International Journal of Food Microbiology 369 (2022) 109616

Contents lists available at ScienceDirect

International Journal of Food Microbiology

journal homepage: www.elsevier.com/locate/ijfoodmicro

A quantitative risk metric to support individual sanitary measure reviews in international trade

Eric D. Ebel [a], Janell R. Kause [a], Michael S. Williams [a,*], Wayne D. Schlosser [a], Stephanie Defibaugh-Chavez [b], Berhanu Tameru [a]

[a] Office of Public Health Science, Food Safety and Inspection Service, United States Department of Agriculture, 2150 Centre Avenue, Building D, Fort Collins, CO 80526, United States of America

[b] Office of Policy and Program Development, Food Safety and Inspection Service, United States Department of Agriculture, 2150 Centre Avenue, Building D, Fort Collins, CO 80526, United States of America

ARTICLE INFO

Keywords:
Risk equivalence
Sample size

ABSTRACT

In order for the United States Department of Agriculture's (USDA) Food Safety and Inspection Service (FSIS) to make an equivalence determination for a foreign meat, poultry or egg products inspection procedure that differs from FSIS inspection procedures (an Individual Sanitary Measure or ISM), a country must demonstrate objectively that its food safety inspection system provides the same level of public health protection as the FSIS inspection system. To evaluate microbiological testing data that such countries may submit to this end, we present a possible risk metric to inform FSIS's assessment of whether products produced under an alternative inspection system in another country pose no greater consumer risk of foodborne illness than products produced under FSIS inspection. This metric requires evaluation of prevalence estimates of pathogen occurrence in products for the foreign country and the U.S. and determining what constitutes an unacceptable deviance of another country's prevalence from the U.S. prevalence, i.e., the margin of equivalence. We define the margin of equivalence as a multiple of the standard error of the U.S. prevalence estimate. Minimizing the margin of equivalence ensures the maximum public health protection for U.S. consumers, but an optimum choice must also avoid undue burden for quantitative data from alternative inspection systems in the foreign country. Across a wide range of U.S. prevalence levels and sample sizes, we determine margin of equivalence values that provide high confidence in conclusions as to whether or not the country's product poses no greater risk of foodborne illness from microbiological pathogens. These margins of equivalence can be used to inform FSIS's equivalence determination for an ISM request from a foreign country. Illustrative examples are used to support this definition of margin of equivalence.

This approach is consistent with the World Trade Organization's concept of risk equivalence and is transparent and practical to apply in situations when FSIS makes an equivalence determination for an ISM requested by a foreign country.

1. Introduction

The United States Department of Agriculture's (USDA) Food Safety and Inspection Service (FSIS) is responsible for ensuring that the supply of meat, poultry and egg products are safe and properly labeled and packaged. FSIS allows products to be imported into the U.S. (after inspection at the port of entry) from countries that have equivalent food safety inspection systems (FSIS, 2012). In keeping with the World Trade Organization's (WTO) *Agreement on the Application of Sanitary and Phytosanitary Measures* principle of equivalence (WTO, 1998), FSIS considers an equivalent food safety inspection system as one that provides the same level of public health protection as the FSIS inspection system for meat, poultry and egg products. FSIS' decision to allow product from another country to be imported into the U.S. ultimately depends on determining that the country implements a food safety inspection system that results in product that is at least as safe, i.e., poses no greater public health risk, as similar product produced under FSIS inspection.

* Corresponding author.
E-mail address: mike.williams@fsis.usda.gov (M.S. Williams).

https://doi.org/10.1016/j.ijfoodmicro.2022.109616
Received 10 September 2021; Received in revised form 28 January 2022; Accepted 4 March 2022
Available online 11 March 2022

Regulatory requirements for equivalence of imported products are set forth in Title 9 of the Code of Federal Regulations (CFR) – §327.2 (for meat products), §381.196 (for poultry products), §557.2 (for fish of the Order Siluriformes) and §590.910 (for egg products) (FSIS, 2019). FSIS organizes these requirements into six components when evaluating a country's food safety system: (1) Government Oversight, (2) Government Verification of Food Safety and Other Consumer Protection Requirements, (3) Government Sanitation Verification, (4) Government Hazard Analysis and Critical Control Point (HACCP) System Verification, (5) Government Chemical Residue Program, and (6) Government Microbiological Pathogen and Process Control Programs.

Countries are not required to develop and implement the same inspection procedures as FSIS. When a country implements inspection procedures that are fundamentally different from FSIS' approach to food safety inspection, the country must objectively demonstrate that its alternative approach provides the same level of public health protection as the FSIS food safety system if it is to export meat, poultry, or egg products to the U.S. FSIS evaluates the country's request for alternative inspection procedures through Individual Sanitary Measure (ISM) equivalence determinations (FSIS, 2020a). Objective evidence includes laws, regulations and procedures that support the country's alternative approach. A country may also submit scientific studies, inspection data, or microbiological testing data in support of their request and FSIS may request specific inspection or testing data when additional evidence is needed to demonstrate the comparable performance of the country's food safety inspection system. These data, if determined to be representative and comparable to FSIS testing data, can augment an assessment of the public health risk posed by a country's product relative to that of product produced in the U.S. The outcome of a microbiological assessment can be used, in conjunction with other information provided by the exporting country, to inform FSIS' determination of ISM equivalence.

Currently, there is no consistent approach for using data to quantitatively assess the performance of a country's food safety control system in mitigating consumer risk of foodborne illness. Many have proposed risk-based approaches to guide national food safety control strategies (Christensen et al., 2013), but proposals for broader use of these decision-support tools to inform trade decisions has not been widely adopted (FAO/WHO, 2016). This may be due, in part, to the complexity and substantive data needs of traditional risk assessment approaches and difficulty in obtaining relevant data, including national foodborne illness surveillance data (Gallagher et al., 2013; Havelaar et al., 2015; Li et al., 2019). To overcome these barriers and increase the application of risk assessments, some countries have begun developing risk-based tools that are less data-intensive and primarily rely on available microbiological testing data (e.g., U.S. Food and Drug Administration's (FDA) iRisk and the swift Quantitative Microbiological Risk Assessment Tool) (Bassett et al., 2012; Chardon and Evers, 2018; Dearfield et al., 2014; Evers and Chardon, 2010; FAO/WHO, 2021; FDA, 2012). Nevertheless, there is no practical risk assessment tool that countries use to evaluate quantitatively the comparative performance of food safety inspection systems. In this paper, we discuss a quantitative risk metric that could inform FSIS' ISM equivalence determinations. The risk metric provides a consistent, transparent and practical approach to evaluate quantitative data that countries with alternative inspection procedures could submit as evidence of the level of public health protection achieved by their food safety inspection system. We demonstrate the use of a non-inferiority statistical test in this risk metric to assess, with clear criteria and a high level of confidence, the public health risk of meat, poultry, or egg products produced under alternative inspection in another country relative to those produced under FSIS inspection. This assessment, along with other laws, regulations, procedures and supporting documentation submitted by the foreign country, is being considered as the basis for FSIS's equivalence determination whenever a foreign country requests an ISM. We also explore the effects of statistical criteria used in this risk metric on the level of evidence needed and our confidence in correctly assessing the relative safety of a country's product.

2. Methods

2.1. Measuring the performance of a food safety system: product pathogen prevalence

The level of public health protection a country's food safety inspection system provides its own consumers is reflected in estimates of its burden of foodborne disease (Havelaar et al., 2015; Kirk et al., 2015). Analysis of these estimates suggests that each country provides its consumers differing levels of protection from foodborne illness associated with particular pathogens in its food supply, although differences in surveillance systems may confound comparisons between countries (Havelaar et al., 2008; Havelaar et al., 2004). Many countries do not have sufficient foodborne illness surveillance data to make an assessment and limitations in food source attribution methods also make it difficult to assess the comparative performance of food safety systems (Batz et al., 2005; NRC, 2009).

One direct measure of the performance of a food safety inspection system that reflects the effectiveness of a country's food safety policies, oversight and enforcement, given the extent of foodborne hazards present at preharvest, is the prevalence of a foodborne pathogen in product at the point of production. Prevalence data can be collected by countries using comparable sample collection and equivalent test methods (FSIS, 2020b; ISO, 2017).

