Journal of Interdisciplinary History, XXXVI:2 (Autumn, 2005), 209–232.

Gidon Cohen

Propensity-Score Methods and the Lenin School

A recent resurgence of interest in prosopography, or collective biography, has seen increasing numbers of historians using computers to construct databases. With growing frequency, these databases are built not from regular sources containing standardized data but from multiple sources containing diverse information. Because these databases often contain extensive amounts of information about particular populations or issues, they appear to offer the promise of more rigorous quantitative examination of important historical questions. However, they usually cannot make any claim to knowledge of a complete population or representative sample, nor permit much regularity in the information stored, thus making quantitative analysis difficult. How can the apparent promise of specific quantitative examinations be fulfilled when the data available to historians can be so problematical?

Epidemiologists and others often try to discern the causal effects of drugs or other treatments from the analysis of observational databases by using "propensity scores" to create groups of treated and nontreated patients who are similar in other respects. Usually the data comprise observational information collected under conditions over which the investigators have no control. Broadly, the problems faced by epidemiologists are, in many respects, similar to those encountered in historical research. Although historians have not as yet taken advantage of propensity-score methods, despite their potentially wide applicability, they might well find these observational techniques immensely helpful in the study of a variety of historical issues.

Gidon Cohen is Senior Lecturer in Politics, University of Northumbria. He is the author of, with Kevin Morgan and Andrew Flinn, *Communists in British Society 1920–1991* (London, 2005); editor of, with Morgan and Flinn, *Agents of the Revolution: New Biographical Approaches to the History of International Communism in the Age of Lenin and Stalin* (Oxford, 2005).

Parts of this research were components of the collaborative CPGB Biographical Project at the University of Manchester funded by the ESRC award number R000 23 7924. Much of the information on which this article is based came from Kevin Morgan and Andrew Flinn. The author is grateful to them, as well as to an anonymous referee, for helpful comments and ideas on earlier drafts of this article.

This article uses propensity-score matching to assess whether study at the International Lenin School in Moscow increased the chances of British students assuming the leadership roles that the Communist International (Comintern) expected of them. It deals with many of the most challenging statistical issues concerning bias from unrepresentative samples and missing data. The results indicate that, in contrast to the more extended influence of Lenin School students elsewhere, their influence in Britain appears largely to have been confined to a period before 1945.

OBSERVATIONAL STUDIES AND MATCHING ON THE PROPENSITY SCORE

Roughly speaking, in an ideal experiment, subjects are randomly assigned to treatment and control groups. When the samples are large enough, random assignment balances all of the disturbing factors, and any difference in outcome between the groups presumably stems from the treatment itself. "Observational studies" of causal relationships in human populations are attempts to maintain an experimental logic even though direct experimentation is either impossible or undesirable. In such studies, a group that has received a "treatment" of some kind is compared to a "control," or "comparison," group that has not received the treatment. In many observational studies, it is not practical to obtain random samples.[1]

Because investigators have no control over the behavior of their subjects in these situations, they must control for naturally occurring systematic differences in background characteristics, such as age or sex. For example, the now-famous observational studies of the relationship between mortality rates and smoking had to contend with such problems as the older average age of pipe and cigar smokers. One important method of dealing with such biases is "case matching," in which each member of a treatment group is matched with a similar member of the remaining population, In the case at hand, pipe smokers could be matched with the observed non-smokers and cigarette smokers who were closest to them in age to establish a matched sample permitting

1 William G. Cochran "The Planning of Observational Studies of Human Populations," *Journal of the Royal Statistical Society,* CXX (1965) 234–255; Sonja McKinlay "The Design and Analysis of the Observational Study: A Review," *Journal of the American Statistical Association,* LXX (1975), 503–520.

differences between the treatment and comparison groups to be minimized.[2]

Observational studies usually contain many different background variables, or covariates, that may introduce bias. Straightforward attempts at matching with many covariates would be impractical even with large sets of data. However, Rosembaum and Rubin showed that use of a propensity score could remove the bias arising from differences in many characteristics. Individuals' propensity scores represent the probability of them being treated based on their background characteristics. A propensity score reduces the entire collection of background characteristics into a single summary characteristic. Matching the sample on the propensity score permits differences between treatment and comparison groups to be estimated in a way that reflects adjustment for differences in all of the observed background characteristics.[3]

For example, smokers and non-smokers do not only differ in terms of age; smokers also tend to be male, to drink more alcohol, to exercise less, to be less educated, and to be more likely to work in blue-collar occupations. A proper estimate of the increased likelihood of death caused directly by smoking requires a comparison between smokers and non-smokers in which the only difference concerns smoking. First, the propensity of each individual to smoke must be calculated—that is, the probability of a particular person being a smoker, given his/her age, education, occupation, and alcohol consumption. These propensity scores can be used to create a matched sample; each smoker is matched with a non-smoker with a similar propensity to smoke. As already mentioned, the propensity score reduces the entire collection of background characteristics to a single one. Thus does matching create a balance on all the background variables, as can be verified by comparing the distribution gender, education, alcohol consumption, and so on of the matched smokers and non-smokers. With these matched samples, directly comparing the mortality of smokers and non-smokers allows a less biased estimate of the increased health risks of smoking.[4]

2 Cochran, "Observational Studies."

3 Paul Rosembaum and Donald Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika,* LXX (1983), 41–55.

4 JoAnne Micale Foody et al., "A Propensity Analysis of Cigarette Smoking and Mortality

In general, these methods should be seen as a supplement to, rather than a replacement of, existing techniques. However, so far as the analysis of observational data is concerned, propensity score methods are often thought to have a number of advantages over such model-based statistical analyses as linear or logistic regression. Most important, matching methods directly reveal the suitability of any particular database for answering specific questions by directly exposing the overlap between treatment and comparison groups. Although extrapolating characteristics between a group of forty-year-old non-smokers and a group of seventy-year-old smokers would not seem to have much to teach about the effects of smoking, model-based approaches would do so without warning, even without sufficient overlap between the groups. Furthermore, the case-matching approach does not rely on the often-unwarranted assumptions about the linearity or log-linearity of the relationship between the covariates (such as age) and outcomes (such as death) that underlie model-based methods.

