Appendix B
Technical Notes

A Small-Scale Examination of the Quality of the Publication Data

Because several individuals can share the same first initial and last name, a verification procedure was performed so as not to incorrectly attribute an article to an individual. The procedure involved matching certain information on an article's authors reported by the ISI with information provided by the Survey of Doctorate Recipients (SDR), which comprised the frame for the publication samples. Characteristics of a publication's authors that were provided for each IS record included first initial, middle initial, and last names of up to 15 authors, along with their institutional affiliations, cities, states, and zip codes. Address data were also provided by SAR staff for the individuals included in the samples.

An article was attributed to an individual if it met certain criteria; (1) the first initial, middle initial (where available), and the last names of an article's author matched those of the individuals as reported in the SAR; (2) the state of residence reported in the SAR and the state for one of the article's authors matched; (3) the zip codes in the SAR and the IS were within plus or minus 100 digits; and (4) the publication year of the article was within five years of the survey year. Applying this algorithm to the Institute for Scientific publication data files was conducted by Quantum Research Corporation.

Small-scale studies assessing the quality of these data were performed for random samples of 100 from each cohort. For those publications that had been attributed to the individual by the computerized matching process, those publications were located and copied and the relevant information (e.g., full author name, institutional affiliation, and mentioning of a training grant as sponsoring the research) was extracted and compared to information on these individuals obtained from the 1981-95 waves of the SAR and the Trainee-Fellow File. In cases where the survey data were incomplete, additional sources were used. For example, if the article of interest had multiple authors, other articles attributed to the person were located and assessed as to which appeared to be accurately attributed (e.g., the sharing of common coauthors). The results indicated that 82 percent of the articles were correctly identified as authored by a person in the sample; this percentage increased to 85 percent when cases were included that appeared to be reliably attributed but the level of certainty was lower (e.g., due to missing information for validation). The percentages for the NRSA, NIH training institution, and non-NIH training institution groups were 85, 78, and 77 percent, respectively.

A sample of 40 individuals who were included in both this study and the MSTP evaluation (which used curriculum vita) was also used to examine the quality of the article attribution process. Articles assigned to the individual by the search strategy used on the ISI data were compared to those listed on the vita. Of the publications identified in the ISI data and attributed to the individuals, 87% were correctly assigned (i.e., listed on the vita). Of the 507 articles identified on the vita, the matched records from the ISI data identified 71 percent. The correlation between the total number of articles found by the ISI and those on the vita for the same time period was 0.84.

Information on Multivariate Analyses

As described in Chapter 1, the logic of the analysis was as follows:

(1) *Outcomes for NRSA trainees and fellows were contrasted separately with those of each comparison group*. Rather than omnibus tests across all three groups, pairwise (or two-group) comparisons were performed. This approach was judged to be more informative, given that the

ways in which NRSA predoctoral trainees and fellows differed from graduates of the same institutions were not the same as those relative to Ph.D.s from universities without NRSA training grants.

(2) *Those between-group differences warranting discussion satisfied two simple criteria — statistical significance and magnitude.* This strategy was used to better identify what represents a "potentially meaningful" group difference between NRSA recipients and their comparison group counterparts -- i.e., one that is not further adjusted statistically to control for other confounds and rival explanations. For example, does a group difference of five percentage points or two journal articles, even when statistically significant, warrant further examination, particularly when it is based on large samples? Consequently, a group difference had to be statistically significant ($p < 0.05$) and translate into an effect size of at least 0.20.[1]

(3) *Where possible, between-group differences satisfying these two criteria were further examined as to the plausibility that other variables were primarily responsible for the observed results.* These additional analyses were of two types. First, when sample sizes permitted, simple comparisons of outcomes were conducted for specific subgroups. These typically involved different disciplinary clusters due to the fact that career opportunities and paths vary among fields. Second, multivariate analyses were performed in order to better ascertain the contribution of NRSA predoctoral support relative to other variables that may be related to career progress.

