WILEY ECOLOGICAL APPLICATIONS

Sequential Testing of Complementary Hypotheses About Population Size

Journal:	*Ecological Applications*
Manuscript ID	EAP24-0517
Wiley - Manuscript type:	Article
Date Submitted by the Author:	26-Jul-2024
Complete List of Authors:	Rincon, Diego; Washington State University, Department of Entomology; AGROSAVIA, Centro de Investigacion Tibaitata Mcdonald, Izzy; Washington State University, Data Analytics Department Crowder, David; Washington State University, Entomology
Substantive Area:	Statistics and Modeling < Theory < Substantive Area, Statistics (general) < Statistics and Modeling < Theory < Substantive Area, Sampling Design and Analysis < Statistics and Modeling < Theory < Substantive Area, Management < Substantive Area
Organism:	
Habitat:	
Geographic Area:	
Key words/phrases:	Bayes' theorem, conditional probability, population management, probability theory, variable-sample-size sequential probability ratio test
Abstract:	Making inferences about population size is paramount in ecology and pest management, with managers often seeking to determine if a given population is above or below a pre-established threshold through sampling or monitoring. Sequential data analysis is an appealing statistical field for monitoring and decision-making as it is more cost-efficient than fixed-sample-size approaches. However, a key limitation of existing sequential testing procedures is that they require the specification of two non-complementary competing hypotheses to allow for sequential calculation of probability ratios as sampling proceeds. Here we develop an approach to overcome this limitation by using Bayes' theorem to sequentially update the posterior probability of a tested hypothesis against its complementary as new data is collected. The new test can explicitly consider simple or composite hypotheses and process either purely sequential (one-at-a-time) or group sequential data to produce a trajectory of posterior probabilities that can aid in making decisions related to the tested hypothesis. The efficiency of our new test is demonstrated with three case studies that involve inferences about populations in a single or multiple time points and detection of rare species through monitoring. We conclude that the sequential test of Bayesian posterior probabilities offers an equivalent or superior

	alternative to probability ratios or fixed-sample-size designs that is easily interpretable for decision-making even when the sampling is stopped before a predefined decision threshold is reached.

Journal: Ecological applications

Manuscript type: Article

Sequential Testing of Complementary Hypotheses About Population Size

Diego F. Rincon [1,2*], Izzy L. Mcdonald [3], David W. Crowder [1]

[1] Department of Entomology, Washington State University, Pullman, Washington 99163, U.S.A.

[2] Corporación Colombiana de Investigación Agropecuaria (AGROSAVIA), Centro de Investigación Tibaitatá, Mosquera, Cundinamarca 250047, Colombia

[3] Data Analytics Department, Washington State University, Pullman, Washington 99163, U.S.A.

*Corresponding author: diego.rincon@wsu.edu

Open research statement

The code produced for this study is available at https://github.com/rincondf/STBP.

Key words: Bayes' theorem, conditional probability, population management, probability theory, variable-sample-size sequential probability ratio test.

Abstract

Making inferences about population size is paramount in ecology and pest management, with managers often seeking to determine if a given population is above or below a pre-established threshold through sampling or monitoring. Sequential data analysis is an appealing statistical field for monitoring and decision-making as it is more cost-efficient than fixed-sample-size approaches. However, a key limitation of existing sequential testing procedures is that they require the specification of two non-complementary competing hypotheses to allow for sequential calculation of probability ratios as sampling proceeds. Here we develop an approach to overcome this limitation by using Bayes' theorem to sequentially update the posterior probability of a tested hypothesis against its complementary as new data is collected. The new test can explicitly consider simple or composite hypotheses and process either purely sequential (one-at-a-time) or group sequential data to produce a trajectory of posterior probabilities that can aid in making decisions related to the tested hypothesis. The efficiency of our new test is demonstrated with three case studies that involve inferences about populations in a single or multiple time points and detection of rare species through monitoring. We conclude that the sequential test of Bayesian posterior probabilities offers an equivalent or superior alternative to probability ratios or fixed-sample-size designs that is easily interpretable for decision-making even when the sampling is stopped before a predefined decision threshold is reached.