Previous studies have shown that it is reasonable to assume that the fraction of product samples that are positive for a specific pathogen is proportional to the probability of foodborne illness among consumers of that product (Bartholomew et al., 2005; Ebel and Williams, 2015; Ebel et al., 2012; Williams et al., 2011). The probability of foodborne illness per serving, $P(ill)$, represents the risk of illness for a product-pathogen pairing, while the fraction of samples positive for a specific pathogen is an estimate of pathogen prevalence of contaminated product. Formally, we can define the risk of foodborne illness as:

$$P(ill) = \int_0^\infty R(d)f(d)\partial d \tag{1}$$

where d is the dose of pathogen consumed; R is a dose-response function that predicts the probability of illness for a given d in a serving; and $f(d)$ is a probability function that describes how dose varies across all product servings consumed. If the distribution of dose is partitioned into the fraction where $d = 0$ and the remaining part where $d > 0$, then we can also describe the risk of foodborne illness as:

$$P(ill) = P(ill|exp) \times P(exp),$$

where $P(ill|exp)$ is the probability of foodborne illness conditioned on consumer exposure to product containing doses of one or more pathogen cells (essentially Eq. (1) with the lower limit of integration truncated at 1) and $P(exp)$ is the fraction of product servings with pathogen levels greater than zero.

Although an exposure distribution at consumption (i.e., $f(d)$) cannot be observed directly, estimates of the prevalence of pathogen-positive food samples at a point before consumption (e.g., end of processing or slaughter) are practical and these estimates are the basis of a so-called "prevalence-based" risk assessment model (Williams et al., 2011).

2.2. Measurable comparative performance of different food safety systems

In its development of a prevalence-based risk assessment model to estimate the predicted effectiveness of a policy option to reduce the risk of foodborne illness from a baseline level, FSIS examined the conditions

where:

$$\frac{P_{new}(test+)}{P_{baseline}(test+)} \cong \frac{P_{new}(ill)}{P_{baseline}(ill)} \quad (2)$$

where $P_*(test+)$ is the prevalence of a pathogen in product samples, usually collected at the end of production, that tested positive before (baseline) or after (new) implementation of a policy option (Ebel and Williams, 2015). The relationship in Eq. (2) was determined to hold unless the quantity of food analyzed per sample was so small that microbiological hazard surveillance would only detect positive samples from product with an exceptionally high concentration of the target pathogen.

Similar prevalence levels in two populations imply similar risk if the underlying concentration distribution is similar and describes similar pathogens (i.e., similar strains of the pathogens in both populations). If a lognormal distribution describes how concentration varies across units produced by an industry, then prevalence will be the share of those units above the limit of detection (LOD) for a qualitative test. Although it is possible that the same prevalence can result despite the relative frequencies of concentrations above the LOD not being equivalent exactly, we have not found such differences to be dramatic enough to result in substantial risk differences. In Ebel and Williams (2015), we explored how changes to the concentration distribution at the end of slaughter and processing affect observed prevalence and estimated risk per serving. That concentration distribution reflects the incoming variability in contamination among carcasses and the variable effectiveness of interventions during processing. In log10 units, that distribution is generally symmetric and defined by a mean and standard deviation. Relative to a reference distribution, the same prevalence might be observed in another population either by increasing the mean and reducing the standard deviation, or vice versa. Based on our experience, however, the standard deviation of this distribution is not dramatically different across a wide range of pathogen-pairs (e.g., generally this standard deviation ranges between 1.2 and 2 log10 units, see Ebel and Williams (2015) and Williams et al. (2015) for lognormal parameters for pathogen product pairs ranging from *Campylobacter*-chicken, *Salmonella*-pork and *E. coli* O157-beef). Therefore, absent extreme flexibility to that underlying distribution, it is reasonable to assume that populations with similar prevalence levels are not likely to have dramatically different underlying concentration differences.

Similar reasoning supports that the change in pathogen prevalence of positive samples associated with a policy is proportional to the change in the risk of illness (i.e., $P_{new}(test+) - P_{baseline}(test+) \propto P_{new}(ill) - P_{baseline}(ill)$). This equation is adaptable to other risk management decisions, such as assessing the difference in the risk of foodborne illness among U.S. consumers who eat foods sourced from another country with alternative inspection procedures compared to those who eat similar product produced under FSIS inspection.

When assessing a U.S. product-pathogen pair, assume that $P_{U.S.}(test+)$ is the prevalence of positives among samples collected at the final point that FSIS inspects the product. If comparable and contemporaneous sampling evidence is available from a country seeking an ISM review, then interest lies in measuring the difference in observed prevalence of pathogen-positive samples between the countries. This measure can be used to assess the difference in consumer risk of foodborne illness between the two different approaches to inspection, i.e.,

$$P_{foreign}(test+) - P_{U.S.}(test+) \propto P_{foreign}(ill) - P_{U.S.}(ill). \quad (3)$$

Nevertheless, both prevalence values are estimates from sampling data and this imposes a probability distribution on the difference in the risk of foodborne illness among U.S. consumers.

Conceptually, if the food product from an exporting country has the same or lower occurrence of a pathogen as similar product produced in the U.S., then it poses no greater risk of foodborne illness than the U.S. product. This is because imported and domestically produced meat, poultry, and egg products typically are not differentiated in the U.S. market and typically are handled, prepared and consumed the same way by U.S. consumers (Kuchler et al., 2010). That is, imported meat, poultry, and egg products have the same downstream exposure pathways (i.e., retail and consumer food handling practices) and, given the same consumer population, the same dose-response relationships, as similar product produced in the U.S. under FSIS inspection.

2.3. Evaluating food safety system performance: non-inferiority test

Our definition of a product intended for export to the U.S. as posing "no greater risk" is that the country's mean pathogen prevalence, $P_{foreign}(test+)$, is as low as, or lower than, the mean pathogen prevalence in FSIS-inspected products produced in the U.S., $P_{U.S.}(test+)$.

Initially, we propose a null hypothesis that the pathogen prevalence in a country's product intended for export to the U.S. is greater than, by some margin, the observed pathogen prevalence in products produced under FSIS inspection. If the country's sampling data provides evidence inconsistent with this null hypothesis (i.e., the country's pathogen prevalence in products intended for export to the U.S. is below some margin), then we can conclude that the country's product pathogen prevalence, and the corresponding public health risk associated with an alternative inspection system, is no greater than that posed by similar product produced under FSIS inspection. In statistical parlance, this hypothesis is considered testing for "non-inferiority," which is routinely used in the medical and pharmaceutical disciplines (Kaul and Diamond, 2006; Schumi and Wittes, 2011; Walker and Nowacki, 2011).

Formally, the null hypothesis for this non-inferiority assessment is $P_{foreign} - P_{U.S.} \geq \delta$, where we drop the *test*+ qualifier for notational brevity and δ is technically referred to as the "margin of equivalence," such that differences between a country's and U.S. product pathogen prevalence greater than δ are unacceptable generally. The alternative hypothesis, $P_{foreign} - P_{U.S.} < \delta$, implies that a country's pathogen prevalence ($P_{foreign}$) is "non-inferior" to the FSIS-inspected U.S. product pathogen prevalence ($P_{U.S.}$).

To test for non-inferiority, we can calculate the following statistic:

$$Z = \frac{\widehat{P}_{foreign} - \widehat{P}_{U.S.} - \delta}{\sigma_{difference}} \quad (4)$$

where Z is a random variable whose distribution is assumed to follow a standard normal distribution ($Z \sim Normal(0,1)$), the prevalence values are estimates from pathogen testing data and $\sigma_{difference}$ is the standard error of $P_{foreign} - P_{U.S.}$ under the null hypothesis. From this statistic, the p-value is the probability of a value of Z more extreme than that observed. If the p-value is less than some acceptable error (e.g., 0.05), then the null hypothesis that $P_{foreign} - P_{U.S.} \geq \delta$ is rejected.

Because this Z statistic applies to the null hypothesis, the standard error in its denominator should reflect the expected difference in the probability of positive samples in the two countries when the null hypothesis is true. Therefore,

$$\sigma_{difference} = \sqrt{\frac{(\widehat{P}_{U.S.} + \delta)(1 - (\widehat{P}_{U.S.} + \delta))}{n_{foreign}} + \frac{\widehat{P}_{U.S.}(1 - \widehat{P}_{U.S.})}{n_{U.S.}}} \quad (5)$$

where we assume a country's product pathogen prevalence is larger than the U.S. product pathogen prevalence by a prescribed marginal difference. Practically, non-inferiority testing is often conducted by calculating the confidence limits of the difference in prevalence levels. In this case, non-inferiority is concluded if:

$$\left(\widehat{P}_{foreign} - \widehat{P}_{U.S.}\right) + Z_{1-\alpha} \times \sigma_{difference} < \delta \quad (6)$$

Eq. (6) highlights two factors that reflect the risk tolerance of risk managers. The first factor is the Z-value, which is the $1 - \alpha$ quantile from a standard normal distribution. The α chosen is the Type 1 error; it is

conventionally set at values of 0.05 or less. The second factor is the margin of equivalence and does not have a default setting. Its value can be considered a risk management decision.