The selection of these variables depended both on the results of prior research and the availability of data on the relevant measures. For example, studies on academic employment among scientists have shown that prior postdoctoral training, along with the reputation of one's doctoral institution, influenced obtaining a faculty position and the type and location of this job. Also included in the analyses were background variables on which the groups involved in the comparisons initially differed (e.g., field of training). This set of predictors was used in the multiple and logistic regressions used to predict outcomes and determine whether NRSA predoctoral training report helped to explain the observed group differences. It should be noted that the final regression models included all the predictors regardless of their significance levels; this decision was based on the belief that it was important to document not only what contributed to an outcome but also what did not. This is because the role of variables found to be instrumental in earlier studies on scientific careers may have increased or decreased over time (e.g., gender differences on some outcomes have disappeared with the growth of women Ph.D.s).

Within-NRSA study group differences. Not only is it important to identify differences in career outcomes between those who did and did not receive NRSA predoctoral training support, but it is also

[1] As a reminder, Cohen (1988) established the criteria as: (a) an effect size of |0.20| is considered as indicative of a small effect; (b) an effect of |0.50| - |0.79| is viewed as a medium effect; and (c) effect sizes that are at least |0.80| are considered as signaling a large effect. Stated in other ways, an effect size of 0.20 corresponds to a correlation of 0.10 between two variables (e.g., receipt of NRSA predoctoral support and subsequent award of an NIH research grant), an r^2 of 0.01 (e.g., NRSA training support explains 1 percent of the differences between groups in being awarded NIH research funds), or a 10 percentage point difference between two groups on success rates (e.g., 55 percent of NRSA trainees and fellows versus 45 percent of their graduate student counterparts had obtained NIH research grants). Although the text does not report significance levels and effect sizes for observed group differences, the interested reader can find them in the relevant appendix tables.

important to examine how variations in NRSA training support are related to career progress. This is particularly relevant to those outcomes that are closest in time to training experiences. These variables were intended to describe aspects associated with NRSA support, including: Thus, multivariate analyses were performed to explore the nature of the variation within the NRSA study group participants and its relationship to a set of variables used to characterize the nature of the NRSA training support. The strategy for selecting and entering possible predictors was the same as that previously described for the pair-wise comparisons; however, the set of variables examined were those related to NRSA support (e.g., duration and timing) rather than group membership.

(a) *Length of NRSA predoctoral support.*

(b) *Receipt of both an NRSA traineeship and fellowship.* Although only a small minority received both awards, obtaining a predoctoral fellowship may signal individual commitment to and success in progressing through a doctoral program.

(c) *Receipt of MSTP support.* Not only is MSTP support typically provided for longer periods of time, but the requirements of earning both degrees may affect degree progress.

(d) *The "maturity" of the training grant.* The number of years that the training grant had been in operation at the institution is one indicator of the youth of a training program, which may not have all of its components (e.g., special classes) fully "in place" during its earliest years. For example, whereas 38 percent of former biomedical trainees were supported by a training grant that had been in operation for 5 or fewer years, this was true for 70 percent of the former trainees in psychology, sociology, and anthropology. Another 25 percent of behavioral science trainees received their support from grants that had seen 6-10 years of funding, and the remaining 5 percent were appointed to grants that had been in place for more than a decade (as contrasted to 20 percent of the former trainees in the biomedical disciplines). The number of doctorates produced by a grant also was chosen, based on the fact that those with larger numbers of Ph.D. graduates may have a "critical mass" -- a factor that may influence degree progress. Approximately half of the former trainees had been supported on training grants that had produced 30 or fewer biomedical sciences Ph.D.s. In contrast, this was true for nine-tenths of behavioral science trainees.

(f) *Having one's first support begin within the first three years of entering graduate school.* This variable could be an indicator of recognized talent early on (i.e., training directors select the best incoming students) and also signify the opportunity to focus entirely on graduate classes and requirements without being disrupted by teaching assistantship or other responsibilities that may not be directly related to one's graduate work and research interests.

(h) *Being supported on a training grant that was awarded to a program in a medical school or college of arts and sciences.* Questions have been raised about differential outcomes for these two types of training environments. It also is the case that the context and available sources of graduate student support differ, with teaching assistantships much less likely to be used by medical schools.