doi: 10.1255/tosf.77

The role of inference in food safety

Charles A. Ramsey
EnviroStat, Inc., PO Box 657, Windsor, CO 80550, USA chuck@envirostat.org

Concerned individuals have been trying to determine the safety of their food since ancient times. In ancient times, people themselves were the ultimate test of food safety, but as human evolution progressed, other techniques such as sensory perception and experimentation on animals were used. Today sophisticated analytical techniques and models are available to measure and predict food safety. These sophisticated techniques and models are dependent not only on the quality of samples that are collected and analyzed but also on how inferences are made from the analytical results to the food being sampled. Unfortunately, the Theory of Sampling and the role of inference have not been fully integrated into prediction of food safety.The basis for many "modern" food sampling protocols was developed prior to the development of the Theory of Sampling. Many of these sampling protocols were based on concepts of acceptance sampling procedures and associated inference. The Theory of Sampling enables the representative sampling of bulk materials and eliminates the reliance of acceptance sampling as the only method for the characterization of food and utilizes a different type of inference than for acceptance sampling. This contribution addresses the differences between inference for acceptance sampling and inference for the sampling of bulk materials and the implications of these differences for food safety.

Introduction

The testing of food for poison has occurred since ancient times. Until recently (and even some today), most food testing was performed by having someone taste the food and waiting to see if there were any ill effects. This process worked for fast acting poisons but was ineffective for slower acting poisons. Through the use of sophisticated analytical techniques and better understanding of toxins, the use of humans to make inference regarding the safety of food has greatly diminished. However, it has been reported that several notable people, including Vladimir Putin[1] and Barack Obama[2] have recently used food tasters to ensure their food is not poisoned.

Food safety today is mostly dependent on manufacturing practices that focus on critical contamination points in the manufacturing process. These are Hazard Analysis and Critical Control Points (HACCP) conceived in the 1960s when Pillsbury developed food for the first space flights[3] and Good Manufacturing Practices (GMP). However, there is still a need for inspection of food to determine the adequacy of HACCP and GMP, assess contamination after manufacturing, respond to outbreaks and a variety of other reasons. Since the amount of food produced is very large compared to the number of samples collected, it is critical that sampling protocols be very efficient. Because the consequences of contaminated food are extreme, it is also critical that inferences are correct and that correct decisions are made.

Inference is the process of estimating parameters of a Decision Unit[4] based on analytical results of samples from the Decision Unit[5]. The most common parameter estimated is the true mean concentration of an analyte of interest. The requirement to enable inference is that the sample is from an equiprobabilistic (random) selection of the elements within the Decision Unit. The error (closeness of the estimation to the true value) in the inference is controlled through the application of principles of the Theory of Sampling[6,7] (TOS). See Figure 1.

In cases where multiple Decision Units exist, inference can also be used to estimate the percent (portion) of Decision Units that possess a specific concentration or characteristic. Equi-probabilistic (random) selection is also required for this type of inference, but it is random selection of the Decision Units, not random selection of the elements within the Decision Unit as above, that must be equiprobable. Figure 2.

These two inferences (to an individual Decision Unit and to unsampled Decision Units) are sometimes used individually and sometimes combined, depending on the Sample Quality Criteria (SQC)[8]. Understanding of the differences in the types of inference is critical for the design of sampling protocols as well as for the interpretation of analytical results and final decision-making. In both types of inference, inference is made from what is sampled (collected) to what is not sampled (not collected). While this paper focuses on food safety, the inference principles are generic and applicable to all sampling and analysis.

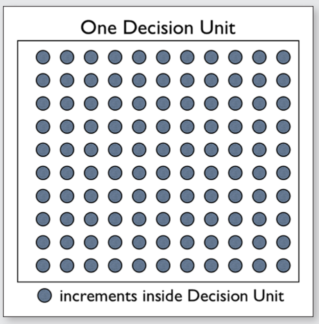

Figure1. Sampling from within a Decision Unit allows inferences to be made with respect to the entire Decision Unit.

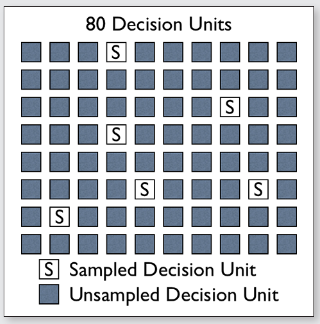

Figure 2. Sampling individual Decision Units allows inferences to be made with respect to all the Decision Units.

