ARTS Contract No.: ART4-0003

Statistical Methodology

Revised September 2017

NATIONAL FORENSIC LABORATORY INFORMATION SYSTEM

Prepared for

Ms. DeMia Pressley, DEA Government Project Manager
Ms. Artisha Polk, DEA Mathematical Statistician
Diversion Control Division
Drug Enforcement Administration
U.S. Department of Justice

Prepared by

RTI NFLIS Staff
Peter Frechtel
David Heller
Paul Mosquin

RTI International
3040 E. Cornwallis Road
Research Triangle Park, NC 27709

September 2017

NFLIS
NATIONAL FORENSIC LABORATORY INFORMATION SYSTEM

DRUG ENFORCEMENT ADMINISTRATION
DIVERSION CONTROL DIVISION

Acknowledgments

This publication was originally prepared in 2014 under DEA Contract No. DJD-13-C-0035 and was revised in 2017 under ARTS Contract No. ART4-0003 for the Drug Enforcement Administration, U.S. Department of Justice. Points of view or opinions expressed in this document do not necessarily represent the official position of the DEA or the U.S. Department of Justice.

Public Domain Notice

Obtaining Additional Copies of This Publication

Electronic copies of this publication can be downloaded from the NFLIS website at https://www.nflis.deadiversion.usdoj.gov/.

STATISTICAL METHODOLOGY

Overview

Since 2001, National Forensic Laboratory Information System (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 provides an overview of the methods used for producing estimates, including sample selection, weighting, imputation, and trend analysis procedures.

Since the last NFLIS Statistical Methodology report was published in April 2014, two methodological changes have been implemented. First, beginning with the 2016 Annual Report in 2017, all drugs reported in an item are now counted in the estimation process. This change ensures that the estimates will take into consideration all reported substances, including emerging drugs of interest. Further details are provided in the "Other Changes to Estimation" section of this report. Second, a change in which the covariances of the totals are directly estimated in the long-term trend analysis was also introduced in the 2016 Annual Report in 2017. Additional details are included in the "Statistical Techniques for Trend Analysis" section of this report.

Original Sampling Design

RTI International, under contract to the U.S. Drug Enforcement Administration (DEA), began implementing NFLIS in 1997. Results from a 1998 survey (updated in 2002, 2004, 2008, and 2013) provided laboratory-specific information, including annual caseloads, which was used to establish a national sampling frame of all known State and local forensic laboratories that routinely perform drug chemistry analyses. For sampling and estimation purposes, State systems (and the multi-laboratory local systems known to exist) were treated as a single laboratory; so, if a State system was selected, all laboratories in the system were selected. The sampling frame of laboratories was divided into four strata by two stratifiers: (1) type of laboratory (State system or municipal or county laboratory) and (2) determination of "certainty" laboratory status. 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). To ensure that the NFLIS sample had strong regional representation, U.S. census regions were used as the geographical divisions to guide the 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. A total of 25 certainty systems or laboratories were identified. A probability proportional to size (PPS) sample of 35 laboratory systems and laboratories 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, including the certainty laboratories. In NFLIS publications released before 2011, data reported by nonsampled laboratories were not used in national or regional estimates.[1] However, as the

[1] The case and item loads for the nonsampled laboratories were used in calculating the weights.

number of nonsampled laboratories reporting to NFLIS increased,[2] it began to make sense to consider ways to use the data they submitted.

The remaining sections of this document highlight the current methodology and estimation processes accounting for additional laboratories recruited into NFLIS and adjusted for changes to the incoming data from participating NFLIS laboratories.

NEAR Methodology

A new statistical method used to calculate estimates for drug reports, referred to as NEAR (National Estimates Based on All Reports), was introduced in the 2010 Midyear Report to more fully exploit the high rate of reporting laboratories.[3] Under this NEAR method, the "volunteer" laboratories (i.e., the reporting nonsampled laboratories) represent themselves and are no longer represented by the reporting sampled laboratories. The volunteer laboratories are assigned weights of one; hence, the weights of the sampled and responding laboratories are appropriately adjusted downward. The outcome is that the estimates are more precise, especially for recent years, which include a large number of volunteer laboratories. 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 reporting period, which results in missing monthly data. If a laboratory reports fewer than six months of data for the annual estimates (fewer than three months for the semiannual estimates), it is considered nonreporting, 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 is used for cases, items, and drug-specific reports and accounts for the typical month-to-month variation and the size of the laboratory requiring imputation. The general idea is to use the nonmissing 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 nonmissing months in laboratory L,

[2] In 2016, for example, out of 113 nonsampled laboratories and laboratory systems, 86 (or 76%) reported.

