U.S. Department of Energy
Office of Health, Safety and Security
Office of Security Policy

Random Selection for Drug Screening

Random Selection for Drug Screening

Prepared for
Office of Security Policy
Office of Health, Safety and Security
U. S. Department of Energy
Washington, D.C.

by
Center for Human Reliability Studies
Oak Ridge Institute for Science and Education
Oak Ridge, Tennessee

ORISE 07-NSEM-0545

May 2007

Random Sampling

Sampling is the process of choosing some members out of a group or population. Probability sampling, or random sampling, is the process of selecting members by chance with a known probability of each individual being chosen. Nonprobability sampling may include volunteers or a "quota" of purposely chosen members that are thought to be "representative" of a certain population according to subjective criteria. Nonprobability sampling introduces biases into the sample that cannot be quantified.

Simple random sampling (SRS) is generally the starting point for a random sampling process. This sampling technique ensures that each individual within a group (population) has an equal chance of being selected. Let the number of members of the population be N, and let the sample contain n distinct elements out of this N. Then with SRS any possible subset of size n from the population of size N is equally likely to be chosen as the sample. This means that every member has the same probability of being selected for the sample as every other member, and the joint probabilities of sets of elements being selected are also equal. SRS is appropriate only when all members of the population can be identified and are uniform, or homogeneous, in the characteristic of interest. It is not widely used, but often underlies more complex sampling designs that are developed for specific applications. Sampling may be done either with or without replacement. If sampling is carried out with replacement, a member selected for the sample is "thrown back into the selection pool" and becomes eligible to be selected again. Because this procedure can result in the same member being selected multiple times, the outcome sample may contain fewer than n distinct elements.

There are a variety of valid ways to implement random sampling in a practical situation. Two potential ways of applying randomization in the context of drug screening are: (1) using daily SRS of the program participants; and (2) randomly selecting a new drug testing date for a participant when the current drug test has been carried out. The first approach randomly picks individual workers on a specific date, while the second method randomly picks a specific date for an individual worker. Although both of these approaches utilize randomization procedures, selecting random dates is much more viable for application in a drug screening program.

Daily simple random sampling of individuals from a group

With daily SRS, on each drug testing day a given number n of the N members of the group are chosen to be tested, and each participant has the same probability, n*(1/N), of being picked. For example, if there are N=200 participants in the group and n=4 are to be tested that day, then each worker has a 4*(1/200), or 2%, chance of being selected. Each participant has a known probability (e.g. 2%) of being picked on the next drug testing day and every other drug testing day that follows. If a worker selected for testing is placed back into the pool and is eligible to be selected again the next testing day, then the probability of being selected the next day would be exactly the same for the already-selected worker as for a worker who has not been selected. Daily SRS with replacement into the pool the next day is not efficient for the following reasons:

- This approach does not accomplish the goal of guaranteeing that every participant receives a random drug test every 12 months.
- A single worker may be selected for multiple drug tests during 12 months, which can disrupt continuity of work processes, affect production, and damage employee moral.

These two problems are interrelated. The larger the size of the daily sample, the more likely a worker is to be picked for at least one drug test and thus meet the 12-month requirement. However, a larger sample also causes an increase in the chance of being selected multiple times during 12 months. Because of these underlying problems, the daily SRS is difficult to implement effectively whether the group consists of all participants enrolled in the drug testing program or a subgroup determined by shift, work duties, or other criteria. To diminish problems such as the two described above, the SRS approach is often modified to meet specific application needs.

Daily stratified random sampling of individuals

A frequently employed alternative to SRS is stratified random sampling, in which simple random sampling is applied within separate pools, or strata. These strata can be determined by whether the worker has or has not already been tested. Stratified random sampling reduces the chance that an individual would be missed by dividing the population into two groups or strata, e.g., drug-tested and not drug-tested. A sample is then drawn from these two groups. A unique feature using this methodology is that more can be selected from one group than another and unannounced randomness can still be maintained. For example, 4% could be selected from the not-yet-tested population versus 0.5% from the already-tested population. However, in the context of random drug screening there are also major problems with this approach. Stratified sampling by pools of not-yet-tested or already-tested participants is problematic for the following reasons:

- This approach does not accomplish the goal of guaranteeing that every participant receives a random drug test every 12 months.
- There is no feasible random and effective method to return a participant from the already-tested to the not-yet-tested pool within the required 12-month period.

