(6,1637)=1.31$, p-value=0.25). The size of the largest city in the county

[6] A wife's decision to start smoking or stop smoking is assumed to have the same effect on her husband's decision to stop smoking but in the opposite way. This assumption is obviously a weakness in the analysis, but hopefully it can be relaxed in an analysis with more observations.

($F(3,1637)=1.82$, p-value=9,14), the geographical region in question ($F(3,1637)=0.92$, p-value=0.4290), and year-specific effects ($F(6,1637)=1.05$, p-value=0.25) also show no significant effects.

The statistical model is now reduced so that dummy variables derived from the same variable are discarded if they are insignificant as a group. This means that dummy variables representing occupational status, changes in education over time, the size of the largest city in a county, a geographical region, and year-specific effects are dropped. Columns two to four in Table 3 show the new regression coefficients, i.e., the OLS coefficients for the first differences, and column five shows the standardized coefficients for the first differences.

The results are similar to those previously reported. A husband's decision to quit smoking is still influenced by his wife's smoking habits and the influence is of the same magnitude (p-value=0.004). The inverse Mills ratio also remains insignificant. However, the husband is not significantly affected by his wage (p-value=0.15) but remains positively affected by his income (p-value=0.0004). Also, the marginal effect of wage is still insignificant (p-value=0.56) and, likewise, the marginal effect of income remains significant (p-value=6.27). Finally, the trend for husbands who are black remains as does the tendency to a trend for husbands who are atheist.

The husbands are now significantly affected by prices. If the price of a packet of cigarettes increases by one USD, the probability that a husband older than 25 quits smoking increases by 0.15 percentage units (p-value=0.091). For a husband no older than 25, this probability increases by another 0.03 percentage units (p-value<0.0001). This corresponds to a price elasticity of 0.207 for husbands older than 25 and an additional price elasticity of 0.046 for husbands no older than 25. We would expect these price elasticities to be higher if we do not control for peer effects. They are higher, i.e., 0.266 and 0.051, but not significantly higher.

Finally, the husband's decision to quit smoking will change more (by more standard deviations) as a result of one standardized change in his wife's smoking habits (0.67) than from a standardized change in any other covariates. Under the assumption that these standardized changes are equally easy to generate, a husband's decision to quit could be influenced the most through his wife's smoking behavior.

6. Conclusion

Under the condition that both husbands and their wives smoke, we could evaluate how the husbands are influenced by their wives' decision to quit by randomly assigning the wives to participate (treated group) or not to participate (control group) in a smoking cessation program (Section 3.1). In this paper, such an experiment is replicated as closely as possible. We assume: (1) that there are exogenous shocks, i.e., instruments, that can "assign" the husbands into the treated or control conditions; (2) that exogenous characteristics that define a husband's choice of wife or their environments are constant over time and that a difference-in-difference regression can consequently be used to control for these characteristics (Section 3.2); (3) that the inverse Mills ratio can be used to control for the differences between husbands and wives that can and cannot be included in our analyses (Section 3.4). Under these assumptions, we can evaluate how a husbands' decision to quit smoking is influenced by his wife's decision using a linear model (Section 3.3).

We find that peer effects can be evaluated without instrumenting for peer behaviors and that the probability that a husband quits smoking increases by 24 percentage units (p-value=0.004) if his wife quits smoking. This means, for example, that a husband who must quit smoking, perhaps for medical reasons, may have a greater chance of success in a smoking cessation program if not only he but also his wife are enrolled in a smoking cessation program. Also, it means that a policy maker need to consider not only the direct effects of a policy on a husband's smoking but also the possible indirect effects resulting from its direct effects on his wife's smoking. Otherwise, the effects of the policy will be underestimated.

Moreover, we find that a husband's decision to quit is significantly affected by price. The price elasticity is 0.207 for husbands who are older than 25, with an additional price elasticity of 0.046 for husbands who are 25 or younger. These elasticities are however not significantly higher when peer effects are not controlled for, i.e., 0.266 and 0.051, as might have been expected (Section 5.2). Finally, we find that the inverse Mills ratio is not significant, which means that the differences we find between husbands and wives that can and cannot be included in our panel may not be important (Section 5.1).

There are a number of reasons why we should interpret these results with caution. For example, we base our conclusions on the critical assumption that the wife's smoking habits are not endogenous. The validity of this assumption cannot, however, be verified because it is only possible to evaluate the hypothesis of no exogenous effects

and not the hypothesis of no endogenous effects. Moreover, we assume that there is a linear relationship between a husband's decision to quit smoking and his wife's change in smoking habits. This relationship could very well be non-linear. We also assume that the percentage units' increase in the husband's probability to quit following his wife's decision to quit smoking is the same as the percentage units' decrease in the husband's probability to quit following his wife's decision to begin smoking. Such an assumption is excessively restricted. Finally, we also assume that we are able to control adequately for exogenous characteristics that can influence both a husband's and his wife's smoking habits and for differences between the couples that can and cannot be included in our panel. However, exogenous characteristics that vary over time are not controlled and unobserved differences between couples who can and cannot be included in the panel also are left uncontrolled. None of these assumptions are innocuous and we should consequently evaluate how robust our results are to them. An ideal way to do this would be to run the random experiment described in the beginning of this section and in Section 3.1.

7. Tables

Table 1: Descriptive statistics reported for all 359 marriages and for the 152/207 marriages whose wives smoked/did not smoke. The statistics describe the marriages when they first and last appeared in the panel.

	Mean All	Std All	Mean Wife non-smoker	Std Wife non-smoker	Mean Wife smoker	Std Wife smoker	Difference t-value
Smoke first (wife)	0.42	0.49	0	0	1	0	−Inf
Smoke last (wife)	0.35	0.48	0.01	0.1	0.82	0.39	−35.38
Smoke first (husband)	1	0	1	0	1	0	
Smoke last (husband)	0.82	0.38	0.81	0.39	0.84	0.37	−1.07
Year married first	12.51	9.64	12.67	10.06	12.3	9.06	0.51
Year present	5.88	2.18	5.94	2.18	5.81	2.17	0.78
White (husband)	0.96	0.19	0.95	0.21	0.98	0.14	−2.16
Black (husband)	0.03	0.17	0.04	0.2	0.01	0.11	2.53
Atheist (husband)	0.09	0.28	0.09	0.29	0.08	0.27	0.61
Age first (husband)	37.07	9.47	37.85	9.35	36.02	9.55	2.55
Age first (wife)	34.66	9.23	35.24	9.23	33.88	9.2	1.95
Education first (husband)	12.84	2.17	13.02	2.19	12.59	2.12	2.7
Education last (husband)	12.84	2.17	13.02	2.19	12.59	2.12	2.7
Education first (wife)	12.79	1.88	13.1	1.92	12.38	1.74	5.25
Education last (wife)	12.79	1.88	13.09	1.93	12.38	1.74	5.22
Family income first	41.94	25.85	43.6	29.47	39.69	19.71	2.13
Family income last	58.2	42.73	60.88	49.16	54.54	31.68	2.1
Labor income first (husband)	25.58	16.34	25.59	17.57	25.57	14.53	0.02
Labor hours first (husband)	2229	623.2	2210	642.8	2255	595.5	−0.95
Labor income last (husband)	32.82	25.91	33.47	29.16	31.93	20.71	0.83
Labor hours last (husband)	2147	705.3	2126	744.2	2175	648.7	−0.95
Labor income first (wife)	11.36	10.2	12.48	11.5	9.83	7.85	3.66
Labor hours first (wife)	1467	732	1544	705.3	1363	755.6	3.26
Labor income last (wife)	17.46	17.88	19.03	20.63	15.32	12.98	2.95
Labor hours last (wife)	1575	663.4	1627	639.4	1504	689.5	2.42

Continued on next page

Table 1 continued

	Mean All	Std All	Mean Wife non-smoker	Std Wife non-smoker	Mean Wife smoker	Std Wife smoker	Difference t-value
Working first (husband)	0.95	0.22	0.95	0.22	0.95	0.21	−0.43
Unemployed first (husband)	0.03	0.17	0.04	0.19	0.02	0.14	1.52
Working last (husband)	0.91	0.28	0.9	0.3	0.93	0.26	−1.17
Unemployed last (husband)	0.06	0.23	0.07	0.25	0.04	0.2	1.69
Working first (wife)	0.76	0.43	0.8	0.4	0.71	0.45	2.8
Unemployed first (wife)	0.04	0.21	0.04	0.19	0.05	0.22	−0.88
Working last (wife)	0.88	0.33	0.89	0.31	0.86	0.35	1.08
Unemployed last (wife)	0.04	0.21	0.03	0.17	0.07	0.25	−2.24
Children at home first	1.31	1.17	1.3	1.15	1.33	1.19	−0.33
Children at home last	1.16	1.14	1.14	1.15	1.18	1.14	−0.38
City 10'-24' inhabitants first	0.21	0.4	0.18	0.39	0.24	0.43	−1.72
City 25'-99' inhabitants first	0.28	0.45	0.29	0.45	0.27	0.44	0.59
City 100' inhabitants first	0.32	0.47	0.33	0.47	0.31	0.46	0.68
City 10'-24' inhabitants last	0.2	0.4	0.2	0.4	0.21	0.41	−0.41
City 25'-99' inhabitants last	0.28	0.45	0.29	0.45	0.26	0.44	0.79
City 100' inhabitants	0.31	0.46	0.31	0.46	0.32	0.47	−0.19
Midwest first	0.36	0.48	0.35	0.48	0.38	0.49	−0.79
South first	0.28	0.45	0.29	0.46	0.26	0.44	1.13
Northeast first	0.19	0.4	0.18	0.39	0.21	0.41	−0.89
Midwest last	0.34	0.47	0.32	0.47	0.37	0.48	−1.24
South last	0.29	0.45	0.32	0.47	0.25	0.43	2.03
Northeast last	0.2	0.4	0.19	0.39	0.21	0.41	−0.73

Table 2: The difference-in-difference for the OLS and 2SLS regressions is based on 1680 and 1239 observations, respectively. The reported standard errors are robust against autocorrelation and heteroscedasticity.

Variable	**OLS**		(Std. Err.)	**2SLS**		(Std. Err.)
Spouse smoke	0.23986	**	(0.07943)	0.65456	*	(0.29892)
Wage/hour (husband)	0.00383	†	(0.00212)	0.00005		(0.00176)
(Wage)2/hour (husband)	-0.00001		(0.00002)	0.00001		(0.00001)
Family income in $1000	-0.00168	**	(0.00061)	0.00018		(0.00066)
(Family income)2 in $1000	0.00001	**	(0.00000)	0.00000		(0.00001)
Temporary unemployed (husband)	-0.00527		(0.04639)	0.02466		(0.06333)
Unemployed (husband)	-0.00231		(0.02038)	0.00613		(0.02028)
Retired (husband)	0.12412		(0.13562)	-0.96973	**	(0.01973)
Disabled (husband)	0.01590		(0.02542)	-0.00534		(0.02554)
Other reason for not working	-0.06263		(0.03942)	-0.00996		(0.03165)
Cigarette price*Youth (husband)	0.02991	**	(0.00924)	0.01363		(0.00931)
Cigarette price	0.11997		(0.09283)	-0.01483		(0.13136)
10'-24' inhabitants	0.00088		(0.03141)	0.02987		(0.06091)
25'-99' inhabitants	-0.07550		(0.06086)	-0.14935		(0.09690)
More than 100' inhabitants	-0.06370		(0.04209)	-0.09360		(0.06034)
Midwest	0.00821		(0.03516)	-0.02948		(0.03898)
South	-0.04672		(0.05411)	-0.01026		(0.03958)
Northeast	0.21886		(0.17942)	0.02526	†	(0.01468)
Mills ratio	0.37758		(0.27160)	0.35142		(0.37587)
1992	-0.10051		(0.19372)			
1991	-0.27075	†	(0.15831)	-0.07172		(0.09380)
1990	-0.24494	†	(0.14314)	-0.05503		(0.10237)
1989	-0.18285	†	(0.10207)	0.00406		(0.10120)
1988	-0.05309		(0.05996)	0.06024		(0.07257)
1987	-0.03544		(0.04551)	0.05884		(0.05525)
Education*1992 (husband)	0.01575		(0.01427)	0.01613		(0.01584)
Education*1991 (husband)	0.02734	*	(0.01196)	0.01920		(0.01332)
Education*1990 (husband)	0.02459	*	(0.01112)	0.01496		(0.01222)
Education*1989 (husband)	0.01732	*	(0.00811)	0.00619		(0.00876)
Education*1988 (husband)	0.00548		(0.00460)	-0.00157		(0.00385)
Education*1987 (husband)	0.00437		(0.00370)	-0.00201		(0.00377)
Atheist*1992 (husband)	-0.15858	*	(0.06887)	-0.08700		(0.08868)
Atheist*1991 (husband)	-0.10896		(0.06687)	-0.07093		(0.08810)
Atheist*1990 (husband)	-0.06003		(0.06372)	-0.00492		(0.07448)
Atheist*1989 (husband)	-0.01080		(0.06204)	0.03352		(0.07309)
Atheist*1988 (husband)	0.01754		(0.06077)	0.05548		(0.07176)
Atheist*1987 (husband)	0.02646		(0.06019)	0.04817		(0.07163)
Black*1992 (husband)	0.04657		(0.17645)	0.28872		(0.27854)
Black*1991 (husband)	-0.04877		(0.11516)	0.05985		(0.16682)
Black*1990 (husband)	-0.14618	**	(0.03259)	-0.10127	**	(0.03527)
Black*1989 (husband)	-0.09000	**	(0.02920)	-0.05238	*	(0.02520)
Black*1988 (husband)	-0.05194	*	(0.02441)	-0.01798		(0.01736)

Continued on next page

Table 2 continued

Variable	**OLS**		(Std. Err.)	**2SLS**	(Std. Err.)
Black*1987 (husband)	-0.04511	*	(0.02113)	-0.02612	(0.01605)
Intercept	0.01756		(0.01175)	-0.00791	(0.03444)

Table 3: The difference-in-difference regression that results from an exclusion of those dummy variables in Table 2 that represent the same variable and are insignificant as a group.

Variable	**OLS**		(Std. Err.)	**2SLS**		(Std. Err.)	**Standardized OLS**		(Std. Err.)
Spouse smoke	0.23416	**	(0.08028)	0.68650	*	(0.29817)	0.67103	**	(0.23005)
Wage/hour (husband)	0.00330		(0.00209)	0.00039		(0.00186)	0.18277		(0.11594)
(Wage)2/hour (husband)	-0.00001		(0.00002)	0.00000		(0.00001)	-0.05220		(0.09026)
Family income in $1,000	-0.00176	**	(0.00061)	0.00011		(0.00068)	-0.30181	**	(0.10475)
(Family income)2 in $1,000	0.00001	**	(0.00000)	0.00000		(0.00001)	0.28327	**	(0.04516)
Cigarette price*Youth	0.03244	**	(0.00640)	0.02307	**	(0.00711)	0.03437	**	(0.00679)
Cigarette price	0.14587	†	(0.08627)	-0.02649		(0.12264)	0.13139	†	(0.07770)
Mills ratio	0.14389		(0.17531)	0.41061	†	(0.22967)	0.03633		(0.04426)
Atheist*1992 (husband)	-0.16050	*	(0.06654)	-0.11404		(0.08828)	-0.09210	*	(0.03819)
Atheist*1991 (husband)	-0.11503	†	(0.06369)	-0.08835		(0.08581)	-0.07074	†	(0.03917)
Atheist*1990 (husband)	-0.06416		(0.06075)	-0.01674		(0.07230)	-0.03772		(0.03571)
Atheist*1989 (husband)	-0.03029		(0.05951)	0.00089		(0.07069)	-0.01695		(0.03329)
Atheist*1988 (husband)	0.00095		(0.05870)	0.02412		(0.06973)	0.00047		(0.02921)
Atheist*1987 (husband)	0.02988		(0.05829)	0.04602		(0.06928)	0.01627		(0.03175)
Black*1992 (husband)	0.05967		(0.17897)	0.27192		(0.27362)	0.02032		(0.06095)
Black*1991 (husband)	-0.04242		(0.11747)	0.04811		(0.16247)	-0.01638		(0.04534)
Black*1990 (husband)	-0.12962	**	(0.01998)	-0.09233	**	(0.02824)	-0.05003	**	(0.00771)
Black*1989 (husband)	-0.08823	**	(0.01520)	-0.07144	**	(0.02119)	-0.03589	**	(0.00618)
Black*1988 (husband)	-0.05758	**	(0.01081)	-0.04938	**	(0.01540)	-0.02342	**	(0.00440)
Black*1987 (husband)	-0.02460	**	(0.00574)	-0.02945	**	(0.00880)	-0.01001	**	(0.00233)
Intercept	0.02907	**	(0.00627)	0.02976	**	(0.00852)	0.17052	**	(0.03677)

References

Abadie, Alberto (2003) "Semiparametric Instrumental Variable Estimation of Treatment Response Models" *Journal of Econometrics*, 113(2):231-263

Angrist, Joshua D. (2001) "Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice" *Journal of Business & Economic Statistics*, 19(1), 2-16

Chaloupka, Frank J. and K. E. Warner (2000) "The Economics of Smoking" *Handbook of Health Economics*, 1539-1565

Environmental Protection Agency (2003) "Respiratory Health Effects of Passive Smoking: Lung Cancer and Other Disorders" *EPA/600/6-90/006F*

Gaviria, Alejandro and Steven Raphael (2001) "School-Based Peer Effects and Juvenile Behavior". *The Review of Economics and Statistics*, 83(2), 257-268

Grossman, M., Frank J. Chaloupka (1997) "Cigarette taxes: The straw to break the camel's back" *Public Health Reports*, 112(4), 290-7

Harris, Jeffrey, E, (2004) "Asymmetric social interaction in economics: Cigarette smoking among young people in the United States, 1992-1999" *NBER Working Paper nr 10409*, April

Heckman J. (1979) "Sample Selection Bias as a Specification Error" *Econometrica*, 47: 153-161

Hirano, K., G. Imbens, D. Rubin, X. Zhou (2000) "Assessing the Effect of an Influencza Vaccine in an Encouragement Design" *Biostatistics*, 1, 69-88

Kooreman, Peter, Adriaan Soetevent (2002) "A Discrete Choice Model with Social Interactions; an Analysis of High School Teen Behavior" *Working Paper IDEAS*, January

Krauth, Brian V (2005) "Peer effects and selection effects on smoking among Canadian youth" *Canadian Journal of Economics*,38(3):735-757

Krauth, Brian (2004) "Simulation-based estimation of peer effects" *Working Paper IDEAS*, July

Laibson, David (2000) "A Cue-Theory of Consumption" *Quarterly Journal of Economics*, 66(1), 81-120

Lewit E. M., D. Coate (1981) "The effects of government regulation on teenage smoking" *Journal of Law and Economics*, 24(3), 545-69

Manski, Charles F. (1993) "Identification of Endogenous Social Effects: The Reflection Problem" *Review of Economic Studies*, 60, 531-542

Mullahy, John (1997) "Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behaviour" *Review of Economics and Statistics*, 11:586-593

Nakajima, Ryo (2004) "Measuring peer effects on youth smoking behavior" *Osaka University Social and Economic Research Paper No 600*, March

Norton, Edward C., Richard C. Lindrooth, and Susan T. Ennett (1998) "Controlling for the endogeneity of peer substance use on adolescent alcohol and tobacco use" *Health Economics*, 7, 439-453

Powell, Lisa A., John A. Tauras, and Hana Ross (2003) "Peer Effect, Tobacco Control Policies, and Youth Smoking Behavior" *Journal of Health Economics*, 24(5):950-968

Rubin, Donald B (1974) "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies" *Journal of Educational Psychology*, 66(5): 688-701

Tamer, Elie (2003) "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria" *Review of Economic Studies*, 70, 147-165

Wooldridge, Jeffrey M (2002) "Econometric analysis of cross section and panel data", The MIT Press, Cambridge

Designing a study using Rubin's Causal Model (Part I)

Peer effects and smoking roommates at Harvard College

ABSTRACT. In quasi-experimental and observational studies, the causal effect of treatment cannot be estimated without controlling for the systematic differences between treated and control subjects. In a sequence of two papers, we suggest an approach for these studies that first, without having access to outcome data, balances these differences and defines appropriate models for the key outcome analyses. Then, when access to outcome data is gained, these models can be fitted without repeated attempts. In this document, we design a hypothetical randomized experiment, studying peer effects related to smoking, that is close to an actual quasi-experimental study. We find that the balance of the observed covariates in the designed study is better than would be expected in a randomized experiment. Furthermore, we define the models of the key outcomes concerning smoking behavior that we commit to run in our next document. Finally, we gain an understanding of which treatment effects we can expect to be real and important in the real outcome analyses by conducting practice analyses on intermediate outcomes using our design.

1. Introduction

The causal effect of a treatment on a subject is defined as a comparison of two potential outcomes: the outcome that would be observed in the world with treatment (treated condition) and the outcome that would be observed in the world without treatment (control condition). It is not possible to observe both of these potential outcomes at the same time – one or the other is always missing. Hence, causal inference can be addressed only once this missing data problem has been addressed (Rubin 1974).

In principle, the missing data problem can be solved if the rule used to assign treatments is stochastic. This stochastic rule may be known, as it is in randomized experiments. In a completely randomized design, the subjects differ systematically only with respect

[0] This research would not have been possible without the support and encouragement of my adviser Donald Rubin, who provided many insights and gave detailed feedback on all aspects of this work. It has also benefited from discussions with and comments by my adviser Magnus Johannesson. Jim Greiner and Björn Persson also contributed helpful comments. All remaining errors are mine. I would like to thank Justin Ashmall and Peter Brown for their technical expertise in running the freshman study on-line. Also, I am grateful to the Marcus Wallenberg Foundation for Advanced Education in International Industrial Entrepreneurship and the Apoteksbolagets Fond för Forskning och Studier i Hälsoekonomi och Socialfarmaci for financial support.

to their treatments. The subjects also differ systematically in their values of the covariates used in the randomization in experiments with randomized block design. In such randomized experiments, researchers can estimate the treatment effects simply by comparing treated subjects and controls, or treated subjects and controls who have similar values of the covariates used in the randomization. This stochastic rule may also be completely or partly unknown, as it is in studies that because of time constraints, cost considerations, or even ethical concerns cannot rely on randomization assignment. In such observational studies, the subjects will not only differ in their treatments but also in the covariates, known or unknown, believed to have influenced their treatments.

The traditional approach among many researchers, including many economists, is to estimate the treatment effects while making adjustments for the factors believed to differ systematically between treated subjects and controls. Elaborate models are used for this purpose. These models may often fail to provide adequate fits to observed data when they are first tried, and may need to be adjusted. They may even need repeated adjustments before their fit to observed data is considered satisfactory. Inevitably, researchers will adjust their models based on what they learned in their previous attempts about the apparent relationship between outcome variables and exposure to treatments. Unless they allow for the consequences of having re-adjusted their models, e.g., on the estimated treatment effects' magnitudes and precisions, the researchers could be viewed as capitalizing on random variation in an attempt to obtain results that are either publishable, or favorable to some a priori viewpoint.

The intention here is not to accuse some researchers of deception to obtain desired results. Rather, it is to emphasize the challenges to overcome when this traditional approach is used for causal inference; either researchers must succeed in fitting their models at their first attempts, or they will need to make the necessary allowances. This document suggests an alternative approach in which causal inference is less subject to these criticisms.

We split the traditional approach, which assesses treatment effects at the same time as it controls for systematic differences between treated subjects and controls, into a two-step process. *Step 1:* Without using outcome data, we design a study by matching the treated students with control students who are similar in their observed covariates. Thereby, we control for the systematic differences between the treated and the control students. *Step 2:* We use the designed study to estimate the treatment effects on outcome data, possibly in conjunction with pre-specified model-based adjustments. These

two steps should be distinct, just as the design and analysis of a randomized trial use two distinct steps separated in time. In this setup, design can influence analysis, but analysis can never influence design. The only exception may be the requirements that certain analyses put on the design, e.g., studying certain primary outcomes when some covariate values are missing requires, for fair estimation of treatment effects, that these missing values are dealt with correctly. Henceforth, we will refer to these two steps as the "design" and the "outcome analysis" phases.

We illustrate our approach in the context of an assessment of peer influences on cigarette smoking among college freshmen based on a quasi-randomized experiment conducted at Harvard College during the academic year 2003/2004 (Part 3). More precisely, we look at those peer influences that the freshmen under because of their roommates' smoking habits. In order to highlight the two-step "design" and "outcome analysis" nature of our approach, we present the two respective steps in two separate documents. In the present paper, we do not use, nor see, any outcome data for any analyses.

In Section 2, we discuss the tools we use to control the observed systematic differences between treated and control students. The data generating process in the Freshman Study is described in Section 3, because a clear understanding of this process and of how it differs from an ideal "template" randomized experiment help us use design tools correctly. After a short description of the preparation of covariate data in Section 4, we use these tools in Section 5 to replicate, as closely as possible, a randomized study from our quasi-randomized study. We learn that our tools have created better balance in observed covariates in the treated and control groups than would be expected to be observed in a completely randomized experiment. Of course, a randomized experiment would still be superior because it would also balance unobserved covariates. Finally, we conduct "practice" outcome analyses in Section 6. These not only help us define the models to use when the actual treatment effect is estimated, but also provide us with benchmarks for evaluating the magnitudes of peer effects that are to be regarded as real and important. Section 7 concludes.

2. Theory and perspectives

2.1. General remarks regarding causal inference. Causal inference requires us to imagine that the units of interest can be treated in alternate ways. For example, a non-smoking freshman who enrolls could either be assigned a suite with at least one roommate who smokes or a suite with only non-smokers. For each unit of interest, i.e.,

each such non-smoker, causal inference for a outcome variable, such as smoking behavior in six months, then involves a comparison of the "potential outcomes" that would be observed under the alternate treatments. The potential outcome when sharing a suite with at least one roommate who smokes, and the potential outcome when sharing a suite with only non-smokers, must be estimated and compared for each freshman, either explicitly or implicitly.

The causal effect of treatment for a single individual i, τ_i, is then a comparison of the potential outcomes when i is assigned the treated condition, y_{1i}, or the control condition, y_{0i}, respectively:

$$\tau_i \equiv y_{1i} - y_{0i}. \tag{2.1}$$

For a group of individuals, the causal effect of treatment also involves a comparison of the potential outcomes for each of the individuals constituting the group. However, there must be no interference between the individuals and no hidden version of the treatments. These assumptions are known as SUTVA – the Stable Unit Treatment Value Assumption (Rubin 1980). In addition, the method used to learn about the potential outcomes cannot influence the potential outcomes or the covariates.

The objective is often to estimate the causal effect for a particular subgroup of individuals. In our study, for example, we are interested in the causal effect on the treated students, those non-smokers living with smokers. If this causal effect prevail, we can prevent non-smokers from beginning to smoke by not letting them share a suite with smokers. Consequently, our policy recommendation would be that Harvard College changed its system of mixing smokers with non-smokers in favor of a system of separating smokers and non-smokers. The treatment effect referred to in this document is thus the treatment effect on the treated, i.e., the non-smokers living with smokers. This type of question goes back to at least Peters (1941).

2.2. Causal inference in Rubin's Causal Model (RCM). In causal inference, it is necessary to draw inferences about at least half of the potential outcomes that are missing. Many assumptions may be needed for this amount of missing information, and some sensitivity of the causal inference to these assumptions may be inevitable. For that reason, it is important to present these assumptions in a clear way so that the basis for conclusion not only can be easily understood but can also critically evaluated.

In the RCM, the systematic differences that are created between the treated and the control students because of the way the treatments are assigned are modeled separately

from the theoretical relationships on which the causal inferences are based, i.e., the relationships between the potential outcomes, treatment exposure, and other covariates. These models are known as the assignment mechanism and the science, respectively (e.g., Rubin 2005). Once these models are derived, they can be used to derive the posterior predictive distribution of the missing potential outcomes (Rubin 1978). The missing data problem is then solved, because the missing potential outcomes can then be drawn from this posterior predictive distribution (Rubin 2005).

Thus, the RCM clearly distinguishes between the assumptions made with regard to the assignment mechanism and those made with regard to the science, because it models these two processes separately but in a unified framework. The implications of the assumptions that are made on these distinct processes then become transparent. Furthermore, it allows the explicit derivation of the missing potential outcomes. Thereby, the assumptions regarding the missing potential outcomes become transparent as well. The models traditionally used to estimate the treatment effects consider the assignment mechanism and the science in the same model. In addition, such models never explicitly derive the the missing potential outcomes. Consequently, the conclusions drawn using the traditional models can be both confusing and confused. For a further discussion of the RCM, see Holland (1983) and Angrist et al. (1996).

We will start off this sequence of documents by describing the model for the assignment mechanism, since this is the only model needed for the design. The model for the science is deferred to the next document, because it is only used for some outcome analyses. Causal inference that relies on the variation in treatments created by the assignment mechanism can often be drawn with less restrictive assumptions than causal inference that relies on the variation in treatments created by the science, because the assignment mechanism is often under the control of the researcher whereas the science never is. For that reason, much of causal inference is based on the model for the assignment mechanism only, simply consider examples of randomization-based analyses (Rubin 1977), and the model merits to be discussed in a separate document.

2.3. Model for the assignment mechanism. The model for the assignment mechanism determines which outcomes will be revealed for individuals in the population $i = 1, ..., n$:

$$Pr(\mathbf{w} \mid \mathbf{X}, \mathbf{y}_1, \mathbf{y}_0) \propto \prod_{i=1}^{n} Pr(w_i \mid \mathbf{x}_i, y_{1i}, y_{0i}), \tag{2.2}$$

where, for individual i, w_i indicates the assignment, where $w_i = 1$ implies treatment and $w_i = 0$ implies control; $\mathbf{x}_i = [x_{i1}, ..., x_{im}]$ reflects the m covariate values used in the assignment; and y_{1i} and y_{0i} represent the potential outcome when assigned the treated and the control condition, respectively. Consequently, the assignment mechanism gives the probability of a vector of assignments $\mathbf{w} = [w_1, ..., w_n]$ conditional on a matrix of covariate values $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_n]'$ and the vectors of potential outcomes when assigned the treated condition, $\mathbf{y}_1 = [y_{11}, ..., y_{1n}]$, and the control condition, $\mathbf{y}_0 = [y_{01}, ..., y_{0n}]$, respectively[1]. The observed and missing potential outcomes for individual i in the population are:

$$y_i^{obs} = w_i y_{1i} + (1 - w_i) y_{0i}, \tag{2.3}$$

$$y_i^{mis} = (1 - w_i) y_{1i} + w_i y_{0i}, \tag{2.4}$$

respectively. The assignment mechanism is ignorable[2] (Rubin 1976, 1978) if it only depends on observed potential outcomes in addition to the values of covariates used in the assignment of treatments:

$$Pr(\mathbf{w} \mid \mathbf{X}, \mathbf{y}_1, \mathbf{y}_0) = Pr(\mathbf{w} \mid \mathbf{X}, \mathbf{y}^{obs}), \tag{2.5}$$

where $\mathbf{y}^{obs} = [y_1^{obs}, ..., y_n^{obs}]$. It is unconfounded if it does not depend on any potential outcomes at all:

$$Pr(\mathbf{w}|\mathbf{X}, \mathbf{y}^{obs}) = Pr(\mathbf{w}|\mathbf{X}). \tag{2.6}$$

An unconfounded assignment mechanism, which is an ignorable assignment mechanism, implies that all students with the same values of the covariates have the same probability of being assigned the treated or the control condition. The treatment effect can then be estimated non-parametrically by matching treated students with control students whose covariate values are the same.