[3] Currently, laboratories representing more than 98% of the national drug caseload participate in NFLIS, with about 97% of the national caseload reported for each reporting period.

$c_{L,m}$ = 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 nonmissing monthly item counts within the laboratory. The imputed value for the missing monthly number of items in each laboratory is calculated by multiplying $c_{L,m}$ 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 nonmissing 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 presented previously for the case and item counts. 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 consistently 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. Drug report-to-case ratios for each drug are produced for the similarly sized laboratories, and these drug-specific ratios are then used to adjust the drug report counts for the relevant laboratories.

NEAR weighting procedures

Each NFLIS reporting laboratory is assigned a weight to be used in calculating design-consistent, nonresponse-adjusted estimates. Two weights are created: one for estimating cases and one for estimating drug reports. The weight used for case estimation is based on the caseload for every laboratory in the NFLIS population, and the weight used for drug report estimation is based on the item load for every laboratory in the NFLIS population. For reporting laboratories, the caseload and item load used in weighting are the reported totals. For nonreporting laboratories, the caseload and item load used in weighting are based on completion-based data obtained from an updated laboratory survey administered in 2013, or, in some cases, via direct communication with laboratories or other external sources.

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[4] represented by the individual laboratory or laboratory system. This step takes advantage of the original PPS sample design and provides precise estimates as long as the drug-specific case and report counts are correlated with the overall caseload and item load.[5]

During the weighting process, laboratories are further categorized into 16 strata by region (Northeast, Midwest, South, and West), in addition to type of laboratory (State system or municipal or county laboratory) and certainty status, which were both used in defining the sampling strata. For noncertainty reporting laboratories in the sample (and reporting laboratories in the certainty strata with nonreporting 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 = ith laboratory or laboratory system;
- A = sum of the case (item) counts for all of the laboratories and laboratory systems (sampled and nonsampled) within a specific stratum, excluding certainty strata and the volunteer stratum; and
- B = number of sampled laboratories and laboratory systems within the same stratum, excluding certainty strata and the volunteer stratum.

Certainty laboratories were assigned a design weight of one.[6]

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

$$NR_j = C/D,$$

where

- j = stratum;
- C = number of sampled laboratories and laboratory systems in the stratum, excluding the volunteer stratum; and

[4] See the introduction of the most recent NFLIS Annual Report for a description of the NFLIS universe.

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

[6] With respect to the design weight, reporting laboratories and laboratory systems in certainty strata with nonreporting laboratories and laboratory systems are treated the same way as reporting noncertainty 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 regardless of their size.

D = number of laboratories and laboratory systems in the same stratum that were sampled and reporting.

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

NEAR estimation

The estimates in the NFLIS Annual Reports and Midyear Reports are the weighted sum of the counts from each laboratory. The weighting procedures make the estimates more precise by assigning large weights to small laboratories and small weights to large laboratories.[7] 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 tend 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 nonreporting, nonsampled 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 nonsampled laboratories.

Other Changes to Estimation

In addition to the NEAR method, two other changes to the estimation methodology were introduced in the 2010 Midyear Report. First, estimates are now based on cases and items *submitted* to laboratories during the reporting period and *analyzed* within three months of the end of the reporting period. Analysis has shown that at least 95% of cases submitted during a six-month or one-year period were analyzed within three months of the end of the period (not including the approximately 30% of cases that are never analyzed). In prior years, estimates were based on completed cases and items. Completion-based data are still used to create final item and case weights for reporting laboratories and laboratory systems.