2.4. Setting a margin of equivalence: benchmarking against the FSIS food safety inspection system

In applications that examine the non-inferiority of new therapeutics or diagnostics relative to an active control, the margin of equivalence (synonymous with "non-inferiority margin" or "equivalence margin") describes the absolute difference in outcomes considered unacceptable. In these other applications, slightly worse therapeutic or diagnostic performance of a new drug or test could be acceptable based on economic or other clinical considerations (e.g., side effects). In addition, the effectiveness of a therapeutic drug is often studied relative to a placebo; the magnitude of a standard treatment's effect relative to a placebo might be a starting point for considering how much worse a new treatment might perform (e.g., the new treatment should perform better than a placebo but could perform somewhat worse than the standard treatment) (Schumi and Wittes, 2011).

In the context of informing food safety decisions in international trade, setting a margin of equivalence is not obvious. For our purposes, the margin of equivalence describes an unacceptable deviance of another country's prevalence from the U.S. prevalence, such that FSIS might consider an unfavorable equivalence decision for an ISM request for an alternative inspection procedure. The unacceptable deviance should not be so large as to allow product into the U.S. that is substantially more contaminated than similar product produced under FSIS inspection. Therefore, the choice for the margin of equivalence should not allow higher prevalence levels in another country's product that are rarely or never observed in the applicable U.S. industry. The margin of equivalence also should not be so small that it imposes an undue burden on a country to objectively demonstrate that its product is as safe as similar FSIS-inspected U.S. product. For example, setting the margin of equivalence such that another country would need to collect more samples than collected in the U.S., even though the proportion of positive samples is the same for both countries, could be considered an undue burden on the other country.

Given a wide range in consumer risk of illness estimates among FSIS-inspected products (Hsi et al., 2015), an absolute (or "fixed") definition for the margin of equivalence is not feasible. For example, annual FSIS product sampling data illustrates that the prevalence of *Salmonella* among ground chicken, beef, pork and turkey samples ranges from 3% to 30% depending on the commodity (Sampling Results for FSIS Regulated Products | Food Safety and Inspection Service (usda.gov)). Among comminuted chicken samples, prevalence of *Salmonella* and *Campylobacter*-positives is nearly 30% and 5%, respectively. The occurrence of *Escherichia coli* O157:H7 isolated from ground beef has been between 0.1 and 0.2% in FSIS-sampled product. Statistical uncertainty about U.S. pathogen prevalence is not consistent across FSIS-regulated products, partly because the number of FSIS-collected samples is not the same for each product. Also, for a particular product-pathogen pair, prevalence is expected to decrease across time as the U.S. strives to achieve a higher level of protection based on its national food safety goals (i.e., Healthy People food safety goals (HHS, 2020)).

Instead, a relative, or "benchmark", definition of the margin of equivalence seems appropriate when comparing the pathogen prevalence of another country to the pathogen prevalence in U.S. product. Such a benchmark should adjust with (a) the pathogen prevalence observed in the U.S. for a given time period and (b) the number of samples used to estimate the U.S. prevalence. Because the standard error of the U.S. estimate is a function of both the U.S. prevalence and its sample size, a benchmark margin of equivalence based on this metric is self-adjusting. When the U.S. prevalence estimate is precise (i.e., its standard error is small), then the unacceptable prevalence is closer to the U.S. estimate and the margin of equivalence is smaller, i.e., the U.S. requires smaller deviations from its prevalence when it is very confident about its prevalence. When the U.S. prevalence estimate is imprecise, i.e., its standard error is large, then what constitutes an unacceptably large prevalence will be farther from the estimate, i.e., the U.S. tolerates larger deviations from its prevalence because there is more uncertainty around its prevalence.

Interest lies in determining a multiple (m) of the standard error of the U.S. prevalence ($\sigma_{U.S.}$) that would define the margin of equivalence (i.e., $\delta = m \times \sigma_{U.S.}$). The standard error is a function of the observed prevalence (e.g., $P_{U.S.}$) and the number of samples used to estimate that prevalence ($n_{U.S.}$); therefore, $\sigma_{U.S.} = \sqrt{\left(\frac{\widehat{P}_{U.S.}[1-\widehat{P}_{U.S.}]}{n_{U.S.}}\right)}$.

An undue burden would be placed on another country if m were chosen such that the country could not demonstrate its non-inferiority to the U.S. despite generating the same testing results in its population as those observed in FSIS testing (i.e., $\widehat{P}_{foreign} = \widehat{P}_{U.S.}$ and $n_{foreign} = n_{U.S.}$). If we were testing that $\widehat{P}_{foreign} = \widehat{P}_{U.S.}$, then the standard error is $\sqrt{2} \times \sigma_{U.S.}$. In this case, we would construct Eq. (6) as $\left(\widehat{P}_{foreign} - \widehat{P}_{U.S.}\right) + Z_{1-\alpha} \times \sqrt{2}\sigma_{U.S.} < m\sigma_{U.S.}$ and conclude that m must be greater than $\sqrt{2} \times Z_{1-\alpha}$.

For a standard value of $Z_{1-\alpha} = 1.96$ (applicable to a 95% confidence interval or a one-sided hypothesis when $\alpha = 0.025$), this constraint implies that the margin of equivalence should be greater than 2.8 standard error units based on FSIS sampling results (i.e., $m = 2.8$). Anything less than this margin of equivalence would result in a conclusion that another country's product produced under an alternative inspection system may pose a greater risk than similar FSIS-inspected product (i.e., the product is deemed "inferior" with respect to pathogen prevalence), despite evidence that the pathogen prevalence in the country's product is exactly the same as observed in U.S. product.

To explore the appropriateness of the margin of equivalence, we examine the implied unacceptable prevalence ($P_{U.S.} + \delta$) relative to the estimated variability in within-establishment product pathogen prevalence in the U.S. As outlined above, the margin of equivalence should not allow pathogen prevalence levels in another country that are exceptionally large relative to levels observed in the applicable U.S. industry.

In its development of pathogen reduction performance standards, FSIS estimates the distribution of within-establishment product pathogen prevalence for the targeted industry. A beta-binomial fitting algorithm is used that considers the binomial sampling error in data generated from an underlying beta distribution that reflects how the probability of positive sampling results varies across an industry (Williams et al., 2013). The maximum likelihood estimate for the beta distribution is used here to examine different choices of the margin of equivalence. Parameters of some of these beta distributions have been summarized in the peer- reviewed literature (Lukicheva et al., 2016).

2.5. Setting a margin of equivalence: confidence in assessing comparative product safety

The choice of a margin of equivalence in the non-inferiority decision should account for the desired performance characteristics of the evaluation, i.e., the level of confidence in correctly assessing the comparative safety of another country's product. Sensitivity is defined as the probability of correctly assessing whether a country's product produced under an alternative inspection system poses a greater public health risk, i.e., is assessed as being inferior when it is truly inferior. The complement of sensitivity is the Type 1 error of the decision (Fig. 1); the probability of mistakenly determining that another country's product does not pose a greater risk than FSIS-inspected U.S. product when it actually poses a greater risk, i.e., is "inferior."

Specificity is defined as the probability of correctly assessing that another country's product produced under an alternative inspection system poses no greater risk, i.e., it is non-inferior to FSIS-inspected U.S.