Randomized experiments have ignorable, usually unconfounded, assignment mechanisms. They have become the gold standard in causal inference, because very few assumptions are needed for the estimation of the treatment effect. The assignment mechanism in such randomized experiments can be described intuitively by tosses of a coin at each particular covariate combination, to determine the assignment of the treated and the control conditions. If the assignment is unconfounded, the bias of the coin depends only on $\mathbf{X}$; if the assignment mechanism is confounded but ignorable, the bias of the coin depends on both $\mathbf{X}$ and $\mathbf{y}^{obs}$, but not on $\mathbf{y}^{mis}$ where $\mathbf{y}^{mis} = [y_1^{mis}, ..., y_n^{mis}]$.

[1] The values of $\mathbf{X}$, $\mathbf{y}_1$, and $\mathbf{y}_0$ are fixed but the values that are revealed are stochastic and determined by $\mathbf{w}$.

[2] The corresponding economic term is selection on observables.

In cases when the dimension of $\mathbf{X}$ is high and/or the number of subjects is few, exact matches for the treated students may not be found among the control students. A nearly unbiased estimate of the treatment effect may, nonetheless, be achieved by matching treated with control students based on their propensity scores. The propensity score is defined as the conditional probability of being treated given the values of a vector of observed covariates:

$$p(\mathbf{x}_i) \equiv pr(w_i \mid \mathbf{x}_i). \tag{2.7}$$

The assignment mechanism will, thus, be dependent on a univariate covariate, i.e., the propensity score, instead of a vector of covariates $\mathbf{x}_i$:

$$Pr(\mathbf{w} \mid \mathbf{X}) \propto \prod_{i=1}^{n} Pr(w_i \mid \mathbf{x}_i) \equiv \prod_{i=1}^{n} p(\mathbf{x}_i). \tag{2.8}$$

Propensity score adjustment is sufficient to remove the bias due to observed covariates (Rosenbaum and Rubin 1983). The propensity scores are however rarely known outside controlled randomized trials, so they must be estimated from the data. In this document the propensity scores, estimated by logistic regression, will be transformed to linear propensity scores. These will be weighted averages of the original covariates and, as such, will give rise to more advantageous distributions for creating different diagnostics[3]:

$$q(\mathbf{x}_i) \equiv log[(1 - p(\mathbf{x}_i))/p(\mathbf{x}_i)]. \tag{2.9}$$

The expected treatment effect will be estimated for treated subjects using their matched controls, i.e., the controls who do not differ systematically in their propensity scores from the treated students.

Propensity score matching makes no distinction between covariates that are strongly or weakly predictive of the outcome. Matching methods that make this distinction can be good supplements to propensity score matching, because we can improve the precision of the treatment effect by emphasizing the balance of covariates thought to be strongly predictive of outcomes. For example, in Mahalanobis-metric matching within calipers defined by the propensity scores, a treated student is first matched with all controls whose propensity scores are within a caliper distance from the propensity score of the treated student (Rosenbaum and Rubin 1985). The treated student is then matched

[3] The propensity score distribution for treated students and controls, for example, is more likely to be symmetric and to have more similar variances on the linear scale.

with the control who have the smallest Mahalanobis distance between the treated student's values of the covariates thought to be strongly predictive of outcome and the corresponding values of the control's covariates (Cochran and Rubin, 1973 and Rubin, 1976a, 1976b). The differences between the s covariates thought to be strongly predictive of outcome for a treated student i, $\mathbf{x}_i^s = [x_{i1}, ..., x_{is}]$, and for a control student j, $\mathbf{x}_j^s = [x_{j1}, ..., x_{js}]$, weighted by the inverse variance-covariance matrix of these covariate values for the control students, $\mathbf{C}_{cntrl}^s$, correspond to the Mahalanobis distance between i and j:

$$d_{ij}^s(\mathbf{x}_i^s, \mathbf{x}_j^s) = (\mathbf{x}_i^s - \mathbf{x}_j^s)'(\mathbf{C}_{cntrl}^s)^{-1}(\mathbf{x}_i^s - \mathbf{x}_j^s). \tag{2.10}$$

In order to find the best match for the treated student i, the above distances also need to be calculated for the rest of the control students. The treated student can then be matched with the closest control student within propensity score calipers in terms of this Mahalanobis distance.

Causal inference in observational studies does not differ fundamentally from causal inference in randomized studies as long as the assignment mechanism is ignorable. In order to make the ignorability assumption more plausible, the dimension of $\mathbf{X}$ may need to be high; hence the great utility of propensity score matching.

2.4. Causal Inference with item nonresponse. For some subjects, the values of the covariates $\mathbf{X}$ may not be completely known. Subjects whose covariates are characterized by such *item nonrespons* are sometimes dropped from the analyses with the consequence that the assignment mechanism may no longer be ignorable. Therefore, these subjects should be kept in the analysis by imputing their missing values.

Analyses based on imputed values must account for the additional uncertainty associated with the use of imputed values instead of the true values they replace. If each missing value is replaced by several imputed ones, the resulting multiply-completed datasets will generate multiple estimates of the treatment effect. Combining these estimates creates one estimate for the treatment effect that incorporates such additional levels of uncertainty. This idea was originally proposed (1977, 1978) and further developed (1987) by Rubin and subsequently by others, e.g., Schafer (1997). The combined causal estimate of τ is determined by averaging the estimates, $\hat{\tau}_l$, in each of the $l = 1, .., M$ datasets completed by imputation:

$$\tau = \sum_{l=1}^{M} \hat{\tau}_l / M. \tag{2.11}$$

Inferences about τ can drawn using associated sampling variance. The variance is the sum of the within variance, U_W, that measures how uncertain the estimates are in each dataset, U_l, on average:

$$U_W = \sum_{l=1}^{M} U_l/M, \tag{2.12}$$

and the between variance, U_B, that measures how the estimates vary between the datasets:

$$U_B = \sum_{1=1}^{M} (\hat{\tau}_l - \tau)'(\hat{\tau}_l - \tau)/(M-1). \tag{2.13}$$

The between variance in this sum needs, however, to be adjusted by a factor $1 + 1/M$. Any hypothesis can then be tested with statistics such as the t- and F-statistics with the exception that the degree of freedom needs to be adjusted. These rules are presented in Rubin (1987); improved degrees of freedom with small samples are developed in Barnard and Rubin (1999).

In this document, we will describe how we imputed the values for the item nonresponses by drawing them from the posterior predictive distribution of missing values given the observed values using IVE-ware, a particular piece of software developed by Raghunathan et al. (www.isr.umich.edu/src/smp/ive).

3. The template experiment

In this section, we describe the actual assignment mechanism in the quasi-randomized experiment. We also describe an "ideal" and controlled assignment mechanism in an experiment that is close to randomized. A comparison of these assignment mechanisms helps us understand the extent and the type of correction we need in the actual quasi-randomized experiment to address its non-randomness. This ideal randomized experiment is referred to in this document as the "template experiment".

The actual quasi-experiment that we use to study peer effects on cigarette smoking was conducted at Harvard College in 2003/2004. On their first day of class, freshmen were asked to participate in a prospective survey. At that time, they had been enrolled in college for one week. The 589 freshmen who decided to participate were asked to describe their health habits during the month prior to their enrollment. They were

also asked to replicate the answers they gave to the questions on the Housing Application Form (HAF) before their enrollment,[4] answers that the Harvard Housing Office (HHO) considered when assigning the incoming freshmen into suites. The interview could be completed on-line or in person. Five weeks after the first day of class, the participants were asked to describe their roommates' health habits during the previous month; 462 participants completed this second interview on-line. The last interview was conducted six months after day one of class. At that time, the students were again asked to describe their health habits during the month that had just passed; this third interview was completed on-line by 411 participants. The freshmen had approximately two weeks to complete each interview.

In the template experiment, the HHO groups most of the students into suites, resulting in some suites that consist of at least one smoker and some suites that consist of non-smokers only. The particular constellation of smokers and non-smokers in a particular suite is known because, in the template experiment, the HAF includes a question about the students' smoking habits. The researcher then assigns the remaining students, representing a sample of non-smoking students, into these HHO-grouped suites. The researcher cannot assign more than one sampled student to any HHO-grouped suite in order for the SUTVA assumption to hold (Section 2.1) but can otherwise assign the sampled students according to any stochastic rule of his preference[5]. The non-smoking students that according to this rule are assigned a HHO-grouped suite consisting of at least one smoker are defined as treated students. Control students are those who are assigned a HHO-grouped suite consisting of only non-smokers. The sampled non-smoking students can thus be said to be randomly assigned to either the treated condition or the control condition conditional on the stochastic rule used by the researcher to assign the students into HHO-grouped suites.

In the actual quasi-randomized experiment, the HAF does not include a question about the students' smoking habits. The researcher must, therefore, collect information about the students' own smoking habits (day one of class) and the smoking habits of their roommates (five weeks after day one of class). Furthermore, the HHO groups all students into suites. Thus, the researcher has no opportunity to assign a sample of non-smoking students into the treated condition (suites with at least one smoker)

[4] The HAF included no question about the students' smoking habits, because the prohibition of smoking in the dormitories was believed to render such information redundant.

[5] Because of the way the treated and the control conditions are defined, the stochastic rule will always be dependent on the number of roommates. Bigger suites will always have higher probability of having at least one smoker than smaller suites.

or the control condition (suites with only non-smokers) according to a stochastic rule chosen by himself. Instead, the researcher is dependent on how the rule used by the HHO to group students into suites implicitly assigns students into the treated and the control condition.

The main difference between this actual quasi-randomized experiment and the template experiment is that the information that is used to define the treated and the control conditions, and the covariate information that influences the assignment process, is not collected until after the assignment. Such information could then possibly be affected by treatment. In that sense, they may be "improper" covariates. Using post-treatment assignment information to define the treated and the control condition and, moreover, to adjust for the assignment mechanism is not ideal. The result could be groups that are more comparable in terms of their outcomes than ideal and, as a consequence, groups that underestimate the treatment effects. Covariates that predate the grouping can have no such effects. They are in that sense "proper".

If we study non-smokers and the effect of sharing a suite with at least one roommate who smokes on their later smoking habits, it may not matter that they report their own and their roommates' smoking habits after assignment; as non-smokers, they can neither be influenced by nor influence their treatment[6]. Furthermore, if we assume that the students remember their answers to the questions on the HAF, it does not matter that we use replicated answers to the questions on the HAF instead of the original answers.

Furthermore, it may not matter that the HHO, and not the researcher, chooses the stochastic rule to use when assigning the sampled students into suites. Because the HHO seems to group the students into suites based on a number of haphazard decisions, and the only information on which they can base these decisions are the students' answers to the HAF, this stochastic rule can be estimated by the researcher.

Thus, we believe that the actual Freshmen Study does not differ substantially from the template randomized experiment, and the existing differences can be addressed using a matched sampling design and relatively innocuous assumptions. The Freshman Study can thus be called a quasi-randomized study.

[6] They could, however, encourage their roommates to quit smoking. This would not influence the way they report their roommates' smoking habits, because they are asked to report these during the month that preceded that interview.

4. Covariate data preparation

In a randomized experiment, the comparison groups are chosen without using outcome data. This means that the study cannot be designed in a way that allows the experimenter to capitalize on random variation to demonstrate certain treatment effects. This conclusion is automatic in a randomized experiment. It is enforced artificially in our experiment by not using outcome data when designing the choice of comparison groups. In fact, outcomes could not be used, because they were immediately separated from the rest of the data.

4.1. Description of covariate data. In our sample, there are 414 students who fulfil our inclusion restrictions, i.e., they were present at the second interview when treatment was defined (462), they reported their prior smoking status (461)[7], and they did not smoke (414). From this pool of non-smoking students, we had to drop one control (non-smoking roommates), because his room number was obviously wrong[8]. We also excluded another eleven controls, because these expressed a preference for having no roommates; these controls are unsuitable matches to any treated student, because no treated student expressed a similar preference. As controls, the exclusion of these twelve students cannot bias the treatment effect.

The remaining 402 freshmen comprise our study population. Treatment (at least one smoking roommate) can be confirmed for 56 freshmen and control (no smoking roommate) can be confirmed for 285 freshmen. Furthermore, treatment is undefined for 61 students, i.e., students who did not report living with a smoker but who did report ignorance with regard to at least one of their roommate's smoking habits[9].

It may be possible to regard the 61 students with undefined treatment as controls if we assume that they cannot be affected by smoking habits they are unaware of. Or, it may be better to drop these students, if we assume that they could be either treated students or controls, i.e., they learn their roommates' smoking habits, as a consequence of the second interview. For example, it can be argued that the attention paid to the general health habits of the students' roommates in the second interview could result in students paying increased attention to their roommates' smoking habits. We will

[7] The student who did not report her smoking habits was sharing a suite with roommates who did not smoke. She can, consequently, not be part of the treated students whose effect of treatment we want to estimate. Hence, excluding her cannot bias the treatment effect.

[8] He seems to have gotten his dorm wrong, but his room number right, according to the Harvard Telephone Directory. We were able to identify this irregularity, because he stated a different number of roommates than his claimed roommates did in the study.

[9] Effectively, they could be either in the treated or in the control condition.

compare how well treated students can be matched with controls when these 61 students are assumed to be controls as opposed to when they are dropped. We will then choose to handle these 61 students in the way that helps us achieve the best matches for the treated students. This procedure cannot be described as "mining" the data or "hunting for a result" because we have not yet introduced or even seen outcome data in our analyses.

4.2. Classification of covariate data. We now need to classify the covariates as proper (not affected by treatment) and improper (possibly affected by treatment). This classification in Tables 1 and 2 is important, because the treated students and the controls must be balanced in terms of the proper covariates in order for the estimated treatment effect to be approximately unbiased. As mentioned previously, the covariates derived from the students' replicated answers to the questions on the HAF can be assumed to be proper; although they are collected after the assignment, they can reasonably be assumed not to be affected by treatment, i.e., exposure to smoking versus non-smoking roommates. Furthermore, we also assume that fathers' and mothers' education, respectively, are proper; answers to such questions are also unlikely to be affected by treatment. They can, therefore, be used to proxy for the information that the HHO looks for in an open-ended question on the HAF. The remaining covariates are by definition improper, i.e., they are collected after the assignment and are possibly affected by treatment.

4.3. Imputation of missing covariate data. We need to impute the missing values for the proper covariates, because the proper covariates will be used to match treated students with controls. First, however, the covariates must be put into an appropriate functional form for studying peer effects. The number of actual and preferred roommates are transformed into logs, because peer effects are likely to be a decreasing function of the numbers of actual and preferred roommates. Ten values are then multiply-imputed for each missing value, and these imputed values are transformed into the closest actually observed values; the result is ten "realistic" datasets. As a way to detect imputed outliers, the Mahalanobis distance between each treated student's continuous covariates and the treated students' means of these covariates is derived. For example, for the treated student i, the differences between this student's c continuous covariates, $\mathbf{x}_i^c = [x_{i1}, ..., x_{ic}]$ and the mean of these covariates for all treated students, $\bar{\mathbf{x}}_{trt}^c$ weighted by the inverse variance-covariance matrix of the treated students' continuous covariates; here $\mathbf{C}_{Trt}^c$, represents the Mahalanobis distance:

$$d_i^c(\mathbf{x}_i^c, \bar{\mathbf{x}}_{trt}^c) = (\mathbf{x}_i^c - \bar{\mathbf{x}}_{trt}^c)'(\mathbf{C}_{trt}^c)^{-1}(\mathbf{x}_i^c - \bar{\mathbf{x}}_{trt}^c). \tag{4.1}$$

We find a Mahalanobis distance that significantly exceeded the mean at a 2.5% level for only between two and four of the 46 freshman in each dataset. Therefore, we find no reason for concern. The statistics obtained by combining the estimates and their standard errors across these ten datasets should capture the degree of uncertainty due to the imputation of the missing values.

5. Replicating a randomized study

In this section, we evaluate how well the treated students can be matched with controls. Our objective is to replicate the balance that would have been achieved in a randomized experiment. This is theoretically possible for the observed covariates, but not necessarily for the unobserved ones. In our quasi-randomized experiment, however, we believe that the unobserved covariates are few: we have replicated all information available to the HHO at the time of assignment except for one open-ended question, for which we use parental educations as proxy covariates.

The analyses presented in this section are carried out for a group of 1,000 samples when students with undefined treatment are considered to be controls and for another group of 1,000 samples when these students are dropped. However, we only discuss the samples that we conclude to be best balanced in each group on the basis of these analyses, i.e., sample number 26 and number 632, respectively. The details about these samples are found at the end of this document, as are the details about two additional samples. Sample number 680 and 55 are added to describe the result of the analyses when we randomly draw one sample from each group rather than select the best from each group.

5.1. Achieving SUTVA. Smoking outcomes for all the sampled freshmen in our study cannot be assumed to be independent, because some sampled freshmen share a suite and, consequently, may have influenced one another's outcomes. SUTVA would then be violated. Smoking outcomes for freshmen who do not share suites, however, would satisfy SUTVA given our hypothesis is true, i.e., that peer effects exist only between roommates and not between other freshmen[10]. Therefore, our sample includes only one sampled freshman from each of the suites represented by our study population.

Billions of samples with only one freshmen per suite could be generated from the study population. We want to focus on those samples that maximize the number of

[10] This assumption is indeed strong but hard to avoid unless we study freshmen from many colleges and only include one freshman from each of those in our analysis.

treated students and, thereby, maximize the sample size. Because billions of such samples can also be generated, we identify thousands of them by always sampling treated students for suites that include at least one treated student and by sampling control students for remaining suites; the sampling is random if there is more than one student that can be sampled. More specifically, we choose 1,000 samples assuming the control condition for the 61 students with undefined treatment; those samples include 46 treated students and 235 controls. Furthermore, we choose another 1,000 samples dropping the 61 students with undefined treatment status; those samples also include 46 treated, but fewer controls, i.e., 204 controls[11].

5.2. Estimating propensity scores. We first need to decide which covariates to include in the propensity score estimation, because the matching methods we use are based on estimated propensity scores. In larger samples, we would be less restrictive when choosing the covariates to include. We would include even weakly predictive covariates because, in larger samples, the biasing effects of excluding them has been shown to dominate the efficiency loss of including them (Rubin and Thomas 1996). But in smaller samples like ours we need to be more restrictive.

There are at least 164 covariates that could be included in the propensity score estimation, including the 16 proper covariates themselves, their squares and their interactions. The covariates that are included, however, are determined iteratively; separate iterations are run for each of the samples' ten multiply-completed datasets. In the first iteration, we include only the number of actual roommates in a linear propensity score estimation[12]. In subsequent iterations, we step-in the covariate that contributes the most to the R^2-value in the linear regression as long as its level of significance is 10 percent or lower, and we step-out the covariates that are no longer significant on this level. The number of actual roommates is never stepped-out because of its importance in determining the probability of treatment, i.e., having at least one smoking roommate. A similar step-in/step-out procedure to estimate propensity scores has been used previously in the literature for larger samples (Bingenheimer et al. 2005). Then, we re-estimate the propensity scores using logistic regression for all the samples' ten multiply-completed datasets based on the covariates chosen in the above step-in/step-out procedure.

[11] We lose fewer than 61 students, because some of these share rooms with one another, with other control students, or with at least one treated student. In the former two cases, only one control student will be randomly included in the sample; in the latter case, one treated student will be included.

[12] For computing efficiency reasons, the propensity scores are estimated with a linear model when deciding which covariates to include. The propensity scores used for matching are however estimated with a non-linear model.

This stepwise procedure results in the same covariates in each multiply-completed dataset, even though the procedure is conducted separately for each of them. In sample number 26, two covariates are included in the propensity score estimation, "actual number of roommates" and "actual number of roommates" interacted with "preference for lively room atmosphere". The "actual number of roommates" and the "actual number of roommates" interacted with "days before the interview" are the only covariates included in sample number 632.

5.3. Evaluation of the potential for bias reduction. We want to assure that the treated students can be matched with the control students so that their respective proper covariates have the same multivariate distributions. The means of these distributions are represented by the propensity score means for the treated and the control students respectively and, consequently, the standardized difference in these means represents the bias, B[13]. In sample number 26, the biases vary between 0.88 and 0.96 (Table 7); for sample 632, they vary between 1.01 and 1.04 (Table 26). If the treated and the control students' propensity score variances are assumed to be similar, biases of these magnitudes imply that the means will always differ at the 5% level in different samples generated from the underlying population. These biases are, therefore, not acceptable. The biases must be less than about 0.30 in order for the samples to be perceived as well-balanced in terms of their means.

Therefore, it is necessary to reduce these biases by matching. We can assess the feasibility of close matching by studying the sample moments in the unmatched data. If the distributions of the proper covariates for the treated students and the control students, respectively, are multivariate normal and have proportional covariances, the maximum achievable fractional percentage bias reduction is $min(1, \theta^*_{max})$ (Rubin 1976b, 2000):

$$\theta^*_{max} = \Omega(r/r_m)(uV + b^2)^{-1/2}, \tag{5.1}$$

where: (1) $\Omega(r/r_m)$ is the upper tail of a standard normal covariate:

$$\Omega(r/r_m) \approx 2^{\pi/2}(2\pi)^{-1/2}(r/r_m)^{(1-\pi/4)}(1 - r_m/r)^{(\pi/4)}, \tag{5.2}$$

which is a function of the number of controls per treated student, $r = N_c/N_t$ and the number of matched controls per treated student, $r_m = N_{mc}/N_t$; (2) u is the number of proper covariates used when matching treated students with control students; (3) V is

[13] $B = (\bar{q}(\mathbf{X}_{trt}) - \bar{q}(\mathbf{X}_{cntrl}))/\sqrt{\frac{var(q(\mathbf{X}_{trt}))+var(q(\mathbf{X}_{cntrl}))}{2}}$ where $\mathbf{X}_{trt} = [\mathbf{x}_1, ..., \mathbf{x}_{n_1}]'$ for the n_1 treated students, and $\mathbf{X}_{cntrl} = [\mathbf{x}_1, ..., \mathbf{x}_{n_0}]'$ for the n_0 control students.

the variance:

$$V = (\sigma^2 + r^{-1})N_t^{-1}, \tag{5.3}$$

which is a function of the ratio of the treated to control variance of the best linear discriminant, σ^2; and (4) b^2:

$$b^2 = (\bar{\mathbf{x}}^u_{trt} - \bar{\mathbf{x}}^u_{cntrl})'(\mathbf{C}^u_{cntrl})^{-1}(\bar{\mathbf{x}}^u_{trt} - \bar{\mathbf{x}}^u_{cntrl}), \tag{5.4}$$

which is the square of the distance between the means of the u proper covariates used in the matching for the treated students, $\bar{\mathbf{x}}^u_{trt}$, and the control students, $\bar{\mathbf{x}}^u_{cntrl}$ weighted by the variance-covariance matrix of these covariates for the control students. Again, we report the $\theta^\star_{max}$ for samples number 26 and 632. The $\theta^\star_{max}$ in the ten completed datasets for these samples vary from 1.48 to 1.65 (Table 4) and from 1.37 to 1.43 (Table 23), respectively. Therefore, we can expect that matching can achieve the same means for the treated and the control students' matched distributions.

For the ten datasets, the multivariate distributions for the treated and the control students not only require similar means but also similar variances. The variances along the direction of the propensity scores should be close, as well as the variances orthogonal to the propensity scores as in Rubin (2001). For that reason, we study the ratios of the treated students' and the control students' propensity score variances in these directions. We find that ratios of variances for the ten multiply-completed datasets for samples number 26 and 632 are between 0.63 and 0.79 and between 0.58 and 0.63, respectively, in the direction of the propensity scores. The ratios should be between 4/5 and 5/4 in order for the samples to be perceived as well-balanced in terms of their variances. These differences in variances should, therefore, be decreased. Because the variance ratios are less than one in the direction of the propensity scores, the chances are good that the differences in variances can be decreased or even eliminated through matching; the variances for the treated students are less than the variance for the control students, which makes it easier to approximate the treated students' distribution by a matched sample of the control students. Orthogonal to the propensity scores, the ratios for the proper covariates in the ten multiply-completed datasets are between 4/5 and 5/4 in 0.60% to 0.70% and in 0.66% and 0.72% of the cases for these respective samples. These differences in variances therefore can be considered to be modest.

The number of roommates is likely to be one of the most important covariates in the assignment, because the more roommates students have, the more likely it is that they are treated. For that reason, we want to assure that variance differences in this number

and in some of its interactions are small. We find variance-ratios that exceed one for some of these covariates, but the biases for these covariates are mostly low. Thus, these variance differences may not be important. Furthermore, we study variance differences for various linear combinations of proper covariates, including the number of roommates. Again, we find that the variance ratios for some linear combinations exceed one, but the directions of these ratios are rarely in the number of roommates. Then, the variance differences in these linear combinations may not be important either. The variance ratios for different linear combinations are represented by the eigenvalues of the matrix resulting from having pre- and post-multiplied the variance-covariance matrix of the treated students' proper covariates, C_{trt}, with the Cholesky decomposition of the inverse variance-covariance matrix of the controls students' proper covariates, C_{cntrl}. The directions of these linear combinations are indicated by the associated eigenvectors. The eigenvectors corresponding to the highest eigenvalue in the first multiply-completed dataset of samples number 26 and number 632 are presented in Tables 3 and 22, respectively.

5.4. Matching treated students with controls. Among these 2,000 samples, we first decide which samples can achieve acceptable biases and relative variances in the ten multiply-completed datasets through propensity score matching. For that reason, each treated student in these samples is matched with one control student[14] so that the difference between their respective estimated propensity scores is as small as possible. The bias reductions we achieve in each sample by means of this matching can be assessed by comparing the biases before and after matching using the propensity scores estimated for the unmatched sample[15]. For the biases in the matched samples, however, we must use the propensity scores estimated for the matched sample. We find that the bias reductions created by matching are substantial and that the biases and the ratios of variances in the matched samples are acceptable for two of the 1,000 samples when students with undefined treatment are classified as controls, and for ten of the 1,000 samples when these students are dropped.

[14] Matching the treated students not only with their best controls but also with, for example, their next best control can indeed improve the efficiency of the treatment effect. At the same time, the bias with which the treatment effect is estimated can also increase, because the next best control is never as good a match as the best control. We are more concerned about bias than about efficiency in our small sample, because a biased but precise estimate can give rise to misleading policy implications. For that reason, we match on a one-to-one basis.

[15] The mean difference of the propensity scores for the treated and the matched control students is standardized with the square root of the average variance of the propensity scores for the treated and the unmatched control students.

Any of these 2 + 10 samples can be used to estimate the treatment effects on students' 6-month smoking habits. However, the precision of the estimates may differ. The best precision for our purpose of estimating peer effects on students' smoking outcomes can be achieved if not only the covariates believed to be predictive of treatment are emphasized in our matching, but also covariates predictive of smoking outcomes. For that reason, we study the biases and the relative variances that we can achieve for these 2+10 samples if each treated student is re-matched with one control so that their propensity scores are within one caliper width of each other and so that the Mahalanobis distance for the covariates believed to be predictive of the outcome variables – in our case, the propensity score and the number of roommates – is as small as possible. The caliper width is chosen to be 0.20 of the squared root of the average variance for the treated and control students' propensity scores in the unmatched sample.

In the next section, we show that, among these 2 + 10 samples, the best samples for our purposes of achieving precise estimates of treatment effects are sample number 26 when students with undefined treatment status are kept as controls and sample number 632 when these students are dropped. The bias reductions in the ten multiply-completed datasets for these samples are presented in Tables 8 and 27. These samples' biases and variance ratios in the matched samples are presented in Tables 9 and 28.

5.5. Choosing the best sample. Mahalanobis metric matching within propensity score calipers results in two samples that are best with regard to the biases and the variance ratios in each of their respective ten multiply-completed datasets: sample 26 of the 1,000 we originally drew when the students with undefined treatment status are confined to the control group (Table 10), and sample 632 of the 1,000 we originally drew when these students were dropped (Table 29). In order to learn how best to treat the students with undefined treatment status, we want to compare biases and relative variances for samples number 26 and 632 for the 20 proper covariates, their cross-products, and their interactions, that is in total 164 covariates. The 164 covariates are compared for the treated students and the controls in all the ten multiply-completed datasets for each sample, so in total there are 1,640 comparisons for each sample.

For sample 26, we find that 45 of these comparisons are significant at the 5% level and 70 covariates at the 10% level (Tables 11 and 12). For sample 632, 4 and 14 covariates are significant at the 5% and the 10% levels, respectively (Table 30). Hence, both of these samples are better balanced with respect to observed covariates than could be expected in a randomized experiment. However, the number of significant covariates seems to be fewer for sample 632. For that reason, we decide to use sample

632 for the final outcome analyses in our next document. The students with undefined treatment status will consequently be dropped.

At the end of this document, we show the propensity score distribution for one multiply-completed dataset for samples number 26 and 632, respectively, before matching, after matching but before re-estimation of the propensity scores, and after matching and re-estimation of the propensity scores for both matching methods, i.e., propensity score matching (Graphs 1 and 5) and Mahalanobis-metric matching within propensity score calipers (Graphs 2 and 6). We only show the distributions of one of the multiply-completed datasets, because the distributions for the remaining datasets are similar.

6. "Practice" outcome analyses

Without having used, or even seen, outcome data, we have decided which assumption to make with regard to the students with undefined treatment status, which sample to use for our final analyses among 2,000 evaluated samples, and which control students to match with the treated students in this sample. We made these decisions so as to achieve the best balance for our purpose of estimating the causal effect on students' six-month smoking habits of having at least one roommate who smokes[16]. We have not yet decided, however, the analyses that we are committing to conduct when estimating these effects. Neither have we discussed the estimators to use. These decisions should also be made before we use and see outcome data. Otherwise, we could always be accused of choosing models and estimators that give us publishable results or results that are favorable to a priori viewpoint. In this section, we make these decisions, thereby setting the bounds for the primary and secondary outcome analyses in the next paper.

6.1. Models for the final analyses. In order to understand how sensitive our matched sample is to different models, that is to different covariate adjustments and estimators, we practice outcome analyses on intermediate outcomes, that is on the improper covariates described in Table 2. Students with missing outcomes are simply dropped in these analyses, because we do not yet have access to any outcomes and can, therefore, not impute them effectively. We estimate the treatment effects on these intermediate outcomes without adjustments for any covariates and after adjustments for those covariates that, according to various stepwise procedures, have been shown to be predictive of these outcomes. We will use estimators that only implicitly define

[16] These were documented in an email sent to Magnus Johannesson at the Stockholm School of Economics July 8, 2005.

the missing potential outcomes, i.e., ordinary-least-square regression and ordinary-least-square regression adding indicators for each matched pairs and an estimator that explicitly derives these outcomes. The assumptions associated with these estimators differ somewhat. The variances for the treated and the control students are assumed to be equal when the first estimator is used and their response surfaces are assumed to be parallel. The variances are still assumed to be equal for these students when the second estimator is used, but the response surfaces need no longer be parallel. For the third estimator, neither the variances need to be similar, nor the response surfaces.

