Preventive Veterinary Medicine 109 (2013) 205–212

Contents lists available at SciVerse ScienceDirect

Preventive Veterinary Medicine

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

Exact alpha-error determination for two-stage sampling strategies to substantiate freedom from disease

I. Kopacka*, J. Hofrichter, K. Fuchs

Austrian Agency for Health and Food Safety (AGES), Division for Data, Statistics and Risk Assessment, Beethovenstraße 8, 8010 Graz, Austria

ARTICLE INFO

Article history:
Received 29 May 2012
Received in revised form
22 September 2012
Accepted 29 September 2012

Keywords:
Alpha-error
Freedom from disease
Two-stage-sampling
Imperfect test
Finite population

ABSTRACT

Sampling strategies to substantiate freedom from disease are important when it comes to the trade of animals and animal products. When considering imperfect tests and finite populations, sample size calculation can, however, be a challenging task. The generalized hypergeometric formula developed by Cameron and Baldock (1998a) offers a framework that can elegantly be extended to multi-stage sampling strategies, which are widely used to account for disease clustering at herd-level. The achieved alpha-error of such surveys, however, typically depends on the realization of the sample and can differ from the pre-calculated value. In this paper, we introduce a new formula to evaluate the exact alpha-error induced by a specific sample. We further give a numerically viable approximation formula and analyze its properties using a data example of Brucella melitensis in the Austrian sheep population.

1. Introduction

In order to meet the standards of trading partners or international organizations, livestock populations are often required to be proven free from certain diseases. Whether or not a population is issued the status of being disease-free depends on the outcome of epidemiological surveys. Absolute proof requires the examination of every individual in a population using an infallible diagnostic test. In reality, however, virtually no diagnostic test is perfect and examining every individual is generally too costly and time consuming. Hence, sample surveys are used, which resort to a representative subset of the population.

Sample surveys cannot provide absolute proof of the absence of a disease. The goal of such a sample survey is rather to provide sufficient evidence to demonstrate – to a level of certainty – that a disease, if present, is present in less than a specified proportion of the population; see, e.g., OIE Terrestrial Code (2012), Article 1.4.6. This prevalence level is referred to as the design prevalence. The reliability of the statement is quantified in terms of the probability of detecting a present disease in the population and is commonly referred to as the confidence level. An equivalent, and widely used, measure of the quality of the statement is given by the error probability (type I error α), i.e., the probability of not being able to detect a present disease. This error is mainly determined by the survey design and the sample size. Cannon and Roe (1982) introduced approximate formulas to compute the sample size required to achieve a preset α. The results were tabulated and have widely been used for the design of surveys. The formulas are easy to apply but they are based on two simplifying assumptions: the population under investigation is assumed to be infinite and the diagnostic test being used is assumed to be perfect, i.e., the test has perfect sensitivity and specificity. MacDiarmid (1988) later developed a modification of the approximate formula which considers imperfect test sensitivity. An exact formula was introduced

* Corresponding author. Tel.: +43 50 555 61412; fax: +43 50 555 61409.
E-mail addresses: ian.kopacka@ages.at (I. Kopacka), jh@derstatistiker.at (J. Hofrichter), klemens.fuchs@ages.at (K. Fuchs).

0167-5877/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.prevetmed.2012.09.016

by Cameron and Baldock (1998a), using a generalization of the hypergeometric distribution to characterize the distribution of the number of test positive cases in a sample. This formula accounts for finite populations and imperfect tests.

In large populations where animals are separated into herds, disease has a tendency to cluster at herd-level. Hence, when an infection occurs, a rather high within-herd prevalence might be observed, while the prevalence between individual herds might be considerably lower. In order to account for the difference in these prevalence levels, Cameron and Baldock (1998b) recommended the use of two-stage schemes, where, instead of sampling individual animals from the entire population, herds are sampled in the first stage and, subsequently, individual animals are chosen from the specified herds in the second stage. Apart from the ability to account for disease clustering, two-stage sampling further has an appeal that is more of a logistic nature. Using a single-stage approach to sample from a large population requires a list featuring all the animals in the population (e.g., the country), each uniquely identifiable, whereas the two-stage sampling approach only requires a list of the herds in the population featuring the herd sizes, i.e., the number of animals per herd.

