Chapter 6 in Upgrading Environmental Radiation Data, Health Physics Society Committee Report HPSR-1 (1980), U.S. Environmental Protection Agency Report EPA 520/1-80-012.

REPORTING OF ENVIRONMENTAL RADIATION MEASUREMENTS DATA

R. Collé (National Bureau of Standards, Washington, DC), H. H. Abee (Union Carbide Corp., Oak Ridge, TN), L. K. Cohen (Nuclear Regulatory Commission, Washington, DC), D. Ed (State of Illinois, Dept. of Public Health, Springfield, IL), E. H. Eisenhower (National Bureau of Standards, Washington, DC), A. N. Jarvis (Environmental Protection Agency, Las Vegas, NV), I. M. Fisenne (Environmental Measurements Laboratory, New York, NY), M. Jackson (Alabama Power Co., Birmingham, AL), R. H. Johnson, Jr. (Environmental Protection Agency, Washington, DC), D. Olson (Radiological and Environmental Sciences Laboratory, Idaho Falls, ID) and J. Peel (Department of Energy, Washington, DC)

This report is intended to serve as a practical guide to treating and reporting environmental radiation measurements data. Recommendations for a uniform method of data reporting are presented and justified. Three primary requisites are considered: proper units, an appropriate number of significant figures, and an unambiguous statement of measurement uncertainty. Present practices are summarized and evaluated, and their deficiencies are examined. To avoid confusion, it is recommended that the existing multiplicity of units used to report various radiation quantities be reduced to a smaller consistent set. Use of the metric system of units is encouraged. Rules for rounding reported values to an appropiate number of significant figures are presented. The appropriate number of significant figures for a reported value is determined by the magnitude of the total uncertainty associated with the value. Guidelines are given for estimating random and systematic uncertainties, and for propagating and combining them to form an overall uncertainty. It is recommended that each reported measurement result include the value, the total random uncertainty expressed as the standard deviation, and the combined overall uncertainty. To avoid possible biases of data, all measurement results should be reported directly as obtained, including negative values. The lower limit of detection (LLD) should serve only as an a priori estimate of detection capability for the instrumentation, and not as an absolute level of activity that can or cannot be detected. The concept of a minimum detectable concentration (MDC) is introduced to serve as an a priori estimate of the capability for detecting an activity concentration with a given measurement instrument, procedure, and type of sample. Neither the LLD nor the MDC is intended to be an a posteriori criterion for the presence of activity.

(Accuracy; activity; data reporting; detection limit; environmental; error; lower limit of detetion [LLD]; measurements; minimum detectable concentration [MDC]; precision; radiation; random uncertainty; significant figures; statistics, systematic uncertainty; uncertainty; units)

Overview

This report highlights the guidelines and recommendations contained within a much lengthier version prepared by the subcommittee. The full report, A Guide and Recommendations for Reporting of Environmental Radiation Measurements Data, is available as a special publication of the National Bureau of Standards [1]. The guide and its recomendations are intended to be useful and specific. It is not intended to be a general or exhaustive theoretical treatment of measurement results, but rather to serve as a practical guide to treating and reporting environmental radiation measurements data.

There are three primal requisites to reporting measurement data. The reported value of a measurement result must always:

1) be unequivocal and properly dimensioned (i.e., the units must be clearly understood and suitable for the value);

2) be expressed in an appropriate number of significant figures; and

3) include an uncertainty statement (whose meaning is unambiguous).

Mistreatments of conditions (1) and (2) are usually not the most serious failings in data reporting. There is, however, considerable variation in the practices recommended and required by the various federal agencies concerned with environmental radiation data reporting. Recommended conventions to satisfy these conditions are covered in Parts I and II, respectively.

Failure to satisfy condition (3) is probably the most frequent abuse and is of greater concern. Unfortunately, almost all environmental radiation data presently being reported fall into one of three categories:

i) the value is reported without stating the uncertainty;

ii) the uncertainty is reported, but it is not specified:

iii) the reported uncertainty is specified, but is based on an incomplete assessment.

Further, federal practices pertaining to the reporting of uncertainties not only contribute to these abuses, but are also as diverse as the number of agencies and requirements specified within

an agency's regulations. Recommendations for assessing, propagating and reporting uncertainties are given in Part III.

Perhaps the most non-uniform practices in environmental radiation data reporting are those involving detection limits. This and the teatment of results at or below these levels are covered in Part IV.

I. Dimensioning (Units)

As outlined in the Overview, the first requirement for the reporting of environmental radiation measurements data is that reported values must "be unequivocal and properly dimensioned." That is, the units for the values must be clearly understood and must be suitable for the value.

A survey and evaluation of typical units currently in use indicates that there are considerable variations in practice. This existing multiplicity is both needless and confusing. Two additional observations can be made. First, compound units (e.g., for radioactivity concentration) are frequently formed by using more than one prefix in forming the multiple of the compound unit (e.g., $\mu Ci \cdot mL^{-1}$). Accepted convention recommends that only one prefix be used [2,3]. Normally the prefix should be attached to a unit in the numerator. The second observation is that the conventional units of radiation quantities (curie, roentgen, rad, rem) are almost exclusively used and that few have converted to use of the International System of Units (SI).

It would be beneficial and helpful (to both laboratories and data users) if the number of units in use were reduced to a smaller consistent set. This goal toward greater uniformity could be accelerated by the encouragement of national use of the metric system, formally called SI, and by the use of standard metric practices and conventions. The familiar "curie," "roentgen" and "rem" are deeply entrenched within the radiological sciences and radiation fields, and will not be easy to forsake. This natural reluctance to change can be appreciated, but the eventual conversion to SI is inevitable and must be recognized.

In addition to our national commitment to the SI, there are many advantages to its use. Some of these are excellently described in the following brief summary from a British committee report[4]:

> The International System of Units (SI) is a rational and comprehensive system of units. It has seven base units. The unit for any physical quantity within the system is derived from one or more of these by multiplication or division and without introducing any numerical factors. The system is therefore coherent. It avoid the chaos of individual choice of units, it breaks down the barriers to communication arising from separate systems of units, and it removes the need for conversion factors that arises when incoherent units are used. The SI is clearly destined to become the universal currency of science and commerce. At first sight the use of the new SI units seems to bring little direct benefit to those...areas in which radiation quantities are extensively used, but their adoption is essential if these areas are not to be cut off from their parent sciences. The use of the new units for ionizing radiations will only extend the existing and growing application of SI....All changes bring their problems but the introduction of SI units does not change existing quantities although it does require people to become familiar with new numerical values.

A number of sources describing the use of the SI are available [2,3,5,6]. For environmental radiation measurements data, we are primarily concerned with the units for radioactivity, exposure, absorbed dose, and dose equivalent. The status of each of these, in terms of both the SI and non-SI units, will be considered in turn.

Radiation Units

Radioactivity is a measure of the rate of those spontaneous, energy-emitting, atomic transitions that involve changes in state of the nuclei of radioactive atoms.[14] It is given by the quotient of dN by dt where dN is the number of spontaneous radioactive events which occur in a quantity of a radioactive nuclide in the time interval dt[7]. The SI unit of activity is the becquerel (Bq), which is equal to one radioactive event per second, i.e.,

$$1\ Bq = 1\ s^{-1}.$$

The special unit of activity is the curie (Ci) which is defined as

$$1\ Ci = 3.7 \times 10^{10}\ Bq \quad \text{(exactly)}.$$

The International Committee for Weights and Measures (CIPM) has authorized, for a limited (unspecified) time, the use of the curie and its multiples. This temporary sanction was provided to allow time for familiarization with the new units.

Exposure, X, is the quotient of dq by dm where dq is the absolute value of the total charge of the ions of one sign produced in air when all the electrons (negatrons and positrons) liberated by photons in a volume element of air

$$X = \frac{dq}{dm}$$

The SI unit for exposure is coulomb per kilogram ($C \cdot kg^{-1}$). The familiar unit for exposure is the roentgen (R) which is defined as

$$1\ R = 2.58 \times 10^{-4}\ C \cdot kg^{-1} \quad \text{(exactly)}.$$

There is no special name for the SI unit of exposure ($C \cdot kg^{-1}$) as a replacement for the roentgen. As for the curie, temporary use of the roentgen has been sanctioned.

Absorbed dose, D, is the quotient of $d\bar{\varepsilon}$ by dm where $d\bar{\varepsilon}$ is the mean energy imparted by ionizing radiation to the matter in a volume element and dm is the mass of the matter in that volume element [7]. The SI unit for absorbed dose and related quantities (specific energy imparted, kerma, and absorbed dose index)* is the gray (Gy). The gray, in SI base units, is a joule per kilogram.

$$1\ Gy = 1\ J \cdot kg^{-1}.$$

The gray is intended to replace the rad which is defined as

$$1\ rad = 10^{-2}\ Gy.$$

The unit rad has also been sanctioned for temporary use.

Dose equivalent accounts for the relative biological effectiveness of a given absorbed dose. It depends on the type of radiation, on the irradiation conditions, and, for a given organ, is inferred by weighting the absorbed dose in that organ by certain modifying factors[7]. The dose equivalent, H, is given as the product of D, Q and N, at the point of interest in tissue, where D is the absorbed dose, Q is the quality factor for the type or radiation, and N is the product of any other modifying factors.

$$H = DQN$$

The SI unit for dose equivalent is the sievert (Sv) which is equal to, in SI base units, $J \cdot kg^{-1}$. Hence, when D is expressed in grays (Gy), H is in sieverts (Sv). The special non-SI unit for the dose equivalent is the rem. When D is expressed in rads, H is in rems. The unit rem has also been sanctioned for temporary use.

Compound Units

Compound units for other quantities are formed by combination of two or more units. The SI unit for radioactivity concentration, for example, could be $Bq \cdot kg^{-1}$ or $Bq \cdot L^{-1}$. Exposure rate, given by the quotient of dX by dt where dX is the increment of exposure in the time interval dt, is expressed in units of any quotient of $C \cdot kg^{-1}$ by a suitable unit of time (e.g., $C \cdot kg^{-1} \cdot h^{-1}$). The unit for absorbed dose rate, $dD \cdot dt^{-1}$, is any quotient of the gray or its multiple or submultiple by a suitable unit of time ($Gy \cdot s^{-1}$, $\mu Gy \cdot h^{-1}$, $nGy \cdot yr^{-1}$, etc.). As will be discussed shortly, there are a number of advantages to using an exponential notation instead of prefixes. The use of prefixes to form multiples and submultiples of the SI units is permissible, however, and for completeness is considered here.

Multiples and Submultiples of Units

Multiples and submultiples of the SI units are formed by adding prefixes to the names or symbols of the units. The SI prefixes are listed in Table 1. The unit of mass, kilogram (kg), is the only base or derived unit of SI that, for historical reasons, contains a prefix. The kilogram (not the gram) is the base unit, but multiples and submultiples of the unit of mass are formed by adding prefixes to the gram (g). Prefixes on the units are normally chosen and used to avoid excessive non-significant digits and leading zeros in numerical values. Two conventions in the use of prefixes should be mentioned.

1. Compound prefixes formed by combining two or more SI prefixes should never be used. For example, use GBq (gigabecquerel) not kMBq, or use μGy (microgray) not mmGy.

2. Only one prefix should be used in forming multiples or submultiples of a compound unit. Normally the prefix should be attached to the first unit listed in a product, or to the unit in the numerator of a quotient. For example, use $MBq \cdot L^{-1}$ not $kBq \cdot mL^{-1}$. The use of the unit kilogram is the only exception to this rule since the kilogram is the base unit of mass.

One should consult more comprehensive guides [2,3,6] for additional details.

TABLE 1
SI Prefixes

FACTOR	PREFIX	SYMBOL
10^{18}	exa	E
10^{15}	peta	P
10^{12}	tera	T
10^{9}	giga	G
10^{6}	mega	M
10^{3}	kilo	k
10^{2}	hecto	h
10^{1}	deka	da
10^{-1}	deci	d
10^{-2}	centi	c
10^{-3}	milli	m
10^{-6}	micro	μ
10^{-9}	nano	n
10^{-12}	pico	p
10^{-15}	femto	f
10^{-18}	atto	a

Special Units

There are a number of units which are not part of the SI, but are so widely used that they are accepted for continued use with the SI. The combination of these units with SI units should be restricted to special uses in order to not lose the advantage of the coherence of SI units. Those of interest for environmental radiation measurements data include the units of time: minute (min), hour (h) and day (d); and the special name liter. The recommended symbol for the liter is "L" in the U.S., not the lower case "l" which may be confused with the numeral

*For definitions of these quantities see, for example, ICRU Report 19[7].

"l"[6]. The SI unit of volume is the cubic meter (m^3), and it or one of its multiples or submultiples is preferred for all applications. In any case, the special name liter is restricted for use with only liquid or gas volumes. No prefix other than milli should be used with liter.

As noted earlier, the common non-SI units for radiation quantities (curie, roentgen, rad and rem) are authorized, for a limited time, for use with the SI. These are destined to be replaced by the becquerel (Bq), coulomb per kilogram ($C \cdot kg^{-1}$), gray (Gy) and sievert (Sv), respectively.

Exponential Notation

There are a number of advantages to expressing results using powers of ten applied to the base and coherent derived SI units. First, it eliminates problems when values lie outside the range covered by the prefixes; second, it minimizes the possibility of computational mistakes; and third, it helps to indicate the number of significant figures in the value thereby reducing ambiguity (see Part II). Some examples of values expressed in SI units and this exponential notation follow:

$2810\ pCi \cdot L^{-1} = 104\ Bq \cdot L^{-1}$
$= 1.04 \times 10^{2}\ Bq \cdot L^{-1}$
$= 1.04\ E{+}02\ Bq \cdot L^{-1}$

$12\ \mu R \cdot h^{-1} = 3.1 \times 10^{-9}\ C \cdot kg^{-1} \cdot h^{-1}$
$= 3.1\ E{-}09\ C \cdot kg^{-1} \cdot h^{-1}$

$123\ mrad = 0.00123\ Gy$
$= 1.23 \times 10^{-3}\ Gy$
$= 1.23\ E{-}03\ Gy$

Conversion to SI Units

It obviously is neither feasible nor economically justifiable to completely and immediately change over from non-SI to SI units. This change will require familiarization and a transition period. Steps must be taken by both laboratories and regulatory agencies to aid and expedite this inevitable transition. This will only be achieved if there is a spirit of cooperation, adequate planning, and logical implementation by all those involved. At the same time, a number of initial steps taken during the transition would immediately further the goal of greater uniformity in data reporting practices.

It is therefore recommended that data reports be converted to the SI base and derived units as soon as technically and economically feasible. In many cases, this cannot be accomplished without changes in governmental regulations and data-reporting requirements imposed on laboratories. During the transition it is recommended that data reports include conversion factors between the SI and non-SI units, and whenever possible include the measurement results in both units. Since the scope of this change is much broader than the parochial scope of environmental radiation measurements data, efforts must be made by concerned laboratories, agencies, and organizations to implement simultaneous changes in all areas concerned with radiation quantities. Broad-based initiatives to plan and implement a safe and orderly transition are highly recommended.

Recommended Units for Use in Data Reporting

Whether temporarily sanctioned non-SI units (Ci, rad, etc.) or SI units (Bq, Gy, etc.) are employed, it is recommended that the number of combined units for radioactivity concentrations, exposure rates, etc., be immediately reduced to a smaller consistent set. The existing multiplicity of units in use is both needless and confusing.

It is recommended that radioactivity concentrations (except for effluent concentrations) be reported in units of activity per kg for mass measurements, and per m^3 or L for volume measurements. Recommended conventions are therefore:

1. Units for activity concentration in solid samples (e.g., soils, vegetation, etc.) should be $Bq \cdot kg^{-1}$ expressed in exponential notation (using the new SI units) or submultiples of $Ci \cdot kg^{-1}$ (using non-SI units).

2. Units for activity concentration in airborne particulate or gas samples should be $Bq \cdot m^{-3}$ expressed in exponential notation or submultiples of the $Ci \cdot m^{-3}$.

