



NIOSH Manual of Analytical Methods (NMAM), 5th Edition

Development and Evaluation of Methods

by Eugene R. Kennedy, Ph.D., Thomas J. Fischbach, Ruiguang Song, Ph.D., Peter M. Eller, Ph.D., Stanley A. Shulman, Ph.D., and R. DeLon Hull, Ph.D., NIOSH

1	Method Development	ME-2
2	Method Evaluation	ME-5
3	Field Evaluation	ME-13
4	Documentation	ME-14
5	Appendix - Accuracy and its Evaluation	ME-15
6	References	ME-17

DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Institute for Occupational Safety and Health





1 Method development

The development and evaluation of analytical methods that are useful, reliable and accurate for industrial hygiene monitoring problems require the application of some general guidelines and evaluation criteria. The guiding objective in this work requires that, over a specified concentration range, the method provide a result that differs no more than $\pm 25\%$ from the true value 95 times out of 100. The application of consistent evaluation criteria and guidelines is particularly important when methods are developed by different individuals and organizations (e.g., contractors or outside laboratories) and compiled into a single manual. Adherence to guidelines should minimize overlooking potential problems in the methodology during its development, as well as provide cohesiveness and uniformity to the method that is developed. This chapter provides an outline of a generalized set of evaluation criteria prepared by NIOSH researchers for the evaluation of sampling and analytical methodology [NIOSH 1995].

In the development of a sampling and analytical method, there is a logical progression of events that cover a search of the literature to gather pertinent information and the preliminary experimentation for selection of analysis technique and sampling medium. To initiate the development of a method, the identity of the analyte must be as fully defined as possible. Physical and chemical properties of the analyte should be defined so that procedures for proper handling and use of the analyte can be prepared. These also aid in establishment of analyte purity. Potential sources of this information include chemical reference books, health hazard evaluation reports, bulk sample analyses, material safety data sheets, chemical process information, etc.

Since innovation is a key element in the sampling and analytical method development process, detailed experiments for the initial development of the sampling approach and optimization of the analytical procedure are better left to the discretion of the researcher. During development, it should be recognized that appropriate, statistically designed experiments will optimize the amount of information obtained. Therefore, consultation with a statistician about appropriately designed experiments will be of value during this phase of the research.

a. Preliminary experimentation

Several key points, including calibration and selection of measurement technique and sampling media, should be studied during the initial method development experiments. The selection of sampling medium and procedure is a decision that usually is made early in the method development process. The physical state of the analyte (i.e., gas, aerosol, vapor, or combination thereof) plays an important factor in the selection of an appropriate sampler. Analytes which can exist in more than one physical state may require a



combination of sampling media in one sampler for efficient collection [NIOSH 1995]. Where possible, commonly available and easily used samplers should be investigated initially. As the preliminary testing of a sampling method progresses, further modification in the sampling medium or sampler design may be required and may affect the measurement procedure. Sampler design and media selection considerations should include U.S. Department of Transportation regulations and restrictions for shipment back to a laboratory for analysis.

Since industrial hygiene analytical methods are geared toward measuring personal exposure, the size, weight, and convenience of the sampler are important elements in sampler design. The personal sampler should allow freedom of movement and should be unobtrusive, unbreakable, and not prone to leakage. The pressure drop across the sampler should not be so great as to limit sample collection times to 10 h with personal sampling pumps. For situations where only a short term sample will be required (i.e., 15 min for ceiling determinations), this 10 h recommendations can be reduced to 1 h. The use of potentially toxic reagents should be avoided unless they can be used safely. Reagents used should not pose any exposure hazard to the worker wearing the sampler or to the industrial hygienist taking the samples.

b. Recovery of the analyte from the medium

During the course of method development experiments, the ability to recover the analyte from the sampling medium should be determined. A suggested experiment to accomplish this entails the fortification of sets of 6 samplers with amounts of analyte equivalent to sampling concentrations of 0.1, 0.5, 1.0, and 2.0 (or higher) times the exposure limit for a minimum of 4 h at the typical sampling rate used for that type of sampler. If the analyte has a ceiling or short-term exposure limit, the amount of analyte fortified should be adjusted for the shorter sampling time required for this type of exposure limit. If the sampler has a backup section, then a like number of separate backup sections should be fortified with amounts of analyte equivalent to 25% of the amount fortified on the front sections of the samplers, since this amount has been used to characterize the breakthrough limit of useful samples [Streicher et al. 1994]. Samples (and backup sections) should be prepared for analysis and analyzed according to previously determined procedures. Results of these analyses should be expressed in terms of estimated percent recovery according to the following formula:

$$\text{Percent Recovery}_{\text{(est.)}} = \left[\frac{\text{Amount of analyte found on sampler}}{\text{Amount of analyte fortified on sampler}} \right] \times 100 \%$$

After initial analyses of the samples, the samples should be resealed and analyzed on the following day, if possible. If the sample workup procedure results in a solution of the



sample, these solutions should be recapped after the initial analysis, if possible, and reanalyzed on the following day using fresh standards.