Inference to a single decision unit

A Decision Unit may be small in size/mass or it may be quite large. There are no size or geometric constraints on a Decision Unit. In the case of a small Decision Unit, it may be possible to collect the entire Decision Unit (DU) as the primary sample. This may be the case for a loaf of bread DU or a cantaloupe DU. However, in most cases the Decision Unit it too large to practically collect in its entirety as a primary sample. This would be the case for a warehouse of bread DU or a truck of cantaloupes DU. In some cases, even if the entire Decision Unit can be collected in its entirety that may not be desired as there would be nothing left.

In the laboratory, it is also possible to analyze the entire primary sample as received or it may be necessary to collect a smaller test portion from the primary sample for subsequent analysis. The possible sampling situations in the field and in the laboratory are very similar, either the entire DU (primary sample) can be taken or the DU (primary sample) must be representatively sampled according to the principles of TOS. In total there are four possibilities:

- Take entire DU in the field, analyze entire primary sample in the laboratory
- Take entire DU in the field, subsample primary sample in the laboratory
- Sample DU in the field, analyze entire primary sample in the laboratory
- Sample DU in the field, subsample primary sample in the laboratory

Each and every one of these possibilities exists in food safety. Inference for each of these possibilities is discussed below.

Take entire DU in the field, analyze entire primary sample in the laboratory

Inference is the simplest in this case. The result from the laboratory is the true concentration of the analyte of interest in the primary sample (except for analytical uncertainty, which will not be discussed). The primary sample in this case is the Decision Unit. No inference is required as everything is taken and analyzed.

Take entire DU in the field, subsample primary sample in the laboratory

The result from the laboratory is used to estimate the true concentration of the analyte of interest in the primary sample. Since the entire primary sample was not analyzed, an inference must be made from the analytical result to the concentration of the analyte of interest in the primary sample. The entire DU in the field was collected as the primary sample, so no inference is required from the primary sample to the DU.

Sample DU in the field, analyze entire primary sample in the laboratory

The result from the laboratory is the true concentration of the analyte of interest in the primary sample. The entire primary sample was analyzed in the laboratory, so there is no inference required from the analytical result to the primary sample. However, the entire DU was not collected in the field as the primary sample, so there will be an inference from the analytical result of the primary sample to the DU.

Sample DU in the field, subsample primary sample in the laboratory

The result from the laboratory is used to estimate the true concentration of the analyte of interest in the primary sample. The primary sample was sampled in the laboratory, so there is an inference from the analytical result to the primary sample. However, the DU was also sampled so there will be another inference from the primary sample to the DU. In this case there are two inferences being made. One inference from the analytical result to the primary sample and one inference from the primary sample to the Decision Unit.

Inference for each of these situations can be made directly or through some type of statistical calculation. Direct inference occurs when an individual analytical result is used to estimate the concentration in the primary sample and/or to the entire Decision Unit. This is very common. Alternatively, several measurements can be made and a statistical calculation used for inference to either the primary sample or to the DU. Examples may be an average or a 95% upper confidence interval of the mean. The type of inference desired (direct or statistical calculation) therefore has an impact on the sampling protocol. For each type of inference it must be determined how that inference is going to be made and the error associated with each inference.

Inference from sampled to unsampled decision units

In some cases the amount of material in the Decision Unit is very small compared to the total amount of material under investigation. In other words, there are many, many Decision Units; so many, in fact, they cannot all be sampled. Even if all the DUs could be sampled, it may be desired not to sample all of them since there would be no Decision Units left for consumption! If every can of tuna fish was tested for mercury or every nut tested for aflatoxin, there would be no canned tuna or nuts left to eat. This type of sampling is actually common, not only in food but in other industries as well. It is commonly known as attribute (or acceptance) sampling[9]. This is the type of sampling used in surveys and quality control. The premise is that if enough Decision Units are sampled, claims can be made about all the Decision Units (especially those not sampled). The claim made is typically based on the percent (or portion) of individual Decision Units that have some specific characteristic or attribute. This characteristic or attribute can also be concentration related as in the case of detection limits.

Survey example to illustrate concepts

Many companies and governments survey (or poll) to determine the percentage of the population that has some opinion, belief, owns a product, etc. For many of these surveys, only several hundred to several thousand people are contacted. The percent of people contacted that have the opinion, belief, product, etc. is used to make an inference to a larger number of people, which can be millions or billions. Surveys can be very accurate even though only a very small amount of people are actually surveyed. The only criteria to make inference from the surveyed people to all the people is that the surveyed people are selected at random (specific types of random are not addressed). The more people surveyed, the better the estimate of the true percentage of people that have that opinion, belief, product, etc. For this example, the individual is the Decision Unit. It is the individual that is "sampled" and information is obtained on the individual. This type of sampling is common and is applicable to food safety where the conditions for implementation are met.