3. Units for activity concentration in liquid samples (e.g., water, milk, etc.) should be $Bq \cdot L^{-1}$ expressed in exponential notation or submultiples of $Ci \cdot L^{-1}$.

The recommended units for these activity concentrations and other radiation quantities, in terms of both the SI and non-SI units, are contained in Table 2. The table also illustrates the conversion factors (in E-notation) from the non-SI to SI units.

Additional Specifications

Obviously, additional specification of the results of measurements may be neessary. Measurements of solid sample, for example, will still require specification as to "dry weight" or "wet weight" and moisture content. Similarly, measutements of gas volumes require specification of the ambient temperature and pressure. One should always explicitly state all of the conditions that are necesssary for the results to be understood and interpreted unequivocally.

Other Unsatisfactory Current Practices

As stated initially, proper dimensioning of measurement results requires that the values and units be clearly understood and suitable for the value. There are a number of still unaddressed current practices that do not satisfy these basic criteria.

Perhaps one of the most serious is the failure to distinguish between measurements of "dose equivalent" and "absorbed dose." The abbreviation "dose" is continually, although incorrectly, used for both quantities. Oftentimes, the only way one can ascertain which is meant is by seeing if the values are reported in units of rems or rads. It should be apparent that not only must the measurement value be properly dimensioned, but the reported quantity itself must also be properly labelled.

A second major concern is the extensive reporting of "gross activity" ("total beta", "total gamma") measurement results in units of radioactivity. The inherent problems and abuses of such practices have been addressed previously[8]. These measurements contain no information on the identity of the radionuclides present in the sample. Therefore, no meaningful conclusions on the dosimetric significance of the results can be obtained. "Gross activity" measurements do serve a useful function for screening purposes to decide if additional, radionuclide-specific measurements should be made. Their value in data reports, however, can be seriously questioned. Reporting values of "gross activity" in units of Bq or pCi (or some other submultiple of the curie) is extremely misleading. At best, the result refers only to the activity that would be obtained from the observed counting rate for the standard radionuclide (that was used to obtain the counting efficiency). This conditional must be clearly described in any report of the result.

Recommendations

1.1 A consistent set of units for various environmental radiation quantities (for both SI and non-SI units) is contained within this guide.(cf., Table 2) and should be employed for data reporting. The existing multiplicity in the number of units in use is both needless and confusing.

1.2 The use of the metric system, formally called the International System of Units (SI), and the use of standard metric practices and conventions should be encouraged. These units and recommended practices are briefly described within this guide. Additional details are available from more comprehensive sources referenced therein.

1.3 Data reports should convert to the SI base and derived units for radiation quantities as soon as technically and

TABLE 2
Recommended Units For Data Reporting

	In Non-SI Units [a]	In SI Units	Conversion Factor From Non-SI to SI Units [b]
Activity Concentrations (Environmental)			
Airborne Particulates and Gas	$pCi \cdot m^{-3}$	$Bq \cdot m^{-3}$	3.70E-02
Liquids (Water, Milk, etc.)	$pCi \cdot L^{-1}$	$Bq \cdot L^{-1}$	3.70E-02
Solids (Soil, Sediment, Vegetation, Foodstuff, etc.)	$pCi \cdot kg^{-1}$	$Bq \cdot kg^{-1}$	3.70E-02
Activity Concentrations (Effluent)			
Gas (Air)	$(\mu Ci \cdot mL^{-1})$*	$Bq \cdot m^{-3}$	3.70E+10
Liquid	$(\mu Ci \cdot mL^{-1})$*	$Bq \cdot L^{-1}$	3.70E+07
Exposure Rate (Environmental)	$\mu R \cdot h^{-1}$	$C \cdot kg^{-1} \cdot h^{-1}$	2.58E-10
Absorbed Dose	mrad	Gy	1.00E-05
Dose Equivalent	mrem	Sv	1.00E-05
Dose Equivalent Rate (Commitment)	$mrem \cdot yr^{-1}$	$Sv \cdot yr^{-1}$	1.00E-05

[a] Sanctioned for temporary use.

[b] To convert non-SI units to SI units, multiply the non-SI units by the conversion factor.

* Adopted because of established convention and use in Maximum Permissible Concentration (MPC) tabulations.

economically feasible. In view of existing governmental regulations and data-reporting requirements, it is, at present, neither feasible nor economically justifiable to completely and immediately change over from non-SI to SI units. This change will require familiarization and a transition period.

II. Significant Figures

The second requirement (outlined in the Overview) for reporting measurement data was that the reported value must be "expressed in an appropriate number of significant figures." Fortunately, this is nearly universally recognized, and is seldom a problem.

The "number of significant figures" refers to the number of numerical digits that is used to express the value. It is obvious that this number must be reasonable, and must not mislead or imply fictitious accuracy in the reported value. An "appropriate" number of digits (or significant figures) is that which is warranted by the accuracy of the reported value. That is, the appropriate number of digits to be retained in the reported value depends on the magnitude of the total uncertainty (see Part III) attached to this value.

For example, a reported value of 1.53365 E+12 Bq(41:45 Ci) for the quarterly tritium release from a nuclear facility is not reasonable. Surely no one would be fooled by the apparent implied accuracy of this value. Suppose further that the estimated total uncertainty in the value is +40%. Then, with the result expressed as

$$(1.53365 \pm 0.61346)\ \text{E+12 Bq}$$
$$(\text{i.e., } 41.45 \pm 16.58\ \text{Ci}),$$

the ludicrousness in the reportd value is even more apparent. This result should be reported as (1.5 + 0.6) E+12 Bq.

Similarly, suppose the annual release of radioactivity from several facilities at a site was reported (to two significant figures) as 21,000 Ci from one, 560 Ci from another, and 1.4 Ci from a third. To then say that the total activity released from the site was 21,561.4 Ci implies an accuracy in the measurement that is both incorrect and unintended.

Another, equally obvious, bad practice is the failure to ensure decimal agreement between the result and its associated uncertainty. For example, results such as 1.2 + 0.002 or 1.234 $\pm$ 0.2 are internally inconsistent. Either the result is more accurate than is given, as indicated by its associated uncertainty, and more significant figures should be retained; or the result is much less accurate than indicated, and the number of significant figures should be decreased to agree with the accuracy indicated by the quoted uncertainty. Care must therefore be taken in the number of significant figures reported both for the value itself and for the uncertainty term.

In general, environmental radiation measurements data seldom justify more than two or three significant figures for the value and one or two significant figures for the uncertainty. More significant figures should only be reported after careful consideration and a decision that the extra figures are indeed reportable. It is recommended that the uncertainty should be reported to no more than two significant figures, and the value itself should be stated to the last place affected by the qualification given by the uncertainty term.

For example, given a measurement result of 123.45 $Bq \cdot L^{-1}$ for an activity concentration with an estimated total uncertainty of $\pm$12% (i.e., $\pm$14.8 $Bq \cdot L^{-1}$), the result would be reported as

$$(1.23 \pm 0.15)\ \text{E+02 Bq} \cdot L^{-1}.$$

In this case, two significant figures for the uncertainty and three for the value might be justified. One might not wish to round the result to

$$(1.2 \pm 0.1)\ \text{E+02 Bq}\ L^{-1}$$

since the latter result implies a greater accuracy (approximately 8% compared to the original estimated uncertainty of +12%). The decision to report either two or three significant figures for a value is particularly critical when the value is close to running from one decade to another, such as for 89, 95, and 103. With estimated uncertainties of $\pm$15%, the above values would be reported as

$$89 \pm 13$$
$$95 \pm 14$$
$$103 \pm 15$$

if two significant figures in the uncertainty term are retained. Otherwise, the values would be rounded to

$$90 \pm 10$$
$$100 \pm 10$$
$$100 \pm 20.$$

The rules for rounding of numbers (i.e., the dropping of insignificant or unwarranted figures) are generally well known and available in a large number of sources [9, 10, 11, and 12], including elementary mathematics texts. These rules are summarized below:

1. To round off the number to n significant figures, truncate the number to n digits and treat the excess digits as a decimal fraction.

2. If the fraction is greater than or equal to 1/2, increase the least significant digit (i.e., the nth) by 1.*

3. If the fraction is less than 1/2 do not increase and leave the nth digit unchanged.

Similarly, rules for the number of significant figures to be retained in successive arithmetic operations have been developed [10, 11]. The rules for addition and subtraction differ from those for multiplication and division, and involve other assumptions (such as, that the values are independent and uncorrelated). Since no general rule can be given for all situations involving different types of arithmetic operations, a recommended procedure is to avoid rounding off until after the final calculation. This can be accomplished by treating all measurement values as exact numbers in the operations and to only round off the final result; or to carry two or more extra (insignificant) figures throughout the computation, and then to round off the final reported value to an appropriate number of significant figures [9]. Carrying along the extra insignificant figures will reduce the possibility of "rounding errors." This practice does not, however, excuse one from rounding the final result to an appropriate number of significant figures for the data reports.

These views are to a large degree currently incorporated into most data reporting practices and regulatory requirements.

Recommendations

II.1 A reported value should be expressed in an appropriate number of significant figures which is determined by the magnitude of the total uncertainty asssociated with the value.

II.2 The uncertainty should be reported to no more than two significant figures, and the value itself should be stated to the last place affected by the qualification given by the uncertainty term.

II.3 Care should be taken in the number of significant figures reported both for the value itself and for its associatd uncertainty. The value and its uncertainty must be in decimal agreement.

II.4 Reported values and their uncertainties should be rounded to an appropriate number of significant figures using the consistent, well-developed rules outlined in the text.

II.5 To avoid "rounding errors" in computations involving successive arithmetic operations, two or more extra (insignificant) figures should be carried on all the values throughout the computation, and then the final reported value should be rounded.

III. Treatment of Uncertainty Statements

Three of the most common abuses concerning the reporting of "errors" for environmental radiation measurements data were outlined in the Overview to this guide. These abuses shall be addressed in turn, but it may be useful to state, at the onset, the necessary conditions for avoiding their failings:

> A REPORTED VALUE MUST INCLUDE AN ASSESSMENT OF ITS UNCERTAINTY.
>
> THE REPORTED UNCERTAINTY MUST BE CLEARLY UNDERSTOOD AND MUST CONVEY SUFFICIENT INFORMATION SO THAT ITS MEANING, USING CORRECT TERMINOLOGY, IS UNAMBIGUOUS.
>
> THE REPORTED UNCERTAINTY MUST BE BASED ON AS NEARLY COMPLETE AN ASSESSMENT AS POSSIBLE.

To help avoid ambiguity, the use of the words "error" and "uncertainty" (which are frequently used interchangeably) should be clarified. The word "error" has familiar usages which range from "a mistake" or "an oversight" to "a deviation from what is correct, right, or true," "the condition of having incorrect or false knowledge," and "the difference between a measured or computed value and a correct value." Further, its use in even the statistical sense is often confusing. Therefore, it is recommended that it not be used except:

1) in those cases where its meaning is unambiguous or of no portent, e.g., "absolute error";

2) when used in uniquely defined statistical terms, e.g., "standard error of the mean"; and

3) in commonly recognized phrases, e.g., "propagation of errors" or "statistical theory of errors."

As such the word "error" should be used as infrequently as possible. The word "uncertainty" is a preferred substitute which can be used to refer to the values following the ± symbol and all of its component parts (random and systematic uncertainties). Terms such as "uncertainty," "accuracy," etc. are more likely to be clearly defined, better understood, and unambiguously used. Data reports are less apt to be confusing if the frequent and loose use of "error" is avoided. This practice has been incorporated into this guide, and it is recommended that it

*This rule will introduce a slight rounding bias due to the treatment of the result when the fraction is exactly equal to 1/2. An unbiased procedure consists of incrementing the nth digit if it is an odd number (leaving the nth digit unaltered if it is an even number).

be followed for all environmental radiation measurement data reports and documents.

Consider now the necessary conditions, given above, for reporting an uncertainty.

A reportd value without an accompanying uncertainty statement is for nearly all purposes worthless. The value is rendered useless because it cannot be put to use with any confidence. Although the particular form of the uncertainty statement may be dependent on the intended or ultimate use of the result, it is without debate that the value must include one.

The absolute error or uncertainty of a reported value, which is the deviation from the true value, is unknowable because the true value can never be known exactly. Limits to this uncertainty, however, can be inferred and estimated from the measurement process itself. Foremost, the uncertainty should be a statement, based on a credible assessment, of the likely inaccuracy or the likely limits to the "absolute error" in the reported value. This uncertainty assessment also includes an incumbent risk of being incorrect.

Similarly, a reported value, say 15 ± 3 Bq, without further information is equally troublesome. The user would be forced to speculate whether the quoted 20% uncertainty is a standard error of the mean based on multiple measurements, or an estimated "statistical or counting error." Also unknown is whether it includes possible systematic uncertainties or is only the random uncertainty, and to what level of confidence the value can be ascribed. The possibilities are innumerable. In short, the user might ask: "How certain am I that the value is between 12 and 18 Bq?" To be useful, the uncertainty statement must convey sufficient information so that its meaning, based upon correct terminology, is unambiguous.

Suppose the above quoted uncertainty was reported as being a "counting error at 2 sigma (95%) confidence interval" [Cf., Environmental Radiation Data, quarterly compilations by the U.S. Environmental Protection Agency's Office of Radiation Programs]. Such reporting is certainly clearer. The user could infer that it is derived from an estimate of the standard deviation of the measurement or counting process (assuming it follows a Poisson distribution) taken to some higher confidence limit ("2 sigma").* This inference can be problematic however. First, although it is frequently assumed, twice the standard deviation is not necessarily a 95% confidence limit (see subsequent section on confidence limits). And second, the reported uncertainty is only an estimate of the random uncertainty or precision in the measurement process which does not address the accuracy of the value. The user might legitimately ask: "Is this the sole contribution to the overall uncertainty?"

Precision, Accuracy, Bias, and Random and Systematic Uncertainties

Precision and Random Uncertainty. The random uncertainty is a statement of precision (or more correctly, of imprecision) and is a measure of the reproducibility or scatter in a set of successive independent measurements [Ref. 13, p. 23-1]. Precision refers to the closeness of the set of results among themselves. When differences in the magnitudes of the observations are small, precision is said to be high; when the differences are large, precision is low. Random uncertainties are assessed and propagated by statistical methods. The treatment of these uncertainties is most familiar to physical scientists and those concerned with environmental radiation data.

Accuracy. The accuracy of a measurement process is a measure of the ability to obtain closeness to the true value. The absolute error of a particular result is just the difference between the measured or reported result and the true value. The exact difference of course is unknowable because the true value can never be known exactly. Although the absolute error is unknowable, limits to its magnitude can be inferred and estimated from the measurement process itself. The estimate of these limits to the absolute error is referred to as the uncertainty, and is used to estimate the inaccuracy of the measurement process.

Bias. A bias is a deviation from the true value which is always of the same magnitude and direction. It cannot be estimated or calculated from a given set of replicate measurements since each and every measurement is affected by the systematic bias in the same way [Ref. 13, p. 23-1 and Ref. 14]. There can be many contributing sources of bias in a given measurement process. They are introduced by the process and are characteristic of it. Such biases are not amenable to statistical treatments. They should be estimated upper limits for each conceivable (or assessable) source of inaccuracy in the measurement process. Their magnitudes would preferably be based on experimental verification, but may have to be estimated from experience and judgement.

The very familiar "bull's eye" example shown in Figure 1 should help illustrate the distinction between precision and systematic biases, and their relation to accuracy. The bull's eye of the target corresponds to the true value, and the six shots represent individual measurement results. The figure illustrates the concept of an inaccurate measurement due largely to imprecision, a precise but inaccurate measurement, and an accurate measurement. Inasmuch as accuracy requires precision, there is no such case as an accurate but imprecise measurement.

*The term "sigma," in this case, is used incorrectly. Sigma is a parameter for the population, and should not be used to refer to the calculated sample statistic "standard deviation."