The recovery of the analyte should be calculated for the primary and backup media in the sampler. Although complete recovery of the analyte from the sampler is most desirable, at a minimum, the estimated recovery of the analyte from the primary collection medium should be greater than or equal to 75% for concentrations equivalent to sampling 0.1, 0.5, 1.0, and 2.0 times the exposure limit. If recovery varies with analyte loading, results should be graphed as recovery versus loading during calibration of the method, so that appropriate correction can be made to sample results, as long as recovery is greater than 75% [Melcher et al. 1978]. If estimated recovery does not exceed 75%, the method is not suitable for monitoring at this limit.

Estimated recovery from any backup media should be noted so that appropriate corrections can be applied if breakthrough of the sampler has occurred during sampling. The recovery of the analyte from the medium in the backup section of a sampler may be different from that of the front section, since the backup section of a sorbent-based sampler usually contains only half of the sorbent of the primary section. If the same volume of desorption solvent is used for both the primary and backup sections of the sampler, the desorption equilibrium can be shifted, since the backup section is being desorbed by twice the volume (i.e., on a mL solvent/mg sorbent basis) [Saalwaechter et al. 1977].

Reanalysis of the samples on the day after initial analysis indicates if immediate analysis after sample preparation is required. Often when processing a large number of samples, it may be necessary to prepare the samples for analysis as a batch. In these instances, the last samples may not be analyzed for up to 24 h or more after preparation because of the time required for analysis. If samples prepared for analysis exhibit time-dependent stability after desorption, analyses must be conducted within acceptable time constraints. Analysis and reanalysis results should agree within 5% of each other.

c. Stability of the analyte on the medium

An extension to the experiment described above may be performed to investigate potential stability problems early in the experimentation. An additional set of fortified samples at each of the 4 concentrations should be prepared and analyzed after 7-days' storage at room temperature. Recovery should be similar to the above results within experimental error. Discrepancies larger than those expected by experimental error indicate sample stability problems that will need correcting by additional developmental effort (e.g., refrigerated storage). Comparison of results can be performed with statistical tests, such as an analysis



of variance (ANOVA) [Posner and Okenfuss 1981] test of the “Day” difference or a paired t-test [Box et al. 1978] of the means of the Day 1 and Day 7 storage results.

2 Method evaluation

After the initial development experiments for the method have been completed and a method has been proposed, the sampling and analysis approach should be evaluated to ensure that the data collected provides reliable, precise, and accurate results. Specifically, the goal of this evaluation is to determine whether, on the average, over a concentration range of 0.1 to 2 times the exposure limit, the method can provide a result that is within $\pm 25\%$ of the true concentration 95% of the time. For simplification, the true concentration is assumed to be represented by an independent method. An experimental approach for collecting the data necessary for this determination is described below.

As part of the evaluation of a method, the sampling of a generated atmosphere is needed to more adequately assess the performance of a method [NIOSH 1984; Nelson 1971; Nelson 1992]. This allows the determination of 1) the capacity of the sampler; 2) the efficiency of analyte collection by the sampler; 3) the repeatability of the method; 4) the bias in the method; 5) interferences in the collection of the sample. Concentration ranges to be used in the evaluation of the method should be based on several factors. These ranges, at a minimum, should cover 0.1 to 2.0 times the exposure limit. In some instances, higher multiples of the exposure limit can be added if needed (e.g., 10 times the exposure limit). In situations where multiple exposure limits (i.e., from different authorities) exist for an analyte, the lowest exposure limit should be used to set the lower limit of the evaluation range (0.1 times lowest exposure limit) and the highest limit used to calculate the upper limit of evaluation range (2 times the highest exposure limit). Intermediate evaluation concentrations should be within these exposure limits. The toxicity of an analyte (e.g., suspected carcinogenicity) may indicate that a concentration lower than that calculated by the exposure limit should be included in the measurement and evaluation ranges. Previous monitoring information from other methods may indicate that typical concentrations of the analyte may be below or above a concentration range based on the exposure limit. In this case, this lower or upper level may be included in the method evaluation.

a. Feasibility of analyte generation

In order to provide a realistic test of the method under study, air concentrations covering the range from 0.1 to 2 times the exposure limit of the analyte should be generated. The generated atmospheres should be homogeneous in concentration and representative of the environment encountered when sampling for the analyte in the workplace.



