

A NOVEL RISK-BASED SAMPLING CALCULATOR

Hamid Ragab Orban (DVM)

Specialist, Abu Dhabi Agriculture & Food Safety Authority, P. O. Box 93251, Abu Dhabi, United Arab Emirates

ABSTRACT: *In addition to conventional sample size tables, few formulae were developed according to risk-based approaches and used for calculating the size of food commodity samples for inspection purposes. The current paper hypothesized the dependency of the sample size on both the risk level of the commodity or establishment and the confidence level of sampling. Accordingly, a sampling formula was developed using the commonly used 95% confidence level as fixed attribute. Application of the developed formula on populations and lots selected from three different sources, sample size tables, official authorities' information, and calculated number of units for a fixed lot of hay (2,400 tonnes), revealed no significant difference between the sample sizes at the two selected risk probabilities 0.99 and 0.75. The findings of the current paper strongly support the use of lot units as basis for calculating the sample size rather than the lot weight, further the use of individual lot sizes to calculate sample sizes is more realistic than the common use of groups of lot and populations. The developed formula could confidently be used for calculating sample sizes for commodities of known risk probabilities.*

KEYWORDS: risk analysis, risk-based sampling (RBS), inspection, sample size, confidence level, risk probability

INTRODUCTION

Risk mitigation of different daily life and societal activities by governments and communities is an old practice, however Covello and Mumpower (1985) reported that it was not until the emergence of probability theory in the 17th century that the intellectual tools for quantitative risk analysis became available. The field of risk assessment and risk management is considered a relatively young branch of science with less than 50 years of age (Aven, 2016).

The Codex Alimentarius (2005) defined food inspection as the examination of food or systems for control of food, raw materials, processing, and distribution to verify that they conform to the requirements. Guidelines for sampling food commodities were set (FAO/WHO, 2004). Recently, inspection authorities and control bodies started to adopt the risk analysis approach in planning for and conduction of both inspection and sampling (Maudoux et al, 2006; FAO, 2006; Nychas, Panagou and Mohareb, 2016) thence the term Risk-Based Sampling (RBS) become of common use.

In addition to the conventional sampling tables, some authorities and organizations developed risk-based methodologies to estimate the size of the units of a lot to be sampled or inspected, nevertheless to the best of the author's knowledge, there is no

methodology in place that ensembles all the foodstuff categories and establishments wherein the fixed or dynamic risk probability of a commodity or establishment together with the confidence level of sampling or inspection are compiled to estimate the size of the sample or predict the frequency of inspection. The current paper sets the attributes of a novel risk-based formula to calculate the sample size of a lot or population for inspection purposes including sampling for laboratory analyses.

LITERATURE

Food inspection is a process that aims at minimizing foodborne illnesses by assuring the safety of food or feed for human or animal consumption (FAO/WHO, 2004). The number of visits is an influencing element of inspection despite its type, traditional or risk-based. The scheduled annual inspection visits to establishments ultimately vary between the different authorities with some debates on their effect on the compliance percentage of an establishment. Zaki, Miller, McLaughlin and Weinberg (1977) concluded that the frequency of inspection of food establishments could best be determined by its compliance history whereas Bader, Blonder, Henriksen and Strong (1978) reported that increasing the number of yearly inspection visits from 6 to 12 resulted in no increased performance scores. In contrast, Riben et al. (1994) suggested that restaurants could be inspected routinely at a frequency of one to three inspections per restaurant per year. The importance of training and education of food handlers has early been emphasized by Corber, Barton, Nair and Dulberg (1984). The authors concluded that improving the sanitary situation of establishments might require the adoption of methods other than increasing the frequency of inspections, these could include education of food handlers and amending the Food Premises Regulations.

Moreover, upon applying the modern definition of risk to documented historical practices researchers concluded that the practise of risk mitigation is reported to be an old one. The efforts of governments and communities to mitigate the risks of the different daily life and societal activities were the first areas covered by risk mitigation. Grier (1980) reported that the practices of the Asipu who lived in the Tigris-Euphrates valley 3200 B.C. mark the first noted risk analysis methodology. Henley and Kumamoto (1981) and Grier (1981) and Covello and Mumpower (1985) listed more than a few activities; including food contamination and adulteration; which had been managed in the past according to risk framework.

