~~Designation:E141–91(Reapproved 2003)ε1~~ Designation: E141 – 10

An American National Standard

Standard Practice for Acceptance of Evidence Based on the Results of Probability Sampling[1]

This standard is issued under the fixed designation E141; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

~~ε1 NOTE—Editorial changes were made throughout in November 2003.~~

1. Scope

1.1 This practice presents rules for accepting or rejecting evidence based on a sample. Statistical evidence for this practice is in the form of an estimate of a proportion, an average, a total, or other numerical characteristic of a finite population or lot. It is an estimate of the result which would have been obtained by investigating the entire lot or population under the same rules and with the same care as was used for the sample.

1.2 One purpose of this practice is to describe straightforward sample selection and data calculation procedures so that courts, commissions, etc. will be able to verify whether such procedures have been applied. The methods may not give least uncertainty at least cost, they should however furnish a reasonable estimate with calculable uncertainty.

1.3 This practice is primarily intended for one-of-a-kind studies. Repetitive surveys allow estimates of sampling uncertainties to be pooled; the emphasis of this practice is on estimation of sampling uncertainty from the sample itself. The parameter of interest for this practice is effectively a constant. Thus, the principal inference is a simple point estimate to be used as if it were the unknown constant, rather than, for example, a forecast or prediction interval or distribution devised to match a random quantity of interest.

~~1.4~~

1.4 A system of units is not specified in this standard.

1.5 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

E105 Practice for Probability Sampling of Materials

E122 Practice for Calculating Sample Size to Estimate, With Specified Precision, the Average for a Characteristic of a Lot or Process ~~E178 Practice for Dealing With Outlying Observations~~

E456 Terminology Relating to Quality and Statistics ~~Note1—Practice E105 provides a statement of principles for guidance of ASTM technical committees and others in the preparation of a sampling plan for a specific material. Practice E122 aids in deciding on the required sample size. Practice E178 helps insure better behaved estimates. Terminology E456 provides definitions of statistical terms used in this standard.~~

E1402 Guide for Sampling Design

E2586 Practice for Calculating and Using Basic Statistics

3. Terminology

~~3.1 *Definitions:*~~

3.1 *Definitions*—Refer to Terminology E456 for definitions of other statistical terms used in this practice.

3.1.1 ~~*Equal Complete Coverage Result*~~audit subsample, *n*~~—the numerical characteristic (θ) of interest calculated from observations made by drawing randomly from the frame, all of the sampling units covered by the frame.~~

[1] This practice is under the jurisdiction of ASTM Committee E11 on Quality and Statistics and is the direct responsibility of Subcommittee E11.10 on Sampling / Statistics.

~~Current edition approved Oct. 1, 2003. Published November 2003. Originally approved in 1959. Last previous edition approved in 1991 as E141–91. DOI: 10.1520/E0141-91R03E01.~~

Current edition approved May 15, 2010. Published August 2010. Originally approved in 1959. Last previous edition approved in 2003 as E141 – 91 (2003) ε1. DOI: 10.1520/E0141-10.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