When no covariate adjustment is made, one of the 23 improper covariates is significant at the 10% level, i.e., the number of drinks the students recalled they drank on days when they consumed alcohol during the month before their enrollment. The magnitude of this estimated effect does not seem to vary much with the different estimators, nor does its precision (Tables 41, 42, and 43). For this particular intermediate outcome, we now study various covariate adjustments. The covariate adjustments to make in each of the ten datasets can be chosen by means of different stepwise procedures. One stepwise procedure that we try adds one covariate at a time and chooses the covariate that is the most predictive of outcome given that it is significant at a fixed pre-defined level α. Previously stepped-in covariates that are no longer significant at this level are stepped-out. We use 5% as the fixed pre-defined significance level and evaluate all the proper covariates, the propensity score, and the derivatives of these covariates, i.e., their cross-products and interactions (209 covariates in total) (Table 5). The disadvantage of this procedure is that the significance level for including one covariate differs from α as long as there are more than one covariate evaluated for inclusion; the level increases with the number of evaluated covariates. Therefore, it is hard to understand on which basis the covariates are included. For that reason, we try another stepwise procedure that does not have this disadvantage. It adjusts the level of significance at which a covariate should be added to α_{bf} so that the significance level for including one covariate always is α. This Bonferroni adjusted level of significance (Shaffer 1995) is derived as follows:

$$\alpha_{bf} = 1 - (1 - \alpha)^{\frac{1}{q-p}} \approx \frac{\alpha}{(q-p)}, \tag{6.1}$$

where q is the number of covariates available for inclusion and p is the number of covariates already included. We try this stepwise procedure when all 209 covariates are evaluated for inclusion (Table 45). We also restrict the number of evaluated covariates to the six covariates believed to be the strongest predictors of outcome (Table 4). The α, however, is always set at 20%. The results of these stepwise regressions do not

indicate that covariate adjustments have a clear advantage.

Because there seems to be no clear benefit of covariate adjustments and because stepwise regressions are always arbitrary to some extent, our primary outcome analysis using real outcomes will not include any covariate adjustments. Furthermore, it will be conducted for all three estimators, because the assumptions of equal variances and parallel response surfaces do not seem to be important. The magnitude and the precision of the treatment effect seem to be robust against the different assumptions of these estimators. It is the results from this primary analysis that we will rely on. The results from secondary analyses that consider covariate adjustments can, however, be good complements to the results from the primary analysis. If the magnitudes of the treatment effects are the same in the primary and secondary analyses but their precisions improve following covariate adjustments, we will learn that the results are sensitive to covariate adjustments. Furthermore, if the particular covariates that are included as adjustments in the secondary analyses are believed to be strongly predictive of smoking, we will be able to conclude that the improved precision following covariate adjustments is real.

The secondary analyses will use two kinds of covariates adjustments. We will adjust for covariates significant at $\frac{1}{5(q-p)}$ level in the above described stepwise procedure, because this adjustment involves less arbitrariness in choosing the covariates than would be involved in an adjustment for covariates significant at a fixed pre-defined significance level. The covariates that we evaluate for inclusion, q, will be preference for "lively room atmosphere", "physically active", "active in religious groups", "propensity score", and "propensity score squared", and "parent's average level of education". However, we will also adjust for covariates significant at a fixed pre-defined level, because adjustments of this kind are often made in praxis. These secondary analyses will also be conducted for the three different estimators.

6.2. New research ideas and benchmarks for treatment effects. Our outcome analyses indicate a close to significant treatment effect for one of the 23 improper covariates. This result can possibly be attributed to the random variation in our dataset, because about one significant improper covariates out of sixteen can be expected at the 5% level. It can, however, be suggestive of a true treatment effect. For example, the students' perceptions of their pre-college drinking intensity and frequency might be influenced by treatment assignment. The smokers with whom the treated students share a suite may drink to a greater extent than the non-smokers with whom the control students share a suite. If students want to be like their peers, the

treated students may then report a higher pre-college alcohol consumption than the controls. Furthermore, if students are more prone to honestly report pre-college alcohol consumption in a drug-tolerant environment, the treated students may also report a higher alcohol consumption than the controls. Another study, however, has to confirm this effect and preferably also explain it.

In such a study, the students would first be interviewed before their enrollment. They would be asked to describe the intensity and the frequency of their cigarette and alcohol consumptions at that time. Students' actual consumptions would also be measured: smokers would be detected by measuring carbon monoxide in their breath. Drinkers would be detected, at least to some extent, by measuring alcohol in their blood and breath. After their enrollment, the students would be interviewed again. They would be asked to report the intensity and the frequency of their cigarette and alcohol consumption at the time of the previous interview[17], as well as their roommates' current smoking and drinking habits. This experiment would help us to understand if freshmen report differently depending on their roommates' health habits. In addition, it would help us understand the reasons for any such reporting differences. To practice outcome analyses can, in other words, give rise to new research ideas. It can also provide necessary benchmarks for evaluating the importance of the estimated treatment effects on final outcomes. In our case, for example, peer effects on smoking should be greater in magnitude than peer effects on recalled alcohol consumption prior to college[18]. Otherwise, the estimated peer effects could be explained by the way in which a behavior is reported rather than by the behavior itself. Although these estimates indicate interesting effects, they are not the peer effects of interest in this study.

7. Conclusion

Above, we document and illustrate an approach that can be used to replicate the benefits of a randomized experiment when only a quasi-experimental study is available.

First, we define the data generating process for the quasi-experimental study and compare it with a close randomized experiment, i.e., the template experiment (Section 3). This comparison clarifies how close the former experiment is to the latter and the extent of adjustments needed in the former because of these differences. For example, it helps us define the population that can be used to study peer effects on cigarette smoking in our quasi-experimental study, i.e., non-smokers who share a suite with at least one

[17] The questions should, however, be framed differently so that they would not be perceived as identical by the students.

[18] The magnitudes are measured in terms of their respective p-values, because the dataset is fixed.

roommate who smokes. It also defines the covariates whose multivariate distribution should be approximately the same for the treated students and the controls in order for the treatment effect estimate to be approximately unbiased.

We then maximize the sample size in our study by letting treated students represent suites with at least one student who smokes and by randomly drawing controls to represent the other suites (Section 5.1). There are 46 treated students and 204 or 235 controls depending on whether students with undefined treatment status are kept as controls or dropped. We draw 1,000 samples in which students with undefined treatment are kept as controls, and another 1,000 samples in which they are dropped.

Subsequently, we impute the missing values of the proper covariates for our sample (Section 4.3). Ten values are imputed for each missing value, resulting in ten multiply-completed datasets. The treatment effect and standard error found by combining the estimated treatment effects and their standard errors in these completed datasets will capture uncertainty related to the use of imputed values instead of true values.

Finally, we match treated students with controls based on the proper covariates in all samples, i.e., for the 1,000 samples when students with undefined treatments are kept as controls and for the 1,000 samples when students with undefined treatments are dropped (Section 5.4). For each set of samples, propensity score matching identifies well-balanced samples, and Mahalanobis-metric matching within calipers defined by the propensity scores picks out the sample with the best balance for the purpose of estimating peer effects relating to smoking. The resulting two samples are compared, and the sample that gives us the best overall balance is chosen for the final outcome analyses (Section 5.5). Thus, we let the covariate data help us decide which assumption to make about the students with undefined treatment status and, hence, which sample to use. We have this opportunity because we set aside the outcome data when designing our study. We choose to drop the students with undefined treatment and to use sample number 632 for our final outcome analyses.

In sample number 632, the treated students and their matched controls are more similar with respect to observed covariates than would be expected from a completely randomized experiment. The standardized differences in their propensity scores after Mahalanobis-metric matching within calipers defined by the propensity scores, i.e., the biases, are less than 0.15 in all ten datasets completed by multiple imputation, and the ratios of the propensity score variances for the treated students and the controls

are always acceptable ranging between 0.8 and 1.25. The variance ratios orthogonal to the propensity score are also acceptable. Furthermore, few proper covariates are significant in the ten multiply-completed datasets; among a total of 1,640 covariates (164 proper covariates in each imputed datasets and 10 multiply-completed datasets), only 4 are significant at the 5% level, and only 14 are significant at the 10% level.

Finally, we estimate the treatment effect on improper covariates, meaning covariates that can be regarded as short term outcomes. This exercise in outcome analyses helps us define the models that we commit to use for the final outcome analyses in the next paper (Section 6.1). As our primary analysis, we decide not to make any covariate adjustments, because such adjustments do not seem to be associated with any clear benefits. Furthermore, we also determine the estimators that we will use. We will use estimators that only implicitly define the missing potential outcomes, i.e., ordinary-least-square regression and ordinary-least-square regression adding indicators for each matched pairs and an estimator that explicitly derives these outcomes, because the assumptions that distinguish these estimators do not seem to be important. We also specify the secondary analyses that we plan to run. These analyses should be perceived as complements to the primary analysis.

The practice outcome analyses also give us ideas for further research, and, more importantly, help us evaluate the magnitudes of treatment effects on the real outcomes that could be real and important (Section 6.2). In this document, for example, we find that treated freshmen tend to recall a higher intensity of alcohol consumption during the month prior to their enrollment than their matched controls after an average of one week of classes. Because we did not design our study to analyze peer effects of this kind, this result can only be suggestive. It needs to be replicated in an experiment that can confirm and explain similar effects. Will freshmen recall their prior consumption of substances more honestly if they live in a drug-tolerant environment? Or will freshmen choose to report a consumption levels similar to that of their peers, because they want to be like them? Also, the practice outcome analyses helps us understand the magnitude of the treatment effect we should expect in the final outcome analysis of smoking outcomes. Smoking peer effects should be at least as large as the alcohol peer effects found in the practice outcome analyses. Otherwise, the estimated effect can be explained by the way students report their cigarette consumption rather than the way students change their behavior.

In conclusion, we believe that the field of causal inference would benefit if the study

designs were published before outcome analyses were undertaken. The study designs could, for example, be published in an on-line database. If the same database also publishes the associated outcome analyses, perhaps even the references to published articles, the database would be a valuable source of information. Because the database would not only publish studies with significant results and results aligned with our prior expectations, but also studies with insignificant results and results opposed to our priors, it would contribute to a fair understanding of our economic environment. In addition, the database would help defend the field of research against claims of publishing only studies with significant results.

8. Tables

FIGURE 1. Sample 26 (students with undefined treatment are assumed to be controls): propensity score distributions before and after propensity score matching for multiply-completed dataset I. Similar distributions are found for multiply-completed datasets II-X.

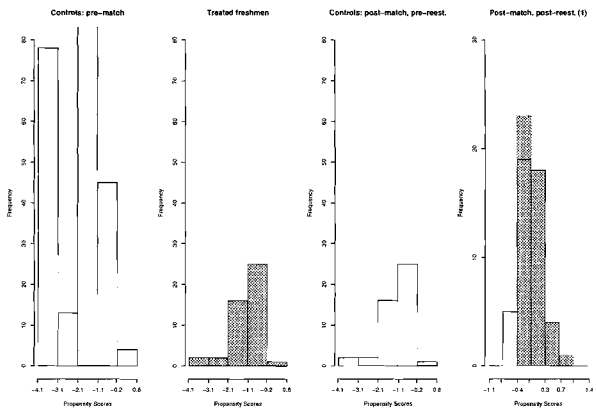

FIGURE 2. Sample 26 (students with undefined treatment are assumed to be controls): propensity score distributions before and after Mahalanobis-metric matching within calipers defined by the propensity scores for multiply-completed dataset I. Similar distributions are found for multiply-completed datasets II-X.

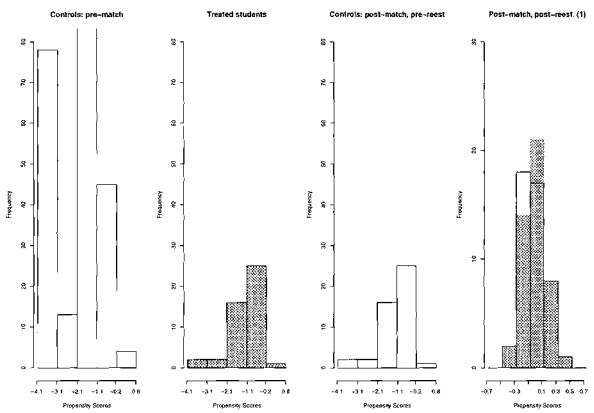

FIGURE 3. Sample 680 (students with undefined treatment are assumed to be controls): propensity score distributions before and after propensity score matching for multiply-completed dataset I. Similar distributions are found multiply-completed datasets II-X.

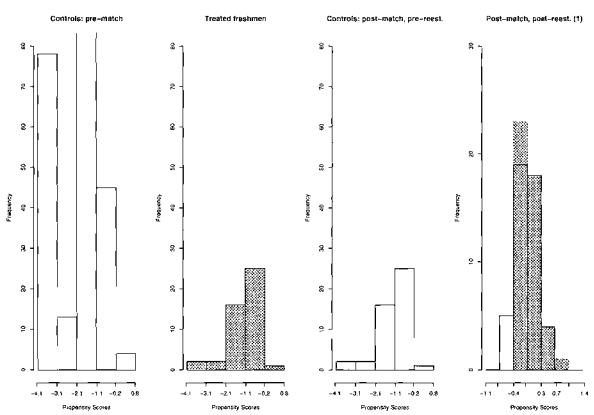

FIGURE 4. Sample 680 (students with undefined treatment are assumed to be controls): propensity score distributions before and after Mahalanobis-metric matching within calipers defined by the propensity scores for multiply-completed dataset I. Similar distributions are found for multiply-completed datasets II-X.

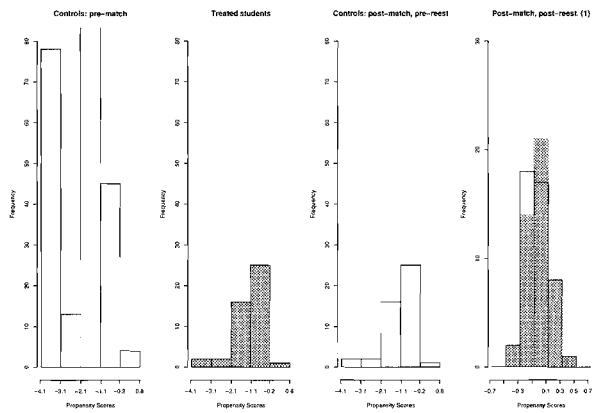

FIGURE 5. Sample 632 (students with undefined treatment are assumed to be controls): propensity score distributions before and after propensity score matching for multiply-completed dataset I. Similar distributions are found for multiply-completed datasets II-X.

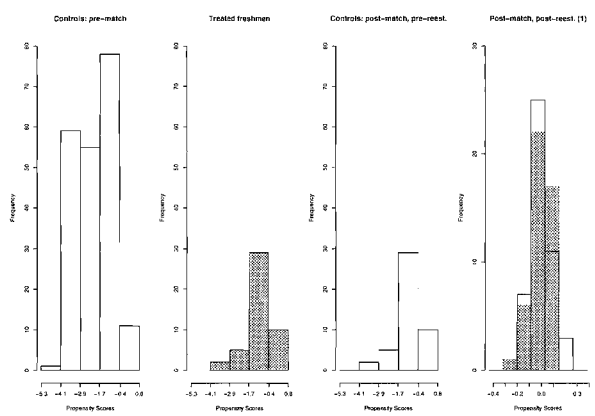

FIGURE 6. Sample 632 (students with undefined treatment are assumed to be controls): propensity score distributions before and after Mahalanobis-metric matching within calipers defined by the propensity scores for multiply-completed dataset I. Similar distributions are found for multiply-completed datasets II-X.

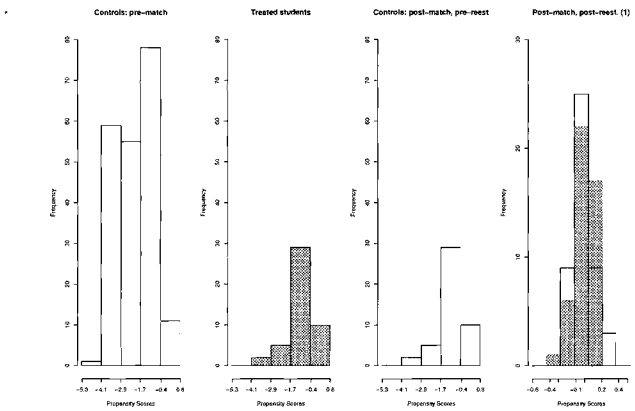

FIGURE 7. Sample 55 (students with undefined treatment are assumed to be controls): propensity score distributions before and after propensity score matching for multiply-completed dataset I. Similar distributions are found for multiply-completed datasets II-X.

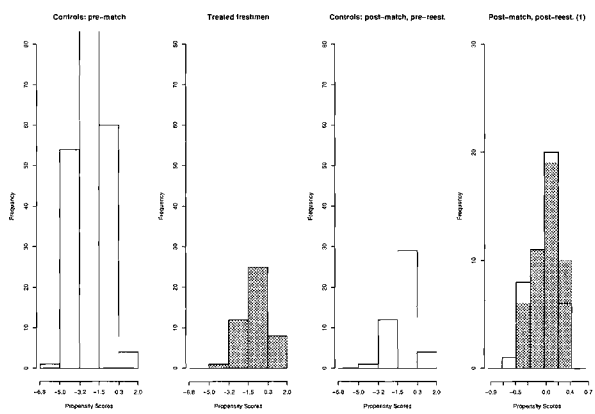

FIGURE 8. Sample 55 (students with undefined treatment are assumed to be controls): propensity score distributions before and after Mahalanobis-metric matching within calipers defined by the propensity scores for multiply-completed dataset I. Similar distributions are found for multiply-completed datasets II-X.

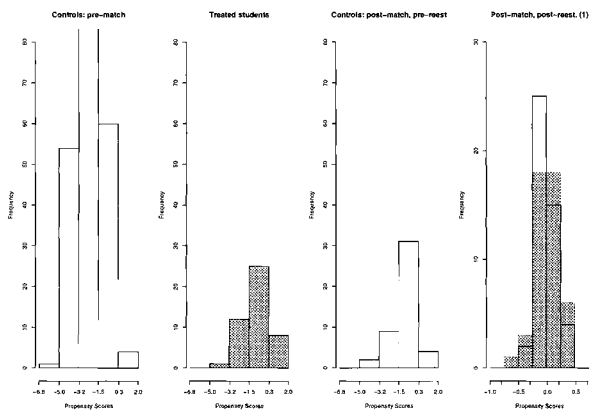

TABLE 1. Description of the proper covariates (third column) and the time for their collection (first column)

Interview 1	Proper	Number of preferred roommates reported (incl. HAF)
Interview 1	Proper	Reported preferences for a room that was lively social center as opposed to quiet/serious (incl. HAF)
Interview 1	Proper	Reported preferences for a room that was disordered as opposed to neat (incl. HAF)
Interview 1	Proper	Reported preference for hours kept during the week: when day begins (incl. HAF)
Interview 1	Proper	Reported preference for hours kept during the week: when day ends (incl. HAF)
Interview 1	Proper	Ranked social science as academic interest (incl. HAF)
Interview 1	Proper	Ranked fine arts as academic interest (incl. HAF)
Interview 1	Proper	Ranked social studies or English as academic interest (incl. HAF)
Interview 1	Proper	Ranked psychology as academic interest (incl. HAF)
Interview 1	Proper	Ranked religious activities as extracurricular interest (incl. HAF)
Interview 1	Proper	Ranked intramural/recreational sports or intercollegiate athletics as extracurricular interests (incl. HAF)
Interview 1	Proper	Ranked alternative, folk music, jazz, rock, gospel as types of music they enjoyed (incl. HAF)
Interview 1	Proxy for proper	Level of education completed by father/mother
Interview 2	Proper	Number of actual roommates
Interview 2	Proper	Race

TABLE 2. Description of the improper covariates (third column) and the time for their collection (first column)

Interview 1	Improper	Carefully completed HAF
Interview 1	Improper	Truthfully completed HAF
Interview 1	Improper	Religious tolerance if questioned about it in the HAF
Interview 1	Improper	Received help completing the HAF
Interview 1	Improper	Smoked a pipe or cigars 30 days prior to enrollment
Interview 1	Improper	Used chewing tobacco or snuff 30 days prior to enrollment
Interview 1	Improper	Took drugs 30 days prior to enrollment
Interview 1	Improper	Number of times consumed alcohol 30 days prior to enrollment
Interview 1	Improper	Number of drinks when consumed alcohol 30 days prior to enrollment
Interview 1	Improper	Number of times consumed ≥ 5 drinks 30 days prior to enrollment
Interview 1	Improper	Number of times devoted to moderate physical activities during 30 days prior to enrollment
Interview 1	Improper	Number of times devoted to vigorous physical activities during 30 days prior to enrollment
Interview 1	Improper	Father/mother smoked cigarettes, a pipe, or cigars or used chewing tobacco, or snuff during childhood
Interview 1	Improper	Father's alcohol consumption during childhood
Interview 1	Improper	Mother's alcohol consumption during childhood
Interview 2	Improper	Ever used alcohol before college
Interview 2	Improper	Ever smoked a cigarette before college
Interview 2	Improper	Ever smoked a cigar before college
Interview 2	Improper	Ever smoked a pipe before college
Interview 2	Improper	Ever chewed tobacco before college
Interview 2	Improper	Ever used illegal drugs or misused prescription drugs before college

TABLE 3. Sample 26 (students with undefined treatment are assumed to be controls): the eigenvector associated with the highest eigenvalue of 3.47 in multiply-completed dataset I.

Mates	0.16
Father edu	−0.00
Mother edu	−0.12
Pref. mates	0.03
Lively	0.20
Disorderly	−0.40
Wake up late	0.01
To bed late	−0.42
Interview	−0.25
Male	−0.37
Black	−0.13
West of US	0.26
Social science	0.06
Fine arts	−0.26
Social studies	−0.17
Psychology	0.08
Religious activities	0.17
Sports activities	0.28
Indep. music	−0.28
Indep. music etc	0.12

TABLE 4. Sample 26 (students with undefined treatment are assumed to be controls): θ^*_{max}.

	θ^*_{max}
Dataset I	1.62
Dataset II	1.59
Dataset III	1.54
Dataset IV	1.61
Dataset V	1.56
Dataset VI	1.64
Dataset VII	1.48
Dataset VII	1.62
Dataset IX	1.65
Dataset X	1.60

TABLE 5. Sample 26 (students with undefined treatment are assumed to be controls): Imbalances, related to actual number of roommates, before matching for multiply-completed dataset I; similar imbalances are found for multiply-completed datasets II-X.

	X_{trt}	σ^2_{trt}	X_{cntrl}	σ^2_{cntrl}	t-test	B	R
Mates	1.09	0.38	0.73	0.56	5.54	77.36	0.46
Mates*Mates	1.33	0.72	0.84	0.76	4.26	67.34	0.89
Mates*Father edu	2.78	1.24	1.94	1.59	3.98	58.67	0.61
Mates*Mother edu	2.69	1.14	1.76	1.50	4.79	70.09	0.58
Mates*Pref mates	1.53	0.64	0.95	0.80	5.39	80.13	0.63
Mates*Lively	3.76	1.50	2.21	1.87	6.11	91.19	0.65
Mates*Disorderly	3.04	1.46	1.94	1.69	4.54	69.41	0.74
Mates*Wake up late	2.35	1.07	1.46	1.20	5.02	77.77	0.79
Mates*To bed late	1.79	0.74	1.18	1.01	4.78	68.76	0.53
Mates*Interview	15.26	6.28	9.66	8.20	5.25	76.82	0.59
Mates*Male	0.51	0.58	0.34	0.51	1.87	31.22	1.26
Mates*Social science	0.63	0.62	0.43	0.56	1.98	33.09	1.25
Mates*Fine arts	0.62	0.61	0.40	0.55	2.26	37.59	1.22
Mates*Social studies	0.56	0.61	0.40	0.55	1.67	27.91	1.23
Mates*Sports activities	0.79	0.62	0.47	0.58	3.32	54.72	1.15
Mates*Indep music etc	0.61	0.61	0.44	0.55	1.68	27.99	1.22

TABLE 6. Sample 26 (students with undefined treatment are assumed to be controls): Imbalances, not related to actual number of roommates, before matching, for multiply-completed dataset I; similar imbalances are found for multiply-completed datasets II-X.

	X_{trt}	σ^2_{trt}	X_{cntrl}	σ^2_{cntrl}	t-test	B	R
Pref mates	1.38	0.27	1.25	0.33	2.87	43.30	0.69
Lively	3.42	0.75	3.04	0.83	3.11	48.24	0.80
Father edu*Lively	8.87	3.13	7.97	2.99	1.79	29.22	1.09
Mother edu*Pref mates	3.41	1.11	3.02	1.24	2.12	32.81	0.79
Mother edu*Lively	8.34	2.70	7.30	2.99	2.33	36.27	0.81
Pref mates*Pref. mates	1.97	0.74	1.66	0.77	2.58	41.00	0.92
Pref mates*Lively	4.77	1.55	3.89	1.67	3.47	54.48	0.85
Pref mates*Disorderly	3.91	1.65	3.35	1.43	2.17	36.59	1.34
Pref mates*Wake up late	2.93	1.03	2.51	0.94	2.61	43.36	1.20
Pref mates*To bed late	2.32	0.85	2.02	0.89	2.18	34.46	0.90
Pref mates*Interview	19.26	6.35	16.30	6.30	2.90	46.84	1.02
Lively*Lively	12.27	4.70	9.95	5.17	3.01	46.90	0.82
Lively*Disorderly	9.75	4.11	8.41	4.14	2.02	32.52	0.98
Lively*Wake up late	7.43	2.90	6.18	2.53	2.74	46.06	1.31
Lively*To bed late	5.83	2.25	5.01	2.41	2.23	35.08	0.87
Lively*Interview	48.05	16.38	40.08	16.60	3.01	48.38	0.97
Disorderly*Interview	39.91	17.31	35.04	15.12	1.78	30.01	1.31
Wake up late*Interview	30.17	12.57	26.30	9.93	1.97	34.21	1.60
To bed late*Interview	23.80	9.79	20.86	9.09	1.89	31.21	1.16
PS	−1.27	0.73	−2.02	0.90	6.11	91.22	0.65

TABLE 7. Sample 26 (students with undefined treatment are assumed to be controls): balances before matching.

	PS		Covariates orthogonal to PS				
	B	R	[0,1/2]	(1/2,4/5]	(4/5,5/4]	(5/4,2]	(2,∞]
Dataset I	0.91	0.65	0.00	0.05	0.66	0.26	0.02
Dataset II	0.92	0.68	0.00	0.09	0.62	0.27	0.01
Dataset III	0.94	0.72	0.00	0.05	0.60	0.32	0.01
Dataset IV	0.92	0.65	0.00	0.05	0.70	0.23	0.01
Dataset V	0.92	0.74	0.00	0.05	0.66	0.27	0.01
Dataset VI	0.91	0.63	0.00	0.06	0.63	0.29	0.01
Dataset VII	0.96	0.79	0.00	0.05	0.66	0.28	0.01
Dataset VIII	0.91	0.65	0.00	0.07	0.66	0.26	0.01
Dataset IX	0.88	0.68	0.00	0.05	0.68	0.26	0.01
Dataset X	0.92	0.68	0.00	0.06	0.68	0.25	0.01

TABLE 8. Sample 26 (students with undefined treatment are assumed to be controls): bias reductions following propensity score matching.

	Pre-match B	Post-match B
Dataset I	0.91	0.00
Dataset II	0.92	0.00
Dataset III	0.94	0.00
Dataset IV	0.92	−0.00
Dataset V	0.92	0.00
Dataset VI	0.91	0.00
Dataset VII	0.96	0.01
Dataset VIII	0.91	0.00
Dataset IX	0.88	0.01
Dataset X	0.92	−0.00

TABLE 9. Sample 26 (students with undefined treatment are assumed to be controls): balances after propensity score matching.

	PS		Covariates orthogonal to PS				
	B	R	[0,1/2]	(1/2,4/5]	(4/5,5/4]	(5/4,2]	(2,∞]
Dataset I	0.25	1.08	0.00	0.09	0.73	0.16	0.02
Dataset II	0.14	0.95	0.00	0.13	0.70	0.15	0.01
Dataset III	0.11	0.97	0.00	0.07	0.74	0.18	0.00
Dataset IV	0.20	1.10	0.00	0.05	0.74	0.19	0.01
Dataset V	0.11	0.99	0.01	0.10	0.74	0.14	0.01
Dataset VI	0.18	1.09	0.01	0.12	0.72	0.13	0.02
Dataset VII	0.13	0.99	0.00	0.09	0.74	0.16	0.01
Dataset VIII	0.25	1.07	0.00	0.09	0.74	0.16	0.01
Dataset IX	0.19	1.09	0.00	0.08	0.76	0.15	0.01
Dataset X	0.25	1.07	0.00	0.05	0.77	0.16	0.01

TABLE 10. Sample 26 (students with undefined treatment are assumed to be controls): balances after Mahalanobis-metric matching.

	PS		Covariates orthogonal to PS				
	B	R	[0,1/2]	(1/2,4/5]	(4/5,5/4]	(5/4,2]	(2,∞]
Dataset I	0.15	0.99	0.01	0.12	0.71	0.15	0.00
Dataset II	0.09	0.96	0.01	0.16	0.72	0.10	0.01
Dataset III	0.07	1.02	0.01	0.12	0.73	0.13	0.00
Dataset IV	0.06	1.02	0.01	0.11	0.69	0.18	0.01
Dataset V	0.07	1.04	0.02	0.13	0.68	0.15	0.02
Dataset VI	0.04	1.04	0.01	0.12	0.66	0.19	0.02
Dataset VII	0.09	1.04	0.01	0.13	0.73	0.12	0.01
Dataset VIII	0.15	0.97	0.01	0.09	0.74	0.15	0.00
Dataset IX	0.13	0.98	0.01	0.10	0.68	0.19	0.01
Dataset X	0.16	0.96	0.00	0.12	0.72	0.16	0.00

TABLE 11. Sample 26 (students with undefined treatment are assumed to be controls): Imbalances after matching for multiply-completed datasets I-V.

