Element 2

Source Lists

The purpose of this element is to increase the coverage of the source list from which prospective jurors' names are selected. The courts should determine the percentage of those citizens eligible to serve on juries, which is covered by the existing source list, extend the coverage if it does not meet the standard, and make that improvement part of the Jury Plan.

APPLICABLE ABA STANDARD

STANDARD 2: JURY SOURCE LIST

(a) THE NAMES OF POTENTIAL JURORS SHOULD BE DRAWN FROM A JURY SOURCE LIST COMPILED FROM ONE OR MORE REGULARLY MAINTAINED LISTS OF PERSONS RESIDING IN THE COURT JURISDICTION.

(b) THE JURY SOURCE LIST SHOULD BE REPRESENTATIVE AND SHOULD BE AS INCLUSIVE OF THE ADULT POPULATION IN THE JURISDICTION AS IS FEASIBLE.

(c) THE COURT SHOULD PERIODICALLY REVIEW THE JURY SOURCE LIST FOR ITS REPRESENTATIVENESS AND INCLUSIVENESS OF THE JURY SOURCE LIST, APPROPRIATE CORRECTIVE ACTION SHOULD BE TAKEN.

(d) SHOULD THE COURT DETERMINE THAT IMPROVEMENT IS NEEDED IN THE REPRESENTATIVENESS OR INCLUSIVENESS OF THE JURY SOURCE LIST, APPROPRIATE CORRECTIVE ACTION SHOULD BE TAKEN.

In some states, the voters list is the source of names for the selection of prospective jurors.[1] This list consists of those registered to vote or those who voted in the most recent election and covers a much wider cross-section of the population than source lists previously prepared by keymen.[2] Many courts accept the voters list as the best single list to use because, in their opinion, it provides an adequate cross-section of a jurisdiction and has many desirable features, the most notable being that it includes all those who displayed civil responsibility by participating in elections.

The question has been raised whether exclusive use of the voters list meets the goals of jury selection. Since voter registration tends to peak every four years before a presidential election, the voters list in the intervening years may not fully represent significant portions of the population, particularly among the young and transient. Further, a substantial proportion of those eligible for jury service may not register to vote. The drivers list usually has greater coverage than the voters list and, as will be discussed, is the list used with the voters list in about half of the states. Since 1979, a number of jurisdictions have gone to exclusive use of the drivers list with excellent results.[3] These changes to the use of the drivers lists or the single use of the drivers list has resulted in improved representativeness at the expense of a reduced yield, as will be addressed in Element 3.[4]

Two measures against which the voters list, or any source list, should be tested are balance and inclusiveness. *Balance* is the degree to which the list reflects the eligible population based on its demographic characteristics (e.g., age, race, gender, occupation, etc.). *Inclusiveness* is the completeness of the list or combined lists.

If the inclusiveness is high, that is, if nearly everyone eligible appears on the master selection list, it follows that balance will be good. But the opposite is not necessarily true. Because of the difficulty of determining the balance of the population and source lists, the

---

[1] Material in this section is extracted from C. H. Mount, G. T. Munsterman, and W. R. Pabst, *Multiple Lists for Juror Selection: A Case Study for San Diego Superior Court* (1978), a study supported by the American University, Criminal Courts Technical Assistance Project. Supt. of Documents, Washington, D.C. 20402, a Stock No. 027-000-00665-5.

[2] *Keymen* were appointed community members who had the responsibility of compiling a list of qualified citizens for consideration as jurors. Keymen based their selections on any source of information about or simply their own personal knowledge of potential jurors. This is the origin of the office of *jury commissioner.* Keymen are no longer used in any state.

[3] A list of the various jury system functional parameters, such as the source list used, qualifications, exemptions, jury sizes used, fees, etc., can be found in David B. Rottman, Carol R. Flango, and R. Sheldine Lockley, *State Court Organization, 1993* (Williamsburg, VA: National Center for State Courts, 1995).