Further, it was not until the emergence of probability theory in the 17th century that the intellectual tools for quantitative risk analysis became available (Covello and Mumpower,1985). The Codex Alimentarius Commission (2005) adopted the Principles for Food Safety and Risk Analysis to be used in the Codex framework. Additionally, in the last decades inspection authorities and control bodies started to embrace the risk analysis context in planning for both inspection and sampling and the term Risk-Based Sampling (RBS) became a common one. Griffin, Bloem and Hurtado (2020) defined Risk-Based Sampling (RBS) as an inspection design that takes account of the probability of detection to determine the sample size for an inspection. Among the advantages of RBS, the authors listed consistent achievement of a specific level of

detection and confidence and adjustment to correspond to different levels of risk. Furthermore, the outcomes of the risk assessment signify an excellent input for the development of risk management strategies (FAO/WHO, 2006a,b; Kiely et al ,2019).

Koutsoumanis and Aspidou (2016) mentioned that food safety classical hazard-based approaches relying on regulatory inspection and sampling regimes cannot ensure consumer protection sufficiently. The authors concluded that upon the use of risk-based inspection, in high-risk food/hazard combinations, safety will be improved, and the burden of food-borne disease will be reduced.

Hurtado, Griffin and Hong (2020) developed a sample size calculator that estimates the sample size of agricultural products for phytosanitary inspection. Risk-based concepts were used to develop the calculator. The pest detection level which is used in the calculator should be predetermined by the official authority.

The advantages of risk-based inspection over traditional inspection were identified by different authors and organizations with the outcomes solicited according to the perspective of the reporter. Though the objective of sample size calculation is to select the minimum sample size (Kotz et al ,2014) , it is apparent that the required resources to inspect small size samples are in general lesser than those required for large size samples. The reduction in sample sizes would eventually be a systematic outcome upon adoption of risk-based sampling where the focus will be on the high-risk commodities without neglecting the low-risk one. Sareen (2014) compared between the traditional and risk-based inspection, the elements of her comparison are illustrated in Table (1).

Table 1. Traditional versus Risk-Based Inspection

Traditional Inspection	Risk-Based Inspection
Corrective/reactive	Preventive
Inspection planned randomly	Prioritization based on risk factors
Emphasis on product/premises inspection	Emphasis on process inspection/controls in place to address risk factors
Sample collection for assurance purposes	Sample collection for verification purposes

Although examining populations is done by sampling a portion of that population however the objective is usually to draw inference on the whole population (Altman, 2005). Sample confidence level and sample error are commonly used terms in statistics and are jointly used. Lininger and Riemann (2016) defined statistical confidence as the percentage of repeated samples that would indicate the true effect in the population whereas Hazra (2017) reported that selection of the acceptable confidence level is a matter of convention with the 95% level frequently used. Sample size tables are available wherein the size of the sample from a population could be picked using two main variables, the confidence level, and the sampling error. Sample size calculators based on sample size tables are available online.

METHODOLOGY

The aim of the current paper is to set the attributes of a novel risk-based formula to estimate the sample size of a commodity to be inspected and/or sampled.

Hypothesis:

In conventional sampling procedures the higher the level of confidence or the lower the margin of error of the sampling the larger the sample size. The use of confidence in conjunction with the sampling error is not uncommon in statistics. These facts set the basis of the current article hypothesis which is that the final sample size to be inspected is proportional to both the predetermined confidence level and the risk probability of the commodity or establishment.

Sampling confidence level: the confidence level is fixed at the commonly used 95% level.

Risk probability: in reality, the risk probability is estimated considering different factors which include and not limited to the establishment compliance, the commodity inherent risk as well as the risk mitigation factors (Racicot et al, 2020). In the current article the risk probability of the tested commodities was not calculated, as an alternative two risk probabilities were used, these are 0.99 and 0.75 based on random selection for numbers between 0.99 and 0.01.

Sample Size Calculator: The sample size (n) which will be inspected and/or sampled for laboratory analysis is computed according to the following equation:

$$n = \text{ceil} \left(\left| RP \left(\frac{1}{1-S_{err}} \right) * \left(1 + \frac{n0 - n1}{2} \right) \right| \right)$$

Where:

RP : risk probability of the commodity/establishment to be inspected

S_{err} : sampling error, equals to 1 minus the confidence level (CL).

