RICHARD H. SANDER, Ph.D.
PROFESSOR OF LAW
Phone (310) 206-7300
Fax   (310) 825-6023
E-mail: sander@law.ucla.edu

UCLA SCHOOL OF LAW
405 HILGARD AVENUE
Box 951476
LOS ANGELES, CALIFORNIA 90095-1476

May 29, 2008

Judy Johnson, Executive Director
State Bar of California
San Francisco (Main Office)
180 Howard Street
San Francisco, CA 94105

Dear Ms. Johnson,

Please consider this letter to be a formal request for public information currently in the possession of the State Bar of California ("Bar"). Our right to this information is established in Article I of the California Constitution (as amended by Proposition 59), as well as by a common law right to records held by the judicial branch (*Copley Press, Inc. v. Superior Court*, 6 Cal.App.4th 106 (1992); *Pantos v. City & County of San Francisco*, 151 Cal. App. 3d 258 (1984)). Alternatively, our right to this information is established by the Public Records Act (Government Code §§6250 *et seq.*). This letter also serves as a formality to document my claim of right to this information, and to foreclose the possibility that the data will be destroyed before my request has been resolved. (*Laughlin v. IRS,* 103 F.Supp.2d 1219 (S.D. Cal, 1999)("A government improperly destroys agency records under FOIA only if it destroys them after a FOIA request has been made to the agency.").)

As you know, I (Richard Sander) made an earlier request for similar information on November 16, 2007.   Since then, my colleagues and I have done additional research both on the Bar's data holdings and on the best practices for anonymizing publicly-held data. We have revised our request to address the concerns raised by the Bar. This new request is similar to my November 16th request, but it includes an important additional variable (related to transfer students) and removes several variables that the Bar rarely collected. This request also tries to make as clear and easy to follow as possible our suggested procedures for thoroughly protecting the anonymity of individuals in the Bar's database.   This request supersedes my prior request.

We believe there are two key steps that will ensure that individual identities are not jeopardized. The first step is to remove from the data any "identity" variables, such as name, social security number, birth date, and other variables

that might directly disclose an individual's identity.

The second step involves so-called "publicly knowable category" variables. These are variables that are not themselves unique to any individual, but that a member of the public might be able to discover, and that might be combined in ways to deduce an individual's identity. For example, suppose that a law school had only one Native American student in its 2001 graduating class. Disclosing an individual-level database that included information on law school, year of graduation, and detailed race (including "Native American" as a category) would make it possible for a diligent member of the public to deduce this student's identity, and thereby know that the other information in the dataset applied to him or her.

The standard and accepted way of avoiding this problem is to make sure that no combination, or "cell," of all "publicly knowable" variables (like school, race, and graduation year) produces few enough cases that an individual's identity can be determined. The concept is illustrated in the article "Protecting the Confidentiality of Education Records in State Databases"by Barbara Clements (available at http://www.educationadvisor.com/documents/OCIO2001/CONFIDENTIALITY.DOC)

> If assessment scores are presented by race/ethnicity within a school and there is only one Asian/Pacific Islander in the school, then it is obvious that student received the score reported in that portion (cell) of the table. If there are scores for two students within a cell, then knowing one score would enable a person to figure out the other student's score. Including a minimum of three students in a group (cell) should prevent this type of inappropriate release of information from occurring. Higher minimums increase the number of students that must be known before the value for the final student can be calculated, thus are safer.

We are adopting the more stringent standard of five.

We are requesting the release of records containing individual-level data on the list of variables described in the chart at the end of this letter (Appendix A). We are reliably informed that the Bar maintains these records. The chart categorizes each variable as either "publicly knowable" or "non-public." It also provides instructions for how each variable should be grouped. We believe that even under a very conservative interpretation (*i.e.*, an interpretation most protective of privacy), there are only five variables in our request that fall into the "publicly knowable" category: race, law school, year of graduation, bar exam pass/fail outcome, and transfer status (whether the student transferred from one law school to another before receiving a JD). As noted above, privacy concerns arise if some combination of these variables produces fewer than five cases. In

technical terms, one cross-tabulates the population by these variables to create "cells." If all cells have five or more members, then privacy is assured; if a cell has fewer than five members, then some data categories are merged to incorporate individuals from the sensitive cells into cells with more members.