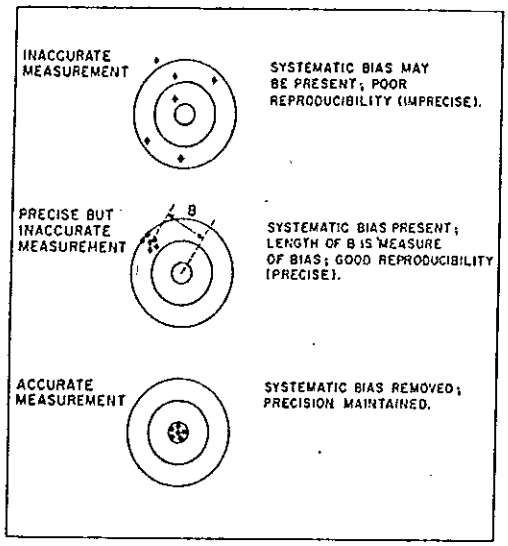


Figure 1. Illustration of the distinction between precision and systematic biases, and their relation to accuracy.

Systematic Uncertainty. For practical purposes, we may define systematic uncertainty to consist of those sources of inaccuracy which are biased, and those which may be due to random causes (stochastic processes) but cannot be or are not assessed by statistical methods. This broader definition which includes sources of inaccuracy in addition to biases is necessary to obtain a complete accounting of all sources of inaccuracy in the total uncertainty. It is the total uncertainty (obtained by combining the random and systematic components) that is ultimately used to estimate the inaccuracy.

"Blunders" and Data Rejection

Before proceeding to discussions of random and systematic uncertainties, it may be useful to consider the other source of inaccuracy. These are the outright mistakes or "blunders" in the measurement process. If one or more blunders are known to exist, then the value should be discarded, or a correction should be applied to the measurement. Parratt [Ref. 11, p. 69] described some possible blunders and their sources. A few examples which could easily be applied to environmental radiation measurements follow:

1. Incorrect logic, or misunderstanding what one is or should be doing. For example, applying a decay correction in the wrong direction.
2. Misreading of an instrument. Perhaps a field survey meter is an exigent situation, such as on a cold rainy day.
3. Errors in transcription of data which might occur in recording a scaler reading, transferring a value from one page to another, or even mis-entering a number into a calculator or computer terminal.
4. Calculational and arithmetical mistakes, which include misplaced decimal points, etc.

Interestingly, Parratt considered confusion of units and listing of an improper number of significant figures to be blunders.

One must recognize that blunders do and will occur. With an appreciation of this fact, every effort should be made to detect or avoid them. Many can be detected by establishing a system of independent checks on each stage of the measurement and data reduction system. A simple but effective system is an informal arrangement with a partner, colleague, or co-worker to check each other's work. It should include checking all stages of the data reduction (arithmetic, etc.) and an independent appraisal of the logic and the measurement process used to obtain the reported value. This checking is also useful in trying to assess all of the sources of systematic error. In effect, the checker serves as a "devil's advocate" to discover where the blunders may be.

Lastly, there is the question of rejecting suspicious or questionable data. When replicate results are available, standard statistical tests may be applied to decide if outliers should be rejected. Details on applying rejection tests can be found in Natrella [13, p. 17-1], Dixon and Massey [15], and other standard statistical texts. Although this is an acceptable practice, caution must be exercised. Rejection of data may result in an unrealistic self-consistent set and may actually bias the result. A further problem is that it may mask a real, anomolous and unexpected effect. Had Lord Rutherford's graduate students, Geiger and Marsden, applied data rejection tests to the small number of "unexpected" large scatterings in their α-particle scattering experiments, they might have failed to discover or interpret the existence of the atomic nucleus.

Jaffey[16] outlined three reasons why one should be conservative in rejecting moderate outliers:

a) Apparent patterns in sequences of random data are often startling. In the long run, averaging only bunched results gives averages that deviate more from true values than do means of all values.

b) As the number of measurements increases, the probability of an outlier increases.

c) In a large group of measurements, omission of the outlier has little effect on the average although it makes the data look better by decreasing scatter.

One of the purposes of environmental radiation surveillance is to detect (perhaps "unexpected") changes in trends. Judgement is necessary in order to avoid rejecting a result which in fact is an important effect. A practical guide is to let the suspicious result serve as a stimulant to find out what went wrong or what happened.

Statistical Treatment of Random Uncertainties

As described in the preceding section, random uncertainties are those which can be treated by statistical methods, and are derived from an analysis of replicate observations of a random or stochastic process. Numerous excellent sources on statistical theory are available [Cf., Ref 13 and 15] and should be consulted for greater detail or a more rigorous treatment. Only the results and a small amount of background information are presented here.

Before proceeding, some terminology is necessary. The term variate (or random variable) is used to denote the quantity which may take on any of the observed values. The agregate of these observations is termed a sample of some parent population, and may be described by a frequency distribution. This distribution of the population is a specification of the way in which the number of observations (frequencies) are distributed according to the values of the variates. The parameters of a population are the descriptive measures of the distribution. The mean (μ), a measure of the center or location of the distribution, and the standard deviation (σ) a measure of the spread or scatter of the distribution, are examples of parameters. The mean (μ) is also termed the first moment of the distribution, and the square of the standard deviation (σ^2), called the variance, is the second central moment. In the absence of an infinite population, one must make estimates of the parameters from finite populations (the sample of observations). A sample statistic is this estimator of the population parameter. The values of sample statistics are computed entirely from the sample and are the basic measures of the central tendency (location) and dispersion (variation). The mean ($\bar{x}$) and standard deviation (s) are widely known examples of statistics. Unfortunately, the distinction between population parameters and sample statistics is frequently ignored and the two are often confused, and incorrectly spoken of interchangably. The following diagram attempts to clarify the distinction.

STATISTICS		PARAMETERS
Calculated from sample (e.g., $\bar{x}$ and s^2)	used to estimate →	for the population (e.g., μ and σ^2)

In practice, the parameters of the population are denoted by Greek alphabet characters, and the corresponding estimators of these parameters (the statistics) by Roman alphabet characters. Table 3, taken from Ref. [14] lists a number of commonly used parameters and statistics.

The population distribution must be known before one can proceed with the treatment of random uncertainties. A rigorous analysis would require confirmation that the sample of observations is a Normal or some other known distribution. Numerous statistical tests, such as the χ^2-, t- and F-tests, are available for this. Standard statistical sources such as References [11, 13 15, 16, and 19] may be consulted for details. These tests are not always practical, particularly since they are not very applicable with samples of less than about 30 observations. With fewer observations, a Normal (or Gauss) distribution, which is completely characterized by the mean and variance, is assumed. For some other distributions, further parameters, such as skewness (third central moment) or peakedness (fourth central moment), may be necessary [Cf., Ref.[11], p. 94]. The justification for this assumption of normality is based on precedent. The Normal distribution can be viewed as a mathematical result empirically shown to be valid for a large number of different experimental situations. It is still an assumption and it is well worthwhile to make a visual examination of the data for any marked departures from normality. There are some simple procedures to do this. They include construction of a histogram or graphical tests using probability paper [Cf., Ref. [17], p. 6-141 and Ref. [15], p. 55]. The discussion of random uncertainties which follows assumes that a Normal distribution is justifiable. It can be shown that this subsequent treatment is not absolutely dependent on a Normal population distribution. The Central Limit Theorem states this, provided the departures are not too great, and further predicts that the convolution or folding-together of non-Normal distributions tends to form Normal distributions. The probabilities for some typical intervals in the Normal distribution are provided in Table 4. As stated before, an analysis of the observed values will be used to estimate μ and σ^2.

Sample Mean and Standard Deviation. For n measurements of x, the best estimate of the parameter μ is obtained from the mean ($\bar{x}$) of the sample: and the best estimate of σ^2 from the variance (s_x^2), where

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \longrightarrow \mu \qquad (1)$$

$$\text{and } s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 \longrightarrow \sigma_x^2. \qquad (2)$$

TABLE 3

Population Parameters		Sample Statistics (Estimators of Parameters)
μ_x	(mean = first moment)	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
σ_x^2	(variance = second central moment)	$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2$
σ_x	(standard deviation of x about μ_x)	$s_x = \sqrt{s_x^2}$
$\sigma_{\bar{x}}$	(standard error of the mean, or standard deviation of the average)	$s_{\bar{x}} = \frac{1}{\sqrt{n}} s_x$
$\sigma_{xy} = \sigma_{yx}$	(covariance)	$s_{xy} = s_{yx} = \frac{1}{n-1}\sum_{i=1}^{n} (x_i-\bar{x})(y_i-\bar{y})$
$\frac{\sigma}{\mu}(100)$	(coefficient of variation, or relative standard deviation, expressed in per cent)	$v_x = \frac{s_x}{\bar{x}}(100)$

The sample standard deviation is the square root of the variance or the quantity s_x. It refers to the standard deviation computed from a sample of measurements.

TABLE 4

Interval $(\mu-\zeta\sigma_x)$ to $(\mu+\zeta\sigma_x)$	Probability of x having a value within this interval.
ζ	%
0.6745	50.
1.000	68.269
1.960	95.
2.000	95.450
2.576	99.
3.000	99.73

Standard Error of the Mean. Any mean $\bar{x}$ is determined from a finite number of measurements. If the determination is repeated, one can obtain a series of slightly different $\bar{x}$ values. According to the Central Limit Theorem, for large n, the distribution of these $\bar{x}$ values will be close to Normal for any distribution of x. Thus, a standard deviation of this distribution could be obtained from repeat determinations of $\bar{x}$. It may, however, also be estimated from just the measurements used in a single determination of x. This estimate of the precision on the mean is termed the standard error of the mean ($s_{\bar{x}}^2$) which is given by

$$s_{\bar{x}}^2 = \frac{s_x^2}{n} = \frac{1}{n(n-1)}\sum_{i=1}^{n} (x_i - \bar{x})^2. \qquad (3)$$

The quantity $s_{\bar{x}}^2$ is termed the variance of the mean. The standard error of the mean ($s_{\bar{x}}$) must not be confused with the sample standard deviation (s_x). The standard deviation s_x is only dependent on the measurement precision, while $s_{\bar{x}}$ depends on both the precision and the number of observations.

Propagation of Random Uncertainities (Measurements Involving Several Physical Quantities). The uncertainties that have been described up till now are those derived by statistical methods from replicate measurements. A reported value is seldom measured directly, and its associated total random uncertainty is usually not entirely due to random fluctuations made just during that measurement. The reported value may be derived from quantities that were constant during the measurement (e.g., an instrument calibration

factor) or from several measured quantities. These quantities, including contants, have uncertainties in their numerical values due to random causes. Estimates of their uncertainty may have been previously derived from replicate measurements at the time the values were determined. It is necessary to combine (or propagate) these individual contributions into a total random uncertainty term for the reported value.

The statistical theory of errors to combine the separate variances of the quantities involved is usually based on the assumption of Normal distribution functions. This results in the very familiar "propagation of error formula" which is based on just the first order Taylor series expansion of the functional form of x about the mean of x (see Ku [20] for a derivation and greater detail). If the reportd value of x is determined by a combination of independent variables, u, v, w. ... which is given by

$$x = f(u, v, w, \ldots), \tag{4}$$

then the estimated variance on x is

$$s_x^2 = \left(\frac{\partial f}{\partial u}\right)^2 s_u^2 + \left(\frac{\partial f}{\partial v}\right)^2 s_v^2 + \left(\frac{\partial f}{\partial w}\right)^2 s_w^2 + \ldots + \text{terms} \tag{5}$$

covariance

where s_u^2, s_v^2, s_w^2, ... are the estimated variances on u, v, w, respectively. The partial derivatives are evaluated with all other variables fixed at their mean values. For simplicity, let α_i correspond to the entity $(\partial f/\partial i)^2 s_i^2$ for the ith component. Then,

$$s_x^2 = \sum_i \alpha_i \tag{6}$$

This treatment assumes that the terms of higher order than the first partial derivatives are negligible,* and that all of the component quantities are independent and not correlated to each other. This is frequently the case, but not necessarily always true. For details of treating partially correlated components, the reader is referred to Ku[20] or other standard statistical sources. It is always necessary to consider the basic functional form of the component variables before assuming a particular propagation formula. Error propagation formulae for some simple, common functions are provided in Table 5.

Grand Averaging and Weighted Means. The preceding discussion concerned only the calculation and propagation of mean values and their variances from a set of independent measurements of the same quantity. Frequently, one may have estimates of the variance on each individual value in the set. This may arise from estimates of the "counting errors" derived from the total number of counts in single determinations (as will be described shortly), or when computing grand averages of previously averaged values where a variance was calculated for each individual mean. In this case, one may intuitively feel that the averaging process should be performed by giving greater weight to those values that have greater precision. It can be shown (Cf., Ref.[21], p. 44) that for a set of m independent values of quantity x (all taken from the same population), each with mean $\bar{x}_j$ and variance $s_{x_j}^2$ the best estimate of the population mean is obtained with a weighting factor of $1/s_{x_j}^2$ for each x_j. Thus, the overall mean (grand average) is given by

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{m} \bar{x}_j / s_{x_j}}{\sum_{j=1}^{m} 1/s_{x_j}^2} \tag{7}$$

The use of this weighting factor results in a mean that has the least variance. This variance is

$$s_{\bar{x}}^2 = 1 / \sum_{j=1}^{m} 1/s_{x_j}^2 \tag{8}$$

This treatment is not applicable if any of the xj values are correlated to any of the others. Therefore, weighting factors, which are based on the reciprocals of individual uncertainties, must not include any contribution from systematic biases, but only those uncertainties due to independent random causes. One could have, just as well, grand averaged the results without weighting. That is,

$$\bar{\bar{x}} = \frac{1}{m} \sum_{j=1}^{m} \bar{x}_j \quad , \tag{9}$$

with

$$s_{\bar{x}}^2 = \frac{1}{m(m-1)} \sum_{j=1}^{m} (\bar{x}_j - \bar{\bar{x}})^2 \tag{10}$$

*This is based on the assumption that over the range of the variables given by their errors, only small changes in the partial derivatives occur. If the errors are large, the second partial and partial cross derivatives should be included. Cf., Ku[20].

This latter approach results in a mean that does not have the least variance. It can, however, serve as a useful guide. In general, one should always treat with suspicion any result that has a significantly different weighted and unweighted mean.

Statistics of Radioactive Decay (Poisson Distribution). In addition to the usual random uncertainties associated with replicate observations (discussed above), radioactivity measurements are also subject to a random variation arising from the nature of the radioactive decay process itself. The rate of radioactive decay is not a constant with time, but fluctuates randomly about a mean or expectation value. This decay can be described by the Poisson probability distribution function (see Ref. [14] and references therein for a derivation of the Poisson function as a limiting case of the Binomial or Bernoulli distribution).