The results of propensity-score analysis have the added advantage of being relatively straightforward to communicate to those unfamiliar with quantitative techniques, even though the matching process itself may be statistically complex. That it can deal with the most pressing statistical issues and effect a comparison between two groups without the intervention of recondite terminology and calculation may recommend it to those historians who are not conversant, or comfortable, with quantitative approaches.[5]

A number of articles provides tutorials for the propensity-score methods for bias removal; they are now widely used in epidemiology. The techniques are also increasingly employed in a range of social-science disciplines, including economics and sociology. They commonly assist in the analysis of large data sets, even in cases where the data "are not based on the results of carefully conducted randomized trials, but rather represent data collected through the observations of systems as they operate in normal practice." However, the observation of systems in their normal

With Consideration of the Effects of Alcohol," *American Journal of Cardiology,* LXXXVII (2001), 706–711.

5 Rubin, "Estimating Causal Effects from Large Data Sets Using Propensity Scores," *Annals of Internal Medicine,* CXXVII (1997), 757–763.

operation rarely produces "perfect" results, missing data generally being the most acute problem.[6]

ESTIMATING PROPENSITY SCORES WITH MISSING DATA The estimation of propensity scores and their use in matching when data are incomplete is based on well-developed statistical methods that stress the importance of determining patterns in what is missing. The proper approach depends on whether the mechanism responsible for the missing data can be ignored without introducing bias. Missing-data mechanisms are expendable in two situations: The first involves data missing completely at random (MCAR). When each value of a variable is equally likely to be missing, then the missing data mechanism can be ignored. In this unusual situation, deletion of the missing data will introduce no bias. The second, more plausible, assumption that enables the mechanism to be ignored involves a situation in which the values for data are missing at random (MAR), *given* the observed data. For example, people with particular characteristics—say, belonging to a particular occupational class or having a particular educational background—may be more reluctant than others to reveal their voting preferences. The telling variables in this case may well be intrinsically related to the (sometimes hidden) voting preference. Under the assumption that the missing data are MAR, complete analysis needs to take account of the relationships between the variables to ensure that statistical inferences do not introduce bias and that they make efficient use of the information available (crucial for historical research). The assumption that the values are MAR, however, is not always appropriate. When values are absent in a way that depends on the actual values of the missing data, the mechanisms are described as Non-Ignorable (NI).[7]

One way to take account of the relationship between different variables in cases of MAR is to use maximum likelihood approaches, which estimate the model that gives the maximum likelihood (ML) of generating the actually observed data, as well as the

6 Ralph B. d'Agostino "Propensity Score Methods for Bias Reduction in the Comparison of a Treatment to a Non-Randomized Control Group," *Statistics in Medicine,* XVII (1998), 2265–2281; Rubin "Estimating Causal Effects," 757.

7 Garry King et al., "Analysing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation," *American Political Science Review,* XCV (2001), 49–69; Rubin, "Inference and Missing Data," *Biometrika,* LXIII (1976); Rubin, "Multiple Imputation after 18+ Years," *Journal of the American Statistical Association,* CXXIV (1996), 473–489.

ML expectations of missing information for each individual case in the observed data. Expectation Maximization (EM) algorithms are the most common means of estimating such maximum likelihood models. Sometimes creating a full EM estimate is not practical, largely because the algorithm may be extremely slow to converge, especially in cases with many terms to be estimated relative to the size of the data set. Placing constraints on the model to be estimated—that is, reducing the number of terms to be estimated—can improve convergence properties. When constraints are placed on the interactions between variables, an Expectation Constrained Maximization (ECM) algorithm can generate expectations. In *Analysis of Incomplete Multivariate Data,* Schafer provides a clear introduction to both EM and ECM algorithms. His explanation of the computational details provided the basis for the missing-data module of the S-Plus statistical software package, which has handled the calculations in this article.[8]

Recent work by d'Agostino and Rubin, using such computations, shows how to estimate and use propensity scores for a treatment with partially missing data. They use either EM or ECM algorithms to estimate a probability of treatment (Z=1) versus control (Z=0), given both the observed data and the patterns of missing data. After the algorithm provides an estimate of the parameters for an ML model, and convergence has been reached, the model is used to generate probabilities of treatment, given that Z is missing, by running one final E(xpectation) step of the algorithm. As d'Agostino and Rubin recently stressed, the problem of propensity-score estimation is different from most other missing-data problems, since propensity scores serve only in intermediate calculations to balance populations.[9]

Following this estimation, matched samples can be chosen according to the nearest available propensity score. Hence, the first treatment case is selected and matched to the non-treatment case with the closest propensity score, and both cases are removed from

8 A. P. Dempster, N. M. Laird, and Rubin, "Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society,* XXXIX (1977), 1–38; Xiao-Li Meng and Rubin, "Maximum Likelihood Estimation via the ECM Algorithm: A General Framework," *Biometrika,* LXXX (1993), 267–278; Joseph Schaffer, *Analysis of Incomplete Multivariate Data* (London, 1997); Insightful Corporation, *Analysing Data With Missing Values in S-Plus* (Seattle, 2001).