When attempting to generate a concentration of an analyte, the impact of environmental conditions, such as temperature, pressure, humidity, and interferences, on sampler performance and/or generation should be considered. The effect of elevated temperature on the collection medium of a sampler may decrease the capacity of the sampler or may decompose the analyte during generation and sampling. Reduced pressure may also reduce the capacity of a sampler. High relative humidity in many instances has been observed to reduce sampler capacity [Melcher et al. 1978]. In other instances it has increased sampler capacity [Cassinelli 1991]. A typical interference(s) should be generated along with the analyte to approximate a typical workplace sampling environment.

Generation of particulate material can be extremely complex [Willeke 1980; Hinds 1982], especially if particles of a required size range must be generated for the evaluation of a specified sampler inlet design. The aerodynamic performance of the generator is a factor in the generation of this type of atmosphere and should be evaluated carefully. Appropriate, independent methods should be available to verify particle size, if this is a critical element in the generation.

The concentration of the generated atmosphere should be verified either by well characterized gravimetric/volumetric means or by analysis of replicate samples (if possible) by an independent method at each concentration used. Further details on this verification are included in the literature [NIOSH 1995; Ashley 2015]. A statistician should be consulted for advice on the design and sample sizes to accomplish this validation. Ideally, the independent method should not be biased and should provide an accurate estimate of the concentration generated, assuming error is randomly distributed around the mean. Also the precision and bias of the independent method should be homogeneous over the concentrations investigated. (See NIOSH [1995] for the definitions of these attributes.) In instances where the concentration of the generator can be based only on calculations using flow rates in the generator and the amount of analyte injected, the generation system should be well characterized so that analyte losses are minimized.

In some instances, generation of an analyte may be difficult and even hazardous. As an alternative to direct generation in these cases, samplers may be fortified with an amount of analyte expected to be sampled over a specified period of time at a specific flow rate. When this is necessary, fortification of the sampler by vaporization of a known amount of analyte onto the sampling medium is a more appropriate method, since this approach more closely approximates a generated atmosphere. The alternative of direct application of a solution of analyte onto the collection medium is less desirable but may be necessary in some instances. After fortification, air, conditioned at both high and low humidity, should be drawn through samplers at the flow rate and time period used in the calculations for the



amount of analyte expected to be collected. In the method report, the fact that samples were not collected from a generated atmosphere should be discussed.

b. Capacity of the sampler and sampling rate

To determine the applicability of the sampling method, the capacity of the sampler should be determined as a function of flow rate and sampling time. This is particularly important if the analyte has both a short-term exposure limit (STEL) and a time-weighted average. Flow rates typical for the media selected should be used. These may range from 0.01 to 4 L/min, depending on sampler type. At extremely low flow rates (ca. 5 mL/min), the effect of diffusion of the analyte into the sampler must be considered. Flow rates should be kept at a high enough rate to prevent diffusion from having a positive bias in the sampler. Sampling should be performed at three different flow rates covering the range appropriate for the particular sampler type, unless the sampler is designed to operate at only one flow rate.

Sampling times should range from 22.5 min for STELs to 900 min (15 h) for time-weighted averages. Shorter sampling times (e.g., 7.5 to 22.5 min) may be used for ceiling (C) measurements. Flow rates should be based on accurately calibrated sampling pumps or critical orifices. The amount of analyte collected at the lowest flow rate and shortest sampling time should be greater than the limit of quantitation of the method. The generated concentration used for capacity determination should be at least 2 times the highest published exposure limit and verified by an independent method.

Sampling should be conducted at ambient, elevated ($>35\text{ }^{\circ}\text{C}$), and low ($<20\text{ }^{\circ}\text{C}$) temperatures to assess the effect of temperature on sampling. To assess the effect of humidity on capacity, sampling should be performed at both low and high humidity (20% and 80%), since both have been observed to affect capacity [Cassinelli 1991; Melcher et al. 1978]. Triplicate samplers at three different flow rates should be included to verify capacity at each of the six different humidity and temperature levels. For samplers which contain backup sampling media, only the front section of the sampler should be used. A means is required to quantitate analyte in the effluent from the sampler. This may involve the use of a backup sampler, continuous monitor or other appropriate means which can provide a measure of analyte concentration in the sampler effluent (ca. 1 to 5% of the influent concentration). If the mass of analyte found on a backup sampler totals 5% of the mass found on the front sampler or if the effluent concentration of the sampler contains 5% of the influent concentration, breakthrough has occurred and the capacity of the sampler has been exceeded.



If the analyte is a particulate material and collected with a filter, the capacity of the filter is defined by the pressure drop across the sampler or by the loading of the filter. For 37-mm filter-based samplers, pressure drop should be less than 1000 mm of water for total loading less than 2 mg. Larger filters and especially filter capsules [Harper and Ashley 2013; O'Connor et al. 2014] will tolerate higher loadings (e.g, up to 5 mg).