The Poisson assumption allows one to estimate, from a single measurement, the spread or scatter of a measured number of counts (c) about the mean value η. Assuming that the total number of counts c obtained in the measurement is the best estimate (or is a good estimate of

TABLE 5

Propagation of Error Formulas for Some Simple Functions

Function Form†	Approximate variance	Additional term if u and v are correlated
$x = au \pm bv$	$s_x^2 = a^2 s_u^2 + b^2 s_v^2$	$\pm 2ab s_{uv}$
$x = auv$	$\frac{s_x^2}{x^2} = \frac{s_u^2}{u^2} + \frac{s_v^2}{v^2}$	$2\frac{s_{uv}}{uv}$
$x = \frac{au}{v}$	$\frac{s_x^2}{x^2} = \frac{s_u^2}{u^2} + \frac{s_v^2}{v^2}$	$-2\frac{s_{uv}}{uv}$
$x = u^2$	$s_x^2 = 4u^2 s_u^2$	
$x = \sqrt{u}$	$s_x^2 = 1/4 \frac{s_u^2}{u}$	
$x = au^{\pm b}$	$\frac{s_x^2}{x^2} = b^2 \frac{s_u^2}{u^2}$	
$x = ae^{\pm bu}$	$\frac{s_x^2}{x^2} = b^2 s_u^2$	
$x = cu^{\pm a}v^{\pm b}$	$\frac{s_x^2}{x^2} = a^2 \frac{s_u^2}{u^2} + b^2 \frac{s_v^2}{v^2}$	$2(\pm a)(\pm b)\frac{s_{uv}}{uv}$
$x = \ln u$	$s_x^2 = \frac{s_u^2}{u^2}$	
$x = a \ln(\pm bu)$	$s_x^2 = a^2 \frac{s_u^2}{u^2}$	

†Where a, b and c are constants; u and v are assumed to be statistically independent; and the value of x if finite and real [e.g., $v \neq 0$ for ratios with v as denominator, $u > 0$ for $\sqrt{u}$ and ln u, and $(\pm bu) > 0$ for $\ln(\pm bu)$].

the mean value that would be obtained from a large number of replicate measurements) of the mean value η, i.e.

$$c \longrightarrow \eta ,$$

then the best estimate of the standard deviation in η may be calculated by

$$s_c = \sqrt{c} \longrightarrow \sigma_\eta = \sqrt{\eta} \qquad (11)$$

where the arrows are read as "used to estimate." Since the counting rate (r) measured over a given counting time T is determined by c, the mean rate (r) and its standard deviation are estimated by

$$r = \frac{c}{T} \qquad R = \frac{\eta}{T} , \qquad (12)$$

and $$s_r = \frac{\sqrt{c}}{T} \qquad \sigma_R = \frac{\sqrt{\eta}}{T} \qquad (13)$$

The relative standard deviation in r (s_r/r) obviously decreases with increasing total number of counts. Similarly, an increase in the counting time of a source with a given rate, decreases s_r.

Background Subtraction. For most environmental radiation measurements, the counting rate of a source (r_s) is determined by subtracting a background rate (r_b) from a gross (source plus background) rate. That is,

$$r_s = r_{s+b} - r_b .$$

By the propagation of errors, as described earlier, the standard deviation in r_s is estimated by

$$\left.\begin{aligned} s_{r_s} &= \sqrt{s_{r_{s+b}}^2 + s_{r_b}^2} \\ &= \sqrt{\frac{r_{s+b}}{T_s} + \frac{r_b}{T_b}} \\ &= \sqrt{\frac{c_{s+b}}{T_s^2} + \frac{c_b}{T_b^2}} \end{aligned}\right\} \qquad (14)$$

where T_s and T_b refer to the gross and background counting times, respectively. These and similar equations which can be derived form the basis for the experimental design of radioactivity measurements (e.g., optimizing the division of time between T_s and T_b. See, for instance, Ref. [14] and [21] for greater details on radioactivity measurement procedures and processing of counting data.

Random Uncertainty Components. The standard deviations described above, which are estimated from the total number of counts in a single determination of the counting rate, are usually termed the "counting errors." Although the use of this term will probably continue to prevail, the term is a misnomer. It takes into consideration only the random scatter about the mean from the radioactive decay process itself. To presume that this is the only source of random fluctuation in the overall measurement or counting process is fatuous. Unfortunately, federal guidance, as it pertains to reporting environmental radiation data, does not suggest otherwise. Other sources may be random timing uncertainties, random variations in the sample preparation, positioning of the sample at the detector, etc. The list is nearly endless. Some of these are addressed in Ref. [1].

The only way to realistically assess the overall random uncertainty is to make replicate determinations and calculate the standard deviation (or variance) of the mean by the usual statistical methods (e.g., by Eq. 2). For reported values derived from several component quantities, the individual uncertinty estimates (obtained from replicate measurements for each independent variable) must be propagated (e.g., by Eq. 5) to obtain the overall standard deviation (or variance) of the final value.

It is recognized that replicate measurements of environmental samples are uncommon and may not be feasible because of time and cost constraints. Yet an uncertainty assessment based on some procedure beyond calculating the square root of the total number of counts is clearly needed. A number of procedures, such as the Poisson index of dispersion test, are available to test for the presence of extraneous measurement variations beyond that inherent in the radioactive decay process itself (Cf., Ref. [14]; Ref. [16]; Ref. [22], p. 789; Ref. [23], p. 64). Again, such tests require multiple measurements, at least occasionally, to assess the magnitude of the overall variability from all random causes. Various control charts (Cf. Natrella [13], p. 18-1 and Ref. [24], p. 32) can also be used as a continuous graphical record of this variability. It must be emphasized, however, that these estimates of random uncertainty should not be based exclusively on the information derived from just the present measurements. Presently derived information should be added to the information accumulated in the past on the variability of the measurement process. In this way, more realistic and reliable canonical values of the random uncertainty estimates may be established over time. Ideally, every major step or component of the measurement process should be independently assessed. This would include not only the variability inherent in the particular measurement of concern but also the imprecision arising from corrections, constants, calibration factors and any other measurements that make up the final result.

A practical procedure for making a more complete uncertainty assessment will be discussed later. So far, we have only addressed the calculation of statistics (e.g., the standard deviation) from replicate measurements, and the estimation of the Poisson counting error from a single measurement. In either case, in order to utilize these pieces in an overall uncertainty estimate, we need to consider the assessment of systematic uncertainties.

Assessment of Systematic Uncertainties

The distinction between random and systematic uncertainties was demonstrated earlier. In practice, the systematic uncertainties can be considered to be those sources of inaccuracy which are biased and not subject to random fluctuations, and those which may be due to random causes but can not be or are not assessed by statistical methods. Although a general guideline for the approach to the assessment can be formulated, there are, unfortunately, no rules to objectively assign a magnitude to the systematic uncertainties. For the most part, it is a subjective process. Their magnitudes would preferably be based on experimental verification, but may have to rely on the judgement and experience of the experimenter. The general approach is:

Step 1

Consider and identify all of the conceivable sources of inaccuracy. This requires a careful scrutiny of every detail and aspect of the measurement that can affect the reported value. The "devil's advocate" can be a useful aid in this step.

Step 2

From the above set, extract the blunders and those sources which can and have been (or will be) assessed by statistical methods (i.e., the random uncertainty contributions).

Step 3

Assign a magnitude to the conceivable limit of uncertainty for each of the remaining sources.

There are at least two general reference frames for estimating the magnitude of the systematic uncertainties. One is to consider them as upper bounds, overall or maximum limits[25]. The other is to express them in terms of some probability (e.g., the limits in which the true value is expected to lie in two out of three cases). The latter approach is not recommended since it implies some knowledge of their probability distribution, which is not likely [26].

In general, each systematic uncertainty contribution is considered as a quasi-absolute upper bound, overall or maximum limit on its inaccuracy value. Its magnitude is typically estimated in terms of a semi-range of plus or minus δ about the mean of the measurement result. If the reported value or mean x is determined by a combination of independent variables [as in Eq. (4)]

$$x = f(u, v, w, \ldots .),$$

then the relative contribution to the systematic uncertainty in x due to the estimated systematic uncertainty in one of the variables (δ_u) is

$$\delta = \left(\frac{\partial f}{\partial u}\right)^2 \cdot \delta_u^2 \qquad (15)$$

(assuming that the variables are independent of each other).

By what method then should the magnitude of these maximum limits to the systematic uncertainties be asigned? It may be based on a comparison to a standard or verification with two or more independent and reliable measurement methods. Additionally, it may be

based on judgement;

based on experience;

based on intuition;

based on measurements and data (e.g., by "varying the factor at issue to its extreme range and noting the change on the results"[27]; or comparison to other methods of measurement, other instruments, etc.[14]; or just a pure guess[28] (which is not recommended, but may be better than nothing).

Or it may include combinations of, or all of the above factors.

The fact that a conceivable source of inaccuracy is ascribable to random causes is not a sufficient condition for treating it as a random uncertainty in error reporting. It must be assessable by statistical methods derived from an analysis of repeated measurements. In the absence of these data, the effect of this random cause must be treated as a systematic uncertainty. A program to design and execute experiments (Cf., Ref.[13, 14, and 29] for statistical designs of experiments) to determine the effects of all such factors and considerations requires resources beyond the capacity of most laboratories. Yet a conscientious effort must be made to assess every source of inaccuracy.

The process of assessing the systematic uncertainties is not a simple task. Because of the complexity and its rather subjective nature, it has frequently been ignored. This has resulted in optimistic and underestimated overall uncertainty assessments. It is due as much to overlooking and ignoring possible sources of inaccuracy, as to underestimating those that are known. To aid the experimenter in this process, a list of conceivable sources of inaccuracy and some suggestions to treat them is provided in Ref.[1].

Confidence Limits

The net result from a single counting experiment (x) or the calculated mean from multiple measurements ($\bar{x}$) is used as the best estimate of the population parameter μ_x and hence the true value (in the absence of bias). Because of the associated statistical fluctuations about x or $\bar{x}$, it may not be exactly equal to μ_x. The confidence interval is the range of possible values on either side of x or $\bar{x}$ within which μ_x can be expected to fall. Confidence limits are the numerical values at the limits of this range. The parameter μ_x can be expected to lie within the confidence limits with a given probability. This probability, usually expressed as a percentage, is called the confidence coefficient.* Alternatively, the confidence coefficient can be considered to be the probability that the confidence interval will include the parameter μ_x. These concepts are illustrated below for a mean of $\bar{x}$. When addressing the subject of confidence limits, what is sought is a procedure to calculate the confidence limits for a given confidence coefficient (probability), or to determine the probability for some confidence interval.

Confidence Limits for Systematic Uncertainty. As discussed in the preceding Section, a systematic uncertainty is an estimated upper or maximum bound for each conceivable and assessable contributing systematic source of inaccuracy. That is, there is a high degree of confidence (high likelihood or large probability) that the systematic uncertainty due to this contributing source of inaccuracy would not exceed the numerical value of this stated systematic uncertainty. Therefore, a systematic uncertainty which is estimated in terms of a semi-range may be regarded as a 99% or greater confidence interval. Although the entire subject of systematic uncertainty is characterized by an uneasy nebulousness, the above approach, which is unsupported by statistical or physical principles, does contain a degree of decisiveness. One can hardly argue that a better estimate than that provided by the conscientious experimenter is obtainable.

Confidence Limits for Random Uncertainty (Normal Distribution). Fortunately, one can rely on statistical theory when considering the confidence limits for random uncertainties. A common practice is to express the confidence limits in terms of

CONFIDENCE LIMITS

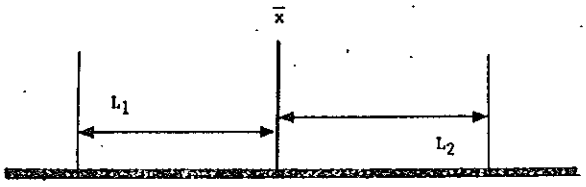
are the values $(\bar{x} - L_1)$ and $(\bar{x} + L_2)$

CONFIDENCE INTERVAL

is the set of possible values between $(\bar{x} - L_1)$ and $(\bar{x} + L_2)$

CONFIDENCE COEFFICIENT* (P)

is the probability (in %) that μ_x will lie within the confidence interval or between the confidence limits



measured variable x

*The confidence coefficient is often referred to as the "confidence level". Although the latter usage is probably more familiar to most readers, it is in conflict and often confused with the term "significance level" which is employed in all statistical literature. The confidence or significance (α) associated with the specification of a confidence interval is the probability that the interval will not include the true value. The confidence coefficient ($1-\alpha$), which is often expressed as a percentage P, is the term used to describe the probability that a confidence interval will include the true value. For example, a 95% confidence interval is one for which the confidence coefficient (P), not the significance level (α), is 95%. In this case, the significance level is 5%. To avoid this confusion, the term "confidence level" is deliberately avoided and, therefore, not employed.

some multiple of the standard deviation or standard error of the mean. For results from multiple measurements, the confidence limits for the mean μ may be expressed in terms of some integral multiple of the calculated standard deviation s_x, i.e.,

$$\bar{x} \pm s_x \ ,$$
$$\bar{x} \pm 2s_x \ ,$$
$$\bar{x} \pm 3s_x \ , \quad \text{etc.}$$

If the estimated statistic s_x was derived from a large number of measurements n, and if the distribution of x values is Normal, then these are 68.3%, 95.5% and 99.7% confidence intervals, respectively. It must be emphasized that these confidence coefficients apply only when n is very large and when the distribution is Normal. As the number of measurements n decreases, the confidence coefficient or probability that x lies within a given confidence interval gradually diminishes. Table 6, taken from [25], demonstrates the effect of various values of n on the confidence coefficients. In order to determine the magnitude of a confidence interval ($L_1 + L_2$) at a given confidence coefficient (P), we need to first consider the number of measurements or what is termed the number of degrees of freedom.

Degrees of Freedom. For purposes of this guide, the number of degrees of freedom (ν) for a calculated statistic (e.g., s_x) is the number of measurements (n) in excess of those needed to determine the estimate of the population parameter of interest. For example, previously we noted that the precision on the mean ($\bar{x}$) may be estimated in terms of the standard error of the mean ($s_{\bar{x}}$). Recalling the definition of the variance, given by Eq. (2), the number of degrees of freedom in $s_{\bar{x}}$ is the number of independent terms in the residual sum of the squares, $\sum_{i=1}^{n} (x_i - \bar{x})^2$. Thus, in calculating the mean of n observations, the number of degrees of freedom (ν) in $s_{\bar{x}}$ is (n-1).

If $\bar{x}$ is derived from a combination of more than one variable [see Eq. (5)], the number of degrees of freeedom does not have an exact meaning. An approximation, derived by Welch[30], can be used to obtain an effective number of degrees of freedom (ν_{eff}) in $s_{\bar{x}}^2$. This is given by

$$\frac{1}{\nu_{eff}} = \frac{\sum_i a_i^2/\nu_i}{\sum_i a_i^2} \qquad (16)$$

where ν_i is the number of degrees of freedom in $s_{\bar{x}}$ for the ith component of a_i is defined by Eq. (6). The value of ν_{eff} calculated in this way is sufficiently accurate in most applications. When ν_i for the individual components becomes small, the approximation becomes less valid[31]. Examples illustrating this type of calculation for n measurements of a sample and k measurements of a standard are provided in NCRP Report 12[32].

Confidence Limits for Random Uncertainty (t-Distribution). With a value of ν (or ν_{eff}) and $s_{\bar{x}}$ for a mean x, one can determine the confidence interval for any given confidene coefficient P. The confidence interval about the mean $\bar{x}$ is symmetrical for a Normal distribution, and the confidence limits may be calculated by

$$L_1 = L_2 = t_\nu(P) s_{\bar{x}} \qquad (17)$$

where t is the Student t- value for a confidence coefficient P and ν degrees of freedom. Table 7 contains values of the t-statistic as a function of ν for various values of the confidence coefficient P. When more than one variable is involved, the value of ν_{eff} [from Eq. (16)]

TABLE 6

Confidence coefficients P for the s_x, $2s_{\bar{x}}$ and $3s_x$ confidence intervals for the Normal distribution with various values of n[16].

CONFIDENCE INTERVAL	CONFIDENCE COEFFICIENT (P) IN %			
	n=5	n=10	n=20	n=∞
$\bar{x} \pm s_x$	62.6	65.7	67.0	68.269
$\bar{x} \pm 2s_x$	88.4	92.3	94.0	95.450
$\bar{x} \pm 3s_x$	96.0	98.5	99.3	99.730
$\bar{x} \pm 4s_x$				99.994

is usually fractional. The value of t may then be obtained from the table by interpolation, or ν_{eff} may be rounded down to the next integral value. Thus, the random uncertainty could be expressed, in terms of the calculated or propagated s_x and ν, as the confidence limits ($\pm ts_x$) for a given confidence coefficient P. As before, this process of going to higher confidence intervals requires an assumption about the form of the distribution, and is dependent on the number of measurements. Table 7 illustrates that the common practice of equating $3s_x$ with the 99% confidence interval is erroneous for ν less than about 9 or 10.

