

# National Estimates Methodology

September 2011

# NFLIS

**NATIONAL FORENSIC LABORATORY  
INFORMATION SYSTEM**

U.S. DEPARTMENT OF JUSTICE  
DRUG ENFORCEMENT ADMINISTRATION  
OFFICE OF DIVERSION CONTROL

# NATIONAL ESTIMATES METHODOLOGY

## Overview

Since 2001, NFLIS publications have included National and regional estimates for the number of drug reports and drug cases analyzed by State and local forensic laboratories in the United States. This document discusses the current methods used for producing estimates, including sample selection, weighting, and imputation procedures. Results from a 1998 survey of forensic laboratories (updated in 2002, 2004, and 2008) provided laboratory-specific information, including annual caseloads, which were used to establish a National sampling frame of all State and local forensic laboratories that routinely perform drug chemistry analyses. A representative probability proportional to size (PPS) sample was drawn on the basis of annual cases analyzed per laboratory, resulting in a NFLIS National sample of 29 State laboratory systems and 31 local or municipal laboratories, for a total of 168 individual laboratories.

Three changes to the initial procedures were implemented with the 2010 NFLIS publications. These changes are mentioned in the announcement in the 2010 Midyear Report and the 2010 Annual Report. These three changes were applied to all previous data reference periods to maintain the ability to compare drug trends. The first change is to a new statistical method, referred to as NEAR (National Estimates Based on All Reports), which more fully exploits the high rate of reporting laboratories. As of December 2010, laboratories representing over 92% of the National drug caseload participate in NFLIS, with about 88% of the National caseload reported for each data reference period.

The second change is that estimates are based on cases and items *submitted* to laboratories during the data reference period and *analyzed* within three months of the end of the data reference period. Analysis has shown that at least 95% of cases submitted during a six-month or one-year period are analyzed within three months after the end of the period (not including the approximately 30% of cases that are never analyzed).

The third change involves the accounting of multiple drugs per item. For each item (or exhibit) analyzed by a laboratory in the NFLIS program, up to three drugs can be reported to NFLIS and counted in the estimation process. A drug-specific case is one for which the specific drug was identified as the first, second, or third drug report for any item associated with the case. A drug-specific report is the total number of reports of the specific drug.

## NEAR Methodology

In NFLIS publications released before 2011, data reported by non-sampled laboratories were not used in National or regional estimates.<sup>1</sup> However, as the number of non-sampled laboratories reporting to NFLIS increased, it became logical to consider ways to utilize the data submitted by the non-sampled laboratories. Under NEAR, the “volunteer” laboratories (i.e., the reporting non-sampled laboratories) are allowed to represent themselves and are no longer represented by the reporting sampled laboratories. The volunteer laboratories are assigned weights of one, and hence the weights of the sampled and responding

---

<sup>1</sup> The case and item loads for the non-sampled laboratories were used in calculating the weights.

laboratories are appropriately adjusted downward. The outcome is that the estimates are more precise, especially for recent years when the number of volunteer laboratories is large. More precision allows for more power to detect trends and fewer suppressed estimates in NFLIS publications.

## NEAR imputations and adjusting for missing monthly data in reporting laboratories

Because of technical and other reporting issues, some laboratories do not report data for every month during a given data reference period, which results in missing monthly data. A laboratory that reports fewer than half of the data for a reporting period is considered non-reporting, and its reported data are not included in the estimates. Otherwise, imputations are performed separately by drug for laboratories that are missing monthly data, using drug-specific proportions generated from laboratories that are reporting all months of data. This imputation method, which is used for cases, items, and drug-specific reports, accounts for both the typical month-to-month variation and the size of the laboratory requiring imputation. The general approach is to use the non-missing months to assess the size of the laboratory requiring imputation and then to apply the seasonal pattern exhibited by all laboratories with no missing data. Imputations of monthly case counts are created using the following ratio ( $r_L$ ):

$$r_L = \frac{\sum_{m \in R_L} c_{L,m}}{\sum_{m \in R_L} c_{.,m}},$$

where

$R_L$  = set of all non-missing months in laboratory  $L$ ,

$c_{L,m}$  = case count for laboratory  $L$  in month  $m$ , and

$c_{.,m}$  = mean case counts for all laboratories reporting complete data.