Introduction

Inferences about population size are key to guiding interventions for pest control, biological invasions, and conservation (Binns et al. 2000, Williams et al. 2002, Barnes et al. 2021). Data from sampling or monitoring is often used to infer whether populations are above a threshold where an action should be taken, such as economic thresholds for applying pesticides (Binns and Nyrop 1992). Sequential data analysis is an appealing approach for assessing population size data against thresholds because hypothesis testing is performed whenever new information is collected without requiring a fixed sampling size specified in advance (Burr 1976, Lai 2001, Lai et al. 2012). Sequential designs are typically more cost-efficient than fixed-sample-size approaches because they require substantially smaller sample sizes to achieve a given statistical power (Wald and Wolfowitz 1948, Schönbrodt et al. 2017).

Most sequential tests monitor the evidence in favor of two competing models by updating some version of the likelihood ratio under one model compared to another each time new data is collected. Models are often specified as simple hypotheses (i.e., single values for population size) and sampling continues as long as the monitored likelihood ratio falls between predefined stopping thresholds. The sequential probability ratio test is a widely used sequential hypothesis testing method due to its proven optimality in sampling costs vs losses derived from an incorrect terminal decision, and superiority to other sequential and fixed-sample tests (Wald 1945, Wald and Wolfowitz 1948, Schnuerch and Erdfelder 2020). Yet, a key limitation of this approach (and other current methods) is that it requires specification of two competing non-complementary hypotheses, which is often not relevant when monitoring is designed to assess if a population is above or below a threshold (Lai et al. 2012). Furthermore, the optimality properties of current sequential data analysis models are restricted to testing simple hypotheses and to designs where

single data are collected every time (i.e., purely sequential sampling), and these properties may not hold for designs that imply the sequential collection of groups of data (i.e., group sequential sampling) (Morgan and Cressie 1997, Lai 2001).

Modifications of the sequential probability ratio test involve a test where the alternative hypothesis is not explicit but implicitly derived from a predefined effect and maximum sample size (Pramanik et al. 2021). Other modified versions of the sequential probability ratio test account for group sequential sampling with constant sample size or include weighting parameters to account for composite hypotheses (Wald 1945, 1947, Binns et al. 2000). Researchers have also proposed variable-sample-size probability ratio tests for purely sequential and group sequential designs, which apply Bayes' theorem to sequentially calculate the posterior probability of a hypotheses relative to an alternative (Cressie and Morgan 1993). Stopping thresholds or sample sizes for this approach are not predefined but are calculated sequentially based on the balance between the cost of new samples and that of making an incorrect decision. The variable-sample-size probability ratio test has been shown to inherit most of sequential probability ratio test's optimality properties, and offer a more general and cost-efficient approach to sequential analysis (Morgan and Cressie 1997).

Here we use the variable-sample-size probability ratio test approach to handle explicit complementary hypotheses (i.e., above or below a threshold) on dynamic or static populations, without requiring specifying a pair of competing, non-complementary hypotheses. Our approach, the sequential test of Bayesian posterior probabilities, proposes that stopping thresholds should be predefined based on the maximum available sampling effort and desired statistical power. We demonstrate the utility of the new test with three case studies that involve management of biological populations, two involving pests and one involving detection of a rare species. We

show that the sequential test of Bayesian posterior probabilities offers a superior alternative to probability ratios or fixed-sample-size designs to make inferences about population size that are more interpretable for decision-making even when sampling is stopped before a predefined decision threshold is reached.

