UNIVERSITÀ DEGLI STUDI FIRENZE

DOTTORATO DI RICERCA IN STATISTICA APPLICATA

CICLO XXIV

COORDINATORE Prof. Fabio Corradi

EVALUATION OF ALTERNATIVE SAMPLING SCHEMES FOR THE DETECTION OF ENTEROBACTER SAKAZAKII IN THE PRODUCTION OF POWDERED INFANT FORMULAE

Settore Scientifico Disciplinare SECS/01

Dottorando
Dott.ssa Biagi Giulia

Tutor
Prof. Bertaccini Bruno

Coordinatore
Prof. Corradi Fabio

Anni 2009/2013

Introduction

It is internationally recognized that breast milk is the best source of nutrition for infants, both from a hygiene and completeness of nutrient point of view. Just before weaning and also immediately after, however, infants need to receive, in addition to breast milk, also supplementary food products to meet their growing nutritional needs. Furthermore, there are instances where breast milk may be insufficient or not available and thus, may need to be supplemented or replaced. In those instances, nutrition consists of infant formula (liquid and powder), prepared according to internationally recognized standards.

The use of powdered milk requires a proper knowledge of the correct methods of preparation and hygiene risks that can arise from improper handling and storage, both at home and in the hospital. Indeed, unlike liquid infant formula, which is sterilized by technological treatments applied before commercialization, the powdered formulations have a residual microbial flora generally composed of saprophytic bacteria and Enterobacteriaceae.

Cow's milk powder is not a sterile product and, once rehydrated, is a good breeding ground for microorganisms. The technology of production of infant formula basically follows the production of dried milk with spray-drying system,

except for slight changes the addition of nutritional complements, usually mixed with milk before heat recovery.

The presence of these microorganisms in milk powder may occur in two ways:

- contamination through the use of ingredients not subjected to heat treatment during the manufacturing process;
- contamination during the so-called "dry" process passages, such as contamination post heat treatment, typically during dehydration or packaging of the product.

Assuming that the heat treatment that the milk undergoes during the process accomplishes a perfect recovery of the product from a microbiological point of view, the contamination will probably occur at some point between the spray-drying and the packaging processes. Therefore, to prevent bacterial contamination of the finished product, it is critical the control of microorganisms of drying, post-drying and pre-packaging environments.

Both FAO and WHO considered cases of illnesses in infants associated with powdered infant formula (PIF) consumption (FAO/WHO, 2004). In particular, a bacterium belonging to the genus Enterobacteriaceae has become more and more relevant: the *Enterobacter Sakazakii* (ES).

Recently, the ICMSF (*International Commission on Microbiological Specifications for Foods*) has described the Enterobacter Sakazakii as "a serious threat to particular categories of people, for which the bacterium could compromise life or at least it could significantly impair the quality of life because of the after-effects, even in the long term, due to infection". Therefore, this bacterium has been included among the most feared food-borne pathogens, like Listeria monocytogenes, Clostridium botulinum type A and B and Cryptosporidium parvum.

The FAO/WHO expert meetings have identified all infants (< 12 months of age) as the population at particular risk for E. Sakazakii (Cronobacter species) infections. Among this group, those at greatest risk are infants less than 2 months of age, particularly neonates (< 28 days), particularly pre-term, low-birthweight (< 2500 g), and immunocompromised infants, and those less than 2 months of

age. The microorganism is easily able to overcome the gastric barrier and cause infection because the pH of the stomach of infants is less acid than that of adults.

While the incidence of these E. Sakazakii (Cronobacter species) infections in infants appears to be low, the consequences can be severe. Reported fatality rates vary considerably with rates as high as 50 percent reported in at least one outbreak. In addition, a portion of surviving infants has permanent disabilities such as retardation and other neurological conditions.

Among the infections caused by *E. Sakazakii* meningitis is the most severe form. In the advanced stages of the disease, the pathological manifestations most commonly observed are: ventriculites, brain abscesses, cysts, hydrocephalus, quadriplegia, delayed mental development and cerebral infarction (Bar-Oz, Preminger, Peleg, Block, & Arad, 2001) (Lai, 2001). Another important pathological neonatal manifestation caused by *ES* is the necrotizing enterocolitis (NEC). The prerequisites involved in the pathogenesis of this disease seem to be: neonatal intestinal ischemia, microbial colonization of the intestine and an increase of the protein substrate present in the intestine as can be derived from the assumption of milk formulae (Lucas & Cole, 1990). The NEC is characterized by necrosis and pneumatosis intestinalis and is the most common gastro-enteric disease of the newborn, with a mortality rate ranging 10-55% (Peter, et al., 1999).

In industrialized countries sixty cases of infection and several deaths caused by ES infection among infants fed with milk powder have been reported. These cases have always led to the recall of the product from the market. In Italy, until now, have not been reported cases of infection from *E. Sakazakii*. However, the problem is so important that the National Institute of Health has recently published a report inherent the microbiological hazards associated with the feeding of the newborn (Fiore, Casale, & Aureli, 2004).

The International Commission on Microbiological Specifications for Foods (ICMSF) provided in 1974 urgently needed guidance on the use of sampling plans and Microbiological Criteria (MC) for foods in international trade (ICMSF, 1974). In particular, ICMSF proposed to use attribute sampling plans; these kind of plans are of two types: a two-class plan used to classify the test samples as “accepta-

ble" or "defective" and a three-class plan used to classify the test samples as "acceptable", "marginally acceptable" or "defective".

These sampling plans have been widely adopted by public and private parties. They have been incorporated into specifications in commercial trading contracts and have been included in food law in different countries (Legan, 2001). Assessing the compliance of a batch to a MC depends on the criteria of the sampling plans. In general, two-class sampling plans are used when the health hazard is severe and direct (ICMSF, 1974) (ICMSF, 2002) (Legan, 2001).

The EU (art. 4 CE n° 852/2004) impose the operators of the food industry to respect specific microbiological criteria. The EU n° 2073/2005 regulation on microbiological criteria applicable to food products requires that the search of Enterobacter Sakazakii has to be performed in "in dried infant formulae and dried dietary foods for special medical purposes intended for infants below 6 months of age ". To this end, 30 sample units per lot must be analyzed during the period of shelf life of the product placed on the market. The results are considered satisfactory, if all the values observed indicate the absence of the bacterium and unsatisfactory, if the presence of the bacterium is detected in any of the sample units. To assess the hygiene of the production process the search of Enterobacteriaceae is performed and it must be absent in 10 sample units of 10 g of product. The analysis is performed at the end of the manufacturing process. If Enterobacteriaceae are detected in any of the sample units, the batch has to be tested for E. Sakazakii and Salmonella.

This dissertation is the outcome of a specific request made by a worldwide famous company, named "ACME" - we keep its name hidden for confidentially agreement with the management - that produces powdered infant formula. ACME asked us to assess if the sampling procedures they adopt internally to decide if a batch can be released on the market is better (and how much better) than the one required by Law for the detection of ES contamination.

In the first chapter, we present a short overview of Acceptance Sampling, given the importance that this type of quality control plays as an audit tool to en-

sure that the output of a process conforms to specific requirements. In the second chapter, we treat the problem of powdered infant formula and the way the microorganisms are physically distributed in foods. The third chapter is initially dedicated to the definition of Autosampling and Law Sampling procedures in the production of Powered Infant Formulae. The second part of the chapter discusses the ACME company, their production process and the three questions the ACME management submitted to a Control Quality Organization about the validity of their sampling procedures. The chapter ends with a description of the data provided by ACME to answer the three mentioned questions. Chapter four is dedicated to the reconstruction of the analysis performed by the Control Quality Organization that has been serving ACME for many years, analysis that has been deemed unsatisfactory by the ACME management. With the results of the analysis conducted on the basis of the procedures normally adopted in the literature on the subject, it was not possible to answer the questions addressed by the management. For this reason, in chapter five we propose a completely different and innovative approach to analyze the problem. The work ends with some proposals to the management on what to do in the future to monitor the performance of their PIF production process and to give better answers to their questions.

1 - Lot-by-Lot Acceptance Sampling

1. Lot-by-Lot Acceptance Sampling for Attributes

Acceptance sampling is an important field of statistical quality control that was popularized by Dodge and Romig[1] and originally applied by the U.S. military to the testing of bullets during World War II. If every bullet was tested in advance, no bullets would be left to ship. If, on the other hand, none were tested, malfunctions might occur in the field of battle, with potentially disastrous results. Dodge reasoned that a sample should be picked at random from the lot, and on the basis of information that was yielded by the sample, a decision should be made regarding the disposition of the lot. In general, the decision is either to accept or reject the lot. This process is called Lot Acceptance Sampling or just Ac-

[1] H. F. Dodge and H. G. Romig developed a set of sampling inspection tables for lot-by-lot inspection of product by attributes using two types of sampling plans: plans for lot tolerance percent defective (LTPD) protection and plans that provide a specified average outgoing quality limit (AOQL). For each of these approaches to sampling plan design, there are tables for single- and double-sampling.

ceptance Sampling. Acceptance sampling is concerned with inspection and decision making regarding products, one of the oldest aspects of quality assurance.

Three aspects of sampling are important (Montgomery, 2009):

1. It is the purpose of acceptance sampling to sentence lots, not to estimate the lot quality. Most acceptance-sampling plans are not designed for estimation purposes;
2. Acceptance-sampling plans do not provide any *direct* form of quality control. Acceptance sampling simply accepts and rejects lots. Even if all lots are of the same quality, sampling will accept some lots and reject others, the accepted lots being no better than the rejected ones. Process controls are used to control and systematically improve quality, but acceptance sampling is not;
3. The most effective use of acceptance sampling is *not* to “inspect quality into the product” but rather as an audit tool to ensure that the output of a process conforms to requirements.

Acceptance sampling is "the middle of the road" approach between no inspection and 100% inspection and is useful in several situation as when testing is destructive, when the cost of 100% inspection is too high; when scheduling production is compromised from 100% inspection etc. it is important to underline that the main purpose of acceptance sampling does not concern with the estimation of the quality of the lot but only with the decision whether or not the lot is likely to be acceptable

Advantages	Disadvantages
- It is usually less expensive because there is less inspection; - There is less handling of the product, hence reduced damage; - It is applicable to destructive testing; - Fewer personnel are involved in inspection activities; - It often greatly reduces the amount of inspection error; - The rejection of entire lots as opposed to the simple return of defectives often provides a stronger motivation to the supplier for quality improvements.	- There are risks of accepting "bad" lots and rejecting "good" lots; - Less information is usually generated about the product or about the process that manufactured the product; - Acceptance sampling requires planning and documentation of the acceptance-sampling procedure whereas 100% inspection does not.

1.1 Types of Sampling Plans

There are two major classifications of lot acceptance sampling plans (LASP): *by attributes* and *by variables*. The attribute case (attributes are quality characteristics that are expressed on a "go, no-go" basis) is the most common for acceptance sampling.

In a *single-sampling plans* one sample of items is selected at random from a lot and the disposition of the lot is determined from the resulting information. These plans are usually denoted as (*n,c*) plans for a sample size *n,* where the lot is rejected if there are more than *c* defectives.

In a *double-sampling plan*, following an initial sample, a decision based on the information in that sample is: (1) accept the lot, (2) reject the lot, or (3) no decision. If the outcome is (3) a second sample is taken and the procedure is to combine the results of both samples and make a final decision based on that information.

A *multiple sampling plans* is an extension of the double sampling plans where more than two samples are needed to reach a conclusion. The advantage of multiple sampling is smaller sample sizes.

A *sequential sampling plans* is the ultimate extension of multiple sampling where items are selected from a lot one at a time and after inspection of each item a decision is made to accept or reject the lot or select another unit.

All these plans will be analyzed in depth in the following paragraph.

The units selected for inspection from the lot should be chosen at random, and they should be representative of all the items in the lot. The random-sampling concept is essential in acceptance sampling. if random samples are not used, bias is introduced. The technique often suggested for drawing a random sample is to first assign a number to each item in the lot. Then *n* random numbers are drawn, where the range of these numbers is from 1 to the maximum number of units in the lot. This sequence of random numbers determines which units in the lot will constitute the sample. If products have serial or other code numbers, these numbers can be used to avoid the process of actually assigning numbers to each unit. In situations where we cannot assign a number to each unit, utilize serial or code numbers, or randomly determine the location of the sample unit, some other technique must be employed to ensure that the sample is. Sometimes the inspector may *stratify* the lot. This consists of dividing the lot into strata or layers and then subdividing each strata into aliquots. Units are then selected from within each aliquot.

1.2 Lot Acceptance Sampling Plan Properties

The choice of the type of plan depends on the characteristics and properties that the plan should have. An acceptance sampling plan can be characterized by:

- *Acceptable Quality Level (AQL)*: the AQL represents the poorest level of quality for the supplier's process that the consumer would consider to be acceptable as a process average. The producer would like to design a sampling plan such that there is a high probability of accepting a lot that has a defect level less than or equal to the AQL;

- *Lot Tolerance Percent Defective (LTPD)*: the LTPD is a designated high defect level that would be unacceptable to the consumer. The consumer would like the sampling plan to have a low probability of accepting a lot with a defect level as high as the LTPD;
- *Type I Error (Producer's Risk)*: this is the probability, for a given sampling plan, of rejecting a lot that has a defect level less than or equal to the AQL. The producer suffers when this occurs, because a lot with acceptable quality was rejected. The symbol α is commonly used for the Type I error and typical values for α range from 0.2 to 0.01;
- *Type II Error (Consumer's Risk)*: this is the probability, for a given sampling plan, of accepting a lot with a defect level greater than or equal to the LTPD. The consumer suffers when this occurs, because a lot with unacceptable quality was accepted. The symbol β is commonly used for the Type II error and typical values range from 0.2 to 0.01;
- *Operating Characteristic (OC) Curve*: this curve plots the probability of accepting the lot (Y-axis) versus the lot fraction or percent defectives (X-axis). The OC curve is the primary tool for displaying and investigating the discriminatory power of a LASP;
- *Average Outgoing Quality (AOQ)*: a common procedure, when sampling and testing is non-destructive, is to inspect 100% of the rejected lots and replace all defectives with good units. In this case, all rejected lots are made perfect and the only defects left are those in lots that were accepted. AOQ's refer to the long term defect level for this combined LASP and 100% inspection of rejected lots process. If all lots come in with a defect level of exactly *p*, and the OC curve for the chosen LASP indicates a probability p_a of accepting such a lot, over the long run the AOQ can easily be shown to be:

$$AOQ = \frac{p_a \cdot p\left(N-n\right)+\left(1-p_a\right)\cdot 0}{N} = \frac{p_a \cdot p\left(N-n\right)}{N}$$

where N is the lot size;

- *Average Outgoing Quality Level (AOQL)*: a plot of the AOQ (Y-axis) versus the incoming lot p (X-axis) will start at 0 for p = 0, and return to 0 for p = 1 (where every lot is 100% inspected and rectified). In between, it will rise to a maximum. This maximum, which is the worst possible long term AOQ, is called the AOQL.
- *Average Total Inspection (ATI)*: when rejected lots are 100% inspected, it is easy to calculate the ATI if lots come consistently with a defect level of p. For a LASP with a probability pa of accepting a lot with defect level p, we have

  $$ATI = n + (1 - p_a)(N - n)$$

  where N is the lot size.
- *Average Sample Number (ASN)*: for a single sampling LASP (*n*, *c*) we know each and every lot has a sample of size n taken and inspected or tested. For double, multiple and sequential LASP's, the amount of sampling varies depending on the number of defects observed. For any given double, multiple or sequential plan, a long term ASN can be calculated assuming all lots come in with a defect level of p. A plot of the ASN, versus the incoming defect level p, describes the sampling efficiency of a given LASP scheme.

Making a final choice between single or multiple sampling plans that have acceptable properties is a matter of deciding whether the average sampling savings gained by the various multiple sampling plans justifies the additional complexity of these plans and the uncertainty of not knowing how much sampling and inspection will be done on a day-by-day basis.

1.3 Single-Sampling Plan for Attributes

A single-sampling plan is defined by the sample size *n* and the acceptance number *c* (where a lot of size *N* has been submitted for inspection). Since the quality characteristic inspected is an attribute, each unit in the sample is judged to be either conforming or nonconforming. From a lot of size *N*, a random sample of *n* units is inspected and the number of nonconforming or defective items *d* ob-

served. If the number of observed defectives *d* is less than or equal to *c*, the lot will be accepted. If the number of observed defectives *d* is greater than *c*, the lot will be rejected. Generally, a unit that is nonconforming to specifications on one or more attributes is said to be a defective unit. When we refer to this procedure we talk about a single-sampling plan because the decision about the rejection or acceptance of the lot is based on the information contained in one sample of size *n*.

One way of picking (*n*,*c*) is to specify two desired points on the OC curve and solve for the (*n*,*c*) that uniquely determines an OC curve going through these points. The *operating characteristic* (OC) *curve* is one important measure of a LASP performance**.** It shows the probability that a lot submitted with a certain fraction defective will be either accepted or rejected.

Suppose that the lot size *N* is large (theoretically infinite). Under this condition, the distribution of the number of defectives *d* in a random sample of *n* items is binomial with parameters *n* and *p*, where *p* is the fraction of defective items in the lot. An equivalent way to conceptualize this is to draw lots of *N* items at random from a theoretically infinite process, and then to draw random samples of *n* from these lots. This is the same way of sampling directly from the process. The probability of observing exactly *d* defectives is

$$P\{d \text{ defectives}\} = f(d) = \frac{n!}{d!(n-d)!} p^d (1-p)^{n-d}$$

The probability of acceptance is simply the probability that *d* is less than or equal to *c*, or

$$P_a = P\{d \le c\} = \sum_{d=0}^{c} \frac{n!}{d!(n-d)!} p^d (1-p)^{n-d} \qquad (1)$$

The OC curve is developed by evaluating equation (1) for various values of *p* and shows the *discriminatory power* of the sampling plan.

Figure 1.1 - OC curve of the single-sampling plan *n* = 89, *c* = 2

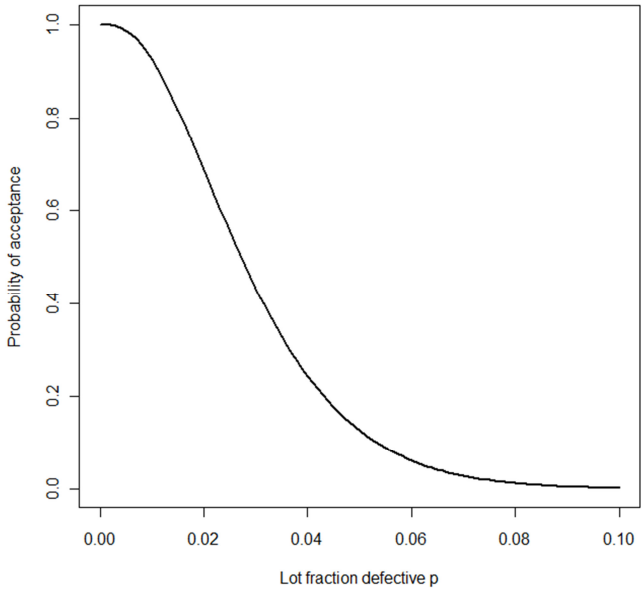


Figure 1.2 - Average outgoing quality curve for *n* = 89, *c* = 2

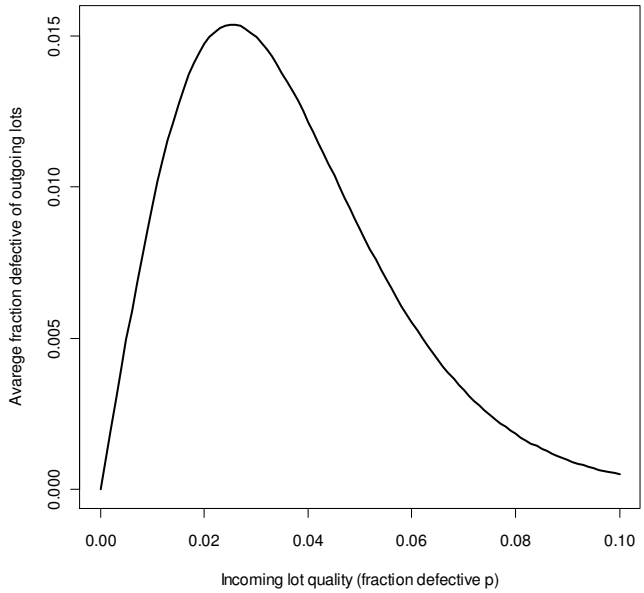


Figure 1.1 and figure 1.2 show, respectively, the OC curve and the AOQ curve for the sampling plan $n = 89$, $c = 2$.

A sampling plan that discriminates perfectly between good and bad lots would have a degenerative OC curve that runs horizontally at a probability of acceptance $P_a = 1.00$ until a level of lot quality that is considered "bad" is reached; from that point the curve drops vertically to a probability of acceptance $P_a = 0.00$, and then the curve runs horizontally again for all lot fraction defectives greater than the undesirable level. If such a sampling plan could be employed, all lots of "bad" quality would be rejected, and all lots of "good" quality would be accepted.

Unfortunately, this kind of ideal OC curve is almost never be obtained in practice. It could be realized by 100% inspection, if the inspection were error-free. The ideal OC curve shape can be approached, however, by increasing the sample size. Thus, the precision with which a sampling plan differentiates between good and bad lots increases with the size of the sample. The slope of the OC curve indicates the discriminatory power (Figure 1.3).

In figure 1.4 is shown how the OC curve behaves to the change of the acceptance number. Generally, changing the acceptance number does not dramatically change the slope of the OC curve; the OC curve is shifted to the left or to the right. Plans with smaller values of c provide discrimination at lower levels of lot fraction defective than do plans with larger values of c.

Figure 1.3 - OC curves for different sample sizes

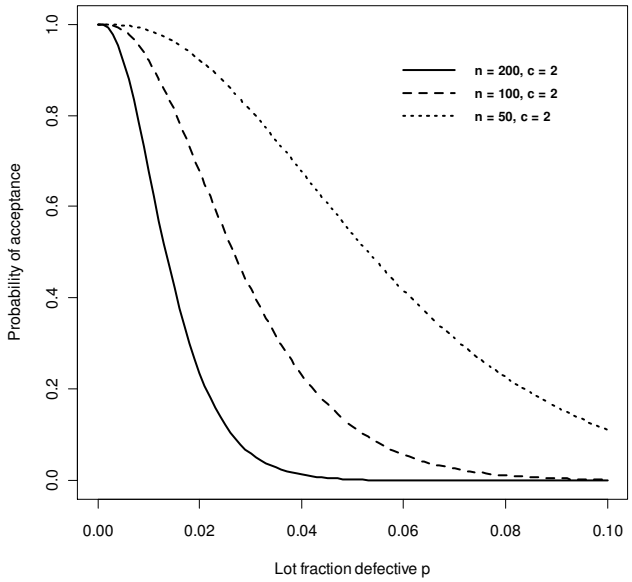
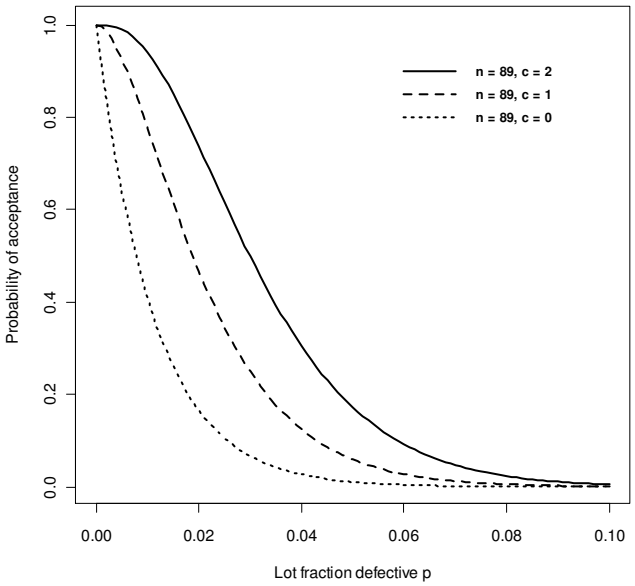


Figure 1.4 - The effect of changing the acceptance number on the OC curves



In order to design a sampling plan with a specified OC curve one needs two designated points. Suppose to design a sampling plan such that the probability of acceptance is 1-α for lots with fraction defective p_1 and the probability of acceptance is β for lots with fraction defective p_2. Assuming that the binomial is

appropriate, then the sample size n and the acceptance number *c* are the solution to

$$1-\alpha = \sum_{d=0}^{c} \frac{n!}{d!(n-d)!} p_1^d \left(1-p_1\right)^{n-d}$$

$$\beta = \sum_{d=0}^{c} \frac{n!}{d!(n-d)!} p_2^d \left(1-p_2\right)^{n-d}$$

These two simultaneous equations are nonlinear so there is no simple, direct solution. There are however a number of iterative techniques available that give approximate solutions.