If the collection process is based primarily on adsorption, breakthrough time should be proportional to the inverse of the flow rate [Jonas and Rehrmann 1973]. This relationship can be checked by plotting the 5% breakthrough time versus the inverse of the flow rate. If the resulting plot is a straight line, then this relationship should hold for all flow rates in the flow rate range studied. Some nonlinearity in the plot may be noted due to experimental variability and assumptions made to simplify the relationship of breakthrough time and flow rate. Results from these experimental trials should provide a prediction of the capacity of the sampler at various flow rates and sampling times. If the flow rates and sampling times used in the experiment do not provide for sufficient capacity, a lower flow rate range may have to be studied and the experiment repeated.

With samplers which use reagents for collection of the analyte, the amount of the reagent in the sampler will also be a limiting factor in the capacity of the sampler, based on the stoichiometry of the reaction. Other factors, such as residence time in the sampler and kinetics of reaction between analyte and reagent, may affect the capacity of this type of sampler.

The combined temperature and humidity conditions that reduce sampler capacity to the greatest extent should be used in all further experiments. The Maximum Recommended Sampling Time (MRST) for a specific flow rate is defined as the time at which sampler capacity was reached, multiplied by 0.667. This adds a measure of safety to this determination. The relationship of breakthrough time with flow rate can be used to adjust flow rates to optimize specific sampling times.

c. Sampling and analysis evaluation

To assess the performance of a method, certain additional experimental parameters should be evaluated through a series of defined experiments. The effect of environmental conditions (e.g., pressure, interferences) on sampling efficiency of the sampling medium can be evaluated by a factorial design [Box et al. 1978]. The temperature, relative humidity, flow rate, and sampling times, determined in the experiment described above to have most severely limited sampler capacity, should be used in these experimental runs [Ashley 2015]. At a minimum, the effect of concentration on method performance should be investigated. Three sets of 12 samples should be collected from an atmosphere containing



concentrations of 0.1, 1.0, and 2.0 times the exposure limit at the humidity determined above to have reduced sampler capacity for the MRST determined in the preceding experiment.

If the analyte has a short-term or ceiling exposure limit in addition to an 8-hour time-weighted average, an additional 12 samplers should be collected at the STEL or C limit for the recommended sampling period at the appropriate flow rate. Potential interferences in the work environment should be included in the generation experiments to assess their impact on method performance. Concentrations up to 2 times the exposure limit value for the interference should be included. Other environmental factors may be studied, but will require a more comprehensive experimental design.

The effects of environmental conditions on analyte recovery should be assessed. A factorial design can be used to evaluate these factors to determine which exert a significant effect on analyte recovery. Those factors which are found to influence analyte recovery should be investigated further to determine if their impact is predictable. If these effects are not predictable, the utility of the method will be limited, based on the conditions defined by this experiment. If only concentration is evaluated, the analyte recovery should be the same at all concentrations after correctable biases have been included, such as desorption efficiency.

d. Sample stability

To assess sample stability, samples should be collected from a generated atmosphere, stored under defined conditions (i.e., ambient or refrigerated, light or dark), and analyzed at specified time periods. A concentration of 0.5 times the lowest exposure limit should be sampled with 30 samplers for a minimum of $\frac{1}{2}$ the MRST. The humidity and temperature of the generator should be at the same level as defined in the sample capacity experiment to reduce sample capacity. The samplers should be divided randomly into one group of 12, one group of 6, and four groups of 3, with the group of 12 analyzed as soon after collection as possible (Day 0). The group of 6 samplers should be analyzed after 7 days. The four remaining sets of 3 samplers should be analyzed after 10, 14, 21, and 30 days. The conditions of storage are determined by the nature of the analyte. If there is an indication of analyte instability on the sampling medium, refrigeration of the samplers may be required. However, storage for the first 7 days should be at room temperature.

Samples should normally be stable for a minimum of 7 days under ambient conditions to simulate shipping to a laboratory for analysis. If the average analysis results of the samplers analyzed on day 7 differs from the set analyzed on day 0 by more than 10%, the method does not meet the sample stability criterion. Either additional precautions, such as



shipment on ice and refrigerator storage, may be required or the method may have to be modified to address this problem. (Note: In practice, reactive or unstable samples should be shipped by overnight mail and analyzed as quickly as possible.) If a plot of recovery versus time indicates that recovery decreased by more than 10% after the initial 7-day storage period, sample instability is a problem. If samples need to be stored for longer periods, more restrictive storage conditions are required. Remedial action, such as cold storage may solve this longer term storage problem. After remedial precautions have been instituted in the method, the sample stability of the method must be determined anew.