Our request uses broad categories to insure that few "cell size" problems arise. In particular, we request that data on each student's year of law school graduation be reported in broad chronological bands (e.g., 2003 through 2007), and that race data be reported in four broad categories (white, Hispanic, black, and other). These redactions will eliminate detail that could potentially be used to identify individuals. For the most part, these steps alone will be sufficient to thoroughly anonymize the data. As explained in Appendix B, we are sending along with this letter an Excel file that illustrates how the Bar's data for a typical California law school would look, both before and after anonymization, and how to perform a cell size analysis to insure there are no problems by a particular school's data.

However, there are a number of situations in which cell sizes issues would arise. For example, some California law schools (especially those not accredited by the ABA) have very small enrollments, and most out-of-state schools have such small numbers of students taking the California Bar that cell sizes in the disclosed information could be quite small. We address all of these situations in Appendix C, which examines each problem in turn and provides simple procedures for combination or elimination of data to keep the cell sizes in the safe zone.

We would like to emphasize that these procedures are extremely conservative in applying the generally-accepted rules for data anonymization. For example, we have included the three different variations on bar exam outcomes (passed, failed but later passed, never passed) as publicly-knowable values, even though as a practical matter it is difficult, if not impossible, to deduce from any public source whether an individual failed (let alone failed repeatedly) the bar exam. It is also generally difficult to deduce from public sources whether a student is a transfer. We have taken this conservative approach because we strongly value privacy concerns, and we know those concerns are important to the Bar as well. We hope that the effort and care with which the cell size problem is addressed in these materials conveys these values as well as our familiarity with the relevant issues.

A few matters of detail. First, we would prefer to receive the data in electronic format, if possible. We are willing to pay all reasonable costs associated with providing this data, including, of course, the Bar's costs involved in coding the data so as to avoid disclosures of confidential information. We are also willing to

pay all reasonable fees for an agreed-upon third party data confidentiality expert to work under contract with the Bar to create the anonymous public database. Second, we have assumed in this request that the Bar has responsive information useful to our research going back as far as 1972; this is based on my knowledge that Dr. Stephen Klein's studies date back to around that period. If this is not so, we would of course be happy to modify the request. Third, it would be helpful for the Bar's response to include a new identification number – not the number used in the Bar's official records, but one created for our request – so that if any questions arise about particular data points, it is possible for us to follow-up with the Bar with specific questions.

If the Bar believes that this request is deficient in some way, or might lead to the disclosure of confidential information, please provide us with a detailed explanation and all relevant material that led the Bar to come to such a belief. And of course, please do not hesitate to get in touch if we can be helpful in clarifying this request in any way.

As you know, the purpose of seeking this data is to conduct careful research, using the best social science methods, on the effect that attending particular law schools has upon students who have been admitted to those schools with the use of significant preferences. In the course of our (ultimately unsuccessful) efforts to persuade the Committee of Bar Examiners and the Board of Governors to do an internal study of these issues, the Bar heard from leading academics around the country, law school deans, and the United States Civil Rights Commission about the importance of this research. Indeed, many of the Bar's own leaders agreed that this was very important research to perform. The concerns raised by the Bar dealt exclusively with privacy issues and the Bar's legal authority to do such research. This request fully addresses the Bar's articulated concerns. I hope we will be able to work together to fulfill this request and permit this research to go forward. The result will be of great civil importance, and will provide information that will be immensely useful to both the public and the Bar.

Sincerely yours,

Richard H. Sander
Professor of Law
University of California, Los Angeles

Joe R. Hicks
Vice President
Community Advocates, Inc.