In some cases there are multiple Decision Units, but they can all be sampled. There may be three trucks (Decision Units) of grain, and it is possible and desirable to sample all three, obtaining specific information on each truck. The amount of material taken for the samples is negligible compared to the total mass of the three trucks. However, if these three trucks each contained 5,000 packages (Decision Units), it may be impractical to collect 15,000 primary samples. Even if it was practical, there would be little material left.

Inference to unsampled Decision Units is a function of only the number of Decision Units sampled, not about mass, increments, tools, etc. as in the case of inference to a DU (where mass, increments, tools, etc. are critical for primary sample collection). Inference to unsampled Decision Units assumes the attribute, characteristic, or concentration of the Decision Unit is known (or can be known) and all that is required is random selection of enough Decision Units to meet a specified confidence. Inference to unsampled DUs is based solely on probability (the reason random selection of Decision Units is required). The importance of random selection of Decision Units cannot be over stressed. Since the entire inference scheme is based on randomness, no compromises can be made. Attitudes like "This looks random to me," and "I can't get to those Decision Units so I will skip them" are unacceptable.

A special case of attribute sampling exists when the desire is to claim the absence of a particular attribute or characteristic. While it is impossible to determine for certain that no DUs have a specified attribute or characteristic, if enough DUs are sampled and the characteristic or attribute is absent from each DU, an inference can be made that there is a XX% confidence that no more than YY% of the DUs have a particular characteristic or attribute. The details of the calculations are addressed in most introductory texts on statistics[9-12].

Example

Lima beans are sold in a variety of packaging including frozen, bagged, bulk, and canned. Two packaging examples, one frozen and the other bulk, will be considered to illustrate the two different types of inference.

Frozen lima beans

Lima beans can be sold in frozen packages. For some reason (e.g., routine surveillance, customer complaint) it was decided to test frozen packages of lima beans to see if a certain contaminant is present above a specified detection limit in any of the packages of lima beans. If any of the packages contains a detectable concentration of the contaminant, one course of action will follow. If none of the packages contain a detectable concentration of contamination, another course of action (no action) will follow. In this case the individual package would be the Decision Unit. It would be easy (and desirable) to select an entire package (DU) as the primary sample and send it to the laboratory. This is a perfect primary sample as no sampling error exists (as long as the sample integrity is maintained). The laboratory, however, cannot analyze the entire primary sample. Instead, the laboratory will have to process the primary sample and remove a small portion (subsampling) for analysis. The act of sample processing and subsampling will contain some error. An inference will have to be made from the analytical result to the primary sample. This inference may be performed with just one analysis (direct), or there could be multiple analysis and some type of statistical calculation could be used for inference. These details would be addressed during the SQC process. For this example direct inference will be used.

The obvious next question is which packages of lima beans are of concern. Just one package, all the packages at the local grocery store, all the packages in the warehouse, all the packages in Europe or something else? From a sampling and inference point of view, it does not matter (as long as random selection is achieved). For this example the choice will be a specific warehouse at a specific point in time. In the case of surveillance sampling or exposure assessment, the packages of lima beans could be sampled over the course of a year or some other time frame.

There will be two types of inference in this example: one will be from the analytical result to the package of lima beans and one will be from sampled packages of lima beans in the warehouse to all the packages of lima beans in the warehouse. The quality of the inference to the package is a function of the error in the sample processing and subsampling. The quality of the inference to all the packages in the warehouse is a function of how many packages (DUs) are sampled. There is no set number for quality. It should be a function of the consequences of an incorrect inference (and resulting incorrect decision). This would be addressed in the SQC process.

It is important to understand these inferences and their impact on the sampling protocol. For instance, the laboratory may receive 300 packages of lima beans and decide to combine them in groups of ten and only perform 30 analysis to save money. If this happened, information would be lost on the individual Decision Units and it would be impossible to determine a course of action.

Bulk lima beans

This example is the same as above except the lima beans are in 10 kg bulk containers (Decision Unit). In this case the entire DU cannot be taken as a primary sample, so the DUs (individual 10 kg bulk containers) will have to be sampled and an inference made from the primary sample back to the DU. In this example there are many DUs (more than can be sampled) and information is required on all the DUs, therefore another inference must be made from the sampled DUs to the unsampled DUs. In other words several, but not all, of the 10 kg packages will be sampled using the principles of TOS. The results from the sampled DUs will be used to infer (estimate) the percent of all the DUs in the warehouse that have a detectable concentration of the specified contaminant. If none of the 10 kg packages have a detectable concentration, one course of action will follow, and if any of the 10 kg packages have a detectable concentration, then another course of action will follow.