$n0$: corrected number of lot units at RP , equals to $n1 * 1/RP$

$n1$: number of lot units at CL , equals to $N * CL$

N : the total number of units of the lot/population.

CL : predefined inspection confidence level, here it is fixed at 95%.

The absolute arithmetical result of the calculation is rounded up to the next largest integer.

FINDINGS

The developed risk-based sampling equation was implemented on different population/lot sizes to test its effectiveness and accuracy. The population sizes were obtained from three diverse sources, these are sample size tables, official authorities' information, and calculated number of units for a fixed lot of hay (2,400 tonnes).

Sample Size Table:

Nine population sizes were selected from the Sample Size Table of the Research Advisor on the internet. The sample sizes of selected populations were reported at two different confidence levels, 95% and 99%, and two different errors, 5% and 1% for each of the confidence levels. The author's developed risk-based equation was used to calculate the sample size of the selected population sizes at two risk probabilities of 0.99 and 0.75 and a fixed sampling confidence level of 95% which is an integer of the equation. Table (2) shows the different sample sizes as per the selected attributes.

Table 2. Sample Sizes as per a Sample Size Table and the Developed Risk-Based Sample Size Formula

Population Size	Sample Size					
	Sample Table				Current Article Risk-Based Formula	
	95% Confidence level		99% Confidence level		Risk Probability	
	5% Error	1% Error	5% Error	1% Error	0.99	0.75
75	63	74	67	75	3	2
300	169	291	207	295	8	6
800	260	739	363	763	20	16
2,500	333	1,984	524	2,173	60	47
25,000	378	6,939	646	9,972	593	462
100,000	383	8,762	662	14,227	2,370	1,848
250,000	384	9,248	662	15,555	6,185	4,618
500,000	384	9,423	663	16,055	12,369	9,235
2,500,000	384	9,423	663	16,478	59,216	46,171

Initial visualization of the sample sizes obtained from the developed risk-based formula does not indicate a significant difference between the sample sizes at 0.99 and 0.75 risk probabilities. This is further proved by the non-significant statistical finding shown in Table (3). Further, when the sample sizes obtained at a confidence level of 95% and risk probabilities of 0.99 and 0.75 from both the sample size table and the developed risk-based formula were statistically tested for significance the result was not significant ($p = 0.198$) as depicted in Table (3). Despite the non-significant statistical result, inspecting lower number of units of a lot is practically more easier and resource reserving than inspecting bigger number of units of the same lot.

Table 3. Comparison t-test of Sample Size Numbers Obtained from a Sample Size Table and the Developed Risk-Based Sample Size Formula

Variables	df	p-value (Two Tailed)	alpha
Risk-Based Formula Sample Sizes at 0.99 and 0.75 Risk Probabilities	8	0.187	0.05
Sample Sizes of the Risk-Based Formula (at 0.99 and 0.75 Risk Probabilities) and the sample size table (at 95% confidence)	17	0.198	0.05

Food Standards Agency, U.K.:

The Food Standards Agency of the United Kingdom (2016) developed a mycotoxins sampling guidance which laid down the methods of sampling for the official control of the levels of mycotoxins in foodstuffs. The guidance followed two main approaches for setting the sample sizes, these are fixed number of samples for specified ranges of numbers and mathematical equations when and if the lot number is greater or less than a specified number. Two essential points to mention are that the guidance estimates the size of the incremental samples composing the final sample and the second point is that the lot size is the physical weight of the lot rather than the number of the units composing it. The confidence level of sampling is not stated by the guidance.

The sample size of selected different food commodity lots was calculated using the Food Standards Agency guidance as well as the risk-based equation developed by the author at a fixed confidence level of 95% and two risk probabilities of 0.99 and 0.75. Results are depicted in Table (4).