Typical choices for these points are: p_1 is the AQL, p_2 is the LTPD and α, β are the Producer's Risk (Type I error) and Consumer's Risk (Type II error), respectively.

1.4 Double, Multiple and Sequential Sampling Plans

Double and multiple sampling plans were considered when a lot is questionable to give to this lot another chance.

If in double-sampling the results of the first sample are not conclusive with regard to accepting or rejecting, a second sample is taken. Application of double sampling requires four parameters:

n_1 = sample size on the first sample

c_1 = acceptance number of the first sample

n_2 = sample size on the second sample

c_2 = acceptance number for both sample

A random sample of n_1 items is selected from the lot, and the number of defectives in the sample, d_1, is observed. If $d_1 \le c_1$ the lot is accepted on the first sample. If $d_1 > c_2$ the lot is rejected on the first sample. If $c_1 < d_1 \le c_2$, a second random sample of size n_2 is drawn from the lot, and the number of defectives in this second sample, d_2, is observed. Now the combined number of observed defectives from both the first and second sample, $d_1 + d_2$, is used to determine the lot sentence. If $d_1 + d_2 \le c_2$, the lot is accepted. However, if $d_1 + d_2 > c_2$, the lot is rejected.

Respect to single-sampling, the double-sampling plan may reduce the total amount of required inspection. Suppose that the first sample taken under a double-sampling plan is smaller than the sample that would be required using a single-sampling plan that offers the consumer the same protection: the cost of inspection will be lower for double-sampling than it would be for single-sampling . It is also possible to reject a lot without complete inspection of the second sample (this is called *curtailment* on the second sample) (Montgomery, 2009). Consequently, the use of double-sampling can often result in lower total inspection costs. Furthermore, in some situations, a double sampling plan has the psychological advantage of giving a lot a second chance.

The potential disadvantages are: unless curtailment is used on the second sample, under some circumstances double-sampling may require more total inspection than would be required in a single-sampling plan that offers the same protection (thus, unless double-sampling is used carefully, its potential economic advantage may be lost); it is administratively more complex, which may increase the opportunity for the occurrence of inspection errors.

A multiple-sampling plan is an extension of double-sampling. In multiple-sampling plan more than two samples can be required to sentence a lot. If, at the completion of any n^{th} stage of sampling, the number of defective items is less than or equal to the acceptance number, the lot is accepted. If, during any stage, the number of defective items equals or exceeds the rejection number, the lot is rejected; otherwise the next sample is taken. The procedure continues until the last sample is taken and the lot disposition decision is made. The first sample is usually inspected 100%, although subsequent samples are usually subject to curtailment. The construction of OC curves for multiple-sampling is a straightforward extension of the approach used in double-sampling.

The principal advantage of this kind of plans is the smaller samples required at each stage in comparison with the other plans seen above; thus, some economic efficiency is connected with the use of the procedure. However, multiple-sampling is much more complex to administer.

In sequential-sampling, we take a sequence of samples from the lot and allow the number of samples to be determined entirely by the results of the sampling process. In practice, sequential-sampling can theoretically continue indefinitely, until the lot is inspected 100%. In practice, sequential-sampling plans are usually truncated after the number inspected is equal to three times the number that would have been inspected using a corresponding single-sampling plan. If the sample size selected at each stage is greater than one, the process is usually called *group* sequential-sampling.

In particular, if the sample size inspected at each stage is one, the procedure is usually called *item-by-item sequential-sampling*.

Item-by-item sequential-sampling is based on the sequential probability ratio test (SPRT), developed by Wald (1947). The operation of an item-by-item sequential-sampling plan is illustrated in Figure 5. In the chart the cumulative observed number of defectives is plotted. For each point, the abscissa is the total number of items selected up to that time, and the ordinate is the total number of observed defectives. With this plan another sample must be drawn only If the plotted points stay within the boundaries of the acceptance and rejection lines,. The lot is rejected if a point falls on or above the upper line and is accepted if a sample point falls on or below the lower line.

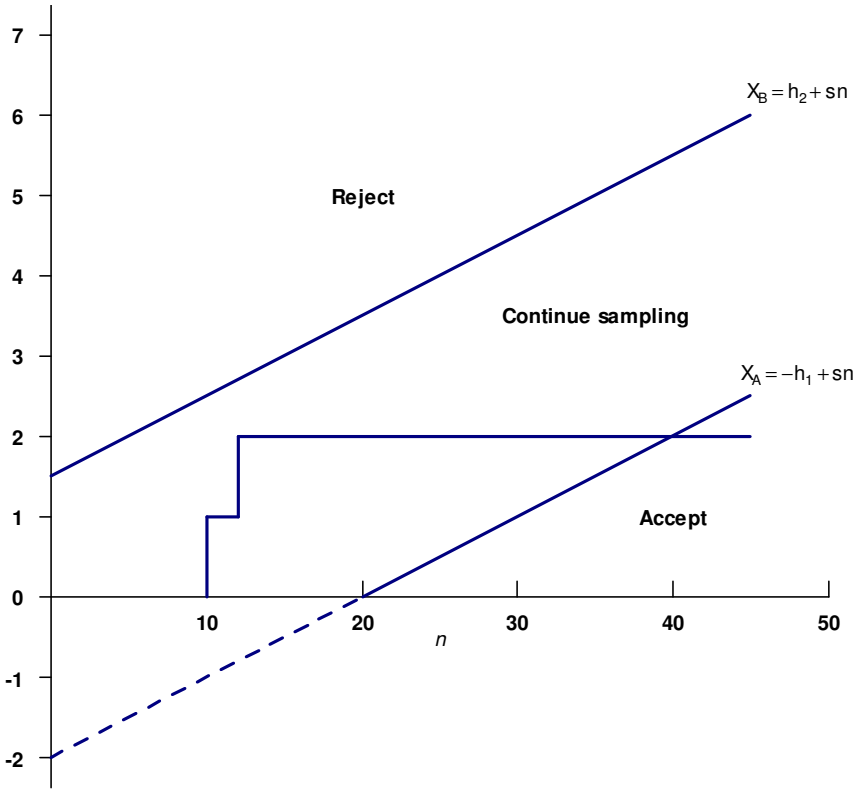


Figure 1.5 - Graphical performance of sequential sampling

The equations for the two limit lines for specified values of p_1, 1-α, p_2, and β are

$$X_A = -h_1 + sn \qquad \text{(acceptance line)}$$
$$X_R = h_2 + sn \qquad \text{(rejection line)}$$

where

$$h_1 = \left(\log \frac{1-\alpha}{\beta} \right) / k$$

$$h_2 = \left(\log \frac{1-\beta}{\alpha} \right) / k$$

$$k = \log \frac{p_2(1-\mathrm{p}_1)}{p_1(1-\mathrm{p}_2)}$$

$$s = \log \left[\frac{(1-p_1)}{(1-p_2)} \right] / k$$

The OC curve for sequential-sampling can be easily obtained. Two points on the curve are (p_1, 1-α) and (p_2, β). A third point, near the middle of the curve, is $p = s$ and $P_a = h_2/(h_1 + h_2)$.

2 – Enterobacteriaceae in powdered infant formula

2.1 Powdered formulae for infants and young children

It is internationally recognized that breast milk is the best source of nutrition for infants. The World Health Organization recommends that infants should be exclusively breast-fed for the first 6 months of life. However, there are instances where it may be insufficient or not available and thus it may need to be supplemented or replaced. In those instances, one of the dietary options is the use of powdered infant formulae (PIF). Infants who are not breast-fed should be provided with a suitable breast milk substitute, formulated in accordance with Codex Alimentarius Commission standards. To reduce the risk of infection in infants fed PIF, recommendations have been made for the preparation and storage of PIF (D. Drudy, 2006).

Breast milk substitutes are formulated to resemble the nutrient composition of breast milk and are a more satisfactory substitute for breast milk than any other product. Cow's milk, which is used in the production of dried-infant formula contains a higher amount of fat, protein and minerals than breast milk. Therefore, it is first skimmed and then diluted in order to achieve levels more commonly found in breast milk.

Care must be taken in the manufacture of infant formula to safeguard the microbiological quality of the finished product. The manufacture of dried-infant formula can follow two procedures (M. Nazarowec-White, 1997). The 'dry procedure' involves dry mixing of ingredients into powder after spray-drying. This method should be avoided, as it can lead to bacteriological contamination (Lambert-Legace, 1982). In the second manufacturing method, dried-infant formula is prepared using a 'wet procedure' where the following heat treatments are used: (1) liquid skim milk is heat-treated before processing at 82°C for 20 s; (2) the pre-mix consisting of skimmed milk and fat components is heat-treated at 80°C for 20 s; (3) the total mixture containing all ingredients is heat-treated at 107-110°C for 60 s; (4) the liquid mixture is concentrated using a falling film evaporator and (5) the concentrate is heat-treated again at 80°C and then immediately spray-dried. Often, a combined procedure is used where water soluble components are added to the milk before drying and the less soluble components are added to the blend after drying (Caric, 1993).

As all dehydrated products, it is not possible using current technology to produce powdered formulae that are devoid of low levels of microorganisms, i.e., the products cannot be sterilized. Thus, their microbiological safety requires strict adherence to good hygienic practices during both manufacture and use.

Both FAO and WHO considered cases of illnesses in infants associated with PIF consumption either epidemiologically or microbiologically (FAO/WHO, 2004). They identified three categories of microorganisms based on the strength of evidence of a causal association between their presence in PIF and illness in infants (FAO/WHO, 2008):

- microorganisms with a clear evidence of causality, namely, *Salmonella enterica* and *Enterobacter Sakazakii*[2];

[2] Enterobacter Sakazakii is the microorganism whose name refers to the Japanese researcher Ricki Sakazakii for the great contribution he has given to the understanding of the biology of Enterobacteriaceae and Vibrionaceae

- microorganisms for which the causality is plausible but not yet demonstrated, i.e., they are well-established causes of illness in infants and have been found in PIF, but contaminated formula has not been convincingly shown, either epidemiologically or microbiologically, to be the vehicle and source of infection, e.g., other *Enterobacteriaceae*;
- microorganisms for which causality is less plausible or not yet demonstrated, including microorganisms, which despite causing illness in infants, have not been identified in PIF, or microorganisms which have been identified in PIF but have not been implicated as causing such illness in infants.

2.2 Enterobacteriaceae in powdered formulae for infants and young children

Salmonella is a well-known long-standing foodborne human pathogen. The incidence of salmonellosis among infants, originating from various sources, was reported to be more than eight times greater than the incidence across all ages in the United States of America (FAO/WHO, 2008). Infants are also more likely to experience severe illness or death from salmonellosis, and infants with immune compromising conditions are particularly vulnerable. It is unclear whether the increased incidence of salmonellosis among infants results from greater susceptibility, or whether infants are more likely than persons in other age groups to seek medical care or have stool cultures performed for symptoms of salmonellosis. At least 6 reported outbreaks of salmonellosis involving approximately 287 infants have been associated with PIF between 1985 and 2005. Most of these outbreaks involved unusual *Salmonella* serotypes, which likely aided in recognition of those outbreaks. It is recognized that outbreaks and sporadic cases of salmonellosis due to powdered infant formula are likely to be under-reported (FAO/WHO, 2006).

Enterobacter Sakazakii is a member of the family Enterobacteriaceae, genus Enterobacter, and is a motile peritrichous, gram-negative bacillus (J.J. Farmer III, 1980). The organism, which was initially referred to as “yellow-pigmented cloacae,” was reclassified as *“E. Sakazakii”* in 1980 on the basis of differences in DNA-DNA hybridization, biochemical reactions, pigment production, and antibi-

otic susceptibility, compared with *Enterobacter cloacae* (J.J. Farmer III, 1980). *Enterobacter Sakazakii* (Cronobacter species) has recently emerged as a pathogen of infants. The FAO/WHO expert meetings have identified all infants (<12 months of age) as the population at particular risk for *E. Sakazakii* infections. Among this group, those at greatest risk are neonates (<28 days), particularly pre-term, low-birthweight (<2500 g), and immunocompromised infants, and those less than 2 months of age (FAO/WHO, 2006). *E. Sakazakii* was first implicated in a case of neonatal meningitis in 1958, when an outbreak in England resulted in the deaths of 2 infants. Since that time, there have been about 70 reported cases of *E. Sakazakii* infection (M. Nazarowec-White, 1997).

A listing of the reported cases and outbreaks of neonatal infections caused by *E. Sakazakii* found in the literature can be seen in Table 2.1. While the incidence of *E. Sakazakii* infections in infants appears to be low, the consequences can be severe. The primary manifestations of *E. Sakazakii* infection in infants, i.e., meningitis and bacteraemia, tend to vary with age. *E. Sakazakii* meningitis tends to develop in infants during the neonatal period, while *E. Sakazakii* bacteraemia tends to develop in premature infants outside of the neonatal period with most cases occurring in infants less than 2 months of age. However, infants with immunocompromising conditions have developed bacteraemia as late as 10 months of age and previously healthy infants have also developed invasive disease outside the neonatal period. Infections have occurred in both hospital and outpatient settings. It was noted that as older infants generally live at home in the community, infections in such infants may be more likely to be under-reported.

Table 2.1 - Sporadic cases and outbreaks of Enterobacter Sakazakii infection for which powdered infant formula (PIF) was implicated as the source agent (D. Drudy, 2006)

Location (year)	No. of cases	No. of deaths	Source	Reference(s)
England (1958)	2	2	Unknown	[8]
Denmark (1958)	1	1	Unknown	[9]
Georgia (1958)	1	0	Unknown	[10]
Oklahoma (1958)	1	1	Unknown	[11]
Indiana (1981)	1	0	Unknown	[12]
Denmark (1983)	8	6	Suspected PIF	[13]
Greece (1977–1981)	1	1	NS	[14]
Greece (1984)	11	4	Unknown	[15]
Missouri (1984)	1	0	Unknown	[16]
Massachusetts (1984)	2	1	Unknown	[17]
Iceland (1986–1987)	3	1	PIF[a]	[18]
Tennessee (1988)	4	0	PIF, blender	[19, 20]
Maryland (1990)	1	0	PIF, blender	[21]
Ohio (1990)	1	0	NS	[22]
Belgium (1998)	12	2	PIF	[23]
Israel (1999–2000)	2	0	PIF and blender	[5, 6]
Tennessee (2001)	10	1	PIF	[24]
Belgium (2002)	1	1	PIF	[25]
New Zealand (2004)	5	1	PIF	[25]
France (2004)	4	2	PIF	[26]

NOTE. Data are based on [20, 27, 28]. NS, not specified.

[a] One of the causal factors responsible may have been reconstituted formula was held at 35°C–37°C for lengthy periods.

Reported fatality rates of *E. Sakazakii* infections in infants vary considerably with rates as high as 50 percent reported in at least one outbreak. Mortality rates of 33% – 80% have been reported (Lai, 2001). *E. Sakazakii* infections are also associated with significant morbidity. Most children who survive *Enterobacter*-associated meningitis (94%) develop irreversible neurological sequelae resulting in quadriplegia, developmental impedance, and impaired sight and hearing (D. Drudy, 2006).

Although all known outbreaks have involved infants, sporadic cases have been reported in children and adults, however these have not been linked to PIF (FAO/WHO, 2004).

PIF is not a sterile product, and current Codex Alimentarius Commission specifications for PIF permit 1–10 coliform bacteria per gram of formula. It

should be noted that *E. Sakazakii* belongs to this group of organisms. Nevertheless, PIF manufacturers implement a policy of zero tolerance for both *Salmonella* and *Listeria* species in products. Current drafting of microbiological specifications for *E. Sakazakii* is under consideration by the International Committee for the Microbiological Safety of Food and the Codex Alimentarius Commission (D. Drudy, 2006).

There are four routes by which *E. Sakazakii* and *Salmonella* can enter PIF:

1. through the ingredients added in dry mixing operations during the manufacturing of PIF;
2. through contamination of the formula from the processing environment in the steps during or following the drying;
3. through contamination of the PIF after the package is opened;
4. through contamination during or after reconstitution by the caregiver prior to feeding.

E. Sakazakii may be found in many environments such as food factories, hospitals, institutions, day-care facilities and homes. In manufacturing, the organism may gain access to the processing line and product, since current technology cannot completely eliminate this organism from the manufacturing environment.

2.3 Mechanisms influencing spatial distributions of microorganisms

The distributions of how microorganisms are physically distributed in foods determine both the likelihood that a foodstuff will cause illness and the consequential public health burden, but not much is known about these, yet.

The spatial distribution of the microorganisms in foods determines the value of the data on prevalence and/or concentration, obtained through sampling and testing, for informing food safety management decision-making (e.g., for lot acceptance or for process control) and, ultimately, their value for determining the associated public health burden. Understanding spatial distributions of harmful microorganisms is important for establishing proper microbiological criteria

and obtaining a realistic view of the performance of the associated sampling plans.

Throughout the food processing, food products (from the raw material to the output) are exposed to a series of processes and related mechanisms that influence the level and spatial distribution of microorganisms. Usually there are six mechanisms (*contamination*, *microbial growth*, *microbial death*, *joining*, *mixing* and *fractionation*) that affects the final microbial distribution of a product (ILSI Europe, 2010).

Contamination is the transfer of microorganisms onto a foodstuff from an external source. The contamination of foodstuffs generally occurs on the surface of a product, and often results in an uneven spatial distribution of microorganisms. Contaminants are often coming from different sources (equipment and utensils, humans, water used for rinsing, cleaning and cooling, packaging materials *etc*.) but also the contact with contaminated surfaces, air or water can cause the contamination.

Once a food product has been contaminated, *microbial growth* can transform an initially homogeneous distribution into a more clustered distribution on or within a foodstuff. In contrast with contamination, which occurs on external surfaces, growth can cause the distribution of microorganisms inside the product. Often, during growth through reproduction, microbial cells form cell clumps or micro-colonies and this is due to particular growth characteristics of the microorganisms or to physical constraints of the food matrix. Different conditions of the product (for example cooling or thawing of the foodstuff) can cause microbial growth in different parts of the product and then an uneven distribution of microorganisms.

As well as some situations may lead to microbial growth, some of the same situations can cause *microbial death*. In fact, situations of cooling or thawing, or the application of lethal processes or the adverse effects of changing environmental conditions could lead to inhibition of microbial growth or (at lethal levels) even complete inactivation (death) of microbial cells.

Microbial death can result from the application of lethal processes (such as thermal processing or the addition of lethal levels of preservatives) or from the adverse effects of changing environmental conditions. Intrinsic product characteristics (e.g., water activity, pH and nutrient availability), and extrinsic product characteristics (e.g., storage temperature or storage atmosphere). In this situation however is very unlikely that all the microbial cells die, so the final distribution of microbial cells would be even more clustered.

Joining two or more materials (e.g., ingredients or food products), each with different microbial distributions, will result in a joined product with a distribution, which is different from the initial microbial populations of the merged materials, but a function of the way in which joining occurs.

When materials or product units are *mixed*, the original microbial population is relocated throughout the product mass. In this situation, that can be an active process or the results of spontaneous movements, the spatial distribution is likely more random because, in general, mixing will disperse the microbial populations.

Fractionation, like mixing, reallocates microorganisms over the resulting product units. Fractionation can also encompass procedures that may result in the removal of contaminating microorganisms, for instance when a portion of a food product is discarded or removed by peeling or rinsing.

It is easy to understand how each of the mechanisms described may have an impact on the spatial distribution of microorganisms in a food.

2.4 Stochastic distributions of microorganisms and clustering

It is unlikely that every food portion of a larger bulk contains the same number of microorganisms, so the simple average number of microorganisms per portion is not an adequate representation of microbial status (ILSI Europe, 2010). But which is the right sizes of the portions and batches of interest? This information is necessary before considering distributions that might be used to model portion-to-portion variation as well as overall average. The portion of interest differs from the perspective of public health or microbiological criteria (acceptance sampling plans).

On a very small scale, comparable to the size of a microorganism (perhaps 10^{-12} cm^3) there are only two kinds of portion, containing an organism or not, so that all possible distributions are clustered. Conversely, large portions can be expected to "average out" small scale clustering, but to reveal larger scale clustering, for example by production runs or production within a particular country. In principle, the presence of clustering can be defined, independently of scale, in terms of the probability of points (organisms) depending on the presence of nearby points. The distribution is deduced from, and its effect mediated by, numbers (or presence) in finite-sized samples.

From the perspective of public health, the portion of interest is that which is actually consumed and could determine the exposure of individual consumers. The batch of interest is that which might be the subject of a risk assessment or be responsible for an outbreak. In an industrial setting, usually ranges from one to hundreds of tons.

In the case of acceptance/rejection, the portion of interest is the amount analyzed, often smaller than the sample taken. The batch of interest is that subject to the acceptance/rejection decision, probably of the order of tons.

For this kind of analysis portions range from 0.1 g to 500 g within batches of tons.

The final distribution of microorganisms in a food is usually the result of multiple distinct mechanisms and is often a mixture of simpler distributions.

Physical or spatial distributions are different from, although related to, frequency distributions. The differences and relationships are illustrated in Figure 2.1, Figure 2.2 and Figure 2.3 (ILSI Europe, 2010). Three different spatial distribution of 100 points over 25 portion are represented in Figure 2.1; chart (a) shows points quite regularly spread, chart (b) represents a very clustered situation and chart (c) shows points randomly spread (the distributions are represented in two rather than three dimensions for illustrative purpose).

Figure 2.1 - Three different spatial distribution of 100 points over 25 portions

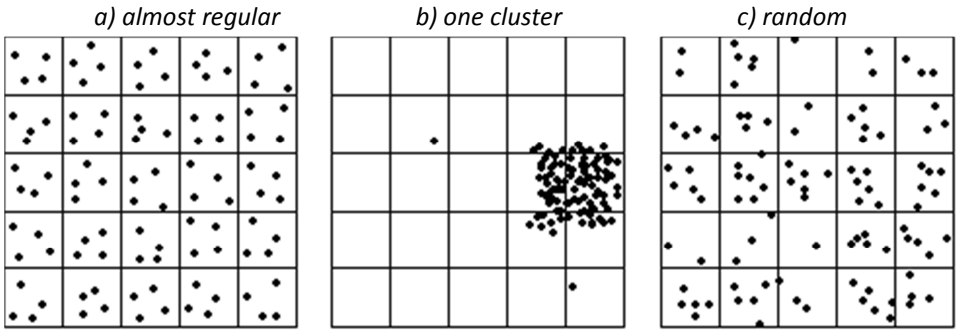


Figure 2.2 - Numbers of points in individual portions for the three spatial distributions depicted in Figure 2.1

a) almost regular

4	4	4	4	4
4	4	4	4	4
4	4	4	4	4
4	4	4	4	4
4	4	4	4	4

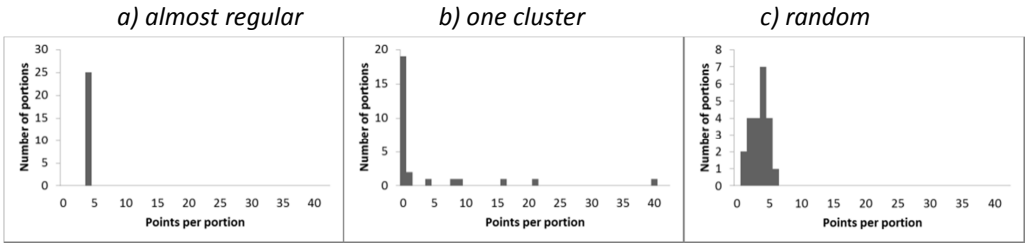
b) one cluster

0	0	0	0	0
0	1	0	4	9
0	0	0	21	40
0	0	0	8	16
0	0	0	0	1

c) random

2	5	1	2	3
4	4	2	5	3
5	7	6	6	6
2	3	1	5	5
5	5	3	6	4

Figure 2.3 - Frequency distributions for the three spatial distributions depicted in Figure 2.1 and Figure 2.2



In the food industry, each portion could be considered a 'unit' and the set of 25 portions a 'lot', so Figure 2.1 and Figure 2.2 represent 'within-a-unit' and 'within-a-lot' variation. Alternatively, each portion could be considered a lot so the figures

represent ‘within lot’ and ‘between lot’ variation. In Figure 2.2 it is shown the number of points belonging to each portion and Figure 2.3 shows the resulting frequency distributions. Figure 2.1 is the one that contains most information because it contains the spatial distributions of points and so the location of the points (not the values). This means that Figure 2.2 and Figure 2.3 can be deduced from Figure 2.1, but not vice versa. Figure 2.2 (spatial distributions of values) contains values (the concentrations in each portion) and locations (of portions, not of individual points). Figure 2.3 (frequency distributions) contains information on values, but no information on location. Thereof it follows that the same frequency distribution may come from different spatial distributions.