Confidence Limits for Random Uncertainty (Poisson Distribution). An analogous situation exists when considering the confidence limits for the Poisson counting error from single determinations. In this case, one assumes the counting process can be described by a Poisson distribution, and the dispersion statistic (e.g., the standard deviation) is estimated from the number of counts obtained in the measurement [such as by Eqs. (11), (13), and (14)]. As described before, the Poisson distribution approximates the Normal for very large counts (>100). Then, the confidence limits for the number of counts c could be given by

$$L_1 = L_2 = t_\infty(P)\sqrt{c} \qquad (18)$$

where t_∞ is the Student-t value for infinite degrees of freedom corresponding to a given confidence coefficient for the Normal distribution. Such as, t is 1.96 and 2.58 for the 95% and 99% confidence coefficients, respectively (refer to Tables 4 and 7). This estimate will, of course, be more correct the larger the number of counts. For small numbers of counts, the Poisson distribution no longer approximates the Normal very well; the confidence limits become asymmetrical and the approximation by Eq. (18) poorly estimates the confidence interval. For example, the lower and upper confidence limits for a low number of counts c may be given by $(c-L_1)$ and $(c+L_2)$, respectively where

TABLE 7

Student t-Values Corresponding to a Given Confidence Coefficient P for Use with an Estimated Standard Deviation for a Normal Distribution Based on ν Degrees of Freedom.

P= / ν	68.269%	95.000%	95.450	99.000	99.730%
1	1.837	12.706	13.968	63.657	235.80
2	1.321	4.303	4.527	9.925	19.207
3	1.197	3.182	3.307	5.841	9.219
4	1.142	2.776	2.869	4.604	6.620
5	1.111	2.571	2.649	4.032	5.507
6	1.091	2.447	2.517	3.707	4.904
7	1.077	2.365	2.429	3.499	4.530
8	1.067	2.306	2.366	3.355	4.277
9	1.059	2.262	2.320	3.250	4.094
10	1.053	2.228	2.284	3.169	3.957
11	1.048	2.201	2.255	3.106	3.850
12	1.043	2.179	2.231	3.055	3.764
13	1.040	2.160	2.212	3.012	3.694
14	1.037	2.145	2.195	2.977	3.636
15	1.034	2.131	2.181	2.947	3.586
16	1.032	2.120	2.169	2.921	3.544
17	1.030	2.110	2.158	2.898	3.507
18	1.029	2.101	2.149	2.878	3.475
19	1.027	2.093	2.140	2.861	3.447
20	1.026	2.086	2.133	2.845	3.422
25	1.020	2.060	2.105	2.787	3.330
30	1.017	2.042	2.087	2.750	3.270
40	1.013	2.021	2.064	2.704	3.199
60	1.008	2.000	2.043	2.660	3.130
120	1.005	1.980	2.023	2.617	3.069
∞	1.000	1.960	2.000	2.576	3.000

Adapted from Brian L. Joiner, "Student t-Deviate Corresponding to a Given Normal Deviate," J. Res. NBS 73C, 15 (1969); and CRC Handbook of Mathematical Tables, First Ed. Chemical Rubber Publishing Co., Cleveland, Ohio (1962), p. 271.

$$L_1 = \tau_1\sqrt{c}$$

and $$L_2 = \tau_2\sqrt{c} \qquad (19)$$

Values of τ as a function of the number of counts for confidence coefficients of P=68%, P=95%, and P=99% are illustrated in Figure 2 and tabulated in Table 8. As indicated, for low counts, (under 100 or so), the Normal approximation of Eq. (18) can give a result which is in error. As a result, special tables [33] are needed to obtain appropriate confidence limits for small numbers of counts.

From Eq. (15) it follows that, for a counting rate r_s with standard deviation s_{r_s}, the confidence limits for a higher confidence coefficient P would be given by $r_s \pm L$ where

$$L = \sqrt{\tau_{c_{s+b}}(P)s_{r_{s+b}}^2 + \tau_{c_b}(P)s_{r_b}^2}$$

$$= \sqrt{\tau_{c_{s+b}}(P)\frac{c_{s+b}}{T_s^2} + \tau_{c_b}(P)\frac{c_b}{T_b^2}} \qquad (20)$$

Or, alternatively, the τ values in Eq. (20) could be replaced with values of t_∞ when the Normal approximation is valid (i.e., very large counts).

In summary, any consideration of confidence limits for single determination or multiple measurement situations is, first, distribution dependent requiring an assumption about the underlying population distribution; and second, is size dependent requiring knowledge of the number of measurements, degrees of freedom or magnitude of the counts, etc.

Reporting the Overall Uncertainty

If the reported uncertainty is to be a credible assessment of the likely accuracy, it must be based on as nearly a complete an assessment as possible and must consider all conceivable sources of inaccuracy.

Categorization of the Random and Systematic Uncertainty Components. A clear distinction between random and systematic uncertainties is often difficult and troublesome. In part, this is because many experimental processes embody both systematic and stochastic (random) elements.

In general, the random uncertainty contributions can be considered to be those sources of inaccuracy which can be and are assessed and propagated by statistical methods. Estimates of population parameters, or statistics, are computed entirely from the measurements data.

The systematic uncertainty components can be considered to be the conceivable sources of inaccuracy which are biased and arise from non-stochastic systematic effects, as well as those which may be due to random causes but can not be or are not assessed by statistical methods. Hence, it is meaningful to speak of a random uncertainty contribution only if one has a computed statistic for the magnitude of the random variation. Further, this does not imply that every conceivable source of inaccuracy (say, the chemical yield) lies in either just the random or systematic uncertainty category. One might obtain an estimate of the random uncertainty contribution by calculating a standard deviation from the results of multiple determinations of the chemical yield, and

TABLE 8

Values of τ Corresponding to a Given Confidence Coefficient P for use with an Estimated Standard Deviation for a Poisson Distribution.

COUNTS	P=68%		P=95% *		P=99%*	
	τ_1	τ_2	τ_1	τ_2	τ_1	τ_2
1	0.83	2.29	0.975	4.57	0.995	6.43
4	0.95	1.60	1.46	3.12	1.66	4.30
9	0.97	1.37	1.63	2.69	1.96	3.67
16	0.98	1.28	1.71	2.50	2.11	3.37
25	0.98	1.22	1.76	2.38	2.20	3.20
50	0.99	1.15	1.82	2.25	2.31	3.01
∞	1.00		1.96		2.58	

* Adapted from E.S. Pearson and H. O. Hartley, Biometrika Tables for Statisticians, Vol. 1, Cambridge Univ. Press (1954), Table 40.

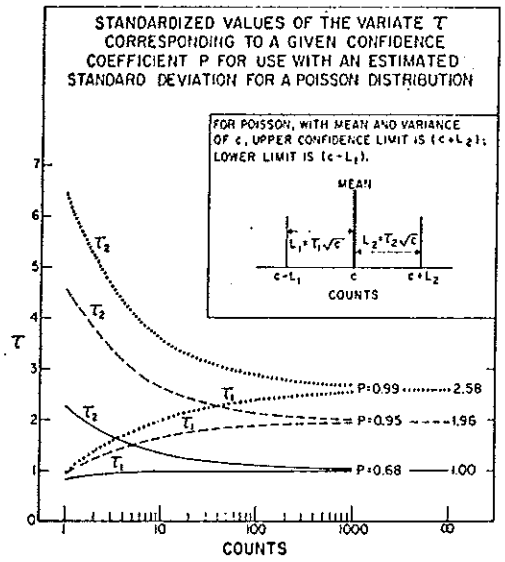


Figure 2.

still have a systematic uncertainty in the chemical yield due to a bias in all of the mass measurements made in the determination of the yields. The diagram below attempts to illustrate the concept of making a complete assessment of all of the conceivable sources of inaccuracy, and their division between random and systematic components.

Total Random Uncertainty. From preceding sections, we can obtain a combined total random uncertainty contribution, stated in terms of some dispersion statistic, such as the standard deviation s_x for a measurement result x. The value of s_x may be evaluated from the results of multiple measurements using, for example, Eq. (3). The Poisson counting error contribution may be estimated from Eq. (13), (14) or comparable equations. Except for this contribution which may be evaluated from single measurements (due to the Poisson assumption), all other contributions must be evaluated by calculating statistics from the results of replicate determinations. As noted, control charts may aid this process. The total random uncertainty, s_x, would include not only the Poisson counting error and random uncertainty derived from the particular measurement under concern, but also the random components from corrections, constants, calibration factors and any other measurements that also make up the final result x. These contributions may be combined by propagation of error formulae, such as Eq. (5).

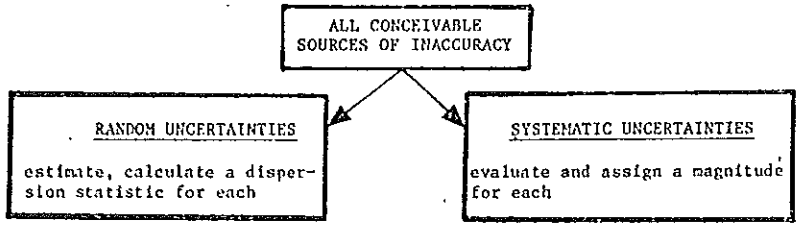
Conceptually, one can propagate and express the total random uncertainty in terms of some confidence limits at any chosen confidence coefficient. This requires a number of assumptions, such as to the underlying population distributions, and can very rapidly become exceedingly complicated requiring computations based on the number of measurements involved in each random uncertainty component, the number of degrees of freedom, the magnitude of the counts in the counting error contribution, etc. One is therefore confronted with the choice between making simplifying assumptions (which ultimately may provide results that are misleading), or performing complex calculations (whose efficacy, in terms of the time and cost required can be seriously questioned).

Systematic Uncertainty Components. Lastly, we can obtain a list of each conceivable source of systematic uncertainty expressed as a confidence limit ($\pm\delta_j$ which can be considered to correspond to some 99% or greater confidence interval.

Approaches to an Overall Uncertainty Statement. One must now consider 1) to what confidence coefficient should the total random uncertainty s_x be reported; 2) how should the individual systematic uncertainty components (δ_j) be combined; and 3) how should the random and systematic uncertainties be combined to form an overall uncertainty. All three questions are interrelated. Underlying this entire discussion is the fundamental consideration that the way uncertainties are reported is dependent on their intended or ultimate use.

Methods of Combining the Random and Systematic Uncertainty Components. Purists would argue that it is incorrect from a theoretical point to view to combine random and systematic uncertainties in any fashion [34]. Their approach is to separately state the random and systematic uncertainty components. For the random contribution, this is to include the confidence limits, the confidence coefficient, and the number of degrees of freedom. For the systematic uncertainty contribution, each component is to be listed, and the method of combination clearly stated. This approach requires a detailed uncertainty statement usually at least a paragraph in length. For additional discussion and references see Eisenhart [35] Ku[36], Natrella [13], and Campion, et al.[37].

This purist approach does not seem practical. First, it is not amenable to large compilations



(e.g., the quarterly EPA Environmental Radiation Data reports) of data from numerous sources. And second, it shifts the burden of evaluating the uncertainties to users. "Users now need, and will do so for many years to come, a single error value resulting from the combination of all the uncertainties. Indeed, a compromise leading to a single error value seems to be better than a perfect solution which in reality shifts the problem of combination of errors to the users" [34].

Several arguments can be advanced for favoring the standard deviation as a measure of the random uncertainty over higher confidence limits. It facilitates subsequent data handling since all weighting and uncertainty propagation is performed with the variances. A stronger argument is that the variance is the only statistic which is derived directly and unambiguously from the measurements. It is unbiased and distribution-free. Converting the measured standard deviation to some other confidence interval requires an assumption that the distribution is known. This equally applies to the use of the Student-t value. "Since an arbitrary multiple of a standard deviation(s) does not contain more information than s itself, the simplest and most reasonable choice of indicating the (random) uncertainty of an experimental result seems to me to go on stating its standard deviation, eventually completed by the number of measurements it is based on"[38].

The arguments for retaining the use of the standard deviation do not address the question of combining it with systematic uncertainties. If one agrees that the systematic uncertainties are to be expressed as estimated maximum limits (the semi-range), then some higher confidence interval for the random uncertainty part (e.g., 99.7% for $3s_x$ for a Normal distribution) provides an acceptable or reasonable level for direct combination with the systematic uncertainty. One can also invoke an argument for which confidence coefficient will be of greatest value to the user of the data. Returning to the original question posed in the example of Section III for a reported value of 15 ± 3 Bq (How certain is the user that the value is between 12 and 18 Bq?"), a high confidence coefficient may be the only unambiguous and unequivocal choice. How sophisticated is the user in ascertaining the meaning and correctly interpreting a 67% confidence interval? If the reported uncertainty is to be a statement of the "likely limits to the accuracy," then a high confidence interval may be the best choice.

Several acceptable models for propagating the systematic uncertainty components and combining them with the total random uncertainty may be considered. The two most commonly employed methods are:

A. Add linearly all components of the systematic uncertainty and add it to the random part. Since the systematic uncertainties (δ_j) are considered to be 99% or greater confidence intervals, it logically should be added to the total random uncertainty at a similar confidence coefficient. Using this model, the overall uncertainty for a reported value x is

$$t_\nu(P=99^+\%)s_x + \sum_j \delta_j \qquad (21)$$

where s_x is an overall standard deviation based on ν degrees of freedom and t_ν ($P=99^+\%$) is the Student-t value corresponding to a 99% or greater confidence coefficient. This is the approach favored by metrology and standardizing laboratories. It probably overestimates the total uncertainty, but can be considered as an estimate of the maximum possible limit. For example, if you estimated that five contributions of about equal magnitude made up the systematic error, you would have to be very unlucky if all five were minus. Yet, if there was one dominant contributor, it might be a very valid approximation.

B. Add in quadrature all of the systematic uncertainty components, and either add it linearly to s_x,

$$s_x + \sqrt{\sum_j \delta_j^2} \quad , \qquad (22a)$$

or add it in quadrature to s_x,

$$\sqrt{s_x^2 + \sum_j \delta_j^2} \qquad (22b)$$

These are frequently considered (erroneously) to be overall 68% confidence intervals. They are probably the most widely employed approaches (if an attempt was made to assess systematic uncertainties at all).

Actually, without making assumptions about the distributions of the systematic uncertainties it is impossible to proceed except to arbitrarily select some method, such as those above. One of the simplest set of assumptions that can be made is that 1) the component systematic uncertainties are all independent, and 2) they are distributed such that all values within the estimated limits are equally likely (rectangular distribution). This approach, often termed the PTB approach[39], is gaining in popularity.

With the above two assumptions, the rectangular systematic uncertainty distributions can be folded together to obtain a combined probability distribution for which the variance may be computed. This may then be combined in quadrature with that for the random uncertainty. In effect, the assumed Normal distribution of the random uncertainty is convoluted with the combined systematic uncertainty distribution to obtain an overall distribution. With this, the overall uncertainty limits at a given confidence coefficient can be evaluated. For the case of a number of comparably-sized com-

ponent systematic uncertainties, a very reasonable estimate of the confidence interval $\pm u_x$ for any confidence coefficient P is

$$u_x = \sqrt{[t_\nu(P)s_x]^2 + \frac{\zeta_N(P)}{3}\sum_j \delta_j^2} \quad (23)$$

where $t_\nu(P)$ and $\zeta_N(P)$ are the Student-t factor and the value of the variate in a Normal distribution, respectively, for the same confidence coefficient P (see Table 9).* For the special case when one of the δ_j's is much larger than the others, then

$$u_x = \sqrt{[t_\nu(P)s_x]^2 + \zeta_R(P)\sum_j \delta_j^2} \quad (24)$$

is a better approximation, where $\zeta_R(P)$ is the value of the variate in a rectangular distribution for a confidence coefficient P. For additional details see Wagner [39] and Williams, et al.[40]

Recommended Statement of Overall Uncertainty and Final Result. There is no one clearly superior method out of the range of approaches to reporting an overall uncertainty. Yet, the diversity of current practices and methods either recommended or required by governmental bodies, or used by those who report data, make it difficult if not impossible to combine or compare the data in a manner that would be useful for determining trends or possible necessary actions. There is, therefore, a great need for uniformity in the method used to report the uncertainties of environmental radiation data. In hope of achieving greater uniformity, the PTB approach, which was outlined above, is recommended. The detailed recommendations for the assessment and propagation of uncertainties which follow are given in this spirit. The recommended method requires that each reported measurement result include the value, the total random uncertainty expressed as the standard deviation, and a combined overall uncertainty. Examples illustrating the method are given in Ref. [1]. This recommended approach may not be best for all purposes. It is intended to be useful and practical without imposing unnecessary burdens on the time, money and personnel resources of laboratories. Adoption of this method may eventually demonstrate its most serious shortcomings and ultimately lead to better methods.