Monthly item counts are imputed for each laboratory using an estimated item-to-case ratio ( $s_L$ ) for non-missing monthly item counts within the laboratory. The imputed value for the missing monthly number of items in each laboratory is calculated by multiplying by  $c_{L,m}$  by  $s_L$ .

$$s_L = \frac{\sum_{m \in R_L} i_{L,m}}{\sum_{m \in R_L} c_{L,m}},$$

where

$R_L$  = set of all non-missing months in laboratory  $L$ ,

$i_{L,m}$  = item count for laboratory  $L$  in month  $m$ , and

$c_{L,m}$  = case count for laboratory  $L$  in month  $m$ .

Drug-specific case and report counts are imputed using the same imputation techniques for the case and item counts listed above. The total drug, item, and case counts are calculated by aggregating the laboratory and laboratory system counts for those with complete reporting and those that require imputation.

## **NEAR imputations and drug report-level adjustments**

Most forensic laboratories classify and report case-level analyses in a consistent manner in terms of the number of vials of a particular pill. A small number, however, do not produce drug report-level counts in the same way as those submitted by the vast majority. Instead, they report as items the count of the individual pills themselves. Laboratories that consider items in this manner also consider drug report-level counts in this same manner. For those laboratories, drug report-to-case ratios for each drug were produced for the similarly sized laboratories, and these drug-specific ratios were then used to adjust the drug report counts for the relevant laboratories.

## **NEAR weighting procedures**

Each NFLIS reporting laboratory was assigned a weight to be used in the calculation of design-consistent, nonresponse-adjusted estimates. Two weights were created: one for estimating cases and one for estimating drug reports. The weight used for case estimation was based on the caseload for every laboratory in the NFLIS population, and the weight used for drug report estimation was based on the item load for every laboratory in the NFLIS population. For reporting laboratories, the caseload and item load used in weighting were the reported totals. For non-reporting laboratories, the caseload and item load used in weighting were obtained from an updated laboratory survey administered in 2008.

When the NFLIS sample was originally drawn, two stratifying variables were used: type of laboratory (State system or municipal or county laboratory) and (2) determination of “certainty” laboratory status. To ensure that the NFLIS sample had strong regional representation, U.S. census regions were also used as the geographical divisions to guide selection of certainty laboratories and systems. Some large laboratories were automatically part of the original NFLIS sample because they were deemed critically important to the calculation of reliable estimates. These laboratories are called “certainty laboratories.” The criteria used in selecting the certainty laboratories included (1) size, (2) region, (3) geographical location, and (4) other special considerations (e.g., strategic importance of the laboratory).

Each weight has two components, the design weight and the nonresponse adjustment factor, the product of which is the final weight used in estimation. After imputation, the final item weight is based on the item count, and the final case weight is based on the case count of each laboratory or laboratory system. The final weights are used to calculate National and regional estimates. The first component, the design weight, is based on the proportion of the caseload and item load of the NFLIS universe represented by the individual laboratory. This step takes advantage of the original PPS sample design, which provides

precise estimates as long as the number of drug-specific case estimates and report estimates are correlated with the overall caseload and item load.<sup>2</sup>

For non-certainty reporting laboratories in the sample (and reporting laboratories in the certainty strata with non-reporting laboratories), the design-based weight for each laboratory is calculated as follows:

$$\text{Design Weight}_i = A / (B \times \text{Case [item] Count for Laboratory or Laboratory System } i),$$

where

$i$  =  $i$ th laboratory or laboratory system,

$A$  = sum of the case (item) counts for all of the laboratories and laboratory systems (sampled and non-sampled) within a specific stratum, and

$B$  = number of sampled laboratories and laboratory systems within the same stratum.

Laboratories in certainty strata with no non-reporting laboratories were assigned a weight of one.<sup>3</sup>

The second component, the nonresponse adjustment factor, adjusts the weights of the reporting and sampled laboratories to account for the non-reporting sampled laboratories. The nonresponse (NR) adjustment, for both certainty and non-certainty laboratories, is calculated as follows:

$$NR_j = C/D,$$

where

$j$  = stratum;

$C$  = sum of the case (item) counts of all sampled laboratories and laboratory systems within the stratum, excluding the volunteer stratum; and

$D$  = sum of the case (item) counts for all sampled reporting laboratories and laboratory systems within the same stratum.