9 D'Agostino and Rubin, "Estimating and Using Propensity Score with Partially Missing Data," *Journal of the American Statistical Association,* XCV (2000), 749–759.

the pool of subjects. The process repeats until all of the subjects have matches. The success of this matching in removing bias can be assessed by comparing the distribution of observed data for both treatment and comparison groups.[10]

THE INTERNATIONAL LENIN SCHOOL AND THE COMMUNIST PARTY OF GREAT BRITAIN Situated in Moscow, and shrouded in secrecy, the International Lenin School (ILS) was founded in 1926 as an instrument for the "Bolshevization" of the Comintern and its national sections. Roughly summarized from what may be the only authorized public description of the school, the ILS' aims were the formation of a revolutionary elite; the induction of this elite into the disciplines of Marxism-Leninism; its indoctrination with the vigilance, discipline, and commitment of the Bolsheviks; and the making of a decisive break with any lingering social-democratic traditions within the communist movement. This last goal also involved the removal of an older leadership cohort tainted by "petty bourgeois" influences and its replacement with trained Leninist cadres drawn from the core sections of the working class.[11]

Between 1926 and 1938, the school graduated about 3,000 communists, most of whom were from European and American communist parties. The Communist University of Toilers of the East, another Comintern-affiliated institution, catered to the majority of students from colonial countries. Both sets of students took courses in working-class history, the political economy of imperialism, Marxist theory, and the experience of proletarian dictatorship, complemented by practical work in a Soviet economic enterprise. In the early days of the Lenin school, students also became members of the Soviet Communist Party. The significance of their training in clandestine practices, sometimes dramatized as "Red Army training," is uncertain: No doubt, this training was of more immediate relevance to students from countries other than Britain. Students' characterizations of, and correspondence with, the Party in Britain suggest that they expected to be deployed systematically by the Party in "mass" work and to become its leading functionaries at national and district levels. Historians of the school have described it as "a long-term investment" by the Comintern,

10 D'Agostino and Rubin, "Estimating and Using."

11 J.T. Murphy, "The First Year of the Lenin School," *Communist International,* 30 Sept. 1927, 267–269.

designed to fill important positions in their national parties and ensure undeviating commitment to the Soviet *diktat*. It is difficult to think of any comparable attempt to shape a generation of national political leaders from a single center.[12]

Due to the secrecy surrounding the ILS, standard histories of the Comintern and its constituent parties could contain little real discussion of the role that it played in the contentious processes of "Bolshevization" and "Stalinization." The Comintern archives in Moscow, accessible since 1991, provide reasonably comprehensive lists and, to a lesser extent, biographical profiles for a number of different national cohorts. However, key questions regarding the school's effectiveness have yet to be fully addressed. Given that the students were selected for their potential to serve in prominent Party cadres, how much difference did attendance at the school make to their subsequent political careers and their functions within their respective parties? Did the Comintern's long-term investment, which was certainly substantial in financial and other terms, pay the intended dividends?[13]

12 See Irina Filatova, "Indoctrination or Scholarship? Education of Africans at the Communist University of the Toilers of the East in the Soviet Union, 1923–1938," *Paedagogica Historica,* XXXV (1999), 53–94. On expectations of how the British students were to be used, witness the comments of the CPGB's Moscow representative in 1931: "It is my opinion that all is going very well in the development of the Lenin School, that its line is correct and that it has been able to have very good results for all of the Parties. Why, half our organizers are L. students and this number, I should imagine, would increase in the future" (495/100/739, R. Page Arnot to Harry Pollitt, February 26, 1931, Russian State Archive of Socio-Political History [hereinafter RGASPI]). Barry McLoughlin, "Proletarian Academics or Party Functionaries? Irish Communists at the International Lenin School, Moscow, 1927–1937," *Saothar,* XXII (1997), 63; Serge Wolikow, "Internationale communiste 1919–1943," in José Gotovitch and Mikhaïl Narinski (eds.), *Komintern; L'histoire et les hommes* (Paris, 2001), 53–54.

13 As late as 1996, the Lenin School barely figures in an account faithfully representing the findings of the most up-to-date international scholarship—Kevin McDermott and Jeremy Agnew, *The Comintern: A History Of International Communism from Lenin to Stalin* (Basingstoke, 1996). For recent work, see, for example, Brigitte Studer, *Un parti sous influence. Le parti communists suisse, une section du Komintern 1931 à 1939* (Lausanne, 1994), 234–249; McLoughlin, "Proletarian Academics"; Wolikow and Jean Vigreux, "L'école léniniste internationale de Moscou: une pépinière de cadres communistes," *Cahiers d'histoire, revue d'histoire critique,* LXXIX (2000), 45–56; Gotovitch and Narinski, *Komintern,* 595–604; Cohen and Kevin Morgan, "Stalin's Sausage Machine: British Students at the International Lenin School 1926–37," *Twentieth Century British History,* XIII (2002), 327–355; John McIlroy et al., "Forging the Faithful: The British at the International Lenin School," *Labour History Review,* LXVIII (2003), 99–128; Joni Krekola, "The Finnish Sector at the International Lenin School," in Morgan, Cohen, and Andrew Flinn (eds.), *Agents of the Revolution: New Biographical Approaches to the History of International Communism in the Age of Lenin and Stalin* (Bern, 2005), 289–308.

Impressionistic evidence from a number of countries suggests that the school may have succeeded in achieving at least some of its aims. Lenin School students can be traced internationally well into the 1960s, exercising significant responsibilities either as heads of communist governments—like Yugoslavia's Marshal Tito, Poland's Wladyslaw Gomulka, and the German Democratic Republic's Erich Honecker—or as leaders of significant oppositional parties—like the general secretaries of the French and South African communist parties, Waldeck Rochet and Moses Kotane. Finland shows a high correlation between key Party leadership positions and attendance at the school. However, given the often-polarized nature of the literature, impressionistic observations have been at best inconclusive and at worst inconsistent, even within a single account. For example, although Studer seemed to describe the school as meeting its formal objectives, she also reported that the majority of Swiss students resigned from the Party within a decade of returning from the school. Though Studer attributes this defection to a lack of fitting permanent positions, the implications of such a high number of resignations for an assessment of the school's success are not discussed.[14]

A number of severely conflicting interpretations have emerged regarding the school's impact on the Communist Party of Great Britain, ranging from little to significant and long-term. The research for this article discovered the possibility of a more distinctive pattern; the influence of the cohort from the school appeared to decline during World War II. The opening of the archives in Moscow and Manchester permits an investigation of this matter in a rigorous, quantitative fashion. The evidence about the Lenin School's students used for this analysis derives from biographical information collected in the author's initial investigation of the Lenin School, extended and revised by data drawn from the biographical list of Lenin School students presented in *Labour History Review,* which is, in some respects, more complete. Subsequent data and research supply even more detailed biographical information and a number of corrections to the lists. The combined and slightly amended data provide the basis for this article.[15]