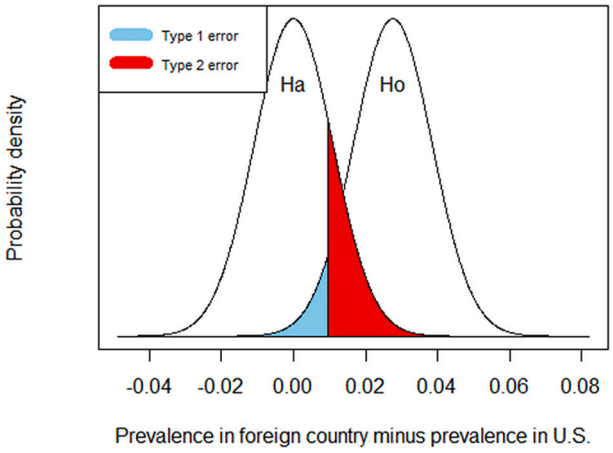

Fig. 1. An illustrative example of the distributions for the null (Ho) and alternative (Ha) hypotheses for non-inferiority testing. In this case, the Type 1 (blue) and Type 2 (red) errors are 0.05 and 0.16, respectively. For this example, the pathogen prevalence in FSIS-inspected U.S. product is 5% based on 1000 product samples (with another 1000 similar product samples collected by the other country), the margin of equivalence (Ho mean) is 4 times the standard error of the U.S. product pathogen prevalence (δ= 0.028) and the critical value that defines the Type 1 and 2 border equals $\delta - Z_{1-\alpha} \times \sigma_{difference} \approx 0.01$.

product, when it is truly non-inferior. The complement of specificity is the Type 2 error of the decision (Fig. 1), the probability of mistakenly deciding another country's alternative inspection system results in product that is riskier than FSIS-inspected U.S. product when it is not. The best choice for setting the margin of equivalence is one that provides risk managers with the ability to protect the public health in the U.S. (high sensitivity) without placing an undue burden on our trading partners (high specificity).

The risk management choice of Type 1 error determines the sensitivity of the decision and the critical value of the difference in the observed product pathogen prevalence values that result in rejection of the null hypothesis. This critical value is $crit = \delta - Z_{1-\alpha} \times \sigma_{difference}$.

Specificity depends on the alternative hypothesis that the consumer risk posed by products from the two different sources are the same, or the country's product pathogen prevalence is lower than the observed pathogen prevalence in similar FSIS-inspected U.S. product. The probability of correctly deciding that another country's product pathogen prevalence is non-inferior under the alternative hypothesis corresponds to the p^{th} percentile of the alternative hypothesis distribution at the critical value (Fig. 1). By convention, the distributions for the null and alternative hypotheses are assumed to follow a normal distribution.[1] Therefore,

$$Specificity = \Phi\left(crit|\mu = 0, \sigma_{same} = \sqrt{\widehat{P}_{U.S.}(1-\widehat{P}_{U.S.})\left(\frac{1}{n_{foreign}} + \frac{1}{n_{U.S.}}\right)}\right), \quad (7)$$

where $\Phi(x|\mu, \sigma_{same})$ is the cumulative probability for a value of x from a normal distribution with its mean equal to μ and its standard deviation equal to σ_{same}. Because specificity is a function of the critical value, we can specify the following equivalency:

$$critical\ value(crit) = 0 + Z_{1-\beta}\sigma_{same} = \delta - Z_{1-\alpha}\sigma_{diff},$$

where $Z_{1-\beta}$ is the value from a standard normal distribution that corresponds to the $1-\beta$ percentile. For a specificity of 80% (Type 2 error of 20%), $Z_{1-\beta} = 0.84$ (see Appendix). In this case, the standard error σ_{same} is based on the pathogen prevalence in FSIS-inspected U.S. product, except that the sample size may differ between the countries. This standard error is necessarily smaller than $\sigma_{difference}$.

To illustrate the effects of the U.S. product pathogen prevalence and FSIS sample size on the margin of equivalence, as well as the sensitivity and specificity of the non-inferiority decision, we consider ranges of product pathogen prevalence (0.1% to 10%) and sample sizes (100 to 10,000 samples) that are representative of annual FSIS pathogen testing results.[2] For example, in Fiscal Year 2020, the *E. coli* O157:H7 occurrence was 0.1% from 10,500 samples of ground beef, *Campylobacter* occurrence was ~5% among 2000 samples of ground chicken, and *Salmonella* occurrence was ~10% among roughly 60 quarter- or half-carcass chicken samples. All calculations were completed using the R statistical software (R Development Core Team, 2018).

3. Results

If we consider the scenario of a pathogen prevalence of 5% in FSIS-inspected U.S. product, as estimated from 1000 samples, then we can assess the probability that the country's product has a larger prevalence (i.e., the *p*-value) for a range of product pathogen prevalence levels in product from another country based on the same number of samples (Fig. 2). Selecting a Type 1 error of 5% ($\alpha = 0.05$) will result in rejecting the null hypothesis if the *p*-value is less than 0.05. In this example, the *p*-value equals 0.05 for a country's product pathogen prevalence values of 5.3, 5.9 and 6.6% for margin of equivalence values that are m= 3, 4 and 5 times the standard error of the FSIS-inspected U.S. product-pathogen prevalence, respectively.[3] Larger margins of equivalence correspond to allowing a higher pathogen prevalence in a country's product, while still concluding non-inferiority (i.e., rejecting the null hypothesis and determining the product poses no greater risk than U.S. product).

For reference, the standard error of the U.S. prevalence estimate in Fig. 2 is 0.69% $\left(\sqrt{\frac{0.05\times0.95}{1000}}\right)$, so the margins of equivalence are 2.1%, 2.76% and 3.45% for m = 3, 4, 5, respectively. These margins define an unacceptable prevalence in the other country as 7.1% (5% + 2.21%), 7.76% (5% + 2.76%) and 8.45% (5% + 3.45%). Nevertheless, the critical values for observed prevalence that result in non-inferiority conclusions (i.e., 5.3, 5.9 and 6.6% from Fig. 2) are substantially less than these unacceptable prevalence levels. In fact, for this example, the critical values imply that the observed prevalence in the other country must be less than 0.43 = $\left(\frac{5.3\%-5\%}{0.69\%}\right)$, 1.30 = $\left(\frac{5.9\%-5\%}{0.69\%}\right)$ or 2.32 = $\left(\frac{6.6\%-5\%}{0.69\%}\right)$ standard error units above the U.S. prevalence to conclude they are non-inferior.

Previously, FSIS estimated how *Salmonella* prevalence varies across the U.S. establishments that produce FSIS-inspected ground beef (Lukicheva et al., 2016) (Fig. 3). This example illustrates the relative

[1] This assumption holds, according to the Rule of Sample Proportions, when the number of positive samples used to estimate prevalence is at least 10 (https://online.stat.psu.edu/stat200/lesson/8/8.1).

[2] Sampling results for FSIS inspected products (https://www.fsis.usda.gov/science-data/sampling-program/sampling-results-fsis-regulated-products).

[3] For a multiple (m) of the standard error of the U.S. prevalence, a conclusion of non-inferiority will result if the observed prevalence in another country is less than a maximum prevalence ($P_{foreign\_max}$) that is calculated by assuming, based on an infinite sample size in the other country $\left(P_{foreign\_max} = P_{U.S.} + (m - Z_{1-\alpha})\sqrt{\frac{P_{U.S.}(1-P_{U.S.})}{n_{U.S.}}}\right)$. For example, if the margin of equivalence is based on m = 4 standard error units of the U.S. prevalence, then a conclusion of non-inferiority will result as long as the maximum observed prevalence in the other country is less than about 2 standard error units above the U.S. prevalence. Practically, the critical value for the observed prevalence in another country will always be less than $P_{foreign\_max}$ because the number of samples for that country will be limited rather than infinite.

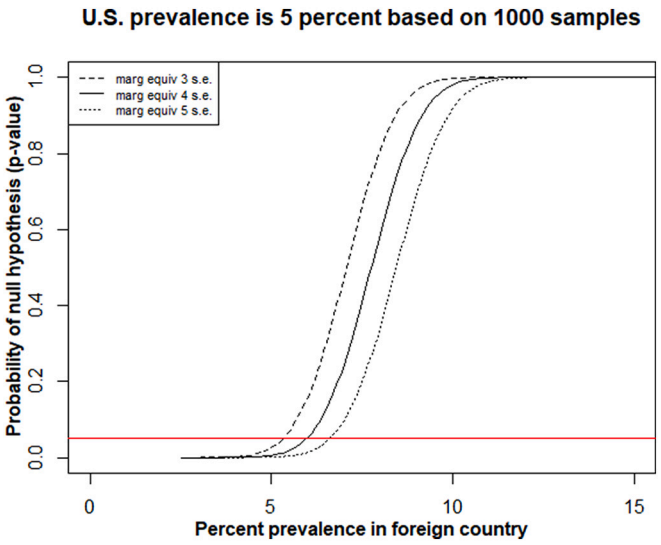

Fig. 2. Illustration of *p*-values for a range of pathogen prevalence levels in a country's product, given an FSIS-inspected U.S. product pathogen prevalence of 5%, based on 1000 product samples and three alternative scenarios for a margin of equivalence. The horizontal line is for a *p*-value of 0.05. If the *p*-value for a particular country's product pathogen prevalence and margin of equivalence scenario is less than the acceptable Type 1 error, then the null hypothesis is rejected in favor of concluding non-inferiority; otherwise, the null hypothesis that the country's product pathogen prevalence is inferior to that of the U.S. product is not rejected.