	X_{trt}	σ^2_{trt}	X_{cntrl}	σ^2_{cntrl}	t-test	B	R	
Social science	1.00	0.54	0.50	0.76	0.43	−2.22	0.71	1.36
Father edu*Social science	1.00	1.37	1.39	2.04	1.30	−2.40	−50.13	1.14
Mother edu*Social science	1.00	1.35	1.35	1.87	1.20	−1.95	−40.73	1.26
Mother edu*Fine arts	1.00	1.50	1.33	1.00	1.25	1.86	38.79	1.14
Pref mates*Social science	1.00	0.73	0.70	1.02	0.61	−2.11	−44.04	1.33
Lively*Social science	1.00	1.82	1.79	2.61	1.65	−2.21	−46.01	1.18
Disorderly*Social science	1.00	1.52	1.56	2.15	1.35	−2.07	−43.23	1.34
To bed late*Social science	1.00	0.90	0.90	1.38	0.90	−2.54	−52.96	1.01
Interview*Fine arts	1.00	8.15	7.48	5.54	6.86	1.74	36.34	1.19
Social science	2.00	0.54	0.50	0.76	0.43	−2.22	0.71	1.36
Father edu*Social science	2.00	1.37	1.39	2.02	1.29	−2.33	−48.66	1.16
Mother edu*Social science	2.00	1.35	1.35	1.80	1.20	−1.71	−35.64	1.26
Pref mates*Social science	2.00	0.74	0.71	1.03	0.62	−2.12	−44.23	1.31
Lively*Social science	2.00	1.84	1.80	2.59	1.63	−2.10	−43.69	1.23
Disorderly*Social science	2.00	1.53	1.57	2.17	1.37	−2.09	−43.48	1.31
To bed late*Social science	2.00	0.90	0.90	1.35	0.90	−2.37	−49.40	1.01
Wake up late*Interview	3.00	30.17	12.57	25.75	11.19	1.78	37.17	1.26
Social science	3.00	0.54	0.50	0.74	0.44	−1.98	0.74	1.29
Father edu*Social science	3.00	1.39	1.41	2.00	1.33	−2.13	−44.43	1.11
Mother edu*Social science	3.00	1.37	1.37	1.83	1.22	−1.69	−35.21	1.27
Pref mates*Social science	3.00	0.76	0.72	0.99	0.62	−1.67	−34.88	1.32
Lively*Social science	3.00	1.85	1.82	2.51	1.67	−1.82	−37.95	1.18
Disorderly*Social science	3.00	1.54	1.57	2.11	1.39	−1.83	−38.12	1.29
To bed late*Social science	3.00	0.93	0.95	1.33	0.91	−2.01	−41.92	1.08
Wake up late*Wake up late	4.00	4.91	2.65	4.03	2.13	1.76	36.60	1.54
Wake up late*Interview	4.00	30.17	12.57	25.20	10.26	2.08	43.39	1.50
Social science	4.00	0.57	0.50	0.78	0.42	−2.26	0.72	1.44
Father edu*Social science	4.00	1.43	1.39	2.09	1.26	−2.35	−49.07	1.22
Mother edu*Social science	4.00	1.41	1.36	1.91	1.17	−1.89	−39.42	1.35
Pref mates*Social science	4.00	0.76	0.70	1.05	0.59	−2.20	−45.79	1.38
Lively*Social science	4.00	1.86	1.75	2.66	1.59	−2.30	−48.03	1.21
Disorderly*Social science	4.00	1.58	1.55	2.21	1.31	−2.11	−43.94	1.40
Wake up late*Social studies	4.00	1.22	1.19	0.80	1.00	1.80	37.52	1.41
To bed late*Social science	4.00	0.95	0.91	1.41	0.88	−2.50	−52.16	1.06
Wake up late*Interview	5.00	30.17	12.57	25.40	11.10	1.93	40.23	1.28
Social science	5.00	0.57	0.50	0.74	0.44	−1.76	0.76	1.27
Father edu*Social science	5.00	1.43	1.39	2.00	1.33	−1.99	−41.46	1.09
To bed late*Social science	5.00	0.93	0.90	1.32	0.92	−2.01	−41.82	0.98

TABLE 12. Sample 26 (students with undefined treatment are assumed to be controls): Imbalances after matching for multiply-completed datasets VI-X.

	X_{trt}	σ^2_{trt}	X_{cntrl}	σ^2_{cntrl}	t-test	B	R	
Wake up late*Interview	6.00	30.17	12.57	25.85	10.24	1.81	37.73	1.51
Social science	6.00	0.57	0.50	0.76	0.43	−2.01	0.74	1.35
Father edu*Social science	6.00	1.43	1.39	2.04	1.30	−2.17	−45.20	1.15
Mother edu*Social science	6.00	1.41	1.36	1.87	1.19	−1.72	−35.80	1.32
Pref mates*Social science	6.00	0.76	0.70	1.03	0.61	−1.92	−40.07	1.32
Lively*Social science	6.00	1.87	1.77	2.61	1.64	−2.08	−43.31	1.17
Disorderly*Social science	6.00	1.61	1.58	2.15	1.35	−1.77	−36.93	1.38
To bed late*Social science	6.00	0.93	0.90	1.37	0.90	−2.31	−48.10	1.00
Father edu*Social science	7.00	1.43	1.39	1.96	1.37	−1.81	−37.83	1.04
Social science	8.00	0.52	0.51	0.74	0.44	−2.19	0.71	1.29
Father edu*Social science	8.00	1.33	1.40	1.98	1.32	−2.30	−47.86	1.12
Mother edu*Social science	8.00	1.30	1.36	1.83	1.22	−1.94	−40.37	1.26
Pref mates*Social science	8.00	0.72	0.71	1.00	0.63	−2.01	−42.00	1.29
Lively*Social science	8.00	1.76	1.79	2.51	1.67	−2.08	−43.29	1.16
Disorderly*Social science	8.00	1.48	1.57	2.10	1.38	−2.01	−41.85	1.30
To bed late*Social science	8.00	0.86	0.90	1.35	0.92	−2.57	−53.67	0.95
Wake up late*Interview	9.00	30.17	12.57	26.02	10.84	1.70	35.37	1.35
Social science	9.00	0.54	0.50	0.76	0.43	−2.22	0.71	1.36
Father edu*Social science	9.00	1.37	1.39	2.04	1.30	−2.40	−50.13	1.14
Mother edu*Social science	9.00	1.35	1.35	1.87	1.19	−1.97	−41.01	1.30
Pref mates*Social science	9.00	0.73	0.70	1.01	0.61	−2.06	−42.87	1.32
Lively*Social science	9.00	1.77	1.78	2.54	1.64	−2.16	−45.12	1.17
Disorderly*Social science	9.00	1.51	1.55	2.13	1.36	−2.04	−42.48	1.30
Disorderly*Fine arts	9.00	1.61	1.61	1.09	1.33	1.69	35.30	1.47
To bed late*Social science	9.00	0.90	0.90	1.35	0.90	−2.37	−49.40	1.01
Social science	10.00	0.54	0.50	0.76	0.43	−2.22	0.71	1.36
Father edu*Social science	10.00	1.39	1.41	2.02	1.29	−2.24	−46.71	1.19
Mother edu*Social science	10.00	1.37	1.37	1.83	1.18	−1.71	−35.68	1.35
Pref mates*Social science	10.00	0.74	0.71	1.03	0.61	−2.06	−42.88	1.35
Lively*Social science	10.00	1.83	1.78	2.58	1.63	−2.11	−43.95	1.20
Disorderly*Social science	10.00	1.52	1.56	2.18	1.36	−2.18	−45.38	1.32
To bed late*Social science	10.00	0.91	0.91	1.37	0.90	−2.41	−50.22	1.03

TABLE 13. Sample 680 (students with undefined treatment are assumed to be controls): the eigenvector associated with the highest eigenvalue of 3.15 in multiply-completed dataset I.

Mates	0.21
Father edu	−0.17
Mother edu	0.16
Pref mates	−0.00
Lively	0.12
Disorderly	−0.12
Wake up late	0.17
To bed late	−0.12
Interview	−0.01
Male	−0.31
Black	−0.53
West of US	0.15
Social science	0.22
Fine arts	−0.13
Social studies	−0.05
Psychology	0.15
Religious activities	0.23
Sports activities	0.34
Indep music	−0.12
Indep music etc	0.40

TABLE 14. Sample 680 (students with undefined treatment are assumed to be controls): θ^*_{max}.

	θ^*_{max}
Dataset I	1.22
Dataset II	1.16
Dataset III	1.22
Dataset IV	1.35
Dataset V	1.22
Dataset VI	1.39
Dataset VII	1.30
Dataset VIII	1.26
Dataset IX	1.19
Dataset X	1.20

TABLE 15. Sample 680 (students with undefined treatment are assumed to be controls): Imbalances, related to actual number of roommates, before matching for multiply-completed dataset I; similar imbalances are found for multiply-completed datasets II-X.

	X_{trt}	σ^2_{trt}	X_{cntrl}	σ^2_{cntrl}	t-test	B	R
Mates	1.09	0.38	0.73	0.56	5.54	77.36	0.46
Pref mates	1.39	0.26	1.24	0.33	3.34	49.34	0.62
Mates*Mates	1.33	0.72	0.84	0.76	4.26	67.34	0.89
Mates*Father edu	2.78	1.19	1.91	1.59	4.31	62.53	0.56
Mates*Mother edu	2.59	1.12	1.74	1.49	4.41	64.07	0.57
Mates*Pref. mates	1.54	0.63	0.94	0.80	5.58	82.49	0.62
Mates*Lively	3.82	1.43	2.22	1.88	6.58	95.88	0.57
Mates*Disorderly	3.15	1.52	1.92	1.68	4.96	77.12	0.81
Mates*Wake up late	2.48	1.13	1.48	1.22	5.44	85.43	0.86
Mates*To bed late	1.88	0.89	1.14	0.99	5.08	78.91	0.81
Mates*Interview	15.04	6.69	9.63	8.24	4.82	72.14	0.66
Mates*Male	0.51	0.58	0.34	0.51	1.87	31.22	1.26
Mates*Social science	0.59	0.64	0.42	0.56	1.76	29.59	1.31
Mates*Fine arts	0.65	0.61	0.38	0.55	2.70	45.11	1.25
Mates*Social studies	0.63	0.60	0.40	0.54	2.40	39.88	1.20
Mates*Sports activities	0.79	0.62	0.46	0.57	3.38	55.81	1.16

TABLE 16. Sample 680 (students with undefined treatment are assumed to be controls): Imbalances, not related to actual number of roommates, before matching for multiply-completed dataset I; similar imbalances are found for multiply-completed datasets II-X.

	X_{trt}	σ^2_{trt}	X_{cntrl}	σ^2_{cntrl}	t-test	B	R
Pref mates	1.39	0.26	1.24	0.33	3.34	49.34	0.62
Lively	3.48	0.60	3.05	0.85	4.09	57.91	0.49
Wake up late	2.24	0.60	2.02	0.51	2.34	39.62	1.37
To bed late	1.74	0.53	1.56	0.52	2.11	34.29	1.04
Father edu*Pref mates	3.63	1.17	3.24	1.28	2.04	31.96	0.85
Father edu*Lively	9.05	2.82	7.88	3.09	2.54	39.64	0.83
Father edu*Wake up late	5.83	2.11	5.19	1.92	1.88	31.27	1.21
Father edu*To bed late	4.57	1.89	4.04	1.81	1.74	28.51	1.09
Mother edu*Pref. mates	3.33	1.13	2.98	1.27	1.85	28.58	0.79
Mother edu*Lively	8.28	2.72	7.24	3.01	2.34	36.40	0.82
Mother edu*Wake up late	5.39	2.27	4.76	1.88	1.79	30.57	1.46
Mother edu*To bed late	4.20	1.87	3.70	1.72	1.68	27.83	1.18
Pref mates*Pref mates	1.99	0.73	1.65	0.77	2.90	45.81	0.88
Pref mates*Lively	4.86	1.39	3.89	1.68	4.19	63.18	0.69
Pref mates*Disorderly	4.06	1.65	3.34	1.47	2.76	46.16	1.26
Pref mates*Wake up late	3.10	1.05	2.52	0.97	3.49	57.65	1.15
Pref mates*To bed late	2.42	0.89	1.95	0.86	3.25	53.08	1.07
Pref mates*Interview	19.04	6.51	16.14	6.52	2.76	44.54	1.00
Lively*Lively	12.45	4.17	10.04	5.32	3.41	50.33	0.62
Lively*Disorderly	10.13	3.79	8.47	4.35	2.66	40.77	0.76
Lively*Wake up late	7.87	2.83	6.28	2.69	3.51	57.48	1.10
Lively*To bed late	6.07	2.22	4.91	2.42	3.18	49.75	0.84
Lively*Interview	47.62	15.18	39.82	16.92	3.12	48.51	0.80
Disorderly*Wake up late	6.63	3.09	5.47	2.53	2.40	41.14	1.49
Disorderly*To bed late	5.20	2.57	4.21	2.12	2.44	41.74	1.47
Disorderly*Interview	40.65	18.23	34.90	15.92	2.00	33.63	1.31
Wake up late*Wake up late	5.37	2.69	4.33	2.11	2.47	42.92	1.61
Wake up late*To bed late	3.98	1.76	3.21	1.47	2.77	47.27	1.42
Wake up late*Interview	31.07	12.98	26.39	10.50	2.30	39.63	1.53
To bed late*To bed late	3.30	1.85	2.70	1.69	2.06	34.15	1.20
To bed late*Interview	23.83	9.66	20.32	9.32	2.27	36.98	1.07
Indep music	0.41	0.50	0.56	0.50	−1.80	0.74	1.00
Indep music etc	0.48	0.51	0.63	0.49	−1.82	0.76	1.08
Father edu*Indep music	1.04	1.37	1.44	1.38	−1.79	−28.81	0.99
Father edu*Indep music etc	1.22	1.40	1.60	1.36	−1.73	−28.09	1.06
Lively*Sports activities	2.33	1.78	1.85	1.63	1.67	27.73	1.20
To bed late*Fine arts	1.11	0.99	0.83	0.86	1.75	29.54	1.33
PS	−1.03	0.93	−2.20	1.07	7.61	116.79	0.76

TABLE 17. Sample 680 (students with undefined treatment are assumed to be controls): balances before matching.

	PS		Covariates orthogonal to PS				
	B	R	[0,1/2]	(1/2,4/5]	(4/5,5/4]	(5/4,2]	(2,∞]
Dataset I	1.17	0.76	0.01	0.10	0.64	0.24	0.01
Dataset II	1.13	1.02	0.01	0.08	0.67	0.23	0.01
Dataset III	1.18	0.75	0.01	0.06	0.69	0.23	0.01
Dataset IV	1.07	0.69	0.01	0.09	0.65	0.24	0.01
Dataset V	1.18	0.72	0.01	0.07	0.68	0.23	0.00
Dataset VI	0.97	0.83	0.00	0.05	0.68	0.26	0.01
Dataset VII	1.03	0.91	0.01	0.05	0.68	0.26	0.00
Dataset VIII	1.16	0.66	0.01	0.09	0.65	0.24	0.01
Dataset IX	1.19	0.79	0.01	0.10	0.66	0.23	0.01
Dataset X	1.18	0.75	0.01	0.09	0.65	0.24	0.01

TABLE 18. Sample 680 (students with undefined treatment are assumed to be controls): bias reductions following propensity score matching.

	Pre-match B	Post-match B
Dataset I	1.17	0.03
Dataset II	1.13	0.07
Dataset III	1.18	0.04
Dataset IV	1.07	0.04
Dataset V	1.18	0.05
Dataset VI	0.97	0.04
Dataset VII	1.03	0.05
Dataset VIII	1.16	0.04
Dataset IX	1.19	0.01
Dataset X	1.18	0.02

TABLE 19. Sample 680 (students with undefined treatment are assumed to be controls): balances after propensity score matching.

	PS		Covariates orthogonal to PS				
	B	R	[0,1/2]	(1/2,4/5]	(4/5,5/4]	(5/4,2]	(2,∞]
Dataset I	0.17	0.65	0.01	0.20	0.64	0.10	0.04
Dataset II	0.42	1.33	0.02	0.16	0.62	0.18	0.02
Dataset III	0.26	0.98	0.01	0.13	0.63	0.15	0.07
Dataset IV	0.11	0.78	0.00	0.09	0.65	0.21	0.05
Dataset V	0.29	2.08	0.01	0.13	0.68	0.15	0.02
Dataset VI	0.14	1.40	0.00	0.09	0.76	0.15	0.01
Dataset VII	0.15	1.26	0.00	0.06	0.79	0.10	0.04
Dataset VIII	0.20	0.65	0.00	0.06	0.71	0.21	0.01
Dataset IX	0.36	1.83	0.01	0.16	0.64	0.17	0.01
Dataset X	0.25	0.83	0.01	0.02	0.73	0.22	0.01

TABLE 20. Sample 680 (students with undefined treatment are assumed to be controls): balances after Mahalanobis-metric matching.

	PS		Covariates orthogonal to PS				
	B	R	[0,1/2]	(1/2,4/5]	(4/5,5/4]	(5/4,2]	(2,∞]
Dataset I	0.15	1.01	0.02	0.18	0.70	0.07	0.03
Dataset II	0.27	1.33	0.03	0.12	0.71	0.12	0.01
Dataset III	0.21	1.39	0.01	0.18	0.65	0.11	0.04
Dataset IV	0.14	1.04	0.04	0.16	0.52	0.21	0.05
Dataset V	0.22	1.20	0.01	0.10	0.70	0.18	0.00
Dataset VI	0.20	1.13	0.00	0.04	0.82	0.12	0.01
Dataset VII	0.16	1.16	0.00	0.04	0.78	0.13	0.04
Dataset VIII	0.23	1.51	0.01	0.13	0.68	0.16	0.01
Dataset IX	0.18	0.94	0.01	0.13	0.78	0.08	0.00
Dataset X	0.22	1.11	0.01	0.10	0.73	0.12	0.04

TABLE 21. Sample 680 (students with undefined treatment are assumed to be controls): Imbalances after matching for all multiply-completed datasets.

	X_{trt}	σ^2_{trt}	X_{cntrl}	σ^2_{cntrl}	t-test	B	R	
Father edu*Social studies	1.00	1.72	1.42	1.22	1.41	1.69	35.29	1.02
Social science	2.00	0.50	0.51	0.70	0.47	−1.93	0.72	1.18
Fine arts	2.00	0.61	0.49	0.43	0.50	1.68	1.40	0.97
Father edu*Fine arts	2.00	1.65	1.40	1.05	1.36	2.08	43.30	1.06
Father edu*Social studies	2.00	1.72	1.42	1.13	1.39	2.00	41.67	1.05
Mother edu*Fine arts	2.00	1.50	1.31	0.96	1.19	2.08	43.36	1.21
Mother edu*Social studies	2.00	1.52	1.31	1.02	1.29	1.84	38.42	1.03
Pref mates*Social science	2.00	0.69	0.71	0.93	0.65	−1.70	−35.39	1.19
Pref mates*Fine arts	2.00	0.85	0.71	0.60	0.71	1.67	34.92	1.00
Wake up late*Social science	2.00	1.11	1.20	1.52	1.09	−1.73	−36.08	1.21
To bed late*Social science	2.00	0.87	0.96	1.26	0.93	−1.99	−41.48	1.06
Pref mates*Indep. music etc	3.00	0.71	0.71	0.47	0.62	1.75	36.50	1.29
Father edu	4.00	2.61	0.68	2.85	0.35	−2.12	−44.16	3.86
Wake up late	4.00	2.24	0.60	2.02	0.46	1.95	40.56	1.73
Father edu*Father edu	4.00	7.26	2.77	8.23	1.75	−2.00	−41.78	2.50
Father edu*Interview	4.00	35.26	13.49	40.60	12.61	−1.96	−40.87	1.14
Pref mates*Wake up late	4.00	3.08	1.08	2.72	0.93	1.70	35.51	1.36
Disorderly*Wake up late	4.00	6.60	3.07	5.41	2.36	2.07	43.26	1.69
Disorderly*To bed late	4.00	5.17	2.57	4.25	2.02	1.92	39.97	1.61
Wake up late*Wake up late	4.00	5.37	2.69	4.29	1.92	2.21	46.12	1.97
Wake up late*To bed late	4.00	3.98	1.76	3.35	1.63	1.77	36.86	1.16
Disorderly*Fine arts	4.00	1.78	1.69	1.16	1.47	1.88	39.20	1.32
Wake up late*Fine arts	4.00	1.33	1.21	0.91	1.07	1.73	36.11	1.28
Wake up late*Social studies	4.00	1.39	1.20	0.93	1.06	1.93	40.25	1.28
To bed late*Fine arts	4.00	1.07	1.00	0.65	0.85	2.14	44.59	1.38
Mates*Indep music etc	5.00	0.55	0.63	0.34	0.50	1.83	38.10	1.56
Father edu*Sports activities	5.00	1.65	1.37	2.15	1.31	−1.79	−37.28	1.09
Father edu*Indep music	6.00	1.02	1.34	1.53	1.45	−1.76	−36.62	0.86
Father edu*Indep music etc	6.00	1.20	1.38	1.73	1.43	−1.82	−37.97	0.93
Mother edu*Fine arts	6.00	1.50	1.31	1.04	1.15	1.77	36.95	1.29
Lively*Indep music	6.00	1.33	1.63	1.97	1.85	−1.76	−36.80	0.78
Lively*Indep music etc	6.00	1.57	1.70	2.20	1.81	−1.73	−36.00	0.88
Interview*Fine arts	6.00	8.22	7.30	5.65	6.35	1.80	37.51	1.32
Father edu*Indep music	7.00	1.04	1.37	1.55	1.45	−1.74	−36.32	0.89
Mother edu*Fine arts	7.00	1.43	1.31	0.96	1.11	1.89	39.31	1.38
Lively*Indep music	7.00	1.37	1.68	1.99	1.85	−1.68	−35.07	0.82
Interview*Fine arts	7.00	7.98	7.39	5.17	5.98	2.00	41.72	1.52
Mother edu*Fine arts	8.00	1.50	1.31	1.04	1.26	1.70	35.43	1.08
Pref mates*Disorderly	10.00	4.03	1.68	3.44	1.43	1.83	38.13	1.38

TABLE 22. Sample 632 (students with undefined treatment are dropped): the eigenvector associated with the highest eigenvalue of 3.08 in multiply-completed dataset I.

Mates	0.13
Father edu	0.01
Mother edu	0.05
Pref. Pref. mates	0.05
Lively	0.24
Disorderly	−0.14
Wake up late	0.37
To bed late	−0.32
Interview	−0.09
Male	−0.39
Black	−0.19
West of US	0.27
Social science	−0.07
Fine arts	−0.42
Social studies	0.16
Psychology	0.08
Religious activities	0.28
Sports activities	0.20
Indep music	−0.24
Indep music etc	0.01

TABLE 23. Sample 632 (students with undefined treatment are dropped): θ^*_{max}.

	θ^*_{max}
Dataset I	1.40
Dataset II	1.38
Dataset III	1.40
Dataset IV	1.42
Dataset V	1.37
Dataset VI	1.40
Dataset VII	1.37
Dataset VIII	1.40
Dataset IX	1.43
Dataset X	1.41

TABLE 24. Sample 632 (students with undefined treatment are dropped): Imbalances, related to actual number of roommates, before matching for multiply-completed dataset I; similar imbalances are found for multiply-completed datasets II-X.

	X_{trt}	σ^2_{trt}	X_{cntrl}	σ^2_{cntrl}	t-test	B	R
Mates	1.09	0.38	0.68	0.56	6.11	86.89	0.45
Mates*Mates	1.33	0.72	0.77	0.75	4.73	75.97	0.90
Mates*Father edu	2.74	1.20	1.78	1.60	4.59	68.03	0.57
Mates*Mother edu	2.66	1.15	1.60	1.46	5.35	80.58	0.62
Mates*Pref mates	1.52	0.63	0.88	0.80	5.98	89.90	0.61
Mates*Lively	3.72	1.46	2.08	1.86	6.56	98.72	0.62
Mates*Disorderly	3.11	1.54	1.87	1.78	4.77	74.28	0.75
Mates*Wake up late	2.38	1.12	1.40	1.24	5.25	83.08	0.83
Mates*To bed late	1.82	0.77	1.07	1.01	5.61	83.41	0.58
Mates*Interview	15.33	6.42	8.64	7.84	6.12	93.43	0.67
Mates*Male	0.51	0.58	0.32	0.51	2.05	34.88	1.30
Mates*Social science	0.66	0.63	0.45	0.57	2.11	35.63	1.24
Mates*Fine arts	0.62	0.61	0.40	0.55	2.18	36.89	1.26
Mates*Social studies	0.53	0.59	0.36	0.52	1.76	29.78	1.28
Mates*Sports activities	0.76	0.62	0.46	0.57	2.97	49.71	1.17
Mates*Indep music	0.56	0.61	0.37	0.53	1.94	33.10	1.35
Mates*Indep music etc	0.61	0.61	0.40	0.54	2.10	35.54	1.27

TABLE 25. Sample 632 (students with undefined treatment are dropped): Imbalances, not related to actual number of roommates, before matching for multiply-completed dataset I; similar imbalances are found for multiply-completed datasets II-X.

	X_{trt}	σ^2_{trt}	X_{cntrl}	σ^2_{cntrl}	t-test	B	R
Pref mates	1.37	0.27	1.24	0.33	2.82	42.87	0.65
Lively	3.40	0.74	3.05	0.80	2.83	45.12	0.86
Interview	14.04	3.79	12.81	3.72	2.00	32.85	1.04
Mother edu*Pref mates	3.37	1.15	2.98	1.28	2.05	32.30	0.80
Mother edu*Lively	8.23	2.81	7.25	3.05	2.09	33.20	0.85
Mother edu*Interview	34.57	13.99	30.16	12.98	1.95	32.63	1.16
Pref mates*Pref mates	1.96	0.73	1.65	0.78	2.50	39.87	0.87
Pref mates*Lively	4.72	1.53	3.90	1.62	3.26	52.22	0.88
Pref mates*Disorderly	3.96	1.67	3.44	1.57	1.95	32.39	1.13
Pref mates*Wake up late	2.95	1.04	2.55	0.96	2.40	40.21	1.18
Pref mates*To bed late	2.34	0.84	1.99	0.93	2.50	39.64	0.83
Pref mates*Interview	19.25	6.34	15.73	5.98	3.43	56.97	1.12
Lively*Lively	12.11	4.69	9.96	4.88	2.79	44.95	0.92
Lively*Disorderly	9.84	4.11	8.64	4.28	1.77	28.52	0.92
Lively*Wake up late	7.46	2.90	6.32	2.45	2.47	42.37	1.40
Lively*To bed late	5.87	2.24	4.99	2.46	2.38	37.63	0.83
Lively*Interview	47.84	16.23	39.07	15.53	3.34	55.19	1.09
Disorderly*Interview	40.78	17.90	35.31	16.49	1.90	31.78	1.18
Wake up late*To bed late	3.74	1.61	3.30	1.54	1.69	28.01	1.10
Wake up late*Interview	30.65	13.05	26.20	10.11	2.17	38.12	1.67
To bed late*Interview	24.24	9.88	20.14	8.96	2.58	43.46	1.22
Interview*Interview	211.30	108.91	177.86	104.71	1.89	31.31	1.08
Pref mates*Psychology	0.10	0.32	0.21	0.51	−1.92	−26.78	0.41
PS	−1.02	0.84	−2.03	1.10	6.87	102.65	0.59

TABLE 26. Sample 632 (students with undefined treatment are dropped): balances before matching.

	PS		Covariates orthogonal to PS				
	B	R	[0,1/2]	(1/2,4/5]	(4/5,5/4]	(5/4,2]	(2,∞]
Dataset I	1.03	0.59	0.01	0.08	0.66	0.23	0.01
Dataset II	1.04	0.60	0.01	0.10	0.67	0.21	0.00
Dataset III	1.03	0.60	0.01	0.05	0.70	0.24	0.00
Dataset IV	1.02	0.58	0.01	0.09	0.70	0.20	0.00
Dataset V	1.04	0.63	0.01	0.07	0.69	0.23	0.00
Dataset VI	1.03	0.59	0.01	0.06	0.70	0.23	0.00
Dataset VII	1.04	0.62	0.01	0.06	0.69	0.24	0.00
Dataset VIII	1.03	0.59	0.01	0.07	0.68	0.23	0.00
Dataset IX	1.01	0.60	0.01	0.06	0.72	0.21	0.00
Dataset X	1.03	0.58	0.01	0.07	0.71	0.21	0.00

TABLE 27. Sample 632 (students with undefined treatment are dropped): bias reductions following propensity score matching.

	Pre-match B	Post-match B
Dataset I	1.03	0.02
Dataset II	1.04	0.02
Dataset III	1.03	0.02
Dataset IV	1.02	0.02
Dataset V	1.04	0.03
Dataset VI	1.03	0.03
Dataset VII	1.04	0.03
Dataset VIII	1.03	0.02
Dataset IX	1.01	0.03
Dataset X	1.03	0.03

TABLE 28. Sample 632 (students with undefined treatment are dropped): balances after propensity score matching.

	PS		Covariates orthogonal to PS				
	B	R	[0,1/2]	(1/2,4/5]	(4/5,5/4]	(5/4,2]	$(2,\infty]$
Dataset I	0.09	0.85	0.02	0.13	0.71	0.09	0.04
Dataset II	0.19	0.87	0.01	0.16	0.64	0.13	0.04
Dataset III	0.10	0.83	0.00	0.12	0.75	0.10	0.02
Dataset IV	0.13	0.83	0.00	0.13	0.76	0.10	0.01
Dataset V	0.09	0.87	0.01	0.10	0.72	0.12	0.04
Dataset VI	0.12	0.96	0.02	0.13	0.68	0.10	0.05
Dataset VII	0.16	0.83	0.00	0.13	0.66	0.16	0.04
Dataset VIII	0.11	0.91	0.00	0.15	0.69	0.12	0.04
Dataset IX	0.08	0.85	0.01	0.16	0.71	0.07	0.04
Dataset X	0.18	0.85	0.01	0.09	0.73	0.13	0.04

TABLE 29. Sample 632 (students with undefined treatment are dropped): balances after Mahalanobis-metric matching.

	PS		Covariates orthogonal to PS				
	B	R	[0,1/2]	(1/2,4/5]	(4/5,5/4]	(5/4,2]	$(2,\infty]$
Dataset I	0.12	0.85	0.02	0.16	0.68	0.12	0.02
Dataset II	0.11	0.87	0.02	0.22	0.70	0.06	0.00
Dataset III	0.08	0.86	0.02	0.17	0.72	0.08	0.00
Dataset IV	0.04	0.97	0.03	0.21	0.68	0.07	0.01
Dataset V	0.13	0.92	0.03	0.19	0.67	0.10	0.00
Dataset VI	0.06	0.97	0.04	0.15	0.73	0.07	0.01
Dataset VII	0.14	0.90	0.01	0.15	0.70	0.13	0.01
Dataset VIII	0.12	0.87	0.02	0.21	0.70	0.06	0.01
Dataset IX	0.11	0.88	0.01	0.14	0.69	0.13	0.02
Dataset X	0.05	0.88	0.02	0.23	0.69	0.05	0.00

TABLE 30. Sample 632 (students with undefined treatment are dropped): Imbalances after matching for multiply-completed datasets I-X .

	X_{trt}	σ^2_{trt}	X_{cntrl}	σ^2_{cntrl}	t-test	B	R	
Father edu*Sports activities	1.00	1.57	1.36	2.13	1.24	−2.08	−43.42	1.20
Father edu	4.00	2.57	0.69	2.80	0.45	−1.97	−41.05	2.31
Father edu*Father edu	4.00	7.04	2.82	8.07	2.10	−1.97	−41.08	1.80
Father edu*Lively	4.00	8.72	3.15	9.86	2.97	−1.79	−37.31	1.12
Father edu*Sports activities	5.00	1.57	1.36	2.05	1.28	−1.77	−36.98	1.12
Interview*Indep music etc	6.00	8.24	7.90	5.59	7.22	1.68	35.06	1.20
Father edu*Sports activities	7.00	1.61	1.34	2.11	1.23	−1.86	−38.81	1.18
Father edu	8.00	2.57	0.69	2.82	0.44	−2.08	−43.31	2.45
Father edu*Father edu	8.00	7.04	2.82	8.11	2.03	−2.09	−43.57	1.93
Father edu*Social science	8.00	1.37	1.39	1.92	1.35	−1.94	−40.49	1.06
Father edu	9.00	2.57	0.69	2.77	0.47	−1.68	−35.10	2.16
Social science	9.00	0.57	0.50	0.74	0.44	−1.76	0.76	1.27
Father edu*Social science	9.00	1.41	1.38	2.01	1.28	−2.15	−44.92	1.15
Father edu*Sports activities	9.00	1.50	1.36	1.99	1.33	−1.74	−36.37	1.05

TABLE 31. Sample 55 (students with undefined treatment are dropped): the eigenvector associated with the highest eigenvalue of 3.03 in multiply-completed dataset I.