It is possible to characterize a spatial distributions by stating how the chance of finding a point depends upon the closeness of other points. So we have to analyze some of the terms often used in spatial distributions: ‘regular’, ‘clustered’ and ‘random’. *Regular* distributions (e.g., Figure 2.1a) are not so usual in food microbiology but they can occur where contamination follows more or less regular patterns: in this situation points are *less likely* close to other points, so that points are relatively far apart from each other. *Clustered* distributions (e.g., Figure 2.1b) are instead quite common in food microbiology because contamination often occurs in clusters, due to initial contaminants multiplying into micro-colonies, localized growth of microorganisms in non-liquid foods, etc. Points are here *more likely* close to other points, so that points are relatively close to each other. *Uniform random* distributions (e.g., Figure 2.1c) sometimes result from other patterns by perfect mixing. Points are *equally likely* close to or far from other points and the chance of finding a point is independent from the others points and the closeness to them. While the points in a random pattern are equally likely everywhere (so, the distribution of probability is uniform), they cannot actually be everywhere (so, the distribution of points is not uniform). Uniform random patterns are quite common in food microbiology, for instance in the case of well-mixed liquids or powders (ILSI Europe, 2010).

To describe spatial distributions in quantitative terms can be quite difficult; the statistics of ‘spatial processes’ is sophisticated. Several approaches could be used. For instance, the positions of the points could be described by their X-Y coordinates, or by the distances between neighboring points. One way of charac-

terizing spatial distributions is by stating how the chance of finding a point depends upon the closeness of other points.

What is the information available in the real situations of food industry? Data describing actual spatial positions of individual microorganisms (e.g., as in Figure 2.1) contains most information, as mentioned above, but such information is very rarely available. Data describing spatial positions of portions and their concentrations (e.g., as in Figure 2.2) contains some direct spatial information that can be converted to frequency distribution form. Also this kind of information is not common and where it is available, the concentration data is often presence/absence rather than counts. As often as not data are available only in frequency distributions form (e.g., as in Figure 2.3), without any spatial content. So the only information available is how often particular concentrations were observed. Again, the concentration data is often presence/absence rather than counts, so that histograms such as Figure 2.3 would have only two bars, 0 and >0.

When we use the word 'dispersed' we need to know that it can assume different and opposite meanings depending if we are describing spatial or frequency distributions. Indeed, the most spread out spatial distribution (Figure 2.1a) gives the smallest variation in points per cell, while the most compact spatial distribution (Figure 2.1b) gives the greatest variation and the intermediate spatial distribution (Figure 2.1c) gives an intermediate variation. So, "a more dispersed, less clustered, spatial distribution gives a less dispersed, more clustered, frequency distribution" and otherwise "a less dispersed, more clustered, spatial distribution gives a more dispersed, less clustered, frequency distribution" (ILSI Europe, 2010).

The degree of spatial clustering can often be assessed by comparing the variance and mean of the corresponding frequency distributions. Generally, a regular spatial distribution has a frequency distribution with variance smaller than its mean; a clustered spatial distribution has a frequency distribution with variance greater than its mean and a uniform random spatial distribution has a frequency distribution with variance equal to its mean.

Table 2.2 (ILSI Europe, 2010) - Relationship between spatial and frequency distributions

Spatial distribution (relative to uniform random)	**Frequency distribution (relative to Poisson)**			**Example**
more spaced	more concentrated	underdispersed	variance < mean	regular contamination due to contaminated filler head
uniform random	poisson		variance = mean	perfect mixing
more clustered	more right skewed	overdispersed	variance > mean	local contamination from hand contact

2.4.1 Frequency distributions modelling microorganisms in food

Distributions used to model frequency distributions of microorganisms should satisfy five criteria if they are used to represent or approximate spatial distributions in real situations:

i. *The model outcome should not be negative;*
ii. *The model should allow zero as an outcome;*
iii. *The model outcome should be discrete numbers only;*
iv. *The frequency distribution should reduce to, or at least approximate, the Poisson distribution;*
v. *The frequency distribution should be similar to, or approximate, the Lognormal distribution at high numbers of microorganisms (when there is negligible probability of zero microorganisms).*

The reasons why distributions have to satisfy these criteria are easy to understand. Indeed, it is not possible to have negative numbers of microorganisms in a food; it is possible to have no microorganisms in a portion of food; it is not possible to have parts of microorganisms in a portion as viable units. So the first three criteria can be satisfied considering frequency distribution will give zero probability to negative values; gives a finite probability to zero values and should not as-

sign probability to fractional numbers. For the last two criteria it can be shown that the Poisson distribution is the best distribution that models a uniform, random, spatial distribution (obtained throughout a perfect mixing) and although the frequency distribution of microorganisms must be discrete when we are talking about very high values the difference between consecutive integers is so small to be approximated by continuous distributions and the Lognormal distribution (described below) has been widely and successfully used to model microorganisms frequency distributions in many circumstances.

Factual insight into the actual spatial distribution of microorganisms in foods is lacking and often generalizing assumptions are made that have become commonplace in day-to-day food safety management. Understanding spatial distributions of (harmful) microorganisms is vital for establishing proper microbiological criteria and obtaining a realistic view of the performance of the associated sampling plans (Jongenburger, Bassett, Jackson, Zwietering, & Jewell, 2012). An assumption often used is that microorganisms are distributed lognormally, since this distribution appears to fit actual observations in foods (Kilsby & Baird-Parker, 1983), or according to the Poisson distribution. However, while there is some mechanistic support for the use of these two statistical distributions, irregular clustering of microorganisms, for example, will impact on the frequency distribution and needs to be considered as well, so also another frequency distribution (Poisson-LogNormal) is analyzed. There are also other two frequency distributions sometimes used to model microorganism, the Gamma and the Negative Binomial, that are not described here.

1. Poisson distribution (including generalized Poisson distributions)
2. Lognormal distribution
3. Poisson-Lognormal distribution (another type of generalized Poisson)

Poisson distribution (including generalized Poisson distributions)

The Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time and/or space if these events occur with a known average rate and independently of the time since the last event (Haight, 1967).

The location is enough to define a single-parameter Poisson frequency distribution. Its dispersion as measured by variances, is equal to the mean.

A Poisson frequency distribution models a uniform random spatial distribution, but it is also used in the absence of anything more appropriate (e.g., based on specific knowledge of the likely spatial distribution), even if a uniform random spatial distribution cannot be assumed. Problems concerning the use of Poisson distribution are related to the fact that this distribution is the correct choice for well-mixed products with low concentrations of microorganisms, but does not have the flexibility to model the variations in microbial concentrations seen in practice. Furthermore, when we consider high concentrations (e.g., above 20 CFU (colony forming units)/portion) a Poisson distribution is essentially symmetrical, while observed distributions of microbial concentrations are often skewed to the right. For these reasons, generalized Poisson distributions are more flexible.

The dispersion of a Poisson frequency distribution is measured by variance and it is equal to its mean; 'over-dispersed' distributions are the ones with variance less than the mean and reflect situations with clustering in the spatial distribution; 'under-dispersed' are the ones with variance greater than the mean, then reflects separation in the spatial distribution, that is situations more regular than a uniform random distribution (this kind of situation are however less common than over-dispersion in foods).

Poisson frequency distributions are commonly used to assess microbiological risk and to explain microbiological criteria. The degree of over- or under-dispersion (clustering or spacing) of a particular distribution can be compared to a Poisson distribution, throughout the ratio between the variance and the mean:

- ratio = 1 for uniform random spatial distributions,
- ratio > 1 for clustered spatial distributions, and

- ratio < 1 for over-spaced distributions.

A statistical test (Stoyan, 1994) for the presence of spatial clustering or over-spacing is based on the 'dispersion index', *I*:

$$I = \frac{ns^2}{\bar{x}}$$

where

n = is the number of portion

s^2 = is the variance of points in each portion (with *n*-1) in the denominator

$\bar{x}$ = is the mean number of points in each portion

For a set of concentrations taken from a Poisson distribution (e.g. where the spatial distribution is uniform random) s^2 is expected to be about equal to $\bar{x}$, so *I* is about equal to *n*. In fact, for such a sample, *I* is distributed according to a χ^2 distribution with *n*-1 degrees of freedom and if *n* is greater than 6 and $\bar{x}$ is greater than 1 then *I* can be tested against the χ^2 distribution (ILSI Europe, 2010). If the cumulative χ^2 probability is very small (e.g. less than 0.05) there is statistically significant evidence of over-spacing, and if it is very big (e.g. more than 0.95) there is statistically significant evidence of spatial clustering.

Basing on the discussion above we can assume that the five criteria that distributions should satisfy to model frequency distributions of microorganisms are not all satisfied for the Poisson distribution:

a. non-negative YES
b. allows zeros YES
c. discrete YES
d. approximates Poisson YES
e. approximates Lognormal NO

One way to have more flexibility is using Generalized Poisson distributions than single-parameter Poisson distributions. In a generalized distribution a parameter of the simple distribution (the only parameter for a Poisson) itself follows a distribution. Expressed mathematically, a distribution containing a parameter θ, say $f(x \mid \theta)$, can be generalized by weighting it by a distribution for θ, say $p(\theta)$, and then integrating with respect to θ to obtain the marginal distribution

$$g(x) = \int_{-\infty}^{+\infty} f(x \mid \theta)\, p(\theta)\, d\theta$$

For generalized Poisson distributions, the generalizing distribution, $p(\theta)$, describes the mean of the Poisson distribution, $f(x \mid \lambda = \theta)$, so that it is not limited to integer values, although it cannot be negative.

So, a way to model clustering is to describe the number of clusters by a Poisson distribution and the number of points within each cluster by another distribution; this may also be viewed as a mixture of Poisson distributions with different means, where the means follow another distribution. The '*generalized Poisson distribution*' models the total number of points in a given volume.

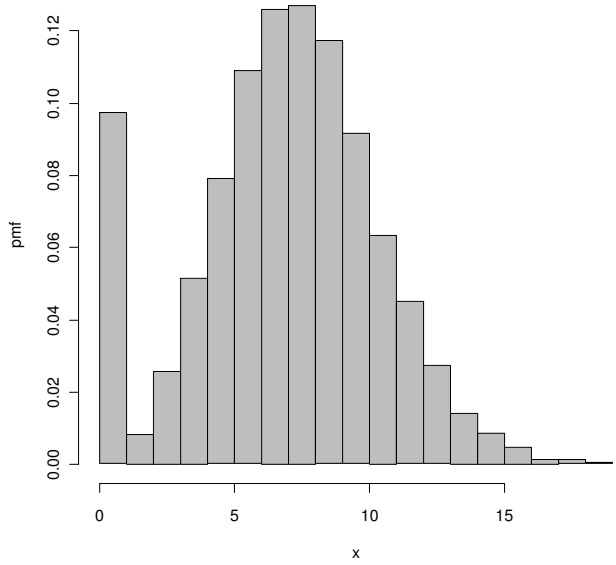
The advantage of the generalized Poisson frequency distribution, in terms of the five proposed criteria, is the matter that preserve the advantages of the single-parameter Poisson, but not necessary having variance equal to its mean it can model the skewness associated with a Lognormal frequency distribution.

a. non-negative YES
b. allows zeros YES
c. discrete YES
d. approximates Poisson YES
e. approximates Lognormal if the generalizing distribution is appropriate YES, otherwise NO.

The *'zero-inflated' Poisson distribution* is another kind of generalized Poisson distribution. This frequency distribution generates more zero values than a single parameter Poisson. As we can see in the example in Figure 2.4, this distri-

bution has a fixed proportion of zero values (10% in the example), with the remainder distributed according to a Poisson with a fixed mean, λ. Because the generalizing frequency distribution is discrete (i.e., Binomial or two valued; the mean of the Poisson is either 0 or λ) the resultant generalized distribution can have more than one peak.

Figure 2.4 - Zero-inflated Poisson frequency distribution, characterized with a Poisson distribution with an mean of 8 and with 10% of the values being zero



This frequency distribution is useful in a particular situation, that is when the overall batch of food product can be considered to be a mixture of two different groups of portions, one with no contaminated portions, one contaminated in a uniform random pattern. The distribution, however, suffers, in terms of the five criteria, of the same problems of the single-parameter Poisson distributions.

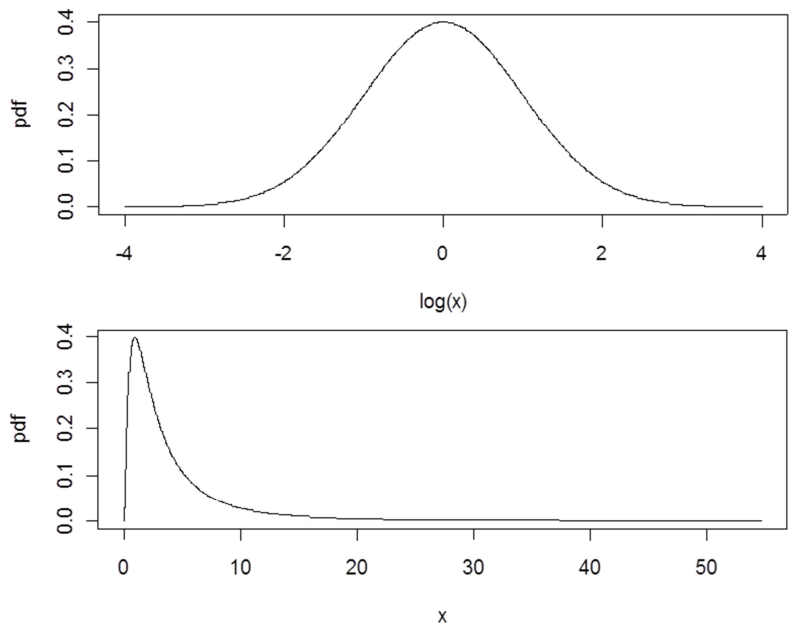
a. non-negative YES
b. allows zeros YES
c. discrete YES
d. approximates Poisson YES
e. approximates Lognormal NO

Lognormal distributions

A log-normal distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed. Thus, if the random variable Y is log-normally distributed, then X = log(*Y*) has a normal distribution. Likewise, if X has a normal distribution, then Y = exp(*X*) has a log-normal distribution. A random variable which is log-normally distributed takes only positive real values. It is defined by two parameters: the 'location' and the "scale" (values must be >0)[3].

As illustrated in Figure 2.5, when the logarithms of values follow a Normal distribution (top panel), the values follow a Lognormal distribution (bottom panel).

Figure 2.5 - Lognormal distribution



[3] Conventionally, parameters for the location and scale are the mean and standard deviation of the natural logs of the values. These can be converted to the log10 value which is more usually used in microbiology by dividing by ln(10) = 2.303.

There are several reasons that leads the Lognormal distribution to be of the most used distribution to model frequency distributions of microbial concentrations. First of all, microbiologists deal with numbers ranging from a few CFU to many billion and they are frequently represented by scientific notation (e.g., 1.23×10^8). So, working with decimal logarithms (e.g., 8.09) is more natural. Normal distributions are used very widely and successfully to represent distributions of values. The Normal distribution is well understood from a large number of people. Furthermore the Lognormal distribution has a technical advantage: the Central Limits Theorem says that (subject to some conditions) a value resulting from the sum of many independent effects will follow a Normal distribution. At last, the Lognormal distribution has some empirical justification, because it is nonnegative and reflects the tail to high values often associated with microbial concentrations.

However, this frequency distribution present also some disadvantages. The most important is that it gives zero probability for zero concentration, so it does not allow complete absence of microorganisms. The second one is that it is continuous.

Considering the five criteria, the Lognormal distribution presents the following characteristics:

a. non-negative YES
b. allows zeros NO
c. discrete NO
d. approximates Poisson NO
e. approximates Lognormal YES

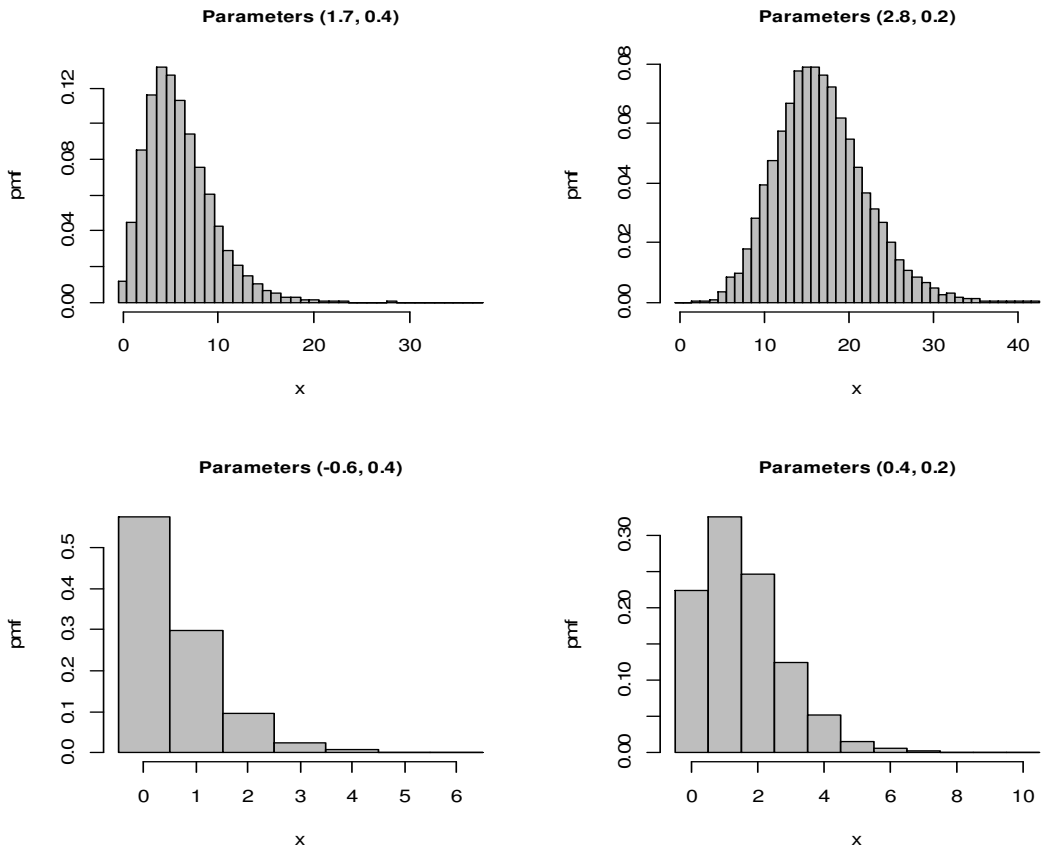
Clearly, the first limitation is relevant and not negligible when the microorganisms are present in food at very low levels while is almost negligible for very high concentrations (when we are talking of an average level of microorganisms of 1.000.000 CFU per portion, the difference between 1.000.000 and 1.000.001 CFU is negligible and the probability of zero is not important). Therefore, the Lognormal distribution is not the right choice to model low numbers of microorganisms; foodborne pathogens are usually present in very low concentrations.

Poisson-Lognormal distributions

It is possible to model the mean of a discrete Poisson frequency distribution with the continuous Lognormal frequency distribution. The resulting frequency distribution is a discrete Poisson-Lognormal distribution (Bulmer, 1974).

The advantage of this frequency distribution is that complies with all five of the criteria proposed for suitability of a frequency distribution to model spatial distribution of microorganisms. Indeed the Poisson-Lognormal distribution is discrete, non-negative and allows zeros. Furthermore it converges to the Poisson or to the Lognormal distribution.

Figure 2.6 - Poisson-Lognormal distributions



a. non-negative YES
b. allows zeros YES
c. discrete YES
d. approximates Poisson YES
e. approximates Lognormal YES

The similarities between the distributions under different combinations of mean and over dispersion can be summarized as (Jongenburger, Bassett, Jackson, Zwietering, & Jewell, 2012):

1. at high means, there is little difference between a continuous distribution and its discrete generalization of the Poisson distribution. While the discrete distribution may be more theoretically correct, the continuous distribution is easier to use and gives practically the same results;
2. at low means, the continuous statistical distributions can differ substantially from their generalizations of the Poisson, and the generalized Poisson distributions should be preferred for low numbers of microorganisms.

We know that pathogens generally occurs at low levels in food and their distribution is almost always clustered. These peculiarities influence the choice of the frequency distribution; indeed, at high means and with little clustering, the choice of model statistical distribution has little effect; the simple Poisson is inappropriate in the presence of any substantial clustering; the continuous distributions (Lognormal) is inappropriate when there is substantial probability of zeros, especially at low means; the family of generalized Poisson distributions is appropriate under a wide range of circumstances.

Beyond theoretical considerations, the choice of the distribution that better describe the pathogens will depend on how effectively the different distributions fit to the real data. According to Kilsby & Baird-Parker and Gale, at high levels of microorganisms, there is substantial positive experience supporting the use of the Lognormal distribution while at low levels of microorganisms, other distri-

butions may be superior to relative to the Poisson distribution. Unfortunately there are little study about the use of the Poisson-Lognormal distribution. Some findings suggest the use of the Negative Binomial distribution. It has been found that this distribution fit microbial data characterized by a relatively high occurrence of zero counts better than the Poisson distribution (Gonzales-Barron, Kerr, Sheridan, & Butler, 2010)

Among the ones described above, we can conclude that the Poisson-Lognormal is the most suitable distribution with regard to the five proposed criteria. The only biggest disadvantage of the Poisson-Lognormal is its mathematical complexity; the advantage is its capability to well-suit the distributions of microbial concentrations despite of the Lognormal that fail the suitability criteria that are important for being able to model low numbers of microorganism and the Poisson distribution that cannot model clustering.

3 – A case study: the ACME PIF production

3.1 “Law sampling” and “Autosampling” in powered infant formulae

A worldwide famous company, named “ACME” - we keep its name hidden for confidentially agreement with the management - produces powdered infant formula in two plants, one in England (ENACME) and one in Italy (ITACME).

The company pursue a rigid strategy of control quality given the type of commercialized product and the possible consequences on the consumer of contaminated products. Commercialized contaminated product furthermore will have an effect on the prestige of the company and the image cost incurred. These actions of quality control are realized with the support of internationally recognized certification authorities.

ACME, in controlling its quality production, operates different sampling plans in Italy or England. Both sites use “Autosampling” (where automatic mechanisms take many small aliquots throughout the production stream, which are combined into one larger aliquot for analysis) and “discrete sampling” (10 x 10g aliquots are taken and analyzed). Both sites examine and accept/reject product in quantities smaller than a complete batch ("sub-lot sampling") as well as in complete batches. We refer to these sampling plans as "*ACME Sampling*".