[4] G. Thomas Munsterman and Janice T. Munsterman, "The Search for Jury Representativeness," *Justice System Journal,* 11 (1986), 59-78.

Incentive Program chose as the standard an inclusiveness of at least 85 percent of the eligible population as the standard of source list adequacy. In most states, this "eligible" population will be the population aged 18 and over.

Courts who wish to meet this standard will necessarily use one or more other lists in addition to their present source list. The method of list selection and combination is presented in the following sections.

Before committing to the use of multiple lists, courts whose source list coverage is below 85 percent may wish to explore the demographic necessity for using multiple lists and, hence, to consider the balance of their lists; that is, the agreement between the source list and population. Some balance problems will be obvious, such as when a source list contains 75 percent male in contrast to a population of 49 percent male. However, when the population contains 5 percent of some minority and the source list contains only 2 percent, the significance of the difference and hence the question of the need for multiple lists is difficult to resolve. The literature contains measures of these disparities.[5]

List Selection

The process of determining which list or lists to use must include the practical aspects of acquisition and combination. For instance, the list of licensed drivers, usually the best list in terms of coverage, may not be separable by county or jurisdiction. Zip codes may also cross jurisdictional lines in some states, yet the voter/driver combination is the one most frequently used. Some lists are not available; e.g., the Social Security and federal census lists (see Table 2-1).

Courts considering the use of multiple lists should consider the following.

1. *Availability*. The best lists (Social Security, federal census, and income tax) cannot be used. However, a state or local census and state income tax lists are available in some jurisdictions. Courts should be prepared to both pay the costs of lists and ensure the privacy of the lists used.
2. *Efficiency*. Combining lists can be costly. This is particularly true if the individual lists are updated at different times, in which case the combined list should be recompiled each time one of the lists is revised. It is also very

[5] David Kairys, Joseph B. Kadane, and John P. Lechoczky, "Jury Representativeness: A Mandate for Multiple Source Lists," *California Law Review* 65 (1976), 776; and Peter W. Sperlich and Martin L. Jaspovice, "Methods for the Analysis of Jury Panel Selections: Testing for Discrimination in Series of Panels," *Hastings Constitutional Law Quarterly* 6 (1979), 787.

Table 2-1

Possible Source Lists for Juror Selection Process

Lists	Inherent Limitations
Social Security	Not available, the Social Security Number is available
Voter registration	Not up to date; not complete
City directory	Not complete; low income missing
Census	Federal list not available
Motor vehicle registration	Institutional and corporate listings, no age indentification
Driver's license	Not up to date; jurisdiction difficult to ascertain
Real estate tax	Commercial properties; mortgage companies; male bias
State income tax	Not available; male bias
Welfare	Not available in most jurisdictions
Telephone directory	Jurisdictions not always apparent; male bias
Utility customers	Jurisdictions not always apparent; not resident owner; male bias

inefficient to generate a large, nonduplicative master list when only a very small number of names is required (e.g., 10,000 selected out of 1,000,000). Kadane and Lehoczky present methods for directly combined samples of lists without having to handle combination of the entire lists. (See "Random Selection Without Full List Combination," page 11.)[6]

3. *Bias.* Some lists are heavily biased. For example, property tax and utility lists are biased toward property holders.
4. *Duplications.* Because of difficulties in eliminating duplicated names in multiple lists, an individual named on several lists has a greater probability of being selected than those named on only one list. Courts confronted with this problem accept some level of duplication rather than the risk of excluding a qualified citizen. The elimination of duplicates, either by computer or by hand, creates two types of errors: exclusion of a name that is not a duplicate and retention of a name that is a duplicate. The best method for removing

[6] J. B. Kadane and J. P. Lehoczky, "Random Juror Selection from Multiple Lists," *Operations Research,* 24 (1976), 207.

duplicates is to use a unique individual identification in each list, such as a Social Security number.[7]

5. *Geocoding.* Some lists are not easily resolved into court jurisdictions; i.e., county or district. This may require a manual verification of those few summoned who reside in the undefined areas, such as in one zip code or census tract, which are not limited to one court's jurisdiction.