Second, the estimation procedures began to account for multiple drugs per item. Instead of only counting the first drug listed as in prior reports, for each drug item (or exhibit) analyzed by a laboratory in the NFLIS program, up to three drugs could be reported to NFLIS and counted in the estimation process. These two changes were applied to all previous reporting periods to maintain the ability to compare drug trends.

A further enhancement to account for multiple drugs per item was introduced in 2017 with the 2016 Annual Report. All drugs reported in an item are now counted in the estimation process. This change ensures that the estimates will take into consideration all reported substances, including emerging drugs of interest, that may typically be reported as the fourth or fifth drug within an item. This change was implemented in the 2016 data processing cycle and will be used in future years. Although this change could not be applied to reporting periods before 2016, the 2016 data showed that 99.97% of drug reports

[7] See footnote 5 on Lohr (2010).

are captured in the first, second, or third drug report for any item; therefore, no statistical adjustments were deemed necessary to maintain the trend with prior years.

Suppression of Unreliable Estimates

For some drugs, such as cannabis/THC and cocaine, 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, a suppression rule was 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 greater than 50% are suppressed and not shown in the tables.

Statistical Techniques for Trend Analysis

Two types of analyses to compare estimates across years are used. The first is called *prior-year comparisons* and compares national and regional estimates between the current reporting period and the reporting period from one year ago (e.g., from January 2015 through December 2015 with those from January 2016 through December 2016 for the 2016 Annual Report). The second is called *long-term trends* and examines trends in the annual national and regional estimates from January 2001 through the end of the most recent reporting period. The long-term trends method described as follows was implemented beginning with the 2012 Midyear Report. The new method offers the ability to identify linear and curved trends, unlike the method used in previous NFLIS publications. Both types of trend analyses are described as follows. For the region-level prior-year comparisons and long-term trends, the estimated drug reports are standardized to the most recent regional population totals for persons aged 15 years or older.

Prior-year comparisons

For selected drugs, the prior-year comparisons statistically compare estimates from the current reporting period with estimates from the reporting period one year ago (e.g., estimates in Table 1.1 of the 2016 Annual Report with estimates in Table 1.1 of the 2015 Annual Report). The specific test examines whether the difference between any two estimates is significantly different from zero. A standard *t*-test is completed using the statistic

$$t_{df} = \frac{a\hat{T}_{YYYY} - b\hat{T}_{YYYY-1}}{\sqrt{a^2 \text{var}\left(\hat{T}_{YYYY}\right) + b^2 \text{var}\left(\hat{T}_{YYYY-1}\right) - 2ab\,\text{cov}\left(\hat{T}_{YYYY-1}, \hat{T}_{YYYY}\right)}},$$

where

df = appropriate degrees of freedom (number of laboratories minus number of strata),

$\hat{T}_{YYYY}$ = estimated total number of reports for the given drug for the current reporting period,

$\hat{T}_{YYYY-1}$ = estimated total number of reports for the given drug for the reporting period from one year ago,

$\text{var}\left(\hat{T}_{YYYY}\right)$ = variance of $\hat{T}_{YYYY}$,

$\text{var}\left(\hat{T}_{YYYY-1}\right)$ = variance of $\hat{T}_{YYYY-1}$, and

$\text{cov}\left(\hat{T}_{YYYY-1}, \hat{T}_{YYYY}\right)$ = covariance between $\hat{T}_{YYYY-1}$ and $\hat{T}_{YYYY}$.

For the national prior-year comparisons, $a = b = 1$. For the regional prior-year comparisons, a = 100,000 divided by the regional population total for the current reporting period, and b = 100,000 divided by the regional population total for the reporting period from one year ago.

The percentile of the test statistic in the t distribution determines whether the prior-year comparison is statistically significant (a two-tailed test at $\alpha = 0.05$).