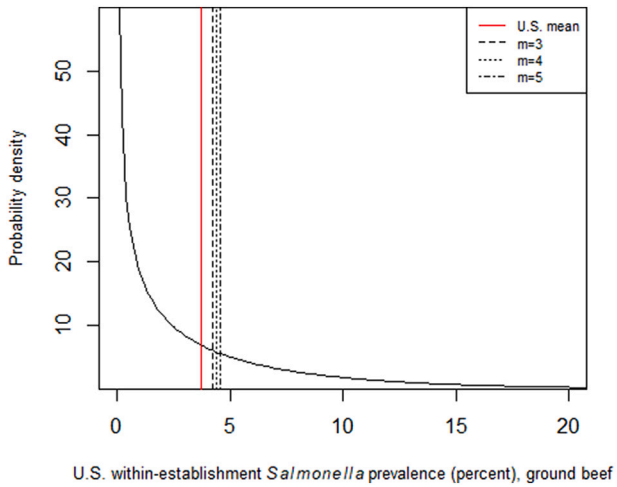

Fig. 3. A beta distribution (beta (0.535, 13.953)) for the U.S. ground beef industry is shown. This distribution describes how Salmonella prevalence – among 325 g samples – varies across FSIS-inspected establishments in this industry (Lukicheva et al., 2016). The solid (red) vertical line represents the mean of this distribution and the dashed vertical lines (black) are the mean, plus a margin of equivalence that is m = 3, 4 or 5 times the standard error of the mean U.S. prevalence (based on 11,791 samples).

positions of the mean product pathogen prevalence and the mean product pathogen prevalence plus a margin of equivalence, in the context of this distribution. In this case, the mean product pathogen prevalence is estimated to be 3.7% (based on 11,791 samples). Given the skewed distribution for *Salmonella* in ground beef across the industry, its mean equates to the 67th percentile of the distribution. For $m = 3$, 4 or 5, the margin of equivalence is $\delta = 0.005$, 0.007 or 0.009. The addition of these margins of equivalence to the mean product pathogen prevalence generates prevalence values that are the 70th, 71st or 72nd percentiles of this beta distribution.

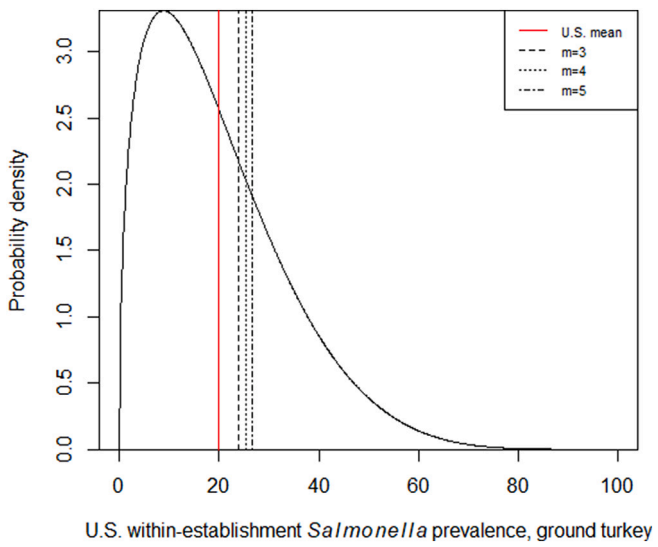

Fig. 4. A beta distribution (beta (1.48, 5.95)) for the U.S. comminuted turkey industry describes how Salmonella prevalence – among 325 g samples – varies across establishments in this industry (Lukicheva et al., 2016). The solid (red) vertical line represents the mean of this distribution and the dashed (black) vertical lines are the mean, plus a margin of equivalence that is m = 3, 4 or 5 times the standard error of the mean prevalence (based on 876 samples).

Another example – for *Salmonella* in comminuted turkey samples produced in the FSIS-inspected U.S. establishments – has a mean product pathogen prevalence of 20%, but is based on a smaller sample size (n = 876 samples) (Fig. 4). The mean product pathogen prevalence represents the 57th percentile of this U.S. industry's beta distribution. In this case, for $m = 3$, 4 or 5, the margin of equivalence is $\delta = 0.04$, 0.05 or 0.07. The addition of these margins of equivalence to the mean product pathogen prevalence generates prevalence values that are the 67th, 70th or 73rd percentiles of this beta distribution.

The sensitivity and specificity of the non-inferiority decision depend on the pathogen prevalence in FSIS-inspected product, sample size and margin of equivalence. For example, if the U.S. pathogen prevalence is 0.1% based on 10,000 samples, then sensitivity and specificity are affected by the selection of the margin of equivalence (Fig. 5). For a sensitivity of 95% (i.e., Type 1 error (1 − Sensitivity) = 0.05), the specificity in this example is 55, 77 and 91% when the margin of equivalence is $m = 3$, 4 or 5 times the standard error of the U.S. product pathogen prevalence, respectively. Alternatively, for a specificity of 80%, the sensitivity in this example is 85, 94 and 98% when the margin of equivalence is 3, 4 and 5 times the standard error of the U.S. product pathogen prevalence, respectively. Other displays (not shown) of the risk metric performance characteristics using larger U.S. product pathogen prevalence and sample size assumptions (e.g., 5% with 1000 samples, 10% with 100 samples), illustrate similar behavior.

When a country collects and tests fewer product samples than the U.S., this affects the assessment of whether a country's alternative inspection system results in product that poses no greater risk compared to FSIS-inspected product.[4] Using the same scenario of a U.S. pathogen prevalence of 0.1% based on 10,000 samples, if another country tests fewer samples, it will worsen the sensitivity and specificity of the decision about non-inferiority of that country's product (Fig. 6). The risk metric demonstrates that if a country collects 5000 samples to estimate its prevalence, then the specificity is only 53% for a sensitivity of 95%. This improves, achieving a specificity of 77% at a 95% sensitivity, when

[4] In the context of an ISM equivalence determination, the U.S. would not be asking other countries to collect samples. Instead, the U.S. would ask for sample data that the other country may have already collected.

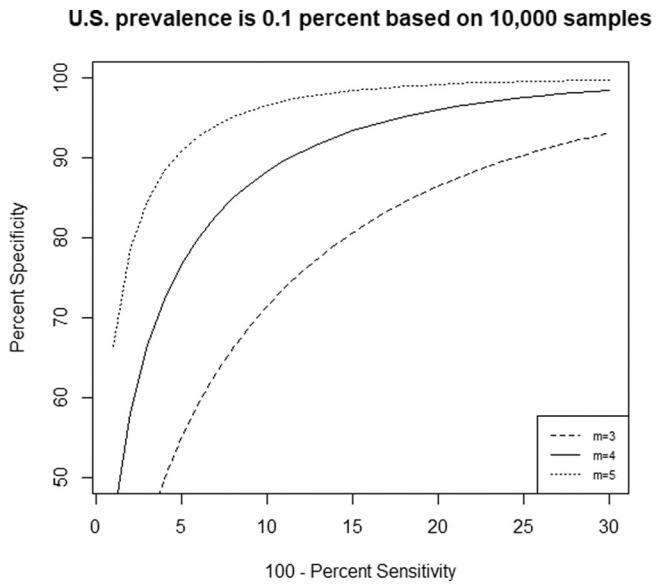

Fig. 5. Risk metric performance characteristics of a decision about non-inferiority of another country's product. For this illustrative example, the U.S. pathogen prevalence in FSIS-inspected product is 0.1% based on 10,000 samples. It is further assumed that the country has collected a similar number of samples to estimate its product pathogen prevalence. The margin of equivalence is varied from m = 3 times to m = 5 times the standard error of the U.S. product pathogen prevalence. Each curve is derived by considering Type I error values ranging from 0.01 to 0.3 (1 − Sensitivity) and using Eq. (7) to determine the corresponding Specificity (1 − Type II error).