TABLE 9

P	$t_{\nu=\infty}$	ζ_N	ζ_R
50%	0.6745	0.6745	1
68.3%	1	1	1
95%	1.960	1.960	1
95.4%	2	2	1
99%	2.576	2.576	1
99.7%	3	3	1

Recommendations

III.1 Every reported measurement result (x) should include an estimate of its overall uncertainty (u_x) which is based on as nearly a complete an assessment as possible.

III.2 The uncertainty assessment should include every conceivable or likely source of inaccuracy in the result.

III.3 Every conceivable source of inaccuracy should be classified into one of two categories depending on how the uncertainty is estimated.

The random uncertainty contributions are estimated by a statistical analysis of replicate measurements. The random counting error contribution may be estimated from single measurements by making the Poisson assumption.

The systematic uncertainty contributions are estimated by less rigorous methods as described in the text.

Until combined for the measurement result, the individual uncertainties in both categories should be retained separately.

III.4 Each random uncertainty component should be estimated in terms of its standard deviation (s_i), or the standard error of the mean ($s_{\bar{x}}$) for independent multiple measurements.

*S. R. Wagner [PTB-Mitteilungen 89, 83-89 (1979)] recently pointed out that the quadratic sum of confidence intervals [as in Eqs. (23) and (24)] is without theoretical basis. The overall uncertainty u_x for any confidence coefficient P is more properly obtained by

$$u_x = k_p\sqrt{s_x^2 + \frac{1}{3}\sum_j \delta_j^2}$$

where k_p is the standardized variate corresponding to P for the combined random and systematic uncertainty probability distribution.

III.5 The total random uncertainty (s_x) should be obtained by propagating the individual variances (s_i^2) for each random uncertainty component. This would include not only the Poisson counting error and random uncertainty derived from the particular measurement under concern, but also the random components from corrections, constants, calibration factors and any other measurements that make up the final result x. Many of these sources of inaccuracy are tabulated and discussed in Ref. [1]. These contributions may be combined by summing in quadrature or by the use of propagation of error formulae, such as Eq. (5) or those in Table 5. This overall random uncertainty may be represented by

$$s_x^2 = \sum_{i=1}^{p} a_i \qquad (25)$$

which was given as Eq. (6) for p independent random uncertainty components.

III.6 Systematic uncertainties are considered to be independent and each is expected to be uniformly distributed over its range ($\pm\delta_j$). Each systematic component (δ_j) should be estimated in terms of the semi-range about the measurement result for the contributing source of inaccuracy [as in Eq. (15)].

III.7 The random and systematic uncertainties should be combined to form an overall uncertainty on the result x by

$$u_x = \sqrt{s_x^2 + \frac{1}{3}\sum_{j=1}^{q} \delta_j^2} \qquad (26)$$

where s_x is the standard deviation corresponding to the overall random uncertainty and δ_j is the magnitude of the estimated systematic uncertainty for each of the q systematic uncertainty components.

Alternatively, the overall uncertainty on the mean for averaged results should be given by

$$u_x = \sqrt{s_{\bar{x}}^2 + \frac{1}{3}\sum_{j=1}^{q} \delta_j^2} \qquad (27)$$

where $s_{\bar{x}}$ is the standard error of the mean [Eq. (3)]. Similarly, the overall uncertainty derived from p random uncertainty components and q systematic uncertainty components should be given by

$$u_x = \sqrt{\sum_{i=1}^{p} a_i + \frac{1}{3}\sum_{j=1}^{q} \delta_j^2} \qquad (28)$$

For the special case if one of the systematic components (δ_1) is greater than one-third the sum of all the components, i.e.,

$$\delta 1 > 1/3 \sum_{j=1}^{q} \delta_j^2 ,$$

then the overall uncertainty should be given by

$$u_x = \sqrt{s_x^2 + \sum_{j=1}^{q} \delta_j^2} \qquad (29)$$

(or comparable equations analogous to Eqs. (27) and (28) which do not include the factor of 1/3.

III.8 Every reported measurement result x (or mean $\bar{x}$ for averages of results) should include a three-part requirement for the reporting of uncertainties:

(a) the value of the result x (or mean $\bar{x}$);

(b) the propagated total random uncertainty expressed as the standard deviation s_x (or standard error of the mean $s_{\bar{x}}$ for results based on multiple determinations of x or for averaged results of x); and

(c) the combined overall uncertainty u_x (or $u_{\bar{x}}$) as in Eqs. (26), (27), (28), or (29).

When short-hand expressions are necessary, the measurement result should be reported in the format

$$x \pm s_x;\ u_x.$$

For example, a result of x=123 $Bq \cdot L^{-1}$ with s_x=15 $Bq \cdot L^{-1}$ and u_x=18 $Bq \cdot L^{-1}$ should be reported in exponential notation as

$$(1.23 \pm 0.15;\ 0.18)\ E{+}02\ Bq \cdot L^{-1}$$

III.9 Confidence intervals based on the overall uncertainty should not suggest a particular confidence coefficient. Rather, the two reported uncertainties

may be referenced as:

"The total random uncertainty is the propagated standard deviation (or standard error of the mean) of all sources of random uncertainty."

"The overall uncertainty was propagated by adding in quadrature the total random uncertainty and one-third of the estimated upper limits of all conceivable sources of systematic uncertainty."

Confidence limits for a measurement result at a particular confidence coefficient cannot be obtaind by merely multiplying the uncertainty by arbitrary constants and adding and subtracting this value to and from the result. Determination of limits for higher confidence coefficients requires knowledge of the underlying population distribution, and knowledge of the number of measurements, degrees of freedom, or magnitude of the counts.

IV. DETECTION LIMITS

A myriad of vastly different expressions and definitions of "detection limits" are frequently encountered. Their meanings are often ambiguous, inconsistent and incorrectly interpreted. Some of these include "detection sensitivity," "minimum detectable activity (or level)," "lower limit of detection," and "background equivalent activity." Currie[14] commented on a number of such terms and addressed some of the problems and inconsistencies.

Pragmatically, a detection limit is useful as a criterion for experiment design, comparison and optimization purposes, such as in selecting among alternative measurement procedures. Additionally, detection limits may serve as guides which are set by regulatory bodies for establishing minimum acceptable detection capabilities for a given type of analysis. For the purposes of this report, the intent of detection limit calculations is to satisfy both of the uses. It must be emphasized that any calculation of a detection limit is at best only an estimate. Their use is "limited to that of serving as guideposts only, and not as absolute levels of activity that can or cannot be detected by a counting system."[14]

Much of the existing confusion with detection limits for environmental ionizing radiation measurements arises from not only the large number of different expressions that are in use, but also from incorrect interpretations and misapplications of some of the original definitions. In order to satisfy both of the above-mentioned purposes, two distinctly different concepts are required.

The first is an estimated detection limit that is related to the characteristics of the counting instrument. It is not dependent on other factors involved in the measurement method or on the sample characteristics. It is a lower limit in the true sense of the word. Because of its current wide usage, the recommended term is the ESTIMATED LOWER LIMIT OF DETECTION (LLD).

The second concept is that most useful for regulatory purposes. It corresponds to a level of activity concentration that is practically achievable with a given instrument, method and type of sample. It depends not only on the instrument characteristics, but also on many other specific factors involved in the measurement process, as well as the characteristics of the sample being measured. As such, it is not a limit at all, but only an estimated level achievable under a given set of practical conditions. The recommended expression for this concept is the ESTIMATED MINIMUM DETECTABLE CONCENTRATION (MDC).

It is recommended that only these two expressions be employed for all environmental radiation measurement detection limits. Both the LLD and MDC concepts can be based on a uniform consistent methodology. The use of this methodology with only these two expressions should help alleviate and avoid much of the existing ambiguity and misapplications. Both the LLD and MDC concepts will be considered in turn. If the word "estimated" is emphasized and continually used in reporting a LLD and MDC, then their limited nature should be more apparent, and hopefully avoid the implication of an absolute significance in their numerical values.

The Estimated Lower Limit of Detection (LLD)

THE LLD may be defined on the basis of statistical hypothesis testing for the presence of activity. This approach is common to both that of Pasternack and coauthors[42,43], and Currie[41]. Procedures for calculating a LLD based on this approach have also been described in the EML Procedures Manual[44], the EPA quality control program report[24], and the NCRP Handbook of Radioactivity Measurements Procedures[14].

Pasternack and Harley[43] defined the LLD as "the smallest amount of sample activity that will yield a net count for which there is a confidence at a predetermined level that activity is present."[44] In theory, this approach for calculating a LLD requires the number of counts to be sufficient for the Poisson distribution to approach the Normal distribution so that Gaussian statistics can be applied. It has been noted, however, that in practice "the approximation is good down to a few total counts."[44] The LLD is an "a priori ESTIMATE of the detection capabilities of a given measurement process."[41] It is based on the premise that from a knowledge of the background count and measurement system parameters (i.e., detection efficiency), an a priori (before the fact) limit can be established for a particular measurement. This limit does not depend on the sample activity, but rather on the detection capability of the measurement process itself. It is important in the application of the LLD to make the distinction between

it and other limits that are directly applicable to the net sample activity. The latter are applied as a posteriori limits (after the fact) and can be determined only after the sample has been counted.

The LLD is derived from the approximation

$$LLD = K\,(k_\alpha s_0 + k_\beta s_D) \qquad (30)$$

where

- K is the proportionality constant relating the detector response (counts) to the activity, such as, $K=1/\epsilon$ where ϵ is an overall detection efficiency, or $K=1/I_\gamma \epsilon_\gamma$ where I_γ is the gamma ray-emission probability per decay and ϵ_γ the detection efficiency for the gamma ray.
- k_α and k_β are the upper percentiles of the standardized normal variate corresponding to the preselected risk for concluding falsely that activity is present (α) and the predetermined degree of confidence for detecting its presence ($1-\beta$)
- s_0 is the estimated standard deviation of the net sample count (N_n) when the limiting mean of N_n equals zero
- s_D is the estimated standard deviation of the net sample count (N_n) when the limiting mean of N_n equals the LLD.

The basis for this approximation for the LLD is illustrated in Figure 3. Additional details and discussion may be obtained in References [41], [42] and [43]. In statistical hypothesis testing, α and β are the probabilities for what are frequently referred to as Type I (false detection) and Type II (false non-detection) errors, respectively. Values for k_α and k_β corresponding to the risks for false detection (α) and non-detection (β) can be found in most statistical texts. As stated before, this assumes that the random uncertainties are normally distributed. In general, both α and β should be reasonably small in order to provide a high degree of confidence that neither type of error will be made. If $\alpha=\beta=0.5$ (i.e., a 50% risk for each type of error), then the LLD would always be zero. In this case, activity detection near the LLD will be wrong 50% of the time, and "the experiment could be performed equally well by the flipping of a coin."[41] Conversely, if α and β are set very small (say $\alpha=\beta=0.001$), then one would rarely be incorrect. At the same time, however, one would seldom attribute significance to anything but very large activity measurements. A convenient compromise and most common practice is to set both risks equal, and to accept a 5% chance of incorrectly detecting activity when it is absent ($\alpha=0.05$) and a 95% confidence that activity will be detected when it is present ($1-\beta=0.95$). Then

$$k = k_\alpha = k_\beta = 1.645 \quad (\alpha = \beta = 0.05), \qquad (31)$$

which is recommended for use in all environmental radiation LLDs. It is incorrect to refer to this as a 5% confidence level. First, the LLD cannot be characterized by a single confidence level; and second, its use can lead to confusion with the confidence level for an a posteriori decision on the presence of activity after the measurement is made. Preferred language is THE ESTIMATED LLD FOR 5% RISKS OF FALSE DETECTION AND FALSE NON-DETECTION.

Using the recommended approximation of Eq. (30) and the convention of Eq. (31), several typical and specific applications can be developed.

Consider measurement processes in which the net activity is derived by subtracting a background from a gross activity measurement. The standard deviation of the net activity is

$$s_n = \sqrt{s_g^2 + s_b^2} \qquad (32)$$

where s_g and s_b are the standard deviations of the gross activity and background, respectively. If the gross activity and background counts are nearly equal (which is a reasonable approximation near the LLD), then Eq. (32) reduces to

$$s_n = \sqrt{2}\; s_b \qquad (33)$$

Further, if one assumes that s_n over the small range of net activity from zero to the LLD is

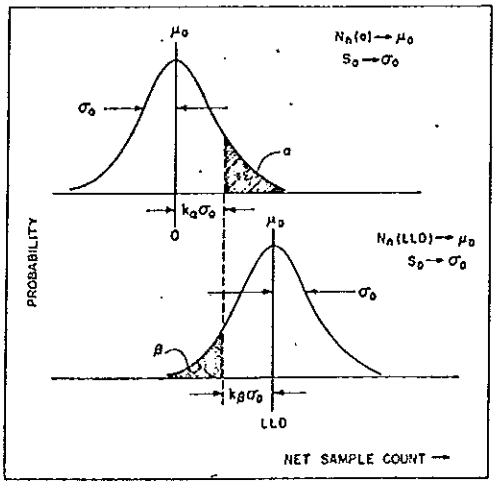


Figure 3. Probability distributions of the net sample count (N_n) at zero and at the LLD illustrating the relations $k_\alpha s_0$ and $k_\beta s_D$.

approximately constant, i.e., $s_n = s_0 = s_b$, then from Eqs. (30) and (33)†

$$LLD = Kk2\sqrt{2}\, s_b = 4.65\, K\, s_b \quad . \tag{34}$$

Alternatively, if one does not make the above assumption ($s_0 \neq s_b$), Currie [41] has shown that the expression for the LLD becomes

$$LLD = K\,(k^2 + 2\sqrt{2}\, ks_b) + K\,(2.71 + 4.65\, s_b) \quad . \tag{35}$$

Except at extremely small counts, the difference in the two calculations is trivial (see Table 10). For a Poisson distribution, s_n is not as independent of the activity level and Eq. (35) may be a better approximation. In most cases, both calculations will be comparable. When they are not (e.g., extremely small background counts), Eq. (35) is recommended.

In many applications, the measurement procedure would frequently consist of a paired sequential observation of the gross and background counts. One could envisage obtaining a single background value which is not exactly equal to the known long-term average background. This single measurement would be used as a check, and perhaps to continually update the average. To calculate a new LLD from a single observation is without merit. It would ultimately lead to a LLD for each and every background measurement. This approach would defeat the spirit of the LLD concept which is to provide an A PRIORI ESTIMATE or a guidepost of the detection capability of the instrument. Similarly, the quantities contained within the proportionality constant K, such as the detection efficiency, should be average values for the instrument. If the values vary substantially from measurement to measurement, or if they abruptly change, then this would suggest that there is a serious instrument or procedural problem. This erratic behavior may be due to an instrument malfunction, or may be an indication of poor laboratory practice. Under any circumstance, it demonstrates the existence of a problem, and should serve as a stimulant to ascertain the cause of the problem (see "blunders," in Part III). If the measurement process or procedure is modified, or if substantive changes in the background or in any of the parameters comprising the proportionality constant K occur, then the LLD should be recalculated.

TABLE 10

TOTAL BACKGROUND COUNTS	s_b	ESTIMATED LLD (K units) Eq. (34)	Eq. (35)
0	0	0	2.7
1	1.	4.7	7.4
4	2.	9.3	12.
10	3.2	15.	17.
35	5.9	28.	30.
100	10.	47.	49.
350	19.	87.	90.
1,000	32.	150.	150.
3,500	59.	280.	280.
10,000	100.	470.	470.