e. Precision, bias, and accuracy

Results from four sets of samplers used in the analyte recovery experiment, the sampling and analysis experiments (e.g., the environmental parameters experiments), and the sample stability experiment can be used for the estimation of precision, bias, and accuracy of the method. A more exacting treatment of this is described elsewhere [NIOSH 1995]. Sampler results from the multi-level factorial design at the 0.1, 1.0, and 2.0 times the exposure limit value; the sampler stability experiment (at 0.5 times the exposure limit); and the environmental factors experiment are used in the calculations of method precision. The calculations for the estimated method precision, \hat{S}_{rT} , have been described previously [NIOSH 1995; Anderson et al. 1981; Busch and Taylor 1981; NIOSH 1980]. Before obtaining a pooled estimate of method precision from the four sets of samplers listed above, the homogeneity of the precision over the range of concentrations studied should be checked using a test, such as Bartlett's test [NIOSH 1995; Anderson et al. 1981; Busch and Taylor 1981]. If the precision is not found to be constant over concentrations, the sample set collected at 0.1 times exposure limit should be removed and Bartlett's test recalculated. Homogeneity of the method precision at all concentration levels is an assumption required to obtain pooled estimate of method precision.

Bias is assumed to be homogeneous over the evaluation range. This assumption should be tested by estimating the bias at each concentration and testing these for homogeneity using the procedures described in the literature [NIOSH 1995]. Method bias should be less than 10%. A test for this is also described [NIOSH 1980].

The bias and precision estimates can be used with the graph presented in Figure 1 or in Table I to estimate accuracy [NIOSH 1995]. The bias and precision estimates are plotted on the x- and y-axes of the graph. The intersection of these points on the parabolic grid in the graph can be used to estimate the accuracy of the method. This procedure gives an estimate of method accuracy but does not yield the statistic required to test compliance of the method with the $\pm 25\%$ accuracy criterion. Techniques for the latter determination are discussed in the Appendix and elsewhere [NIOSH 1995].

If the results for 4 concentrations fail the 25% accuracy criterion, then the set of samples collected at the lowest concentration level should be excluded from the data set. The pooled \hat{S}_{RT} and the bias should be recalculated on this reduced data set before performing the accuracy analysis described in the previous paragraph.

For the 12 samplers collected at the ceiling limit, the accuracy analysis described above should be repeated using only the data collected at the ceiling limit.

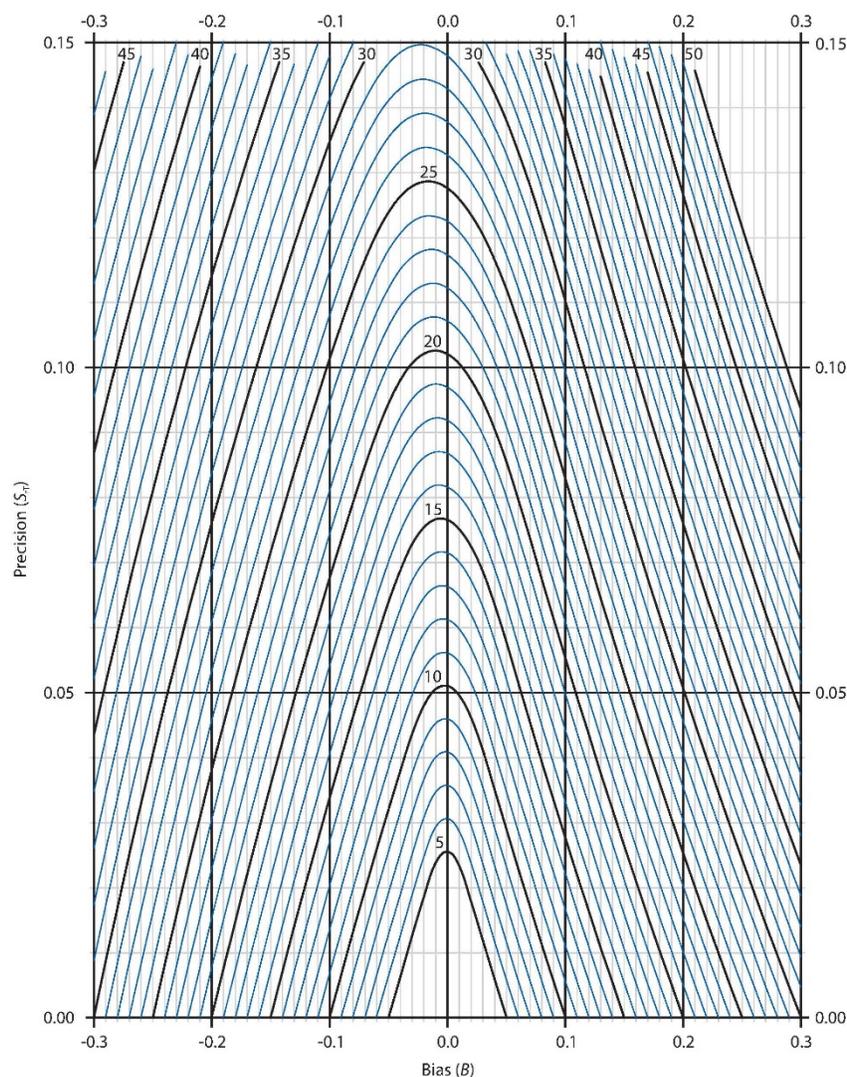


Figure 1. Nomogram relating accuracy to precision and bias. Accuracy (A), in percentage units, is a function of the bias (B) and the precision (S_{RT}). Each curve is the locus of all points (B, S_{RT}) that yield the value of A indicated on the curve.