Mates	0.04
Father edu	−0.12
Mother edu	0.22
Pref mates	0.04
Lively	−0.14
Disorderly	0.28
Wake up late	0.21
To bed late	0.29
Interview	0.38
Male	0.12
Black	−0.33
West of US	−0.32
Social science	0.27
Fine arts	0.26
Social studies	0.18
Psychology	−0.15
Religious activities	−0.09
Sports activities	0.07
Indep music	0.21
Indep music etc	0.28

TABLE 32. Sample 55 (students with undefined treatment are dropped): θ^*_{max}.

	θ^*_{max}
Dataset I	1.11
Dataset II	1.01
Dataset III	1.14
Dataset IV	1.42
Dataset V	1.33
Dataset VI	1.43
Dataset VII	1.11
Dataset VIII	1.04
Dataset IX	1.01
Dataset X	1.06

TABLE 33. Sample 55 (students with undefined treatment are dropped): Imbalances, related to actual number of roommates, before matching for multiply-completed dataset I; similar imbalances are found for multiply-completed datasets II-X.

	X_{trt}	σ^2_{trt}	X_{cntrl}	σ^2_{cntrl}	t-test	B	R
Mates	1.09	0.38	0.68	0.56	6.11	86.89	0.45
Mates*Mates	1.33	0.72	0.77	0.75	4.73	75.97	0.90
Mates*Father edu	2.72	1.20	1.76	1.59	4.59	68.14	0.57
Mates*Mother edu	2.69	1.14	1.64	1.50	5.29	78.89	0.58
Mates*Pref mates	1.54	0.64	0.88	0.79	6.03	91.56	0.65
Mates*Lively	3.74	1.43	2.06	1.85	6.75	101.22	0.60
Mates*Disorderly	3.06	1.45	1.90	1.80	4.67	70.99	0.65
Mates*Wake up late	2.39	1.06	1.39	1.22	5.62	87.63	0.76
Mates*To bed late	1.82	0.77	1.08	1.01	5.55	82.47	0.57
Mates*Interview	15.27	6.58	9.06	8.24	5.50	83.32	0.64
Mates*Male	0.51	0.58	0.32	0.51	2.05	34.88	1.30
Mates*Social science	0.59	0.63	0.42	0.56	1.70	28.77	1.27
Mates*Fine arts	0.64	0.61	0.39	0.54	2.54	42.93	1.26
Mates*Social studies	0.60	0.60	0.33	0.51	2.84	48.65	1.39
Mates*Sports activities	0.79	0.62	0.43	0.56	3.64	61.17	1.20
Mates*Indep music	0.58	0.63	0.37	0.53	2.05	35.37	1.44
Mates*Indep music etc	0.63	0.63	0.42	0.54	2.06	35.01	1.33

TABLE 34. Sample 55 (students with undefined treatment are dropped): Imbalances, not related to actual number of roommates, before matching for multiply-completed dataset I; similar imbalances are found for multiply-completed datasets II-X.

	X_{trt}	σ^2_{trt}	X_{cntrl}	σ^2_{cntrl}	t-test	B	R
Pref mates	1.39	0.26	1.25	0.33	3.14	47.59	0.64
Lively	3.42	0.71	3.04	0.80	3.20	50.29	0.79
Father edu*Pref mates	3.56	1.19	3.20	1.27	1.83	29.27	0.88
Father edu*Lively	8.78	3.13	7.74	3.04	2.06	33.94	1.06
Mother edu*Pref mates	3.42	1.06	3.03	1.24	2.20	34.06	0.73
Mother edu*Lively	8.34	2.66	7.31	2.94	2.31	36.55	0.82
Pref mates*Pref mates	1.99	0.73	1.66	0.77	2.76	44.21	0.89
Pref mates*Lively	4.77	1.43	3.89	1.62	3.69	57.90	0.78
Pref mates*Disorderly	3.96	1.59	3.48	1.58	1.83	29.91	1.02
Pref mates*Wake up late	2.99	0.98	2.53	0.92	2.94	48.97	1.14
Pref mates*To bed late	2.37	0.85	1.99	0.88	2.77	44.66	0.93
Pref mates*Interview	19.41	6.67	16.32	6.24	2.87	47.85	1.14
Lively*Lively	12.22	4.55	9.89	4.89	3.09	49.32	0.87
Lively*Wake up late	7.57	2.87	6.24	2.45	2.90	49.55	1.37
Lively*To bed late	5.91	2.22	4.93	2.33	2.70	43.40	0.91
Lively*Interview	47.84	16.28	40.16	16.24	2.89	47.23	1.00
Wake up late*Wake up late	5.09	2.65	4.33	2.03	1.81	31.97	1.71
Wake up late*To bed late	3.78	1.59	3.26	1.44	2.06	34.69	1.22
Wake up late*Interview	30.61	12.44	26.68	10.05	2.00	34.71	1.53
To bed late*Interview	24.02	9.89	20.74	9.07	2.06	34.58	1.19
Lively*Male	1.76	1.77	1.28	1.55	1.72	29.12	1.30
PS	−0.81	1.03	−2.26	1.29	8.16	123.57	0.64

TABLE 35. Sample 55 (students with undefined treatment are dropped): balances before matching.

	PS		Covariates orthogonal to PS				
	B	R	[0,1/2]	(1/2,4/5]	(4/5,5/4]	(5/4,2]	(2,∞]
Dataset I	1.24	0.64	0.01	0.05	0.66	0.26	0.01
Dataset II	1.35	0.64	0.01	0.12	0.64	0.23	0.00
Dataset III	1.17	0.73	0.01	0.07	0.66	0.26	0.00
Dataset IV	1.01	0.59	0.01	0.07	0.66	0.24	0.01
Dataset V	1.03	0.72	0.01	0.06	0.65	0.27	0.01
Dataset VI	1.00	0.56	0.01	0.06	0.66	0.25	0.02
Dataset VII	1.19	0.79	0.01	0.07	0.67	0.25	0.00
Dataset VIII	1.32	0.62	0.01	0.10	0.65	0.23	0.00
Dataset IX	1.29	0.66	0.01	0.08	0.68	0.23	0.00
Dataset X	1.23	0.70	0.01	0.08	0.69	0.22	0.00

TABLE 36. Sample 55 (students with undefined treatment are dropped): bias reductions following propensity score matching.

	Pre-match B	Post-match B
Dataset I	1.24	0.07
Dataset II	1.35	0.06
Dataset III	1.17	0.08
Dataset IV	1.01	0.01
Dataset V	1.03	0.01
Dataset VI	1.00	0.01
Dataset VII	1.19	0.11
Dataset VIII	1.32	0.07
Dataset IX	1.29	0.11
Dataset X	1.23	0.04

TABLE 37. Sample 55 (students with undefined treatment are dropped): balances after propensity score matching.

	PS		Covariates orthogonal to PS				
	B	R	[0,1/2]	(1/2,4/5]	(4/5,5/4]	(5/4,2]	$(2,\infty]$
Dataset I	0.21	0.96	0.01	0.18	0.74	0.05	0.01
Dataset II	0.16	1.07	0.01	0.13	0.73	0.12	0.01
Dataset III	0.18	1.00	0.05	0.05	0.71	0.16	0.02
Dataset IV	0.14	1.23	0.02	0.11	0.74	0.10	0.01
Dataset V	0.01	0.97	0.02	0.20	0.66	0.11	0.01
Dataset VI	0.09	1.14	0.01	0.12	0.74	0.11	0.01
Dataset VII	0.24	1.49	0.00	0.05	0.77	0.13	0.04
Dataset VIII	0.31	0.74	0.00	0.09	0.82	0.08	0.01
Dataset IX	0.25	0.60	0.00	0.18	0.65	0.13	0.03
Dataset X	0.20	1.00	0.02	0.09	0.74	0.13	0.01

TABLE 38. Sample 55 (students with undefined treatment are dropped): balances after Mahalanobis-metric matching.

	PS		Covariates orthogonal to PS				
	B	R	[0,1/2]	(1/2,4/5]	(4/5,5/4]	(5/4,2]	$(2,\infty]$
Dataset I	0.18	1.50	0.01	0.15	0.75	0.08	0.00
Dataset II	0.35	1.25	0.01	0.18	0.68	0.12	0.00
Dataset III	0.24	1.21	0.03	0.12	0.75	0.09	0.01
Dataset IV	0.09	1.04	0.03	0.14	0.70	0.12	0.01
Dataset V	0.01	0.97	0.02	0.17	0.66	0.14	0.01
Dataset VI	0.04	1.00	0.02	0.18	0.68	0.11	0.01
Dataset VII	0.29	1.18	0.01	0.13	0.72	0.12	0.02
Dataset VIII	0.35	1.68	0.01	0.21	0.70	0.07	0.00
Dataset IX	0.35	2.86	0.00	0.15	0.78	0.07	0.00
Dataset X	0.17	1.47	0.01	0.16	0.68	0.13	0.00

TABLE 39. Sample 55 (students with undefined treatment are dropped): Imbalances after matching for multiply-completed datasets I-V.

	X_{trt}	σ^2_{trt}	X_{cntrl}	σ^2_{cntrl}	t-test	B	R	
Father edu*To bed late	2.00	4.39	1.82	5.09	1.71	−1.89	−39.40	1.13
PS	2.00	0.06	0.37	−0.06	0.33	1.67	34.89	1.25
Social science	4.00	0.52	0.51	0.80	0.40	−2.97	0.65	1.59
Social studies	4.00	0.59	0.50	0.33	0.47	2.57	1.80	1.10
Mates*Social science	4.00	0.61	0.63	0.83	0.53	−1.82	−37.93	1.40
Mates*Social studies	4.00	0.60	0.60	0.37	0.56	1.93	40.30	1.17
Father edu*Social science	4.00	1.33	1.40	2.12	1.20	−2.92	−60.92	1.36
Father edu*Social studies	4.00	1.59	1.41	0.92	1.36	2.30	47.94	1.07
Mother edu*Social science	4.00	1.30	1.36	1.93	1.10	−2.44	−50.81	1.53
Mother edu*Social studies	4.00	1.50	1.33	0.87	1.29	2.31	48.08	1.06
Pref. mates*Social science	4.00	0.71	0.71	1.08	0.59	−2.76	−57.60	1.44
Pref. mates*Social studies	4.00	0.78	0.71	0.47	0.69	2.19	45.67	1.05
Lively*Social science	4.00	1.73	1.76	2.71	1.50	−2.87	−59.76	1.38
Lively*Social studies	4.00	1.99	1.75	1.12	1.72	2.40	50.13	1.03
Disorderly*Social science	4.00	1.45	1.54	2.37	1.44	−2.98	−62.16	1.14
Wake up late*Social science	4.00	1.11	1.16	1.64	0.96	−2.40	−50.08	1.46
Wake up late*Social studies	4.00	1.33	1.19	0.62	0.95	3.14	65.50	1.58
To bed late*Social science	4.00	0.86	0.90	1.30	0.81	−2.50	−52.21	1.24
To bed late*Social studies	4.00	0.99	0.93	0.55	0.87	2.31	48.16	1.15
Interview*Social science	4.00	7.54	7.80	10.76	6.45	−2.16	−44.98	1.46
Interview*Social studies	4.00	7.83	7.26	4.00	6.24	2.71	56.55	1.35
Social science	5.00	0.52	0.51	0.80	0.40	−2.97	0.65	1.59
Social studies	5.00	0.59	0.50	0.35	0.48	2.34	1.69	1.07
Mates*Social science	5.00	0.61	0.63	0.85	0.54	−1.89	−39.49	1.34
Father edu*Social science	5.00	1.33	1.40	2.07	1.21	−2.71	−56.43	1.33
Father edu*Social studies	5.00	1.59	1.41	0.97	1.38	2.13	44.51	1.05
Mother edu*Social science	5.00	1.30	1.36	1.91	1.07	−2.38	−49.63	1.62
Mother edu*Social studies	5.00	1.50	1.33	0.89	1.29	2.23	46.54	1.07
Pref mates*Social science	5.00	0.72	0.71	1.07	0.60	−2.56	−53.46	1.41
Pref mates*Social studies	5.00	0.79	0.70	0.50	0.70	1.97	40.99	1.00
Lively*Social science	5.00	1.78	1.83	2.74	1.54	−2.71	−56.47	1.42
Lively*Social studies	5.00	1.98	1.74	1.26	1.83	1.92	40.11	0.91
Disorderly*Social science	5.00	1.46	1.54	2.41	1.47	−3.04	−63.45	1.11
Wake up late*Social science	5.00	1.11	1.16	1.65	0.97	−2.44	−50.83	1.42
Wake up late*Social studies	5.00	1.33	1.19	0.65	0.95	3.00	62.53	1.59
To bed late*Social science	5.00	0.85	0.89	1.29	0.81	−2.51	−52.35	1.23
To bed late*Social studies	5.00	0.98	0.93	0.59	0.88	2.07	43.12	1.11
Interview*Social science	5.00	7.54	7.80	10.54	6.26	−2.03	−42.43	1.55
Interview*Social studies	5.00	7.83	7.26	4.52	6.67	2.27	47.41	1.18

TABLE 40. Sample 55 (students with undefined treatment are dropped): Imbalances after matching for multiply-completed datasets VI-X.

	X_{trt}	σ^2_{trt}	X_{cntrl}	σ^2_{cntrl}	t-test	B	R	
Father edu*Disorderly	6.00	7.30	3.06	8.47	3.40	−1.72	−35.95	0.81
Social science	6.00	0.52	0.51	0.72	0.46	−1.95	0.73	1.23
Social studies	6.00	0.57	0.50	0.33	0.47	2.35	1.73	1.12
Mates*Social studies	6.00	0.58	0.60	0.37	0.56	1.73	36.13	1.17
Father edu*Social science	6.00	1.33	1.40	1.92	1.33	−2.10	−43.74	1.10
Father edu*Social studies	6.00	1.52	1.41	0.93	1.37	2.02	42.18	1.06
Mother edu*Social studies	6.00	1.43	1.33	0.85	1.28	2.16	44.98	1.07
Pref. mates*Social science	6.00	0.72	0.71	0.96	0.64	−1.70	−35.54	1.22
Pref. mates*Social studies	6.00	0.77	0.71	0.46	0.68	2.16	45.12	1.09
Lively*Social science	6.00	1.74	1.78	2.37	1.64	−1.77	−36.84	1.18
Lively*Social studies	6.00	1.93	1.77	1.13	1.72	2.21	46.10	1.06
Disorderly*Social science	6.00	1.48	1.57	2.11	1.54	−1.94	−40.51	1.05
Wake up late*Social studies	6.00	1.28	1.20	0.61	0.93	3.00	62.61	1.68
To bed late*Social studies	6.00	0.93	0.93	0.57	0.89	1.95	40.73	1.10
Interview*Social studies	6.00	7.59	7.33	4.00	6.20	2.53	52.82	1.40
PS	8.00	0.07	0.41	−0.06	0.32	1.69	35.29	1.68
Father edu*Disorderly	9.00	7.27	3.11	8.51	3.53	−1.79	−37.28	0.77

TABLE 41. The treatment effects on different improper covariates estimated using OLS regressions. No covariate adjustments.

	τ	σ^2	p-value
Smoke a pipe or cigars - unpair	0.09	0.31	0.12
Chewing tobacco or snuff - unpair	0.00	0.00	
Use drugs - unpair	−0.01	0.24	0.87
Times consumed alcohol/month - unpair	0.28	3.61	0.15
Drinks/time when consuming alcohol - unpair	0.25	2.24	0.09
≥ 5 drinks/time - unpair	0.11	1.04	0.27
Moderate physical exerice - unpair	0.27	4.05	0.18
Vigorous physical exercise - unpair	0.02	4.87	0.91
Father smoked - unpair	−0.02	0.49	0.83
Mother smoked - unpair	0.05	0.35	0.41
Father consumed alcohol - unpair	−0.06	2.47	0.70
Mother consumed alcohol - unpair	0.08	2.17	0.60
Carefully complete HAF - unpair	−0.02	5.11	0.92
Truthfully complete HAF - unpair	−0.14	1.94	0.33
Religious tolerance - unpair	0.14	1.03	0.18
Help completing HAF - unpair	0.24	5.62	0.31
Ever used alcohol before college - unpair	0.14	0.73	0.11
Ever smoked a cigarette before college - unpair	−0.02	0.49	0.80
Ever smoked a cigar before college- unpair	0.06	0.59	0.47
Ever smoked pipe before college - unpair	0.00	0.10	0.99
Ever used chewing tobacco before college - unpair	0.00	0.10	0.99
Ever used snuff before college - unpair	−0.03	0.07	0.30
Ever used illegal drugs before college- unpair	0.10	0.77	0.27

TABLE 42. The treatment effects on different improper covariates estimated using OLS regressions. No covariate adjustments but indicators for matched pairs used.

	τ	σ^2	p-value
Smoke a pipe or cigars - pair	0.09	0.34	0.15
Chewing tobacco or snuff - pair	0.00	0.00	
Use drugs - pair	−0.01	0.27	0.82
Times consumed alcohol/month - pair	0.27	3.91	0.18
Drinks/time when consuming alcohol - pair	0.28	2.88	0.10
$\geq$ 5 drinks/time - pair	0.12	0.99	0.22
Moderate physical exercise - pair	0.27	4.14	0.19
Vigorous physical exercise - pair	0.02	5.94	0.92
Father smoked - pair	−0.02	0.55	0.84
Mother smoked - pair	0.05	0.32	0.37
Father consumed alcohol - pair	−0.06	2.59	0.71
Mother consumed alcohol - pair	0.08	2.24	0.60
Carefully complete HAF - pair	0.05	5.65	0.82
Truthfully complete HAF - pair	−0.11	1.73	0.39
Religious tolerance - pair	0.15	1.06	0.16
Help completing HAF - pair	0.26	6.08	0.29
Ever used alcohol before college - pair	0.14	0.84	0.13
Ever smoked a cigarette before college - pair	−0.02	0.50	0.78
Ever smoked a cigar before college - pair	0.05	0.71	0.53
Ever smoked a pipe before college - pair	0.00	0.10	1.00
Ever used chewing tobacco before college - pair	0.00	0.10	0.89
Ever used snuff before college - pair	−0.03	0.07	0.30
Ever used illegal drugs before college - pair	0.11	0.65	0.20

TABLE 43. The treatment effects on different improper covariates estimated using regressions that multiply-impute the missing potential outcomes. No covariate adjustments.

	τ	σ^2	p-value
Smoke a pipe or cigars - multiply impute	0.09	0.34	0.13
Chewing tobacco or snuff - multiply impute	0.00	0.00	
Use drugs - multiply impute	−0.01	0.25	0.88
Times consumed alcohol/month - multiply impute	0.28	4.07	0.17
Drinks/time when consuming alcohol - multiply impute	0.26	2.39	0.10
$\geq$ 5 drinks/time - multiply impute	0.11	1.18	0.30
Moderate physical exercise - multiply impute	0.27	4.53	0.20
Vigorous physical exercise - multiply impute	0.02	5.39	0.92
Father smoked - multiply impute	−0.02	0.52	0.83
Mother smoked - multiply impute	0.05	0.37	0.42
Father consumed alcohol - multiply impute	−0.06	2.61	0.71
Mother consumed alcohol - multiply impute	0.08	2.27	0.60
Carefully complete HAF - multiply impute	−0.02	5.54	0.92
Truthfully complete HAF - multiply impute	−0.13	2.06	0.35
Religious tolerance - multiply impute	0.14	1.09	0.18
Help completing HAF - multiply impute	0.24	5.92	0.32
Ever used alcohol before college - multiply impute	0.14	0.77	0.12
Ever smoked a cigarette before college - multiply impute	−0.02	0.59	0.81
Ever smoked a cigar before college - multiply impute	0.06	0.64	0.49
Ever smoked a pipe before college - multiply impute	0.00	0.10	0.98
Ever used chewing tobacco before college - multiply impute	0.00	0.10	0.98
Ever used snuff before college - multiply impute	−0.03	0.08	0.33
Ever used illegal drugs before college - multiply impute	0.10	0.84	0.29

TABLE 44. The effect of treatment on the number of reported drinks consumed *prior to* college enrollment when the stepwise procedure keeps stepping-in the most predictive covariate that is significant at the 5% level.

	τ	σ^2	p-value
Unpaired OLS-regression	0.22	2.13	0.14
Paired OLS-regression	0.22	2.89	0.21
RCM	0.23	2.20	0.14

TABLE 45. The effect of treatment on the number of reported drinks consumed *prior to* college enrollment when the the stepwise procedure that keeps adding the most predictive covariate given that it is significant at the $\frac{1}{5(q-p)}$ level where q is the number of proper covariates, their cross-products, and their interactions and p is the number of covariates already included in the regression.

	τ	σ^2	p-value
Unpaired OLS-regression	0.25	2.17	0.09
Paired OLS-regression	0.27	2.99	0.13
RCM	0.25	2.42	0.11

TABLE 46. The effect of treatment on the number of reported drinks consumed *prior to* college enrollment when the the stepwise procedure keeps adding the most predictive covariate significant at the $\frac{1}{5(q-p)}$ level where q is the 6 covariates believed to be most predictive of outcome and p is the number of covariates already included in the regression.

	τ	σ^2	p-value
Unpaired OLS-regression	0.24	2.02	0.09
Paired OLS-regression	0.24	2.70	0.15
RCM	0.24	2.16	0.11

References

Angrist, Joshua D, Guido W Imbens, and Donald B Rubin (1996) "Identification of Causal Effects Using Instrumental Variables" (with discussion) *Journal of the American Statistical Association*, 91:444-472

Barnard, John, Donald B Rubin (1999) "Small-sample Degrees of Freedom with Multiple Imputation" *Biometrica*, 86(4): 948-955

Bingenheimer, Jeffrey B, Robert T Brennan, Earls J Felton (2005) "Firearm Violence Exposure and Serious Violent Behavior" *Science*, 308(5726): 1323-1326

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

Peters, C C (1941) "A method of matching groups for experiment with no loss of population" *Journal of Educational Research* 34:606-612

Raghunathan, E Trivellore, James M Lepkowski, John Van Hoewyk, Peter Solenberger (2001) "A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models" *Survey Methodology*, 27:85-95

Rosenbaum, Paul R, Donald B Rubin (1985) "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score" *American Statistical Association*, 39(1): 33-38

Rosenbaum, Paul R, Donald B Rubin (1983) "The Central Role of the Propensity Score in Observational Studies for Causal Effects" *Biometrika*, 70(1): 41-55

Rubin, Donald B (2005) "Causal Inference Using Potential Outcomes: Design, Modeling, Decision" *Journal of the American Statistical Association*, 100: 322-331

Rubin, Donald B (2001) "Using Propensity Scores to Help Design Observational Studies: Application to Tobacco Litigation" *Health Services and Outcomes Research Methodology*, 2, 169-188

Rubin, Donald B (2000) "Combining Propensity Score Matching With Additional Adjustments for Prognostic Covariates" *Journal of the American Statistical Association*, 95(450): 573-585

Rubin, Donald B (1987) "Multiple Imputation for Nonresponse in Surveys" John Wiley & Sons, Inc

Rubin, Donald B (1980)"Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment" *Journal of the American Statistical Association*, 75(371): 591-593

"Inference and Missing Data" *Biometrika*, 63(3):581-592

Rubin, Donald B (1978) "Multiple Imputations in Sample Surveys - Phenomenological Bayesian Approach to Nonresponse" *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 20-34. Also in *Imputation and Editing of Faulty or Missing Survey Data*, U.S. Dept. of Commerce, Bureau of the Census, 1-23

Rubin, Donald B (1978) "Baysian Inference for Causal Effects: The Role of Randomization" *The Annals of Statistics*, 6(1): 34-58

Rubin, Donald B (1977) "Assignment to Treatment Group on the Basis of a Covariate" *Journal of Educational Statistics*, 2(1):1-26

Rubin, Donald B (1977) "The Design of a General and Flexible System for Handling Non-response in Sample Surveys" Manuscript Prepared for the U.S. Social Security Administration, July 1, 1977

Rubin, Donald B (1976a) "Multivariate Matching Methods that are Equal Percent Bias Reducing, I: Some Examples" *Biometrics*, 32(1): 109-120

Rubin, Donald B (1976b) "Multivariate Matching Methods that are Equal Percent Bias Reducing, II: Maximums on Bias Reduction for Fixed Sample Sizes" *Biometrics*, 32(1): 121-132

Rubin, Donald B (1974) "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies" *Journal of Educational Psychology*, 66(5): 688-701

Rubin, Donald B, Neal Thomas (1996) "Matching Using Estimated Propensity Scores: Relating Theory to Practice" *Biometrika*, 52(1): 249-264

Schafer, Joseph L (1997) Analysis of Incomplete Multivariate Data, Chapman & Hall

Shaffer, Julie P (1995) "Multiple Hypothesis Testing" *Annual Review of Psychology* 46:561-84

Analyzing the results using Rubin's Causal Model (Part II)

Peer effects and smoking roommates at Harvard College

ABSTRACT. With this document, we conclude our demonstration of how Rubin's Causal Model (RCM) can be used to draw causal inferences in a two-step procedure. In the first step, we designed a study to evaluate if Harvard freshmen were more prone to start smoking when sharing a suite with at least one smoker than they would have been when sharing a suite with only non-smokers. Treated students were matched with control students, and models for the outcome analyses were specified. In this second step, we fit these models and evaluate the treatment effects. We also discuss how robust the effects are to various assumptions, as demonstrated by the variation in the effects across the different models. Our main result is that our effect of treatment is small and insignificant when we fit our statistical models on a well-balanced study. Also, this result is robust to the assumptions we make both with regard to the missing potential outcomes and to the various covariate adjustments. Our secondary result is that we would have found peer effects had we instead fitted a model on a less balanced sample, as has been done previously in the peer effect literature, using the traditional approach of causal inferences. However, this secondary result is not robust to the covariate adjustments we make. This exercise illustrates that it is difficult to replicate the results we find when we evaluate peer effects using a well-balanced study (RCM) when we evaluate peer effects using a less-balanced study (traditional approach). The result is reminiscent of the classic results of LaLonde (1986).

1. Introduction

In a sequence of two documents, we evaluate if Harvard freshmen are more likely to begin smoking when they share a suite with at least one smoker (observed potential outcomes) than they would have been when they shared a suite with only non-smokers (missing potential outcomes). This comparison of potential outcomes represents the causal effect of sharing a suite with at least one smoker.

[0] This research would not have been possible without the support and encouragement of my adviser Donald Rubin, who provided many insights and gave detailed feedback on all aspects of this work. It has also benefited from discussions with and comments by my adviser Magnus Johannesson. Jim Greiner and Björn Persson also contributed helpful comments. All remaining errors are mine. I would like to thank Justin Ashmall and Peter Brown for their technical expertise in running the freshman study on-line. Also, I am grateful to the Marcus Wallenberg Foundation for Advanced Education in International Industrial Entrepreneurship and the Apoteksbolagets Fond för Forskning och Studier i Hälsoekonomi och Socialfarmaci for financial support.

We use a two-step procedure for causal inference that is based on Rubin's Causal Model. We did not use any outcome variables in the first step, described in our previous document, when we used the model for the assignment mechanism to design a study in which the treated students are similar to, or balance, the control students. We defined different statistical models to use in order to evaluate peer effects and their sensitivity to various assumptions made with regard to covariate adjustments and missing potential outcomes. The design of the study consequently could not be done in a way that would enable us to capitalize on particular random variation in our sample's outcome variables in order to demonstrate certain effects. In the design, we also discussed which magnitudes of peer effects we could interpret as real and important, i.e., our benchmarks.

In the second step, described in this document, we need not only the potential outcomes that are observed for Harvard freshmen but also the potential outcomes that are missing for them before we can draw our causal inferences. We use the Bayesian approach to impute these outcomes in the RCM framework, because these outcomes are then explicitly derived, not implicitly defined as in the frequentist's approach. As a consequence, the assumptions made with regard to the missing potential outcomes become more transparent in the Baysian than in the frequentist's approach. In addition, the Baysian approach also allows us to use more flexible estimators and to better handle data that are missing for other reasons than would the frequentist's approach.

Then, we fit these statistical models using our well-balanced designed study. We report the resulting estimates of peer effects that we obtain on our first attempt and discuss how robust the peer effects are to various assumptions related to missing potential outcomes and covariate adjusmtents. Also, we fit statistical models that are similar to models previously used in the literature to evaluate peer effects and report the estimates that we achieve not only when we fit these models at our first attempt but also when we re-fit them repeatedly.

In broad outline, we do not find any peer effects when we use the two-step approach based on RCM. This result is robust to the assumptions that we make with regard to covariate adjustments and missing potential outcomes. We do find peer effects, however, when we use the traditional approach previously used in the peer effect literature. This result is not robust to the assumptions we make with regard to the covariate adjustments and it is unclear if it is robust against the assumptions we make with regard to the missing potential outcomes. Consequently, we find that it is hard to replicate

the result we obtain when we use a well-balanced study in the RCM approach when we instead use the less well-balanced study in the traditional approach. This finding is reminiscent of the classic results of LaLonde (1986).

The different assumptions that the Bayesian and the frequentist make with regard to the missing potential outcomes are discussed in Section 2. In Section 3, we describe the preparations that are needed before peer effects can be evaluated. Peer effects can then be evaluated in Section 4 according to already specified models. Previous models to fit peer effects are also used to evaluate peer effects in Section 5. Section 6 concludes this document.

2. Theory

The causal effect of treatment for a group of treated students involves the comparison of the observed potential outcomes under the actual treated condition and the missing potential outcomes under the alternative control condition. Because only the observed potential outcomes are revealed by the assignment mechanism, the missing potential outcomes must either be derived according to the Baysian perspective or defined according to the frequentist's perspective. These outcomes are stochastic in the Bayesian approach, whereas they are assumed to be fixed in the frequentist's approach, i.e., they are given a probability distribution in the Bayesian approach but not in the frequentist's approach. Consequently, the methods used in these two perspectives for causal inference differ.

2.1. The frequentist approach to causal inference. Fisher developed a summary, known as a p-value, to test the hypothesis that the treatment effect is the same for all subjects (Rubin 1980). In order to obtain Fisher's p-value, the treatment effects under all possible randomizations of treatments must first be obtained given that the hypothesis is true. Then, the p-value can be calculated as the ratio of the number of those estimated treatment effects that are more extreme than or equally extreme as the observed estimated treatment effect to the number of possible randomizations of treatments. The closer this p-value is to 0.5, the more likely it is that the hypothesis is true. Fisher's p-values can also be used to derive an interval of possible treatment effects by evaluating hypotheses of varying treatment effects. Those hypotheses with p-values between 0.025 and 0.975 are typically used to indicate possible treatment effects because they exclude those hypotheses that result in the 2.5% most extreme treatment effects in either direction.

Neyman developed a method for testing the hypothesis that the average treatment

effect for the subjects corresponds to a certain value (Rubin 1990); the average is taken across all possible randomizations of treatments. Neyman demonstrated that this process corresponds to the usually estimated average treatment effect, but that the usually estimated confidence interval is upwardly biased unless the treatment effect is additive.