Long-term trends

Long-term trend analyses are performed on the estimates from January 2001 through the most recent reporting period and on regional estimates of rates for selected drug reports. The models allow for randomness in the totals and rates due to the sample and the population. That is, for the vector of time period totals over that time,

$$\mathbf{Y}^T \equiv \left(Y_1, \ldots, Y_n\right),$$

and for the estimates,

$$\hat{\mathbf{Y}}^T \equiv \left(\hat{Y}_1, \ldots, \hat{Y}_n\right),$$

the regression model is

$$\hat{\mathbf{Y}} = \boldsymbol{X\beta} + \boldsymbol{\eta} + \boldsymbol{\varepsilon},$$

where

n = number of estimates in the time series;

$\boldsymbol{\eta} = \hat{\mathbf{Y}} - \mathbf{Y}$, an $n \times 1$ vector of errors due to the probability sample; and

$\boldsymbol{\varepsilon}$ = an $n \times 1$ vector of errors due to the underlying model.

Randomness due to the sample exists because only a sample of all eligible laboratories has been randomly selected to be included. Randomness due to the population exists because many factors that can be viewed as random contribute to the specific total reported by a laboratory in a time period. For example, not all drug seizures that could have been made were actually made, and there may have been some reporting errors. If rates (per 100,000 persons aged 15 years or older) and not totals are of interest, the aforementioned model can be applied to $\hat{\mathbf{Y}}^* = c\hat{\mathbf{Y}}$, where c equals 100,000 divided by the 15-or-older regional population size as given by the U.S. Census Bureau.

The regression model used to perform the analysis is

$$Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_m t^m + \varepsilon_t \qquad t = 1, \ldots, T,$$

where

Y_t = population total value, considered to be a realization of the underlying model; and

ε_t = one of a set of *n* independent normal variates with a mean of zero and a variance of σ^2.

The model allows for a variety of trend types, depending on the maximal polynomial degree *m* of the analysis, such as the following: linear (straight-line; m = 1), quadratic (U-shaped; m = 2), and cubic (S-shaped; m = 3), and quartic (higher-order shape; m = 4). Because it is a model for Y_t but the sample estimates $\hat{Y}_t$ differ by the sampling error, estimation is performed by restricted maximum likelihood (REML), allowing for the two sources of error.

To implement the regression model, point estimates of totals $\hat{Y}_t$ and their standard errors are obtained for all *n* reporting periods. Sampling standard errors are estimated as the full sampling variance-covariance matrix **S** over these *n* time periods. The **S** matrix contains variances in totals at any time period and covariances in totals between any two time periods, thus giving a very general modeling of the sampling variance structure. The variance-covariance matrix of the totals is then $Var[\hat{\mathbf{Y}}] = \sigma^2\mathbf{I} + \mathbf{S}$, where **I** is the identity matrix.

Before the 2016 Annual Report, the variance and covariance components of the **S** matrix for the means were estimated simultaneously. The variance-covariance matrix for the means was then converted into a variance-covariance matrix for the totals. A change was introduced in 2017 in which the covariances of the totals are directly estimated, and the estimation of the covariance of the means is no longer necessary. This change in the computation of the covariance of totals provides an incremental improvement over the previous approach and theoretically provides more valid statistical inferences. In addition, it creates consistency in the covariance estimation between these long-term trends and the prior-year comparisons.

Regression coefficients are estimated using the REML method. Because higher-order polynomial regression models generally show strong collinearity among predictor variables, the model is reparameterized using orthogonal polynomials. The reparameterized model is

$$Y_t = \beta_0 X_0(t) + \beta_1 X_1(t) + \beta_2 X_2(t) + \cdots + \beta_m X_m(t) + \varepsilon_t \qquad t = 1, \ldots, T,$$

where

$X_0(t) = 1/\sqrt{T}$ for all *t*, and

$X_1(t), \ldots, X_m(t)$ provide contributions for the first-order (linear), second-order (quadratic), and higher-order polynomials, respectively.

Note that the error term is the same in the original model and the reparameterized model because the fitted surface is the same for both models. The model is further constrained to have regression residuals sum to zero, a constraint that is not guaranteed by theory for these models but is considered to improve model fit because of an approximation required to estimate **S**. Standard errors of the regression trend estimates are obtained by simulation.

Final models are selected after testing for the significance of coefficients at the $\alpha = 0.05$ level ($p < .05$), which means that if the trend of interest (linear, quadratic, cubic) was in fact zero, then there would be a 5% chance that the trend would be detected as statistically significant when in fact it is not. Final fitted models are most easily interpreted using graphical plots.