Table I. Values of the bias (B) and the precision (S_{rt}) required to obtain designated values of accuracy (A) in percentage units[#]

A (%)	B (%)	S_{rt} (%)
5	-3.5	0.945*
5	-2.5	1.559*
5	0.0	2.551*
5	2.5	1.483*
5	3.5	0.881*
10	-7.5	1.643*
10	-5.0	3.200
10	0.0	5.102
10	5.0	2.895*
10	7.5	1.414*
15	-10.0	3.378*
15	-5.0	6.381
15	0.0	7.653
15	5.0	5.774
15	10.0	2.764*
20	-10.0	6.755
20	-5.0	9.448
20	0.0	10.20
20	5.0	8.548
20	10.0	5.527
25	-10.0	10.13
25	-5.0	12.39
25	0.0	12.75
25	5.0	11.21
25	10.0	8.287
30 [§]	-15.0 ^{&}	10.73
30 [§]	-7.5	14.55
30 [§]	0.0	15.31
30 [§]	7.5	12.52
30 [§]	15.0 ^{&}	7.930
35 [§]	-15.0 ^{&}	14.31
35 [§]	-7.5	17.59
35 [§]	0.0	17.86
35 [§]	7.5	15.14
35 [§]	15.0 ^{&}	10.57

[#] Note: the values shown in this table are population or theoretical values.

* Below the minimum attainable precision with a 5% pump correction.

[§] Does not fulfill the Accuracy Criterion ($\pm 25\%$ of the true value).

[&] Does not fulfill the bias criterion ($\pm 10\%$).



3 Field evaluation

While field evaluation is not required in method evaluation, it does provide a further test of the method, since conditions which exist in the field are difficult to reproduce in the laboratory. Also unknown variables may affect sampling results when field samples are taken. This type of evaluation is recommended to further study the performance of the method in terms of field precision, bias, interferences and the general utility of the method. Both the collection of area samples and personal samples should be included in the field evaluation of the method.

Area samples should provide an estimate of field precision and bias. Personal samples may confirm these values and also provide a means to assess the utility of the method. A statistical study design should be prepared, based on the variability of the method and the statistical precision required for estimates of the differences in analyte concentrations yielded by the independent method and the method under evaluation [CEN 2015].

If this type of statistically designed study is not feasible, a minimum of 20 pairs of samples of the method under study and an independent method should be used for personal sampling. Placement of the samplers on the workers should be random to prevent the biasing of results due to the "handedness" of the worker. Workers sampled should be in areas where both low and high concentrations of the analyte may be present.

As a minimum, sets of 6 area samplers paired with independent methods should be placed in areas of low, intermediate, and high analyte concentration. If the atmosphere sampled is not homogeneous, precautions may have to be taken to ensure that all samplers are exposed to the same concentrations. This can be done by using field exposure chambers, such as those described in the literature [Cassinelli et al. 1985; Kennedy et al. 1985].

Field precision and bias of the area sampler results of the method under study should compare with laboratory evaluation results, provided that precautions have been taken to ensure that all samplers have been exposed to the same homogeneous atmosphere. Differences in precision and bias should be investigated. Sources of variation should be studied and corrections implemented where necessary. Evaluation of personal sampler results should be done cautiously, because observable differences may be due to work practices or other situations which are beyond the control of the method.

A field evaluation of a method also allows the developer of the method to determine its ruggedness. Although this may be a subjective judgement, first-hand experience with the method in the field may suggest changes in the sampler or method that may make the method more easily used in the field and less subject to variability.



4 Documentation

Development and evaluation research on a sampling and analytical method should be documented in a final report. The report should describe what was determined about the method. If the results of the statistical analysis of the data indicate there is not 95% confidence that the accuracy of the method is less than or equal to $\pm 25\%$, the report should state this fact. In some instances, the method may actually have an accuracy of less than 25%, but a larger sample size must be used to prove this statistically (See Appendix 1 of NIOSH 1995).

The final report can be either a technical report or a failure report. The technical report (acceptable method developed) documents the successful development of the method. This report may be prepared in a format appropriate for submission to a peer-reviewed journal for publication. The failure report (no acceptable method developed) documents the research performed on an attempted method development for an analyte or analytes. The report should describe the failure of the method, as well as other areas of the method research that were successful. Recommendations to solve the failure of the method may be included.