Methods

Model derivation

In the simplest case, the aim is to test

$H_0{:}\Omega \geq \psi$ against its complementary $H_1{:}\Omega < \psi$

where Ω is the true size of a sampled population and ψ is a size threshold above which an intervention is required. By applying the Bayes' theorem to determine the credibility for H_0 under the condition of an independent and identically distributed instance $x \in X$ from the population Ω, we get:

$$P(H_0|X=x) = \frac{P(X=x|H_0)P(H_0)}{P(X=x|\neg H_0)P(\neg H_0)+P(X=x|H_0)P(H_0)} \quad \text{(Eq. 1a).}$$

As H_0 and H_1 are complementary, $\neg H_0 \equiv H_1$, and Eq. 1 may also be expressed including explicitly H_0 and H_1:

$$P(\Omega \geq \psi|X=x) = \frac{P(X=x|\Omega \geq \psi)P(\Omega \geq \psi)}{P(X=x|\Omega < \psi)P(\Omega < \psi)+P(X=x|\Omega \geq \psi)P(\Omega \geq \psi)} \quad \text{(Eq. 1b).}$$

If we assign a probability p to the prior $P(\Omega \geq \psi)$, then $P(\Omega < \psi) = (1-p)$, and Eq. 1 expressed in terms of the parameter space of Ω, we get:

$$P(\Omega \geq \psi|X=x) = \frac{p\int_{\psi}^{\sup\Omega} f_X(x;\mu)\,d\mu}{(1-p)\int_{\inf\Omega}^{\psi} f_X(x;\mu)\,d\mu + p\int_{\psi}^{\sup\Omega} f_X(x;\mu)\,d\mu} \quad \text{(Eq. 2)}$$

where $f_X(\cdot)$ is a probability density or mass function with mean μ and known dispersion and p could be seen as a uniform density distribution that assigns the same probability to all possible values of Ω. Notice that likelihood of x given $f_X(\cdot)$ is integrated over the parameter space of $\mu = \{\inf\Omega,...,\sup\Omega\}$ according to Ho and Ha, such that $0 \leq \inf\Omega < \sup\Omega$ and that $\sup\Omega$ can be ∞.

If a sample x_i is collected from a total of I bouts and $i \in \{0,...,I\}$, then $p_{i+1} = P(\Omega \geq \psi | X = x_i)$ and a sequence of levels of credibility for H_0 can be produced across sampling bouts by:

$$p_{i+1} = \frac{p_i \int_{\psi}^{\sup\Omega} f_X(x_i;\mu)\, d\mu}{(1-p_i) \int_{\inf\Omega}^{\psi} f_X(x_i;\mu)\, d\mu + p_i \int_{\psi}^{\sup\Omega} f_X(x_i;\mu)\, d\mu} \quad \text{(Eq. 3)},$$

which is similar to the decision maker's posterior probability (Cressie and Morgan 1993), except that Eq. 3 includes explicitly the parameter space associated with H_0 and its complementary H_1.

Ideally, sequential sampling should continue as long as $p_L < p_i < p_U$, where p_L and p_U are predefined upper and lower thresholds to decide in favor or against H_0, but sampling may also be terminated before when a maximum sampling size is reached. In the latter case, the value of p_i can inform decisions as it refers to the absolute level of credibility of H_0 with the data at hand and can be directly associated with the cost of triggering an action derived from accepting H_0 or otherwise. If one knows (i) L_{00}, the cost of deciding in favor of H_0 when H_0 is true ($\Omega \geq \psi$); (ii) L_{11}, the cost of deciding in favor of H_1 (against H_0) when H_1 is true ($\Omega < \psi$); (iii) L_{01}, the cost of deciding in favor of H_0 when H_1 is true ($\Omega < \psi$); and (iv) L_{10}, the cost of deciding in favor of H_1 when H_0 is true ($\Omega \geq \psi$), then the optimal decision rule in favor of H_0 occurs when (following Cressie and Morgan 1993):

$$L_{00}p_I + L_{01}(1-p_I) > L_{10}p_I + L_{11}(1-p_I) \quad \text{(Eq. 4a)},$$

105 or in favor of H_1 (against H_0) when:

106 $$L_{00}p_I + L_{01}(1 - p_I) \leq L_{10}p_I + L_{11}(1 - p_I) \quad \text{(Eq. 4b).}$$