3.1.1.1—*a small subsample of a sample selected for review of all sample selection and data collection procedures.*
3.1.2 *equal complete coverage result*, *n*—the numerical characteristic of interest calculated from observations made by drawing randomly from the frame, all of the sampling units covered by the frame.
3.1.2.1 *Discussion*—Locating the units and evaluating them are supposed to be done in exactly the same way and at the same time as was done for the sample. The quantity itself is denoted θ. The equal complete coverage result is never actually calculated. Its purpose is to serve as the objectively defined concrete goal of the investigation. The quantity θ may be the population mean, ($\bar{Y}$)($\bar{Y}$), total *(Y)*, median (*M*), the proportion (*P*), or any other such quantity.
3.1.2
3.1.3 *frame*, *n*—a list, compiled for sampling purposes, which designates all of the sampling units (items or groups) of a population or universe to be considered in a specific study.
3.1.2.1*Discussion*—The list may cover a specific shipment or lot, all households in a county, a state, or country; for example, any population of interest. Every sampling unit in the frame (1) has a unique serial number, which may be preassigned or determined by some definite rule, (2) has an address—a complete and clear instruction (or rules for its formulation) as to where and when to make the observation or evaluation, (3) is based on physically concrete clerical materials such as directories, dials of clocks or of meters, ledgers, maps, aerial photographs, etc., referred to in the addresses.
3.1.3 **E1402**
3.1.4 *probability sample*, *n*—a sample in which the sampling units are selected by a chance process such that a specified probability of selection can be attached to each possible sample that can be selected. **E1402**
3.1.5 *replicate subsamples*, *n*—a number of disjoint samples, each one separately drawn from the frame in accord with the same probability sampling plan.
3.1.6 *sample*, *n*—a group of items, observations, test results, or portions of material, taken from a larger collection of such items; it provides information for decisions concerning the larger collection.
3.1.3.1*Discussion*—A particular sample is identified by the set of serial numbers from the randomization device and by the addresses on the frame generated by those serial numbers.
3.1.4*sampling unit, n*—an item, test specimen or portion of material that is to be subjected to evaluation as part of the sampling plan.
3.1.4.1*Discussion*—If it is not feasible to select test specimens or laboratory samples individually, the sampling unit may be a group of items, for example, a row, an entire case of items, or a prescribed area (as in the examination of a finishing process).
3.1.4.2By a more expensive method of measurement (future time, more elaborate frame) it may be possible to define a quantity, θ′, as a target parameter or ideal goal of an investigation. Criticism that holds θ to be an inappropriate goal should demonstrate that the numerical difference between θ and θ′ is substantial. Measurements may be imprecise but so long as measurement errors are not too biased, a large size of the lot or population, N, insures that θ and θ′ are essentially equal. —a group of observations or test results, taken from a larger collection of observations or test results, which serves to provide information that may be used as a basis for making a decision concerning the larger collection. **E2586**
3.1.7 *sampling unit*, *n*—an item, group of items, or segment of material that can be selected as part of a probability sampling plan. **E1402**

4. Significance and Use

4.1This practice is designed to permit users of sample survey data to judge the trustworthiness of results from such surveys. Section Significance and Use
4.1 This practice is designed to permit users of sample survey data to judge the trustworthiness of results from such surveys. Practice E105 provides a statement of principles for guidance of ASTM technical committees and others in the preparation of a sampling plan for a specific material. Guide E1402 describes the principal types of sampling designs. Practice E122 aids in deciding on the required sample size.
4.2 Section 5 gives extended definitions of the concepts basic to survey sampling and the user should verify that such concepts were indeed used and understood by those who conducted the survey. What was the frame? How large (exactly) was the quantity *N*? How was the parameter θ estimated and its standard error calculated? If replicate subsamples were not used, why not? 4.2Adequate answers should be given for all questions. There are many acceptable answers to the last question.
4.3 If the sample design was relatively simple, such as simple random or stratified, then good fully valid estimates of sampling variance are easily available. If a more complex design was used then methods such as discussed in Ref.Ref **(1)**[3] or in Guide E1402 may be acceptable. Replicate Use of replicate subsamples is the most straightforward way to estimate sampling variances when the survey design is complex.
4.3Once4.4 Once the survey procedures that were used satisfy Section 5, consult Section 4 to see if any increase in sample size is needed. The calculations for making it are objectively described in Section 4.
4.4Refer to Section , see if any increase in sample size is needed. The calculations for making it objectively are described in Section 6to guide in the interpretation of the uncertainty in the reported value of the parameter estimate, θ, i.e. the value of its

[3] The boldface numbers in parentheses refer to a list of references at the end of this standard.

standard error, se(θ). The quantity se(θ) should be reviewed to verify that the risks it entails are commensurate with the size of the sample. .

4.5 Refer to Section 7 to guide in the interpretation of the uncertainty in the reported value of the parameter estimate, $\hat{\theta}$, that is, the value of its standard error, *se*($\hat{\theta}$). The quantity *se*($\hat{\theta}$) should be reviewed to verify that the risks it entails are commensurate with the size of the sample.

4.6 When the audit subsample shows that there was reasonable conformity with prescribed procedures and when the known instances of departures from the survey plan can be shown to have no appreciable effect on the estimate, the value of $\hat{\theta}$ is appropriate for use.

5. Descriptive Terms and Procedures Concepts and Procedures of Sampling

5.1 *Probability Sampling Plans*—include instructions for using either:

5.1.1carefully prepared tables of random number,

5.1.2computer algorithms, carefully programmed and run on a large computer, to generate pseudo-random numbers or,

5.1.3certifiably honest physical devices, such as coin flips, to select the sample units so that inferences may be drawn from the test results and decisions may be made with risks correctly calculated by probability theory.