14 Studer, *Un parti,* 234–249.

15 For the view that attendance at the School made little impact, see Andrew Thorpe, "Comintern 'Control' of the Communist Party of Great Britain 1920–1943," *English Historical Review,* CXIII (1998) 637–662; Barry McLoughlin, "Proletarian Academics or Party Func-

This study assesses the specific impact of attendance at the Lenin School by comparing the careers of its students with those of a similar group of communists of the 1920s and 1930s who did not attend the school. The comparison group comes from the CPGB Biographical Database—initially constructed, and subsequently updated, by a team of researchers at the University of Manchester between 1999 and 2001—which includes records on more than 4,300 British communists. The database provides excellent and extensive information about many of these individuals, including their leisure pursuits, accents, and personal appearance in addition to more routine dates of birth, occupations, and political activities.

The project draws from a wide variety of sources, with differing levels and types of information: more than 150 life-history interviews with former Party members; short autobiographies collected by the CPGB, like other communist parties, starting from the 1930s; nominations to positions within the Party; published and unpublished biographies and memoirs; information derived from correspondence between the CPGB and Moscow; and personal archives/correspondence. The sources used for each individual's entry were recorded systematically in the database in considerable detail. All of the information discovered about every documented member of the Communist Party entered the database. As a result, the database contains an unparalleled collection of biographical information about British political activists of this period, creating the potential for quantitative investigation of many historical questions relating to the CPGB.

However, two discrete problems prevent a straightforward statistical analysis of this data. First, the individuals about whom substantial information is available are extremely unlikely to have been representative of British communists. In general, the communists described in any one source were probably from the upper echelons, or atypical in other ways. Furthermore, even though the use of sources was systematic, the choice of sources was less so, reflecting specific areas of interest—such as the ILS. The sheer

tionaries? Irish Communists at the International Lenin School, Moscow, 1927–1937," *Saothar,* XXII (1997), 63–79. For the opposite view, see Alan Campbell et al., "The International Lenin School: A Response to Cohen and Morgan," *Twentieth Century British History,* XV (2004), 51–76; Cohen and Morgan (2002) "'Stalin's Sausage Machine"; McIlroy et al., "Forging the Faithful," 121–124. The most significant differences surround alternative definitions of such terms as *full-time party worker.*

number of these sources, with their vastly divergent orientations and emphases, only compounded the difficulty of constructing an image of a representative British communist. Second, substantial evidence is missing from the database. The information about a considerable number of British communists is fragmentary, and only enough for a complete profile in a small proportion of cases. From a statistical perspective, the drawback is that both unrepresentative samples and missing data create bias; quantitative estimates derived from the database could vary significantly from the true (unknown) values in the underlying population.[16]

CREATING MATCHED SAMPLES FOR LENIN SCHOOL STUDENTS Despite the problematical nature of the available data, propensity-score methods allow a quantitative investigation of the Lenin School's impact on the British Communist Party via a comparison of the achievements of its students with those of a similar comparison group.

The assumption behind observational studies based on matched populations is that if the treatment and comparison groups are balanced in all of the relevant background covariates, their differences must stem from their respective treatments. However, the way in which historical sources define the individuals recorded in such documentary sources as the CPGB biographical database normally renders this assumption invalid. For one thing, the database includes many people who were ineligible to attend the Lenin School. For obvious reasons, those known not to have been members of the CPGB in 1926 would not be suitable matches for the students who attended the Lenin School in that year. Any individual failing to meet this condition in any year was excluded from the analysis of that year. More significant, many sources will have been constructed after the treatment has occurred, or even after the outcome of interest has transpired. In this case, a relationship between an outcome being measured and the likelihood of being included in a particular source will often exist. Since these sources affect the composition of the database and are related to outcomes, estimates of the prevalence of the outcome in both treatment and comparison groups will be biased.

16 For an overview of these issues, see Cohen, "Missing, Biased and Unrepresentative: The Quantitative Analysis of Multi-Source Biographical Data," *Historical Methods,* XXXV (2002), 166–176.

As an example, imagine the database as constructed solely from a list of Britons who died fighting in the Spanish Civil War. In this case, regardless of how closely the comparison group matches the students in background characteristics, since the entire comparison group died before 1939, none will have been prominent in the Party after 1945. If, instead, the comparison group were to consist only of those known from lists of the postwar CPGB Central Committee, every member would appear as prominent in the postwar party. The bias in these sources is glaringly obvious. In most other sources compiled after the event, similar, although less extreme biases, will be present. For example, the date at which a source was compiled may have implications for an individual's behavior. The extensive autobiographical files on British Communists stored at the Labour History Archive and Study Centre, Manchester, complied since 1942, would include only those communists of the 1920s and 1930s who chose to remain in the CPGB. Whatever the nature of these biases, and in whatever different directions they point, simply adding the populations found from such sources together gives no guarantee that the biases will be eliminated.[17]

These problems can be mitigated by restricting the sources used to define the population from which the comparison group derives to those predating the selection of students for the Lenin School. This proviso does not create a representative sample of communists from the period in question, but it testifies to the importance of ensuring that the reasons why certain groups might be mentioned in a source has nothing to do with events that occurred after the Lenin School students were selected. Thus, at least in principle, the ways in which the population is atypical can be captured in a set of variables that relates solely to the period prior to enrollment in the school. That this population overrepresents those who belonged to the CPGB Central Committee during the 1940s is inescapable, because, for example, active communists of the 1930s were both more likely to be recorded in contemporary sources and more likely to be prominent during the 1940s. However, this study's findings assume that those who went onto the CPGB Central Committee during the 1940s are not overrepresen-

17 William Rust, *Britons in Spain: The History of the British Battalion of the XVth International Brigade* (London, 1939).

ted *given* the variables being measured, such as communist activity during the 1930s.