†A value of 4.66 has also been frequently used for the quantity $2\sqrt{2}\, k$ [Cf. Ref. 44]. The difference is merely one of rounding.

The Estimated Minimum Detectable Concentration (MDC)

The MDC is a level (not a limit) of activity concentration which is practically achievable by an overall measurement method. As distinguished from the LLD, the MDC considers not only the instrument characteristics (background and efficiency), but all other factors and conditions which influence the measurement. It is an a priori estimate of the activity concentration that can be practically achieved under a specified set of typical measurement conditions. These include the sample size, counting time, self-absorption and decay corrections, chemical yield and any other factors that comprise the activity concentration determination. It cannot serve as a detection limit per se, for any change in measurement conditions or factors will influence the value of the MDC. Its use is limited to establishing, for regulatory purposes, that some minimum overall measurement conditions are met. Any of several factors, such as sample size or counting time, could be varied to satisfy these regulatory values.

Expressions for the MDC can be derived analogously to those for the LLD using the approximation of Eq. (30) and convention of Eq. (31). The results are

$$MDC = 4.65\, K^*\, s_b \tag{36}$$

or

$$MDC = K^*(2.71 + 4.65\, s_b) \tag{37}$$

which are analogous to Equations (34) and (35) for the LLD. The proportionality constant K*, in this case, relates the detector response (counts) to the activity concentration in a sample for a given set of measurement conditions. It may, for example, consist of

$$K^* = \frac{1}{YVTS\,\varepsilon \exp(-\lambda t)} \tag{38}$$

where Y is the fractional yield for the radiochemical separation;

V is the sample size;

T is the counting time interval;

S is the self-absorption correction factor;

ϵ is the detection or counting efficiency;

$\exp(-\lambda t)$ is the correction for radioactive decay between sample collection and counting (time interval, t);

λ is the decay constant for the particular radionuclide with half-life $T_{1/2} = \ln 2/\lambda$.

As discussed previously when considering the K for the LLD, all of the factors contained within the proportionality constant K* for the MDC should be typical or average values for the instrument and measurement procedure.

Misapplication of the LLD or MDC for A Posteriori Decisions

As stated earlier, the LLD is an a priori estimate dependent on only the instrument background and detection efficiency, and the MDC is an a priori estimate for a given type of analysis or measurement process under specified typical conditions. They are not a posteriori decision limits for every measurement. They need not, and should not, therefore, be calculated for each individual measurement. The practice of comparing a unique computed LLD or MDC for each measurement against the measurement result should be avoided. This has sometimes been employed to determine the "significance" of the result for reporting purposes.

If results below a particular computed LLD or MDC are rejected or excluded from data reports, serious errors and distortion in long-term trends could result. Consider the hypothetical situation illustrated in Figure 4. The background is represented by the probability distribution with the mean μ_b and standard deviation σ_b. Similarly, a gross count is given by the distribution with mean μ_g and standard deviation σ_g. Thus the net count, obtained by subtracting the background from the gross count, is the probability distribution with mean $\mu_n = (\mu_g - \mu_b)$ and variance $\sigma_n^2 = (\sigma_g^2 + \sigma_b^2)$. This net count distribution can be compared, as shown, to the estimated LLD calculated with the standard deviation of the background (σ_B) using Eq. (35). In this case, the true or limiting mean of the net count, and a substantial fraction of the distribution are below the computed LLD. A resultant net count (N_n) from a typical individual paired observation of background and gross counts (N_b, N_g), which are both well within $\pm\sigma$, is illustrated. If this value and many others less than the LLD are not reported, then the distribution of the reported values would grossly distort the true situation given by μ_n and σ_n for the population. This point is demonstrated further in the data of Table 11. In this example, twenty-five net count (N_n) measurement results from a population with a mean of 20 were obtained from paired observations of background (N_b) and gross counts (N_g). The average background of 16 counts was used to estimate the LLD:

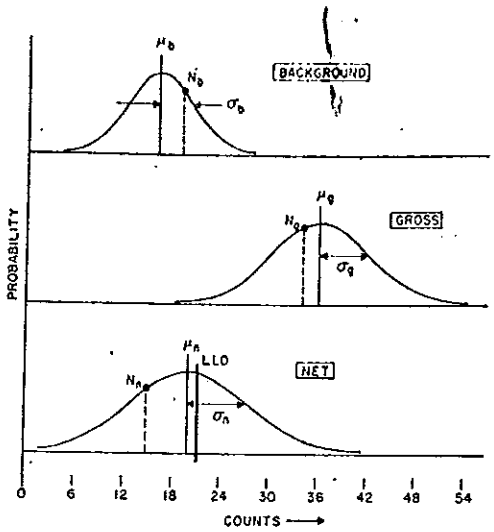


Figure 4. Illustration of the LLD and its relation to the underlying population distributions in a hypothetical measurement situation. See text for details.

$$LLD = 2.71 + 4.65\ \sqrt{16}$$

$$= 21.3 \text{ counts} \quad [\text{from Eq. (35)}]$$

or

$$LLD = 4.65\ \sqrt{16}$$

$$= 18.6 \text{ counts} \quad [\text{from Eq. (34)}].$$

The last two columns in the table contain the tabulated results when values less than 21.3 and 18.6 counts are reported as "less than the LLD" (<LLD). In the first case, only 8 of the 25 measured results would be specified. How does one average this tabulation? The average of the 8 specified values is 25.4 counts, which may be contrasted to the population mean of 20 counts. If the other 17 "<LLD" values are taken as being equal to the LLD* (21.3 counts), the average of the 25 values becomes 22.6 counts. Even if one invokes the less restrictive LLD of 18.6 counts, 9 of the 25 measured values are "<LLD."

This distortion of values near the LLD or MDC will always result in a positive bias when one attempts to make realistic estimates over a period of time. The implications can be serious. When making measurements near background levels, one can expect to frequently obtain values that are less than the estimated LLD or MDC. If these values are not recorded and used in making average estimates, then these estimates are always

* This conservative approach may be the most likely treatment when the tabulated results are ultimately used to make dose assessments.

going to be greater than the "true" representation of the environment. Therefore, it is recommended that every measurement result should be recorded and reported directly as found. "Less than the LLD," "not detected" and similar expressions should not be used in reporting data. This does not imply that the activity in the sample is truly less than some absolute level that can be detected. Rather, it merely indicates that this particular measurement resulted in a single value which was less than an estimated LLD or MDC for the measurement procedure.

The LLD and MDC for Multi-Component and Spectral Analyses

The MDC is conceptually more complex when the net count is a function of the detector response from two or more radionuclides in the sample. In this situation, the proportionality constant K* cannot relate the net count to the activity of just a single radionuclide as was previously done. The simplest case can be represented by

$$N_n = A_1C_1\epsilon_1 + A_2C_2\epsilon_2 \qquad (39)$$

where A_1 and A_2 refer to the activities of two different radionuclides, ϵ_1 and ϵ_2 are the respective detection efficiencies, and C_1 and C_2 are the respective constants which include chemical yields, absorption corrections, decay factors, etc. Then, the LLD for A_1 and A_2 are just given by

$$LLD(A_1) = K_1(2.71 + 4.65s_b)$$
$$LLD(A_2) = K_2(2.71 + 4.65s_b) \qquad (40)$$

where $K_1 = 1/\epsilon_1$ and $K_2 = 1/\epsilon_2$. Obviously, this interpretation results in an LLD for A_1 when A_2 is absent. And conversely, the LLD for A_2 is independent of the amount of A_1. For purposes of the LLD calculation then, Equation (39) in effect reverts to either $N_n = A_1C_1\epsilon_1$ (with A_2 absent) or $N_n = A_2C_2\epsilon_2$ (with A_1 absent). For many procedures, such as the determination of ^{89}Sr and ^{90}Sr, with typical samples, this approach has been subject to the criticism that

TABLE 11

N_b	N_g	N_n	s_n	Distorted N_n Reported Values for LLD=21.3	Distorted N_n Reported Values for LLD=18.6
20	38	18	7.6	<LLD	<LLD
15	36	21	7.1	<LLD	21
18	34	16	7.2	<LLD	<LLD
21	27	6	6.9	<LLD	<LLD
16	36	20	7.2	<LLD	20
17	41	24	7.6	24	24
14	33	19	6.9	<LLD	19
16	44	28	7.7	28	28
17	35	18	7.2	<LLD	<LLD
19	38	19	7.5	<LLD	19
12	32	20	6.6	<LLD	20
15	36	21	7.1	<LLD	21
16	39	23	7.4	23	23
18	34	16	7.2	<LLD	<LLD
23	43	20	8.1	<LLD	20
14	31	17	6.7	<LLD	<LLD
11	28	17	6.2	<LLD	<LLD
16	35	19	7.1	<LLD	19
10	40	30	7.1	30	30
15	37	22	7.2	22	22
22	38	16	7.7	<LLD	<LLD
16	30	14	6.8	<LLD	<LLD
13	37	24	7.1	24	24
17	44	27	7.8	27	27
9	34	25	6.6	25	25
16	36.	20.		(22.6) $^{25.4}$	(21.1) $^{22.6}$

All values in counts.

it does not reflect realistic detection capabilities. It is sometimes argued that the LLD must consider that any net count N_n results from contributions from both ^{89}Sr and ^{90}Sr. This argument is another misapplication of the LLD concept. If it is accepted, then the LLD would no longer be a limit and would be dependent on something other than the instrument background and efficiency characteristics.

The MDC, however, is intended to serve as a practical level that is achievable under a given set of specified measurement conditions, and must consider the effect of multi-components in the sample. Referring again to the case of Equation (39), the MDC for A_1 can be written

$$MDC(A_1) = K_1^* 2 k s_1 \qquad (41)$$

where $k = k_\alpha = k_\beta = 1.645$, and s_1 is the estimated standard deviation of the net sample count corresponding to A_1 over the range of net activities from $A_1 = 0$ to $A_1 = MDC$, i.e. $s_1 = s_o(1) = s_b(1)$ [refer to Eq. (30)]. Since the net sample count corresponding to A_1 is given by

$$N_1 = N_{1+2+b} - N_2 - N_b \quad ,$$

the estimated standard deviation s_1 is

$$s_1 = \sqrt{s^2_{1+2+b} + s^2_2 + s^2_b} \quad ,$$

and, with $s^2_{1+2+b} = s^2_{1+b} + s^2_2$,

$$s_1 = \sqrt{s^2_{1+b} + 2s^2_2 + s^2_b} \quad .$$

If the gross counts N_{1+b} and background N_b are approximately equal (which is a reasonable assumption near the MDC), then s_1 reduces to

$$s_1 = \sqrt{2s^2_b + 2s^2_2}$$

$$= \sqrt{2}\; s_b \sqrt{1 + \frac{s^2_2}{s^2_b}}$$

Making a Poisson assumption, s^2_2 becomes

$$s^2_2 = \frac{N_2}{T^2} = \frac{A_2 C_2 \epsilon_2}{T^2} \quad ,$$

where C_2 and ϵ_2 were given in Eq. (39) and T is the count time. Hence,

$$s_1 = \sqrt{2}\; s_b \sqrt{1 + \frac{A_2 C_2 \epsilon_2}{s_b{}^2 T^2}} \quad ,$$

and substitution into Eq. (41) yields

$$MDC(A_1) = K_1^* 2\sqrt{2}\; k\; s_b \sqrt{1 + \frac{s_2{}^2}{s_b{}^2}}$$

$$= K_1^*(4.65 s_b) \sqrt{1 + \frac{A_2 C_2 \epsilon_2}{s_b{}^2 T^2}} \qquad (42)$$

Analogously to Eq. (37), the MDC for A_1 may alternatively be given by

$$MDC(A_1) = K_1^* \left[2.71 + 4.65 s_b \sqrt{1 + s^2_2/s^2_b}\right] \qquad (43)$$

$$= K_1^* \left\{2.71 + 4.65 s_b \sqrt{1 + \frac{A_2 C_2 \epsilon_2}{s_b{}^2 T^2}}\right\} .$$

The factor K_1^* is, by definition, given by

$$K_1^* = \frac{A_1}{N_n} = \frac{1}{C_1\epsilon_1 + \left(\frac{A_2}{A_1}\right) C_2 \epsilon_2} \quad , \qquad (44)$$

and is dependent on the relative amount of activity A_2 since the net count is apportioned between that due to A_1 and that due to A_2.

One is therefore confronted with a paradox. The MDC should be an a priori estimate, but its determination requires a posteriori knowledge of the amounts of activity in the sample. This dichotomy can be overcome by calculating the MDC for typical sample conditions. This is subject to criticism that some samples may vary substantially from the typical, and that the MDC may then be in considerable error. Without denying this argument, there are two reasons why the effects from such errors are unimportant. First, the MDC is only to be an ESTIMATE; and second, all measurement results should be reported without a comparison to the MDC. As a result, there is no great need for knowing some absolute detection level for every measurement.

The determination of a LLD and MDC for activity measurements by gamma-ray spectrometry is a good example of the above described situation. It is complicated in that the background count rate for the same radionuclide may vary from sample to sample. This is due to the fact that the background is usually the sum of two separate sources of gamma radiation. First, there is the system background radiation for a blank, corresponding to the sample to be measured, as discussed previously. Second, there is Compton-scattered radiation from other higher energy gamma rays in the sample. Thus, it is readily seen that the measurement of blank backgrounds has meaning only for very specific situations, e.g., when only one radionuclide is present in the sample, or when the radionuclide of interest emits the highest energy gamma ray.

It then follows that the LLD, which is independent of sample conditions, is a function

of only the instrument performance (as contained within K) and the uncertainty in this blank background, i.e.

$$LLD = f(K, s_{\substack{\text{blank} \\ \text{bkgnd}}}) \quad . \tag{45}$$

Blank backgrounds are also useful for evaluating the adequacy of shielding and system performance. A large deviation in blank background values may indicate instrument malfunction.

For environmental samples one can assume that almost always two or more gamma-emitting radionuclides will be present in any given sample, thus a more realistic background must be determined for calculating a MDC. This sample background will depend upon the relative amounts of other radionuclides that are present in the sample and the energy of each of their gamma rays. One can then represent the MDC as a function of K* and the uncertainty in this sample background, i.e.

$$MDC = f(K^*, s_{\substack{\text{sample} \\ \text{bkgnd}}}) \tag{46}$$

Pasternack and Harley [43], for a number of typical counting situations, considered the influence of other radionuclides on the detection limit in mock multi-component samples. Their experimental comparisons were made with large NaI(Tl) detectors. This was extended by Wrenn, et al.[45] for the effect of the presence of natural radionuclides on the detection limits of man-made radionuclides in environmental samples. They also made comparisons to Ge(Li) spectrometry. The general problem of detecting small photopeaks in Ge(Li) spectra was also addressed by Head[46]. Fisenne, et al.[47] applied the Pasternack and Harley[43] approach to multi-component alpha spectrometry which has pulse-height distribution data that is very similar to that of gamma spectrometry. In all cases, the MDC for a particular radionuclide was shown to be dependent on the composition of the sample. It depends upon the number, quantity, and spectral characteristics of the other radionuclides in the sample.

Therefore, to satisfy the condition that the MDC be an a priori estimate, it is recommended that it be calculated for each radionuclide from typical spectrum backgrounds for each type of sample.

Examples of these more complex MDC calculations such as for multi-component and spectral analyses can be found in Ref. [1].

Interpretations and Restrictions

In summary, the major concepts underlying the Lower Limit of Detection (LLD) and Minimum Detectable Concentration (MDC) are:

1. The LLD should be viewed as an a priori estimate or guidepost of the detection capability for an instrument. Its value is dependent upon only the detection instrument characteristics (e.g., the efficiency) and the uncertainty in the instrument's background.

2. The MDC should be viewed as an a priori estimate or guidepost of the level of activity concentration that is practically achievable by a specific given instrument, measurement method and type of sample. Its value is dependent upon the characteristics and conditions of the overall measurement system (instrument and method) and on the sample characteristics.