Table 4. Sample Sizes as per the Mycotoxins Sampling Guidance of Food Standards Agency and the Developed Risk-Based Sample Size Formula

Foodstuff	Lot Size		Incremental Sample Size		
	Mycotoxins Sampling Guidance Groups	Selected Size from the Group	Mycotoxins Sampling Guidance	Risk-Based Formula	
				0.99 Risk Probability	0.75 Risk Probability
Cereals	< 50 kg	45 kg	7	2	2
Cereals	> 500 kg - ≤ 1 tonne	980 kg	10	25	19
Cereals	> 3 tonnes - ≤ 10 tonnes	9,800 kg	40	243	182
Dried Fruits	≤ 100 kg	75 kg	10	3	2
Dried Fruits	> 500 kg - ≤ 1000 kg	980 kg	30	25	19
Dried Fruits	> 10 tonnes ≤ 15 tonnes	14,420 kg	100	358	267

Fixed weight hay lot with different units:

The author came across different hay bales physical specifications on the internet, these are mainly the weight and the dimensions. The number of bales of a 2,400 tonnes lot was consecutively calculated for 3 different bale weights of 255, 399 and 574 kilograms. Table (5) shows the calculated sample sizes of the different bale numbers using the developed risk-based formula at a fixed confidence level of 95% and two risk probabilities of 0.99 and 0.75 as well as the sample size table at a confidence level of 95% and a sampling error of 5%.

Table 5. Sample Sizes of a 2,400 tons Hay Lot with Different Bale Weights at Two Risk Probabilities Using the Developed Risk-Based Sample Formula and Sample Size Table

Bale Weight (kg)	Calculated Number of Bales	Sample Size		
		0.99 probability risk	0.75 probability risk	Sample Size Table
255	9,411	234	175	370
399	6,015	150	112	365
574	4,181	104	78	357

DISCUSSION

The aim of the current paper is to set the attributes of a sampling formula according to the risk-based approach that could easily be used to estimate the sample size of a lot under inspection. The paper hypothesized that sample size is proportional to both the confidence level of sampling and the risk level of the commodity or population under inspection.

Upon application of the developed risk-based sampling formula on different populations selected from a sample size table, the results of the formula were lower than that of the table, however above a population size of 100,000 size the formula results were higher than that of the table and unlike sample size tables are directly proportional to the population as Table (2) shows. The discrepancy between the sample sizes of the developed risk-based formula and the sample size table for populations more than 100,000 could be attributed to the fact that sample sizes are fixed for population sizes of 100,000 and more in the sample size tables. The difference in sample sizes from a lot with different risk probabilities was not statistically significant as Table (3) shows, however the relatively large sample size at 0.99 risk probability compared to that at 0.75 risk probability strongly support the perception that risk-based sampling (RBS) is more risk focussed (Sareen, 2014).

When the current risk-based formula is applied to different lot sizes of foodstuff enlisted in the mycotoxins sampling guidance of the Food Standards Agency (FSA), the results of the formula were lower for lot sizes less than 100 and higher for lots bigger than 100 as Table (4) illustrates. Among the advantages of the current developed risk-based formula over the mycotoxin sampling guidance of the FSA is that the formula considers each lot independently unlike the guidance wherein foodstuff are grouped based on the lot size and sample sizes are fixed to the groups in an analogous manner to sample size tables. Moreover, in the guidance the lot size is expressed in weight rather than number of units. The guidance calculates the size of the incremental sample that composes the final sample, this limits the use of the guidance to foodstuff only whereas the attributes of the current developed risk-based formula render it suitable for estimating the sample size for any commodity or establishment for inspection purposes.

It is not quite uncommon that lots of the same commodity and weight could have different unit numbers according to the variation in the physical weight of the units comprising the lot. Upon virtual calculation of the units of the 2,400 tonnes lot according to different unit physical weights, the results revealed that the same lot has 3 different unit numbers. Calculating the sample size of the original lot (2,400 tonnes) using the sample size table and the FSA guidance resulted in 332 and 100 incremental sample numbers, respectively. This result together with the results reported in Table (5) strongly support the necessity of considering the number of units composing a lot in calculating the sample size rather than the total physical weight of the lot.

The specificity of the formula developed by Hurtado, Griffin and Hong (2020) to sample for pest inspection limits its use to plant commodities whereas in contrast, the

current developed risk-based formula could easily be used for commodities as well as establishments.