Restricting the population will lead to unbiased estimates of outcomes only if the comparison of the groups balances on an appropriate set of variables. This condition depends on a theoretical model of what factors distinguished Lenin School students from other communists or what factors were important in predicting the attainment of leadership positions within the CPGB. As in most areas of historical research, previous attempts at statistical investigation are nonexistent; any such model must be based on a formalization of qualitative evidence. Precise instructions from the Comintern about who was worthy of being sent to the ILS varied, and the loosely followed criteria frequently led to complaints that "almost all parties fail[ed] to fulfil completely ECCI's [Executive Committee of the Comintern] directives on the requirements expected of everyone entering the school." However, the principles remained fairly constant, also providing a reasonable indication of who might have been expected to form the next generation of communist leaders.

As Studer summarized the qualifications, the students were to be of working-class or peasant origin, in perfect health, aged no more than thirty-five, members of the Communist Party or Young Communist League for at least a year, and proven in some form of class struggle. Other sources, such as the testimony of the students, state that factors, such as gender, also played a part in selection to both the Lenin School and leadership positions within the Communist Party. Although selection on these factors may have been imperfectly fulfilled, the Lenin School students were certainly not typical communists in these respects.[18]

For various technical reasons, a model that covers all of the potential difficulties is impossible to create. Based on the Comintern criteria, this study gives priority to four factors that appear to have been critical in the attainment of leadership positions within the CPGB: level of previous communist activity, gender, class, and generation. Table 1 contains the full set of variables used to represent these concepts and to estimate the propensity score. Many of

18 ECCI Political Secretariat, Letter to the sections on the selection of students for the International Lenin School for the new school year, 531/1/18/1, RGASPI; Studer, *Un parti,* 235. See also instructions for acceptance into the ILS, May 1932; Cohen and Morgan, "Stalin's Sausage Machine," 336–341.

Table 1 Covariates Contributing to the Estimation of the Propensity Score

FACTOR	VARIABLE	DESCRIPTION	CODING (FOR CATEGORICAL DATA)
Gender	Sex	Sex	0-Female 1-Male
Record of communist activity (see also Generation)	CPAct	Highest level of CPGB activity attained	1-Local 2-District 3-National 4-International
	CPActYear	Highest level of current CPGB activity	1-Local 2-District 3-National 4-International
Generation	JoinCP	Date of Joining CPGB	
	DecadeJoin	Date of Joining CPGB	In decades
	DOB	Date of Birth	
	DecadeBirth	Date of Birth	In decades
Class	OccClass	Occupational class	
	Education	Highest level of formal education	1-Elemenatry/Primary 2-Secondary 3-Higher/Further

these factors, like communist activity and class, are dynamic, in that they can change over time. Because what the students were doing immediately before attending the Lenin school is significant, the calculations take this date into account. In order to create data suitable for matching students with non-students, separate calculations on each of the variables were performed for every year in which students were sent to the Lenin School. Non-students were considered as potential matches for students according to individual data at the time when students were about to enter the ILS.[19]

19 Balancing all of the factors did not prove possible because no systematic information was available about certain factors, such as state of health. For other factors, such as prior non-communist political activity or trade union activity, about which missing data was more likely to mask a negative response, the assumption that data were MAR was unreasonable.

Much of the data may be described as ordinal—for example, the hierarchy of positions within communist organizations from local, through district and national, to international levels. The missing-data techniques used in this article make no provision for the direct inclusion of ordinal data. However, as King et al. recommend for situations in which the data distribution is (or can be transformed to be) approximately normal, "ordinal variables should be coded to be as close to an interval scaling as information indicates." The occupational data in this article was coded in accordance with ILO's International Standard Classification of Occupations (ISCO-88). This occupational information was combined with other evidence about employment status, roughly indicating whether individuals were employees, supervisors, managers, or self-employed proprietors, to determine a standard measure of social class—one of the five groups (one subdivided) until recently used by the British government as an official classification. Individuals were assigned to the occupational class in which they had spent the most time until the date under consideration. Since this occupational data is distributed in an approximately normal fashion throughout the CPGB data, it merits treatment as continuous. The other nominal data in the data set was distributed in such a way that it could not be considered approximately normal and had to be treated as categorical data.[20]

As previous studies have noted, one major problem with prosopographical data is its lack of precision, especially with regards to dates. On the CPGB Biographical Database, a date of birth might be recorded exactly—as, for example, 20/10/1892—or much less precisely—as, for example, 1901–3, early 1890s, 1890s, and so on. To restrict the study only to exact birth dates, or at least birth years, would ignore the substantial amounts of "fuzzy" information in the database and suggest greater uncertainty than is actually the case. In order to overcome this problem about dates, the student and comparison groups are balanced on both exact dates as a continuous variable and specified date ranges as a categorical variable.[21]

20 King et al., "Analysing Incomplete Political Science Data," 49. See www.iser.essex.ac.uk/ons/ns-sec/derivations.php.

21 Evan Mawdsely and Stephen White, *The Soviet Political Elite From Lenin to Gorbachev—The Central Committee and Its Members, 1917–1991* (Oxford, 2000).