The International Commission on Microbiological Specifications for Foods (ICMSF) provided in 1974 urgently needed guidance on the use of sampling plans and the Microbiological Criteria (MC) for foods in international trade (ICMSF, 1974). In particular, ICMSF proposed to use *attribute* sampling plans; these kind of plans consists of two types: a two-class plan used to classify the test samples as “acceptable” or “defective” and a three-class plan used to classify the test samples as “acceptable”, “marginally acceptable” or “defective”. The microbiological criteria define the acceptability of a product or a food lot, based on the absence or presence or number of microorganism and/or quantity of their toxins/metabolites, per unit(s) of mass, volume, area or lot (Codex, 1997).

These sampling plans have been widely adopted by public and private parties. They have been incorporated into specifications in commercial trading contracts and have been enshrined in food law in different countries (Legan, 2001). Assessing the compliance of a batch to a MC depends on the criteria of the sampling plans. In general, two-class sampling plans are used when the health hazard is severe and direct (ICMSF, 1974) (ICMSF, 2002) (Legan, 2001) and its plan stringency depends on the number of samples tested (*n*) and the upper limit (*m*). Plans become more stringent as *n* increases and/or *m* decreases (Jongenburger, Reij, Boer, Gorris, & Zwietering, 2011).

The powdered infant formula is subject to Commission Regulation (EC) No 2073/2005 on microbiological criteria for foodstuffs (2005) which specifies "microbiological criteria defining the acceptability of the processes, and also food safety microbiological criteria setting a limit above which a foodstuff should be considered unacceptably contaminated with the microorganisms for which the criteria are set".

Relevant criteria are the following two:

Food safety criteria

Food category	Micro-organisms/their toxins, metabolites	Sampling-plan[(1)]		Limits		Analytical reference method[(2)]	Stage where the criterion applies
		n	c	m	M		
Dried infant formulae and dried dietary foods for special medical purposes intended for infants below six months of age	*Enterobacter Sakazakii (ES)*	30	0	Absence in 10gr		ISO/DTS 22964	Products placed on the market during their shelf-life

[(1)] n = number of units comprising the sample; c = number of sample units giving values over m or between m and M.

[(2)] The most recent edition of the standard shall be used.

Interpretation of the test results

Enterobacter Sakazakii in dried infant formulae and dried dietary foods for special medical purposes intended for infants below 6 months of age:

— satisfactory, if all the values observed indicate the absence of the bacterium;

— unsatisfactory, if the presence of the bacterium is detected in any of the sample units.

Process hygiene criteria - Milk and dairy products

Food category	Micro-organisms/ their toxins, metabolites	Sampling-plan[(1)]		Limits		Analytical reference method[(2)]	Stage where the criterion applies	Action in case of unsatisfactory results
		n	c	m	M			
Dried infant formulae and dried dietary foods for special medical purposes intended for infants below six months of age	*Enterobacteriaceae (EB)*	30	0	Absence in 10gr		ISO 21528-1	End of the manufacturing process	Improvements in production hygiene to minimize contamination. If Enterobacteriaceae are detected in any of the sample units, the batch has to be tested for E. Sakazakii and Salmonella

[(1)] n = number of units comprising the sample; c = number of sample units giving values over m or between m and M.

[(2)] The most recent edition of the standard shall be used.

Interpretation of the test results

Enterobacteriaceae in dried infant formulae and dried dietary foods for special medical purposes intended for infants below 6 months of age:

— satisfactory, if all the values observed indicate the absence of the bacterium;

— unsatisfactory, if the presence of the bacterium is detected in any of the sample units.

Also relevant is Codex Alimentarius Commission Code of hygienic practice for powdered formulae for infants and young children. CAC/RCP 66 - 2008. Annex I, Microbiological Criteria For Powdered Infant Formula, Formula For Special Medical Purposes And Human Milk Fortifiers.

Two sets of criteria are provided below, one for pathogens and a second for process hygiene indicators.

Criteria for pathogenic microorganisms

These are to be applied to the finished product (powder form) after primary packaging or anytime thereafter up to the point when the primary package is opened.

Microorganisms	n	c	m	Class Plan
Enterobacter Sakazakii (Cronobacter species)*	30	0	0/10 g	2

Where n = number of samples that must conform to the criteria: c = the maximum allowable number of defective sample units in a 2-class plan. m = a microbiological limit which, in a 2-class plan, separates good quality from defective quality.

*The mean concentration detected is 1 CFU in 340g (if the assumed standard deviation is 0.8 and probability of detection is 95%) or 1 CFU in 100g (if the assumed standard deviation is 0.5 and probability of detection is 99%)

Criteria for process hygiene

These are to be applied to the finished product (powder form) or at any other previous point that provides the information necessary for the purpose of the verification.

Microorganisms	n	c	m	M	Class Plan
Enterobacteriaceae**	10	$2^{(1)}$	0/10 g	Not applicable	2

Where n = number of samples that must conform to the criteria: c = the maximum allowable number of defective sample units in a 2-class plan or marginally acceptable sample units in a 3-class plan: m = a microbiological limit which, in a 2-class plan, separates good quality from defective quality

** The mean concentration detected is 1 CFU in 16g (if the assumed standard deviation is 0.8 and probability of detection is 95%) or 1 CFU in 10g (if the assumed standard deviation is 0.5 and probability of detection is 99%).

$2^{(1)}$ This 2 class plan is proposed because a 3 class plan with equivalent performance would not be practical analytically, given the low levels of EB typically occurring when stringent hygiene conditions are maintained.

It may seem that peak contaminations in up to 2 samples are tolerated in this Microbiological criterion (MC). However, it is assumed that the product is sufficiently homogeneous that high level contaminations will fail the MC. It is further assumed that, in practice, under sufficiently strict hygienic operation, the manufacturer will normally not find positives and that if, occasionally, positives are found the manufacturer will take appropriate actions.

Finding 1 or 2 positives should indicate to the manufacturer a trend toward potential loss of process control and appropriate actions would include further microbial evaluation of the implicated end product (i.e. re-evaluation of the EB content; when EB MC fails, evaluation of product safety using the proposed MCs for Salmonella and E. Sakazakii (Cronobacter species (Cronobacter species) before its release as well as evaluation of the hygiene programme to confirm it is suitable to maintain ongoing hygiene control or to amend the programme such that is suitable to do so).

Finding 3 or more positives should signal to the manufacturer loss of process control and appropriate actions should be the evaluation of product safety using the proposed MCs for Salmonella and E. Sakazakii (Cronobacter species) before release of the implicated product as well as evaluation of the hygiene programme to amend the programme such that it is suitable to maintain high hygiene control on an ongoing basis before production is resumed.

Noting that CAC/RCP 66 - 2008 requires action on positive Enterobacteriaceae, even when only 1 or 2 positive aliquots are found in a sample of 10 aliquots, The EC and Codex criteria are consistent with each other and may be summarized as:

a. Food safety criteria: *Enterobacter Sakazakii*: 30 x 10g aliquots, all negative
b. Process hygiene criteria: Enterobacteriaceae: 10 x 10g aliquots, all negative

These sampling plans are the ones called "*Law Sampling*". Law Sampling does not explicitly specify random aliquots, but EC 2073/2005 defines ‘representative sample’ as "a sample in which the characteristics of the batch from which it is drawn are maintained. This is in particular the case of a simple random sample where each of the items or increments of the batch has been given the same probability of entering the sample". In practice, it seems likely that Law Sampling will be intended to be random.

Apart from the technical aspects, it is important to remark that all the companies producing powdered infant formula have to comply with these legal requirements.

Recently, “ACME” management decided to assess possible differences in effectiveness between Law Sampling and “ACME” Sampling. In particular ACME wishes to asses if its sampling procedure a better (and how much better) than the Law one in detecting contaminated batches. The organization that usually supports the company in quality control released a technical report containing a lot of mathematical and statistical analysis very hard to understand from the management. For this reason the management asked two Ph.D students, one in math and one in applied statistics, to check the content of the report and verify the accuracy and the clarity of their statements.

ACME released internal production and sampling data with the agreement that the data can be used only for scientific purpose without mentioning the company directly.

3.2 ACME requests

Three specific questions were addressed from ACME management to the quality control organization and therefore to the two Ph.D students:

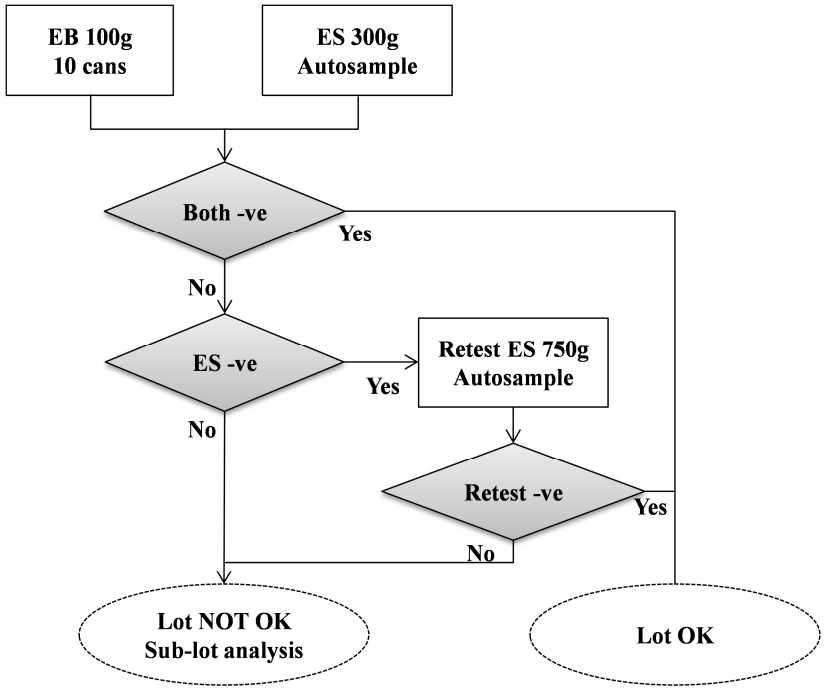
1) Is ACME Sampling better than Law Sampling?
   Is the ACME Sampling, especially Autosampling, demonstrably more likely to detect contamination than Law Sampling (10 g aliquots taken randomly)?
2) How much better?
   If ACME Sampling is demonstrably better than Law Sampling, can the difference be quantified?
3) What is the residual risk of in-market positive Law Sampling?
   No sampling scheme is perfectly effective in detecting contamination, so there is always the possibility that a contaminated batch will be 'missed' on one occasion to be detected later. What is the residual risk that a lot released following ACME Sampling will be found positive if later examined by Law Sampling? This should consider that later Law Sampling might be applied to a complete batch, or to an approximately contiguous part of a full batch.

3.3 Available data from the production process

In this work we try to answer to the questions addressed by the management using only data coming from the Italian plant.

Generally, ITACME produces two batches of powder in a day. The dimension of a batch varies between 3.600 and 17.000 cans of milk powder depending on the size of the cans (3.600-12.000 for the 800 gram cans and 6.000-17.000 for the 350 gram cans). As soon as the cans come out of the processing pipeline they are placed in pallets containing 200 of the larger cans (160 Kg) or 252 (88 Kg) of the smaller ones. In one hour either 7 pallets containing 1.400 big cans or 10 pallets containing 2.520 small cans are produced. This means 23.3 large cans or 42 small cans every minute.

Figure 3.1 - Flow diagram of ITACME sampling plan

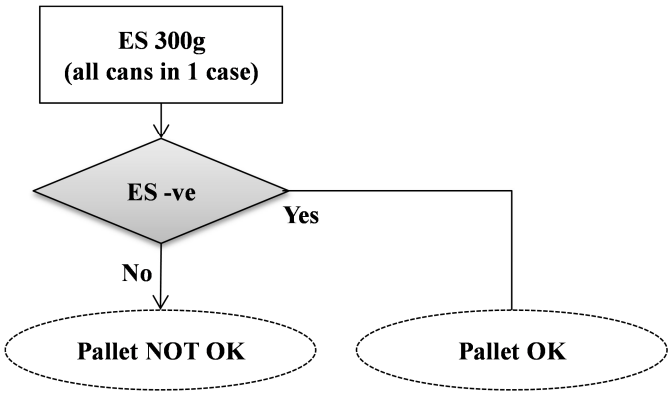


In the Figure 3.1 above is described “ACME” sampling plan of the Italian plant: 10 cans are randomly selected from the produced batch, 10 g from each can is se-

lected, and the mixed 100g are tested for EB. 10g every 2 minutes are collected during the process (autosample), the collected powder is mixed and 300g are tested for ES:

- if both tests are negative, the batch is released for marketing;
- if test for ES is positive, the batch is not OK and will go to the sub-lot analysis;
- if test for ES is negative and for EB is positive, from the autosample collected powder 750g are retested for ES;
- if the 750g retesting is negative the batch is released, if positive the batch will go to the sub-lot analysis.

Figure 3.2 - Sub-lot analysis per pallet flow diagram



For the sub-lot analysis (Figure 3.2), for each pallet, a case is chosen from the top of the pallet: the powder of all the cans is mixed and 300 g are tested for ES. Pallets are made up of:

a) 50 cases = 160Kg – 4 x 800g cans per case
b) 60 cases = 192 Kg – 4 x 800g cans per case
c) 42 cases = 88.2 Kg – 6 x 350g cans per case
d) 42 cases = 100.8 Kg – 6 x 400g cans per case

that take, at the maximum processing speed, respectively, 3’’20’’, 4’, 3’9’’and 2’45’’ to be produced. If the test for ES is negative the pallet is released, if positive the pallet is discarded.

The available data concern all the test performed by ACME from 2006 to 2011. Data on 1864 batches are organized per year of production, batch identification number, product types ("stage": before 6 months; 6 to 12 months; later), type of packaging, and results of the EB and ES tests.

Table 3.1 - Number of batches, by year and stage (Bag-in-Box and Can only)

	First	Follow up	Growth	
2006	90	106	0	*196*
2007	157	130	0	*287*
2008	176	135	0	*311*
2009	164	107	0	*271*
2010	148	94	0	*242*
2011	73	56	13	*142*
Total	*808*	*628*	*13*	**1449**

Table 3.2 - Number of batches, by year and type of packaging (Bag-in-Box and Can only)

	Bag in box	Can	
2006	196	0	*196*
2007	287	0	*287*
2008	311	0	*311*
2009	36	235	*271*
2010	0	242	*242*
2011	0	142	*142*
Total	*830*	*619*	**1449**

Table 3.3 - Number of batches, by year and stage (Free sample and free sample bag only)

	First	Follow up	Growth	
2006	5	40	0	*45*
2007	22	78	0	*100*
2008	14	58	0	*72*
2009	31	52	0	*83*
2010	37	22	0	*59*
2011	22	32	2	*56*
Total	*131*	*282*	*2*	**415**

Table 3.4 - Number of batches, by year and type of packaging (Free sample and free sample bag only)

	Free sample bag	Free sample bag in box	
2006	40	5	*45*
2007	82	18	*100*
2008	67	5	*72*
2009	82	1	*83*
2010	59	0	*59*
2011	56	0	*56*
Total	*386*	*29*	**415**

4 – THE CURRENT APPROACH FOR THE DETECTION OF ES IN THE PIF PRODUCTION

4.1 Theoretical background

The distributions of how microorganisms are physically distributed in foods determine both the likelihood that a foodstuff will cause illness and the consequential public health burden, but there is relative little knowledge about these, yet. The spatial distribution of the microorganisms in foods determines the value of the data on prevalence and/or concentration, obtained through sampling and testing, for informing food safety management decision-making (e.g., for lot acceptance or for process control) and, ultimately, their value for determining the associated public health burden (ILSI Europe, 2010).

Ideally, a batch of food should be produced under uniform and constant conditions; this means that the microorganisms present in the batch should be homogeneously distributed. Under this hypothesis (when drawn from a perfectly homogenously contaminated batch), the level of microorganisms present in an aliquot of a random samples would be distributed according to the Poisson distribution and would depend only on the microbial concentration within the batch (Jongenburger, Reij, Boer, Gorris, & Zwietering, 2011).

In practice, however, as we saw in the previous chapter, microorganisms are rarely homogeneously distributed within batches of food. Due, for instance, to the heterogeneity of the food matrix, incidental contamination, localized microbial growth or incomplete mixing, microorganisms are heterogeneously distributed, which results in an unequal probability to detect microorganisms in equal amounts of powder drawn throughout different parts of the whole batch. Depending on when and how the contamination has occurred during the production or thereafter, the spatial microbial distribution within the batch may also vary in size and concentration (Jongenburger, Reij, Boer, Gorris, & Zwietering, 2011).

Generally, sampling strategies are based on these assumptions. It has been shown that logarithms of counts from a batch of food are likely to be normally distributed (Kilsby & Baird-Parker, 1983); total viable counts data from batches of frozen meat, frozen vegetable, frozen dairy, and powdered products appeared to be lognormally distributed in 92% of the batches; in 8% of the batches, the total viable count appeared to be not lognormally, with a maximum of 13% for powdered products (Kilsby & Baird-Parker, 1983). Based on studies such as that of Kilsby and Baird-Parker, the International Commission Microbiological Specification for Foods assumed a lognormal distribution in order to evaluate the performance of attribute sampling plans (ICMSF, 2002).

In dried milk product it has been established that clustering and heterogeneity occurs and that there is a substantial stratification of contamination (Habraken, 1986). If the contamination is homogenously distributed, the probability to detect it by definition would be the same for each sample, no matter what sampling scheme is used. However, if the contamination is heterogeneously distributed or clustered in local spots, the sampling strategy becomes important (Jongenburger, Reij, Boer, Gorris, & Zwietering, 2011).

The literature on the subject reports a series of statements, mostly not proven, that we are mentioning only for completeness but on whose truth we have some doubts. We will get back to it at the end of this chapter.

Theoretical considerations suggest that sampling plans incorporating Autosampling should be better at detecting clustered contamination than a plan

analyzing the same amount of material taken as random (Jongenburger, Reij, Boer, Gorris, & Zwietering, 2011).

Systematic sampling was reported to be more effective in detecting a localized contamination (Habraken, 1986). Since systematic sampling improves the probability of detection, this improvement depends on the contaminated fraction and the number of aliquots taken. The improvement reaches a maximum, when exactly one systematic aliquot will be drawn from the contaminated fraction. In this case, the sampling interval equals the size of the contaminated fraction. Estimating the size of the contaminated fraction or the optimal sampling interval is a 'chicken and egg' dilemma. However, if one can estimate the size of the contaminated fraction, the optimal number of systematic samples may be derived from that (Jongenburger, Reij, Boer, Gorris, & Zwietering, 2011). This is surely valid when there is a single localized contaminated fraction. But what does it happen when there are multiple localized contaminated fractions in a batch?

Jongenburger et. al. showed in a further studio that stratified random sampling improved the probability to detect the heterogeneous contamination when there is systematic contaminations (caused, for example, by a contaminated filler head). Taking more and smaller samples and keeping the total sampling weight constant, clearly improved the performance of the sampling plans. Therefore to improve the probability of detection, Autosampling employed just before filling will be a practical way to collect the necessary large number of small samples per batch (Jongenburger, Reij, Boer, Gorris, & Zwietering, Actual distribution of Cronobacter spp. in industrial batches of powdered infant formula and consequences for performance of sampling strategies, 2011).

Again, we do not fully agree with most of the statements made above but we are not yet ready to refute them (more later).

4.2 Executive report from the control quality organization

The full report produced by the control quality organization has not been released by the ACME management. An executive report is the only document that has been granted for consultation. This document summarizes the answers given to ACME by the control quality organization.

1) Is ACME Sampling better than Law Sampling?

Basing on the fact that all batches contain at least one Enterobacteriaceae bacterium, so a perfectly sensitive plan would detect bacteria in practically every batch, the effectiveness of a sampling procedure depends, inter alia, on the total amount of material tested. ACME sampling procedures analyze more material than Law sampling so, all other things being equal, ACME procedures would be more likely to detect contamination than Law sampling.

Furthermore, being the contamination clustered, systematic sampling is more likely to detect contamination that random sampling (Habraken, 1986)*. For these two reasons ACME sampling will detect contamination more likely than Law Sampling.*

2) How much better?

There are several variants of ACME sampling, and different changing circumstances, so any single answer is simplistic. At high contamination levels (over about 30 CFU/kg) all sample sizes are very likely to detect contamination, they are effectively equivalent. At very low levels (less than about 1 CFU/kg) the probability of detection is approximately proportional to the sample size - doubling the sample size doubles the probability of detection.

The different clustering patterns have only small effects on detection probabilities in the contamination ranges of interest; so the differences between systematic and random sampling are small.

In conclusion, at the low contamination levels seen in 2011 ACME sampling is more likely than Law Sampling to detect contamination by a factor of very approximately 2 times; it seems that this is attributable more to total sample size than to systematic sampling. The increases in the higher probability of detecting higher levels of contamination are more limited. Clustering causes only small reductions in detection probability for random (Law) sampling; those small reductions are almost completely compensated for by structured (ACME) sampling.

3) What is the residual risk of in-market positive Law Sampling?

Considering that only small differences are caused by clustered/uniform distributions and by random/systematic sampling, calculations to answer to this question are on the basis of uniform distribution and random sampling.

Considering all batches of released product to be equally contaminated at levels indicated by the effective overall concentrations (0,28 CFU/kg), the detection probability for a 300gr sample is 8,14%. However, clustering has little effect if aliquots are taken at random from the whole batch: if all 30 aliquots in a sample come from a single random clustered stratum, the probability is somewhat reduced. If we assume that all released batches have been tested for EB and found negative (this is not strictly true - sometimes EB positive portions of batches are removed and the remainder released), we know that ES detection probabilities may be one-half to one-quarter of that in all batches, so the probability still decrease. Considering, finally, that the contamination levels in 2011 are substantially lower than earlier years, we conclude that the residual risk of in-market positive law sampling in 2011 is estimated to be in the range 0.25% to 8,14%, probably in the lower half of that range.

4.3 Random versus systematic sampling

To understand the reasons that led the control quality organization to produce those answers we firstly try to reconstruct their computations on the basis of the results published on the executive report and the data provided by ACME.

For any sampling plan, the probability of detection depends on the number of bacteria (CFU), being zero when there are no CFU (it is in general assumed that the analysis gives no false positives) and approaching 100% when the number of CFU is very large. To assess a sampling plan it is fundamental to have some idea of the typical contamination level; should we consider small numbers of CFU per batch (0, 1, 2, …), or millions, or somewhere in between?

The effect of a sampling scheme also depends on the distribution of contamination levels between batches. At one extreme, if all the batches are equally contaminated, there are no "good" and "bad" batches; the sampling plan has an equal chance of detecting contamination in all batches; the plan identifies batch-

es as good or bad at random. At the other extreme, if most of the batches have very low or very high contamination and only few batches have intermediate contamination, it is much easier for the sampling plan to distinguish batches.

To assess and compare sampling plans we need to make judgments on the between-batch distribution of contamination levels, as well as the range and typical contamination levels.

The principal influence on the probability of a given sampling plan detecting contamination is the number of bacteria in the batch, but this probability is also influenced by the arrangement of the bacteria in the batch. If the bacteria are randomly spread so that they are equally likely in all regions of the batch (the concentration is uniform throughout the batch) then it doesn't matter where the material in the sample is taken from. If the bacteria are concentrated in one (or a few) regions of the batch then a random sample may miss those regions. A sample made up of systematically taken aliquots may have a higher probability of detection, depending on the sampling plan and the pattern of contamination in the batch.

To assess and compare sampling plans we need to estimate within-batch distributions of contamination levels.

In the first two answers, the control quality organization focuses more on the amount of powder selected for testing that on the sampling schemes adopted. Acme sampling tests more product (300gr + 750gr if enterobacteriaceae tests positive) than Law testing, however the two sampling schemes can be compared taking into account the same amount of powder (i.e. excluding from the analysis the results of the retesting procedures).

Furthermore, while in the first answer they state (quoting the work of Habraken, 1986) that Autosampling is better than Law sampling in detecting contaminated batches (because "*systematic sampling is more likely to detect contamination that random sampling*"), in the second one they state that "*the different clustering patterns have only small effects on detection probabilities in the contamination ranges of interest; so the differences between systematic and random sampling are small*".

Do we have data to prove that? If we would have data on Auto and Law tests done on the same batches and assuming that if test is positive the batch is contaminated, the sampling schema that have a greater percentage of positive test is better.