If an acceptable method is developed, a sampling and analytical method should be prepared in appropriate format. The format of the resulting method should provide clear instructions for the use of the method. Sampling, sample workup, and analysis procedures should be clearly described. The necessary equipment and supplies for the method should be listed clearly in the method. A summary of the evaluation of the method should be included, as well as a discussion of method applicability and lists of interferences and related references. As a check on the clarity and performance, new methods should be reviewed and submitted to a user check (i.e., the method is used to analyze spiked or generated samples of known concentration by someone other than the researcher who developed it) and to a collaborative test, if feasible.



5 Appendix - Accuracy and its evaluation

In the development of a sampling and analytical method, one of the goals is to minimize the measurement error to the lowest feasible and practical levels. It is assumed that all feasible corrections to reduce error have been made in the laboratory experimentation process. Method evaluation requires adequate characterization of the magnitude and distribution of the uncorrectable error that cannot be prevented. One might consider a hypothetical experiment in which a method is used repeatedly to measure the same concentration, T , under the same conditions. These measurements would tend to exhibit a pattern or statistical distribution, here assumed to be normal, with a mean, μ , and standard deviation, \hat{S} . The distribution can be characterized in terms of two components: its location relative to T , which is the systematic error termed bias (B), is given by $(\mu - T)/T$; and its spread, which is the random error termed imprecision (\hat{S}_{rT}), is given by \hat{S}/μ . The bias and imprecision are used to determine the inaccuracy of the method but they are also important characteristics of the error in and of themselves, as will be discussed below.

Accuracy refers to the closeness of the measurements to T but it is defined in terms of the discrepancy of the measurements from T . Inaccuracy (I) is defined as the maximum error, regardless of sign, expressed as a percentage of T that occurs with a probability of 0.95. Thus, an inaccuracy (or accuracy) of 20% means that on the average 95 of every 100 measurements will differ from T by no more than $0.2T$. The accuracy criterion for single measurements mentioned at the beginning of this chapter, often termed the "NIOSH Accuracy Criterion," requires inaccuracy to be less than or equal to 25%.

Accuracy, bias, and imprecision have the following relationship:

$$0.95 = \Phi((1-B)/((1+B)S_{rT})) - \Phi((-1-B)/((1+B)S_{rT}))$$

where Φ denotes the probability that a standard normal random variable is less than or equal to Φ . A practically exact numerical solution to Equation (1) can be readily programmed using statistical software packages [Press et al. 1986]. A DOS program, ABCV.EXE, is also available which solves for I (denoted by A in the program), \hat{S}_{rT} (denoted by CV in the program), or B when the values for the other two quantities are input. An estimate of I can be obtained in either case by entering estimates of B and \hat{S}_{rT} . An approximate solution, which is accurate to about 1.1 percent, is given as follows [NIOSH 1995]:

$$I = 1.57 (B+1)S_{rT} + \sqrt{((0.39 (B + 1) \cdot S_{rT})^2 + B^2)} \quad \text{for theoretical or true } I$$

$$\hat{I} = 1.57 (\hat{B} + 1)\hat{S}_{rT} + \sqrt{((0.39 (\hat{B} + 1) \cdot \hat{S}_{rT})^2 + \hat{B}^2)} \quad \text{for estimates of } I$$



Also, the nomogram in Figure 1 can be used to solve for I or an estimate of I by entering B and S_{RT} or their estimates. Procedures for obtaining “best” single point and 95% confidence interval estimates of B , and S_{RT} and a 90% confidence interval estimate for I are given in NIOSH [1995].

The 90% confidence interval for I can be used to infer whether the method passes or fails the 25% accuracy criterion (AC) for single measurements with 95% confidence as follows:

- 1) The method passes with 95% confidence if the interval is completely less than 25%.
- 2) The method fails with 95% confidence if the interval is completely greater than 25%.
- 3) The evidence is inconclusive if the interval includes 25% (there is not 95% confidence that the AC is true or that it is false).

When researchers interpret the results from analyses of the type described above, it is important to consider that most methods have many uses in addition to individual measurement interpretation. Because accuracy is very important whenever any quantity is to be estimated, the ideal (“other things being equal”) is to use the most accurate estimator regardless of its bias or imprecision. However, it is crucial to distinguish between the accuracy of the source or “raw” measurements and that of the final estimator, which might involve many intermediate analyses or operations. Unfortunately, the most accurate input or raw measurements do not always produce the most accurate final result unless the latter is a single measurement. The bias and imprecision of the source measurements can be differentially affected by intermediate operations in producing the final estimate. For example, if the final estimate is a function of a single average of many source measurements, its bias is not affected by the averaging while imprecision is reduced as a function of the square root of the number of measurements. Thus, a lower biased method might be preferable to another even if the inaccuracy of the latter is less. On the other hand, in comparative studies, the desired estimate is either a difference or ratio of means of measurements in which there can be partial or complete cancellation of the bias in the source measurements. Thus, the bias of the method used for the source measurements may be of little importance. If there are several methods applicable for a given user’s project (regardless of whether all fulfill the accuracy criterion for single measurements), the analyst would be well-advised to consult with the user (preferably in advance of measurement) to determine which of those methods would produce the optimal accuracy for the final results or estimates needed by that particular user. Accuracy, bias, and imprecision jointly form a complete or sufficient set for the efficient description of the measurement error characteristic of any method.