The Fisher-Neyman approaches required that the treatments were randomly assigned subjects because otherwise the potential outcomes that were revealed, and that would subsequently be used in the evaluation of Fisher-Neyman's hypotheses, might not only reflect differences in treatments but also differences in covariate values and potential outcomes. Rubin (1978) showed, however, that the process that determines which potential outcomes are revealed in such non-random studies, i.e., the assignment mechanism, can be modeled. The outcomes that are revealed given this explicit model for the assignment mechanism can then be regarded as random and, consequently, the Fisher-Neyman approaches to estimate the causal effect of treatment can be pursued.

The assignment mechanism can be modeled by means of the propensity scores when the assignment is strongly ignorable (Rosenbaum and Rubin 1983). The propensity score is the probability of treatment given the observed covariates on which the assignment mechanism is based. The potential outcomes that are revealed given the propensity scores then can be regarded as random as long as the observed covariates included in the propensity score estimation can fully account for the assignment mechanism, i.e., the assignment mechanism is strongly ignorable given these covariates. Conditioning the observed potential outcomes on the propensity scores thus can be perceived as a tool by which to replicate the virtue of a randomized experiment. The model for the assignment mechanism is more carefully described in Section 2.2. It is part of a framework known as Rubin's Causal Model.

In the Fisher-Neyman approaches, the covariates and the potential outcomes in all treatment conditions are assumed to be fixed for all subjects constituting the population[1]. Therefore, we need not rely on assumptions with regard to the relationship between the potential outcomes and the covariate values, i.e., the science, when we draw causal inferences according to the frequentist's approach.

In these approaches, however, we are restricted in terms of the hypotheses that we

[1] However, the covariates and the potential outcomes that are observed for the subjects that constitute our sample are stochastic, because a stochastic mechanism determines which subjects to include in the sample and which potential outcomes to reveal for them.

can test, i.e., the Fisher-Neyman hypotheses. Moreover, we do not have an obvious way to incorporate information we deem relevant for deriving the likely values of quantities that are missing, not because they are not revealed by the assignment mechanism, but because the students choose not to respond to certain questions or to drop out of the study prematurely. There are no such restrictions in the Bayesian approach. For these reasons, we prefer the Bayesian approach for causal inference.

2.2. The Bayesian approach to causal inference. If we derive the posterior predictive distributions for the potential outcomes that are missing because they are not revealed by the assignment mechanism, we can draw the missing potential outocomes from these distributions. Then, any possible estimand for the treatment effect can be derived. Furthermore, if we derive these posterior predictive distributions according to the framework offered by the Rubin's Causal Model (RCM), these estimands are often robustly estimated. A posterior predictive distribution of the potential outcomes in RCM combines a model for the science with a model for the assignment mechanism (Rubin 2005).

The science represents the information that we will use for our inferences, i.e., the theoretical relationship between the outcome variables and their explanatory variables. We can model this information by the marginal distribution of the potential outcomes $\mathbf{y_1}$, $\mathbf{y_0}$ and those covariates $\mathbf{X}$ that describe our choice of population, i.e., $Pr(\mathbf{y}_1, \mathbf{y}_0, \mathbf{X} \mid \boldsymbol{\pi}_1, \boldsymbol{\pi}_2)$ where $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ represent unknown vectors of parameters with prior distributions. Thus, the model of the science consists of the conditional distribution of the potential outcomes given the covariates and the marginal distribution of the covariates, i.e., $Pr(\mathbf{y}_1, \mathbf{y}_0 \mid \mathbf{X}, \boldsymbol{\pi}_1) Pr(\mathbf{X}|\pi_2)$. It is challenging to posit $Pr(\mathbf{y}_1, \mathbf{y}_0 \mid \mathbf{X}, \boldsymbol{\pi}_1)$, because we need to model natural processes over which we have little control. However, it is easier to formulate $Pr(\mathbf{X}|\pi_2)$, because we can choose a population for which this distribution is known. Thus, the science is only partly known to us, and the students that are sampled from the population and the potential outcomes that are revealed for them are determined by the assignment mechanism.

The standard procedure is to draw inferences first for a sample and then, by means of additional assumptions, to generalize these inferences for a population. In this paper, we only draw inferences for the students in our sample because, by not making the additional assumptions required to generalize our results, we hope to achieve robust estimates. Consequently, the assignment mechanism does not need to describe how the students are sampled from a particular population. It only needs to describe which of their potential outcomes are revealed, i.e., the probability of a vector of treatments $\mathbf{w}$

conditional on $\mathbf{y}_1$, $\mathbf{y}_0$, $\mathbf{X}$, $\boldsymbol{\pi}_3$, i.e., $Pr(\mathbf{w} \mid \mathbf{X}, \mathbf{y}_1, \mathbf{y}_0, \boldsymbol{\pi}_3)$ where $\boldsymbol{\pi}_3$ represents unknown vectors of parameters with prior distributions. Once we have posited the models for the science and for the assignment mechanism, the posterior predictive distribution of the missing outcomes can be derived:

(2.1)

$$Pr(\mathbf{y}_{mis} \mid \mathbf{X}, \mathbf{y}_{obs}, \mathbf{w}) = \int Pr(\mathbf{y}_1, \mathbf{y}_0, \mathbf{X} \mid \boldsymbol{\pi}_1, \boldsymbol{\pi}_2) Pr(\mathbf{w} \mid \mathbf{X}, \mathbf{y}_1, \mathbf{y}_0, \boldsymbol{\pi}_3) d\mathbf{y}_{mis} d\boldsymbol{\pi}_1, d\boldsymbol{\pi}_2, d\boldsymbol{\pi}_3.$$

This model can be simplified substantially when the probability of the assignments $\mathbf{w}$ is the same independently of the missing potential outcomes $\mathbf{y}_{mis}$, and $\boldsymbol{\pi}_3$ is a priori independent of $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$, because the distribution of the assignment mechanism will then integrate to a constant. The assignment mechanism is then known as ignorable (Rubin 1978), and the posterior predictive distribution is defined by the science:

$$Pr(\mathbf{y}_{mis} \mid \mathbf{X}, \mathbf{y}_{obs}, \mathbf{w}) = \int Pr(\mathbf{y}_1, \mathbf{y}_0, \mathbf{X} \mid \boldsymbol{\pi}_i, \boldsymbol{\pi}_2) d\boldsymbol{y}_{mis} d\boldsymbol{\pi}_1, d\boldsymbol{\pi}_2. \tag{2.2}$$

If the treated and the control students are very different with respect to $\mathbf{X}$, the science may need to be modeled. Otherwise, there are often good chances that the treated students can be matched with control students so that the differences in $\mathbf{X}$ can be balanced. Then, the science does not need to be modeled for other purposes than to improve the precision of the estimates. The treatment effect can be estimated non-parametrically using the Fisher-Neyman perspective, rather than the Bayesian perspective, by comparing the treated students directly with the matched control students.

In the previous paper, we showed that the assignment mechanism used by the Harvard Housing Office (HHO) to assign students into suites can be assumed to be ignorable, and that the model of the assignment mechanism can be used to match the treated students with similar control students. The treatment effect can thus be estimated non-parametrically in this paper, that is by comparing the treated students with their matched control students and avoiding the modeling of the science.

Thus far, we have assumed that there are no missing values other than those potential outcomes that are missing because they are not revealed by the assignment mechanism. There are also values, however, that are missing because the students choose not to answer particular questions or to drop out of the study. From the Bayesian perspective these values, like the values that are missing because the assignment mechanism does not reveal them, can be drawn from their posterior predictive distributions. We do not posit the models to do this explicitly, however, because the burden to posit explicit models for these outcomes and covariates that fit the data satisfactory can be

overwhelming. Complications may arise because there may be many models that need to be posited. The understanding of these models may be limited, because they may differ in their forms, i.e., continuous, binary, count, or mixed, and/or because the models may be exposed to different restrictions and consistency bounds. For these reasons, sequences of regression models were used to impute their missing values (Raghunathan 2001).

We assume that the matrix of covariates $\boldsymbol{X}$ represents the proper covariates, i.e., covariates unaffected by treatment. These covariates are fully observed at this stage, because their missing values are imputed in the design phase of our suggested two-step procedure, not in this analysis phase. Furthermore, we assume that $\boldsymbol{y}$ represents the outcome variables and the covariates possibly affected by treatment, i.e., improper covariates. In the first sequence of regressions, we regress the observed units of the outcome variable with the fewest missing values $\mathbf{y}^1$ on the fully observed $\mathbf{X}$. The resulting posterior predictive distribution given the fully observed $\mathbf{X}$ is then used to complete $\mathbf{y}^1$, i.e., the missing values of $\mathbf{y}^1$ are drawn from this distribution. The outcome variable with the next fewest missing values $\mathbf{y}^2$ is then regressed on the fully observed $\mathbf{X}$ and completed $\mathbf{y}^1$, and the resulting posterior predictive distribution given the fully observed $\boldsymbol{X}$ and the observed values of $\boldsymbol{y}^1$ is used to complete $\mathbf{y}^2$. Repeating this procedure for the rest of the outcome variables $\boldsymbol{y}^3, ..., \boldsymbol{y}^o$ results in their joint density conditional on the fully observed $\boldsymbol{X}$, and the observed values of $\boldsymbol{y}^1, ..., \boldsymbol{y}^o$:

$$Pr(\mathbf{y}^1_{mis}, ..., \mathbf{y}^o_{mis} \mid \mathbf{y}^1_{obs}, ..., \mathbf{y}^o_{obs}, \mathbf{X}, \boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_o) =$$
$$\int Pr(\mathbf{y}^1_{mis} \mid \mathbf{y}^1_{obs}, \mathbf{X}, \boldsymbol{\theta}_1), ..., Pr(\mathbf{y}^o_{mis} \mid \mathbf{y}^1_{obs}, ..., \mathbf{y}^o_{obs}, \mathbf{X}, \boldsymbol{\theta}_o) d\boldsymbol{\theta}_1, ..., d\boldsymbol{\theta}_o,$$

where $\boldsymbol{\theta}_1$,...,$\boldsymbol{\theta}_o$ represent the vectors of parameters with prior distributions that govern the posterior predictive distributions. In the next sequence of regression models, the missing values of the outcomes are again drawn from their posterior predictive distributions, but these distributions are now conditioned on the fully observed X and all the observed values of $\boldsymbol{y}^1, ..., \boldsymbol{y}^o$, i.e., the observed values of each outcome is regressed on the fully observed $\mathbf{X}$ and completed $\boldsymbol{y}^1, ..., \boldsymbol{y}^o$. Thereby, the complex interdependencies among the outcomes can be modeled. The purpose of any subsequent sequence of regressions is to achieve convergence in the distribution of the imputed values.

In this document, we will use the software IVE-ware (www.isr.umich.edu/src/smp/ive) that follows these principles for drawing the missing outcomes from their posterior predictive distribution. Because the outcome variables possibly are affected by treatment, we use separate models for the posterior predictive distributions of the treated and the

control students. When the missing values of the proper covariates $\boldsymbol{X}$ were imputed in the design phase, they were also drawn from their posterior predictive distributions. These distributions were not different for the treated and the control students, however, because proper covariates, by definition, cannot be influenced by treatment.

2.3. The contribution of Rubin's Causal Model to causal inference. We choose the Bayesian approach for causal inference, because this approach allows us to assess more hypotheses and to better handle such complications as missing data than does the frequentist approach. Furthermore, we use RCM for causal inference because its two-step procedure offers some advantages over the traditional one-step procedure. The first step in RCM is to design a study without using outcome variables, i.e., to match the treated students with similar control students and to define the statistical models for evaluating peer effects (design phase). The RCM's next step is to fit the statistical model for evaluating peer effects at the first attempt (analysis phase). In the traditional approach, however, the study is designed at the same time as peer effects are evaluated.

Separating the design of the study from the outcome analyses means that we can re-design the study until we find a design that allows us not only to restrict the number of assumptions in the outcome analyses but also to fit the models for the outcome analyses at our first attempt. Without this separation, we are not allowed to re-design the study. We would then have to fit the models for the outcome analyses under restrictive assumptions if the design did not fit the data well. Or, we would be forced to re-fit the models for the outcome analyses if the design did not fit the data satisfactorily, and we would then inevitably use the information just obtained about the science to find a better design. The results under RCM are thus more robust to assumptions because the assumptions are less restrictive, and because the attempts to fit the model are often fewer the results become more "honest".

In addition, the assumptions with regard to the missing potential outcomes are more transparent in the RCM, because these outcomes can be derived explicitly. The fact that the RCM models two distinct processes separately – the design of the study and the evaluation of the results – also means that the implications of the assumptions on these processes become more transparent.

We present the design and the outcome analyses in two separate documents because we want to highlight that the design and the outcome analyses should be handled separately. Moreover, we do not use any outcomes for the design, only for the analyses.

This is a distinction that should be made more often by researchers, perhaps by publishing the design of the study in an on-line database before the analysis is undertaken. If the subsequently estimated results are also included in this database, and perhaps also references to the journal in which the results are published, this database would be a valuable tool for building an unbiased understanding of our economic environment. The database would also help defend the field against accusations that only significant results are published.

3. Preparation

In the previous paper, we defined the treated students whose effect of treatment we want to estimate. We know their six-month smoking habits when assigned the treated condition (observed potential outcomes) but lack their six-month smoking habits when assigned the control condition (missing potential outcomes). Therefore, we identified the control students who would help us draw inferences about the treated students' missing potential outcomes; these control students were found by matching each of the treated students with one control student. In this section, we describe what remains to be done before we can estimate the final outcome analyses using these students. That is, we need to impute the missing outcomes for these students, and we need to refine the benchmarks for which treatment effects we can regard as real and important.

3.1. Description of outcomes. The outcomes collected in the third interview focus mainly on the students' health habits. Separate questions were asked about the students' exercise habits and about their use of cigarettes, cigars, pipes, chewing tobacco, and drugs during the month prior to the interview. The students reported their use on a four-graded scale, i.e., they experienced the habits not at all (grade=1), one or two days/month (grade=2), several days/month (grade=3), or every day (grade=4) and, because we believe that the importance of the effect of a marginal change in the lower range is more important than the effect of a marginal change in the upper range, the four-graded scale was translated into log. Questions were also asked about the extent of these habits, i.e., the length of time they involved in physical exercise and the quantity they used of alcohol, cigarettes, cigars, pipes, chewing tobacco, snuff, or drugs. The answers to these questions were also translated into logs. However, these outcomes were not used for any other purpose than to impute the missing outcomes.

3.2. Imputation of missing outcomes. We have already imputed the missing values for the proper covariates, because we needed the proper covariates to design our study. For these covariates, the models for the treated students' missing values were the same as the models for the control students' missing values, because proper

covariates by definition are not affected by treatment. We imputed ten values for each missing value in order to account for the uncertainty with which the missing values were imputed. This idea was originally proposed (1977, 1978) and further developed (1987) by Rubin and subsequently by others, e.g., Schafer (1997).

The resulting ten multiply-completed datasets must now be completed with regard to the improper covariates and the outcomes that, until this point, have not even been seen. The models for these improper covariates and outcomes should be allowed to differ for the treated and the control students respectively because, contrary to the proper covariates, the improper covariates and the outcomes can be affected by treatment. Therefore, we separately impute the missing values of the improper covariates and of the outcomes for the treated students (46) and the control students (56)[2]. Only one imputed value is needed for each missing value in the ten multiply-completed datasets, because the ten datasets will give rise to the variations we need to account for imputation uncertainty. The subsequent estimations of the treatment effects in these ten datasets and the aggregation of their estimates will result in an aggregate estimate that considers imputation uncertainty.

In these imputations, we cannot include too many covariates or covariate transformations, because the models for the missing values will only accommodate those covariates and transformations that are most predictive of the observed outcomes. If we include too few covariates or transformations, however, we may fail to account for all the interdependencies between the covariates. We use all covariates that are collected in the first, second, or third interview. However, we do not use indicators for which values are missing for each covariate. If we would have had the opportunity to remake this imputation, we probably would have included such indicators as well, because missing values may be nonignorable.

Finally, we translate the imputed values to values that actually are observed in the collected dataset. This ensures that we do not create any outliers by imputation and that we estimate results that can be interpreted.

We now need to evaluate the causal effect of treatment for our sample of treated students and matched control students. In the previous paper, covariate adjustments were found to be unnecessary, because the treated students and the control students

[2] There are more control students than treated students, because the treated students are matched with different controls in the ten multiply-completed datasets.

were balanced. Covariate adjustments were therefore not specified for the primary analyses. For the secondary analyses, however, they were specified in order to improve the precision of the estimates and to evaluate the sensitivity of the estimates to various covariate adjustments. The primary and secondary approaches are estimated both according to the Neyman approach and to the Bayesian approach. Although the Bayesian approach is our preferred choice, the Neyman approach is added to evaluate the sensitivity of the estimates to the assumptions we make with regard to the missing potential outcomes. In the Neyman approach, we derive the average treatment effect by comparing the treated and the control students' outcomes both when we control for the matched pairs and when we do not, i.e., a paired and an unpaired analysis. In the Bayesian approach, we draw the missing potential outcomes from their posterior predictive distributions and then multiply impute them.

3.3. Refinement of benchmarks. In the previous paper, we found that the students' recall of their pre-college alcohol consumption was possibly affected by treatment, i.e., students recalled a higher pre-college alcohol consumption if they shared a suite with at least one smoker than they would have recalled if they had instead shared a suite with only non-smokers. Therefore, we concluded that the students' six-month smoking habits should be more affected by treatment than the students' recall of their alcohol consumption. Otherwise, it may be the students' perceptions of their behaviors that is influenced by treatment rather than their behaviors.

We derived the benchmarks in our previous document by estimating the effects that the treatment has on 23 improper covariates. The observations with missing values on these improper covariates had to be dropped from these estimations, because these missing values had not yet been imputed. If these values are not missing at random, our benchmarks will be biased. For that reason we re-estimate the benchmarks now, after having imputed the missing values.

We re-estimate the treatment effects on these 23 improper covariates according to the Neyman (paired and unpaired) and the Bayesian perspectives. First, we do not adjust for any covariates (Tables 1, 2, and 3). Again, we find a "significant" treatment effect for one improper covariate, i.e., the students' recall of their pre-college consumption of pipe or cigars[3]. The p-values vary between 0.08 and 0.09. For this improper covariate, we then adjust for the most predictive covariates among a pool of six covariates (Table 4) and among a pool of 209 covariates (Table 5). The p-values now vary in the range

[3] The covariate that was previously significant, i.e., the students' recall of the pre-college alcohol consumption, is no longer "significant".

of 0.08 and 0.11 and of 0.03 and 0.07, respectively.

This practice of outcome analyses not only helps us understand which benchmark we should use for each of the final analyses that we have specified, but also indicates that whatever approach we take, i.e., the Neyman or the Bayesian approach, the estimates differ little when we use our matched data.

4. Analysis

We need to evaluate the prospects of the sample to demonstrate significant results before running the final analyses. If the prospects are good, small and insignificant results may be interpreted as a lack of effect. After the final analyses are run, we summarize what we have learned about peer effects. Although significant results help us build our understanding of the addressed research question, insignificant results can help us better formulate research questions in future work. Also, we should assess whether there is anything about our choice of data or our methods that may invalidate our results.

4.1. Final outcome analyses. The prospects of demonstrating significant peer effects relating to smoking can be evaluated from two perspectives. First, we can evaluate which sample size we need in order to demonstrate significant peer effects given some prior understanding of the relationship between the effects' magnitudes and variances, and the sample size. We believe, for example, that peer effects are at least as large as, and have no larger variation than, the effects of *in utero* exposure to phenobarbital on adult men's intelligence. Given these priors, which we regard as conservative, we can translate the results of Reinisch et al's study (1995) to different sample sizes. We can then conclude that our sample size, i.e., 46 treated and 46 control students, would belong to the range of sample sizes that should be able to demonstrate significant treatment effects[4].

Once the sample size is fixed, we should also have greater prospects of finding significant effects for the hypothesis for which the study is designed than for other hypotheses. That is, having found treatment effects on improper covariates (Section 3.3) we should be able to find treatment effects on cigarette smoking, because our study was not designed to study the former effects, only the latter.

[4] Their adjusted t statistic was -2.92, and our t statistic would be $-2.92/\sqrt{2}$ given the assumption that their effect and ours are of similar magnitude and variation.

The results of the final estimations of the treatment effects on peer effects are consistent across the primary and secondary analyses and across approaches, i.e., the Neyman and the Bayesian approaches: the magnitudes of the effects are close to zero, the effects are not even close to significant, and the significance level achieved is considerably lower than the significance level for the effects on pre-college consumption of pipe or cigars (Tables 6, 7, and 8). These findings leads us to believe that there may not be any treatment effects relating to roommates' smoking habits in our sample, i.e., the students do not seem to be more prone to begin smoking after six months in a suite with at least one smoker than they would have been in a suite with only non-smokers.

4.2. The Contribution of the Work. Even though we were not able to reject the hypothesis of no peer effects, new hypotheses may evolve from this exercise. And these new hypotheses may bring the research one step forward.

It is possible that insignificant and small effects are demonstrated because the Harvard freshmen are resistent to their roommates' smoking habits but freshmen at other colleges are not. Harvard freshmen may be more resistant because they may come from higher socioeconomic background, may have better self-confidence, or may even be more intelligent. If Harvard freshmen are more resistant to peer effects, it could be worthwhile to evaluate the same hypothesis of peer effects at other colleges.

Also, it is possible that college freshmen have passed the age at which they are influenced by their roommates' smoking habits. It could then be worthwhile to evaluate the hypothesis of peer effects at a younger age.

Finally, the freshmen may be influenced by their peers' smoking but not necessarily the peers that their roommates constitute. Unfortunately, it is hard to evaluate the hypothesis of other peer effects, because they cannot be evaluated by relying on variation in treatments created by the assignment mechanism. For example, the Harvard freshmen cannot be "assigned" their friends. Instead, it is necessary to rely on variation in treatments created by the science itself and all the strong assumptions that follow such an identification strategy.

4.3. The validity of our results. It is possible that our results are not valid because we may have used inappropriate methods to evaluate peer effects relating to smoking. One could, for example, argue that we should have evaluated peer effects on the bedroom level instead of peer effects on the suite level, because freshmen may interact more with those roommates with whom they share a bedroom. However, 39%

of all those students who participated in the second interview stated that they planned to rotate rooms during the semester, and 80% of those who participated in the third interview actually did so. Because there seems to be no clear indication that the students spend more time with any particular roommate, we do not think that our results would have changed had we instead studied peer effects on a bedroom level[5].

One might also argue that the reason why we do not detect peer effects is that freshmen may be unfamiliar with their roommates' smoking habits. This may be the case, because smoking is prohibited in the dormitories. In this case, however, we would not expect to find a treatment effect for any covariate. The indication of a treatment effect that we found on some improper covariates speaks against this claim.

Further peer effects may take longer to develop than six months and if so, we might have found different results had we followed the students longer. Although we may never know if this is a valid argument, we believe that this is unlikely. The students should be especially vulnerable to their roommates' smoking habits when they are not yet established in a new environment.

Finally, one might argue that the students were asked to describe their latest month's smoking habits at a time when they were preparing for their first-year exams. At such a time, they may be exposed to fewer incentives to smoke than at other times, e.g., at times when parties are thrown etc. On the contrary, however, we believe that students who begin to smoke before an exam period because of their roommates' influences will continue to smoke during this period because of the same influence; the roommates may smoke even more during an exam period because the may manage stress by smoking.

5. "Hunting for results"

The effects of the smoking habits of those peers that the roommates constitute are shown to be small and non-significant in this study. In previous studies, however, the effects of the smoking habits of other peers have been large and significant. Our results could differ from previous results because the peers whose effects are evaluated differ, or because the methods for evaluation differ.

In our method, we first define an ideal hypothetical randomized experiment. Then we replicate this experiment as closely as possible for our quasi-experimental data by

[5] We did evaluate the prospect to conduct such an analysis, but because only 19 treated students would be available for such an analysis we decided against it.

designing a study. The study is designed without using outcome data by matching treated students with similar control students by means of their propensity scores and by defining the outcome analyses to use when testing for peer effects. Finally we evaluate peer effects. Our approach allows us to handle missing data in a consistent way.

Now, we examine the extent to which previous studies define a template experiment, separate the design of the study from the analysis of the results, match treated students with similar control students, and treat missing data in a consistent way.

5.1. Previous analyses of smoking peer effects. Most of the previous studies could be said to rely on variations that are created by the assignment mechanism to identify peer effects (Gaviria 2001; Norton 1998; Cheryl 2001 etc.) but some studies also rely heavily on variations that are created by the science (Krauth 2005 etc.). We will focus on the approaches of the former class of studies, because they intend to use the same variations as we do to identify peer effects. Also, we will not discuss all peer studies in this class, only a small selection of representative studies.

A template experiment is not discussed in Gaviria's, Norton's, or Cheryl's studies. Given the peers they are studying and the instruments they are using to create exogenous variations in the treatments, their template experiments would be to randomize 10th graders into different schools (Gaviria et al 2001), to randomize students into different neighborhoods (Norton et al 1998), or to randomly allocate 7th to 12th graders to friends (Cheryl et al 2001). The template of Gaviria would be feasible and peer effects could possibly be separated from other confounding effects such as neighborhood effects. Norton's template would also be feasible, but would not separate peer effects from neighborhood effects. Finally, Cheryl's template is not feasible and, consequently, the effects that would be identified are unclear.

Furthermore, none of these three studies separates the design of the study from the analyses of the results. They must fit their statistical model at the first attempt, because any models that they fit subsequently will inevitably incorporate the understanding about the relationship between outcomes and peer exposure gained in previous attempts to find a model that fits the data better. Because it is hard to posit a model that fits the data satisfactorily at the first attempt, it is likely that Gaviria, Norton, and Cheryl had to re-fit their statistical models. They could then capitalize on random variations in a particular sample in order to demonstrate peer effects of certain magnitudes and significances.

Also, Gaviria, Norton, and Cheryl rely on purely observational data, not data from quasi-randomized experiments as we do. Hence, their assignment mechanisms probably involve considerably more covariates than our assignment mechanism does. Considering the difficulty we had in balancing the treated and the control students when we matched them based on the few covariates that were included in our propensity score estimation, it is not likely that Gaviria, Norton, and Cheryl managed to balance them given that they did not use matching and did not take advantage of the propensity scores to control for many covariates.

Finally, missing data is not addressed at all in these three studies. Students with missing data on any of the relevant outcomes or covariates are simply dropped from the analyses. We believe that this omission can have considerable consequences for the results, because the students who are missing these values could differ from the students who are not missing them.

We will now compare what will happen to our results when we follow the same approach as Gaviria, Norton, and Cheryl, i.e., when we do not separate the design of the study from the analyses of outcome data, when we do not match treated students with similar control students, when we do not control for the assignment mechanism appropriately by means of the propensity scores, and when we do not impute missing values.

5.2. Replicating the methods of previous analyses. We use the same sample that we found to be the best balanced sample in the previous paper, but we forgo the matching of the treated and the control students. Furthermore, we forgo the students with missing values, i.e., students who did not report the proper covariates or the outcomes needed for the outcome analyses or who dropped out from the study prematurely. Then, we run an analysis that is comparable to the analyses previously made in the literature, i.e., we regress the students' six-month smoking habits (smoking or not smoking) on the five covariates that we previously defined as predictive of students' six-month smoking habits, i.e., students' preference for a lively atmosphere, students' religious and physical activities, and parents' educational level. We also control for the students who drop out from the study prematurely by including the inverse Mills ratio as an additional covariate. However, we do not control for the assignment mechanism by means of the propensity score because the assignment mechanisms were not adequately controlled for in previous studies.

The results are diametrically opposed to the results we obtained after first designing our study and then estimating the effects. The result indicates that a freshman who does not smoke is 8 percentage unit more likely to begin smoking when he shares a suite with at least one smoker than he would have been if he had instead shared a suite with only non-smokers. Moreover, the results are significant at the 10 percent level (Table 9). Furthermore, the magnitude of the treatment effects can be increased and the significance level decreased by including "the right" covariates in addition to the inverse Mills ratio (Table 10). Two "right" covariates are needed to achieve a treatment effect of 9 percentage units that is significant at 5 percent; an additional "eight" right covariates are needed for a treatment effect of 12 percentage units to be significant at 1 percent. In fact, if there is no restriction on the number and the kind of covariates that can be included, many different effects can be demonstrated and the results are all significant at the 1 percent level.

This exercise has shown that the methods chosen to identify peer effects can be decisive for whether or not peer effects are claimed to be found. Significant "results" are demonstrated when previous methods are followed, i.e., when the researcher does not design the study before the outcomes are analyzed and, consequently, may need to experiment with the relationship between the outcomes and the covariates before the statistical model can satisfactorily fit to the data. Insignificant results, however, are demonstrated when we follow our chosen method, i.e., separating the design of the study from the analysis and, consequently, fitting the statistical model at the first attempt.

6. Conclusion

Before the statistical models for the outcome analyses are fit in this document, we account for the missing potential outcomes explicitly by drawing them from their posterior predictive distribution (the Bayesian's approach) (Section 3.1). Because the outcomes may be influenced by the treatment, these distributions differ for the treated and the control students, respectively (Section 3.2). We also account for the outcomes implicitly by letting them be defined by the hypothesis that the average treatment effect is zero (the frequentist's approach). We can then evaluate how sensitive our results are to various assumptions that we make with regard to the missing outcomes.

We also account for the missing improper covariates. These too are drawn from their posterior predictive distributions, and different distributions are again specified for the

treated and the control students. We can then re-estimate the benchmarks more efficiently, because the students with missing values no longer need to be dropped (Section 3.3).

Then we fit the statistical models for the outcome analyses. The primary analyses are our main concern and they evaluate the treatment effects when no covariate adjustments are made. The missing potential outcomes are derived explicitly, but in order study how sensitive the estimates are to the assumptions about the missing potential outcomes, these outcomes are also defined implicitly. The secondary analyses are posited in order to help us evaluate how sensitive our results are to various covariate adjustments.

We cannot find that Harvard freshmen are more prone to begin smoking when they share a suite with at least one smoker than they would have been had they shared a suite with only non-smokers. The estimates of the treatment effects are close to zero and far from significant. This result is robust against the assumptions we make with regard to the missing potential outcomes and the covariate adjustments (Section 4.1).

This result could be interpreted as a lack of treatment effect on the freshmen's smoking habits, because the estimates are close to zero and because the effects on the outcomes that the study is designed to evaluate should be easier to find than the effects found on other outcomes, e.g., on improper covariates. Of course, this does not mean, however, that peer effects relating to smoking do not exist. Other college freshmen may be influenced by their roommates' smoking habits, because they may be less resistant to peer effects than are Harvard freshmen. Younger students may also be influenced by their roommates' smoking habits, but college freshmen may have become resistant to them. Finally, Harvard freshmen may be influenced by their peers' smoking habits but not necessarily the peers that the roommates constitute. These hypotheses can be evaluated in future research (Section 4.2).

Our results are different from those previously found in the literature. Therefore, we evaluate to what extent the previous results are based on a plausible template experiment, a proper design of the study before the analysis is conducted, and an adequate attention to missing information. We find no such study. Therefore, we decide to analyze our data using the same statistical models as used previously in the literature (Section 5.1).