5.1.4Such plans are defined and their relative advantages discussed in Refs. **Probability sampling** is a procedure by which one obtains a result from a selected set of sampling units that will agree, within calculable limits of variation, with the equal complete coverage result. Probability sampling plans include instructions for using either (*1*) prepared tables of random numbers, (*2*) computer algorithms to generate pseudo-random numbers, or (*3*) certifiably honest physical devices to select the sample units so that inferences may be drawn from the test results and decisions may be made with risks correctly calculated by probability theory.

5.1.1 Such plans are defined and their relative advantages discussed in Guide E1402 and Refs (**1**), (**21-3**), and (**3**) .

5.2

5.2 *Procedures* must be described in written form. Parties interested in collecting data should agree on the importance of knowing θ and its definition including measurement methods. The frame shall be carefully and explicitly constructed. Every sampling unit in the frame (*1*) has a unique serial number, which may be preassigned or determined by some definite rule and (*2*) has an address—a complete and clear instruction (or rules for its formulation) as to where and when to make the observation or evaluation. Address instructions should refer to concrete clerical materials such as directories, dials of clocks or of meters, ledgers, maps, aerial photographs, etc. Duplicates in the frame shall be eliminated. *N* shall be well established. Random numbers (or a certifiably honest physical random device) shall dictate selection of the sample. There shall be no substitution of one sampling unit for another. The method of sample selection shall permit calculation of a standard error of the estimate. The use of replicate subsamples is recommended (see 5.4). An audit subsample should be selected and processed and any departures from prescribed measurement methods and location instructions noted (see 5.5). A report should list $\hat{\theta}$ and its standard error with the degrees of freedom in the *se*($\hat{\theta}$).

5.3 *Parameter Definition*—The equal complete coverage result may or may not be acceptable evidence. Whether it is acceptable depends on many considerations such as definitions, method of test, care exercised in the testing, completeness of the frame, and on other points not to be settled by statistical theory since these points belong to the subject matter, and are the same whether one uses sampling or not. Mistakes, whether in testing, counting, or weighing will affect the result of a complete coverage just as such mistakes will affect the sample result. By a more expensive method of measurement or more elaborate sampling frame, it may be possible to define a quantity, θ′, as a target parameter or ideal goal of an investigation. Criticism that holds θ to be an inappropriate goal should demonstrate that the numerical difference between θ and θ′ is substantial. Measurements may be imprecise but so long as measurement errors are not too biased, a large size of the lot or population, *N*, insures that θ and θ′ are essentially equal.

5.4 *Replicate Subsamples*—a number of disjoint samples, each one separately drawn from the frame in accord with the same probability sampling plan. When appropriate, separate laboratories should each work on separate replicate subsamples and teams of investigators should be assigned to separate replicate subsamples. This approach insures that the calculated standard error will not be a systematic underestimate. Such subsamples were called interpenetrating in Ref. **—When appropriate, separate laboratories should each work on separate replicate subsamples and teams of investigators should be assigned to separate replicate subsamples. This approach insures that the calculated standard error will not be a systematic underestimate. Such subsamples were called interpenetrating in Ref (4) where many of their basic properties were described. See Ref.Ref (25) for further theory and applications.**

5.2.1*Discussion*—For5.4.1 For some types of material, a sample selected with uniform spacing along the frame (systematic sample) has increased precision over a selection made with randomly varying spacings (simple random sample). Examples include sampling mineral ore or grain from a conveyor belt or sampling from a list of households along a street. If the systematic sample is obtained by a single random start the plan is then a probability sampling plan, but it does not permit calculating the standard error as required by this practice. After dividing the sample size by an integer *k* (such as $k = 4$ or $k = 10$) and using a random start for each of *k* replicate subsamples, some of the increased precision of systematic sampling (and a standard error on $k - 1$ degrees of freedom) can be achieved.

5.2.25.5 Audit Subsample—a small An audit subsample of the survey sample (as few as 10 observations may should be adequate)taken for review of all procedures from use of the random numbers through locating and measurement, to editing, coding, data entry and tabulation. Selection of the audit subsample may be done by putting the *n* sample observations in order as they are collected, calculating the nearest integer to $\sqrt{n}$, or some other convenient integer, and taking this number to be the spacing for systematic selection of the audit subsample. As few as 10 observations may be adequate. The review should uncover any gross departures from prescribed practices or any conceptual misunderstandings in the definitions. If the audit subsample is large enough (say 30 observations or more) the regression of audited values on initial observations may be used to calibrate the estimate. This technique is the method of two-phase sampling as discussed in Ref.Ref (**1**). Helpful discussion of an audit appears in Ref.Ref (**2**).