Appendix A
Data Variables Requested

Variable Name	Publicly Knowable?	Reporting Instructions
Race	Yes	Four categories: White, Black, Latino, Other
Law School	Yes	Name of the law school from which the bar-taker earned his or her JD
Transfer Student	Yes	See Appendix B
Year of Law School Graduation	Yes	Cluster into the following bands: 1972 to 1981 1982 to 1986 1987 to 1990 1991 to 1993 1994 to 1996 1997 to 1999 2000 to 2002 2003 to 2007
Total Raw score on first exam	No	
Total Scaled score on first exam	No	
MBE score on exam (raw and scaled)	No	
Essay score on exam (raw and scaled)	No	
Performance test score on exam (raw and scaled, for exams where it was administered)	No	
Bar Passage	Yes	0 = Never; 1 = First Attempt; 2 = Later Attempt
Law School GPA	No	
LSAT score	No	Report all known scores
Undergraduate GPA	No	

Appendix B
Illustrative Data

Accompanying this data request is an Excel file that is intended to illustrate our data request and anonymization process. The file contains imaginary data about students graduating from a fictional California law school, but we think it is helpful in understanding the details of our request and the concept of anonymization.

Sheet 1 in the file contains the sort of data that might exist in the California Bar's archives, for students who graduated from one California law school over a three-year period (2000 through the end of 2002). The data is grouped by graduation year and detailed race category.

Sheet 2 shows the same data in the form that we are requesting it. Note the important changes between Sheet 1 and Sheet 2: personal identifiers (e.g., name, bar identification number, school of origin) have been removed; some racial groups have been combined so only four broad categories remain; graduation year has been combined into a three-year range. For transfer students, race has been removed altogether as a variable, since otherwise there are likely to be cell size issues.

Sheet 3 cross-tabulates the data in Sheet 2 by the various "publicly knowable" variables – which, given that this data is already limited to one school during one period, really amount to the race and bar outcome variables. Each section of the spreadsheet shows the observations that fall within a particular cell – that is, a unique combination of the publicly knowable variables. In this illustration, no cell size problems arise because all cells have at least five members. If any cell had fewer than five members, the Bar would combine some categories of data together to increase cell size, following the procedures laid out in Appendix C (section 5). For example, in Sheet 3 one can see (lines 576-580) that the "transfers who never pass the bar" has a cell size of five. If there were only four people in this cell, then by our protocols the Bar would combine transfers who "never pass" with transfers who "fail but later pass" to create a larger cell (in this case, one with fourteen members).

Appendix C
Additional Database Protocols

The general rules described in the cover letter are, by themselves, enough to ensure confidentiality for the typical California bar exam takers – that is, bar-takers who graduated from accredited schools with substantial enrollments. The simulation described in Appendix B is based on such a typical school. The purpose of this appendix is to identify the variety of atypical situations where cell size problems could arise, and to lay out procedures for solving those problems.

(1) Graduates of small-enrollment California schools, and bar-takers with irregular training. In these cases, it is feasible to simply group schools into cohesive cohorts without significant loss of analytic usefulness. The Bar should therefore group following schools and bar-takers:

a. Four ABA-approved schools: Thomas Jefferson School of Law, University of LaVerne College of Law, Western State University College of Law, and Chapman University School of Law.

b. All of the California-accredited law schools. There are currently eighteen law schools accredited by California's Bar but not by the ABA:

Cal Northern School of Law, Empire College School of Law, Glendale University College of Law, Humphreys College Laurence Drivon School of Law, John F. Kennedy University School of Law, Lincoln Law School of Sacramento, Lincoln Law School of San Jose, Monterey College of Law, New College of California School of Law, San Fransisco Law School, San Joaquin College of Law, Santa Barbara College of Law, Southern Califronia Institute of Law (Santa Barbara), Souhtern California Institute of Law (Ventura), Trinity Law School, University of West Los Angeles School of Law (San Fernando Valley), University of West Los Angeles (West Los Angeles), Ventura College of Law

c. All of the bar-takers whose training comes from "California unaccredited" schools, Correspondence schools, and tutorial programs provided in law offices or judges' chambers. We ask the bar to identify all of these individuals as "Unaccredited and other" in the produced database.

d. All other Bar-takers who are not in the "out-of-state ABA" category. This includes foreign attorneys sitting for the California Bar, United States attorneys taking the general Bar exam, and a few miscellaneous others.

