

## **Sampling Appendix A: Sampling Technical Appendix**

### **1. Overview**

The After the JD study utilizes a two-stage scientific sampling approach, first selecting among metropolitan areas (or non-metropolitan portions to states) to obtain a wide geographic and population size distribution of geographic areas, and second selecting among individuals who meet individual eligibility criteria.

### **2. Target Population**

The target population is persons who first became eligible to practice law in calendar year 2000. The year 2000 is selected as a bar entry year so as to provide a small cohort with similar times of entries into the legal labor market which we will be able to track over time.

(We might have alternatively defined the relevant population as all persons who graduated from law school in a given period. But our focus is on persons who had sufficient intent to practice law that they took the bar exam, and who were admitted to a state bar so that they were in fact eligible to practice law.)

### **3. Sample Population**

The sample population of the After the JD study are persons who **first** became members of a state bar in CY 2000, and who graduated from law school in the period July 1, 1998 through June 30, 2000.

Who among year 2000 admittees is excluded by this definition of the eligible population? First, persons who graduated from law school after June 30, 2000 are excluded. We want to be able to ensure that persons in the sample became bar members at least two years before they are interviewed in the spring and summer of 2002, so we exclude those few persons who might have graduated at the end of the summer 2000 term and were allowed to take the July 2000 bar exam early.

Second, our exclusion of persons who graduated from law school prior to June 30, 1998, excludes persons who had a job of two or more years after law school but before admission to the bar. This restriction should not exclude persons who were judicial clerks for one or two years prior to joining a bar, and it should allow inclusion of persons who took the bar exam three or four times before passing it and joining the bar.

Third, a very small number of persons who did not graduate from a law school, and who entered the bar through a “law office study” program, will also be excluded under this definition. (In 1999, 11 of the 47,000 persons passing a bar exam were not law school graduates).

Finally, our definition of the eligible population excludes persons who joined a given state bar in 2000, but who had previously been members of another state bar. (If they joined the

first state bar in 2000 as well, they would be eligible in that first state). This restriction is intended to include in the cohort only persons who became eligible to practice law at about the same time. It should also be noted that our definition of the eligible population excludes persons who entered the bar on motion, due to their previous membership in another state bar. Except in DC, such motion admissions require 3-5 years of practice in the previous state. DC presents special problems, which we discuss below.

#### **4. Estimation of the eligible population**

In order to properly select the first stage sample of geographic areas, we need to estimate the number of eligible persons in each geographic area of the United States. This is a non-trivial problem with the data available to us prior to the sample selection of individual within areas. Unfortunately, we have incomplete information on numbers of jobs taken by recent law school graduates and a mismatch of those numbers to the numbers of persons admitted to the bar by state. Appendix B explains the problems with the existing sources of data, and our attempts to reconcile the job taker with the bar admissions data. In brief, we used data on job takers to determine intra-state distributions and revised admissions data to determine statewide numbers of eligible persons. The result of our computations was an estimated 48,695 eligible CY 2000 bar admittees.

#### **5. Clustering eligible persons by geographic area**

Because eligibility to practice law is governed by individual state statutes and rules, and admission to the bar requires a two-step process of passing a bar exam administered by state bar examiners and registration with state registrars, our sample is naturally clustered by state. In addition, lawyers tend to locate disproportionately in larger metropolitan areas.

For reasons of sampling efficiency, we created clusters of geographic areas to use as Primary Sampling Areas (PSAs). We relied on the metropolitan areas defined by the US Office of Management and Budget (June 30, 1996 definitions) and the non-metro remainders of states. One exception to the OMB definitions was made: where an MSA crossed state lines, it was split into separate MSAs in each state, since there would be a separate bar admissions process for each state. However, the District of Columbia MSA did include the relevant portions of DC, Maryland, Virginia, and West Virginia. There were a possible 346 metropolitan areas and 50 non-metro remainder areas under this process, for 396 possible areas.

We obtained data on the geographic location of jobs taken by 1999 law school graduates from the National Association for Law Placement. This information, organized as the number of persons who reported taking jobs in each city and state, was recoded to derive totals for metropolitan areas, non-metropolitan areas, "city not recognized", and "city not specified" elements within each state. The "city not recognized/specified" job takers were re-allocated proportionally among the metro areas and the non-metro remainder in each state. These 1999 job taker numbers were one initial basis for our estimates of the number of 2000 eligibles. Because some small metropolitan areas in fact had an estimated zero eligibles, the actual number of geographic areas into which the eligible population was clustered was 348.