Among the available data, we have 41 batches that were tested with both testing procedures; in 2006 a number of randomly selected batches were sampled and tested for ES by both 300g composite autosample and 30x10g random sampling. Results are shown in Table 4.1:

Table 4.1 – Numbers of positive batches by Auto/Law sampling

		30 x 10g Law sample		
		negative	**positive**	**total**
Autosample	**negative**	33	0	33
	positive	5	3	8
	total	38	3	41

The proportion of positives from Autosampling (8 / 41 = 19.5%) is significantly greater than the proportion from Law-sampling (3 / 41 = 7.3%) (one-sided exact Liddell test (Liddell, 1983), p = 0.031). The observed difference is 19.5% - 7.3% = 12.2% or expressed as a ratio 19.5% / 7.3% = 2.67. The lower 95% confidence limit on the ratio is 1.09.

Nevertheless, we have to observe that 41 batches are quite few, that the confidence intervals for the difference are unfortunately very wide due to the small size of the sample and the percentage of contaminated batches is too high (considering the available data concerning all the test performed by ACME from 2006 to 2011 as we will see the following analysis). For these reasons, we have some doubt about the drawing at random of these 41 because they do not seem a sample of the entire population of produced batches.

Despite of our above findings, in our opinion the two answers of the control quality control are quite incongruent: a sampling strategy cannot be considered a priori better than another. As a matter of fact, it is possible to demonstrate that the effectiveness of a sampling strategy is strongly affected by the

spatial distribution of the CFU inside the batch (in terms of typology and numbers of clusters).

Let *N* be the total number of aliquots in a batch and suppose we stratify the *N* aliquots in *K* strata. Each stratum has $\frac{N}{K} = m$ aliquots. Let's denote the aliquots in stratum $(i = 1, \dots, K)$ $a_{i1}, a_{i2}, \dots, a_{im}$. Assume that *r* of the aliquots in the batch are contaminated and that $r < m$. If we draw *randomly n* aliquots from the entire batch, the probability of drawing at least one contaminated aliquot is approximately

$$P(+_r) = 1 - \left(\frac{N-r}{N}\right)^n = 1 - (1-p)^n$$

if we assume that *N* is sufficiently large (i.e. using the binomial rather than the hypergeometric distribution) and $p = \frac{r}{N}$ (i.e. the probability of drawing a contaminated aliquot in a single draw).

If sampling is *systematic* and $K = n$, and assume extreme clustering (all *r* contaminated aliquots are concentrated in one stratum), the probability of testing positive (i.e. the probability of exactly one contaminated aliquot) is

$$P\left(+_{s_c}\right) = \frac{r}{m} = \frac{r}{\frac{N}{K}} = \frac{r \cdot K}{N} = n \cdot p$$

where $+_{s_c}$= positive testing with systematic concentrated. So, it could be demonstrated that

$$P\left(+_{s_c}\right) \geq P(+_r)$$

If the *r* contaminated aliquots are in different strata the probability is the same if the intervals of contaminated aliquots are disjoint. If there is some overlapping of the contaminated intervals in different strata, the $P(+)$ will decrease and the decrement depends on the amount of overlapping, reaching the minimum if every stratum has exactly the same contamination interval (perfect overlapping). The $P(+)$ in this case

$$P\left(+_{s_{po}}\right) = \frac{P\left(+_{s_c}\right)}{K} = \frac{\frac{r \cdot K}{N}}{K} = \frac{r}{N} = \frac{r}{m \cdot K} = \frac{r/K}{m}$$

One suggestion (Jongenburger, Reij, Boer, Gorris, & Zwietering, 2011) could be to draw randomly from each stratum a certain number of aliquots. Following this suggestion and assuming that the number of aliquots drawn in each stratum is the same, if $K = n$, we should draw one aliquot from each stratum. Consider now a sample space $A = \{a_1, a_2, \dots, a_m\}$ where the a_j $(j = 1, \dots, m)$ are the aliquots in a generic stratum. Denote by A_i the subset of contaminated aliquots in stratum *i*. A_i can be considered a subset of *A* and hence an event of the sample space *A*. If we draw one aliquot from A_i

$$P(A_i) = \frac{c_i}{m} \; (c_i = number\ of\ contaminated\ aliquots)$$

Hence,

$$P\left(+_{r_{str}}\right) = 1 - \prod_{i=1}^{K} P(\bar{A}_i) = 1 - \prod_{i=1}^{K} \left(1 - P(A_i)\right)$$

$P\left(+_{r_s}\right)$ reaches the minimum if $\prod_{i=1}^{K} P(\bar{A}_i)$ is maximum.

Since

$$\sum_{i=1}^{K} P(A_i) = \sum_{i=1}^{K} \frac{c_i}{m} = \frac{\sum c_i}{m} = \text{constant}$$

the distribution of the c_i among the strata is irrelevant. It is easy to show that the maximum of $\prod_{i=1}^{K} P(\bar{A}_i)$ is reached if $P(\bar{A}_i)$ is the same for each *i* (i.e. the number of contaminated aliquots is the same in each stratum). In this case

$$P\left(+_{r_{str}}\right) = 1 - \prod_{i=1}^{K} P(\bar{A}_i) = 1 - \prod_{i=1}^{K} \left(\frac{m - \frac{r}{K}}{m}\right) =$$

$$1 - \prod_{i=1}^{K} \left(1 - \frac{r}{K \cdot m}\right) = 1 - \left(1 - \frac{r}{K \cdot m}\right)^K$$

At the other extreme, if all r contaminated aliquots are in one stratum (concentrated), then

$$P\left(+_{r_{str_c}}\right) = 1 - \prod_{i=1}^{K} P(\bar{A}_i) = 1 - \left(\frac{m - r}{m} \cdot 1 \cdot \dots \cdot 1\right) = 1 - 1 + \frac{r}{m} = \frac{r}{m}$$

Table 4.2 summarizes, for certain numbers of contaminated aliquots, the probability values relating to the sampling strategies illustrated above, for an ideal batch of 4,5 tons (450,000 10g aliquots) and a sample of 30 aliquots.

Table 4.2 – Detection probabilities of different sampling strategy for certain number of contaminate aliquots (in a batch of 4,5 tons and a sample of 30 aliquots)

Cont. aliquots	$\boldsymbol{P}(+_{r})$	$\boldsymbol{P}(+_{s_c})$	$\boldsymbol{P}\left(+_{s_{po}}\right)$	$\boldsymbol{P}(+_{r_s})$	$\boldsymbol{P}\left(+_{r_{str_c}}\right)$
300	0.0198	0.0200	0.0007	0.0198	0.0200
500	0.0328	0.0333	0.0011	0.0328	0.0333
800	0.0520	0.0533	0.0018	0.0520	0.0533
1500	0.0953	0.1000	0.0033	0.0953	0.1000
2000	0.1251	0.1333	0.0044	0.1251	0.1333
3000	0.1818	0.2000	0.0067	0.1818	0.2000
5000	0.2848	0.3333	0.0111	0.2848	0.3333
10000	0.4904	0.6667	0.0222	0.4904	0.6667
15000	0.6383	1.0000	0.0333	0.6383	1.0000

In the third answer "*considering that only small differences are caused by clustered/uniform distributions and by random/systematic sampling, calculations to answer to this question are on the basis of uniform distribution and random sampling*" the control quality organization assumes that all the batches are that all the batches are identical and that the aliquots follow a uniform distribution.

As said above, if all the batches are identical and hence equally contaminated, there are no "good" and "bad" batches; the sampling plan has an equal chance of detecting contamination in all batches; the plan identifies batches as good or bad at random. Furthermore, if the aliquots follow a uniform distribution the residual risk of in-market positive Law Sampling will be exactly the same as the one found in the original batches. Notice, however, that detection of contamination is independent of the sampling plan and hence the assumption of random sampling is irrelevant. In this case, assuming a Poisson distribution, the

mean concentration level can be easily computed by the probability of detecting a contaminated batch.

In fact, let's assume that we draw randomly from a batch an aliquot and analyze the aliquot to find how many CFU it contains. Let's denote by *X* the number of CFU found and define negative (-ve) test if $X = 0$, positive (+ve) test if $X \neq 0$:

$$P(-ve\ test) = P(X = 0) = 1 - p$$
$$P(+ve\ test) = P(X > 0) = p$$

Let's suppose that we draw *n* aliquots from the batch and assume that the number of aliquots in the batch is large enough to be assumed infinite and analyze individually the *n* aliquots.

Let's denote by *Z* the number of aliquots that test positive; if sampling is *random*, $Z \sim B(n, p)$ no matter what the distribution of *X* is, and *p* is the probability that an aliquot drawn randomly from the batch tests positive.

Let's go now the other way around i.e. let's suppose we *randomly* draw *n* aliquots from each batch of a certain number of batches. This will produce an empirical distribution of values from 0 to *n* each with frequency f_i ($i = 0, \ldots, n$). If *p* is *constant* across batches then the empirical distribution is a sample from $Z \sim B(n, p)$ and we can also test the empirical distribution described above against $Z \sim B(n, p)$. Rejection of the binomial would imply that our assumptions are wrong, i.e. *p* varies across batches.

If X is a $\mathrm{P}(\lambda)$ for all the batches analyzed and λ is constant across batches,

$$P(X = 0) = e^{-\lambda} = 1 - p$$

where

$$p = 1 - P(X = 0) = 1 - e^{-\lambda}$$

is the probability that the single aliquot tests positive. If you draw *n* aliquots

$$Pr(+ve\ batch) = 1 - P(Z = 0) = 1 - (1 - p)^n = 1 - [P(X = 0)]^n$$

and the probability that the test is positive will depend on the number of aliquots drawn (tends to 1 for *n* that goes to infinity).

Hence if we know *p* we can get λ

$$p = 1 - P(X = 0) \Rightarrow P(X = 0) = 1 - p = e^{-\lambda}$$

and

$$\lambda = -log(1 - p)$$

Furthermore

$$p = 1 - \left(1 - Pr(+ve\ batch)\right)^{\frac{1}{n}}$$

and *p* could be estimated from the results of the batch testing.

If sampling is *random* and we assume that *P*(*X* > 0) is *constant across* batches, it is generally possible to go from the distribution of *Z* to the distribution of *X* if *X* is a one parameter distribution, while this is not possible if *X* is a two or more parameter distribution.

If the drawing of aliquots is *systematic* (Autosampling, for example) *Z* will be a Binomial if the distribution of CFU is a Poisson (i.e. a random spatial distribution). If the spatial distribution is not random (in particular if there is clustering) and the sampling is systematic the distribution of *Z* will still assume values 0, 1, …, *n* but it will not be a binomial even if *p* is constant across batches. Hence, if the drawing is systematic, we should test the empirical distribution against the binomial $(n, \hat{p})$, and if the test rejects the binomial we should conclude that the frequency distribution of CFU is not a Poisson or that *p* is not constant across batches. Presumably, even assuming constant *p* across batches, farther away is the empirical distribution from the binomial the more clustering there is. How to measure the distance between the empirical distribution and the $B(n, p)$ is matter of academic research (value of χ^2 or likelihood ratio statistic? Distance between empirical variance and $np(1 - p)$?).

Let us go back to the data provided by ACME. Assuming that the distribution of CFU in the batch follows a uniform distribution and considering only the results of the auto-test without the retesting, we get the following results:

Table 4.3 - Number of ES positive batches, by year (Bag-in-Box and Can only – 300g testing)

	Negative	Positive	% of +ve	Total
2006	185	11	*5,6%*	*196*
2007	276	11	*3,8%*	*287*
2008	273	38	*12,2%*	*311*
2009	231	40	*14,8%*	*271*
2010	224	18	*7,4%*	*242*
2011	142	0	*0,0%*	*142*
Total	*1331*	*118*	*8,14%*	**1449**

Table 4.4 - Number of EB positive batches, by year (Bag-in-Box and Can only – 100g testing)

	Negative	Positive	% of +ve	Total
2006	169	27	*13,8%*	*196*
2007	262	25	*8,7%*	*287*
2008	246	65	*20,9%*	*311*
2009	221	50	*18,5%*	*271*
2010	206	36	*14,9%*	*242*
2011	131	11	*7,7%*	*142*
Total	*1235*	*214*	*14,8%*	**1449**

Table 4.5 - Probabilities of positive samples and "effective" overall concentrations

			95% confidence interval	
		Estimation	***Lower limit***	***Upper limit***
ES (30 x 10g)	Pr(+)	8,14%	6,73%	9,55%
	CFU/kg[4]	0,28	0,23	0,33
EB (10 x 10g)	Pr(+)	14,80%	12,97%	16,63%
	CFU/kg	1,6	1,39	1,82

The level of contamination ('CFU/kg') of ES that would give the observed proportion of positive 300 g aliquots under the mentioned assumptions, would be 0.28 CFU/kg, 1.6 CFU/kg for EB, as stated in the executive report.

[4] The concentration of CFU was calculated assuming a Poisson distribution

The CFU/kg levels above indicate the order of magnitude of 'typical' contamination levels; that overall contamination levels - averaged across many batches - are in tenths of CFU/kg for ES and several CFU/kg for EB.

Through the ordinary tests of association it is possible to demonstrate that the batches analyzed are not identical and hence p is not constant across batches. Different groups of batches are contaminated differently and there is strong evidence that the concentration of ES and EB, and the probability of a positive aliquot and/or sample, varies between batches. Furthermore, we will show that concentration of CFU varies also within batches. So that there is no single concentration and probability applicable to a single batch or a group of batches. These contamination levels must hence be treated with caution.

The percentage of batches positive for EB (Table 4.6) is on average equal to 14.8%, but quite variable in the different years: very high in 2008 and 2009 (20.9% and 18.5%, respectively) and remarkably low in 2011. Statistically we can conclude that this percentage is certainly different in different years (p-value < 0.001), even if the index of association Cramer's V is not very high (13.5%).

Even more pronounced is the difference in the percentages of batches positive for ES for the different years (Table 4.7): from a high 16.6% for 2009 to 0% in 2011. The conclusion is that the yearly ES percentages, like EB, change from year to year but, from the available data, we cannot show any overall time trend. The association index is higher than that for EB but not remarkably high (18.4%).

The association between results on EB tests and months (Table 4.8) is barely significant ($p \approx 0.02$), but the strength of the association is quite weak (Cramer's V = 12.3%) and the only month for which we notice a clear difference is April.

More pronounced is the association month-positive ES test (Table 4.9) (p-value < 0.0001). As expected, the summer months, particularly August, have a much higher percentage of positive (22.2%).

Table 4.8 - Number of EB positive batches, by year
(Bag-in-Box and Can only)

	Negative	Positive	% of +ve	Total
2006	169	27	*13,8%*	*196*
2007	262	25	*8,7%*	*287*
2008	246	65	*20,9%*	*311*
2009	221	50	*18,5%*	*271*
2010	206	36	*14,9%*	*242*
2011	131	11	*7,7%*	*142*
Total	*1235*	*214*	*14,8%*	**1449**

Association statistics				**Symmetric Measures**		
	Value	**df**	**Asymp. Sig.**		**Value**	**Approx. Sig.**
Pearson Chi-Square	26,293	5	< 0,001	**Cramer's V**	0,135	< 0,001
Likelihood Ratio	27,439	5	< 0,001	N of Valid Cases	1449	

Table 4.7 - Number of ES positive batches, by year
(Bag-in-Box and Can only)

	Negative	Positive	% of +ve	Total
2006	182	14	*7,1%*	*196*
2007	275	12	*4,2%*	*287*
2008	271	40	*12,9%*	*311*
2009	226	45	*16,6%*	*271*
2010	224	18	*7,4%*	*242*
2011	142	0	*0,0%*	*142*
Total	*1320*	*129*	*8,9%*	**1449**

Association statistics				**Symmetric Measures**		
	Value	**df**	**Asymp. Sig.**		**Value**	**Approx. Sig.**
Pearson Chi-Square	48,99	5	< 0,001	**Cramer's V**	0,184	< 0,001
Likelihood Ratio	59,129	5	< 0,001	N of Valid Cases	1449	

Table 4.8 - Number of EB positive batches, by month (Bag-in-Box and Can only)

	Negative	Positive	% of +ve	Total
January	57	14	*19,7%*	*71*
February	111	14	*11,2%*	*125*
March	82	18	*18,0%*	*100*
April	97	6	*5,8%*	*103*
May	95	19	*16,7%*	*114*
June	125	35	*21,9%*	*160*
July	118	25	*17,5%*	*143*
August	37	8	*17,8%*	*45*
September	149	22	*12,9%*	*171*
October	140	17	*10,8%*	*157*
November	136	18	*11,7%*	*154*
December	88	18	*17,0%*	*106*
Total	*1235*	*214*	*14,8%*	**1449**

Association statistics

	Value	df	Asymp. Sig.
Pearson Chi-Square	21,928	11	0,025
Likelihood Ratio	23,013	11	0,018

Symmetric Measures

	Value	Approx. Sig.
Cramer's V	0,123	0,025
N of Valid Cases	1449	

Table 4.9 - Number of ES positive batches, by month (Bag-in-Box and Can only)

	Negative	Positive	% of +ve	Total
January	62	9	*12,7%*	*71*
February	118	7	*5,6%*	*125*
March	93	7	*7,0%*	*100*
April	100	3	*2,9%*	*103*
May	101	13	*11,4%*	*114*
June	132	28	*17,5%*	*160*
July	122	21	*14,7%*	*143*
August	35	10	*22,2%*	*45*
September	160	11	*6,4%*	*171*
October	145	12	*7,6%*	*157*
November	150	4	*2,6%*	*154*
December	102	4	*3,8%*	*106*
Total	*1320*	*129*	*8,9%*	**1449**

Association statistics

	Value	df	Asymp. Sig.
Pearson Chi-Square	51,713	11	< 0,0001
Likelihood Ratio	50,543	11	< 0,0001

Symmetric Measures

	Value	Approx. Sig.
Cramer's V	0,189	< 0,001
N of Valid Cases	1449	

Analyzing the data of the plant it is clear that the quality of the process in terms of contamination was strikingly improved in recent times. It was hence decided to report some of the descriptive statistics using only the more recent data. It was felt that 18 months was an adequate length of time to consider the process stable on the new level of contamination.

If we consider only the last 18 months, positive tests for EB and ES (Table 4.10 and 4.11) are not statistically associated with months. This is perhaps due to the much smaller sample size and the extremely small number of positives, particularly for ES.

Table 4.9 - Number of EB positive batches, by month
(Bag-in-Box and Can only)
(from may 2010 - latest 18 months)

	Negative	Positive	% of +ve	Total
January	10	1	*9,1%*	*11*
February	14	1	*6,7%*	*15*
March	10	0	*0,0%*	*10*
April	22	1	*4,3%*	*23*
May	23	5	*17,9%*	*28*
June	39	6	*13,3%*	*45*
July	29	1	*3,3%*	*30*
August	4	1	*20,0%*	*5*
September	43	7	*14,0%*	*50*
October	41	4	*8,9%*	*45*
November	22	3	*12,0%*	*25*
December	21	0	*0,0%*	*21*
Total	*278*	*30*	*9,7%*	**308**

Association statistics

	Value	df	Asymp. Sig.
Pearson Chi-Square	10,245	11	0,508
Likelihood Ratio	13,207	11	0,280

Symmetric Measures

	Value	Approx. Sig.
Cramer's V	0,182	0,508
N of Valid Cases	308	

Table 4.10 - Number of ES positive batches, by month
(Bag-in-Box and Can only)
(from may 2010 - latest 18 months)

	Negative	Positive	% of +ve	Total
January	11	0	*0,0%*	*11*
February	15	0	*0,0%*	*15*
March	10	0	*0,0%*	*10*
April	23	0	*0,0%*	*23*
May	27	1	*3,6%*	*28*
June	45	0	*0,0%*	*45*
July	30	0	*0,0%*	*30*
August	5	0	*0,0%*	*5*
September	49	1	*2,0%*	*50*
October	40	5	*11,1%*	*45*
November	25	0	*0,0%*	*25*
December	21	0	*0,0%*	*21*
Total	*301*	*7*	*2,3%*	**308**

Association statistics

	Value	df	Asymp. Sig.
Pearson Chi-Square	20,359	11	0,041
Likelihood Ratio	16,991	11	0,108

Symmetric Measures

	Value	Approx. Sig.
Cramer's V	0,257	0,041
N of Valid Cases	308	

There is no association between stage and positive test for EB (Table 4.12) ($p > 0.7$) while it is significant for ES (Table 4.13) ($p \approx 0.003$). The association index is, however, low (9%). Neither, results of EB or ES tests are associated with pack type (Table 4.14 and 4.15).

Table 4.12 - Number of EB positive batches by stage
(Bag-in-Box and Can only)

	Negative	Positive	% of +ve	Total
First	687	121	15,0%	*808*
Follow up	536	92	14,6%	*628*
Growth	12	1	7,7%	*13*
Total	*1235*	*214*	14,8%	*1449*

Association statistics

	Value	df	Asymp. Sig.
Pearson Chi-Square	0,552	2	0,759
Likelihood Ratio	0,644	2	0,725

Symmetric Measures

	Value	Approx. Sig.
Cramer's V	0,02	0,759
N of Valid Cases	1449	

Table 4.13 - Number of ES positive batches by stage (Bag-in-Box and Can only)

	Negative	Positive	% of +ve	Total
First	718	90	11,1%	*808*
Follow up	589	39	6,2%	*628*
Growth	13	0	0,0%	*13*
Total	*1320*	*129*	8,9%	*1449*

Association statistics				Symmetric Measures		
	Value	df	Asymp. Sig.		Value	Approx. Sig.
Pearson Chi-Square	11,865	2	0,003	**Cramer's V**	0,09	0,003
Likelihood Ratio	13,291	2	0,001	N of Valid Cases	1449	

Table 4.14 - Number of EB positive batches by pack type (Bag-in-Box and Can only)

	Negative	Positive	% of +ve	Total
Bag in box	708	122	14,7%	*830*
Can	527	92	14,9%	*619*
Total	*1235*	*214*	14,8%	*1449*

Association statistics				Symmetric Measures		
	Value	df	Asymp. Sig.		Value	Approx. Sig.
Pearson Chi-Square	0,008	1	0,931	**Cramer's V**	0,002	0,931
Likelihood Ratio	0,008	1	0,931	N of Valid Cases	1449	

Table 4.15 - Number of ES positive batches by pack type (Bag-in-Box and Can only)

	Negative	Positive	% of +ve	Total
Bag in box	762	68	8,2%	*830*
Can	558	61	9,9%	*619*
Total	*1320*	*129*	8,9%	*1449*

Association statistics				Symmetric Measures		
	Value	df	Asymp. Sig.		Value	Approx. Sig.
Pearson Chi-Square	1,207	1	0,272	**Cramer's V**	0,029	0,272
Likelihood Ratio	1,199	1	0,273	N of Valid Cases	1449	

Similarly to the other batches (Table 4.6 and 4.7), for the free sample batches there is association between the results of the tests and the years (Table 4.16 and 4.17), particularly for ES for which the percentages of positive increase from 2006 to 2009 (when it reaches a maximum of 32.5%) and decrease in the last two years reaching a minimum of 3.6% in 2011.

Table 4.16 - Number of EB positive batches, by year
(Free sample and free sample bag only)

	Negative	Positive	% of +ve	Total
2006	41	4	*8,9%*	*45*
2007	77	23	*23,0%*	*100*
2008	51	21	*29,2%*	*72*
2009	61	22	*26,5%*	*83*
2010	50	9	*15,3%*	*59*
2011	49	7	*12,5%*	*56*
Total	*329*	*86*	*20,7%*	**415**

Association statistics				Symmetric Measures		
	Value	**df**	**Asymp. Sig.**		**Value**	**Approx. Sig.**
Pearson Chi-Square	12,345	5	0,030	**Cramer's V**	0,172	0,030
Likelihood Ratio	13,152	5	0,022	N of Valid Cases	415	

Table 4.17 - Number of ES positive batches, by year
(Free sample and free sample bag only)

	Negative	Positive	% of +ve	Total
2006	42	3	*6,7%*	*45*
2007	88	12	*12,0%*	*100*
2008	56	16	*22,2%*	*72*
2009	56	27	*32,5%*	*83*
2010	51	8	*13,6%*	*59*
2011	54	2	*3,6%*	*56*
Total	*347*	*68*	*16,4%*	**415**

Association statistics				Symmetric Measures		
	Value	**df**	**Asymp. Sig.**		**Value**	**Approx. Sig.**
Pearson Chi-Square	29,142	5	0,000	**Cramer's V**	0,265	0,000
Likelihood Ratio	29,679	5	0,000	N of Valid Cases	415	

The association between positive tests and month (Table 4.18 and 4.19) is statistically significant for both, EB and ES (lower p-value for EB). The differences between months is much stronger here than the ones shown in the cans and boxes batches (Table 4.10 and 4.11), with particularly high percentages in the summer months.