Study of List Overlaps and Efficiencies

Even after source lists for juror selection are combined, few jurisdictions know the contribution of each list. Courts considering the use of multiple lists should have some estimate to determine if the combination is worthwhile; i.e., how many names not found on the first list does each successive list add. The method used to determine the number of unique names provided by the voters and drivers lists for San Diego County will be described.

To determine the overlap between the voters and drivers lists, a sample of 200 names was drawn from the voters list and manually checked against the entire drivers list. Eighty percent of the voters in San Diego County were also found on the drivers list (see Table 2-2.)

The reliability of these results was tested by reversing the process and checking a sample of 200 names from the drivers list against the voters list (see Table 2-3). The two samples were used to estimate the size of the total combined list less duplicates (see Table 2-4). The two samples show less than 1 percent difference in the percentage of overlap measured (49.6 percent vs. 48.7 percent), a difference that is insignificant.

Based on the estimate of the overlap from the first sample, the result of combining the drivers and voters lists in San Diego would provide a coverage of 93 percent. In contrast, the voters list alone would cover 56 percent of the 18 and over population, and the drivers list alone would cover 83 percent (see Figure 2-1).

---

[7] Under a 1993 revision to the Social Security Act, the Social Security number may be used for juror selection purposes.

Table 2-2

Voter/Driver Overlap Estimate from Voters List Sample

Category	Sample Size	% of Sample	Voters List	Drivers List
Total	200	100	628,217	925,497
Voter/Driver overlap	164	82	515,138	515,138
Voters only	36	18	113,079	
Drivers only				310,359

Table 2-3

Driver/Voter Overlap Estimate from Driver List Sample

Category	Sample Size	% of Sample	Voters List	Drivers List
Total	200	100	925,497	628,217
Voter/Driver overlap	100	55	509,023	509,023
Voters only				119,194
Drivers only	90	45	416,474	

Table 2-4

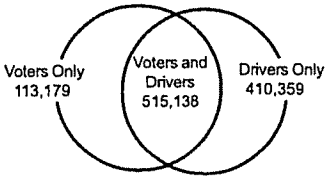
Summary of Sampling Results

Category	Voters List Sample from Table 2-2	Drivers List Sample from Table 2-3
Total combined list minus duplicates	1,038,576	1,044,691

Figure 2-1

Effect of Combining Lists in San Diego County, California, 1976

	Number	Coverage
Registered voters list	628,217	56%
Driver's licenses (18+)	925,497	83%
Total	1,553,714	
Less duplicates	515,138	
Combined master list	1,038,576	93%
Population 18 and over (est.)	1,110,783	



Voters Only	Voters and Drivers	Drivers Only
11%	50%	39%

Combined Master List—1,038,576 = 100%

Methodology for the Combining of Lists

Once the need for using two or more lists is apparent, the actual method of combining the lists comes into focus. The most direct way of doing this (and the most expensive and tiresome), whether done by computer or by hand, is the direct combination of all lists into a single alphabetical or otherwise ordered list from which duplicates have been removed. The intent of such a list is to include only one listing (or one card or token) for each eligible person such that each person has the same and equal chance of being selected.[8]

Misconceptions of Combining Source Lists

Before discussing the method for combining lists, several misconceptions concerning the use of multiple lists should be discussed

1. *Improper Supplementation*—Some courts take a sample of a list or lists and add these names to the voters list. This method violates equal probability of selection unless no duplicates exist between the lists.
2. *Improper Duplicate Recognition*—If a random sample is made from each list

[8] See Standard 3, *Standards Relating to Juror Use and Management,* Appendix A.

and duplicates are checked only between these samples, then equal probabil ity of selection is not maintained.

3. *Improper Combination*—Some courts attempt to duplicate the cross-section of the community by combining certain lists in determined proportions. Athough this achieves a better cross-section, the equal probability of selection is violated.
4. *Improper Timing*—Some courts draw from one list in the spring and another list in the fall, in a sense taking random samples from each list without considering the duplication between the lists.