107 On the other hand, when sampling can be continued because the maximum sample size
108 has not been reached, the definitions of p_L and p_U are analog to the selection of α and β for
109 sequential probability ratio tests such that $\alpha \approx p_L$ and $\beta \approx 1 - p_U$, and indicate probabilities of
110 committing type I or type II error when H_1 is true or when H_0 is true, respectively. Smaller
111 values for p_L and $1 - p_U$ result in greater numbers of sampling bouts and more powerful tests.
112 As is advised when designing sequential probability ratio tests for the selection of α and β, the
113 selection of p_L and $1 - p_U$ for the sequential test of Bayesian posterior probabilities should be
114 based on the operating characteristics curve and the average number of samples function to
115 ensure an appropriate balance between required precision and the logistically possible maximum
116 sampling effort (Wald 1945).

117 Except for p_0, the entire set of values for p_i are estimated sequentially from the data. The
118 chosen value for p_0 is entirely based on prior knowledge about the system and reasonable
119 biological expectations for Ω before the sampling begins. For example, when there is no
120 available information about the system before sampling starts, it is advised to use $p_0 = 0.5$, but
121 when there is reliable information about the credibility for H_0, values different from 0.5 can be
122 assigned to p_0, which may reduce the required sampling effort and increase accuracy.

123 We validated the sequential test of Bayesian posterior probabilities with three case
124 studies that involve management of biological populations. The first case addresses a common
125 situation in agriculture where managers need to assess whether a pest population is above a
126 predefined economic threshold at a single point in time. The second deals with population
127 monitoring through regular sampling to determine whether abundance values are compatible

with, or above, undesirable trajectories from forecast models. The third case focuses on the sample size required to monitor presence of rare species over time with a prespecified probability of detection.

Case 1: Testing static population sizes through purely sequential sampling

Prior research developed a sequential probability ratio test to assist the testing of whether a given tomato leafminer, *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae), population is above a predefined economic threshold at specific time points (Rincon et al. 2021). The model used an economic threshold of 9 larvae per plant, and assumed a negative binomial distribution of tomato leafminer counts among tomato plants within greenhouses. Since a sequential probability ratio test requires the specification of two non-complementary hypotheses to operate, these were set to H_0:$\omega = 8$ larvae per plant, and H_1:$\omega = 10$ larvae per plant, where $\omega \in \Omega$, and probability of committing type I and type II error, α and β, were both set to 0.1. The idea is to count the number of tomato leafminer larvae in a single, randomly selected tomato plant within a tomato greenhouse, calculate the ratio between the probability of getting the data if H_1 is true and if H_0 is true, and repeat the process until one of two predefined stop thresholds is reached. The upper threshold to decide in favor of H_1 and recommend a control to prevent yield loss is given by $(1-\beta)/\alpha$ and the lower to decide in favor of H_0 and recommend doing nothing by $\beta/(1-\alpha)$. To facilitate the sequential calculation of likelihood ratios and comparison with the stop thresholds for sequential probability ratio test sampling plans, two parallel ascending practical stopping thresholds based on cumulative counts can be established for several probability density functions. We used the formulae provided by Binns et al. (2000) for the sequential probability ratio test stopping lines with a negative binomial distribution and the variance-mean model

provided by Rincon et al. (2021) to estimate the parameter k as a function of mean tomato leafminer larvae.

The model for the sequential test of Bayesian posterior probabilities was derived from H_0 :$\Omega \geq 9$ larvae per plant, against its complementary H_1:$\Omega < 9$ larvae per plant, so posterior probabilities were estimated as:

$$p_{i+1} = \frac{p_i \int_9^\infty NB_X(x_i;\mu,k)\, d\mu}{(1-p_i)\int_0^9 NB_X(x_i;\mu,k)\, d\mu + p_i \int_9^\infty NB_X(x_i;\mu,k)\, d\mu} \quad \text{(Eq. 5)}$$

where $NB_X(\cdot)$ denotes a negative binomial density function with mean μ that describes tomato leafminer counts $X = \{x_1, ..., x_I\}$, and k is the dispersion parameter for $\mu = 9$ estimated from the variance-mean model fit by Rincon et al. (2021). The values for p_L and $1 - p_U$ were both set to 0.01, as preliminary trials resulted in similar power levels as those obtained with α and β both set to 0.1 for the sequential probability ratio test.