Table 4.18 - Number of EB positive batches, by month
(Free sample and free sample bag only)

	Negative	Positive	% of +ve	Total
January	23	3	*11,5%*	*26*
February	14	2	*12,5%*	*16*
March	30	11	*26,8%*	*41*
April	18	6	*25,0%*	*24*
May	17	4	*19,0%*	*21*
June	43	10	*18,9%*	*53*
July	32	11	*25,6%*	*43*
August	5	11	*68,8%*	*16*
September	43	9	*17,3%*	*52*
October	54	10	*15,6%*	*64*
November	25	1	*3,8%*	*26*
December	25	8	*24,2%*	*33*
Total	*329*	*86*	*20,7%*	**415**

Association statistics				**Symmetric Measures**		
	Value	**df**	**Asymp. Sig.**		**Value**	**Approx. Sig.**
Pearson Chi-Square	32,559	11	0,001	**Cramer's V**	0,28	0,001
Likelihood Ratio	29,198	11	0,002	N of Valid Cases	415	

Table 4.19 - Number of ES positive batches, by month
(Free sample and free sample bag only)

	Negative	Positive	% of +ve	Total
January	22	4	*15,4%*	*26*
February	16	0	*0,0%*	*16*
March	33	8	*19,5%*	*41*
April	22	2	*8,3%*	*24*
May	17	4	*19,0%*	*21*
June	44	9	*17,0%*	*53*
July	29	14	*32,6%*	*43*
August	10	6	*37,5%*	*16*
September	44	8	*15,4%*	*52*
October	55	9	*14,1%*	*64*
November	25	1	*3,8%*	*26*
December	30	3	*9,1%*	*33*
Total	*347*	*68*	*16,4%*	**415**

Association statistics				**Symmetric Measures**		
	Value	**df**	**Asymp. Sig.**		**Value**	**Approx. Sig.**
Pearson Chi-Square	22,676	11	0,020	**Cramer's V**	0,234	0,020
Likelihood Ratio	24,229	11	0,012	N of Valid Cases	415	

The small samples sizes and the few positive tests means that the tests tend to lose power. In spite of that, the relationship between positive EB tests and month is still weakly significant, with higher percentages positive in the summer months (Table 4.20 and 4.21). Both, EB and ES are associated with stage (Table 4.22 and 4.23). Pack type is not associated with the test results (Table 4.24 and 4.25).

Table 4.20 - Number of EB positive batches, by month
(Free sample and free sample bag only)
(from may 2010 - latest 18 months)

	Negative	Positive	% of +ve	Total
January	3	0	*0,0%*	*3*
February	3	0	*0,0%*	*3*
March	1	0	*0,0%*	*1*
April	8	0	*0,0%*	*8*
May	7	0	*0,0%*	*7*
June	25	3	*10,7%*	*28*
July	5	1	*16,7%*	*6*
August	1	3	*75,0%*	*4*
September	14	6	*30,0%*	*20*
October	10	0	*0,0%*	*10*
November	4	1	*20,0%*	*5*
December	3	0	*0,0%*	*3*
Total	*84*	*14*	*14,3%*	**98**

Association statistics				**Symmetric Measures**		
	Value	**df**	**Asymp. Sig.**		**Value**	**Approx. Sig.**
Pearson Chi-Square	22,361	11	0,022	**Cramer's V**	0,478	0,022
Likelihood Ratio	21,971	11	0,025	N of Valid Cases	98	

Table 4.21 - Number of ES positive batches, by month
(Free sample and free sample bag only)
(from may 2010 - latest 18 months)

	Negative	Positive	% of +ve	Total
January	3	0	*0,0%*	*3*
February	3	0	*0,0%*	*3*
March	1	0	*0,0%*	*1*
April	8	0	*0,0%*	*8*
May	7	0	*0,0%*	*7*
June	28	0	*0,0%*	*28*
July	5	1	*16,7%*	*6*
August	4	0	*0,0%*	*4*
September	17	3	*15,0%*	*20*
October	10	0	*0,0%*	*10*
November	5	0	*0,0%*	*5*
December	3	0	*0,0%*	*3*
Total	*94*	*4*	*4,1%*	**98**

Association statistics

	Value	df	Asymp. Sig.
Pearson Chi-Square	11,581	11	0,396
Likelihood Ratio	11,109	11	0,434

Symmetric Measures

	Value	Approx. Sig.
Cramer's V	0,344	0,396
N of Valid Cases	98	

Table 4.22 - Number of EB positive batches by stage
(Bag-in-Box and Can only)

	Negative	Positive	% of +ve	Total
First	687	121	15,0%	*808*
Follow up	536	92	14,6%	*628*
Growth	12	1	7,7%	*13*
Total	*1235*	*214*	14,8%	*1449*

Association statistics

	Value	df	Asymp. Sig.
Pearson Chi-Square	0,552	2	0,759
Likelihood Ratio	0,644	2	0,725

Symmetric Measures

	Value	Approx. Sig.
Cramer's V	0,02	0,759
N of Valid Cases	1449	

Table 4.23 - Number of ES positive batches by stage (Bag-in-Box and Can only)

	Negative	Positive	% of +ve	Total
First	718	90	11,1%	*808*
Follow up	589	39	6,2%	*628*
Growth	13	0	0,0%	*13*
Total	*1320*	*129*	8,9%	*1449*

Association statistics				Symmetric Measures		
	Value	**df**	**Asymp. Sig.**		**Value**	**Approx. Sig.**
Pearson Chi-Square	11,865	2	0,003	**Cramer's V**	0,09	0,003

Table 4.24 - Number of EB positive batches by pack type (Free sample and free sample bag only)

	Negative	Positive	% of +ve	Total
Free sample bag	306	80	20,7%	*386*
Free sample bag in box	23	6	20,7%	*29*
Total	*329*	*86*	20,7%	*415*

Association statistics				Symmetric Measures		
	Value	**df**	**Asymp. Sig.**		**Value**	**Approx. Sig.**
Pearson Chi-Square	0,000	1	0,996	**Cramer's V**	0,000	0,996
Likelihood Ratio	0,000	1	0,996	N of Valid Cases	415	

Table 4.25 - Number of ES positive batches by pack type (Free sample and free sample bag only)

	Negative	Positive	% of +ve	Total
Free sample bag	321	65	16,8%	*386*
Free sample bag in box	26	3	10,3%	*29*
Total	*347*	*68*	16,4%	*415*

Association statistics				Symmetric Measures		
	Value	**df**	**Asymp. Sig.**		**Value**	**Approx. Sig.**
Pearson Chi-Square	0,830	1	0,362	**Cramer's V**	0,045	0,362
Likelihood Ratio	0,925	1	0,336	N of Valid Cases	415	

The evidence of association between product types and contamination is variable (Table 4.12, Table 4.13, Table 4.22, Table 4.23), and indicated associations are weak. Overall, at least to a first approximation, it seems reasonable to neglect differences between product types. There is no significant evidence of differences between packaging formats (Table 4.14, Table 4.15, Table 4.24, Table

4.25). At least to a first approximation, it seems reasonable to neglect differences between packaging formats.

There is strong evidence of substantial reductions in contamination levels in 2011, so that it may be misleading to apply quantitative conclusions based on earlier years to later production. There is also evidence that contamination levels are higher in the summer than the winter (Table 4.10, Table 4.11, Table 4.20, Table 4.21). Those differences between months and between years are differences in overall contamination levels between groups of many batches, averaged across those groups.

Moreover, from this data it is impossible to know anything about the within-batch distribution. We have, however, some data on sublot sampling that can shed some light on this aspect.

For 42 of the batches analyzed in the sublot analysis, we have the sequence in time of the results of the pallet testing (832 pallets). From this data we can see the spatial distribution of positive and negative pallets within the batch. If pallets within the batch are clustered, positive would be, more often than not, be followed by positive and the same for negatives. This would give an association between the result at time *t* and the result at time *t*-1. The data can be summarized by a 2 x 2 contingency table (Table 4.26):

Table 4.26 – Sub-lot testing: association between ES results for pallets and preceding pallets

		pallet t		
		negative	**positive**	**total**
pallet t-1	**Negative**	244	82	326
	Positive	79	385	464
	Total	323	467	790

We ran a chi-square test and the result is:

Chi-square test			
	Value	**df**	**p-value**
Chi-Square	264,8567	1	< ,000
Phi	0,579	1	

The association within the batches between the result at time *t* and the result at time *t*-1 is highly significant and this shows that there is clustering within batches.

The analysis performed on the available data clearly show that the conclusions reached by the control quality organization are, at best, only partially acceptable and quite often wrong. With the data on the sublot analysis we have proved the presence of clustering within batches. In such situation the uniform Poisson model is not adequate to describe the distribution of CFU inside the batch. As already said, the literature on the subject affirms that in most situations a Poisson LogNormal distribution seems to be more adequate to model the amount of microorganisms in aliquots of milk powder. Furthermore, the assumption of identical batches was clearly rejected by our analysis and since most of the data are across batches it is impossible to give an answer to all the questions posed by ACME.

As seen above, it is impossible to state that a sampling strategy is better than another (it depends on the spatial distribution of the CFUs inside the batch). An additional problem is due to the particular way the sample is drawn in the ACME sampling procedure that make it different from the usual systematic sampling. Drawing 10g every two minutes, we get 1.53 Kg of powder, from which, after intense mixing, only 300g are finally collected and analyzed. In the final stage the aliquots are thus remade and obviously only some of them will be contaminated. The number of aliquots contaminated will depend on the number of CFU present in the 1.53 kg of powder and on chance (allocate *m* CFU in *k* aliquots). This is the classical "occupancy problem" (Feller, 1968), complicated by the fact that *m* can assume values 0, 1, 2, …, that allows the computation of *p*, the probability that an aliquot is contaminated, and finally the probability that the sample is contaminated through the binomial distribution. The wide variety of situations that can arise make the analytical solution of this problem quite complex.

We therefore decided to tackle the problem by simulating the sampling process on a set of 100,000 batches generated by a Poisson Lognormal distribution. The simulations compute both the probability of detection of a contaminated batch and the residual risk of in-market positive Law Sampling. On the basis of the standard deviation value of 0.69539082 estimated by FAO/WHO (see Table 4.27) and imposing a probability of detection equal to 8.14% on the full production, the mean concentration level characterizing all the batch produced by ACME can be computed.

Table 4.27 – Calculated concentration values based on published studies in the scientific literature and unpublished studies provided to FAO/WHO on the frequency of E. Sakazakii contamination of PIF (FAO/WHO, 2006).

Company/Study	Year	Sample size (*g*)	Samples tested	*E. sakazakii* +	$C_{Poisson}$	$Log_{10}C$
A	2004	50	37	0	nd	
A	2004	50	281	5	0.000359	-3.44481
B	2004	10	497	1	0.000201	-3.69592
B	2004	10	2 018	9	0.000447	-3.34971
B	2004	10	1 286	*0.5*	3.89E-05	-4.41019
B	2004	10	437	*0.5*	0.000114	-3.94126
B	2004	10	2 114	1	4.73E-05	-4.325
B	2004	100	293	*0.5*	1.71E-05	-4.76753
C	2004	333	11 558	26	6.76E-06	-5.16986
D	2004	333	12	0	nd	
D	2004	333	12	0	nd	
E	2004	333	67	3	0.000138	-3.86149
F	2004	25	30	0	nd	
G	2004	333	518	1	5.8E-06	-5.23635
G	2004	333	571	8	4.24E-05	-4.37293
H	2004	100	320	7	0.000221	-3.65526
I	2004	250	6	2	0.001622	-2.78999
I	2004	250	121	27	0.00101	-2.99569
I	2004	250	198	23	0.000494	-3.30634
J	2004	65	20	2	0.001621	-2.79024
K	2004	333	316	20	0.000196	-3.70698
L	2004	333	434	5	3.48E-05	-4.45845
L	2004	333	328	*0.5*	4.58E-06	-5.33902
L	2004	100	940	13	0.000139	-3.85616
L	2004	100	888	16	0.000182	-3.74035
L	2004	100	908	2	2.21E-05	-4.65658
L	2004	100	921	4	4.35E-05	-4.36125
L	2004	100	255	15	0.000606	-3.21735
L	2004	100	2 523	25	9.96E-05	-4.00182
L	2004	100	756	8	0.000106	-3.97312
Muytjens, Roelofs-Willemse and Jasper	1988	333	141	20	0.000459	-3.33784
Iversen and Forsythe	2004	25	82	2	0.000988	-3.00537
FDA	2003	1 332	22	5	0.000194	-3.71317
Leuschner *et al.* *	2004	450	58	8	0.00033	-3.48172
Nazarowec-White and Farber	1997	333	120	8	0.000207	-3.68364
Heuvelink *et al.*	2002	25	40	1	0.001013	-2.99451
Heuvelink *et al.*	2003	25	101	2	0.0008	-3.0969
				Mean log concentration (MLC)		**-3.84051**
				Standard deviation		**0.695961**

To do that, it was firstly implemented the “finder” function in order to find the mean concentration level of the PLN that minimizes the square deviation between the probability of discarding the batch (= 1 - PLN(0)^30) and the desired probability of detection.

The simulation scheme was firstly implemented assuming that all batches are contaminated and the probability of detection equals the one found with our testing procedure (8.14%). The simulation then moves to more realistic schemes which assume progressive smaller fraction of contaminated batches (30%, 20%, 10% to arrive at the minimum value 8.14%). For example, if we hypothesize that only 30% of the batches is contaminated, the probability of detection for the contaminated batches rises to 27.133%, because 0.27133333 x 0.30 + 0 x 0.70 = 0.0814. In this case the mean concentration level of the contaminated batches is equal to -5.782467.

```
library(VGAM)
finder = function(y,a){
# sdlog on std FAO
pln0=dpolono(0,meanlog=y, sdlog=1.60119652566498)
return ( (1-pln0^30-a)^2 )
}

opt=optimize(finder, c(-30,0), a=.27133333)
opt
```

```
> opt
$minimum
[1] -5.782467
$objective
[1] 1.357951e-13
```

Independently from the actual spatial distribution of CFUs (impossible to recreate because of the absence of information on its shape), each batch is computationally recreated in vectorial mode, simulating the actual serialization of the product during the processing line (the time in which the process of Autosampling starts).

The batch reproduced in the simulations is a batch of 4.5 tons (450,000 10g aliquots). The code attached below simulates the configuration of a contaminat-

ed batch when only 30% of the produced batches is contaminated, randomizing the inside distribution of the CFUs.

```
tot.cub=450000
res=dpolono(0:1000,meanlog=-5.782467, sdlog=1.60119652566498)
expect=round(res*tot.cub,0)
names(expect)=0:1000
expect=expect[expect>0]

serializz=NULL
for (i in 1:length(expect))
    serializz=c(serializz,rep(i-1,expect[i]))
serializz=c(serializz, rep(0,tot.cub-length(serializz)))
# randomizzazione degli aliquots
serializz=serializz[ order(runif(length(serializz))) ]
```

For each of the hypothesized configurations of contamination (share of contaminated batch equal to 100%, 30%, 20%, 10, 8.14%), it was deemed necessary to make assumptions on the concentration of CFU within the batches, comparing patterns of random distribution with more realistic patterns of clustering. Based on the information supplied by the technicians of ACME, the distribution of CFUs in the case of clustering has been hypothesized of descending intensity starting from the center of the cluster. This configuration has been simulated by means of a function with the unique purpose of increasing the realism of the simulations; from a purely computational point of view, this distribution assumption does not affect in any way the outcome of the simulations, as the tests are based on positivity of the sampled aliquot (and not on the actual number of CFU present in it).

It was therefore implemented a generalized function that, for each of the hypothesized odds of contamination and types of distribution, is able to compare the performance of five different sampling strategies in identifying contaminated batch, extracting 30 aliquots of 10 grams from each batch.

The five different sampling strategies are:

- *Random sampling (Random or Law test)*: 30 aliquots are taken at random, using the urn scheme (drawing 30 values from a vector of indices of length 450,000);

- *Systematic sampling (SYS)*: 30 aliquots are drawn selecting the first at random from the block of aliquots 1 – 15,000 and then selecting the others with the 15,000 step;
- *ACME sampling (ACME300gr)*: an aliquot of 10gr is sampled every 2 minutes. In 2 minutes ACME is able to package 84 cans of 350g, it means manage 2,940 10gr aliquots. In other words, one aliquot is drawn at random from the block of aliquots 1 – 2,940 and the others are selected with the 2,940 step. In this way, we get 1,530gr of powder (153 10g aliquots). After well mixing this amount of powder, 300g of product are randomly selected for testing. The simulation of this last drawing was carried out using a self-generated algorithm that solved the Occupancy Problem mentioned earlier;
- *Stratified sampling (STR random)*: 30 aliquots are sampled at random one in each of the 30 blocks of 15,000 aliquots;
- *ACME stratified sampling (STR-ACME.random300gr)*: this sampling strategy is similar to *ACME300gr* with the difference that the aliquots are not systematically sampled but they are drawn at random, one for each of the 153 block of 2,940 aliquots. Even in this case, the amount of 1,530gr of powder is well mixed and 300g is randomly selected for testing.

The simulation scheme was then repeated assuming a probability of detection equal to 2.3% (the observed probability in the latest 18 months of production).

Results of the whole simulation process are shown in Table 4.28 and Table 4.29. Irrespective of the starting probability of detection (8.14% or 2.3%) and the different rates of contaminated batches (100%, 30%, 20%, 10% and the limiting case) no sampling schema is clearly prevalent in all situations. It is important to underline the fact that how the systematic sampling can be much worse of the

other sampling schemes due to the amount of overlapping while the improvement in case of no overlapping is rather limited[5]. The stratified random sampling can be a compromise between random and systematic sampling even if the improvement with respect to random sampling is quite low.

Table 4.28 – Simulation results for different hypothesis on the distribution of the contaminated aliquots and the five analyzed sampling strategies (considering a probability of detecting positive batches equal to 8.14%)

 # ITACME: P(+)=8.14%
 # all bathes are equally contaminated (var between batches = 0)
 # from finder.r: meanlog=-7.133803 (with sdlog=1.60119652566498 estimated from FAO/WHO)
 # contaminated aliquots = 1,272

	Random	SYS	ACME (300gr)	STR.random	STR-ACME.random (300gr)
random contamination	0,0813	0,0805	0,0832	0,0824	0,0822
one cluster	0,0811	0,0848	0,0864	0,0854	0,0858
two clusters, no overlapping	0,0807	0,0844	0,0848	0,0840	0,0869
two clusters, perfect overlapping every 2,940 aliquots (ACME step)	0,0826	0,0848	0,0774	0,0863	0,0852
two clusters, perfect overlapping every 15,000 aliquots (SYS step)	0,0809	0,0424	0,0816	0,0835	0,0829
			P(L+/released)		
random contamination		0,0812	0,0817	0,0812	0,0812
one cluster		0,0816	0,0813	0,0811	0,0815
two clusters, no overlapping		0,0807	0,0807	0,0809	0,0805
two clusters, perfect overlapping every 2,940 aliquots (ACME step)		0,0831	0,0822	0,0824	0,0824
two clusters, perfect overlapping every 15,000 aliquots (SYS step)		0,0812	0,0812	0,0808	0,0807

 # ITACME: P(+)=8.14%
 # hypothesis: 30% of the batches are contaminated - 70% contamination free
 # from finder.r: meanlog=-5.782467 for the only contaminated batches (with sdlog=1.60119652566498 estimated from FAO/WHO)
 # contaminated aliquots = 4,722

	Random	SYS	ACME (300gr)	STR.random	STR-ACME.random (300gr)
random contamination	0,0814	0,0817	0,0843	0,0816	0,0848
one cluster, perfect overlapping every 2,940 aliquots (ACME step)	0,0803	0,0946	0,0918	0,0946	0,0900
two clusters, perfect overlapping every 2,940 aliquots (ACME step)	0,0815	0,0949	0,0872	0,0940	0,0912
two clusters, perfect overlapping every 15,000 aliquots (SYS step)	0,0807	0,0468	0,0892	0,0866	0,0889
			P(L+/released)		
random contamination		0,0641	0,0636	0,0639	0,0636
one cluster, perfect overlapping every 2,940 aliquots (ACME step)		0,0603	0,0602	0,0603	0,0614
two clusters, perfect overlapping every 2,940 aliquots (ACME step)		0,0614	0,0633	0,0616	0,0620
two clusters, perfect overlapping every 15,000 aliquots (SYS step)		0,0716	0,0623	0,0625	0,0620

[5] As expected, in case of perfect overlapping (with SYS step) with contamination in only two cluster, the probabilities of detection for the systematic sampling is about one half of the same probabilities computed with the other sampling schemes. In the same situation, instead, the disadvantages of the systematic drawing characterizing the ACME sampling are partially compensated by the random drawing of the final 300g of powder carried out after the last mixing.