Direct Combination of Lists

The combining or merging of two lists can be a manual task but is usually done by a computer. The task is complicated by the matching of names to remove duplicates. The computer technique is basically as follows:

1. The entry on each list is formed into a common format upon which the comparison for duplicates can be based. This information is called the *key.*
2. Each list is ordered based on the information in the key. If the key begins with the name of the person, the ordering is alphabetical.
3. The lists are simultaneously scanned, starting at the lowest key value or the top of each list, and a comparison of the entries from each list is made. If the comparison indicates a duplicate, the name is written to the combined list file, and the next name is considered. When duplicates are found, the most recent address or the address from the list found to be most up-to-date should be carried forward. If the names are different, the one of lowest value is written to the combined list file and the higher value name is kept and compared to the next name on the list from which the lower value name was taken.

The result of this process is a single merged list with the level of duplicates determined by the name match criteria used and the degree of duplication in each individual list. This combined list can then be sampled, using some random method, to generate a list for the qualifying or summoning process.

This technique can be used to merge any number of lists, because the process is a successive merging of two lists until all the lists are combined.

The computer time necessary to perform this method can take several hours; however, it is usually done only once a year. If jurisdictions desire a more up-to-date source list, the

process could be performed more often, providing the constituent lists are updated. Regardless of the method of combining the lists, certain data should be obtained, such as the number of duplicates found and the constituent sizes of the lists. Samples of the duplicates rejected can also be used to monitor the accuracy of the duplicate-matching routine.

Random Selection Without Full List Combination

Kadane and Lehoczky describe techniques for sequentially sampling and checking from several lists to arrive at a random sample.[9] These techniques do not require that the lists be combined; instead, only one list need be scanned for duplicates. If the list that is the most easily scanned is chosen, the task of achieving a merged list is simplified.

The following illustrates the methods of direct combination and random sampling (without direct combination) by using a hypothetical list. The combined list consists of twenty-one "names" with nine duplicated "names." The direct combination method described in the previous section would generate such a combined list.

The results of random selection without full combination are illustrated in Table 2-5. This method, which is described below, is the fourth of the five methods described by Kadane and Lehoczky.

1. A random sample is taken from each list given in Figure 2-2 in proportion to its size. Since the lists are of equal length, fifteen names, the samples are also equal, and a random sample of five names is selected from each list (see Table 2-5).
2. List 1 (Table 2-5) is considered the "primary list," and all five names are defined as "good." The five names selected from List 2 are compared to the entire List 1, and the duplicate names (A, L, and Q) are rejected. The remaining names (H and J) are combined with the first five to obtain the random sample of seven names.

This method achieves the same result as a direct combination followed by a random sample by sampling first and then comparing one sample to the entire other list and rejecting duplicates. This rejection of duplicates ensures that the duplicated names are given only a single chance of being selected; i.e., on the primary list. Either list may be the primary list, with the choice usually based on which list is easiest to check.

---

[9] Kadane and Lehoczky, *Operations Research*, 2-3.

Figure 2-2

Example of Combination of Lists

List 1	List 2	Combined List	Duplicates
A	A	A	A
B	C	B	F
D	E	C	I
F	F	D	L
G	H	E	M
I	I	F	P
K	J	G	Q
L	L	H	R
M	M	I	T
N	O	J	
P	P	K	
Q	Q	L	
R	R	M	
T	S	N	
U	T	O	
		P	
		Q	
		R	
		S	
		T	
		U	
15	15	21	9

A second example combines voters and drivers lists (see Table 2-6). The technique is as follows:

1. Indicate the drivers list as List 1 because it is usually larger than the voters list, is available on computer tape, and is accessible at an on-line terminal.
2. Designate the voters list as List 2 because it is generally smaller and usually composed of many different alphabetical section listings.
3. Determine the selection rate as follows:

Number of names needed		50,000	(A)
Number of names on List #1		900,000	(B)
Number of names on List #2		600,000	(C)
Estimated % of "good" names from List #2 (decimal)		.18	(D)
Estimated number of names to come from List #2	=	C x D	
"	=	600,000 x .18	
"	=	108,000 (E)	
Selection Rate	=	A/(B + E)	
"	=	50,000/(900,000 +108,000)	
"	=	.0496 = 5%	

Table 2-5

Random Sampling From Lists 1 and 2 to Get Random Sample of Combined List

Sample of Five Names		Random Samples from Combined Lists
List 1	List 2	List 1 Primary
D	A	D
G	H	G
P	J	H
Q	L	J
T	Q	P
		Q
		T
5	2	7

Table 2-6

List Merging Samples

Item	List 1 (Drivers)	List 2 (Voters)
Number of names	900,000	600,000
Random sample	45,000	30,000

4. Select 5 percent (45,000) names at random from the drivers list and accept all of these as valid names (unless there are internal duplicates, which should be removed). Do not check these against the other list.
5. Select 5 percent (30,000) names at random from the voters list and check each one of these against the entire drivers list. Approximately 82 percent of the 30,000 will be duplicates: these 24,600 duplicated names are dropped, while the remaining 5,400 nonduplicated names are added to the 45,000 drivers names to form a combined list of 50,400 names, near the desired number of 50,000 (A).
6. The 50,400 names retained as a result of this selection and checking proce dure should be randomized because they may retain the order of the original lists. If more or less than 50,400 names are desired, then the percent sampled from each list may be increased or decreased as necessary. The important factor is that the percentage remain constant for both lists; i.e., each sample is proportional to the size of the list. Because the approximate overlap of the lists is known (about 82 percent), the results are predictable within limits.

However, this method does not yield an exact number or the same number of names each time.

If the overlap of the lists is not known, cross-checking a small sample of each list will provide an estimate of the overlap needed.[10] Any error in this estimate can be corrected in future samples.

North Dakota has used this random selection technique without full list combination for several years. It is particularly useful in North Dakota, for the voters list is maintained only in the poll books and the drivers list is computerized. Therefore, it is very easy to look up a name from the voters list on the drivers list, but the reverse is virtually impossible. Before the introduction of this method, North Dakota employed several persons in many counties for several months to combine the two lists. This is no longer necessary.

Many courts with automated systems use this random selection technique because of the reduced computer run time needed to achieve the desired result.

Combination of Three or More Lists

In combining three or more lists as a source for jury selection, the principles and procedures are basically the same as when combining two lists. The first step, is to study the available lists with respect to their overlap. This can be done by sampling as described under the combination of two lists. The lists should then be ordered by size, ease of checking for duplicates, or both. The importance of studying the lists in these terms arises from the checking procedures in which all names selected from List 2 will be checked for duplicates only against the entire List 1; those names selected from List 3 will be checked for duplicates against the entire List 2 and then again against the entire List 1 and so on for as many lists as may be used. Such checking is necessary to retain equal probability of selection for each name on the combined list; i.e., to preserve the randomness of the selection.

Checking for duplication of those names selected from a sample of one list against only the sample from the other list, a shortcut that many jurisdictions have been tempted to adopt, does not produce a random sample from the combined list and, hence, should not be done.

Duplicate Recognition and Decision Errors

Any matching routine, whether manual or computerized, requires some criteria upon

[10] The method described in "Study of List Overlap and Efficiencies," page 7, may be used.

which to base the decision as to whether two records are duplicates or represent two individuals. Associated with each criterion are two possible errors:

1. The probability of rejecting a good name (a mismatch)
2. The probability of keeping a duplicated name (a missed match)

The kinds of decisions possible when two records are matched fall into one of four categories, which include the two error types and the two correct decisions, as illustrated in the following truth diagram.

		Criterion Indicates:	
		Individual Same	Individuals Different
True Situation	Individual Same	True Match A	Missed Match C
	Individuals Different	Mismatch B	True Nonmatch D

An error is made when either:

- The matching criterion does not recognize two records that, in fact, do match (missed match—cell C). The chance of this type of error becomes greater as the matching criterion becomes more strict and uses more information.
- The matching criterion recognizes two different records as representing the same person (mismatch—cell B). The chance of this error occurring in creases as the criterion becomes less strict, thereby enabling two records to be more easily identified as being similar.