5.2.3*Estimate*—a quantity calculated on the n sample observations in the same way as the equal complete coverage result θ would have been calculated from the entire set of N possible observations of the population; the symbol θ denotes the estimate. (In calculating θ, replicate subsample membership is ignored.)

5.2.3.1*Discussion*—An estimate has a sampling distribution induced from the randomness in sample selection. The equal complete coverage result is effectively a constant while any estimate is only the value from one particular sample. Thus, there is a mean value of the sampling distribution and there is also a standard deviation of the sampling distribution.

5.2.4*Standard Error*—the quantity computed from the observations as an estimate of the sampling standard deviation of the estimate; *se* (θ) denotes the standard error.

5.2.4.1*Example 1*—When θ is the population average of the N quantities and a simple random sample of size

5.6 The estimate is a quantity calculated on the *n* was drawn, then the sample average *y* becomes the usual estimate θ, where

$$\theta = \bar{y} = \sum_{i=1}^{n} \quad (1)$$

(1) θ=*y*=*i*=1*nsample observations in the same way as the equal complete coverage result* θ *would have been calculated from the entire set of N* possible observ
denotes the estimate. In calculating $\hat{\theta}$, replicate subsample membership is ignored.

5.6.1 An estimate has a sampling distribution induced from the randomness in sample selection. The equal complete coverage result is effectively a constant while any estimate is only the value from one particular sample. Thus, there is a mean value of the sampling distribution and there is also a standard deviation of the sampling distribution.

5.7 The standard error is the quantity computed from the observations as an estimate of the sampling standard deviation of the estimate; *se*$(\hat{\theta})$ denotes the standard error.

5.7.1 When θ is the population average of the *N* quantities and a simple random sample of size *n* was drawn, then the sample average $\bar{y}$ becomes the usual estimate $\hat{\theta}$, where:

$$\hat{\theta} = \bar{y} = \sum_{i=1}^{n} y_i/n. \quad (1)$$

$$\hat{\theta} = \bar{y} = \sum_{i=1}^{n} y_i / n \quad (1)$$

The quantities y_1, y_2, ..., y_n denote the observations. The standard error is calculated as:

$$se(\theta) = se(\bar{y}) = \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2 / n(n-1)}. \quad (2)$$

(2) *se*(θ)=*se*(*y*)=*i*=1*n*(*yi*−*ydenote the observations. The standard error is calculated as*:

$$se(\hat{\theta}) = se(\bar{y}) = \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2 / n(n-1)} \quad (2)$$

There are $n - 1$ degrees of freedom in this standard error.

5.7.1.1 *Example*—When the observations are:

81.6, 78.7, 79.7, 78.3, 80.9, 79.5, 79.8, 80.3, 79.5, 80.7

then *y*then $\bar{y}$ = 79.90 and *se*$(\bar{y})$=0.32. As this example illustrates, formula (2) is correct when *k* replaces *n* and subsample estimates are used in place of observations.

5.2.4.2*Example 2 on the Finite Population Correction (fpc)*—Multiplying se $(\bar{y})$ se$(\bar{y})$ = 0.32.

5.7.2 *Finite Population Correction (fpc)*—Multiplying *se*$(\bar{y})$ by $\sqrt{1 - n/N}$ is always correct when the goal of the survey is to estimate the finite population mean (θ= *Y*). Using the previous data and if *1* − *n / N is always correct when the goal of the survey is to estimate the finite population mean* (θ = $\bar{Y}$). *If random measurement error exists in the observations, then* θ′ *based on a reference measurement method may be a more appropriate survey goal than* θ *(see 5.3). If so, then se*$(\bar{y})$ would be further adjusted upward by an amount somewhat less than the downward adjustment of the fpc. Both of these adjustments are often numerically so small that these adjustments may be omitted—leaving *se*$(\bar{y})$ of Eq 2 as a slight overestimate.

5.7.2.1 *Example*—Using the previous data and if *N* = 50, then *se*$(\bar{y})$ se$(\bar{y})$ becomes *se*$(\bar{y})$=0.28 after applying the fpc. If random measurement error exists in the observations, then θ′ based on a reference measurement method may be a more appropriate survey goal than θ (see section 4.1.4.1). If so, then *se*$(\bar{y})$ would be further adjusted upward by an amount somewhat less than the downward