## 6. Stratification

It is our intention to interview persons in a wide variety of geographic areas and in metro and non-metro areas with widely varying numbers of new attorneys. A nation-wide sample of persons joining the bar would be ideal for this purpose, but is impractical for two reasons. First, it would require us to obtain names and addresses of CY 2000 entrants to the bar in all 50 states and DC, and it seems likely that at least a few states will effectively refuse. Second, it would be very costly for us to conduct face-to-face interviews (for the roughly 10% sample of persons who complete a mail questionnaire or telephone interview) with persons throughout the country, so we need to limit the number of geographic locations at which those face-to-face interviews will be conducted.

The practice of scientific sampling in two stages – first selecting a sample of geographic areas, and then a second sample persons within those areas – allows us to obtain a nationally representative sample of persons in a relatively limited number of geographic areas. By randomly selecting a limited number of the geographic areas, within carefully defined strata, to be representative of the nation as a whole, our sample will allow us both to generalize accurately to the entire class of 2000, and also to be able to report on the situation of new attorneys in many of the larger geographic areas in our study.

The clusters of geographic areas (PSAs) were stratified along two dimensions: 1) number of estimated eligibles, in order to ensure the representation of legal employment markets of various sizes, and 2) Census Region/Division, to ensure wide geographic representation. With these two dimensions, we formed 18 strata.

The geographic areas were divided into three groupings or “tiers” by number of eligibles: 1) *tier 1*: 2001 or more eligibles, 2) *tier 2*: 751-2000 eligibles, and 3) *tier 3*: 750 or fewer eligibles. Tier 1 areas were treated as self-representing areas because of their sheer size. Tier 2 areas were initially assumed to be separate Primary Sampling Areas eligible on their own for later selection. Tier 3 areas were initially assumed to need to be combined with other areas, to reach a minimum size of 750, for sample selection purposes; however, no combination over state lines was to be permitted, so many small areas (particularly the non-metropolitan areas of small states) had a size of less than 750. (We chose 750 as target minimum size for an area to stand by itself as a PSU because we expect to need to select 375 persons in a given area in order to obtain 250 completed mail questionnaires, and we did not want to select more than half the eligible population in any area.) After initial examination, some PMSAs were combined with other geographically adjacent PMSAs to form CMSAs (Consolidated Metropolitan Statistical Areas, the OMB designation for adjacent metro areas with significant cross-commuting to work), in order to attain the 750 size minimum for inclusion in Tier 2. Other MSAs/PMSAs which did not meet the minimum size for Tier 2 were tentatively consolidated with the non-metro remainder of their state.

The second basis for stratification were modified versions of Census Region and Division. In Tier 2, we began with the four Census Regions, but split the large South region into two parts (Southeast vs West South Central). For Tier 3, we used the 9 Census Divisions.

The resulting tiers and strata into which the 348 geographic areas were divided are as follows:

The **Tier 1** areas – the New York City MSA (n=5666), the District of Columbia MSA (n=3234), the Chicago MSA (n=2497), and the Los Angeles MSA (n=2453) – were so large that we considered them self-representing, and they were selected with certainty. There were an estimated 13,850 eligible persons in this tier.

The **Tier 2** areas– the Boston MSA , the Philadelphia MSA , the Newark MSA , the Atlanta MSA, the Miami-Fort Lauderdale CMSA plus the West Palm Beach MSA, the Houston MSA, the Dallas-Fort Worth CMSA, the Minneapolis MSA, and the San Francisco-Oakland-San Jose CMSA – were grouped into five strata:

*Stratum 2-1:* Northeast (New England plus MidAtlantic): includes the Boston MSA (n=1604), Philadelphia MSA (n=1204), and Newark (n=762), for a total of 3570 eligible persons.

*Stratum 2-2:* Southeast (South Atlantic plus East South Central): includes the Atlanta MSA (n=1173) and the Miami-Fort Lauderdale CMSA plus the West Palm Beach MSA (n=1090), for a total of 2263 eligible persons.

*Stratum 2-3:* West South Central: includes the Houston MSA (n=982) and the Dallas-Fort Worth CMSA (n=913), for a total of 1895 eligible persons.

*Stratum 2-4:* Midwest: includes the Minneapolis MSA (n=908).

*Stratum 2-5:* West: includes the San Francisco-Oakland-San Jose CMSA (n=1904).

After an adjustment noted below, there were an estimated 10,766 eligible persons in this tier.

The 336 **Tier 3** areas included a mix of metropolitan areas and non-metropolitan areas, each of which included less than 751 eligible persons. After adjustments noted below, there were an estimated 23,367 eligible persons in this tier. These areas were grouped into 9 strata, one for each of the Census Divisions:

*Stratum 3-1:* New England, which consisted of Connecticut (n=688), the remainder of Massachusetts outside the Boston MSA (n=164), Maine (n=81), New Hampshire (n=130), Rhode Island (n=151), and Vermont (n=70). There were an estimated 1284 eligibles in this stratum.