The possibility that a good name will be rejected or that a list will contain duplicates is the product of two probabilities. The first probability is that the matching criterion will fail; i.e., reject a good name or keep a duplicated name (as in 1 or 2 above), and the second is the probability that a pair of names is or is not a true duplicate. The first probability is a property of the matching criterion. The second probability is a property of the lists considered and is related to the percentage of duplicates on the list.

Types of Record Errors

The matching criteria used to determine duplicate records for estimating the overlap of San Diego County's voters and drivers list was based upon the information available in each

list.[11] The percentage of records missing information is given below.

	Voter	Driver
Name:		
Last	0%	0%
First	0%	0%
Middle	9%	0%
Address		
Home No.	0%	0%
Street Name	0%	0%
Birth:		
Day	25%	0%
Month	25%	0%
Year	10%	0%

The basic matching criteria to determine a true match manually were as follows.

1. When birth month and day information existed on the voters list, the following must agree:
   a. Last name
   b. First name
   c. Middle initial
   d. Birthday
   e. Birth month

2. When birth month and date information do not exist, the following must agree:
   a. Last name
   b. First name
   c. Middle initial
   d. Home number
   e. Street name

All format and minor spelling discrepancies, such as Av., Ave., or Camto Basswood, Cam Basswood, were recorded but ignored for matching purposes. No attempt was made to check for duplicates that may exist within each list itself.

There were three areas of discrepancies between the 164 matched pairs found in San Diego County (see Table 2-7):

[11] If the Social Security number is available on the lists, this unique identification vastly simplifies the matching process. Unfortunately, this was not the case in the San Diego study cited in the text. While duplicate Social Security numbers can exist, the occurrence is much less than found in the other matching methods.

Table 2-7

Analysis of 164 Pairs of Voter/Driver Records with Matching Last Names

Matching Criteria	Street Name Exactly Same		Street Name Similar But Not Exactly Same		Street Name Different		Total
	Same Birthday & Month	Birth Date Not Recorded	Same Birthday & Month	Birth Date Not Recorded	Same Birthday & Month	Birth Date Not Recorded	
First Name Exactly Same:							
Same middle initial:							
Same street number	64	34	21	6	0	0	125
Different street number	1	0	0	0	20	0	21
Middle initial not recorded:							
Same street number	6	2	2	0	0	0	10
Different street number	0	0	0	0	5	0	5
First Name Similar But Not Exactly Same:							
Same name initial:							
Same street number	2	0	0	0	0	0	2
Different street number	0	0	0	0	1	0	1
Middle initial not recorded:							
Same street number	0	0	0	0	0	0	0
Different street number	0	0	0	0	0	0	0
Totals	73	36	23	6	26	0	164

- *Availability of records*—day and month of birth are missing in approximately 25 percent (42/164) of the voting list records; middle initial is missing in 9 percent (15/164) of the records.
- *Different street addresses*—approximately 16 percent (26/164) of the matched records have different street addresses but are listed due to same name and birth date information.
- *Format and spacing errors*—street name format errors (Ave., Av.; Cam, Camto, etc.) occur in 18 percent (29/164) of the records matched.

In the San Diego sample, only the last name seems to be immune to any discrepancies, possibly because this sample contains only last names that start with the letter A through "Armstrong." A recent study of combined voters/drivers lists with most of the duplicates removed by computer revealed a relatively higher duplication rate among last names starting with the letters D and M due to simple spacing (format) errors; e.g., Mc Hugh versus McHugh, which caused otherwise identical records to be denoted by the computer as two individuals. Other studies have shown last name discrepancies to be a real factor in

record matching.[12] These are often minimized by extraction codes or "Soundex" matching systems.[13]

Recommended Matching Criteria

Agreement of the following criteria is recommended for identifying duplicate records from the voters and drivers lists for the court studied.