The LLD and MDC are only estimates, and not absolute levels of activity or activity concentration that can or cannot be detected. Similarly, they are not intended to be a posteriori criteria for the presence or absence of activity.

The practical significance of the estimated LLD and MDC are only to serve as guideposts or criteria for experiment design, comparison and optimization purposes, and to serve, for regulatory purposes, as approximate guidelines of minimally acceptable levels that can be practically achieved. As such, all measurement results should be reported directly as obtained, and the estimated LLD or MDC should not be employed to exclude some results from reports. Similarly, the practice of calculating a new LLD or MDC value for every measurement defeats the spirit of the concepts which are to provide a priori estimates. A posteriori calculations contain no additional information, and are neither technically or economically justifiable.

The above interpretations place some important restrictions on the use of the LLD or MDC, particularly with respect to satisfying regulatory specifications. The most important restriction is that it is unreasonable for a regulator to establish absolute values for the MDC or LLD for various combinations of radionuclide and type of sample media, and then expect compliance 100% of the time in all situations. Even ideally the MDC and LLD, by convention, involve 5% risks of false-detection and false non-detection.

A further restriction arises from the fact that the MDC for a radionuclide may vary from sample to sample depending on the characteristics of the sample. For example, the MDC for a radionuclide assayed by gamma-ray spectrometry depends upon the number, quantity, and spectral characteristics of other radionuclides which may be present in the sample. It must therefore be recognized that an MDC established for one specific set of conditions may not be applicable for all other conditions.

Detection capabilities, notably those specified by the Nuclear Regulatory Commission, are stated to be "state-of-the-art for routine environmental measurement"[48]. If these state-of-the-art values are established for one set of assumptions (instrument, procedures and sample

variables), then the MDC values should not be expected to be technically achievable unless the assumptions, particularly typical sample composition, are still valid. Although NRC documents reflect this view[48,49], a number of licensees have reported cases where inspectors have interpreted the MDC as an absolute level, and have tested licensees for specification compliance with simulated spiked samples. If an MDC contained within a licensee's Technical Specifications was determined with blanks or typical samples, then there is no reason to expect the same value to be applicable for atypical samples of dissimilar composition.

The frequent use of an MDC or LLD as a criterion for excluding some results from data reports must not continue. The resulting positive biasing effects of this practice were discussed previously. As a result, it is recommended that all measurement results be reported directly as found; and that "less than MDC" and similar terms never be used.

For some measurements, the reporting of all results will not present a problem. Others, like automated gamma-ray-spectrometer systems, can, however, present a practical problem. Some computer-coupled systems which routinely test for upwards of a hundred or so radionuclides in a system library would require a data report consisting of a value for every radionuclide in the library. Obviously, such an approach is not reasonable. The current practice of using MDC values to decide which results from the long list will be reported is not the solution. What is needed is a practical approach that avoids the problems inherent in using a preselected exclusion criterion, but at the same time is reasonable. This approach requires that efforts be made to reflect on the purpose of the measurements. Measurement of a hundred different radionuclides in a surveillance or monitoring program is neither reasonable or justifiable. Fewer reliable measurement results are much better than many questionable results. What is needed are more good measurements whose results can be viewed with confidence, not more measurements per se. The criteria for what should be measured (and hence reported) should be based upon what is actually needed for the purpose of environmental radioactivity monitoring [50] and upon the dosimetric significance of the radionuclides. Attempts to measure everything that is technically achievable do not serve anyone (not the laboratories or the public). Laboratories, as well as regulators, must begin to recognize this. An approach which is more reasonable and justifiable in terms of the dosimetry and purposes of the monitoring must be taken when designing measurement and data reporting programs.

A practical approach to this problem may be for laboratories and regulators to preselect only those measurements which are reasonable and justifiable in terms of the dosimetry and purposes of the monitoring program. The effect of this preselection could be evaluated and its magnitude incorporated as a systematic uncertainty component in the assessment and propagation of an overall uncertainty (see Part III). In gamma-ray spectrometry with automated systems, for example, the magnitude of this systematic uncertainty could be evaluated by analyzing a test spectrum first without any preselection criteria, and then with a limited system library containing only the preselected radionuclides. The difference between the two results could be used as an estimate of the additional systematic uncertainty. This uncertainty component could be evaluated for each preselected radionuclide of interest using test spectra containing a full range of radionuclides found in typical samples. This practical approach may be a reasonable compromise.

Recommendations

IV.1 Only two expressions for detection limits, based on a uniform consistent methodology, should be employed. The recommended terms are the ESTIMATED LOWER LIMIT OF DETECTION (abbreviated LLD) and the ESTIMATED MINIMUM DETECTABLE CONCENTRATION (MDC).

IV.2 The practical significance of the estimated LLD and MDC are only to serve as guideposts or criteria for experiment design, comparison and optimization purposes, and to serve, for regulatory purposes, as approximate guidelines of minimally acceptable detection capabilities.

IV.3 The LLD should be viewed as an a priori estimate or guidepost of detection capability, and not as absolute levels of activity that can or cannot be detected. It is not intended to be an a posteriori criterion for the presence of activity.

IV.4 The MDC should be viewed as an a priori estimate or guidepost of the capability for detecting an activity concentration by a given measurement system, procedure and type of sample, and not as absolute activity concentrations that can or cannot be detected.

IV.5 The estimated LLD or MDC should be based on the approximation of Eq. (30) which is an approach derived from statistical hypothesis testing.

IV.6 The estimated LLD or MDC should be calculated, using the convention of Eq. (31), for "5% risks of false detection and false non-detection."

IV.7 The estimated LLD should be calculated from Eq. (34), (35), (40), (45) or comparable equations which are derived from the approximation of Eq. (30) and convention of Eq. (31).

IV.8 The estimated MDC should be calculated from Eq. (36), (37), (42), (43), (46), or comparable equations which are derivable from the approximation of Eq. (30) and convention of Eq. (31).

IV.9 The estimated LLD should be expressed in units of activity and should include

only the instrument parameter which relates the detector response (counts) to activity. This normally would be only the detection efficiency or calibration factor for the instrument, and would not include other parameters such as the sample size, chemical yield, decay scheme parameters, absorption and attenuation corrections, decay factors, etc.

IV.10 The estimated MDC should be expressed as activity per unit mass or activity per unit volume, and its calculation would include all parameters which relate the detector response (counts) to the activity concentration in a given type of sample. These may include detection efficiencies, chemical yields, absorption and attenuation corrections, decay scheme parameters, decay factors, etc.

IV.11 The estimated LLD should be calculated using average blank backgrounds and efficiencies for the instrument; the estimated MDC should be calculated using average sample backgrounds and parameters for typical samples.

IV.12 The estimated LLD for a given instrument should be calculated for each radionuclide of interest, but it is independent of the sample characteristics; in contradistinction, the estimated MDC for a given procedure should be calculated for each radionuclide for each type of sample.

IV.13 To avoid possible positive biases in long-term data trends, all measurement results should be reported directly as obtained. "Less than LLD," "less than MDC," "not detected" and similar expressions should not be reported. Similarly, the LLD or MDC should not be used as a decision criterion for excluding some results from data reports.

REFERENCES

[1] A Guide and Recommendations for Reporting of Environmental Radiation Measurements Data, Ad Hoc Subcommittee on Data Reporting, National Bureau of Standards Special Publication, to be published (1980).

[2] American Society for Testing and Materials, ASTM Standard for Metric Practice E380-72 (1972). Excerpts from this Standard also appear in the Annual Book of ASTM Standards Part 31 (Water) and Part 45 (Nuclear Standards).

[3] American National Standards Institute, "American Standard for Metric Practice," American National Standard Z210.1-76 (1976).

[4] British Committee on Radiation Units and Measurements, "Draft Recommendations on the Introduction of the New SI Units for Use with Radioactivity and Ionizing Radiations," published by National Physical Laboratory, United Kingdom (January 1978).

[5] National Bureau of Standards, The International System of Units (SI), NBS Special Publication 330, 1977 Edition.

[6] National Bureau of Standards, NBS Guidelines for Use of the Metric System, LC1056 (Revised August 1977).

[7] International Commission on Radiation Units and Measurements, "Radiation Quantities and Units," ICRU Report 19 (July 1971).

[8] National Council on Radiation Protection and Measurements, Environmental Radiation Measurements, NCRP Report No. 50 (1976).

[9] Harry H. Ku, "Statistical Concepts in Metrology," in Precision Measurement and Calibration: Statistical Concepts and Procedures, National Bureau of Standards, Special Publication 300, Vol. 1, 296-330 (1969).

[10] R. C. Pinkerton and C. E. Gleit, "The Significance of Significant Figures," Journal of Chemical Education, 44, No. 4, 2332-34 (April 1967).

[11] L. G. Parratt, Probability and Experimental Errors in Science, Wiley NY (1961), p.69-71.

[12] P. R. Bevington, Data Reduction and Error Analysis for the Physical Sciences, McGraw-Hill Book Co., NY (1969), p. 4-6.

[13] M. G. Natrella, "Experimental Statistics," National Bureau of Standards Handbook 91 (1973).

[14] National Council on Radiation Protection and Measurements, "A Handbook of Radioactivity Measurements Procedures," NCRP Report No. 58 (1978). A revision of National Bureau of Standards Handbook 80.

[15] W. J. Dixon and F. J. Massey, Jr., Introduction to Statistical Analysis, 1st. Ed., McGraw-Hill Book Co., N.Y. (1951).

[16] A. H. Jaffey, "Statistical Tests for Counting," Nucleonics 18, No. 11, 180-4 (1960).

[17] Energy Research & Development Administration, "A Guide for Environmental Radiological Surveillance at ERDA Installations," ERDA 77-24 (March 1977).

[18] H. H. Ku (ed.), "Precision Measurement and Calibration: Statistical Concepts and Procedures," National Bureau of Standards, Special Publication 300, Volume 1 (1969).

[19] A. Goldstein, Biostatistics, MacMillan Publ. Co., N.Y. (1964).

[20] H. H. Ku, "Notes on the Use of Propagation of Error Formulas," J. Research NBS 70C,

No. 4, 236-73 (1966). (Also reprinted in Ref. [18], p. 331-41.)

[21] P. C. Stevenson, "Processing of Counting Data," National Academy of Sciences - National Research Council, Nuclear Science Series NAS-NS-3109, U. S. Atomic Energy Commission (1965).

[22] R. D. Evans, "Statistical Fluctuations in Nuclear Processes," Chap. 5 in C. L. Yuan and C. S. Wu (eds.), Methods of Experimental Physics, Vol. 5, Pt. B, Academic Press (1963).

[23] W. J. Price, Nuclear Radiation Detection, 2nd ed., McGraw-Hill (1964).

[24] Lee H. Ziegler and H. M. Hunt, "Quality Control for Environmental Measurements Using Gamma-Ray Spectrometry," U.S. Environmental Protection Agency, EPA-600/7-77-144 (Dec. 1977).

[25] International Commission on Radiation Units and Measurements, "Certification of Standardized Radioactive Sources," ICRU Report 12 (1968).

[26] A. Williams, P. J. Campion and J. E. Burns, "Statement of Results and Their Accuracy," Nucl. Instr. and Meth. 112, 373-376 (1973).

[27] H. H. Ku, "Discussion on Statement of Data and Errors," Nucl. Instr. and Meth. 112, 391 (1973).

[28] H. H. Ku, "Statistical Methods Applicable to Counting Experiments and Evaluation of Experimental Data," Nucl. Instr. and Meth. 112, 377-383 (1973).

[29] W. J. Youden, eleven papers on statistical experimental designs useful in physical sciences reprinted in Ref. [18].

[30] B. L. Welch, "The Generalization of 'Student's' Problem When Several Different Population Variances Are Involved," Biometrika 34, 28-35 (1947).

[31] A. A. Aspin, "Tables for Use in Comparisons Whose Accuracy Involves Two Variances, Separately Estimated," Biometrika 36, 290-6 (1949). See also Ref. [30] and B. L. Welch, Biometrika 34, 87 (1947).

[32] National Council on Radiation Protection and Measurements, "Tritium Measurement Techniques," NCRP Report No. 12, p. 72-74 (1976).

[33] E. S. Pearson and H. O. Hartley, Biometrika Tables for Statisticians, Vol. 1, Cambridge Univ. Press, London (1954).

[34] Y. Le Gallic, "Discussion on Statement of Data and Errors," Nucl. Instr. and Meth. 112, 393 (1973).

[35] C. Eisenhart, "Expression of the Uncertainties of Final Results," Science 160, 1201-1204 (1968). (Also reprinted in Ref. [18], p. 69-72.)

[36] H. H. Ku, "Expressions of Imprecision, Systematic Error, and Uncertainty Associated with a Reported Value," Measurements and Data, p. 72-77 (July-Aug. 1968). (Also reprinted in Ref. [18], p. 73-78.)

[37] P. J. Campion, J. E. Burns and A. Williams "A Code of Practice for the Detailed Statement of Accuracy," National Physical Laboratory, London (1973).

[38] J. Müller, "Discussion on Statement of Data and Errors," Nucl. Instr. and Meth. 112, 393 (1973).

[39] S. Wagner, PTB Mitteilungen 79, 343 (1969). An English version, "How to Treat Systematic Errors in Order to State the Uncertainty of a Measurement," is available from the Physikalish-Technische Bundesanstalt, Braunschweig, as report number PMRB 31/69.

[40] A. Williams, P. J. Campion and J. E. Burns "Statement of Results of Experiments and Their Accuracy," Nucl. Instr. and Meth. 112, 373-376 (1973).

[41] Lloyd A. Currie, "Limits for Qualitative Detection and Quantitative Determination," Anal. Chem. 40, 586-93 (March 1968).

[42] B. Altshuler and B. Pasternack, "Statistical Measures of the Lower Limit of Detection of a Radioactivity Counter," Health Physics 9, 293-8 (1963).

[43] B. S. Pasternack and N. H. Harley, "Detection Limits for Radionuclides in the Analysis of Multi-Component Gamma-Spectrometer Data," Nucl. Instr. and Meth. 91, 533-40 (1971).

[44] John H. Harley (ed.), "EML Procedures Manual," Department of Energy, Environmental Measurements Laboratory, HASL-300 (1972 ed., revised annually).

[45] McDonald E. Wrenn, S. M. Jinks, L. M. Hairr, A. S. Paschoa and J. W. Lentsch, "Natural Activity in Hudson River Estuary Samples and Their Influence on the Detection Limits for Gamma Emitting Radionuclides Using NaI Gamma Spectrometry," p. 897-916 in The Natural Radiation Environment II, J. A. S. Adams, et al. (eds.) Report CONF-720805 (U.S. ERDA, 1972).

[46] J. H. Head, "Minimum Detectable Photopeak Areas in Ge(Li) Spectra," Nucl. Instrum. & Meth. 98, 419 (1972).

[47] I. M. Fisenne, A. O'Tolle and R. Cutler, "Least Squares Analysis and Minimum Detection Levels Applied to Multi-Component Alpha Emitting Samples," Radiochem. Radioanal. Letters 16, 1, 5-16 (1973).

[48] Nuclear Regulatory Commission, Branch Technical Position on Regulatory Guide 4.8 (March 1978).

[49] Cf., Nuclear Regulatory Commission, Regulatory Guide 4.8 "Environmental Technical Specifications for Nuclear Power Plants" (Dec. 1975); Regulatory Guide 4.14, "Measuring, Evaluating, and Reporting Radioactivity in Releases of Radioactive Materials in Liquid and Airborne Effluents from Uranium Mills" (June 1977); Regulatory Guide 4.16, "Measuring, Evaluating, and Reporting Radioactivity in Releases of Radioactive Materials in Liquid and Airborne Effluents from Nuclear Fuel Processing and Fabrication Plants" (March 1978); and others.

[50] Nuclear Regulatory Commission, Radiological Environmental Monitoring by NRC Licensees for Routine Operations of Nuclear Facilities, NUREG-0475 (October 1978).