*Stratum 3-2:* MidAtlantic, which consisted of New Jersey outside the Newark MSA (n=634), New York outside the New York City MSA (n=752), and Pennsylvania outside the Philadelphia MSA (n=889). There were an estimated 2275 eligibles in this stratum.

*Stratum 3-3:* South Atlantic, which consisted of Delaware (n=202), Florida outside the Miami-Fort Lauderdale CMSA plus the West Palm Beach MSA (n=961), Georgia outside the

Atlanta MSA (n=216) , Maryland outside the DC MSA (n=703), North Carolina (n=627), South Carolina (n=363), Virginia outside the DC MSA (n=737), and West Virginia (n=260). This stratum included an estimated 3809 eligibles.

*Stratum 3-4:* East South Central, which consisted of Alabama (n=470), Kentucky (n=483), Mississippi (n=339), and Tennessee (n=599). This stratum included an estimated 1891 eligible persons.

*Stratum 3-5:* West South Central, which consisted of Arkansas (n=260), Louisiana (n=606), Oklahoma (n=423), and Texas outside the Houston MSA and the Dallas-Fort Worth CMSA (n=917). This stratum included an estimated 2206 eligible persons.

*Stratum 3-6:* East North Central, which consisted of Illinois outside the Chicago MSA (n=111), Indiana (n=707), Michigan (n=994), Ohio (n=1854), and Wisconsin (n=581). This stratum included an estimated 4247 eligible persons.

*Stratum 3-7:* West North Central, which consisted of Iowa (n=276), Kansas (n=227), Minnesota outside of the Minneapolis MSA (n=60), Missouri (n=1150), North Dakota (n=69), Nebraska (n=232), and South Dakota (n=81). This stratum included an estimated 2095 eligible persons.

*Stratum 3-8:* Mountain, which consisted of Arizona (n=544), Colorado (n=728), Idaho (n=146), Montana (n=108), New Mexico (n=181), Nevada (n=231), Utah (n=514), and Wyoming (n=51). This stratum included an estimated 2503 eligible persons.

*Stratum 3-9:* Pacific, which consisted of Alaska (n=121), California outside the Los Angeles MSA and the San Francisco-Oakland-San Jose CMSA (n=1407), Hawaii (n=176), Oregon (n=476), and Washington (n=851). This stratum included an estimated 3031 eligible persons.

Following the initial construction of these Tier 3 strata, some adjustments were made to reduce variations in size of areas within strata, so as to minimize the later effects of disparate selection weights. Ideally, all areas and all strata would be of approximately the same size, but we are limited by the small total numbers in some states and our desire to obtain at least 250 interviews (which we believe required 375 persons in the sample) from each stratum. But we do have some control over the minimum and maximum number of eligible persons in each area through our ability to redefine those areas. The following changes were made after the initial stratification:

First, three small areas (Massachusetts outside the Boston MSA (n=164), Illinois outside the Chicago MSA (n=111), and Minnesota outside the Minneapolis MSA (n=60) were consolidated with their related Tier 1 and Tier 2 areas respectively.

Second, a number of the larger areas (those with more than 750 eligibles and which could be split into areas of 375 or more) in Tier 3 were split to reduce their size:

1) the portion of Texas outside the Dallas-Fort Worth CMSA and Houston MSAs was split into two parts: the Austin MSA (n=435) and the remainder of Texas (n=482).

2) Ohio was split into the Cleveland-Akron CMSA (n=675), the Columbus MSA (n=486), and the remainder of Ohio (n=693).

3) Missouri was split into the St. Louis MSA (n=570) and the remainder of Missouri (n=580).

4) California outside the Los Angeles MSA and the San Francisco-Oakland-San Jose CMSA was split into the San Diego MSA (n=607) and the remainder of the state (n=800).

The resulting 9 Tier 3 strata had 52 areas. See the attached spreadsheet for a listing of these areas.

## 7. Selecting the first stage sample of geographic areas

Within each stratum with more than one PSA, the fraction of the eligible population represented by each area, and a cumulative fraction over areas, was calculated. For each such stratum, a random number between zero and one was generated, and the PSA whose cumulative fraction included the random number was selected as the stratum representative, the Primary Sampling Unit (PSU).