- Last name
- First name
- Middle initial (where record exists)
- Birth month and day (where record exists)
- Street number or post office box number (street name is not included due to format difficulties)

In these criteria the greatest amount of reliable information is used to minimize the chance of a mismatch. The following format considerations are also necessary:

- No name or number should contain internal blank spaces
- Obvious errors in the records (such as incorrect zip or out of county, town, or city; nonalpha names or inconsistent numeric sequences; should be checked or rejected)

In using these criteria, the assumption was made during the sample checking that records matching the last name, first name, middle initial, day, and month of birth are the same, not withstanding street address differences. The court should investigate the validity of this assumption by using the information provided in the qualification questionnaire, which is sent to and returned by prospective jurors, and by sampling the names rejected as being duplicates.

Error Estimates

Based on the results of the 164 matched pairs, the expected percentage of true matches that will be missed is about 18 percent (29 of 164)—2 percent whose first names will differ slightly and 16 percent due to the record's different street numbers.

To determine the percentage of duplicates that will remain in the merged list, it is

---

[12] David M. Nitzberg, "Results of Research Into the Methodology of Record Linkage," and J. E. Fisher and M. R. Hubbard, "A Computer System for Medical Record Linkage," in E. D. Acheson, ed., *Record Linkage in Medicine* (London: E. & S. Livingstone, Ltd., 1968).

[13] A "Soundex" matching system compensates for most spelling errors by removing all vowels. See Acheson, ed., *Record Linkage in Medicine.*

necessary to introduce the second probability; i.e., the probability of a name being duplicated when the criteria are applied. The criteria are applied every time two names are compared. The number of comparisons is based on the routine used and is approximately $n_1 + n_2$, where n_1 and n_2 are the number of names on each list. Using the San Diego data (see Figure 2-1), if the probability of compared names being duplicated is

$$\frac{n_d}{n_1 + n_2}$$

where n_d is the number of duplicated names, the probability is

$$\frac{515{,}138}{1{,}553{,}714} = 0.332$$

The probability of a duplicate remaining in the merged San Diego list is, therefore, 0.177 x 0.332 = 0.059. The merged list will contain about 6 percent duplicates.

The probability of a good name being rejected is almost impossible to measure without actually merging and counting the errors based on manual screening of computer-determined duplicates. However, it can be estimated. The following probabilities of accepting a duplicate and rejecting a good name based on experiments by the Social Security Administration indicate the expected inverse relationship, which their product verifies.

P_{dup}	P_{rej}	$P_{dup} \times P_{rej}$
0.32	0.004	0.0013
0.12	0.01	0.0012
0.05	0.02	0.0010

The probability of retaining a duplicate was calculated to be 0.177. Over the range of 0.32 to 0.05, the product is fairly constant and can be used to estimate a P_{rej} of 0.007. The probability that a comparison will be nonduplicated is 1 - 0.332 = 0.668. The probability of a good name being rejected is 0.007 x 0.668, or 0.005.

Using the data from Figure 2-1, these errors would have the following effect upon the list combination in San Diego County.

Voters only	113,079
Drivers only	410,359
Voters and drivers	515,138
Unrecognized duplicates	+ 65,118
	1,103,694
Good names rejected	- 5,193
Total list size	1,098,501
The apparent coverage could be	$\frac{1{,}098{,}501}{1{,}110{,}783} = 99\%$

This shows how the unrecognized duplicates can inflate the coverage above the expected 93 percent.

A method for reducing the duplicates below the 6 percent level is developed in the referenced San Diego County report. This method uses a question on the qualification questionnaire to verify whether the person has a driver's license and is a registered voter in the state. The answers are used to verify the duplicate-matching technique and to give an estimate of the error rate for the list-merging process.[14]

---

[14] Mount, Munsterman, and Pabst, *Multiple Lists for Juror Selection,* 6-1 to 6-2.

Court Management Library Series

Jury System Management

by G. Thomas Munsterman