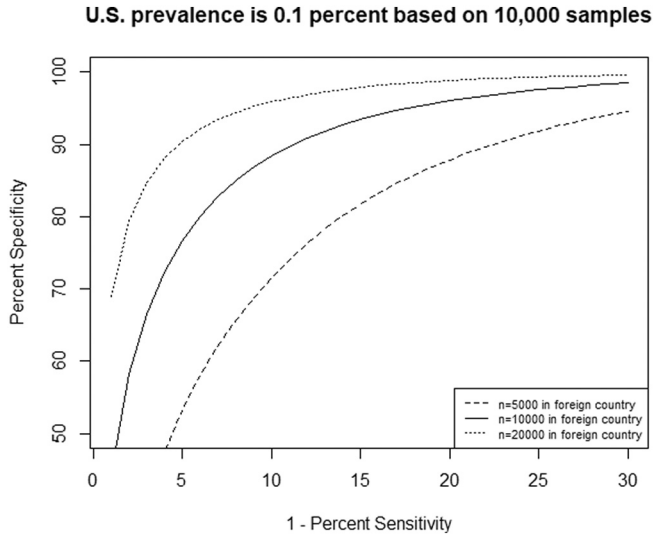

Fig. 6. Risk metric performance characteristics of a decision about non-inferiority of another country's food product. For this illustrative example, the U.S. pathogen prevalence in FSIS-inspected product is 0.1% based on 10,000 product samples and the margin of equivalence set at m = 4 times the standard error of the U.S. pathogen prevalence. Each curve reflects a different number of samples collected in the country; either one-half, equal to, or twice the number of samples collected by FSIS. Each curve is derived by considering Type I error values ranging from 0.01 to 0.3 (1 − Sensitivity) and using Eq. (7) to determine the corresponding Specificity (1 − Type II error).

the country collects the same number (i.e., 10,000) of samples as FSIS. Alternatively, if the country collects even more samples, for example 20,000 product samples (i.e., double the number collected by FSIS), then the specificity improves further, to 90% for a sensitivity of 95%.

Setting default risk tolerances for Type 1 and 2 errors at 0.05 and 0.20 (Walker and Nowacki, 2011), respectively, we explore the margin of equivalence necessary to achieve these targets across a range of U.S. pathogen prevalence levels (0.1% to 10%) and sample sizes (from 100 to 10,000 product samples) for FSIS-inspected product.[5] By modeling each of these inputs as uniform variables – then solving each iteration to determine the multiple of the standard error (m) applicable to that iteration's U.S. product prevalence such that Type 1 and 2 errors equal their targets – a distribution of the multiples of the U.S. standard error is generated that ranges between 3.5 and 5.5, with a mean value of 3.7 (Fig. 7). This range is consistent with theoretic minimum and maximum values of m that can be derived by assuming large sample numbers ($m_{\min} = 3.5$) and small prevalence levels with corresponding limited sample numbers ($m_{\max} = 5.8$) (see Appendix).

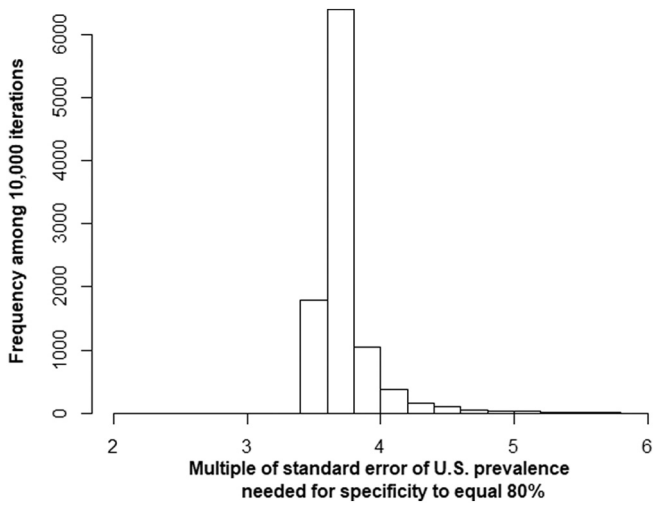

Fig. 7. The multiple of the standard error of the mean U.S. pathogen prevalence in FSIS-inspected product needed to calculate the margin of equivalence such that specificity is 80% when sensitivity is estimated to be at 95% across a range of the U.S. prevalence levels (i.e., 0.1% to 10%) and sample sizes (i.e., 100 to 10,000 samples).

Suggestions for the choice of a general margin of equivalence for a range of applications are typically whole numbers and sample size calculations for low prevalence applications are often higher than necessary to achieve the desired power value (Williams et al., 2007). For this reason, we suggest a reasonable margin of equivalence may be based on four standard error units (rounding 3.7 up to the nearest whole integer), which encompasses 90 to 95% of the scenarios considered.

In Fig. 7, multiples of the standard error of the mean pathogen prevalence in FSIS-inspected product that are less than four are associated with higher product pathogen prevalence and larger sample numbers. For example, a U.S. prevalence of 5%, based on 1000 samples generates $m = 3.78$. In this case, setting $\delta = 4\sigma_{U.S.}$ increases specificity from 80% to 84% if we maintain sensitivity at 95%. Improved specificity about the non-inferiority decision (i.e., high probability of correctly assessing production from another country poses no greater risk than similar product in the U.S.) assures that the assessment of the comparative risk of products is consistent with the WTO concept that food safety systems that are equivalent provide at least the same level of public health protection.

Scenarios associated with multiples greater than four are those with lower U.S. product pathogen prevalence values combined with small sample sizes. For example, a U.S. prevalence of 0.15% based on 1500 samples generates $m \approx 5$. In this case, setting $\delta = 4\sigma_{U.S.}$ reduces specificity from 80% to 63% if we maintain sensitivity at 95%. Nevertheless,

[5] For each iteration, the minimum possible prevalence equals the inverse of the sample number for that iteration.

this scenario implies FSIS found only two positive samples among the 1400 samples it collected. Such a scenario would be unusual based on FSIS experience, although small numbers of *E. coli* O157:H7-positive samples are found typically among the roughly 10,000 ground beef samples collected annually by FSIS.

Given these results, the choice of setting the margin of equivalence at four standard errors of the mean U.S. product pathogen prevalence will generally result in decisions with specificity greater than 80% (while maintaining sensitivity at 95%) because most scenarios we considered achieve the target specificity with a margin of equivalence less than four standard errors. Alternatively, because the margin of equivalence that achieves the targeted sensitivity and specificity can be determined directly as a function of the U.S. prevalence and sample number (Appendix), it is also feasible to choose a value of m on an ad hoc basis for each ISM decision. Such an approach would maintain the same specificity and sensitivity for each decision by adjusting the value of m, but would require more mathematics than a policy that chooses a default value of m.

4. Discussion

In this paper, we present a risk-based approach that would inform an evaluation of the comparative safety of products produced under differing inspection systems based on contemporaneous pathogen testing data when collected using harmonized sample collection and equivalent laboratory testing methods. This approach, based on a risk metric that applies non-inferiority statistical testing, enhances the evaluation of pathogen testing data that a foreign country may provide to FSIS when requesting an ISM equivalence determination for an alternative inspection procedure. The approach we present is streamlined and fit-for-purpose to objectively evaluate if the product produced under an alternative inspection system and intended for export to the U.S. is non-inferior to similar FSIS-inspected product.

Risk managers may choose to standardize the margin of equivalence or determine it on an adhoc basis. If they choose a standardized approach, using four standard errors of the reference prevalence generates margin of equivalence values that are consistent with some non-inferiority testing applications in medicine. For example, in a comparison of alternative stem cell transplants (Tunes da Silva et al., 2009) the chosen margin of equivalence for survival was $\delta = 0.10$. The reference survival probabilities ranged from 38% (n = 117) to 44% (n = 478) and implied standard errors of 0.045 and 0.023, respectively. In these cases, a choice of $m = 4$ standard errors would generate margins of equivalence that range from $\delta = 0.18$ to$\delta = 0.09$.

In our discussion of the risk metric, the choice of target pathogens was outside the scope of this paper, but the importing country's priorities and the system being evaluated should influence the selection of which pathogens will be assessed. For example, in poultry products, FSIS focuses its surveillance attention on *Salmonella* and *Campylobacter*, but the pathogens of interest are *Salmonella* and Shiga toxin-producing *E. coli* in raw beef products. These product-pathogen pairs are also commonly priorities for international trade (Babu Rajendran et al., 2020).

Situations when multiple pathogens are priorities, such as *Salmonella* and *Campylobacter* in poultry products, will require additional consideration by risk managers. Application of the risk metric's non-inferiority test to sampling evidence for each pathogen is straightforward. It is the prerogative of risk managers to decide how to weight the results of multiple comparisons to inform ISM equivalence determinations.