To compare performance of the sequential probability ratio test against the sequential test of Bayesian posterior probabilities, we used a variance-mean model to fit a negative binomial distribution with a varying dispersion parameter k and generate tomato leafminer counts across a range of 13 means, from 1 to 13 tomato leafminer larvae per plant (Rincon et al. 2021). We then sampled randomly from the computer-generated counts using the sequential probability ratio test and the sequential test of Bayesian posterior probabilities frameworks to determine the proportion of correct decisions, and the average required number of samples, for each test across various tomato leafminer population sizes. We then assessed the sequential test of Bayesian posterior probabilities under three scenarios: (i) when the initial prior is naïve and is set $p_0 = 0.5$ for all values of Ω, (ii) when the prior is incorrect (uninformative) so that $p_0 = 0.9$ for $\Omega < 9$ and $p_0 = 0.1$ for $\Omega \geq 9$, and (iii) when the prior is correct (informative) so that and $p_0 = 0.1$ for Ω

< 9, $p_0 = 0.5$ for $\Omega = 9$ and $p_0 = 0.9$ for $\Omega > 9$. We ran 1,000 simulations for each combination of sequential analysis, tomato leafminer density, and initial prior (only for the sequential test of Bayesian posterior probabilities).

Case 2: Testing dynamic population sizes through group sequential sampling

Another study used a time-sequential probability ratio test that dealt with composite hypotheses and group sequential sampling to produce a warning for a potential outbreak of green cloverworm, *Hypena scabra* (F.) (Lepidoptera: Erebidae), on soybeans (Pedigo and van Schaik 1984). The idea was to monitor green cloverworm adults at 2 to 3-day intervals with pheromone traps to determine early enough if the sampled population is consistent with an outbreak that requires additional sampling and eventually a control measure. In this case, data is collected in groups (sampling bouts, *N*) made of *n* pheromone traps recorded at regular intervals, and the goal is to determine after the least possible number of sampling bouts if the sampled population is consistent with an outbreak. Since the time-sequential probability ratio test requires two non-complementary competing hypotheses to operate, Pedigo and van Schaik (1984) described an outbreak, $\boldsymbol{U}_O = \{u_1,...,u_I\}$, and an "endemic", $\boldsymbol{U}_E = \{v_1,...,v_I\}$, population trajectory model, each made of nine sample periods ($I = 9$) of moths captured in pheromone traps per 0.1 ha, such that H_0:$\boldsymbol{\omega} = \boldsymbol{U}_E$, H_1:$\boldsymbol{\omega} = \boldsymbol{U}_O$ (Table S1), where $\boldsymbol{\omega}$ is one of two possible states of the true population trajectory $\boldsymbol{\Omega}$, and α and β were both set to 0.01. Notice that here $\boldsymbol{\omega} = \{X_1,...,X_I\}$, where each X_i is a collection of counts x_{ij} each from the *j*th trap out of a total *n*, and differs from ω in that sampling bouts, denoted as *i*, are collected over time, while the latter refers to a population size at a single point in time.

The time-sequential probability ratio test is similar to the sequential probability ratio test, except that stop thresholds and cumulative counts are weighted according to relative differences

between u_i and v_i. The upper stop threshold for counts described with a negative binomial distribution, such as those for green cloverworm captured in pheromone traps, is then given by:

$$\log\left(\frac{1-\beta}{\alpha}\right) + k\sum_1^I n \log\left(\frac{k+u_i}{k+v_i}\right) \quad \text{(Eq. 6a)}$$

the lower by:

$$\log\left(\frac{\beta}{1-\alpha}\right) + k\sum_1^I n \log\left(\frac{k+u_i}{k+v_i}\right) \quad \text{(Eq. 6b)}$$