Because volunteer laboratories represent only themselves, they were automatically assigned a final weight of one.

---

<sup>2</sup> Lohr, S. L. (2010). *Sampling: Design and analysis* (2nd ed., pp. 231–234). Boston, MA: Brooks/Cole.

<sup>3</sup> With respect to the design weight, laboratories and laboratory systems in certainty strata with non-reporting laboratories and laboratory systems are treated the same way as non-certainty sampled laboratories and laboratory systems. This is done to reduce the variance; otherwise, all reporting laboratories and laboratory systems in certainty strata would get the same weight.

## NEAR estimation

The estimates are the weighted sum of the counts from each reporting laboratory. The weighting procedures make the estimates more precise by assigning large weights to small laboratories and small weights to large laboratories.<sup>4</sup> Because most of the values being estimated tend to be related to laboratory size, the product of the weight and the value to be estimated tends to be relatively stable across laboratories, resulting in precise estimates.

A finite population correction is also applied to account for the high sampling rate. In a sample-based design, the sampling fraction, which is used to create the weights, equals the number of sampled laboratories divided by the number of laboratories in the NFLIS universe. Under NEAR, the sampling fraction equals the number of sampled laboratories divided by the sum of the number of sampled laboratories and the number of non-reporting, non-sampled laboratories. Volunteer laboratories are not included in the sampling fraction calculation. Thus, the NEAR approach makes the sampling rate even higher because volunteer laboratories do not count as non-sampled laboratories.

## Suppression of Unreliable Estimates

For some drugs, such as hydrocodone, thousands of reports occur annually, allowing for reliable National prevalence estimates to be computed. For other drugs, reliable and precise estimates cannot be computed because of a combination of low report counts and substantial variability in report counts between laboratories. Thus, suppression rules were established. Precision and reliability of estimates are evaluated using the relative standard error (RSE), which is the ratio between the standard error of an estimate and the estimate. Drug estimates with an RSE > 50% are suppressed and not shown in the tables of estimates.

## Statistical Techniques for Trend Analysis

Trend analyses can be performed on the estimates covering any period of time for selected drug reports. Typically, models test for mean differences; however, the NFLIS estimates are based on total drug report counts. To work around this challenge, a bootstrapping technique was employed. (Bootstrapping is an iterative technique used to estimate variances when standard variance estimation procedures cannot be used.<sup>5</sup>) All statistical tests were performed at the 95% confidence level ( $p < 0.05$ ). In other words, there is a < 5% probability of detecting a statistically significant linear trend when no linear trend exists.

The bootstrapping method used for trend analysis has four steps. First, estimates and standard errors are obtained for each time period included in the specified trend analysis. Second, a background distribution that assumes no trend is generated using a simulation: For each time point, 1,000 values are drawn from a normal distribution with a mean equal to the mean of all estimates included in the trend analysis and a standard deviation equal to the actual standard error from the first step. Third, the slope of the least-squares trend line is calculated for each of the 1,000 simulated time series. Fourth, the slope of

---

<sup>4</sup> See footnote 2. (Lohr, S.L. 2010)

<sup>5</sup> For more information on this technique, see Chernick, M. R. (1999). *Bootstrap methods: A practitioner's guide*. New York: Wiley.

the observed least-squares trend line is calculated. If the observed slope is  $\geq 975$  of the 1,000 simulated slopes, a significant increasing trend is indicated; if the observed slope is  $< 975$  of the 1,000 simulated slopes, a significant decreasing trend is indicated. Otherwise, the data did not support a significant linear trend.

Note that the trend analyses test for a linear trend is based on a time series of the estimates included in the trend analysis. The tests do not compare the most recent estimate with the oldest estimate included in the trend analysis. Instead, the tests follow the trend across all time points. The trend line may not fit the time series particularly well because, for example, the actual time series shows a curvilinear pattern. If the estimates increased drastically during the early years of the time series but decreased in recent years, the linear trend test may detect an increasing trend, thus oversimplifying the actual pattern. For the regional trends, the estimated drug reports are standardized to the most recent regional population totals for those aged 15 years or older.