This paper's approach assumes a common understanding about the meaning of prevalence. Prevalence measurement issues are beyond the scope of this paper, but should be considered when evaluating the suitability of data to support the equivalence determination. Generally, sampling in both countries should be contemporaneous and representative of the food produced under the inspection systems being compared, samples should be consistent in the quantity and type of material assayed and equivalent laboratory methods should have comparable performance characteristics for the intended use.[6] Estimation of product pathogen prevalence should also account for how the population of product was sampled. For example, FSIS product pathogen prevalence estimates are derived by weighting each establishment's sampling results with respect to its production volume to account for the substantial differences in annual production volume across the U.S. establishments (Williams et al., 2013). These results are also derived from nearly all establishments in the U.S. Because FSIS surveillance programs over-sample smaller establishments, the effective sample size used in the risk metric's non-inferiority test for a U.S. product-pathogen pair may represent an adjusted value relative to actual number of samples collected across the U.S. industry. Conversely, establishments in a country that wishes to export to the U.S. may be just a subset of that country's industry.

This paper only discusses the mechanics of assessing pathogen testing data collected in a country and in the U.S. The underlying credibility of those data is assumed to be reflected by the magnitude of standard errors for each prevalence estimate. Nevertheless, representation of prevalence for either country requires careful examination before and after data are collected. For example, the proposed methods do not account for factors such as differences in pathogen virulence between countries, or differences in pathogen growth and attenuation associated with the increased average transportation time between production and consumption.

Experience may suggest that prevalence is dynamic across time, seasons, demographic or topographic groups, etc. How dynamics are factored into a non-inferiority decision is beyond the scope of this paper, but such behavior should be considered when applying the methods discussed here. For example, if prevalence in the U.S. is expected to continue a downward trend that has been identified previously, then alternative predictions about the current U.S. prevalence may be assessed in a scenario analysis of the non-inferiority test. The output of such an analysis might inform the confidence about a decision on the equivalence of risk between another country and the U.S. The proposed methods also do not account for factors such as differences in pathogen virulence between countries or differences in pathogen growth and attenuation associated the increased average transportation time between production and consumption.

While the framework is relatively straightforward, risk managers are still left with important decisions and details to ensure proper implementation. As with all such decisions, consistency and transparency in the process is crucial to ensuring equitable treatment of both domestic and foreign producers.

Declaration of competing interest

None.

Appendix A

The following explores the theoretic minimum and maximum values for m, the multiplier of the U.S. standard error ($\sigma_{U.S.}$). A quadratic form is

[6] https://www.fsis.usda.gov/sites/default/files/import/Validation_Studies_Pathogen_Detection_Methods.pdf.

derived such that the positive root for m can be calculated directly for a given U.S. prevalence and sample number.

$P_{U.S.}$ = prevalence in U.S.(product − pathogen)

$n_{U.S.}$ = samples collected in U.S.

$n_{foreign}$ = samples collected in other country

$$\sigma_{U.S.} = \sqrt{\left(\frac{P_{U.S.}(1-P_{U.S.})}{n_{U.S.}}\right)}$$

$$\sigma_{same} = \sqrt{2}\sigma_{U.S.}$$

$$\sigma_{diff} = \sqrt{\left(\frac{P_{U.S.}(1-P_{U.S.})}{n_{U.S.}} + \frac{(P_{U.S.}+\delta)(1-(P_{U.S.}+\delta))}{n_{foreign}}\right)}$$

$$critical\ value(crit) = 0 + Z_{1-\beta}\sigma_{same} = \delta - Z_{1-\alpha}\sigma_{diff}$$

$$\delta = Z_{1-\beta}\sqrt{2}\sigma_{U.S.} + Z_{1-\alpha}\sigma_{diff}$$

assume $\delta = m\sigma_{U.S.}$

$$m = Z_{1-\beta}\sqrt{2} + Z_{1-\alpha}\frac{\sigma_{diff}}{\sigma_{U.S.}}$$

If $\sigma_{diff} \approx \sigma_{U.S.}, Z_{1-\beta} = 0.84$ (80%specificity),

$Z_{1-\alpha} = 1.64$ (95%sensitivity), then $m \approx 2.8$.

This is the minimum m can be. But, note that as $\sigma_{U.S.} \to 0$

(e.g., large sample number), then $\frac{\sigma_{diff}}{\sigma_{U.S.}} \approx \sqrt{2}$ and

a practical minimum for $m_{\min} = \sqrt{2}(Z_{1-\beta} + Z_{1-\alpha}) \approx 3.5$.

Next, derive the quadratic to solve for m.

Assume $n_{U.S.} = n_{foreign} = n$; and $P_{U.S.} = P$

$$\text{Then, } \frac{\sigma_{diff}}{\sigma_{U.S.}} = \sqrt{1 + \frac{(P + m\sigma_{U.S.})(1 - P - m\sigma_{U.S.})}{P(1-P)}}$$

$$\therefore \left(m - \sqrt{2}Z_{1-\beta}\right)^2 = \left(Z_{1-\alpha}\sqrt{1 + \frac{(P + m\sigma_{U.S.})(1 - P - m\sigma_{U.S.})}{P(1-P)}}\right)^2$$

$$m^2 - 2\sqrt{2}Z_{1-\beta} + {Z_{1-\beta}}^2 = {Z_{1-\alpha}}^2 + {Z_{1-\alpha}}^2\left(\frac{(P + m\sigma_{U.S.})(1 - P - m\sigma_{U.S.})}{P(1-P)}\right)$$

$$\left[P(1-P) + {Z_{1-\alpha}}^2\sigma^2{}_{U.S.}\right]m^2 - \left[2\sqrt{2}Z_{1-\beta}P(1-P) + {Z_{1-\alpha}}^2(\sigma_{U.S.} - 2P\sigma_{U.S.})\right]m + \left[P(1-P)\left(2{Z_{1-\beta}}^2 - 2{Z_{1-\alpha}}^2\right)\right] = 0$$

Next, explore a theoretic maximum for m.

As $P \to 0$ and a minimum sample number of $n = 1/P$,

we can assume the following;

$P(1-P) \approx P; \sigma_{U.S.} = \sqrt{P(1-P)/(1/P)} \approx \sqrt{P^2} \approx P$; and $\sigma^2 \approx P^2 \approx 0$.

Replacing these terms in the quadratic gives

$$[P]m^2 - \left[P\left(2\sqrt{2}Z_{1-\beta} + Z_{1-\alpha}^2\right)\right]m + P\left(2{Z_{1-\beta}}^2 - 2{Z_{1-\alpha}}^2\right) = 0$$

$$\therefore m_{\max} = \frac{2\sqrt{2}Z_{1-\beta} + Z_{1-\alpha}^2 + \sqrt{Z_{1-\alpha}^4 + 4\sqrt{2}Z_{1-\beta}{Z_{1-\alpha}}^2 + 8{Z_{1-\alpha}}^2}}{2} \approx 5.8$$

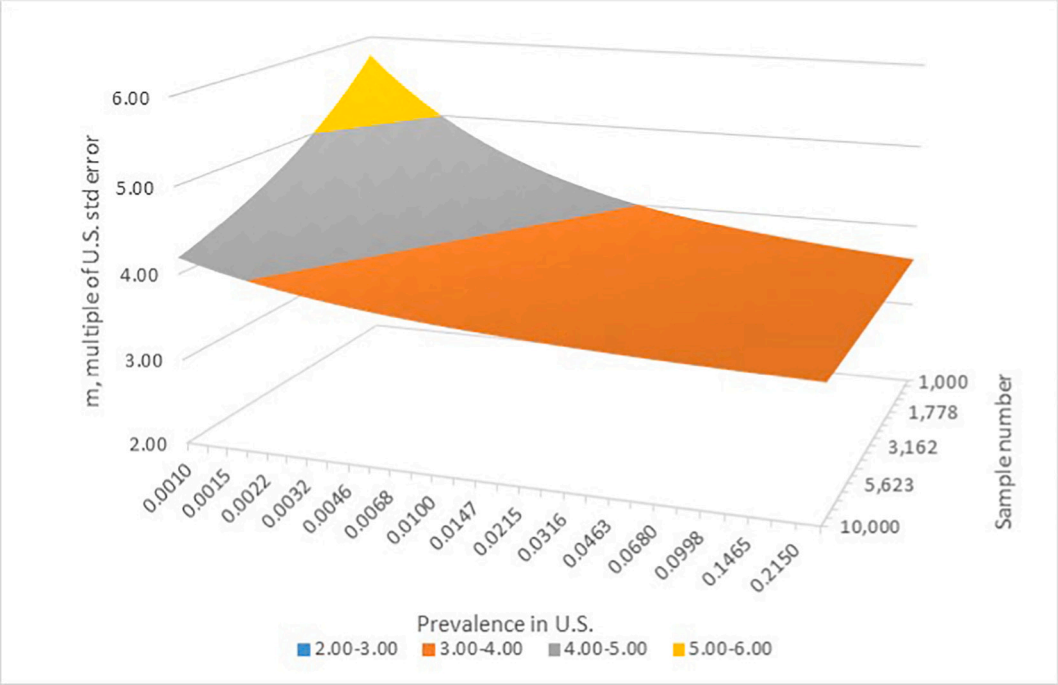

Fig. 1A. A surface map shows how *m*, the multiplier of the U.S. standard error ($\sigma_{U.S.}$), changes as a function of the U.S. prevalence and numbers of samples collected to estimate the prevalence. The theoretic minimum value for *m* is 3.5 while the theoretic maximum is 5.8.