One such statistical model that we fit at our first attempt demonstrates that a freshman who does not smoke is 8 percentage units more likely to begin smoking after six month in college when he shares a suite with at least one smoker than he would have been if he had instead shared a suite with only non-smokers. The model controls for the five covariates that are believed to be the strongest predictor of six-month smoking habits and the attrition caused by the students that are dropped from the analysis by means of the inverse Mills ratio. Other statistical models that we subsequently fit demonstrate an even larger and more significant treatment effect when we allow the number and the kind of covariates to vary. In fact, a range of possible significant treatment effects can be demonstrated (Section 5.2).

Consequently, we find that if we evaluate peer effects on a sample that is balanced in observed covariates, we obtain estimates of peer effects that are robust to the assumptions that we make with regard to covariate adjustments and missing potential outcomes. However, if we instead evaluate peer effects at the same time as we control for imbalances in observed covariates, we do not obtain estimates that are robust to the assumptions we make with regard to covariate adjustments. Moreover, we cannot evaluate how robust the estimates are to the assumptions we make with regard to the missing potential outcomes, because the missing potential outcomes are only implicitly defined, not explicitly derived. This result is reminiscent of the classic results of LaLonde (1986).

Using a two-step procedure based on RCM to evaluate the causal effect of treatment, we have found in a sequence of two papers that causal inference relies on less restrictive and more transparent assumptions than the traditional one-step approach previously used to evaluate peer effects on cigarette smoking. Because the RCM does not use the outcome variables when the assignment mechanism is modeled, this model can be re-fit repetitively without biasing the model for the science. The model for the assignment mechanism can then often be modeled to fit the data better and, because the model for the science can consequently better control for the assignment mechanism, the model for the science can be fit with less restrictive assumptions. Moreover, because the RCM models two distinct processes separately – the assignment mechanism and the science – the implications of the assumptions on these processes become more transparent. Finally, the RCM can derive the two potential outcomes needed for drawing causal inferences explicitly, which makes the assumptions regarding the missing potential outcomes more transparent.

7. Tables

TABLE 1. The effect of treatment on different improper covariates estimated by deriving the missing potential outcomes only implicitly, i.e., by using OLS regressions. No covariate adjustments are made.

	τ	σ^2	p-value
Smoke a pipe or cigars - unpair	0.10	0.34	0.08
Chewing tobacco or snuff - unpair	0.00	0.00	
Use drugs - unpair	0.02	0.29	0.78
Times consumed alcohol/month - unpair	0.25	4.99	0.26
Drinks/time when consuming alcohol - unpair	0.24	3.55	0.22
≥ 5 drinks/time - unpair	0.15	1.44	0.22
Moderate physical exercise - unpair	0.27	4.05	0.18
Vigorous physical exercise - unpair	0.02	4.87	0.91
Father smoked - unpair	−0.02	0.49	0.83
Mother smoked - unpair	0.04	0.37	0.52
Father consumed alcohol - unpair	−0.06	2.47	0.70
Mother consumed alcohol - unpair	0.08	2.17	0.60
Carefully complete HAF - unpair	−0.09	8.03	0.76
Truthfully complete HAF - unpair	−0.19	3.75	0.33
Religious tolerance - unpair	0.13	1.01	0.19
Help completing HAF - unpair	0.21	6.09	0.41
Ever used alcohol before college - unpair	0.14	0.72	0.10
Ever smoked a cigarette before college - unpair	−0.02	0.48	0.80
Ever smoked cigars before college - unpair	0.06	0.60	0.45
Ever smoked a pipe before college - unpair	0.01	0.12	0.85
Ever used chewing tobacco before college - unpair	0.01	0.13	0.81
Ever used snuff before college - unpair	−0.03	0.07	0.30
Ever used illegal drugs before college - unpair	0.11	0.81	0.21

TABLE 2. The effect of treatment on different improper covariates estimated by deriving the missing potential outcomes only implicitly, i.e., by using OLS regressions. No covariate adjustments are made but indicators for matched pairs are used.

	τ	σ^2	p-value
Smoke a pipe or cigars - pair	0.10	0.35	0.09
Chewing tobacco or snuff - pair	0.00	0.00	
Use drugs - pair	0.02	0.31	0.79
Times consumed alcohol/month - pair	0.25	5.45	0.29
Drinks/time when consuming alcohol - pair	0.24	3.79	0.24
≥ 5 drinks/time - pair	0.15	1.37	0.22
Moderate physical exercise - pair	0.27	4.14	0.19
Vigorous physical exercise - pair	0.02	5.94	0.92
Father smoked - pair	−0.02	0.55	0.84
Mother smoked - pair	0.04	0.33	0.50
Father consumed alcohol - pair	−0.06	2.59	0.71
Mother consumed alcohol - pair	0.08	2.24	0.60
Carefully complete HAF - pair	−0.09	8.08	0.76
Truthfully complete HAF - pair	−0.19	3.21	0.30
Religious tolerance - pair	0.13	1.00	0.19
Help completing HAF - pair	0.21	6.62	0.43
Ever used alcohol before college - pair	0.14	0.81	0.12
Ever smoked a cigarette before college - pair	−0.02	0.47	0.80
Ever smoked cigars before college - pair	0.06	0.70	0.49
Ever smoked a pipe before college - pair	0.01	0.12	0.85
Ever used chewing tobacco before college - pair	0.01	0.12	0.80
Ever used snuff before college - pair	−0.03	0.07	0.30
Ever used illegal drugs before college - pair	0.11	0.65	0.17

TABLE 3. The effect of treatment on different improper covariates estimated by deriving the missing potential outcomes explicitly. No covariate adjustments are made.

	τ	σ^2	p-value
Smoke a pipe or cigars - multiply impute	0.10	0.37	0.09
Chewing tobacco or snuff - multiply impute	0.00	0.00	
Use drugs - multiply impute	0.02	0.33	0.79
Times consumed alcohol/month - multiply impute	0.25	6.34	0.31
Drinks/time when consuming alcohol - multiply impute	0.24	4.51	0.27
≥ 5 drinks/time - multiply impute	0.15	1.64	0.25
Moderate physical exercise - multiply impute	0.27	4.57	0.20
Vigorous physical exercise - multiply impute	0.03	5.39	0.91
Father smoked - multiply impute	−0.02	0.53	0.83
Mother smoked - multiply impute	0.04	0.39	0.53
Father consumed alcohol - multiply impute	−0.06	2.68	0.71
Mother consumed alcohol - multiply impute	0.08	2.30	0.61
Carefully complete HAF - multiply impute	−0.08	10.27	0.80
Truthfully complete HAF - multiply impute	−0.19	4.50	0.37
Religious tolerance - multiply impute	0.13	1.13	0.21
Help completing HAF - multiply impute	0.21	6.61	0.42
Ever used alcohol before college - multiply impute	0.14	0.77	0.11
Ever smoked a cigarette before college - multiply impute	−0.02	0.57	0.82
Ever smoked cigars before college - multiply impute	0.06	0.68	0.47
Ever smoked a pipe before college - multiply impute	0.01	0.14	0.86
Ever used chewing tobacco before college - multiply impute	0.01	0.14	0.82
Ever used snuff before college - multiply impute	−0.03	0.08	0.33
Ever used illegal drugs before college - multiply impute	0.11	0.90	0.23

TABLE 4. The effect of treatment on smoking a pipe or cigars when the stepwise procedure keeps adding the most predictive covariate from a pool of six covariates that are significant at Bonferroni adjusted p-levels, i.e., at $\frac{1}{5(q-p)}$ level where q is the 6 covariates believed to be most predictive of outcome and p is the number of covariates already included in the regression.

	τ	σ^2	p-value
Unpaired OLS-regression	0.10	0.33	0.08
Paired OLS-regression	0.10	0.37	0.11
RCM	0.10	0.36	0.09

TABLE 5. The effect of treatment on smoking a pipe or cigars when the stepwise procedure keeps adding the most predictive covariate from a pool of 209 covariates significant at the 5% level.

	τ	σ^2	p-value
Unpaired OLS-regression	0.11	0.25	0.03
Paired OLS-regression	0.10	0.31	0.07
RCM	0.11	0.25	0.03

TABLE 6. Primary analysis: The effect of treatment on 6-month smoking when no covariate adjustments are made.

	τ	σ^2	p-value
Unpaired OLS-regression	-0.02	0.64	0.83
Paired OLS-regression	-0.02	0.62	0.82
RCM	-0.02	0.78	0.84

TABLE 7. Secondary analysis I: The effect of treatment on 6-month smoking when the stepwise procedure keeps adding the most predictive covariate from a pool of six covariates that are significant at Bonferroni adjusted p-levels, i.e., at the $\frac{1}{5(q-p)}$ level where q is the 6 covariates believed to be most predictive of outcome and p is the number of covariates already included in the regression.

	τ	σ^2	p-value
Unpaired OLS-regression	-0.01	0.63	0.85
Paired OLS-regression	-0.01	0.64	0.87
RCM	-0.01	0.74	0.84

TABLE 8. Secondary analysis II: The effect of treatment on 6-month smoking when the stepwise procedure keeps adding the most predictive covariate from a pool of 209 covariates significant at the 5% level.

	τ	σ^2	p-value
Unpaired OLS-regression	-0.03	0.57	0.70
Paired OLS-regression	-0.03	0.58	0.71
RCM	-0.03	0.73	0.74

TABLE 9. Using a statistical model similar to the models previously used to evaluate peer effects relating to smoking. The model is fitted at the first attempt.

	τ	σ^2	p-value
Intercept	−0.0429	0.2312	0.8529
Peer effects	0.0753	0.0452	0.0974
Preference for lively room atmosphere	0.0001	0.0243	0.9976
Engage in physical activities	0.0324	0.0419	0.4399
Engage in religious activities	−0.0262	0.0440	0.5524
Father education	0.0506	0.0303	0.0972
Mother education	−0.0647	0.0311	0.0388
Mills ratio	0.4583	1.2167	0.7069

TABLE 10. Using a statistical model similar to the models previously used to evaluate peer effects relating to smoking. The model is re-fitted repeatedly with the purpose of estimating the treatment effect as precisely as possible, i.e., one covariate at a time is added and the covariate that contributes to the most precise treatment effect is chosen.

Variable	τ	σ^2	p-value	R^2
Mates*Social science	0.08	0.04	0.06	0.04
Mother edu*Lively	0.09	0.04	0.04	0.05
Father edu^2	0.10	0.04	0.03	0.07
To bed late*Exercise	0.10	0.04	0.02	0.10
Mother edu*Interview	0.10	0.04	0.02	0.10
Mother edu*Fine arts	0.11	0.04	0.02	0.11
Wake up late*Psychology	0.11	0.04	0.02	0.13
Pref. mates*Psychology	0.11	0.04	0.01	0.14
To bed late*Psychology	0.12	0.04	0.01	0.14
Mother edu*To bed late	0.12	0.04	0.01	0.15
Mother edu*Pref. mates	0.12	0.04	0.01	0.15
Wake up late*Sports activities	0.12	0.04	0.01	0.15
Parents edu*Psychology	0.12	0.04	0.01	0.15
Mates*Race	0.12	0.04	0.01	0.16
Race	0.13	0.04	0.00	0.18
Pref. mates*Race	0.13	0.05	0.00	0.18
Mother edu*Disorderly	0.13	0.05	0.00	0.18
Father edu*Social studies	0.13	0.05	0.00	0.18
Disorderly*Social studies	0.14	0.05	0.00	0.20
Disorderly*Psychology	0.14	0.05	0.00	0.20
Disorderly*Religious activities	0.14	0.05	0.00	0.20
Religious activities	0.15	0.05	0.00	0.22
Wake up late*Indep. music+	0.15	0.05	0.00	0.22
To bed late*Indep. music	0.15	0.05	0.00	0.22
Father edu	0.15	0.05	0.00	0.23
Parents edu*Religious activities	0.15	0.05	0.00	0.23
Pref. mates*Religious activities	0.15	0.05	0.00	0.23
Go to bed late*Male	0.15	0.05	0.00	0.24
Disorderly*Male	0.16	0.05	0.00	0.26
Wake up late*Male	0.16	0.05	0.00	0.26

References

Angrist, Joshua D, Imbens, Guido W and Rubin Donald B (1996) "Identification of Causal Effects Using Instrumental Variables" (with discussion) *Journal of the American Statistical Association* 91:444-472

Cheryl, Alexander, Marina Piazza, Debra Mekos, Thomas Valente (2001) "Peers, Schools, and Adolescent Cigarette Smoking"

Gaviria, Alejandro, Steven Raphael (2001) "School-based Peer Effects and Juvenile Behavior" *The Review of Economics and Statistics*, 83(2):257-268

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

Krauth, Brian V (2005) "Peer effects and selection effects on smoking among Canadian youth" *Canadian Journal of Eonomics*,38(3):735-757

LaLonde, Robert J (1986) "Evaluating the Econometric Evaluations of Training Programs with Experimental Data" *American Economic Review*, 76:604-620

Norton, Edward C, Richard C Lindrooth, Susan T Ennett (1998) "Controlling for the Endogeneity of Peer Substance use on Adolescent Alchohol and Tobacco Use" *Econometrics and Health Economics*, 7:439-453

Raghunathan, E Trivellore, James M Lepkowski, John Van Hoewyk, Peter Solenberger (2001) "A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models" *Survey Methodology*, 27:85-95

Reinisch, June Machover, Stephanie A. Sanders, Erik Lykke Mortensen, Donald B. Rubin (1995) "In Utero Exposure to Phenobarbital and Intelligence Deficits in Adult Men" *JAMA*, 274:1518-1525

Rosenbaum, Paul R, Donald B Rubin (1983)"The Central Role of the Propensity Score in Observational Studies for Causal Effects" *Biometrica*, 70(1):41-55

Rubin, Donald B (2005) "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions" *Journal of the American Statistical Association*, 100(469):322-331

Rubin, Donald B (1990) "[On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.] Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies" *Statistical Science*, 5(4):472-480

Rubin, Donald B (1987) "Multiple Imputation for Nonresponse in Surveys" John Wiley & Sons, Inc

Rubin, Donald B (1980) "Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment" *Journal of the American Statistical Association*, 75(371):591-593

Rubin, Donald B (1978) "Multiple imputations in sample surveys - phenomenological Bayesian approach to nonresponse" *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 20-34. Also in *Imputation and Editing of Faulty or Missing Survey Data*, U.S. Dept. of Commerce, Bureau of the Census, 1-23

Rubin, Donald B (1978) "Bayesian Inference for Causal Effects: the Role of Randomization" *The Annals of Statistics*, 6(1):34-58

Rubin, Donald B (1977) "The design of a general and flexible system for handling non-response in sample surveys" Manuscript prepared for the U.S. Social Security Administration, July, 1 1977

Rubin, Donald B (1974) "Estimating causal effects of treatments in randomized and nonrandomized studies" *Journal of Educational Psychology*, 66(5): 688-701

Schafer, Joseph L (1997) Analysis of Incomplete Multivariate Data, Chapman & Hall

Part 3

A quasi-experimental study

A freshman study at Harvard College

There are many reasons why we might expect peers to have similar smoking habits. For example, it is possible that peers mutually influence each others' smoking habits (endogenous effects), or that their smoking habits are similarly affected by some exogenous characteristics defining them as a group or their environments that, to the researcher, are either known (contextual effects) or unknown (correlated effects) (Manski 1993). It is hard to separate the endogenous effects from the contextual and correlated ones. For this reason, we choose to study how Harvard freshmen are influenced by their roommates' smoking habits, because the contextual effects are created by the known process that Harvard College uses to assign students into rooms and because the correlated effects are not present in such a known process. By controlling for the contextual effects, we can then separate the endogenous effects from the correlated effects. We enroll the freshmen in the beginning of their fall semester in 2003 for a first interview and ask them to participate in two follow-up interviews administered online. Here, we discuss the approval process of the study (Section 1), the recruitment of students (Section 2), and finally the interviews themselves (Section 3). We also present the questionnaires and the informed consent forms (Section 4 and Section 5, respectively).

1. Applying for approval

All research at Harvard University involving students requires permission not only from the Committee on the Use of Human Subjects (CUHS)[6], but also from the Committee on the Student Research Participation (CSRP). The CUHS ensures that the study respects the rights and the welfare of the human subjects and the CSRP ensures that the study neither interferes with students' education nor violates University policy.

We did not receive permission from the CUHS and CSRP until we had had a careful deliberation with them about how to handle a number of sensitive issues. We needed, for example, information about not only the study participants themselves but also their roommates. This information could not be collected by asking the participants' roommates to enroll in the study, because the roommates might have felt pressured to

[6] CUHS is the Institutional Review Board at Harvard University.

enroll. In discussions with the committees, we therefore decided to ask the participants to report the information we needed about their roommates.

We also needed information about habits of the participants and their roommates that were illegal, i.e., alcohol consumption and drug habits. For that reason, we agreed to encrypt all such sensitive information in addition to the participants' names, room numbers, and roommates' initials. We also agreed not to conduct the study without a valid Certificate of Confidentiality issued by the U.S. government, because the committee wanted to be sure that we would never be obliged to reveal the identities or the answers of our study participants.

We also asked a battery of questions about the roommates' health habits. It could be possible that through this process the students became uncomfortably aware of their unhealthy habits and needed support to deal with this awareness. For that reason, the committees wanted us to ask the students if they were worried about their health habits and wanted referral information.

Finally, not all students were expected to be 18 at the time the study would be carried out. For that reason, we would need not only these students' own informed consent to participate in the study but also their parents'. In deliberations with the committees, we decided to write a separate informed consent form for the students who wanted to participate but were not yet 18, and to fax this form to their parents for signature.

2. Recruiting students

The rules were such that we were only allowed to stand outside the Science Center to enroll freshmen and to post notices on small billboards in the Harvard Yard. Because we knew that the Science Center would be frequented not only by freshmen but also other by undergraduate students and graduate students and that the billboards would be full of other notices, we made an effort to stand out at the Science Center and on the billboards.

We hired four assistants who helped us enroll students after they had gone through appropriate training[7]. We all wore t-shirts printed with the question "Calling all freshmen, can you spare 5-10 minutes of your time?" in order to distinguish ourselves from other students outside the Science Center and to communicate that we were only interested

[7] They underwent on-line training provided by Harvard University and Rice University. We also gave them some additional training in an afternoon session.

in the freshmen's participation. In addition, we had two big posters with information about our study at the exit of the Science Center and we used additional ones around the table we had at our disposal. Small flyers were distributed to all passers-by. We also posted notices in the yard that directed interested students to enroll in the study either outside the Science Center or at the study homepage. All these materials were professionally produced, and their designs were built on a common theme.

We expected the students to be worried about revealing sensitive information about their alcohol consumption and drug habits. Therefore, we wanted the students to understand the precautions we had taken to protect their data before they enrolled, so that they would feel confident responding honestly to the study questions. The students were informed that once they handed in their responses to the questions on the first interview, their answers would immediately be separated from their contact details and would be placed in separate piles. Thereafter, only the principal investigator, i.e., Sophie Langenskiöld, would be able to relate their identities with their answers. They were also informed that the principal investigator was protected by a Certificate of Confidentiality and that she therefore could never be requested to reveal any information about the students, not even under court order or subpoena. In addition, they were informed that sensitive information would be encrypted in the database where it was stored.

After having received this information, the students were asked to sign the informed consent form if they were still interested in enrolling in the study. Then they were given the first questionnaire to complete. We assumed that the students would be more honest in their answers if they completed the questionnaire themselves than they would have been in an interview setting; after all, we were not professional interviewers. Once the students had filled out the questionnaire, they received homemade cookies as a compensation for their participation. They were also enrolled in a lottery where they had a chance to win one of the six prizes[8]. The cookies were popular, and so became the study.

The students could also enroll in the study online. To do so, they were required to first register for the study, to read through and sign the informed consent form, and to wait for their username and password before they could complete the first questionnaire. These students could not be compensated with cookies but were compensated with a lottery ticket.

[8] There were six prizes in the lottery, one prize of USD 500, one of USD 250, and four of USD 50.

3. Interviewing participants

In the first interview, the students were asked to replicate the answers they gave on the Housing Application Form, because these answers were the basis for the assignment of students into suites. Also, they were asked to describe their different health habits 30 days before they enrolled in college, and their parents' health habits when they were still children. In total 589 students enrolled in the study. The first interview was launched at the first day of their classes and continued over a two-week period.

Approximately five weeks later, the students were contacted by email and were asked to complete the second interview's questionnaire online. In the email, they were given the username and password they needed to access the online questionnaire. The students were asked to describe their roommates' health habits during the last 30 days in this interview. Also, they were asked for some additional information that we should have included in the first interview, i.e., the students' race and their room number. Once the second interview was closed two weeks later, 462 students had completed this second questionnaire.

Another 6 months later, the students were again contacted for a final interview. In addition to questions about their health habits during the last 30 days before that interview, the students were questioned whether they still lived with the same roommates and/or had any additional roommates. With this information, we would be able to control for changes in room configuration. This interview period again lasted for two weeks, and 411 students remained once the last interview was over.

4. Questionnaires

Questionnaire 1(3). The first interview was launched at the first day of their classes and continued for two weeks.

> This first section asks questions about your personal profile and the health habits that you expressed during the 30 days prior to your arrival at Harvard College.

1(1) How old are you? .....

2(1) Are you male or female?
1 = Male
2 = Female

3(1) What best describes your ethic group?

1 = White/Caucasian
2 = Black/African American
3 = Native American
4 = Asian/Pacific Islander
5 = Latino origin or descent
6 = Other .....
77= Decline
99= Don't know

4(1) During the 30 days prior to your arrival, did you smoke cigarettes?

1 = Every day
2 = On several days
3 = On one or two days
4 = Not at all
77= Decline
99= Don't know

If you replied 1 = Every day, 2 = On several days, or 3 = On one or two days to the above question, I would like you to consider the following three sub-questions.

During the 30 days prior to your arrival, on approximately how many days did you smoke cigarettes?

.....
77= Decline
99= Don't know

During the 30 days prior to your arrival, on days when you smoked, approximately how many cigarettes did you smoke each day?

.....
77= Decline
99= Don't know

During the 30 days prior to your arrival, approximately how many cigarettes did you smoke each day?

.....
77= Decline
99= Don't know

5(1) During the 30 days prior to your arrival, did you smoke a pipe or a cigar?
1 = Every day
2 = On several days
3 = On one or two days
4 = Not at all
77= Decline
99= Don't know

6(1) During the 30 days prior to your arrival, did you use chewing tobacco or snuff?
1 = Every day
2 = On several days
3 = On one or two days
4 = Not at all
77= Decline
99= Don't know

7(1) During the 30 days prior to your arrival, did you use any other illegal substances? If yes, please specify what you used.
1 = Every day .....
2 = On several days .....
3 = On one or two days .....
4 = Not at all
77= Decline
99= Don't know

8(1) During the 30 days prior to your arrival, did you drink alcoholic beverages like beer, wine, or liquor?
1 = Yes
2 = No
77= Decline
99= Don't know

If you replied 1 = Yes in the above questions, I would like you to consider the following four questions. In the questions, 1 drink = 1 can or bottle of beer, 1 glass of wine, 1 can or bottle of wine-cooler, 1 mixed-drink, or 1 shot of liquor.

During the 30 days prior to your arrival, on approximately how many days did you have at least 1 alcoholic drink?
1 = Number of days each week .....

2 = Number of days over all 30 days .....
3 = No alcoholic drink over all 30 days
77= Decline
99= Don't know

During the 30 days prior to your enrollment, on days when you consumed alcoholic beverages, approximately how many drinks did you have?

.....
77= Decline
99= Don't know

During the 30 days prior to your arrival, on days when you consumed alcoholic beverages, approximately how many times did you have 5 or more drinks on a single occasion?

.....
77= Decline
99= Don't know

During the 30 days prior to your arrival, what did you drink most often: beer, wine, wine-coolers,straight liquor, or mixed drinks?
1 = Beer
2 = Wine
3 = Wine-coolers
4 = Straight liquor
5 = Mixed drinks
6 = Whatever was available
77= Decline
99= Don't know

9(1) During the 30 days prior to your arrival, approximately how much total time per day/per week did you devote to MODERATE physical activities, i.e., brisk walks, bicycling, vacuuming, gardening, or anything else that may have caused an increase in your breathing or heart rate?
1 = Minutes per day.....
2 = Hours per week .....
77= Decline
99= Don't know

10(1) During the 30 days prior to your arrival, approximately how much total time per day did you devote to VIGOROUS physical activities, i.e., running, aerobics, heavy yard work, or anything else that caused a significant increase in your breathing or heart rate?

1 = Minutes per day.....

2 = Hours per week .....

77= Decline

99= Don't know

11(1) If any of your health habits concern you, we would like to give you referral information so that you can discuss them with a clinician. Do any of your health habits concern you?

1 = Yes

2 = No

77= Decline

99= Don't know

In the second section, I would like you to think about the health habits of your mother/father, or of the other person/s acting as your mother/father during your childhood.

12(1) Did your mother/father smoke cigarettes, a pipe or cigars, or use chewing tobacco, or snuff?

Mother:

0 = No Mother/substitute

1 = Every day

2 = On several days each month

3 = On one or two days each month or less

4 = Not at all

77= Decline

99= Don't Know

Father:

0 = No Father/substitute

1 = Every day

2 = Several days each month

3 = On one or two days each month or less

4 = Not at all

77= Decline

99= Don't Know

13(1) Describe your mother's/father's use of alcohol?

Mother:	Father:
0 = No Mother/substitute	0 = No Father/substitute
1 = Abstainer	1 = Abstainer
2 = Infrequent or Light drinker	2 = Infrequent or Light drinker
3 = Moderate Drinker	3 = Moderate Drinker
4 = Problem Drinker	4 = Problem Drinker
77= Decline	77= Decline
99= Don't know	99= Don't know

14(1) What level of education did your mother/father complete?

Mother:	Father:
0 = No Mother/substitute	0 = No Father/substitute
1 = High School	1 = High School
2 = College	2 = College
3 = Graduate School	3 = Graduate School
77= Decline	77= Decline
99= Don't know	99= Don't know

This last section asks you questions about the Dean of Freshmen's Application Form that you filled out for your first-year housing at Harvard. Your answers to these questions will increase the quality of the study data; they help us to adjust for systematic differences between your and other participants' roommates.

15(1) Please indicate the NUMBER of other students you said you preferred to live with; 77 and 99 represent "decline" and "don't know" respectively:

1 2 3 4 5 77 99

16(1) Please indicate the preference you said you had for a room that was; 77 and 99 represent "decline" and "don't know" respectively:

quiet/serious lively social center

1 2 3 4 5 77 99

neat						disordered
1	2	3	4	5	77	99

17(1) Please indicate the HOURS you said you preferred to keep during the week:

I said that my day begins at:	I said that my day ends at:
1 = 6 - 7 a.m	1 = 11-12 p.m
2 = 8 - 9 a.m	2 = 1 - 2 a.m
3 = 10-11 a.m	3 = 3 - 4 a.m
77= Decline	77= Decline
99= Don't know	99= Don't know

18(1) What did you say your ACADEMIC interests were?

.....

.....

19(1) How did you rank (1,2,3) in order of importance THREE EXTRACURRICULAR interests that you said you hoped to pursue?

....Music	....Student publications
....Polictics, student gov't, debate	....Outdoors (camping...)
....Religious activities	....Intercollegiate athletics
....Ethnic organizations	....Drama/performing arts
....Intramural/recreational sports	....Other.....
....Fine arts (painting, photography...)	....Decline
....Social services	....Don't know

20(1) How did you rank the types of MUSIC that you said you usually listened to or enjoyed?

....Broadway tunes	....Dance/top 40	....Jazz
....Classical	....Electronic	....Oldies
....Country/western	....Indie/punk	....Opera

....Rock	....Latin	....Decline
....Rap/R&B	....Other(s).....	....Don't know.
....Reggae		

21(1) Did you anticipate a need for residential or academic accommodations related to a medical condition or a disability?
1 = Yes
2 = No
77= Decline
99= Don't know

22(1) How carefully did you complete the Housing Application Form; 77 and 99 represent "decline" and "don't know" respectively?

carelessly						carefully
1	2	3	4	5	77	99

23(1) How truthfully did you answer the questions on the housing application form; 77 and 99 represent "decline" and "don't know" respectively?

not very truthfully						truthfully
1	2	3	4	5	77	99

24(1) If you had been asked about religious tolerance, how would you have responded at that time; 77 and 99 represent "decline" and "don't know" respectively?

tolerant only of my own religion						tolerant of any religion
1	2	3	4	5	77	99

25(1) How much help did you receive from other people when completing the form; 77 and 99 represent "decline" and "don't know" respectively?

no help						a lot of help
1	2	3	4	5	77	99

Questionnaire 2(3). The second interview was launched five weeks after the first day of their classes and continued for two weeks.

The introductory section asks for information that is needed to complement information given in the previous interview.

1(2) Where did you live before you arrived at Harvard College?

Alabama	Maryland	Pennsylvania
Arizona	Massachusetts	Rhode Island
Arkansas	Michigan	South Carolina
California	Minnesota	South Dakota
Colorado	Missisippi	Tennesse
Connecticut	Missouri	Texas
Delaware	Montana	Utah
DC	Nebraska	Vermont
Florida	Nevada	Virgina
Georgia	New Hampshire	Washington
Idaho	New Jersey	West Virginia
Illinois	New Mexico	Wisconsin
Indiana	New York	Wyoming
Iowa	North Carolina	Alaska
Kansas	North Dakota	Hawaii
Kentucky	Ohio	Guam
Louisiana	Oklahoma	Puerto Rico
Maine	Oregon	Virgin Islands
Alberta:	Labrador	Prince Edward Island
British Columbia	Northwest Territories	Quebec
Manitoba	Nova Scotia	Saskatchewan
New Brunswick	Nunavut	Yukon Territory
Newfoundland	Ontario	

Africa	Australia	South America
Asia	Europe	

77= Decline
99= Don't know

2(2) Before you enrolled at Harvard College, did you ever try any of the following substances? The numbers 77 and 99 represent "decline" and "don't know", respectively. Please reply yes even if you used very little of the substance in question.

	Yes	No	77	99
Alcohol				
Cigarettes				
Cigars				
Pipes				
Chewing tobacco				
Snuff				
Illegal drugs and/or misuse of prescription drugs				

The first section asks questions about your general housing conditions.

3(2) In which dormitory do you live?

Apley Court	Hurlbut	Stoughton
Canaday	Lionel	Straus
Grays	Massachusetts Hall	Thayer
Greenough	Matthews	Weld
Hollis	Mower	Wigglesworth
Holworthy	Pennypacker	

66= Other.....
77= Decline
99= Don't know

4(2) What is your room-number?
Please note that all room numbers will be encrypted during the remaining part of the study and changed to another number once the study has been completed, a number that cannot be linked to any particular room
.....

77= Decline
99= Don't know

5(2) How many roommates do you have?
Please note that all freshmen that have the same room number are defined as roommates.
1 = 1 freshman
2 = 2 freshmen
3 = 3 freshmen
4 = 4 freshmen
5 = 5 freshmen
6 = Other .....
77= Decline
99= Don't know

6(2) Can you please list your roommates by their initials (add the second letter of their first name if two or more roommates share the same initials)? Also specify whether or not you share a bedroom with one or two of them.
The remaining section in this follow-up interview will ask questions about your roommates by referring to these initials.
1 = ...we share bedroom ■ we do not share bedroom ■
2 = ...we share bedroom ■ we do not share bedroom ■
3 = ...we share bedroom ■ we do not share bedroom ■
4 = ...we share bedroom ■ we do not share bedroom ■
5 = ...we share bedroom ■ we do not share bedroom ■
77= Decline
99= Don't know

7(2) Does your suite have a common room in addition to bedrooms?
Please note that all freshmen that have the same room number also share the same suite.
1 = Yes
2 = No
77= Decline
99= Don't know

8(2) How did you and your roommates decide who should sleep in each bedroom?
1 = By flipping a coin or drawing cards

2 = By discussing similarities and differences, e.g., similar sleeping habits.
3 = Other explanations.
77= Decline
99= Don't know

9(2) Have you decided to rotate bedrooms this semester or next semester?
1 = No
2 = Yes
77= Decline
99= Don't know

In this section, do your best to describe your roommates' health habits during the first month of the semester. For each roommate, you will be asked to fill out two tables: one table asks you about the frequency of your roommate's health habits, the other about the intensity of those habits.