As in the frozen package example, understanding of these inferences is critical for developing the sampling protocol. It would be incorrect to select increments from different bulk containers and combine them into a primary sample because information will be lost on the individual bulk containers. For bulk containers, an overall error for both primary sampling and for the sample processing/subsampling in the laboratory need to be established.

Issues

In many cases a single sample can be used to represent a Decision Unit. This is always desirable. However, in some cases it may require multiple samples. If the desire is to estimate the exposure risk from pesticides on tomatoes to all individuals in a country, one could theoretically collect a single sample from tomatoes in time and space (across the entire country for a 30 year period), but this could never happen. In a situation such as this multiple samples of tomatoes within the Decision Unit (entire country for 30 years) would have to be collected.

It is typically not appropriate to combine increments across Decision Units as this will dilute the concentration of the individual Decision Units. There is, however, an exception to this. It is acceptable to combine multiple increments (from different Decision Units) for analytical efficiency (a composite sample) as long as information regarding the individual Decision Units is not lost. A common example is the presence of some prohibited attribute or characteristic that can be detected/measured (analyte is not diluted out) in the composite sample. In this case composting (as described above) is a viable strategy to reduce analytical cost and still achieve the objectives.

Inference to unsampled Decision Units can be made using attributes and concentrations. There are many, many statistical approaches to estimate both attributes and concentrations that are not addressed in this paper. The purpose of this paper is to identify the types of inferences and how they are used, not how the inferences are calculated.

In some cases the average of the Decision Units is calculated for decision-making purposes. In this case, the Decision Unit was incorrectly chosen. There should have been only one Decision Unit that contained all of the material. While it could be argued that the same average result is achieved, it would be more cost effective to treat all the material as one Decision Unit.

Conclusion

Knowledge of inference to Decision Units and to unsampled Decision Units is critical when applying the Theory of Sampling to food safety to make correct and defensible decisions. The sampling protocols for inference to a Decision Unit and to unsampled Decision Units are very different. Inference within a Decision Unit is based on the sampling errors incurred, sample processing and analysis. This error is mitigated and controlled through correct application of the principles of TOS. Confidence is indirectly related to the total sampling plus analysis error. Inference to unsampled Decision Units is based on the number of Decision Units sampled. This number is based on the probability of finding all the Decision Units that possess or lack a specific attribute or characteristic. Confidence in this case is directly related to the number of Decision Units sampled.

As TOS becomes more widely adopted in the food industry, it is imperative that practitioners understand and apply the principles of inference correctly in the development of sampling protocols. This is critical to ensure that defensible and cost effective decisions are made regarding food safety.

References

1. Walsh, John, *Vladimir Putin employs a full-time food taster to ensure his meals aren't poisoned*, The Independent, July 23, 2014.
2. Luthern, Ashley, *Testing for Poison Still a Profession for Some*, Smithsonian.com, June 26, 2009.
3. Ross-Nazzal, Jennifer, *Farm to Fork: How Space Food Standards Impacted the Food Industry and Changed Food Safety Standards*, In Dick, Steven J. and Launius, Roger D. (eds.), Societal Impact of Spaceflight, US Government Printing Office, 2007.
4. Ramsey, Charles A., The decision unit—a lot with objectives, in Proceedings of the 7th *International Conference on Sampling and Blending*, *TOS forum* **Issue 5,** 31–34 (2015). doi: 10.1255/tosf.75
5. Ramsey, Charles A., *Considerations for Inference to Decision Units*, Journal of AOAC International, Vol. 98, No.2, 288-294, 2015.
6. Pitard, Francis F., *Pierre Gy's Sampling Theory and Sampling Practice* 2nd ed., CRC Press, 1993.
7. *DS 3077 (2013) Representative Sampling– Horizontal Standard*, Danish Standards. www.ds.dk
8. Ramsey, Charles A., Wagner, Claas, *Sample Quality Criteria*, Journal of AOAC International, Vol 98, No. 2, 265-268 (2015)
9. Schilling, Edward G., Neubauer, Dean V., *Acceptance Sampling in Quality Control*, CRC, 2009.
10. Montgomery, Douglas C., *Introduction to Statistical Quality Control,* 2nd ed., Wiley, 1991.
11. ANSI/ASQ Z1.4: *Sampling Procedures and Tables for Inspection by Attributes*, American Society for Quality.
12. Walpole, Ronald E., and Myers, Raymond H., *Probability and Statistics for Engineers and Scientists,* 3rd ed., 1985.