(2) Transfer students. A central purpose of our research is to determine

whether students who enter law school with much lower credentials than their classmates have special difficulties on the bar exam (difficulties greater than their relatively low credentials alone would explain). Since this research seeks to understand the effect of a student's law school upon his learning, transfer students who switch law schools before graduation pose an obvious difficulty. By definition, these students have attended more than one law school. It is therefore important to avoid mixing transfer students in with our principal analyses of legal education effects. Fortunately, the California Bar can determine which test-takers were law school transfer students, because the Bar asks all applicants to identify each of the law schools they attended. Because it is conceivable that an industrious researcher could deduce which students who graduated from a particular school were "transfers" and which were not, we treat transfers as a publicly knowable variable. And since the number of transfers at many law schools is relatively small, additional steps are needed to protect the anonymity of transfers. Therefore, we ask that, where the Bar can verify that a graduate attended more than one law school (and is disclosing data for a particular law school), it follow the following protocols:

a. Group transfer students with their school of graduation, not their school of matriculation. When cell sizes for transfer students are below the threshold of five:

b. Suppress the "race" variable for the problematic observations. If small cell sizes remain:

c. In the "pass variable", combine categories "0" (never pass) and "2" (pass after first attempt) for the problematic observations, using a code "3" to signify "did not pass bar on first attempt." If small cell sizes remain:

d. Suppress all variables for transfer students at that school during that time period except for the name of the law school, the year of law school graduation (banded), and the transfer status.

(3) Out-of-state ABA schools. California is a major work destination for law graduates trained all over the United States; nearly two thousand out-of-state law graduates take the California bar each year, and virtually every law school in the nation is annually represented by alumni. However, the number of graduates from individual out-of-state schools is always modest compared to a typical in-state institution, and in many cases the numbers are tiny. To avoid anonymity issues, we ask the bar to group the data according to the following protocols, which should eliminate all or nearly all of these issues:

a. Provide data from out-of-state ABA schools in longer time-bands than

those used for in-state schools, as follows: graduation years 2000 through 2007; 1991 through 1999; 1982 through 1990; 1972 through 1981.

b. Provide school identifiers only for the following out-of-state ABA schools: American University (Washington College of Law), Columbia Law School, George Washington University Law School, Georgetown Law School, Harvard Law School, New York University Law School, and University of Michigan Law School. (These schools are selected strictly on the criterion that they have large numbers of Bar-takers in California, and are likely, for each of the periods specified in 3(a), above, to have over five hundred California Bar-takers.

c. For the other out-of-state ABA schools, use the chart in at the end of this Appendix to assign each school to a non-California law school group, using that group label in place of the school name in the disclosed data (e.g, "Group A", "Group B", etc.).

(4) The special case presented by Stanford Law School. Stanford is smaller than the typical ABA law school in California, and because of its stature among law schools, its graduates are more likely to accept jobs outside California than is the case for any other California law school. Consequently, only around one hundred Stanford graduates take the Bar each year. Given our three-year grouping, it is therefore plausible that for some periods, particular cell sizes might include fewer than five individuals. Where that occurs, we ask that the Bar begin by combining values "0" and "2" in the pass variable into the value of "3" (as described in 5a, below). For any cases where small cell size remains a problem after this step is taken, we ask that the Bar use broader year ranges (as specified in 3(a), above, for the out-of-state ABA law schools).