10(2) During the FIRST MONTH OF THIS SEMESTER, approximately how FREQUENTLY did [repeat name] engage in the following activities[9]? The numbers 77 and 99 represent "decline" and "don't know", respectively.

	FREQUENCY					
	4	3	2	1	77	99
Drinking alcohol						
Smoking cigarettes						
Smoking cigars						
Smoking a pipe						
Chewing tobacco						
Taking snuff						
Using illegal drugs misusing prescription drugs						
VIGOROUS physical exercise						
MODERATE physical exercise						

11(2) On days when [repeat name] engaged in the following activities, approximately how INTENSE was each activity? Please enter 0 if the activity in question never occurred during the first month of this semester. The numbers 77 and 99 represent "decline" and "don't know", respectively.

[9] 1 = not at all, 2 =one or two days/month, 3 =several days/month, 4 =every day

	INTENSITY	77	99
Drinking alcohol	drinks/day		
Smoking cigarettes	cigarettes/day		
Smoking cigars	cigars/day		
Smoking a pipes	pipefuls of tobacco/day		
Chewing tobacco	plugs, wads, chaws/day		
Taking snuff	pinches, dips, rubs/day		
Using illegal drugs mis-using prescription drugs	times/day		
VIGOROUS physical exercise	minutes/day		
MODERATE physical exercise	minutes/day		

The following links were also provided on the hompage:
Link 1 VIGOROUS physical activities: Activities that cause heavy perspiration or a significant increase in breathing or heart rate. Some examples are running, swimming laps, aerobic classes, fast bicycling, e.t.c.
Link 2 MODERATE physical activities: Activities that cause only moderate perspiration or a slight to moderate increase in breathing or heart rate
Link 3 DRINKS/day: 1 drink=1 can or bottle of beer, 1 glass of wine, 1 can or bottle of wine-cooler, 1 mixed-drink, or 1 shot of liquor

Questionnaire 3(3). The third interview was launched six months after the first day of their classes and continued for two weeks.

The introductory part asks questions about your suite-number and/or your dormitory at the time of the previous interview. We must have this information to make use of the data that you have so kindly provided us with. The information is necessary if we are to understand which study participants are sharing suites and adjust the analysis accordingly. Please remember that your suite-number and dormitory will be encrypted during the remaining part of the study; later, once the study has been completed, it will be changed to a number that cannot be linked to any particular room.

1(3) At the time of the previous interview, in which suite did you live, i.e., which room number did you have?
Please note that a group of rooms with one room number is a suite.
Also note that you took the previous interview in November, ...
.....
77= Decline
99= Don't know

2(3) At the time of the previous interview, in which dormitory did you live?
Please note that you took the previous interview in November, ...

Apley Court	Hurlbut	Stoughton
Canaday	Lionel	Straus
Grays	Massachusetts Hall	Thayer
Greenough	Matthews	Weld
Hollis	Mower	Wigglesworth
Holworthy	Pennypacker	

66= Other.....
77= Decline
99= Don't know

The first part asks questions about your present housing conditions and changes in those conditions that may have taken place after the previous interview.

3(3) Do you live in the same suite as you did at the time of the previous interview?
Please note that a group of rooms with one number is a suite.
Also note that you took the previous interview in November, ...
1 = Yes
2 = No
77= Decline
99= Don't know

4(3) In the previous interview, you reported that you were sharing a suite with the following roommates [list the initials reported in the previous interview]. Could you please specify with whom you are still sharing a suite?
SL .....We are still sharing a suite ■ We are no longer sharing a suite ■
BL .....We are still sharing a suite ■ We are no longer sharing a suite ■

5(3) In addition to the roommates you reported sharing a suite with in the previous interview, i.e. [list the initials reported in the previous interview], have you shared a suite with any others since?
If so, could you please specify with how many other roommates you have shared a suite?
1a= Yes, I have shared a suite with one/some other roommates
1b= The number of additional roommates that I have shared a suite with is:
2 = No, I have not shared a suite with any additional roommates.

77= Decline
99= Don't know

6(3) Can you please list the initials of these roommates, i.e., roommates with whom you have shared or now share a suite but whose initials you did not report in the previous interview.
Please add the second letter of all first names that have the same initials as other roommates you listed?

.....
.....

7(3) For how long a period of time have you shared a SUITE and a BEDROOM with each of the roommates listed?

Shared a SUITE with SL:	Shared a SUITE with BL:	Shared a SUITE with LL:
From Sept. 2003	From Sept. 2003	From Sept. 2003
For several months	For several months	For several months
For 1 or 2 months	For 1 or 2 months	For 1 or 2 months
Not at all	Not at all	Not at all
Decline	Decline	Decline
Don't know	Don't know	Don't know

Shared a BEDROOM with SL:	Shared a BEDROOM with BL:	Shared a BEDROOM with LL:
From Sept. 2003	From Sept. 2003	From Sept. 2003
For several months	For several months	For several months
For 1 or 2 months	For 1 or 2 months	For 1 or 2 months
Not at all	Not at all	Not at all
Decline	Decline	Decline
Don't know	Don't know	Don't know

8(3) Have you and your roommates rotated bedrooms since the previous interview?
1 = Yes
2 = No
77= Decline
99= Don't know

9(3) At the time/s you and your roommates rotated bedrooms, how did you and your roommates decide who should sleep in each bedroom?
1 = By flipping of a coin or a drawing cards
2 = By discussing similarities or differences
3 = Other explanations
77= Decline
99= Don't know.

This last part asks questions about your health behavior during the previous month. You will be asked to fill out two tables: one table asks you about the frequency of your health habits, the other about the intensity of those habits.

10(3) During the PREVIOUS MONTH, approximately how FREQUENTLY did you engage in the following activities[10]? The numbers 77 and 99 represent "decline" and "don't know", respectively.

	FREQUENCY					
	4	3	2	1	77	99
Drinking alcohol						
Smoking cigarettes						
Smoking cigars						
Smoking a pipe						
Chewing tobacco						
Taking snuff						
Using illegal drugs, misusing prescription drugs						
VIGOROUS physical exercise						
MODERATE physical exercise						

11(3) On days when you engaged in the following activities, approximately how INTENSE was each activity? Please enter 0 if the activity in question never occurred during the last month of this semester. The numbers 77 and 99 represent "decline" and "don't know", respectively.

	INTENSITY	77	99
Drinking alcohol	DRINKS/day		
Smoking cigarettes	cigarettes/day		
Smoking cigars	cigars/day		
Smoking a pipes	pipefuls of tobacco/day		
Chewing tobacco	plugs, wads, chaws/day		
Taking snuff	pinches, dips, rubs/day		
Using illegal drugs mis-using prescription drugs	times/day misused prescription drug		
VIGOROUS physical exercise	minutes/day		
MODERATE physical exercise	minutes/day		

[10] 1 = not at all, 2 =one or two days/month, 3 =several days/month, 4 =every day

The following links were also provided on the hompage:
Link 1 VIGOROUS physical activities: Activities that cause heavy perspiration or a significant increase in breathing or heart rate. Some examples are running, swimming laps, aerobic classes, fast bicycling, e.t.c.
Link 2 MODERATE physical activities: Activities that cause only moderate perspiration or a slight to moderate increase in breathing or heart rate. Some examples are brisk walks, bicycling, vacuuming, gardening, or anything else that may have caused an increase in your breathing or heart rate.
Link 3 DRINKS/day: 1 drink=1 can or bottle of beer, 1 glass of wine, 1 can or bottle of wine-cooler, 1 mixed-drink, or 1 shot of liquor.

If any of your health habits concern you, we would like to give you referral information so that you can discuss them with a clinician. Do any of your health habits concern you?
1 = Yes
2 = No
77= Decline
99= Don't know

5. Informed consent forms

Consent for for Adults
Invitation to Participate in a Research Study

How Peers Influence Health Habits among Freshmen
Researcher Sophie Langenskiöld

Purpose:

You and your freshman peers will already have a variety of habits that affect your health. The purpose of this study is to understand if, and how, those health habits will change during your freshman year, possibly as a consequence of your roommates' influence. Previous research has indicated that we are influenced by our peer group's habits in a number of ways. A peer group can even influence our health habits and ultimately, our health. Understanding these peer group effects is, therefore, of great importance, since health is one of the prerequisites for our overall sense of well-being. You can help us better understand these effects by participating in this study.

Procedures:

This study will ask the participants to complete questionnaires on three occasions.

If you decide to enroll in the study, your responses to the questionnaires can either be given to the researcher in person, on line (PeerStudy.fas.harvard.edu), or over the telephone. In the first and the last questionnaire, you will be asked questions relating to your own health habits, whereas the questions in the second interview will relate to your roommates health habits. In addition, the first interview will ask you questions about health habits to which you may have been exposed during your childhood. None of these questionnaires is expected to take longer than ten minutes.

Benefits and Risks:

You may not benefit directly from the results of the survey, but perhaps you will derive some satisfaction from knowing that your participation has a scientific value and will further our understanding of peer group effects. As a compensation for your time, you will be entered in a drawing, and your odds of winning are 6:300. One winner will be drawn for the first prize of $200 and for the second prize of $100. For the last and third prize, four winners will be drawn and each will receive $50. You will be asked questions about your smoking, drinking and exercising habits, some of which may conflict with Harvard College policy or Massachusetts state law. I will take precautions to protect this information, see next sections.

Confidentiality

Your name and contact information are requested, since your participation on two further occasions is important for the study. Precautions have be taken to protect your identity:

(1) A Confidentiality Certificate, issued by the U.S. Government, protects the researcher from being forced to identify you, even under court order or subpoena, but this does not mean that the Government has lent its support to the project. You should, however, know that the researcher may provide information to certain individuals or agencies if he/she sees any sign of child abuse of minors;

(2) Your name and contact information will be separated from the questionnaire as soon as the researcher receives it. The code linking your identity to any of the questionnaires will be stored in a locked cabinet during the study and will be destroyed after the drawing has taken place. During the collection of data, it will, therefore, not be possible for a third party to match your identity with your responses, and upon the completion of data collection, there will be no remaining connection between you and

this study.

Voluntary Participation:

Participation in this study is entirely voluntary. At any time, for any reason, you are free to decline to answer questions or to end your participation without penalty or any loss of compensation.

Questions:

If you have questions regarding the research or your participation in it, please do not hesitate to ask the researcher, Sophie Langenskiöld. Cell phone: (617)-504-8515. E-mail contact: langensk@fas.harvard.edu. She will be happy to answer any of your questions. If you have questions about your rights as a research participant, you may contact the Committee on the Use of Human Subjects, Harvard University. Committee phone: (617)-495-5459. E-mail contact: jcalhoun@fas.harvard.edu

Agreement to Participate

I have read the above information, have had the opportunity to ask any questions that I may have about this study and have received satisfactory answers. I am eligible to participate in this study; I am a freshman; and I am 18 years of age. It is my decision to take part in this study.

.....

(date)

.....

(printed name)

.....

(signature)

Consent Form for Minors
Invitation to Participate in a Research Study

How Peers Influence Health Habits among Freshmen

Researcher Sophie Langenskiöld

Purpose:

Your child and his/her freshman peers will already have a variety of habits that affect their health. The purpose of this study is to understand if, and how, those health habits

will change during their freshman year, possibly as a consequence of their roommates' influence. Previous research has indicated that people are influenced by peer group's habits in a number of ways. A peer group can even influence people's health habits and ultimately, their health. Understanding these peer group effects is, therefore, of great importance, since health is one of the prerequisites for our overall sense of well-being. Your child can help us better understand these effects by participating in this study.

Procedures:

This study will ask the participants to complete questionnaires on three occasions. Those enrolling in the study can return completed questionnaires to the researcher in person, on line (PeerStudy.fas.harvard.edu), or over the telephone. In the first and the last questionnaire, participants will be asked questions relating to their own health habits, whereas the questions in the second interview will relate to their roommates' health habits. In addition, the first interview will ask questions about health habits to which the participant may have been exposed during their childhood. None of these questionnaires is expected to take longer than ten minutes.

Benefits and Risks:

Participants may not benefit directly from the results of the survey, but perhaps will derive some satisfaction from knowing that their participation has a scientific value and will further understanding of peer group effects. As a compensation for their time, participants will be entered in a drawing; the odds of winning are 6:300. One winner will be drawn for the first prize of $200 and for the second prize of $100. For the third prize, four winners will be drawn and each will receive $50. The surveys ask questions about smoking, drinking and exercising habits, some of which may conflict with Harvard College policy or Massachusetts state law. I will take precautions to protect this information, see next sections.

Confidentiality

Participants' name and contact information are requested, since participation on two further occasions is important for the study. Precautions have been taken to protect participants' identity and the information they provide:

(1) A Confidentiality Certificate, issued by the U.S. Government, protects the researcher from being forced to identify participants, even under court order or subpoena, but this does not mean that the Government has lent its support to the project. You

should know, however, that the researcher may provide information to certain individuals or agencies if he/she sees any sign of child abuse of minors;

(2) Participants' name and contact information will be separated from the questionnaire as soon as the researcher receives it. The code linking participants' identity to any of the questionnaires will be stored in a locked cabinet during the study and will be destroyed after the drawing has taken place. During the collection of data, it will, therefore, not be possible for a third party to match the participants' identity with their responses, and upon the completion of data collection, there will be no remaining connection between them and this study.

Voluntary Participation:

Participation in this study is entirely voluntary. At any time, for any reason, your child is free to decline to answer questions or to end his/her participation without penalty or any loss of compensation.

Questions:

If you have questions regarding the research or your child's participation in it, please do not hesitate to ask the researcher, Sophie Langenskiöld. Cell phone: (617)-504-8515. E-mail contact: langensk@fas.harvard.edu. She will be happy to answer any of your questions. If you have questions about your child's rights as a research participant, you may contact the Committee on the Use of Human Subjects, Harvard University. Committee phone: (617)-495-5459. E-mail contact: jcalhoun@fas.harvard.edu

Agreement to Participate

I have read the above information, have had the opportunity to ask any questions that I may have about this study and have received satisfactory answers. My child is eligible to participate in this study; he/she is a freshman. I give permission to my child to participate in this study.

.....

(Date)

.....

(Printed Name of Child)

.....

(Printed Name of Parent)

.....

(Signature of Parent)

References

Manski, Charles F. (1993) "Identification of Endogenous Social Effects: The Reflection Problem" *Review of Economic Studies*, 60, 531-542

EFI, The Economic Research Institute

Reports since 2000
A complete publication list can be found at www.hhs.se/efi

Published in the language indicated by the title.

2005

Books

Andersson, Per, Susanne Hertz and Susanne Sweet (eds). *Perspectives on market networks – boundaries and new connections.*

Charpentier, Claes. *IT inom omsorgen. Förväntade effekter av införande av IT-system för utförarna inom äldre- och handikappomsorgen.*

Lind, Johnny och Göran Nilsson (redaktörer). *Ekonomistyrningens metoder, sammanhang och utveckling – En vänbok till Lars A Samuelson.*

Samuelson, Lars A. *Organizational governance and control – a summary of research in the Swedish society.*

Dissertations

Andersson, Martin. *Making a Difference – Project Result Improvement in Organizations.*

Arvidsson, Per. *Styrning med belöningssystem – Två fallstudier om effekter av belöningssystem som styrmedel.*

Bems, Rudolfs. *Essays in International Macroeconomics.*

Berg-Suurwee, Ulrika. *Nya trender, nya nämnder – effekter av en stadsdelsnämndsreform inom kultur och fritid.*

Björkman, Hans. *Learning from members – Tools for strategic positioning and service innovation in trade unions.*

Bodnaruk, Andriy. *Essays in Empirical Corporate Finance and Portfolio Choice.*

Clapham, Eric. *Essays in Real Estate Finance.*

Dareblom, Jeanette. *Prat, politik och praktik – Om individers möten med strukturer i en kommunal satsning på kvinnors företagande.*

Hjelström, Anja, *Understanding International Accounting Standard Setting – A Case Study of IAS 12, Accounting for Deferred Tax.*

Hortlund, Per. *Studies on Swedish Banking 1870-2001.*

Lindahl, Therese. *Strategic and Environmental Uncertainty in Social Dilemmas.*

Linnarsson, Håkan. *Alliance for Innovation. A structural perspective on new business development in cooperative ventures.*

Madestam, Andreas. *Developing Credit Markets*

Nilsson, Roland. *The Market Impact of Short-Sale Constraints.*

Nordfält, Jens. *Is consumer decision-making out of control? Non-conscious influences on the consumer decision-making process for fast moving consumer goods.*

Nordin, Fredrik. *Externalising Services – Walking a Tightrope between Industrial and Service Logics.*

Simbanegavi, Witness. *Price Discrimination, Advertising and Competition.*

Thodenius, Björn. *Användning av ledningsinformationssystem: en longitudinell studie av svenska storföretag.*

Tolis, Christofer. *Framing the Business – Business Modelling for Business Development.*

Östberg, Per. *Corporate Disclosure and Investor Recognition.*

2004

Books

Ahrne, Göran och Nils Brunsson (red). *Regelexplosionen.*

Lind, Johnny. *Strategi och ekonomistyrning. En studie av sambanden mellan koncernstrategi, affärsstrategi och ekonomistyrning.*

Lind, Johnny och Walter Schuster (red). *Redovisningens teori, praktik och pedagogik. En vänbok till Lars Östman.*

Sevón, Guje och Lennart Sjöberg (red). *Emotioner och värderingar i näringslivet.* EFIs Årsbok 2004.

Wijkström, Filip and Stefan Einarsson. *Foundations in Sweden – Their scope, roles and visions.*

Dissertations

Anderson, Anders. *Essays in Behavioral Finance.*

Balsvik, Gudrun. *Information Technology Users: Studies of Self-Efficacy and Creativity among Swedish Newspaper Journalists.*

Blix, Magnus. *Essays in Mathematical Finance – Modelling the Futures Price.*

González Gómez, Andrés. *Nonlinear dynamics and smooth transition models.*

Grönqvist, Erik. *Selection and Moral Hazard in Health Insurance: Taking Contract Theory to the Data.*

Ivarsson Westerberg, Anders. *Papperspolisen – varför ökar administrationen i moderna organisationer.*

Jutterström, Mats. *Att påverka beslut – företag i EUs regelsättande.*

Jönsson, Kristian. *Macroeconomic Aspects of Capital Flows to Small Open Economies in Transition.*

Larsson, Pär. *Förändringens villkor. En studie av organisatoriskt lärande och förändring inom skolan.*

Lagerwall, Björn. *Empirical Studies of Portfolio Choice and Asset Prices.*

Malmsten, Hans. *Properties and Evaluation of Volatility Models.*

Marshall, Cassandra. *Dating for Innovation. Creating and Recognizing Opportunities through Collaborative Interorganizational Relationships in Fluid Environments.*

Mattsson, Susanna. *På gränsen mellan ordning och oordning – tingens betydelse vid marknadsombildningar. En studie av svenska postväsendets ombildning under 1990-talet.*

Nilsson, Charlotte. *Studies in Environmental Economics: Numerical Analysis of Greenhouse Gas Policies.*

Nilsson, Hans. *Medborgaren i styrsystemet – beskrivning av VAD och HUR i styrning av kommunal verksamhet.*

Nystedt, Jens. *Competition, Regulation and Integration in International Financial Markets.*

Pajuste, Anete. *Corporate Governance and Controlling Shareholders.*

Richtnér, Anders. *Balancing Knowledge Creation. Organizational Slack and Knowledge Creation in Product Development.*

Salabasis, Mickael. *Bayesian Time Series and Panel Models – Unit Roots, Dynamics and Random Effects.*

Sandberg, Rickard. *Testing the Unit Root Hypothesis in Nonlinear Time Series and Panel Models.*

Skallsjö, Sven. *Essays on Term Structure and Monetary Policy.*

Strikholm, Birgit. *Essays on Nonlinear Time Series Modelling and Hypothesis Testing.*

Söderström, John. *Från Produkt till Tjänst. Utveckling av affärs- och miljöstrategier i produktorienterade företag.*

Talia, Krim. *The Scandinavian Currency Union, 1873–1924 – Studies in Monetary Integration and Disintegration.*

2003

Books

Lundahl, Mats (editor). *Globalization and Its Enemies.* EFIs Årsbok 2003.

Sundgren, Bo, Pär Mårtensson, Magnus Mähring and Kristina Nilsson (editors). *Exploring Patterns in Information Management. Concepts and Perspectives for Understanding IT-Related Change.*

Dissertations

Andersson, Henrik. *Valuation and Hedging of Long-Term Asset-Linked Contracts.*

Bergman, Malin. *Essays on Human Capital and Wage Formation.*

Damsgaard, Niclas. *Deregulation and Regulation of Electricity Markets.*

Eklund, Bruno. *Four Contributions to Statistical Inference in Econometrics.*

Hakkala, Katariina. *Essays on Restructuring and Production Decisions in Multi-Plant Firms.*

Holgersson, Charlotte. *Rekrytering av företagsledare. En studie i homosocialitet.*

Ivaschenko, Iryna. *Essays on Corporate Risk, U.S. Business Cycles, International Spillovers of Stock Returns, and Dual Listing.*

Lange, Fredrik. *Brand Choice in Goal-derived Categories – What are the Determinants?*

Le Coq, Chloé. *Quantity Choices and Market Power in Electricity Market.*

Magnusson, Peter R. *Customer-Oriented Product Development – Experiments Involving Users in Service Innovation.*

Meisiek, Stefan. *Beyond the Emotional Work Event Social Sharing of Emotion in Organizations.*

Mårtensson, Anders. *Managing Mission-Critical IT in the Financial Industry.*

Nilsson, Göran. *Processorientering och styrning – Regler, mål eller värderingar?*

Sandberg, Robert. *Corporate Consulting for Customer Solutions Bridging Diverging Business Logics.*

Sturluson, Jon Thor. *Topics in the Industrial Organization of Electricity Markets.*

Tillberg, Ulrika. *Ledarskap och samarbete – En jämförande fallstudie i tre skolor.*

Waldenström, Daniel. *Essays in Historical Finance.*

Wallén, Ulrika. *Effektivitet i grundskolan i anslutning till en stadsdelsnämndsreform.*

Ögren, Anders. *Empirical Studies in Money, Credit and Banking – The Swedish Credit Market in Transition under the Silver and the Gold Standards, 1834–1913.*

2002

Books

Schuster, Walter. *Företagets Valutarisk – En studie av horisontella och vertikala styrprocesser.*

Sjöstrand, Sven-Erik och Pernilla Petrelius. *Rekrytering av koncernstyrelsen – Nomineringsförfaranden och styrelsesammansättning med fokus på kvinnors ställning och möjligheter.* EFI/SNS Förlag

Löwstedt, Jan och Bengt Stymne (red). *Scener ur ett företag – Organiseringsteori för kunskapssamhället.* EFIs Årsbok 2002.EFI/Studentlitteratur.

Dissertations

Barinaga, Ester. *Levelling Vagueness – A Study of Cultural Diversity in an International Project Group.*

Berglund, Johan. *De otillräckliga – En studie av personalspecialisternas kamp för erkännande och status.*
Bolander, Pernilla. *Anställningsbilder och rekryteringsbeslut.*
Damjanovic, Tatiana. *Essays in Public Finance.*
Ekman, Mattias. *Studies in Health Economics – Modelling and Data Analysis of Costs and Survival.*
Heyman, Fredrik. *Empirical Studies on Wages, Firm Performance and Job Turnover.*
Kallifatides, Markus. *Modern företagsledning och omoderna företagsledare.*
Kaplan, Michael. *Acquisition of Electronic Commerce Capability – The Cases of Compaq and Dell in Sweden.*
Mähring, Magnus. *IT Project Governance.*
Nilsson, Mattias. *Essays in Empirical Corporate Finance and Governance.*
Schenkel, Andrew. *Communities of Practice or Communities of Discipline – Managing Deviations at the Øresund Bridge.*
Skogsvik, Stina. *Redovisningsmått, värderelevans och informationseffektivitet.*
Sundén, David. *The Dynamics of Pension Reform.*
Ternström, Ingela. *The Management of Common-Pool Resources – Theoretical Essays and Empirical Evidence.*
Tullberg, Jan. *Reciprocitet – Etiska normer och praktiskt samarbete.*
Westling, Gunnar. *Balancing Innovation and Control – The Role of Face-to-face Meetings in Complex Product Development Projects.*
Viklund, Mattias. *Risk Policy – Trust, Risk Perception, and Attitudes.*
Vlachos, Jonas. *Risk Matters – Studies in Finance, Trade and Politics.*

2001

Books
Sjöstrand, Sven-Erik. *Center for Management and Organization 50 (1951–2001).*
Charpentier, Claes. *Uppföljning av kultur- och fritidsförvaltningen efter stadsdelsnämndsreformen.*
Hedlund, Andreas. *Konsumentens erfarenhet – och dess inverkan på livsmedelsinköp på Internet.*
Hvenmark, Johan. *Varför slocknar elden? Om utbrändhet bland chefer i ideella organisationer.*
Jönsson, Bengt (red). *Läkemedel – Kostnad eller resurs för sjukvården?* EFIs Årsbok 2001.
Lange, Fredrik och Richard Wahlund. *Category Management – När konsumenten är manager.*
Ljunggren, Ulrika. *Nyckeltal i grundskolan i Stockholms stad före och efter stadsdelsnämndsreformen.*
Thorén, Bertil. *Stadsdelsnämndsreformen och det ekonomiska styrsystemet – Om budgetavvikelser.*

Dissertations
Adolfson, Malin. *Monetary Policy and Exchange Rates – Breakthrough of Pass-Through.*
Andersson, Patric. *Expertise in Credit Granting: Studies on Judgment and Decision-Making behavior.*
Björklund, Christina. *Work Motivation – Studies of its Determinants and Outcomes.*
Dahlén, Micael. *Marketing on the Web – Empirical Studies of Advertising and Promotion Effectiveness.*
Eckerlund, Ingemar. *Essays on the Economics of Medical Practice Variations.*

Ekelund, Mats. *Competition and Innovation in the Swedish Pharmaceutical Market.*
Engström, Stefan. *Success Factors in Asset Management.*
Ericsson, Daniel. *Kreativitetsmysteriet – Ledtrådar till arbetslivets kreativisering och skrivandets metafysik.*
Eriksson, Rickard. *Price Responses to Changes in Costs and Demand.*
Frisell, Lars. *Information and Politics.*
Giordani, Paulo. *Essays in Monetary Economics and Applied Econometrics.*
Gustavsson, Patrik. *Essays on Trade, Growth and Applied Econometrics.*
Hill, Martin. *Essays on Environmental Policy Analysis: Computable General Equilibrium Approaches Applied to Sweden.*
Hägglund, Peter B. *Företaget som investeringsobjekt – Hur placerare och analytiker arbetar med att ta fram ett investeringsobjekt.*
Höök, Pia. *Stridspiloter i vida kjolar, om ledarutveckling och jämställdhet.*
Johansson, Christer. *Styrning för samordning.*
Josephson, Jens. *Evolution and Learning in Games.*
Kjellberg, Hans. *Organising Distribution – Hakonbolaget and the efforts to rationalise food distribution, 1940–1960.*
Liljenberg, Anders. *Customer-geared competition – A socio-Austrian explanation of Tertius Gaudens.*
Lindkvist, Björn. *Kunskapsöverföring mellan produktutvecklingsprojekt.*
Löf, Mårten. *On Seasonality and Cointegration.*
Martensen, Kaj. *Essays on Entry Externalities and Market Segmentation.*
Matros, Alexandr. *Stochastic Stability and Equilibrium Selection in Games.*
Mårtensson, Pär. *Management Processes – An Information Perspective on Managerial Work.*
Nilsson, Arvid. *Market Transparency.*
Norberg, Peter. *Finansmarknadens amoralitet och det kalvinska kyrkorummet – En studie i ekonomisk mentalitet och etik.*
Persson, Björn. *Essays on Altruism and Health Care Markets.*
Rech, Gianluigi. *Modelling and Forecasting Economic Time Series with Single Hidden-layer Feedforward Autoregressive Artificial Neural Networks.*
Skoglund, Jimmy. *Essays on Random Effects Models and GARCH.*
Strand, Niklas. *Empirical Studies of Pricing.*

2000

Books

Bengtsson, Lars, Johnny Lind och Lars A. Samuelson (red). *Styrning av team och processer – Teoretiska perspektiv och fallstudier.*
Berg-Suurwee, Ulrika. *Styrning före och efter stadsdelsnämndsreform inom kultur och fritid – Resultat från intervjuer och enkät.*
Brodin, Bengt, Leif Lundkvist, Sven-Erik Sjöstrand och Lars Östman. *Koncernchefen och ägarna.*
Charpentier, Claes och Lars A. Samuelson. *Effekter av en sjukvårdsreform – En analys av Stockholmsmodellen.*
Emling, Emil. *Svenskt familjeföretagande.*
Ericson, Mona. *Strategi, kalkyl, känsla.*
Gunnarsson, Jonas, Richard Wahlund och Helena Flink. *Finansiella strategier i förändring: segment och beteenden bland svenska hushåll.*
Ljunggren, Ulrika. *Styrning av grundskolan i Stockholms stad före och efter stadsdelsnämndsreformen – Resultat från intervjuer och enkät.*

Schwarz, Brita och Susanne Weinberg. *Serviceproduktion och kostnader – Att söka orsaker till kommunala skillnader.*
Söderlund, Magnus (red). *I huvudet på kunden.* EFIs Årsbok 2000. EFI/Liber Förlag.

Dissertations
Bergkvist, Lars. *Advertising Effectiveness Measurement: Intermediate Constructs and Measures.*
Bornefalk, Anders. *Essays on Social Conflict and Reform.*
Edman, Jan. *Information Use and Decision Making in Groups – A Study of an Experimental Oligopoly Market with the Use of a Business Game.*
Hellman, Niclas. *Investor Behaviour – An Empirical Study of How Large Swedish Institutional Investors Make Equity Investment Decisions.*
Hyll, Magnus. *Essays on the Term Structure of Interest Rates.*
Håkansson, Per. *Beyond Private Label – The Strategic View on Distributor Own Brands.*
Karlsson Stider, Annelie. *Familjen och firman.*
Ludvigsen, Johanna. *The International Networking between European Logistical Operators.*
Nittmar, Henrik. *Produktutveckling i samarbete – Strukturförändring vid införande av nya informationssystem.*
Robertsson, Göran. *International Portfolio Choice and Trading Behavior.*
Stenström, Emma. *Konstiga företag.*
Sweet, Susanne. *Industrial Change Towards Environmental Sustainability – The Case of Replacing Chloroflouorocarbons.*
Tamm Hallström, Kristina. *Kampen för auktoritet – standardiseringsorganisationer i arbete.*