The selected PSUs areas are:

Stratum number	Name of area representing stratum	Estimated number of eligibles	Estimated number of interviews
11	New York NY PMSA	5666	400
12	District of Columbia PMSA	3234	400
13	Chicago IL PMSA and Illinois	2608	400
14	Los Angeles CA PMSA	2453	400
21	Philadelphia PA PMSA	1204	250
22	Atlanta GA MSA	1173	250
23	Houston TX PMSA	982	250
24	Minneapolis MN MSA and Minnesota	968	250
25	San Francisco-Oakland-San Jose CA CMSA	1904	250
31	Connecticut state	688	250

Stratum number	Name of area representing stratum	Estimated number of eligibles	Estimated number of interviews
32	New Jersey state outside Newark PMSA	634	250
33	Florida state outside Miami-Fort Lauderdale CMSA/West Palm Beach MSA	961	250
34	Tennessee state	599	250
35	Oklahoma state	423	250
36	Indiana state	707	250
37	St Louis MO MSA	570	250
38	Utah state	514	250
39	Oregon state	476	250
	total	25,864	5100

The Tier 1 PSUs include 29% of the eligible population, and 31% of the interviews. The Tier 2 PSUs include 22% of the eligible population, and 24% of the interviews. The Tier 3 PSUs include 49% of the eligible population, and 45% of the interviews.

The first stage sample of 18 PSUs includes portions of 18 states and DC. It includes 53% of the estimated national population of eligible persons.

Within strata 21, 22, 23, and 31-39, an alternative selection can be made if the relevant state authorities are unable or unwilling to cooperate in providing the information needed for the selection of individuals for the sample.

## **8. Selecting individuals in the second stage sample**

The second stage of sampling necessary for the After the JD project requires the identification and selection of eligible individuals within the geographic areas selected. Individuals will be selected on the basis of a) joining a state bar with jurisdiction over one of the areas selected in our geographic sample during calendar year 2000; b) not having been a member of another state bar prior to this admission; and c) graduating from law school in the period July 1, 1998 through June 30, 2000.

It is worth noting here the special problems of the DC area. While we estimate that 3234 CY1999 law school graduates took jobs in DC, only 327 persons were admitted to the DC bar in 1999 on the basis of passing the DC Bar Exam. Because the DC Bar (alone among US bar admitting authorities) imposes no length of practice requirement for being admitted to the DC bar by motion (all other states require at least 3 years), the overwhelming majority admissions to

the DC bar are by motion from persons who passed the bar or were otherwise admitted in other states first; 3027 persons were admitted to the DC bar on motion in 1999, 90% of the total admissions. Thus our ability to locate most of the persons who were admitted in DC depends on our finding them in the admissions records of other areas in our sample of geographic areas. (Many may come from the portions of the DC MSA which are in Maryland, Virginia, and West Virginia, but we expect to find such persons in nearly all of our other PSUs).

a. Constructing a sampling frame

Our expectation is that we will receive from state bar organizations (state bar associations in state with integrated bars, registration authorities in states without integrated bars, and perhaps from state bar examiners for some items of information) a full listing of persons who joined the state bar in 2000, along with information on current (early 2001) zip codes, the basis of joining the bar (exam, diploma privilege, motion), and the date of law school graduation. That will allow us to narrow the list frame to those persons who meet all our eligibility criteria. We expect our largest task will be the translation of zip codes to county identifiers and then to area (MSA and state) identifiers. (In states where not all the above information is available, we will need to select a sample that is larger than we would prefer, in order to account for the fact that some of the persons selected will be ineligible and their returned questionnaires discarded.)

For each sampling area, a separate database will be constructed including all apparently eligible persons. The database will include, for each person, a PSU identifier, a state identifier, an individual identifier (perhaps a state bar ID number), law school graduation date, bar entry date, and an individual name and address; we hope also to be able to obtain a law school identifier.

b. Selecting an individual sample

The target number of interviews for the After the JD study is 400 in each of the Tier 1 areas and 250 in each of the Tier 2 and 3 areas. Given our assumption of a 66.7% response rate, that means we need to start with a sample of 600 potential respondents in each of the Tier 1 areas and 375 potential respondents in each of the Tier 2 and Tier 3 areas. In addition, for any area in which we are unable to determine whether an individual is eligible on the basis of this being her being **first** admitted to a state bar in 2000, or graduating from law school in the period July 1, 1998 through June 30, 2000, we will need to increase the initial sample size by some margin (which we hope to determine from those areas which do provide us with all the required information but which we expect not to exceed 10%).

Following a determination of the number of persons required for the initial sample in each area, a sampling interval can be determined by dividing the number of persons required in the initial sample by the number of persons apparently eligible in the area. Then, a systematic interval sample can be obtained from a random starting point less than the interval.