 # ITACME: P(+)=8.14%
 # hypothesis: 20% of the batches are contaminated - 80% contamination free
 # from finder.r: meanlog=-5.253941 for the only contaminated batches (with sdlog=1.60119652566498 estimated from FAO/WHO)
 # contaminated aliquots = 7,769

	Random	SYS	ACME (300gr)	STR.random	STR-ACME.random (300gr)
random contamination	0,0813	0,0814	0,0849	0,0808	0,0850
two clusters, perfect overlapping every 2,940 aliquots (ACME step)	0,0798	0,1027	0,0910	0,1031	0,0914
two clusters, perfect overlapping every 15,000 aliquots (SYS step)	0,0818	0,0518	0,0912	0,0901	0,0899
			P(L+/released)		
random contamination		0,0525	0,0512	0,0521	0,0514
two clusters, perfect overlapping every 2,940 aliquots (ACME step)		0,0416	0,0468	0,0428	0,0477
two clusters, perfect overlapping every 15,000 aliquots (SYS step)		0,0630	0,0482	0,0498	0,0486

 # ITACME: P(+)=8.14%
 # hypothesis: 10% of the batches are contaminated - 90% contamination free
 # from finder.r: meanlog=-3.963396 for the only contaminated batches (with sdlog=1.60119652566498 estimated from FAO/WHO)
 # contaminated aliquots = 25534

	Random	SYS	ACME (300gr)	STR.random	STR-ACME.random (300gr)
random contamination	0,0808	0,0807	0,0857	0,0814	0,0854
one cluster, perfect overlapping every 2,940 aliquots (ACME step), partial overlapping every 15,000 aliquots (SYS step)	0,0816	0,1000	0,0892	0,0997	0,0895
two clusters, perfect overlapping	0,0816	0,0821	0,0877	0,0964	0,0891
			P(L+/released)		
random contamination		0,0170	0,0130	0,0164	0,0132
one cluster, perfect overlapping every 2,940 aliquots (ACME step), partial overlapping every 15,000 aliquots (SYS step)		0,0000	0,0095	0,0003	0,0095
two clusters, perfect overlapping		0,0163	0,0112	0,0033	0,0102

 # ITACME: P(+)=8.14%
 # hypothesis: 8,14% of the batches are contaminated - 91,86% contamination free
 # from finder.r: meanlog=-4.105704e-05 for the only contaminated batches (with sdlog=1.60119652566498 estimated from FAO/WHO)
 # contaminated aliquots = 26,9562

	Random	SYS	ACME (300gr)	STR.random	STR-ACME.random (300gr)
random contamination	0,0814	0,0814	0,0814	0,0814	0,0814
one cluster, overlapping	0,0814	0,0814	0,0814	0,0814	0,0814
			P(L+/released)		
random contamination		0,0000	0,0000	0,0000	0,0000
one cluster, overlapping		0,0000	0,0000	0,0000	0,0000

Table 4.29 – Simulation results for different hypothesis on the distribution of the contaminated aliquots and the five analyzed sampling strategies (considering a probability of detecting positive batches equal to 2.3%)

 # ITACME: P(+)=2.3%
 # all bathes are equally contaminated (var between batches = 0)
 # from finder.r: meanlog=-8.439253 (with sdlog=1.60119652566498 estimated from FAO/WHO)
 # contaminated aliquots = 349

	Random	SYS	ACME (300gr)	STR.random	STR-ACME.random (300gr)
random contamination	0,0231	0,0227	0,0230	0,0225	0,0228
one cluster	0,0224	0,0225	0,0242	0,0232	0,0240
two clusters, no overlapping	0,0228	0,0237	0,0231	0,0228	0,0224
two clusters, perfect overlapping every 2940 aliquots (ACME step)	0,0224	0,0231	0,0206	0,0235	0,0219
two clusters, perfect overlapping every 9000 aliquots (SYS step)	0,0235	0,0119	0,0235	0,0228	0,0232
			P(L+/released)		
random contamination		0,0231	0,0229	0,0232	0,0232
one cluster		0,0224	0,0224	0,0224	0,0224
two clusters, no overlapping		0,0227	0,0229	0,0227	0,0228
two clusters, perfect overlapping every 2940 aliquots (ACME step)		0,0224	0,0224	0,0224	0,0225
two clusters, perfect overlapping every 9000 aliquots (SYS step)		0,0235	0,0236	0,0235	0,0236

 # ITACME: P(+)=2.3%
 # hypothesis: 30% of the batches are contaminated - 70% contamination free
 # from finder.r: meanlog=-7.197164 for the only contaminated batches (with sdlog=1.60119652566498 estimated from FAO/WHO)
 # contaminated aliquots = 1,195

	Random	SYS	ACME (300gr)	STR.random	STR-ACME.random (300gr)
random contamination	0,0225	0,0233	0,0222	0,0231	0,0234
one cluster	0,0221	0,0248	0,0250	0,0241	0,0238
two clusters, no overlapping	0,0228	0,0235	0,0241	0,0233	0,0242
two clusters, perfect overlapping every 2,940 aliquots (ACME step)	0,0231	0,0234	0,0219	0,0231	0,0236
two clusters, perfect overlapping every 15,000 aliquots (SYS step)	0,0231	0,0123	0,0228	0,0234	0,0240
				P(L+/released)	
random contamination		0,0212	0,0213	0,0211	0,0209
one cluster		0,0211	0,0207	0,0208	0,0207
two clusters, no overlapping		0,0216	0,0214	0,0218	0,0216
two clusters, perfect overlapping every 2,940 aliquots (ACME step)		0,0219	0,0217	0,0215	0,0217
two clusters, perfect overlapping every 9,000 aliquots (SYS step)		0,0223	0,0217	0,0217	0,0217

 # ITACME: P(+)=2.3%
 # hypothesis: 20% of the batches are contaminated - 80% contamination free
 # from finder.r: meanlog=-6.763741 for the only contaminated batches (with sdlog=1.60119652566498 estimated from FAO/WHO)
 # contaminated aliquots = 1,828

	Random	SYS	ACME (300gr)	STR.random	STR-ACME.random (300gr)
random contamination	0,0231	0,0230	0,0232	0,0223	0,0233
one cluster	0,0232	0,0246	0,0246	0,0243	0,0249
two clusters, no overlapping	0,0233	0,0244	0,0242	0,0242	0,0242
two clusters, perfect overlapping every 2,940 aliquots (ACME step)	0,0226	0,0242	0,0224	0,0246	0,0249
two clusters, perfect overlapping every 9,000 aliquots (SYS step)	0,0226	0,0121	0,0234	0,0236	0,0238
				P(L+/released)	
random contamination		0,0209	0,0208	0,0213	0,0209
one cluster		0,0208	0,0209	0,0209	0,0208
two clusters, no overlapping		0,0208	0,0211	0,0211	0,0211
two clusters, perfect overlapping every 2,940 aliquots (ACME step)		0,0204	0,0204	0,0204	0,0203
two clusters, perfect overlapping every 9,000 aliquots (SYS step)		0,0217	0,0207	0,0207	0,0206

 # ITACME: P(+)=2.3%
 # hypothesis: 10% of the batches are contaminated - 90% contamination free
 # from finder.r: meanlog=-5.981853 for the only contaminated batches (with sdlog=1.60119652566498 estimated from FAO/WHO)
 # contaminated aliquots = 3,902

	Random	SYS	ACME (300gr)	STR.random	STR-ACME.random (300gr)
random contamination	0,0235	0,0233	0,0239	0,0236	0,0238
one cluster, overlapping solo su passo ACME	0,0225	0,0258	0,0269	0,0266	0,0259
two clusters, perfect overlapping every 2,940 aliquots (ACME step)	0,0224	0,0261	0,0242	0,0260	0,0244
two clusters, perfect overlapping every 9,000 aliquots (SYS step)	0,0229	0,0125	0,0247	0,0245	0,0251
				P(L+/released)	
random contamination		0,0186	0,0184	0,0186	0,0187
one cluster, overlapping solo su passo ACME		0,0170	0,0165	0,0166	0,0165
two clusters, perfect overlapping every 2,940 aliquots (ACME step)		0,0170	0,0172	0,0165	0,0175
two clusters, perfect overlapping every 9,000 aliquots (SYS step)		0,0203	0,0178	0,0181	0,0176

 # ITACME: P(+)=2.3%
 # hypothesis: 2,3% of the batches are contaminated - 97,7% contamination free
 # from finder.r: meanlog=-4.105704e-05 for the only contaminated batches (with sdlog=1.60119652566498 estimated from FAO/WHO)
 # contaminated aliquots = 26,9562

	Random	SYS	ACME (300gr)	STR.random	STR-ACME.random (300gr)
random contamination	0,0230	0,0230	0,0230	0,0230	0,0230
one cluster, overlapping	0,0230	0,0230	0,0230	0,0230	0,0230
				P(L+/released)	
random contamination		0,0000	0,0000	0,0000	0,0000
one cluster, overlapping		0,0000	0,0000	0,0000	0,0000

4.4 Conclusions

The simulation results do not help us in answering all the ACME questions; as commented above, we can give some suggestions on the best sampling schemes: systematic sampling is certainly a risky procedure and in our opinion, if technically feasible, should be avoided in favor of some random scheme. We cannot, however, say anything on the residual risk of in-market positive Law sampling.

The simulations show that with the same probability of detecting the original contaminated batches, we can get very different percentages of contaminated aliquots in the released batches. The simplistic assumption of a PLN distribution for the number of CFU in all aliquots of the different batches does not solve our problems: all the batches would be contaminated and all should not be released on the marked (absurd). What is then a more realistic distribution? The assumption of zero inflated distributions hypothesized in some of our simulations is certainly more realistic. Unfortunately, even if this assumption were true, we do not know the zero inflation factor. To conclude, to say something on the characteristics of the released batches and hence on the efficacy of the lot by lot acceptance sampling, we would need to know much more on the spatial distribution within the batch and, even more important, the distribution between batches. This is the reason that suggested to abandon the approach followed by the quality control organization, as well as by the literature on the subject, and to propose a completely new and innovative one that will allow to answer all the questions posed by ACME.

5 – A NEW APPROACH FOR THE DETECTION OF ES IN THE PIF PRODUCTION

5.1 A new approach

Since we could not answer the questions posed by ACME with the procedures described in the previous chapters we will try to answer the same questions by a completely different approach.

It is well known that sensitivity and specificity are the most widely used statistics to describe a diagnostic test. In our case, the sensitivity is the probability that the test is positive when the batch is contaminated while the specificity is the probability that the test is negative when the batch is not contaminated. What is the meaning of “contaminated”? To make things clearer we will use wording such as “ES is present in the batch analyzed” even if “contaminated” for us means simply that the batch would not be released on the market if the firm had a perfect knowledge regarding the concentration of ES. Obviously, “not contaminated” means that the firm, with perfect knowledge, would release the batch. This is a necessary clarification since some people believe that at least a few CFU are always present in every batch. ACME management does not think so, but we do not care who is right, as long as a clear definition of “contaminated

batch" is given. The definition could be "at least one CFU is present in the batch" or "the concentration of CFU is above the detection limit" or any other definition.

Since in the following we are going to estimate the probability of many different events, some notation is necessary:

T_+ = the test is positive, ie. ES was found in the milk powder of the batch analyzed;

T_- = the test is negative, ie. ES was not found in the milk powder of the batch analyzed;

C = the batch is contaminated (ES is present in the powder of the batch analyzed);

$\bar{C}$ = the batch is not contaminated

To answer the questions posed by ACME we would need to know, or, more realistically, estimate from the available data, the sensitivity and specificity of the tests performed under the different sampling schemes.
The estimate of the sensitivity and specificity of a test together with the a priori probability, in our case $P(C)$, would also allow the estimates of the posterior probabilities, through the usual Bayes formulas:

$$P(C/T_+) = \frac{P(C) \cdot P(T_+/C)}{P(C) \cdot P(T_+/C) + P(\bar{C}) \cdot P(T_-/\bar{C})}$$

and

$$P(\bar{C}/T_-) = \frac{P(\bar{C}) \cdot P(T_-/\bar{C})}{P(\bar{C}) \cdot P(T_-/\bar{C}) + P(C) \cdot P(T_+/C)}$$

If a test has high sensitivity and high specificity, these probabilities would be very close to one and hence the test would be very reliable.

Among the available data there are some *laboratory tests* (called analytical validation tests) in which the tests were carried out to validate the ACME Lab test. These tests were performed on contaminated and not contaminated powder; i.e. in these tests the knowledge regarding the presence or absence of ES in the powder was perfect. In total we have 68 tests on contaminated powder and 97 on ES-free powder. Both sensitivity and specificity were equal to 100% i.e. the estimate of $P(T_+/C)$ and of $P(T_-/\bar{C})$ were both equal to one. This does not imply that the lab tests are 100% sure since these are only estimates of the true probabilities based on a random sample, and the same results could have been obtained even if the two mentioned probabilities had been less than one. To measure the precision of our estimates we could compute a confidence interval for the two probabilities. Unfortunately, the usual formulas used in constructing a confidence interval for proportions cannot be used since the estimates of $P(T_+/C)$ and of $P(T_-/\bar{C})$ are both equal to one and hence it is not possible to estimate the variances of the estimators (they would both be equal to zero). We can solve the problem using the Wilson Score interval (Wilson, 1927):

$$\frac{\hat{p} + \frac{1}{2n} z_{1-\alpha/2}^2 \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{1-\alpha/2}^2}{4n^2}}}{1 + \frac{1}{n} z_{1-\alpha/2}^2}$$

obtaining the following interval:

		95% confidence interval	
	Estimation	**Lower limit**	**Upper limit**
$P(T_+/C)$	1	0,9465	1
$P(T_-/\bar{C})$	1	0,9619	1

The confidence intervals have been computed only to evaluate the precision of the estimates. Notice, however, that the estimates are not, as normally happens, the midpoint of the intervals, but they are both equal to one. These probabilities, however, are estimated in a laboratory setting and they should be

revised if we operate in the field since they will be affected by the sampling criteria.

Even if the lab test were perfect (100% sensitivity and specificity), the random drawing of the powder to be analyzed could change the sensitivity (not the specificity) of the test performed, since $P(T_+/C)$ depends both on the precision of the lab test and on the capacity of the random drawing to intercept the contaminated powder if present. The way we draw the powder is hence very important: a batch could be contaminated, but the sample drawn is ES-free. A further complication is due to the fact that we do not know the so called prior probabilities, i.e. the probability that ES is present in a batch.

In order to answer the questions posed by ACME, we will assume that sensitivity and specificity of the lab test are equal to their estimates obtained in the analytical validation tests i.e. they are set equal to 1.

With this approach we will try to answer the questions directly with the data we have, making a few assumptions. The main assumption is the following: *if a test, either the ACME test or the Law test, is positive, then the batch is "contaminated"*. The assumption has some rather strong implications; the following statements are all equivalent to our assumption:

a) there are no false positives;
b) the probability that the batch is contaminated when we get a positive test is equal to one (we could relax our assumption a little by saying that this probability is "close to one" and if we work assuming that is equal to one, our conclusions will be approximately true)
c) if a batch is not contaminated the test cannot be positive, i.e. the probability that that the test is positive when the batch is not contaminated is equal to zero (we could assume that it is close to zero, with the same considerations made above).

To make things easier, let us start with some notation:

A_+	test positive with ACME sampling (Autosampling plus retesting) carried out only on the original batches
L_+	test positive with law-sampling on the original batch
$L_{+on}\, A_-$	test positive by Law-sampling carried out on released batch (original batches that tested negative by Autosampling and were released on the market). This test is carried out by the authorities.
C	batch is contaminated (there is at least one Sakazaky bacterium in the batch)
$\bar{C}$	batch is not contaminated
C_R	the set of contaminated batches released on the market (these batches tested negative by Autosampling and were - wrongly - released on the market)
C_S	the set of contaminated batches in the sublot analysis (it is really redundant since all the batches in the sublot analysis are contaminated - they tested positive by Autosampling - however some of the pallets may not be)

Notice the unusual notation $L_{+on}\, A_-$ to denote the conditional event "positive Law-test performed on the batches that tested positive in the Auto-test. We will often use this notation because we feel that sometimes it is more clear.
If ES is not present in the batch of powder, the two testing procedures are equivalent since they depend only on the lab tests. Autosampling is then better than Law-sampling if $P(A_+/C) > P(L_+/C)$.

The probability $P(\mathrm{C})$ is unknown; the only thing we know is that $P(\mathrm{C}) \geq P(A_+)$, given the assumption made above. Now,

- $A_+ = (A_+ \cap C) \cup (A_+ \cap \bar{C})$ i.e. the ACME procedure will test positive and the batch is contaminated or it will test positive and the batch is uncontaminated. Our assumption implies that the second case is not possible or, at least, the probability that a not contaminated batch tests positive is so close to zero that we can assume that it is equal to zero. So we

will assume that $P(A_+ \cap \bar{C}) = 0 \Rightarrow \boldsymbol{P(A_+) = P(A_+ \cap C) = P(C) \cdot P(A_+/C)}$;

- $L_{+_{on}}A_-) = [(L_{+_{on}}A_-) \cap C_R] \cup [(L_{+_{on}}A_-) \cap \bar{C_R}]$ i.e. Law testing on released batches can be positive and the batch is contaminated or can be positive and the batch is not contaminated. Again we exclude this second possibility, that is, we assume $P[(L_{+_{on}}A_-) \cap \bar{\mathrm{C}}_\mathrm{R}] = 0$ and hence $\boldsymbol{P(L_{+_{on}}A_-) = P[(L_{+_{on}}A_-) \cap \mathrm{C_R}] = P(C_R) \cdot P[(L_{+_{on}}A_-)/C_R]}$

But we know that

$$P(C_R) = P(C) \cdot P(A_-/C) = P(C) \cdot [1 - P(A_+/C)]$$

and hence

$$P(L_{+_{on}}A_-) = P(C_R) \cdot [P(L_{+_{on}}A_-)/C_R] = P(C) \cdot [1 - P(A_+/C)] \cdot P[(L_{+_{on}}A_-)/C_R]$$

If we take the ratio

$$\frac{P(L_{+_{on}}A_-)}{P(A_+)} = [1 - P(A_+/C)] \cdot \frac{P[(L_{+_{on}}A_-)/C_R]}{P(A_+/C)}$$

and so, after some algebra, we get

$$P(A_+/\mathrm{C}) = \frac{P(A_+) \cdot P[(L_{+_{on}}A_-)/\mathrm{C_R}]}{P(L_{+_{on}}A_-) + P(A_+) \cdot P[(L_{+_{on}}A_-)/\mathrm{C_R}]}$$

If we could estimate this last equation we would have solved all our problems since the left hand side of the equation is the sensitivity of the ACME testing procedure, that is, its effectiveness.

Notice now that:

- $P(A_+)$ can be estimated from the data on testing the original batches;
- $P(L_{+_{on}}A_-)$ can be estimated from the data on inspectors' testing;
- $P[(L_{+_{on}}A_-)/\mathrm{C_R}]$ cannot be estimated directly from data on released batches, but something can be done with the data on the sublot analysis (as we can see below).

Once $P(A_+/\mathrm{C})$ is estimated we can estimate $P(\mathrm{C})$ by

$$\hat{P}(\mathrm{C}) = \frac{\hat{P}(\mathrm{A}_+)}{\hat{P}(\mathrm{A}_+/\mathrm{C})}$$

It may seem strange that we will be able to estimate the probability of C without even defining it. Take into consideration, however, that we have implicitly defined C by saying that $P(\mathrm{A}_+/\bar{\mathrm{C}}) = 0$ or, equivalently, $P(\mathrm{C}/\mathrm{A}_+) = 1$. If we believe that practically all batches, even the cleanest, have some CFU, we are saying that a batch is not contaminated if the number of CFU is so small that the probability that it tests positive is zero or reasonably close to zero.

Unfortunately the data we have does not allow the estimation of $P(\mathrm{A}_+/\mathrm{C})$, since we do not know $P\left[\left(L_{+_{on}}A_-\right)/\mathrm{C}_\mathrm{R}\right]$. This probability can be estimated using the sublot sampling. We have data on 113 batches that tested positive with Autosampling plus retesting and for which a later analysis on pallets was carried out: for these we know the number of pallets tested (all the pallets of the batch were tested) and the number of positive pallets.

Every batch in the sublot sampling tested positive with the ACME procedure; hence all the batches were contaminated. Some boxes (4 big cans, 6 small ones or 12 bags in the old batches) were drawn from every pallet (really a crate was chosen and crates have 4 or 6 cans or 12 bags), the powder of the crate of the pallet was mixed and 300 grams was tested. If at least one pallet tested positive, the batch tested positive with sublot sampling. The testing on the batch was hence carried out on powder taken from a variable number of cans (calling "cans" also the bags), and a variable weight of the powder analyzed (300g multiplied by the number of pallets). Law testing on the released batches is instead done by randomly drawing 30 cans from the batch (sometimes the cans could be drawn from part of the batch), mixing the powder and analyzing 300grams.

We tried to estimate $P\left[\left(L_{+_{on}}A_-\right)/\mathrm{C}_\mathrm{R}\right]$, by first estimating $P\left[\left(L_{+_{on}}A_+\right)/\mathrm{C}_\mathrm{S}\right]$ from the data on sublot sampling. This was accomplished in two steps: first reducing the number of cans drawn to 30. Drew randomly a number of pallets in each batch to get exactly 30 cans. In order to get 30 cans we have to draw 2.5 pallets if crates are made of 12 bags, 5 pallets if 6 can crates and 7.5 pallets if 4 can crates. To draw 2.5 and 7.5 pallets, 2 and 7 pallets were first drawn and then

another was drawn with probability 0.5. At the end of this drawing, a certain percentage of batches tested positive.

The problem now is that we have drawn 30 cans from every batch but the amount of powder analyzed is not 300 grams, as required by law testing, but 300g multiplied by the number of pallets drawn. So, secondly, a logit model was applied to the data: the result of the acceptance test as the dependent variable and the amount of powder analyzed as the only explicative variable. Finally the forecast for the dependent variable (the estimate of the percentage of positive batches with law testing) was obtained by setting the amount of powder equal to 300g in the estimated model.

This procedure was repeated ten thousand times and the average was taken as our final estimate. The drawing was carried out both with and without replacement getting, obviously, almost the same results. Drawing with replacement was finally chosen because it is in line with bootstrap techniques that allow the estimation of the distribution of the estimator and the construction of a confidence interval.

The final result of this procedure is the estimation of $P\left[\left(L_{+_{on}}A_{+}\right)/\mathrm{C_S}\right]$, where $\mathrm{C_S}$ is the set of contaminated batches in the sublot analysis. In actual fact all batches in the sublot analysis are contaminated: we did not use C because $P(\mathrm{C}) < 1$ while $P(\mathrm{C_S}) = 1$. We will assume that

$$P\left[\left(L_{+_{on}}A_{+}\right)/\mathrm{C_S}\right] = P\left[\left(L_{+_{on}}A_{-}\right)/\mathrm{C_R}\right]$$

The weak point of this reasoning is the usual one: both probabilities refer to contaminated batches, but the contaminated batches in the sublot analysis ($\mathrm{C_S}$) tested positive by ACME procedure and hence presumably have a higher concentration of bacteria than the contaminated batches ($\mathrm{C_R}$) that tested negative using the same procedure. Furthermore, sublot sampling is not completely random: with our procedure we randomly drew some pallets and then drew some cans. Sampling is hence “weakly” systematic. The final estimate reported below is probably an overestimate of the intended probability (batches in the sublot sampling have a higher concentration of bacteria that increase the probability of testing positive and sampling is systematic that could increase or de-

crease this probability). A better way to estimate it in the future will be proposed in the next chapter.

With this assumption and the results of the described simulation, we can compute:

$$P\left[\left(L_{+_{on}}A_{-}\right)/\mathrm{C_R}\right] = 0.675 \; with \; a \; 95\% \; confidence \; interval \; [0.575, 0.771]$$

The available data give us the results of acceptance sampling from May 2006 to October 2011, more than 5 years of testing for a total of 1449 batches. Of these 129 tested positive either with the 300g testing or the successive 750g testing. The estimate of $P(A_+)$ = 0.089 with a 95% confidence interval [0.0744, 0.1037].

Finally, we have the results on inspection testing by the authorities: in the more than 5 years of testing the inspectors carried out at least 600 (probably more than 700) tests, none positive.

The following table summarizes the various probabilities:

		95% confidence interval	
	Estimation	**Lower limit**	**Upper limit**
$P(A_+)$	0.089	0.0744	0.1037
$P\left[\left(L_{+_{on}}A_{-}\right)/\mathrm{C_R}\right]$	0.675	0.5750	0.7710

From the information on the authorities testing we get:

		95% confidence interval	
	Estimation	**Lower limit**	**Upper limit**
$P\left(L_{+_{on}}A_{-}\right)$ (n=500)	0	0	0.0076
$P\left(L_{+_{on}}A_{-}\right)$ (n=600)	0	0	0.0064
$P\left(L_{+_{on}}A_{-}\right)$ (n=700)	0	0	0.0055

where these last probabilities were computed with the Wilson score interval.

These last probabilities really answer the question on the residual risk that a lot released applying the ACME procedure will be found positive by the authorities applying the Law sampling procedure. The estimate is zero but the 95% confidence interval gives a maximum risk of 0.0076 if n = 500 (or 0.0055 if n = 700).

Finally, applying the formula found above

$$P(A_+/\mathrm{C}) = \frac{P(A_+)\cdot P\left[\left(L_{+_{on}}A_-\right)/\mathrm{C_R}\right]}{P\left(L_{+_{on}}A_-\right) + P(A_+)\cdot P\left[\left(L_{+_{on}}A_-\right)/\mathrm{C_R}\right]}$$

we can compute the sensitivity of the ACME test. The estimations of the various probabilities in the best and worst possible cases are reported below:

$\boldsymbol{P(A_+)}$	$\boldsymbol{P\left[\left(L_{+_{on}}A_-\right)/\mathrm{C_R}\right]}$	$\boldsymbol{P\left(L_{+_{on}}A_-\right)}$	$\boldsymbol{P(A_+/\mathrm{C})}$	
0.0890	0.6750	0	1	*best case*
0.0744	0.5750	0.0076	0.8490	*worst case*

The estimate of the sensitivity of the ACME testing procedure, with our assumptions, is 1 i.e. perfect testing. In the worst situation i.e. assuming that two of the computed probabilities are at the lower end of the confidence interval and the $P\left(L_{+_{on}}A_-\right)$ is at the upper limit we get sensitivity equals to 0.849. Take into consideration, however, that $P\left[\left(L_{+_{on}}A_-\right)/\mathrm{C_R}\right]$ could be overestimated. Finally, an estimate of $P(\mathrm{C})$ is

$$\hat{P}(\mathrm{C}) = \frac{\hat{P}(\mathrm{A_+})}{\hat{P}(\mathrm{A_+}/\mathrm{C})} = 0.089$$

$$with\ an\ aproximate\ 90\%\ confidence\ interval\ [0.074, 0.119] if\ n = 600$$

Take into consideration, however, that this refers to the average of the last six years; today the situation is much improved and the probability of the batch testing positive with ACME procedure in the last eighteen months is just above 2% with a similar improvement for $P(\mathrm{C})$. However, since it refers only to contaminated batches, the $P(\mathrm{A_+}/\mathrm{C})$ may be unchanged.

