

Appendix A

Description of Data Sources

Sources of Data

NIH Trainee and Fellow File (TFF). The Trainee and Fellow File includes records for all individuals who have received NIH training grant or fellowship support. Appointment and termination records completed by training grant directors and information on individual fellowship awards provide the data contained in this data base. Maintained and updated regularly by Quantum Research Corporation, this file was used to identify individuals who had received nine or more months of NRSA support at either the pre- or the postdoctoral level, along with other forms of NIH training support.

Doctorate Records File (DRF). The Doctorate Records File is based on the Survey of Earned Doctorates (SED), which had been previously conducted by the National Research Council and is currently being carried out by the National Opinion Research Center. Sponsored by several Federal agencies (including the NIH, this survey has been a census of the research doctorates awarded by U.S. academic institutions during the academic year (beginning July 1 of one year and ending June 30 of the next year) since 1920. Included in the questionnaire, which is filled out by graduates at the time that they complete all requirements for their doctorate, are questions covering: demographic characteristics (gender and race/ethnicity); educational history (e.g., undergraduate, master's, and Ph.D.-granting institutions); degree characteristics (e.g., field and time-to-degree); financial support for graduate study; and employment or postdoctoral study plans. The response rate for the survey is extremely high (about 95% of new Ph.D.s), and commencement programs and institutional sources provide basic data on the remaining 5%. The data in this file provided the majority of information on study and comparison group characteristics (e.g., race/ethnicity), sources of graduate student support, time-to-degree, and postgraduation plans.

Survey of Doctorate Recipients (SDR). The Survey of Doctorate Recipients is a biennial, panel survey conducted by the National Opinion Research Center (and previously carried out by the National Research Council). Initiated in 1973, this data collection effort is sponsored by the National Science Foundation, the National Institutes of Health, and other Federal agencies. The survey population consists of research doctorate holders in the United States, and the sample for each survey is modified by adding individuals from the most recent two-year cohort of new doctorates and eliminating those who have retired or exceeded 70 years of age. Overall, the SDR sample includes about 8% of Ph.D. awardees. In the spring of each survey wave, the questionnaire is mailed to each sampled person (nearly 50,000 in the 1995 SDR). Covered are demographic and employment characteristics of the individuals, with surveys in a particular year also incorporating items on special topics (e.g., the 1995 SDR covered current and previous postdoctoral training). Beginning in 1991, the SDR underwent several changes in the sampling, questionnaire, and follow-up design. In 1991, the sample was reduced due to cost constraints, and a series of concerted follow-up efforts (including the use of computer-assisted telephone interviewing) were initiated. Some questions also were reworded. In contrast to the usual 24 months between surveys, the intervals also were nonstandard between the 1989 and 1991 and between the 1991 and 1993 survey waves (31 and 20 months, respectively). Because the increased follow-up efforts were successful in improving the response rates, the analyses in this report include only those to the mail questionnaire. This was done in order to help make the data across surveys more comparable. At the same time, caution should be used in comparing data from pre-1991 surveys to those conducted during the 1991-95 period, given the other changes identified.

Consolidated Grant Applicant File (CGF). This file is compiled by the NIH and includes records for all applications for NIH research grants and contracts. Each record includes information on the name of the principal investigator, the type of application (e.g., investigator-initiated or program project grant), the

status of the application (funded or not funded), and the dates of review and award. The file is maintained and updated regularly by Quantum Research Corporation.

NSF Master Database of Proposals and Awards. This file is quite similar to the Consolidated Grant Applicant File maintained by the NIH in structure and contents. Records cover all applications for NSF research grants and contracts and include information on the name of the principal investigator, the type of application, and its status. Quantum Research Corporation also is responsible for maintaining this data base.

Institute for Scientific Information (ISI) Publication and Citation Data Base. Publication records were available for the 1981-82 and 1987-88 Ph.D. cohorts sampled in the 1983 and 1989 Surveys of Doctorate Recipients. These data were compiled by the Institute for Scientific Information; each record consisted of an article that had been published between 1981 and 1995 in one of the journals indexed by the ISI. Information on articles included the journal, publication year, authorship, and number of citations it had received through 1995.