Disclaimer

Mention of any company or product does not constitute endorsement by NIOSH. In addition, citations to websites external to NIOSH do not constitute NIOSH endorsement of the sponsoring organizations or their programs or products. Furthermore, NIOSH is not responsible for the content of these websites. All web addresses referenced in this document were accessible as of the publication date.

6 References

Anderson C, Gunderson E, Coulson D [1981]. Sampling and analytical methodology for workplace chemical hazards. In: Choudhary G, Ed. *Chemical Hazards in the Workplace*. Washington: American Chemical Society, pp. 3-19.

Ashley K [2015]: NIOSH Manual of Analytical Methods 5th edition and harmonization of occupational exposure monitoring. *Gefahrstoffe Reinh Luft* 75:7-16 (and references therein).

Box GEP, Hunter WG, Hunter JS [1978]. *Statistics for experimenters*. New York: John Wiley & Sons.

Busch K, Taylor D [1981]. Statistical protocol for the NIOSH validation tests. In: Choudhary G, ed. *Chemical Hazards in the Workplace*. Washington: American Chemical Society, pp. 503-517.

Cassinelli ME [1991]. Development of a solid sorbent monitoring method for chlorine and bromine in air with determination by ion chromatography. *Appl Occup Environ Hyg* 6:215-226.

Cassinelli ME, Hull RD, Cuendet PA [1985]. Performance of sulfur dioxide passive monitors. *Am Ind Hyg Assoc J* 45:599-608.

Comité Européen de Normalisation (CEN) [2015]. EN 482:2012+A1:2015 Workplace exposures - General requirements for the performance of procedures for the measurement of chemical agents. Brussels, Belgium: Comité Européen de Normalization.

Harper M, Ashley K [2013]. Acid-soluble internal capsules for closed-face cassette elemental sampling and analysis of workplace air. *J Occup Environ Hyg* 10:297-306.

Hinds WC [1982]. *Aerosol technology*. New York: John Wiley & Sons.



Jonas LA, Rehrmann JA [1973]. Predictive equations in gas absorption kinetics. *Carbon* 11:59-64.

Kennedy ER, Smith DL, Geraci CL [1985]. Field evaluations of sampling and analytical methods for formaldehyde. In: Turoski V, ed. *Formaldehyde – Analytical Chemistry and Toxicology*, Washington, DC: American Chemical Society, pp. 151-159.

Melcher R, Langner R, Kagel R [1978]. Criteria for the evaluation of methods for the collection of organic pollutants in air using solid sorbents. *Am Ind Hyg Assoc J* 39:349-361.

NIOSH [1980]. Development and validation of methods for sampling and analysis of workplace toxic substances. By Gunderson E, Anderson C, Smith R, Doemeny L, Busch K. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health; DHHS (NIOSH) Publication No. 80-133.

NIOSH [1984]. Gas and vapor generating systems for laboratories. By Woodfin W. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health; DHHS (NIOSH) Publication No. 84-113.

NIOSH [1995]. Guidelines for air sampling and analytical method development and evaluation. By Kennedy ER, Fischbach TJ, Song R, Eller PM, Shulman SA Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health; DHHS (NIOSH) Publication No. 95-117.

O'Connor S, O'Connor PF, Feng HA, Ashley K [2014]. Gravimetric analysis of particulate matter using air samplers housing internal filtration capsules. *Gefahrstoffe Reinh Luft* 74:403-410.

Posner J, Okenfuss J [1981]. Desorption of organic analytes from activated carbon. *Am Ind Hyg Assoc J* 42:643-646.

Press WH, Flannery BP, Teukolsky SA, Vettering WT [1986]. *Numerical recipes – the art of scientific computing*. Cambridge, UK: Cambridge University Press.

Saalwaechter A, McCammon C, Roper C, Carlberg K [1977]. Performance of the NIOSH charcoal tube technique for the determination of air concentrations of organic vapors. *Am Ind Hyg Assoc J* 38:476-486.



Streicher R, Kennedy E, Lorderau C [1994]. Strategies for simultaneous collection of vapours and aerosols with emphasis on isocyanate sampling. *Analyst* 119:89-97.

Willeke K, ed. [1980]. *Generation of Aerosols and Facilities for Exposure Experiments*. Ann Arbor, MI: Ann Arbor Science Publishers.