and the sampled counts are adjusted as:

$$\sum\left[\log\left(\frac{u_i}{v_i}\right) - \log\left(\frac{k+u_i}{k+v_i}\right)\right] \times \sum_1^n x_{ij} \quad \text{(Eq. 7)}$$

where k is the dispersion parameter and was set to a common value for all population densities, $k = 1.16$, as described by Pedigo and van Schaik (1984) and $\sum_1^n x_{ij}$ is the total counts at the ith sampling bout. Notice that the time-sequential probability ratio test uses cumulative total counts, ignoring the variability among subsamples and assumes a constant n across sampling bouts.

Since the sequential test of Bayesian posterior probabilities does not require the specification of an "endemic" population model, to make a fair comparison with the time-sequential probability ratio test, the sequential test of Bayesian posterior probabilities was set to test whether $\boldsymbol{\Omega}$ was above or below a trajectory that is in between $\boldsymbol{U}_O$ and $\boldsymbol{U}_E$. Such trajectory can be defined as:

$$\boldsymbol{U}_T(\xi) = \boldsymbol{U}_E^{(1-\xi)} \times \boldsymbol{U}_O^{\xi} \quad \text{(Eq. 8)}$$

where ξ is the outbreak severity and ranges from 0 to 1 and indicates how close a resulting trajectory $\boldsymbol{U}_T(\xi)$ is from $\boldsymbol{U}_O$, such that $\boldsymbol{U}_T(\xi = 0) = \boldsymbol{U}_E$, and $\boldsymbol{U}_T(\xi = 1) = \boldsymbol{U}_O$. Thus, the

sequential test of Bayesian posterior probabilities was derived from H_0:$\boldsymbol{\Omega} \geq \boldsymbol{U}_T(\xi = 0.5)$, against its complementary H_1:$\boldsymbol{\Omega} < \boldsymbol{U}_T(\xi = 0.5)$, so posterior probabilities were estimated as:

$$p_{i+1} = \frac{p_i \int_{r_i}^{\infty} \prod_1^n NB_X(x_{ij};\mu,k)\, d\mu}{(1-p_i)\int_0^{r_i} \prod_1^n (x_{ij};\mu,k)\, d\mu + p_i \int_{r_i}^{\infty} \prod_1^n NB_X(x_{ij};\mu,k)\, d\mu} \quad \text{(Eq. 9)}$$

where $\boldsymbol{U}_T(\xi = 0.5) = \{r_1,...,r_I\}$. The values for p_L and $1 - p_U$ were both set to 0.1, as preliminary trials resulted in similar power levels as those obtained with α and β both set to 0.01 for the time-sequential probability ratio test. Notice that Eq. 9 explicitly incorporates every observation, x_{ij}, and that sample size is allowed to vary across sampling bouts.

To compare performance of the time-sequential probability ratio test against the sequential test of Bayesian posterior probabilities, we used Eq. 8 to generate test trajectories from $\xi = 0.1$ to $\xi = 1$ at 0.1 intervals. Count data were generated for each test trajectory at each of nine sampling periods using a negative binomial distribution that describes green cloverworm counts with a common $k = 1.16$ (Pedigo and van Schaik 1984). We then sampled randomly from the computer-generated counts using the time-sequential probability ratio test and the sequential test of Bayesian posterior probabilities frameworks to determine the proportion of correct decisions and the average required number of sampling bouts for each of the test trajectories generated across the range of values for ξ. The process was replicated using different subsample sizes, *n* (number of samples within each bout), for each sequential analysis: 5, 10, 20 and 30 subsamples. We tested the sequential test of Bayesian posterior probabilities under three scenarios: (i) when the initial prior is naïve so that is set $p_0 = 0.5$ for all values of ξ, (ii) when the prior is incorrect (uninformative) so that $p_0 = 1 - \xi$, and (iii) when the prior is correct (informative) so that and $p_0 = \xi$. We ran 1,000 simulations for each combination of sequential

235 analysis, subsample size, values for ξ, and initial prior (only for the sequential test of Bayesian
236 posterior probabilities).