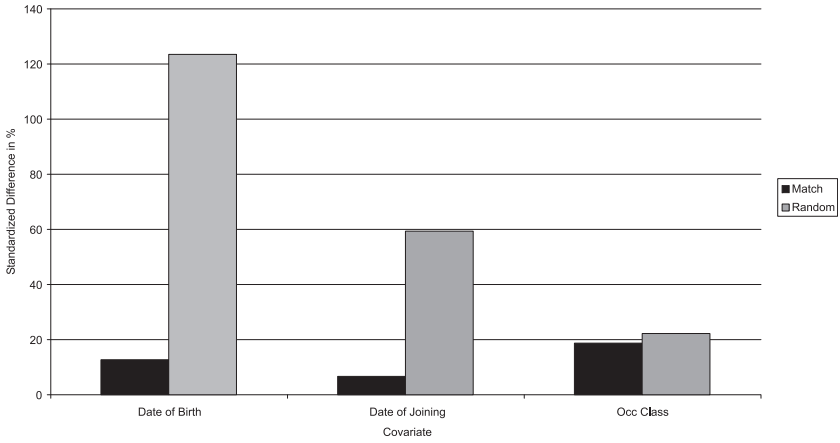
The propensity score—the probability of attending the International Lenin School given the set of background characteristics—was estimated in accord with the method described by d'Agostino and Rubin to compensate for missing data. Given the ratio of cases to necessary parameters in a full model, the EM algorithm was inappropriate. Instead, a constrained maximization using the ECM algorithm was specified. This model included all one-way and two-way interactions on the categorical variables, but did not include three-way or higher-level interactions. Once the ECM algorithm had converged, the expected value of the attendance at the Lenin School was calculated from the parameters of the model. Separate estimates of the propensity score were obtained for each intake to the ILS. Each student was matched to the non-student with the nearest available propensity score from the unmatched population in a particular year until the matched sample was complete.

ASSESSING THE BIAS REDUCTION The relative success of an estimated propensity-score model can be assessed by comparing the balance on the observed values of the covariates–in this case, the background covariates for students and matched populations. For continuous covariates, the standardized difference in means was the measure of success—zero indicating a perfect balance—whereas for categorical covariates, it was the available case proportion for each value of the variable—a value identical to the student group indicating a perfect match. The extent of the bias reduction achieved by the propensity-score matching is evident in a comparison of these results to those involving a randomly selected sample of the database meeting the logical criteria for matching. The balance on continuous covariates is shown in Figure 1 and the balance for categorical covariates in Figure 2.[22]

Studies in medical and epidemiological literatures often feature propensity-score methods alongside the results from randomized experimental studies. In such cases, the existence of alternative quantitative methods of data collection and investigation, which may have demonstrably less bias, requires acceptance of only small levels of bias. Both theoretical and applied studies tend

22 The standardized-percentage difference is defined as the mean difference between groups as a percentage of the standard deviation $\{100(\bar{x}_p - \bar{x}_t) / \sqrt{[(s_p^2 + s_t^2) / 2]}\}$.

Fig. 1 Comparison of Standardized Difference (in %) for Covariates Between ILS Students and Matched Groups

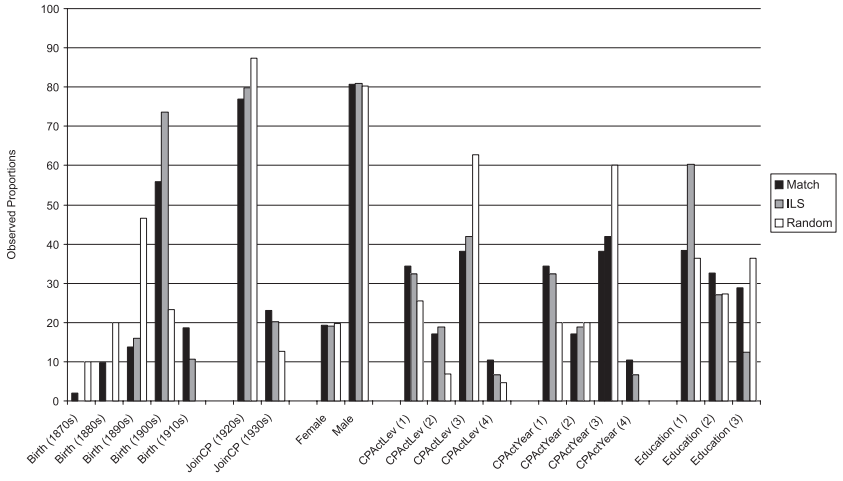


to suggest that a standardized difference in continuous covariates of less than 10 percent can be considered small. However, any assessment of bias reduction is a matter of judgment, dependent on the alternatives available. In historical studies based on multisource biographical data, two factors suggest the possible need for greater toleration of bias: (1) the scarcity of much historical data, which has an inescapably negative effect on accuracy, and (2) the limitation on the types of investigation that data collection may impose. Whenever these conditions apply, particularly the first, historical studies may well have to accept a greater toleration of bias. Both factors, however, suggest the need to consider the congruence of results with findings of other, perhaps more qualitative, forms of investigation.[23]

Figures 1 and 2 show that bias reduction was most obviously effective with regard to dates of birth and dates of joining the CPGB, both of which showed the greatest initial bias. Bias on date of joining was reduced to below the 10 percent of the standardized difference level considered small in epidemiological studies. Bias was also largely eliminated in level of past CP activity, the variable

23 D'Agostino and Rubin, "Estimating and Using," 756–758; S-L.T. Normand et. al., "Validating Recommendations for Coronary Angiography Following Acute Myocardial Infarction in the Elderly: A Matched Analysis Using Propensity Scores," *Journal of Clinical Epidemiology,* LIV (2001), 390.

Fig. 2 Observed Proportions of Categorical Covariates for Students, Matched and Random Samples