Implication to research and practice

The developed risk-based formula was tested against sample size tables, official authorities' information, and calculated number of units for a fixed lot of hay (2,400 tonnes). The results of the developed risk-based formula documented in the current paper indicate its suitability for sampling any population to which the risk analysis concept is applicable. Further, as the compliance result of inspection is changing thence the risk level of a commodity or an establishment could change. This stipulates the critical need of developing a real-time or dynamic tool that captures the changes in any of the risk attributes of the commodity and/or establishment to immediately update its risk level.

CONCLUSION

The aim of the current paper is to develop and test the efficiency of a sampling formula according to the risk-based concept. From the preceding results and discussion, it is justifiable to concluded that sample size tables neglect a major contributor to the sample size which is the risk attributes of the sampled commodity. Furthermore, sample size tables do not consider the increase in the population under inspection when its number exceeds 100,000. An important advantage of the current developed risk-based sampling formula over all the tested sampling tools is that it considers the risk level of the sampled commodity which will, regardless of the non-significant statistical results obtained, sequentially impact the sample size. Moreover, the application of the developed risk-based formula clearly emphasized the inevitability of considering the lot size in any sampling tool specially for populations larger than 100,000.

FUTURE RESEARCH

The outcomes of the current paper necessitate testing the developed risk-based formula at sampling or inspection confidence levels other than 95%. There is a need to manipulate the formula in a way that makes it suitable for predicting the number of inspection visits an establishment should be visited in a year or in other words to use the formula in developing annual risk-based inspection plans.

REFERENCES

- Altman, D.G. (2005). Why we need confidence intervals. *World J Surg*, 29:554-6.
- Aven, T. (2016). Risk assessment and risk management: Review of recent advances on their foundation, *European Journal of Operational Research* , Volume 253, Issue1,2016,Pages1-13,ISSN0377-2217, <https://doi.org/10.1016/j.ejor.2015.12.023>
- Bader, M., Blonder, E., Henriksen, J. and Strong, W. (1978). A study of food service establishment sanitation inspection frequency. *American Journal of Public Health*, 68(4), 408-410
- Codex Alimentarius Food Import and Export Inspection and Certification Systems (2005). Codex Alimentarius Commission, Joint FAO/WHO Food Standards

- Programme, Rome, Italy (2005). <http://www.fao.org/3/y6396e/Y6396E00.htm#TOC>
- Corber, S., Barton, P., Nair, R.C. and Dulberg, C. (1984). Evaluation of the effect of frequency of inspection on the sanitary conditions of eating establishments. *Canadian Journal of Public Health*, 76, 434-438
- Covello, V.T. and Mumpower, J. (1985). Risk Analysis and Risk Management: An Historical Perspective., *Risk Analysis* 5(2), 103–120. doi:10.1111/j.1539-6924.1985.tb00159.x
- Food and Agriculture Organization of the United Nations/World Health Organization (FAO/WHO) (2006a): Food Safety Risk Analysis. A Guide for National Food Safety Authorities. FAO Food and Nutrition Papers 87; 2006
- Food and Agriculture Organization of the United Nations/World Health Organization (FAO/WHO) (2006b): The Use of Microbiological Risk Assessment Outputs to develop Practical Risk Management Strategies: Metrics to improve Food Safety. FAO/WHO Expert Meeting; 2006.
- Food and Agriculture Organization of the United Nations/World Health Organization (FAO/WHO). 2004. Assessing Quality and Safety of Animal Feeds Food and Agriculture Organization of the United Nations/World Health Organization (FAO/WHO): (2004).
- The Codex General Guidelines on Sampling – CAC/GL 50-2004; Food Standards Agency. Mycotoxins Sampling Guidance, Reviewed October 2016. <https://www.food.gov.uk/sites/default/files/media/document/mycotoxins-sampling-guidance.pdf>
- Grier, B. (1980) One Thousand Years of Mathematical Psychology (Paper presented at Society for Mathematical Psychology Convention, Madison, Wisconsin, 1980).
- Grier, B. (1981). The Early History of the Theory and Management of Risk (Paper presented at the Judgment and Decision Making Group Meeting, Philadelphia, Pennsylvania, 1981)
- Griffin, R.; Bloem, S. and Hurtado, M. (2020). RBS: Frequently Asked Questions. In: Risk-Based Sampling (RBS) Manual -Part1, a multi-authored manual on the what, why and how of RBS. North American Plant Protection Organization (NAPPO), 2020. https://nappo.org/application/files/7916/1471/8930/Manual_RBS_Part_I_Eng_Final20200224.pdf
- Hazra, A. (2017). Using the confidence interval confidently. *Journal of thoracic disease*, 9(10), 4125–4130. <https://doi.org/10.21037/jtd.2017.09.14>
- Henley, E.J. and Kumamoto, H. (1981). Reliability engineering and risk assessment, Prentice Hall.
- Hurtado, M., Griffin, B. and Hong, S. (2020). Sample Size Calculator, In: Risk-Based Sampling (RBS) Manual -Part1, a multi-authored manual on the what, why and how of RBS. North American Plant Protection Organization (NAPPO) ,2020. https://nappo.org/application/files/7916/1471/8930/Manual_RBS_Part_I_Eng_Final20200224.pdf
- Kiely, S., Karen LeGrand, K., Chuong, T., Kiesel, K., Trexler, C.J., Buntong, B., and Young, G.M. (2019). Perceptions of Risk and Risk Management Strategies: Identifying Alternative Strategies to Promote Smallholder Vegetable Production in Cambodia *International Journal of Agricultural Extension and Rural Development Studies* Vol.6, No.5, pp.21-43.