237 *Case 3: Detecting rare species through monitoring*

238 One important application of the sequential test of Bayesian posterior probabilities may
239 be the development of monitoring programs aimed to explicitly test hypotheses about the
240 absence of a species in an area. Current standards are based on fixed-sample-size approaches, in
241 which the required sample size, n, is established as a function of a minimum population density
242 per sampling unit and the probability of no collecting (or observing) individuals when the
243 population density is actually > 0 (type II error, β) (Green and Young 1993). However, evidence
244 indicates that detecting rare species requires sampling over time, which is not efficiently
245 accounted for by fixed-sample-size approaches (Ma et al. 2022).

246 In the simplest case for the conventional approach, the sample size, n, required to detect a
247 rare species whose population counts among sampling units follow a Poisson distribution is
248 given by (Green and Young 1993):

249 $$n = -\frac{1}{\mu}\log\beta \qquad \text{(Eq. 10a)}$$

250 where μ is the mean density per sampling unit of the target species and β is the tolerable type II
251 error. Rearranging we obtain the probability β of not collecting any individuals as a function of μ
252 for a given sample size by:

253 $$\beta = \exp(-n\mu) \qquad \text{(Eq. 10b).}$$

254 The sequential test of Bayesian posterior probabilities to detect rare species with a
255 Poisson distribution was derived as H_0:$\Omega = 0$ individuals per sampling unit, against its

complementary H_1:$\Omega > 0$ individuals per sampling unit, so that posterior probabilities were estimated as:

$$p_{i+1} = \frac{p_i \prod_1^n (0^{x_{ij}}/x_{ij}!)}{(1-p_i)\int_0^\infty \prod_1^n \left(\frac{\mu^{x_{ij}} \exp(-\mu)}{x_{ij}!}\right) d\mu + p_i \prod_1^n (0^{x_{ij}}/x_{ij}!)} \quad \text{(Eq. 11)}$$

where x_{ij} is an observation collected from the *j*th subsample in the *i*th bout. The values for p_L and $1 - p_U$ were both set to 0.0001, as preliminary trials resulted in similar sample sizes as those obtained for the fixed-sample-size approach. Notice that the numerator and the second term in the denominator of Eq. 11 can only take two values: 0 if there are any $x_{ij} > 0$ or 1 otherwise.

To compare the performance of the sequential test of Bayesian posterior probabilities against the conventional fixed-sample-size approach, we sampled from Poisson computer-generated count data with $\mu = \Omega$ varying from 0.01 to 0.2 at 0.05 intervals. For each analysis, we computed the probability β of not collecting any individuals as a function of μ for different sample sizes: 10, 20 and 30 samples for the fixed-sample-size approach, and 1, 3, 5 and 10 subsamples for the sequential test of Bayesian posterior probabilities using $p_0 = 0.5$. The probability β across values of μ was also calculated from Eq. 10b for the fixed-sample-size approach. To determine the number of sample bouts required by the sequential test of Bayesian posterior probabilities to decide in favor of the absence of a species, sampling using this sequential analysis was carried out on all-zero counts for 1, 3, 5 and 10 subsamples with $p_0 = 0.1$, $p_0 = 0.5$ and $p_0 = 0.9$. We ran 1,000 simulations for each combination of analysis approach and values for μ.