An individual in the not-yet-tested stratum could fail to be selected by chance for more than 12 consecutive months, even with a high percentage of selection from this pool. After receiving a drug test, a participant is moved from the not-yet-tested pool to the already-tested pool. Because only a small percentage of individuals in the already-tested pool are selected on each drug testing day, this reduces the chance of multiple selection in 12 months for any individual. However, one problem that arises with the two separate pools is deciding when an individual should be returned from the already-tested pool to the not-yet-tested pool to receive a drug test within the required 12 months. Participants in the already-tested pool have different dates of their last drug test that could range from one day ago to months ago. If specific criteria were applied to determine when to return an individual to the not-yet-tested pool, this would invalidate random selection. But if an individual were randomly selected to move back to the not-yet-tested pool, some workers could remain in the already-tested pool beyond their 12-month deadline and very likely not receive their required random drug test.

Randomly selecting a new drug testing date for an individual

An alternative to SRS or stratified random sampling of individuals is random selection of a date that is within 12 months of each drug test conducted for an individual. This methodology, which guarantees that each participant is selected for random drug testing every 12 months, can be implemented to keep multiple drug tests during a 12-month period at a low level. This straightforward approach assigns a randomly selected integer from 1-12 to each participant who received a drug test during a month; this random integer determines the number of months until this individual's next drug test. In this manner everyone will meet the 12-month requirement, but the randomness of the test date is maintained.

The Random Drug Screening System (RDSS)

The Center for Human Reliability Studies developed the Random Drug Screening System (RDSS), a Microsoft Access© application for administering a drug screening program. The RDSS is based on the approach of randomly selecting a new drug testing date for an individual. There are three separate random components to the RDSS process for selecting the next test date for an individual. The first random element determines the month/year of the next drug test by picking an integer from 1-12. The assigned integer is stored in a table within the compiled program so that the value of the integer cannot be accessed by anyone, including the program administrator or a computer programmer. The second and third pieces are involved in randomly choosing the test day during that month. The RDSS provides the ability to enter the days during a month when drug tests can be conducted so that certain days can be excluded if required. From the testing days entered by the program administrator, actual testing days are randomly selected and then randomly assigned to individual participants being tested that month. In addition, the expected number of drug tests per individual per year can be customized to suit program needs. To protect confidentiality, even the program administrator does not know until the morning of a testing day the names of the workers to be tested that day.

Selecting the month/year for the next test

The core of the RDSS randomization process is a Microsoft Access© table, "tblRandomMonth," that contains 200 integers from 1 through 12. The number of times

each of these integers occurs in the table is determined by the percents that each facility selects as appropriate for their particular drug testing program. Table 1 shows the default values for "tblRandomMonth."

Table 1: Default Percents of Each Integer in "tblRandomMonth"

Integer	Percent	How many in table
1	0. 5%	1
2	1%	2
3	3%	6
4	4%	8
5	6%	12
6	8%	16
7	11%	22
8	19%	38
9	18%	36
10	17%	34
11	12%	24
12	0.5%	1
Total	**100%**	**200**

The built-in Microsoft Access© randomization function is utilized to select an integer at random from the 200 integers in the table. The greater the number of occurrences in the table of a given integer, say 9, the greater the chance that an employee will be assigned a next test month/year 9 months from the test just completed. The percents of each integer seen in Table 1 have been used by many drug testing programs. Communications with the users of this system indicate that it works well. Each percent establishes the total number of employees, on average, who will receive their next drug test during the given number of months after their current test. For example, 11 percent is assigned to the integer 7. If there are 389 employees, then 0.11 x 389 or approximately 43 employees, on average, will be tested seven months after their last test. Using the default values, the expected number of participants to have exactly one drug test within 12 months following a confirmed drug test is 87.088%, while 99.649% can be expected to have no more than two tests, and 99.996% no more than three tests within 12 months. It is important to note that the first year after

adopting the RDSS is a transition period when the distribution of previous test dates also influences the number of multiple tests. The percents of individuals receiving multiple tests may be higher than expected during the first year, particularly when the cutoff dates for the next drug test for the entering program participants were not spread out fairly evenly during the 12 months preceding implementation. If a large number of participants require a drug test within a few months of RDSS implementation, then it may take more than a year to move into the expected lower level of multiple drug tests.

Using the default values, approximately one of every 200 drug tests will randomly be assigned to occur 12 months after the current test. It should be noted that the randomly selected day of this next drug test could be any drug testing day during that month. Therefore, it would be possible for an individual to receive a drug test more than 365 days after the previous test. To avoid this possibility, the percents in Table 1 would be changed as follows in only the rows shown below:

Integer	Percent	How many in table
11	12.5%	24
12	0.0%	1

On the last working day of a month, each employee who had a drug test during the month is assigned a next drug test month/year by adding an individual random integer from 1 to 12 to the current month/year. For example, if the current month/year is January 2007, a person tested this month and receiving the random integer 5 would be tested next in June 2007, and another person receiving the random integer 8 would be tested next in September 2007. To enhance security, the month/year of each participant's next drug test is hidden in an internal table that cannot be viewed, even by the program administrator.