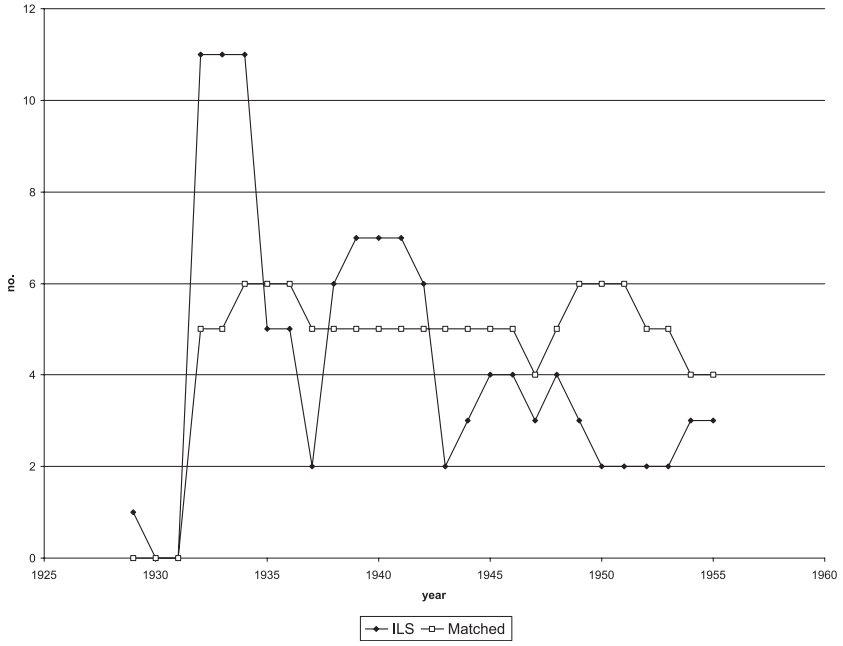


that intuitively seems the best predictor of future level of activity. However, the model was less successful in balancing the variables of class, OccClass, and Education. Despite the overall success of the propensity-score model in reducing bias in this study, and the concession that different levels of residual bias may be acceptable in some historical studies, the areas where substantial bias remains are important to bear in mind. Any differences in outcome between the student and matched sample may be explained by slightly different social origins.

RESULTS The creation of matched populations makes the presentation and analysis of the Lenin School's impact on the careers of British communists comparatively simple. By the 1930s, membership in the executive committee was probably the best indicator of status within the Party, and it was consistently associated with leading positions at the Party newspaper and in the Party's largest districts. Tests were also run on the data to see whether any pattern involving Party organizers and secretaries was discernible in any of the districts, including the smaller ones without representation on the executive committee.

The simplest way to examine the trajectories is through graphical presentation. Figure 3 shows the numbers of ILS and

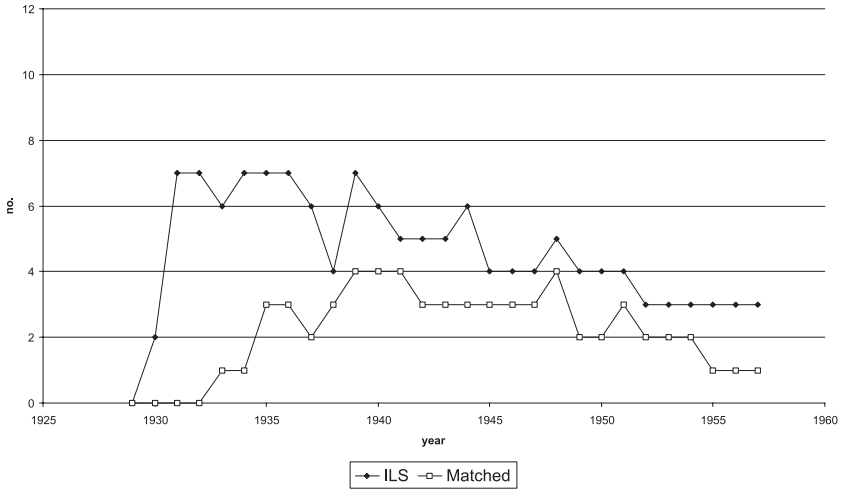
Fig. 3 ILS and Matches on Executive Committee, 1929–1957



matched groups on the executive committee over time. Figure 4 shows the patterns in relation to the numbers of district organizers. Differences are in evidence, both with respect to the clarity of the patterns and the years. However, in both cases, a possible trajectory is suggested first in the 1930s when more Lenin School students seem to have occupied higher positions within the CPGB and later in the postwar period when the trajectory for both students and the matched groups appear to have been approximately the same.

Although it provides a useful starting point, this graphical analysis should be approached with considerable caution. First, although the data are certainly informative, the imprecision with respect to dates, particularly relating to district secretaries, means that figures for particular years cannot be regarded as definitive. Second, the data set is not large enough to enable questions about distinct periods to be addressed as neatly as the graphical presentation might suggest. Third, the presentation of aggregate figures for

Fig. 4 ILS and Matched Samples as District Secretaries/Organizers, 1926–1957



each year masks a considerably faster turnover of ILS students than of matched samples in both the executive committee and the position of district organizer/secretary. Finally, any analysis dependent on inspecting graphical differences is prone to difficulty. Rarely will two samples ever produce exactly the same results, and it can be difficult to determine whether different trajectories are due to expected fluctuations or whether the differences are greater than would be expected by chance alone.

A more rigorous alternative, which circumvents these difficulties, is to compare directly the number of the students and matched samples who eventually achieved positions of prominence within the Communist Party. Table 2 shows the respective numbers elected to the executive committee after graduation during the period to 1945 and later during the postwar years. Table 3 shows the number of matched samples who worked as district organizers or secretaries during the same periods. Whether attendance at the Lenin School was related to the attainment of a prominent position can be determined by testing the hypothesis that no relationship existed.

Inspection of tables 2 and 3 would seem to indicate substantial differences between the student and matched populations during

Table 2 Number of Students and Matched Samples on the Executive Committee during Two Time Periods

	STUDENT (N = 159)	MATCHED (N = 159)	CHI-SQUARE
1926–1945	20	9	0.041
Post-1945	8	9	0.808

Table 3 Number of Student and Matched-Samples District Organizers/Secretaries in Two Time Periods

	STUDENT (N = 159)	MATCHED (N = 159)	CHI-SQUARE
1926–1945	23	7	0.003
Post-1945	5	5	1

the early period but little difference during the postwar years. The patterns can be investigated more rigorously by using a chi-square test. For example, the chi-square test for the executive committee during the period to 1945 gives the probability that a difference as great as 20 of 159 students and 9 of 159 matched cases would have occurred if cases had been randomly assigned from 318 individuals with 29 on the executive committee. The chi-square result of 0.041 is statistically significant at the 0.05 level in the sense; that students' representation on the executive committee would exceed that of matched samples to this extent by chance alone has less than a 5 percent probability.