(5) General combining strategies. As we have discussed, there may be particular cases in which, after following the protocols described in this request, cell sizes fall below the threshold of five individuals. In those cases, we ask that the following steps be taken:

a. Simplify the "bar passage outcome" variable by combining the values for "never passed the bar" (0) and "passed on a subsequent attempt" (2) into the value "3", signifying "did not pass the bar on first attempt." Where cell size problems remain:

b. Combine blacks and Hispanics into a combined racial group, "Underrepresented minorities" (assuming the cell size problem for the particular school and particular period is affecting that group).

c. If any cell size problems remain – which seems unlikely to us – we can

discuss for individual cases the combination of affected institutions.

We ask that these rules be used as narrowly as possible – that is, only for the years, schools, and racial categories where cell size emerges as an issue.

Table for Clustering Out-of-State Schools (see 3c, above)

Group A	University of Chicago, Yale University, University of Pennsylvania
Group B	Northwestern University, University of Virginia, Duke University
Group C	Cornell University, Vanderbilt University, University of Texas – Austin, Illinois University, University of Notre Dame, Washington and Lee University
Group D	Washington University in St. Louis, University of Minnesota The College of William and Mary, Boston University Boston College
Group E	Emory University, Brigham Young University University of Georgia, Fordham University, Wake Forest University University of Indiana – Bloomington, University of Colorado Temple University, George Mason University
Group F	University of Washington, University of North Carolina University of Arizona, Villanova University, University of Alabama Baylor University, University of Iowa, University of Wisconsin Ohio State University, Tulane University, University of Maryland – College Park, Southern Methodist University, Benjamin N. Cardozo School of Law, Brooklyn Law School, Lewis and Clark College Northeastern University , University of Richmond
Group G	University of Utah. University of Houston, University of Cincinnati Rutgers University – Camden, University of Tennessee Case Western Reserve University, University of Florida Chicago Kent School of Law, University of Oregon University of Pittsburgh, University of Kentucky, DePaul University University of South Carolina, Georgia State University

Group H	Arizona State University, University of Miami, University of Kansas Rutgers University – Newark, University of Missouri – Columbia St. John's University, University of Hawai'I, University of Nevada – Las Vegas, Catholic University, University of Oklahoma University of Denver, Marquette University, Seton Hall University Willamette University, Brandeis University, Pennsylvania State, Dickinson, University of Seattle, University of Connecticut SUNY, Buffalo, University of Nebraska, University of Indiana-Indianapolis, Wayne State University, Loyola University – Chicago University of St. Louis, Louisiana State University, McGeorge School of Law, University of the Pacific, Mercer University, Quinnipiac University, University of Akron, Samford University
Gorup I	University of New Mexico, Syracuse University, Hofstra University University of Maine, University of Arkansas – Fayetteville University of Mississippi, New York Law School, Drake University Texas Tech University, Creighton University, University of Idaho
Group I (continued)	University of Toledo, Gonzaga University, University of Memphis Suffolk University, Michigan State University, Northern Illinois University, Florida State University, University of Missouri – Kansas City, University of Vermont, SUNY, Albany, Cleveland State University, Pace University, Stetson University, St Mary's University Washburn University, William Mitchell College of Law Hamline University, John Marshall Law School, Texas Weslyan University, Northern Kentucky University
Group J	West Virginia University, University of Montana, University of Arkansas – Little Rock, University of Baltimore, Duquesne University Southern Illinois University, City University of New York Capital University, South Texas College of Law, Regent University Ave Maria University, Loyola University – New Orleans University of Tulsa, University of South Dakota, University of

	Dayton
	Franklin Pierce University, Widener University, Ohio Northern University, Western New England College, Roger Williams Unviversity, University of Wyoming, Howard University
	University of North Dakota, Valparaiso University, New England School of Law, Nova Southeastern University, Touro College
	University of the District of Columbia, Florida Costal University
	John Marshall – Atlanta, Widener University – Harrisburg
	University of Detroit Mercy, Mississippi College, University of Oklahoma, University of St. Thomas University, Liberty University
	Appalachian School of Law, Barry University, North Carolina central University, Texas Southern University, Southern University
	Thomas M. Cooley Law School, University of Puerto Rico