The formula of Cameron and Baldock can quite elegantly be extended to fit such two-stage sampling by regarding each sampling stage as a separate single-stage sampling scheme with its own design prevalence and confidence level. The "interface" between the stages is given by the test sensitivity of the top-level stage which corresponds to the confidence level of the lower-level stage (i.e., the herd sensitivity); see, e.g., Cameron and Baldock (1998b).

Typically, such sampling schemes are designed to ensure an overall α-error below a predefined (desired) threshold α^* (usually 5% or 1%). The computation of the first-stage sample size (i.e., the number of herds to be sampled) is, however, generally based on the assumption that the herd sensitivities are constant over all herds in the population. In reality, the herd sensitivity varies between the herds. For reasons of computability, approximate values (average or lower bound) are used for sample size calculations. Hence, the actually achieved α of the scheme can differ from the pre-calculated value, and the error ultimately depends on the realization of the sample. In the course of this paper, we introduce a formula to compute the exact α-error induced by a specific sample. We investigate properties of the exact error for the sampling schemes *individual sampling* (Ziller et al., 1999) and *limited sampling* (Selhorst et al., 1999); see also Ziller et al. (2002) for a comparison of the two sampling strategies. We further apply the results to a data example of Brucella melitensis in the Austrian sheep population.

Due to combinatorial issues, the exact evaluation of the error requires high computational effort and, hence, the population size for which the error can be evaluated is limited by the available computational resources. We, therefore, give a numerically viable alternative and discuss its approximation properties, again using the data example of Brucella melitensis in the Austrian sheep population.

2. Materials and methods

2.1. Preliminaries

In Sections 2.1.1 and 2.1.2, we recall some basic theory and notation from Cameron and Baldock (1998a,b) and Ziller et al. (2002).

2.1.1. One-stage sampling

With surveys to substantiate freedom from disease, a sample is taken from a population and the population is said to be disease-free if the number of test-positive individuals in the sample lies beneath a predefined threshold. For sample designs assuming perfect specificity of the diagnostic test, this threshold is set to one, i.e., the population is considered diseased if one or more test-positive individuals are found in the sample. If, however the diagnostic test has an imperfect specificity, even a disease-free population may generate (false) positive test results. In the survey design, this can be accounted for by setting the threshold to a higher level, i.e., by allowing a certain number of positive test results. This, however, generally leads to larger sample sizes. For the analysis in this paper, we will therefore henceforth assume that the diagnostic test being used has perfect specificity and we define a disease-free population as one where all test results are negative. Apart from its appeal from a computational point of view, the assumption of perfect test specificity is plausible, as positive test results might have economic implications and are generally undesirable; see, e.g., OIE Terrestrial Code (2012). It can therefore be assumed that positive results are thoroughly followed-up to establish a final classification that has no false positive outcomes.

Sample sizes are subsequently computed such that the probability of the survey not being able to detect a present disease falls below a predefined error-level α^*. The error-probability generally depends on the sample size n, the population size N, the design prevalence π and the sensitivity Se of the diagnostic test. Let T^+ denote the number of test-positive individuals in the sample. According to the design prevalence π, we assume that the population contains $d \approx N \cdot \pi$ diseased individuals. The probability of finding no test-positives in the sample, given the disease is present in the population at level π, is then given by

$$\alpha = P(T^+ = 0 | \pi > 0, N, n) = \sum_{y=\max(0,n-N+d)}^{\min(d,n)} \frac{\binom{d}{y}\binom{N-d}{n-y}}{\binom{N}{n}} (1-\mathrm{Se})^y, \quad (1)$$

where the summation index y corresponds to the possible number of diseased individuals in the sample; see Cameron and Baldock (1998a).

2.1.2. Two-stage sampling

Let the parameters of the first sampling stage (sampling herds from the population of herds) be denoted by the sub-index 1 and the parameters of the second stage (sampling animals within selected herds) be denoted by the sub-index 2, followed by the index indicating the selected herd. When