Selecting the day of the next test

At the beginning of each month, the available days of the month when drug screening tests can be conducted are determined and how many days, d, that drug tests will be performed is decided. The built-in Microsoft Access$^©$ random function can then choose d specific days from among the available days, which adds a second random component to test date

selection. A third random component can be included when each worker scheduled to receive a drug test during the month is randomly assigned to one of the actual testing days. In the RDSS, for example, in September 2006 drug testing would not occur on the 4th, which is Labor Day, or on weekends. Assume also that drug tests are not carried out on Fridays at this facility because of work scheduling constraints. Therefore, the program administrator would enter 5, 6, 7, 11, 12, 13, 14, 18, 19, 20, 21, 25, 26, 27, and 28 as days when tests could be conducted. The administrator then has the option of selecting the number of actual testing days. This number can be equal to the total number of available testing days if desired, which is 15 in this case. To include an additional random component, the administrator has the alternative option of stipulating that the number of days to conduct testing is less than the total number of available testing days. For example, in this case the administrator could stipulate 10 days for conducting drug tests this month. The RDSS would then randomly choose 10 of the 15 possible testing days to be the actual testing days. There are 3003 possible different combinations of 10 days that can be picked from these 15 days, and one of these 3003 combinations is chosen for actual test days. Next, each participant whose next test month/year was September 2006 would be assigned randomly to one of the 10 actual test days.

Legitimately Unavailable Participants

A decision must be made on how to handle the situation where a worker who was randomly selected for a drug test but was not available for testing that day because of being on travel, on vacation, absent due to illness, or for another valid reason. Participants in this situation will be referred to as "legitimately unavailable." The RDSS has been built using the reasoning that a legitimately unavailable individual has now been randomly selected for a drug test, and that drug test should be carried out as soon as possible. Therefore, legitimately unavailable workers are placed in a separate pool, called the picked-but-not-yet-tested pool. From this pool, participants are added to the daily testing list by the program administrator as soon as they return to work and there is an opening for drug testing. If a participant was randomly selected for drug testing while on personal leave, travel, or absent for other reasons, the best time to test is soon after returning to work, for it would be during these situations that an individual might be more likely to consider using illegal substances. It has been suggested that this procedure cancels the random selection. However, none of

the three separate random components involved in picking the date of each person's random drug test has been affected in any way by the individual's unavailability. The random selection is not in any way invalidated by the unavoidable delay in carrying out the test.

The only difference between one randomly selected participant who receives a drug test on the day chosen and another randomly selected participant who receives a drug test upon returning to work is the fact that the program administrator and supervisor are aware that the legitimately unavailable individual has now been selected. Although the randomness of selection is not an issue here, there is a possibility of a security issue if the program administrator or supervisor cannot be trusted to keep the drug testing date confidential. If it is determined that a potential security issue exists, there are several ways this problem might be mitigated. For example, each morning all supervisors might be required to send the program administrator the names of participants who are not available for drug testing that day due to legitimate reasons. Then the program administrator would not notify supervisors of a drug test selection for any of their absent employees but would confidentially place names in the picked-but-not-yet-tested pool. In this way, only the program administrator would ever have knowledge of the names of individuals in the picked-but-not-yet-tested pool.

Conclusion

Given the above discussion regarding selection and randomness, the issue is how to maintain the integrity of the unannounced random drug testing for individuals who for legitimate reasons are unable to be tested when selected. Though we have shown one way this could be done for the RDSS, it must be pointed out that drug testing systems are based on trust of the individuals who are responsible for its implementation and conduct. No statistical process can ensure the integrity of random unannounced tests if the individual responsible wishes to subvert the system and inform a selected individual. If an alcohol testing program were based on the premise of at least one random alcohol test every 12 months, then the RDSS would also be suitable for selecting program participants for this alcohol testing.

Glossary of Terms

- Sampling – the process of choosing some members out of a larger group, called the population.
- Random Sampling – a sampling technique in which each individual is chosen entirely by chance and each member of the population has a known, but possibly non-equal, chance of being included in the sample
- Simple Random Sampling – random sampling with each member of the population having an equal chance of being included in the sample and every sample of a given size having the same chance of being selected. SRS is appropriate when the population is fairly uniform (homogeneous) in the characteristic of interest.
- Stratified Random Sampling – a sampling technique in which the members of a diverse (heterogeneous) population are first partitioned into more homogeneous sub-groups, called strata, and then simple random sampling is applied within each stratum.
- Sampling with Replacement – sampling members of the population one at a time with the selected member being replaced before the next is sampled. The chance of being selected each draw remains constant, but a member can be selected multiple times for the sample.
- Sampling without Replacement - sampling members of the population one at a time for the sample with the selected member not being replaced before the next is sampled. The chance of an individual being selected depends on all previous outcomes, and a member can be selected only one time for the sample.