The results for the post-1945 period suggest that chance could well explain the small difference between the two groups (eight students versus nine individuals from the matched population) in election to the executive committee at that time. The results for service by members of the two groups as district organizers and secretaries show the same pattern, but in an even stronger fashion—the difference being significant at the 0.01 level (a less than 1 percent likelihood that the differences are due only to chance). In the post-1945 period, identical numbers of students and matched samples were district secretaries. Hence, the findings generally present a consistent picture. During an initial period, lasting approximately until the end of World War II, attendance at

the Lenin School significantly increased an individual's chance of achieving a position of prominence within the CPGB. After this period, however, attendance at the ILS appears to have had no significant effect on the attainment of these positions.

In a context as controlled and politicized as the selection of Communist Party leadership, these surprising results about the limited duration of ILS students' prominence have significant ramifications concerning the movement. At the most basic level, they suggest that the training provided, with its heavy emphasis on theory and "conspiracy," was not well tailored to the more open campaigning requirements of a legal communist organization like the CPGB. Traditionally, however, the leadership of each Comintern section was seen as a "self-perpetuating clique of obedient leaders" whose prior political conditioning gradually superseded the continuous direction of the Comintern as the instrument of Russian control: "Like the Duke of Newcastle's bishops, they could be expected to worship their maker." In reality, the decay of the Comintern's formal structures after the mid-1930s left no alternative mechanisms to secure the positions of these "tested henchmen" within the Party apparatus. At a national level, considerable powers of patronage devolved to general secretaries or secretariats; they acquired the capacity to advance, block, or divert careers without much interference from an external authority. In Britain, graduates of the Lenin School may have been identified with what came to be regarded as the School's "sectarianism." Older Party leaders lacking a formal Moscow training may have regarded them as potential challengers. A "self-perpetuating clique" of British communists, formed prior to the Lenin School, may have begun to promote its own recruits from the 1930s, seemingly at the expense of those whom the Communist International had been formally grooming for leadership roles.[24]

In the past few years prosopography, the study of collective biography, has undergone a resurgence. Major projects dealing with a wide range of historical periods, from ancient Greece to the mod-

24 Henry Pelling, *The British Communist Party: A Historical Profile* (London, 1975; orig. pub. 1958), 73–75; Studer, "More Autonomy for the National Sections? The Reorganisation of the ECCI after the Seventh World Congress," in Mikhail Narinsky and Jürgen Rojahn, *Centre and Periphery: The History of the Comintern in the Light of New Documents* (Amsterdam, 1996), 107.

ern world, have recently emerged, often collecting unprecedented amounts of data and foretelling significant analyses. Relying entirely on qualitative judgment to gauge patterns within such large data sets is an invitation to inaccuracy; collective biography benefits from some form of statistical analysis.[25]

One possible approach is to treat the collective-biographical data as a regular, complete probability sample of the population as a whole. At best, this strategy involves the explicit, and risky, assumptions that a large sample can substitute for a representative sample and that the patchy nature of the data can be ignored. The biases found in the CPGB Biographical Database data—concerning both the individuals described in the sources and the varying types of information about them provided by the sources—are the same as those in almost every other complex biographical database. Given the obvious problems entailed, many scholars believe that the only alternative to assuming regular, complete, and representative samples is to reject the possibility of any systematic statistical investigation of collective biographical data. One aim of this article is to suggest that techniques taken from scientific observational studies, such as propensity-score case matching, can help historians to circumvent these unsatisfactory alternatives.[26]

Propensity-score matching is not a magical technique for historical databases; problems can arise at many stages. The method requires a relatively large initial data set containing information about a substantial group of individuals. In the case examined in this article, the database had the requisite background information, such as prior political activity and educational history. The individuals under examination could be identified as either "treated" or "untreated' according to their having attended the International Lenin School or not. They also could be recorded as having experienced a measurable outcome, election (or non-election) to the CPGB executive committee. At the risk of creating samples with considerable residual difference, treated and untreated indi-

25 R. Kilpatrick et al., "The Daidalos Project," *Literary & Linguistic Computing,* III (1997), 177–184; Katharine Keats-Rohan, "Historical Text Archives and Prosopography: The COEL Database System," *History and Computing,* X (1998), 57–72; John R. Martindale (ed.), *The Prosopography of the Byzantine Empire I (641–867)* (Aldershot, 2001); Gotovitch and Narinsky, *Komintern*.

26 Anna Hillyar and Jane McDermid, *Revolutionary Women in Russia, 1870–1917: A Study in Collective Biography* (Manchester, 2000); Barbara Evans Clements, *Bolshevik Women* (Cambridge, 1997); Mawdsley and White, *Soviet Political Elite*.

viduals should show a reasonable overlap in measured background variables.

Another problem is that the probability of inclusion in certain historical sources may be related to the outcomes of interest, as inclusion in the list of those who died in the Spanish Civil War necessarily precluded attainment of any office in the CPGB after 1945. This pitfall may require that limits be placed on the data derived from the database in question. The step taken herein was to restrict the population to those individuals who appeared in sources created before a particular date or event during their lifetime. The greatest limitation of this methodology is that bias can be removed only from background factors that have been explicitly considered and measured. Furthermore, historians who use case-matching methods, unlike scientists who use them, have no recourse to randomized experimental methods. They will have to remain sensitive to the findings from other, often more qualitative, forms of investigation.

Despite these issues, techniques borrowed from scientific observational studies have great potential for historical research. Propensity-score techniques have proven their worth. They can deal rigorously with such problems of central concern to historians as unrepresentative samples and missing data. Methods, based on the comparison of matched samples, may also be statistically more appropriate to matters of causation than, say, regression, particularly because they directly ask questions about the suitability of the data. The logic of comparing otherwise similar groups who have undergone different "treatment" is relatively easy to grasp, and the results of the matching and the final comparisons, and the operational details of the methods, need not be prohibitively complicated to understand. Recent developments in prosopography and historical computing have suggested, but too infrequently delivered, a wider application of quantitative methods. The techniques developed in scientific observational studies provide one possible route to realising this potential.