- Kotz, S., Read, C.B., Balakrishnan, N., Vidakovic, B., Johnson, N.L., Wang, H. and Chow, S.-C. (2014). Sample Size Calculation for Comparing Proportions. In Encyclopedia of Statistical Sciences (eds S. Kotz, C.B. Read, N. Balakrishnan, B. Vidakovic and N.L. Johnson). <https://doi.org/10.1002/0471667196.ess7211>
- Koutsoumanis, K.P; Aspidou, Z. (2016). Moving towards a risk-based food safety management. *Current Opinion in Food Science*, 12, 36–41. doi: 10.1016/j.cofs.2016.06.008
- Lininger, M., and Riemann, B.L. (2016). Statistical Primer for Athletic Trainers: Using Confidence Intervals and Effect Sizes to Evaluate Clinical Meaningfulness. *J Athl Train*; 51 (12): 1045–1048. doi: <https://doi.org/10.4085/1062-6050-51.12.14>
- Maudoux, J.-P.; Saegerman, C.; Rettigner, C.; Houins, G.; Van Huffel, X. and Berkvens, D. (2006). Food safety surveillance through a risk based control programme: Approach employed by the Belgian Federal Agency for the safety of the food chain. *Veterinary Quarterly*, 28(4), 140–154. doi:10.1080/01652176.2006.9695220
- Nychas, G.-J.E., Panagou, E.Z. and Mohareb, F. (2016). Novel Approaches for Food Safety Management and Communication, *Current Opinion in Food Science*, <http://dx.doi.org/10.1016/j.cofs.2016.06.005>
- Racicot, M, Comeau, G, Tremblay, A, et al. (2020). Identification and selection of food safety-related risk factors to be included in the Canadian Food Inspection Agency's Establishment-based Risk Assessment model for Hatcheries. *Zoonoses Public Health*. 67: 14–24. <https://doi.org/10.1111/zph.12650>
- Riben, P., Mathias, R.G., Wiens, M., Cocksedg, W., Hazlewood, A., Kirshner, B. and Pelton, J. (1994). Routine restaurant inspections and education of food handlers: Recommendations based on critical appraisal of the literature and survey of Canadian jurisdictions on restaurant inspections and education of food handlers. *Canadian Journal of Public Health*, 85(Suppl.1), S67-70.
- Sareen, S. (2014). Introduction to risk-based inspection. Asian-Pacific Economic Cooperation (APEC) Workshop on Improved Food Inspection Capacity Building Based on Risk Analysis, Seoul, Korea 21-23 May 2014.
- Zaki, M., Miller, G.S., McLaughlin, M.D. and Weinberg, S.D. (1977). A progressive approach to the problem of foodborne infections. *American Journal of Public Health*, 67(1), 44-49.