To conclude, the questions that can be answered with this approach are:

1. *Is ACME Sampling better than Law Sampling and, if yes, how much better?* The estimate of the sensitivity of ITACME sampling process is 1, i.e. all contaminated batches are identified by ACME sampling and are not released on the market. Naturally, this is the point estimate; in the least favorable situation, the sensitivity of ITACME sampling is

equal to 0.849. We do not have a direct estimate of the sensitivity of Law sampling. From the data on the sublot analysis we estimated a sensitivity of 0.675, giving ACME sampling a 48% advantage over Law sampling. The advantage is probably much greater considering that the sensitivity found for Law sampling, as already stated, is probably overestimated.

2. *What is the residual risk of in-market positive Law Sampling?*
   From the samples taken by the authorities, the estimate of the residual risk of Law Sampling testing positive is zero with a 95% confidence interval [0% - 0.76%].

5.2 Conclusions

From our new analysis and the results of the simulations of chapter 4 we can conclude that ACME sampling is certainly superior to Law sampling, most probably due to the greater quantity of powder analyzed by the first method and not to the sampling procedure. In point of fact, without the retesting that further analyzes 750 g of powder in presence of positive test to the bacteriaceae, ACME sampling is just slightly better in the best situation (no or very small overlapping of the contaminated areas), but can be much worse.

Since the tests carried out by the Authorities give an estimation of the residual risk equal to 0% with the sample size of 500 – 700, the real rate of contamination should be pretty close the limiting one.

A weak point of our new approach is the estimation of $P\left[\left(L_{+_{on}}A_{-}\right)/\mathrm{C_R}\right]$ that was quite elaborate and that probably provided an overestimation of this probability. In the next chapter, we will propose the collection of new data that will, at least partially, overcome this problem and that will give better answers to all the question submitted by ACME.

6 – PROPOSALS FOR THE FUTURE

In the previous chapter we estimated all the relevant probabilities (the most important was the sensitivity of ACME testing procedure) making an assumption that allowed the estimation (really overestimation) of $P\left[\left(L_{+_{on}}A_{-}\right)/\mathrm{C_R}\right]$. We are now going to propose some further testing on a certain number of batches that would allow the elimination of the mentioned assumption, with the substitution of a much weaker one.

Our proposal to ACME was to test the original batches for an adequate period of time by both sampling procedures. We suggested to collect data by testing a certain number of batches also by Law sampling, so we can also estimate $P(L_{+})$.

As we saw above, tests of this kind were carried out in ITACME in 2006 on only 41 batches and we were able to prove a difference between the two testing procedures; the confidence interval for this difference was, however, too wide. A larger sample is certainly needed, but the decision on how large the sample should be can be delayed: we can follow a sequential testing procedure where the sample size is not fixed in advance. Instead the data are evaluated as they are collected, and further sampling is stopped in accordance with a pre-defined

stopping rule as soon as significant results have been observed. - a recognized and pragmatic approach. After that, a second Law testing should be carried out only on the set that tested differently by the two testing procedures (presumably very few. Obviously, released batches will be Law tested by the authorities. It is difficult to decide how many batches should be tested with Law testing since the right number depends on the probability of positive tests, on the difference between $P(A_+)$ and $P(L_+)$ and on the precision we would like to have.

If we assume, as was done in the previous chapter, specificity equal to 1, i.e.

$$P(A_+/\bar{\mathrm{C}}) = P(L_+/\bar{\mathrm{C}}) = 0 \Leftrightarrow P(\mathrm{C}/A_+) = P(C/L_+) = 1$$

then

$$\frac{P(A_+)}{P(L_+)} = \frac{P(\mathrm{C}) \cdot P(A_+/\mathrm{C})}{P(\mathrm{C}) \cdot P(L_+/\mathrm{C})} = \frac{P(A_+/\mathrm{C})}{P(L_+/\mathrm{C})}$$

will give an estimate of how much better ACME procedure is in terms of the sensitivity of the two testing procedures. If the difference between the two testing procedures is noticeable (like the 41 batches tested in ITACME in 2006), an adequate sample should not take more than a year. This, together with some testing done on the released batches would answer most of ACME questions.

To summarize our proposal consider all the possible results of the two testing procedures and their consequences:

1. If A_+ and L_+ the batch is rejected;
2. If A_- and L_- the batch is released (authorities will carry out Law tests on these batches);
3. If A_- and L_+ or A_+ and L_- the batch is rejected and we propose to perform another Law test on these batches.

Denote by

$L_{+_{on}}A_-$ = Law test positive on batches that tested negative to Autosampling
$L_{+_{on}}L_-$= Law test positive on batches that tested negative to Law sampling

Now,

$$P\left(L_{+_{on}}A_{-}\right) = P\left(\mathrm{C}_{A_{-}}\right) \cdot P\left[\left(L_{+_{on}}A_{-}\right)/\mathrm{C}_{A_{-}}\right]$$
$$= P(\mathrm{C}) \cdot [1 - P(A_{+}/C)] \cdot P\left[\left(L_{+_{on}}A_{-}\right)/\mathrm{C}_{A_{-}}\right]$$

where $C_{A_{-}}$ = set of contaminated batches that tested negative to Autosampling, and

$$P\left(L_{+_{on}}L_{-}\right) = P\left(\mathrm{C}_{L_{-}}\right) \cdot P\left[\left(L_{+_{on}}L_{-}\right)/\mathrm{C}_{L_{-}}\right]$$
$$= P(\mathrm{C}) \cdot [1 - P(L_{+}/C)] \cdot P\left[\left(L_{+_{on}}L_{-}\right)/\mathrm{C}_{L_{-}}\right]$$

where $C_{L_{-}}$ = set of contaminated batches that tested negative to Law sampling.

In order to estimate the complete set of probabilities involved, we shall assume that $P\left[\left(L_{+_{on}}A_{-}\right)/C_{A_{-}}\right] = P\left[\left(L_{+_{on}}L_{-}\right)/C_{L_{-}}\right]$. This is the weaker assumption that substitute the stronger one made in the previous paragraph. The assumption regards two sets of batches, both contaminated.

This, again, could raise the objection that $\mathrm{C}_{A_{-}}$ has a different level of concentration than $\mathrm{C}_{L_{-}}$ due to the more effective Autosampling with respect to Law-sampling. The difference, however, should not be relevant. So, we would have the following equations:

1) $P(A_{+}) = P(\mathrm{C}) \cdot P(A_{+}/\mathrm{C})$
2) $P(L_{+}) = P(\mathrm{C}) \cdot P(L_{+}/\mathrm{C})$
3) $P\left(L_{+_{on}}A_{-}\right) = P(\mathrm{C}) \cdot [1 - P(A_{+}/\mathrm{C})] \cdot P\left[\left(L_{+_{on}}A_{-}\right)/\mathrm{C}_{A_{-}}\right]$
4) $P\left(L_{+_{on}}L_{-}\right) = P(\mathrm{C}) \cdot [1 - P(L_{+}/\mathrm{C})] \cdot P\left[\left(L_{+_{on}}A_{-}\right)/\mathrm{C}_{A_{-}}\right]$

We have 4 equations in 4 unknowns: $P(\mathrm{C})$, $P(A_{+}/\mathrm{C})$, $P(L_{+}/\mathrm{C})$, $P\left[\left(L_{+_{on}}A_{-}\right)/\mathrm{C}_{A_{-}}\right]$. Denote by:

$$\begin{array}{ll} P(\mathrm{C}) = x & \\ P(A_{+}/\mathrm{C}) = y & \\ P(L_{+}/\mathrm{C}) = z & \text{and} \\ P\left[\left(L_{+_{on}}A_{-}\right)/\mathrm{C}_{A_{-}}\right] = w & \end{array} \qquad \begin{array}{l} P(A_{+}) = a \\ P(L_{+}) = b \\ P\left(L_{+_{on}}A_{-}\right) = c \\ P\left(L_{+_{on}}L_{-}\right) = d \end{array}$$

a, b, c, d can be estimated and x, y, z, w are the unknowns.

The 4 equations are now written:

1) $a = x \cdot y$
2) $b = x \cdot z$
3) $c = x \cdot (1 - y) \cdot w$
4) $d = x \cdot (1 - z) \cdot w$

Solve by substitution:

1) $x = \frac{a}{y}$
2) $z = \frac{b}{x} = \frac{b}{\frac{a}{y}} = \frac{b}{a} y$
3) $c = \frac{a}{y}(1 - y)w \Rightarrow w = \frac{c}{\frac{a}{y}(1-y)} = \frac{c}{\frac{a-ay}{y}} = \frac{cy}{a-ay}$
4) $d = \frac{a}{y}\left(1 - \frac{b}{a}y\right)\frac{cy}{a(1-y)} = \frac{a-by}{a} \cdot \frac{c}{1-y} = \frac{c}{a} \cdot \frac{a-by}{1-y} \Rightarrow ac - bcy - ad + ady = 0 \Rightarrow y = \frac{ad-ac}{ad-bc}$ $con\ y \neq 0\ e\ y \neq 1.$

So:

1) $x = \frac{a}{y} = \frac{a}{\frac{ad-ac}{ad-bc}} = \frac{a^2d-abc}{ad-ac} = \boxed{\frac{ad-bc}{d-c}}$
2) $y = \boxed{\frac{a \cdot (d-c)}{ad-bc}}$
3) $z = \frac{b}{a}y = \frac{b}{a} \cdot \frac{ad-ac}{ad-bc} = \frac{abd-abc}{a^2d-abc} = \boxed{\frac{b \cdot (d-c)}{ad-bc}}$
4) $w = \frac{c}{a} \cdot \frac{y}{1-y} = \frac{c}{a} \cdot \frac{\frac{ad-ac}{ad-bc}}{\frac{ad-bc-ad+ac}{ad-bc}} = \frac{c}{a} \cdot \frac{ad-ac}{ac-bc} = \frac{acd-ac^2}{a^2c-abc} = \boxed{\frac{d-c}{a-b}}$

Coming back to original variables:

1) $P(A_+/\mathrm{C}) = \frac{P(A_+)\cdot[P(L_{+on}L_-)-P(L_{+on}A_-)]}{P(A_+)\cdot P(L_{+on}L_-)-P(L_+)\cdot P(L_{+on}A_-)}$

2) $P(L_+/\mathrm{C}) = \frac{P(L_+)\cdot[P(L_{+on}L_-)-P(L_{+on}A_-)]}{P(A_+)\cdot P(L_{+on}L_-)-P(L_+)\cdot P(L_{+on}A_-)}$

3) $P\left[\left(L_{+on}A_-\right)/\mathrm{C}_{A_-}\right] = \frac{P(L_{+on}L_-)-P(L_{+on}A_-)}{P(A_+)-P(L_+)}$

4) $P(\mathrm{C}) = \frac{P(A_+)\cdot P(L_{+on}L_-)-P(L_+)\cdot P(L_{+on}A_-)}{P(L_{+on}L_-)-P(L_{+on}A_-)}$

In conclusion, with the proposed testing we can: a) estimate the sensitivity of ACME and Law testing procedures and hence allow the comparison of their effectiveness; b) directly estimate the residual risk on the released batches; c) estimate the probability of contamination.

APPENDIX

1. Normal distribution

$$X \sim N(\xi, \sigma) \qquad f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{x-\xi}{\sigma}\right)^2}$$

2. LogNormal distribution

$$Y \sim LN\left(e^{\xi+\frac{\sigma^2}{2}}, e^{\xi+\frac{\sigma^2}{2}}\left(e^{\sigma^2}-1\right)\right) \quad \text{se} \quad X = \log Y \sim N(\xi, \sigma)$$

$$\boxed{f(y) = \frac{1}{y} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{\log y-\xi}{\sigma}\right)^2}}$$

$$\begin{aligned}
E(Y) &= \int_0^{+\infty} y \cdot \frac{1}{y} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{\log y-\xi}{\sigma}\right)^2} dy \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{k-\xi}{\sigma}\right)^2} \cdot e^k dk \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{k-\frac{1}{2}\left(\frac{k-\xi}{\sigma}\right)^2} dk \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{k-(\xi+\sigma^2)}{\sigma}\right)^2 + \frac{2\xi\sigma^2+\sigma^4}{2\sigma^2}} dk \\
&= e^{\xi+\frac{\sigma^2}{2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{k-(\xi+\sigma^2)}{\sigma}\right)^2} dk \\
&= e^{\xi+\frac{\sigma^2}{2}}
\end{aligned}$$

$$E\left(Y^2\right)=\int_0^{+\infty} y^2\cdot\frac{1}{y}\cdot\frac{1}{\sqrt{2\pi\sigma^2}}\cdot e^{-\frac{1}{2}\left(\frac{\log y-\xi}{\sigma}\right)^2}dy$$

$$=\int_{-\infty}^{+\infty} e^k\cdot\frac{1}{\sqrt{2\pi\sigma^2}}\cdot e^{-\frac{1}{2}\left(\frac{k-\xi}{\sigma}\right)^2}\cdot e^k dk$$

$$=\int_{-\infty}^{+\infty} e^k\cdot\frac{1}{\sqrt{2\pi\sigma^2}}\cdot e^{-\frac{1}{2}\left(\frac{k-\xi}{\sigma}\right)^2}\cdot e^k dk$$

$$=\int_{-\infty}^{+\infty} e^{2k}\cdot\frac{1}{\sqrt{2\pi\sigma^2}}\cdot e^{-\frac{1}{2}\left(\frac{k-\xi}{\sigma}\right)^2}dk$$

$$=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\cdot e^{\left[-\frac{(k-\xi)^2}{2\sigma^2}+2k\right]}dk$$

$$=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\cdot e^{-\left[\frac{(k-(\xi+2\sigma))^2}{2\sigma^2}\right]}\cdot e^{2\xi+2\sigma^2}dk$$

$$=e^{2\xi+2\sigma^2}$$

$$VAR\left(Y\right)=e^{2\xi+2\sigma^2}-e^{2\xi+\sigma^2}=e^{2\xi+\sigma^2}\left(e^{\sigma^2}-1\right)$$

3. Poisson distribution

$$X\sim Poi\left(\lambda\right)\qquad P\left(X=\frac{x}{\lambda}\right)=\frac{\lambda^x}{x!}e^{-\lambda}$$

4. Poisson LogNormal (PLN) distribution

$$X \sim PLN\left(e^{\xi+\frac{\sigma^2}{2}}, e^{\xi+\frac{\sigma^2}{2}} + e^{2\xi+\sigma^2}\left(e^{\sigma^2}-1\right)\right) \qquad \text{se} \qquad \lambda \sim LN\left(e^{\xi+\frac{\sigma^2}{2}}, e^{\xi+\frac{\sigma^2}{2}}\left(e^{\sigma^2}-1\right)\right)$$

$$f(x) = \int \frac{\lambda^x e^{-\lambda}}{x!} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{\log\lambda-\xi}{\sigma}\right)^2} d\lambda = \frac{1}{x!} \cdot \int \frac{\lambda^{(x-1)} e^{-\lambda}}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{\log\lambda-\xi}{\sigma}\right)^2} d\lambda$$

$$\begin{aligned}
E(X) &= \sum_{x=1}^{\infty} x \cdot \frac{1}{x} \cdot \int_0^{+\infty} \frac{\lambda^{(x-1)} e^{-\lambda}}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{\log\lambda-\xi}{\sigma}\right)^2} d\lambda \\
&= \int_0^{+\infty} \frac{e^{-\frac{1}{2}\left(\frac{\log\lambda-\xi}{\sigma}\right)^2}}{\sqrt{2\pi\sigma^2}} \cdot \left(\sum_{x=1}^{\infty} \frac{\lambda^{(x-1)} e^{-\lambda}}{(x-1)!}\right) d\lambda \\
&= \int_{-\infty}^{+\infty} \frac{e^k}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{k-\xi}{\sigma}\right)^2} dk \\
&= \int_{-\infty}^{+\infty} \frac{e^k}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{k-\left(\xi+\sigma^2\right)}{\sigma}\right)^2} dk \\
&= e^{\xi+\frac{\sigma^2}{2}}
\end{aligned}$$

$$
\begin{aligned}
E\left(X^2\right) &= \sum_{x=1}^{\infty} x^2 \cdot \frac{1}{x} \cdot \int_0^{+\infty} \frac{\lambda^{(x-1)} e^{-\lambda}}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{\log\lambda-\xi}{\sigma}\right)^2} d\lambda \\
&= \sum_{x=1}^{\infty} \frac{x(x-1)+x}{x!} \cdot \int_0^{+\infty} \frac{\lambda^{(x-1)} e^{-\lambda}}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{\log\lambda-\xi}{\sigma}\right)^2} d\lambda \\
&= \sum_{x=1}^{\infty} \frac{x(x-1)}{x!} \cdot \int_0^{+\infty} \frac{\lambda^{(x-1)} e^{-\lambda}}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{\log\lambda-\xi}{\sigma}\right)^2} d\lambda + E(X) \\
&= \int_0^{+\infty} \frac{e^{-\frac{1}{2}\left(\frac{\log\lambda-\xi}{\sigma}\right)^2}}{\sqrt{2\pi\sigma^2}} \cdot \left(\sum_{x=2}^{\infty} \frac{\lambda^{(x-1)} e^{-\lambda}}{(x-2)!} \right) d\lambda + E(X) \\
&= \int_0^{+\infty} \frac{\lambda e^{-\frac{1}{2}\left(\frac{\log\lambda-\xi}{\sigma}\right)^2}}{\sqrt{2\pi\sigma^2}} d\lambda + E(X) \\
&= \int_{-\infty}^{+\infty} \frac{e^k}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{k-\xi}{\sigma}\right)^2} dk + E(X) \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{\left[-\frac{1}{2}\left(\frac{k-\xi}{\sigma}\right)^2 + 2k\right]} dk + E(X) \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{\left[\frac{\left(k-\left(\xi+2\sigma^2\right)\right)^2}{2\sigma^2}\right]} \cdot e^{2\xi+2\sigma^2} dk + E(X) \\
&= e^{2\xi+2\sigma^2} + e^{\xi+\frac{\sigma^2}{2}}
\end{aligned}
$$

$$
VAR(X) = e^{2\xi+2\sigma^2} + e^{\xi+\frac{\sigma^2}{2}} - e^{2\xi+\sigma^2} = e^{\xi+\frac{\sigma^2}{2}} + e^{2\xi+\sigma^2}(e^{\sigma^2} - 1)
$$

BIBLIOGRAPHY

Bar-Oz, B., Preminger, A., Peleg, O., Block, C., & Arad, I. (2001). Enterobacter sakazakii infection in the newborn. *Acta Paediatrica*, 356-358.

Bulmer, M. (1974). On fitting the Poisson lognormal distribution to species–abundance data. *Biometrics*, 101–110.

Caric, M. (1993). *Concentrated and dried dairy products. In: Dairy science and technology handbook.* New York: VCH Publishers Inc.

Codex. (1997). *General Requirements (Food Hygiene). Codex Alimentarius (Supplement to Volume 1B).* Roma.

D. Drudy, N. M. (2006, April 1). Enterobacter sakazakii: an emerging pathogen in powdered infant formula. *Food Safety*, pp. 996-1002.

FAO/WHO. (2004). Enterobacter sakazakii and other microorganisms in powdered infant formula: meeting report. *Microbiological Risk Assessment series 6*.

FAO/WHO. (2006). Enterobacter sakazakii and Salmonella in powdered infant formula: meeting report. *Microbiological Risk Assessment Series 10*.

FAO/WHO. (2008). *Code of hygienic practice for powdered formulae for infants and young children.*

Feller, W. (1968). *An Introduction to Probability Theory and its Applications, Vol. I.* New York: John Wiley & Sons, Inc.

Fiore, A., Casale, M., & Aureli, P. (2004). *Pericoli microbiologici emergent nell'alimentazione del neonato: il caso Enterobacter sakazakii.* Rapporti ISTISAN 04/13.

Gale, P. (2005). Matrix effects, nonuniform reduction and dispersion in risk assessment for Escherichia coli O157. *Journal of Applied Microbiology*, 1640-1648.

Gonzales-Barron, U., Kerr, M., Sheridan, J. J., & Butler, F. (2010). Count data distributions and their zero-modified equivalents as a framework for modelling microbial data with a relatively high occurrence of zero counts. *International Journal of Food Microbiology*, 268-277.

Habraken, C. M. (1986). Management of Salmonella risk in the production of powdered milk product. *Netherlands Milk Diary Journal*, 99-116.

Haight, F. A. (1967). *Handbook of the Poisson Distribution.* New York: John Wiley & Sons.

ICMSF. (1974). *Microorganisms in foods 2. Sampling for microbiological analysis; principles.* Toronto: University of Toronto Press.

ICMSF. (2002). *Microorganisms in foods 7. Microbiological testing in food safety management.* New York: Kluwer Academic/Plenum Publishers.

ILSI Europe. (2010). *Impact of Microbial Distributions on Food Safety.* Brussels: ILSI Europe a.i.s.b.l.

J.J. Farmer III, M. A. (1980, July). Enterobacter sakazakii: A New Species of "Enterobacteriaceae" Isolated from Clinical Specimens. *IJSEM*, pp. 569-584.

Johnson, N., Kotz, S., & KempA.W. (1992). *Univariate discrete distributions.* New York: Wiley.

Jongenburger, I., Bassett, J., Jackson, T., Zwietering, M., & Jewell, K. (2012). Impact of microbial distributions on food safety I. Factors influencing microbial distributions and modelling aspects. *Food control*, 601-609.

Jongenburger, I., Reij, M., Boer, E., Gorris, L., & Zwietering, M. (2011). Actual distribution of Cronobacter spp. in industrial batches of powdered infant formula and consequences for performance of sampling strategies. *International Journal of Food Microbiology*, 62-69.

Jongenburger, I., Reij, M., Boer, E., Gorris, L., & Zwietering, M. (2011). Random or systematic sampling to detect a localised microbial contamination within a batch of food. *Food contol*, 1448-1455.

Kilsby, D., & Baird-Parker, A. (1983). Sampling programmes for microbiological analysis of food. *Food microbiology: advances and prospects. Society for applied Bacteriology Symposium series*, 309-315.

Lai, K. (2001). Enterobacter sakazakii infections among neonates, infants, children and adults: case reports and a review of the literature. *Medicine*, 113-122.

Lambert-Legace, L. (1982). Other milks before and after six months . *Feeding your child*, 70-77.

Legan, D. (2001). Determining the concentration of microorganisms controlled by attributes sampling plans. *Food control*, 137-147.

Liddell, F. D. (1983). Simplified Exact Analysis of Case-Referent Studies: Matched Pairs; Dichotomous Exposure. *Journal of Epidemiology and Community Health*, 82-84.

Lucas, A., & Cole, T. (1990). Breast milk and neonatal necrotising enterocolitis. *Lancet*, 1519-1523.

M. Nazarowec-White, J. F. (1997). Enterobacter sakazakii: a review. *International Jounrnal of Food Microbiology*, pp. 103-113.

Montgomery, D. C. (2009). *Introduction to Statistical Quality Control.* New York: Wiley.

Peter, C. S., Feuerhahn, M., Bohnhorst, B., Schlaud, M., Ziesing, S., Von der Hardt, H., & Poets, C. F. (1999). Necrotising enterocolitis: is there a relationship to specifi cpathogens? *European Journal of Pediatrics*, 67-70.

Rohatgy, V. K., & Md. Eshanes Saleh, A. K. (2000). *An Introduction to Probability and Statistics.* New York: Wiley.

Stoyan, D. a. (1994). *Fractals, Random Shapes and Point Fields.* Chichester: J. Wiley and Sons.

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 209–212.