References

Babu Rajendran, N., Mutters, N.T., Marasca, G., Conti, M., Sifakis, F., Vuong, C., Voss, A., Baño, J.R., Tacconelli, E., 2020. Mandatory surveillance and outbreaks reporting of the WHO priority pathogens for research & discovery of new antibiotics in European countries. Clin. Microbiol. Infect. 26, 943.e941–943.e946.

Bartholomew, M.J., Vose, D.J., Tollefson, L.R., Travis, C.C., 2005. A linear model for managing the risk of antimicrobial resistance originating in food animals. Risk Anal. 25, 99–108.

Bassett, J., Nauta, M., Lindqvist, R., Zwietering, M., 2012. Tools for microbiological risk assessment, ILSI Europe. In: ILSI Europe Report Series, p. 40.

Batz, M.B., Doyle, M.P., Morris Jr., J.G., Painter Jr., J., Singh Jr., R., Tauxe Jr., R.V., Taylor Jr., M.R., Wong Jr., D.M.L.F., Group F.A.W., 2005. Attributing illness to food. Emerging Infect. Dis. 11, 993.

Chardon, J., Evers, E., 2018. Burden of disease as a metric for risk-based sampling of imported foods. In: IAFP 2018 Annual Meeting. IAFP.

Christensen, B.B., Nauta, M., Korsgaard, H., Sørensen, A.I.V., Rosenquist, H., Boysen, L., Perge, A., Nørrung, B., 2013. Case-by-case risk assessment of broiler meat batches: An effective control strategy for Campylobacter. Food Control 31, 485–490.

Dearfield, K.L., Hoelzer, K., Kause, J.R., 2014. Review of various approaches for assessing public health risks in regulatory decision making: Choosing the right approach for the problem. J. Food Prot. 77, 1428–1440.

Ebel, E.D., Williams, M.S., 2015. When are qualitative testing results sufficient to predict a reduction in illnesses in a microbiological food safety risk assessment? J. Food Prot. 78, 1451–1460.

Ebel, E.D., Williams, M.S., Golden, N.J., Marks, H.M., 2012. Simplified framework for predicting changes in public health from performance standards applied in slaughter establishments. Food Control 28, 250–257.

Evers, E.G., Chardon, J.E., 2010. A swift Quantitative Microbiological Risk Assessment (sQMRA) tool. Food Control 21, 319–330.

FAO/WHO, 2016. Statistical aspects of microbiological criteria related to foods: a risk manager's guide. In: Microbiological Risk Assessment Series, No 24, Rome, p. 120.

FAO/WHO, 2021. Risk Management Tool for the Control of Campylobacter and Salmonella in Chicken Meat (verison 1.0), 1.0 ed. Geneva, Switzerland.

FDA, 2012. FDA-iRISK®, 4.2 ed. Food and Drug Administration, Washington, D.C.

FSIS, 2012. In: Imported Products. U.S. Department of Agriculture, Washington, D.C., p. 9. CFR 327:331-350.

FSIS, 2019. In: Publication Method for Lists of Foreign Countries Eligible to Export Meat, Poultry, or Egg Products to the United States. U.S. Department of Agriculture, Washington, D.C., pp. 65265–65269

FSIS, 2020. Individual Sanitary Measure Equivalence, Washington, D.C.

FSIS, 2020. Microbiology Laboratory Guidebook. U.S. Department of Agriculture, Washington, D.C.

Gallagher, D., Ebel, E.D., Gallagher, O., LaBarre, D., Williams, M.S., Golden, N.J., Pouillot, R., Dearfield, K.L., Kause, J., 2013. Characterizing uncertainty when evaluating risk management metrics: Risk assessment modeling of Listeria monocytogenes contamination in ready-to-eat deli meats. Int. J. Food Microbiol. 162, 266–275.

Havelaar, A.H., Evers, E.G., Natua, M.J., 2008. Challenges of quantitative microbial risk assessment at EU level. Trends Food Sci. Technol. 19, S26–S33.

Havelaar, A.H., Kirk, M.D., Torgerson, P.R., Gibb, H.J., Hald, T., Lake, R.J., Praet, N., Bellinger, D.C., De Silva, N.R., Gargouri, N., 2015. World Health Organization global estimates and regional comparisons of the burden of foodborne disease in 2010. PLoS Med. 12, e1001923.

Havelaar, A.H., Nauta, M.J., Jansen, J.T., 2004. Fine-tuning food safety objectives and risk assessment. Int. J. Food Microbiol. 93, 11–29.

HHS, 2020. Healthy people topics & objectives: food safety. In: Reduce Infections Caused by Salmonella — FS-04, Washington, D.C.

Hsi, D.J., Ebel, E.D., Williams, M.S., Golden, N.J., Schlosser, W.D., 2015. Comparing foodborne illness risks among meat commodities in the United States. Food Control 54, 353–359.

ISO, 2017. General Requirements for the Competence of Testing and Calibration Laboratories. EN ISO/IEC, Geneva, Switzerland.

Kaul, S., Diamond, G.A., 2006. Good enough: a primer on the analysis and interpretation of noninferiority trials. Ann. Intern. Med. 145, 62–69.

Kirk, M.D., Pires, S.M., Black, R.E., Caipo, M., Crump, J.A., Devleesschauwer, B., Döpfer, D., Fazil, A., Fischer-Walker, C.L., Hald, T., 2015. World Health Organization estimates of the global and regional disease burden of 22 foodborne bacterial, protozoal, and viral diseases, 2010: a data synthesis. PLoS Med. 12, e1001921.

Kuchler, F., Krissoff, B., Harvey, D., 2010. Do consumers respond to country-of-origin labelling? J. Consum. Policy 33, 323–337.

Li, M., Havelaar, A.H., Hoffmann, S., Hald, T., Kirk, M.D., Torgerson, P.R., Devleesschauwer, B., 2019. Global disease burden of pathogens in animal source foods, 2010. PLoS One 14, e0216545.

Lukicheva, N., Ebel, E.D., Williams, M.S., Schlosser, W.D., 2016. Characterizing the concentration of pathogen occurrence across meat and poultry industries. Microb. Risk Anal. 4, 29–35.

NRC, 2009. Letter Report on the Review of the Food Safety and Inspection Service Proposed Risk-based Approach to and Application of Public-Health Attribution, Wahington, D.C.

R Development Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Schumi, J., Wittes, J.T., 2011. Through the looking glass: understanding non-inferiority. Trials 12, 1–12.

Tunes da Silva, G., Logan, B.R., Klein, J.P., 2009. Methods for equivalence and noninferiority testing. Biol. Blood Marrow Transplant. 15, 120–127.

Walker, E., Nowacki, A.S., 2011. Understanding equivalence and noninferiority testing. J. Gen. Intern. Med. 26, 192–196.

Williams, M.S., Ebel, E.D., Allender, H.D., 2015. Industry-level changes in microbial contamination on market hog and broiler chicken carcasses between two locations in the slaughter process. Food Control 51, 361–370.

Williams, M.S., Ebel, E.D., Cao, Y., 2013. Fitting distributions to microbial contamination data collected with an unequal probability sampling design. J. Appl. Microbiol. 114, 152–160.
Williams, M.S., Ebel, E.D., Vose, D., 2011. Framework for microbial food-safety risk assessments amenable to Bayesian modeling. Risk Anal. 31, 548–565.
Williams, M.S., Ebel, E.D., Wagner, B.A., 2007. Monte Carlo approaches for determining power and sample size in low-prevalence applications. Prevent. Vet. Med. 82, 151–158.
WTO, 1998. Understanding the WTO Agreement on Sanitary and Phytosanitary Measures, Geneva, Switzerland.