Results

Purely sequential sampling

The sequential probability ratio test and the sequential test of Bayesian posterior probabilities produced incorrect decisions at different ranges around the threshold, ψ (Fig. 1a). The sequential probability ratio test error was symmetrical around ψ due to testing the probability ratio between equidistant values above and below. The fall in proportion of correct decisions around ψ becomes narrower as the values tested in the probability ratios are closer, with a pay toll in the required sample size (Binns et al. 2000). In contrast, most of the sequential test of Bayesian posterior probabilities model's incorrect decisions were produced below ψ (type I error) and decreased when the initial prior p_0 provides reliable information about H_0 (Fig. 1a). In general, the sequential probability ratio test produced fewer incorrect decisions, with an average of 0.052 across the range of tested tomato leafminer population sizes, compared to 0.116, 0.163 and 0.220 from sequential tests of Bayesian posterior probabilities with informative, naïve, and uninformative initial priors, respectively. The average proportion of incorrect decisions for the sequential probability ratio test when the tomato leafminer population was $< \psi = 9$ was 0.014, and for sequential tests of Bayesian posterior probabilities was 0.184, 0.256, and 0.323 with informative, naïve, and uninformative initial priors, respectively. Yet, type II error (incorrectly deciding in favor of $\Omega < 9$) was more common for the sequential probability ratio test with average incorrect decisions of 0.113 when the tomato leafminer population was ≥ 9, compared with 0.008, 0.014, and 0.056 with the sequential test of Bayesian posterior probabilities with informative, naïve, and uninformative initial priors, respectively (Fig. 1a).

Like the proportion of correct decisions, the average number of samples required to reach a decision for both the sequential probability ratio test and the sequential test of Bayesian

298 posterior probabilities peaked at different ranges around $\psi = 9$ (Fig. 1b). While the average
299 numbers of required samples for the sequential probability ratio test peaked symmetrically
300 around ψ, those for the sequential test of Bayesian posterior probabilities peaked earlier and
301 remained relatively constant across the range of tomato leafminer population sizes. In general,
302 the sequential test of Bayesian posterior probabilities required about half the number of samples
303 to reach a decision compared to the sequential probability ratio test. The sequential test of
304 Bayesian posterior probabilities averaged 4.531, 5.763, and 6.141 across the range of tomato
305 leafminer population sizes with informative, naïve, and uninformative initial priors, respectively,
306 while the average number of samples required by the sequential probability ratio test was 13.972
307 (Fig. 1b).

308 *Group sequential sampling*

309 We found that both the time-sequential probability ratio test and the sequential test of
310 Bayesian posterior probabilities tend to produce similar patterns of misclassification around the
311 midpoint between a full outbreak and a non-outbreak population. However, the sequential test of
312 Bayesian posterior probabilities is more sensitive to the number of subsamples in each sampling
313 bout (Fig. 2a-c). For example, the average proportion of incorrect decisions across the entire
314 range of outbreak severities for the time-sequential probability ratio test varied with the
315 subsample size between 0.106 and 0.124, while it varied between 0.087 and 0.253 for the
316 sequential test of Bayesian posterior probabilities with naïve initial priors. In general, the time-
317 sequential probability ratio test produced slightly less incorrect decisions than the sequential test
318 of Bayesian posterior probabilities, with an average of 0.113 across all subsample sizes and
319 outbreak severities, while the sequential test of Bayesian posterior probabilities produced 0.134,
320 0.149 and 0.176 with informative, naïve, and uninformative initial priors, respectively. However,

the differences between the overall rate of incorrect decisions were heavily influenced by the poor performance of the sequential test of Bayesian posterior probabilities with $n = 5$, and similar rates to that observed for the time-sequential probability ratio test are obtained after excluding the smallest sample size (0.106, 0.114 and 0.128 for informative, naïve, and uninformative initial priors, respectively). The time-sequential probability ratio test performed better avoiding type I error with an overall rate of incorrect decisions of 0.075 when the outbreak severity is < 0.5, compared to 0.131, 0.167 and 0.209 obtained for the sequential test of Bayesian posterior probabilities using informative, naïve, and uninformative initial priors, respectively. However, most of the recorded type I error for the sequential test of Bayesian posterior probabilities was influenced by the poor performance of the test when $n = 5$, and similar values to that observed for the time-sequential probability ratio test are obtained after excluding this sample size from analysis (0.072, 0.091 and 0.111 for informative, naïve, and uninformative initial priors). Both tests performed similarly well dealing with type II error with 0.135, 0.137, 0.154 for the sequential test of Bayesian posterior probabilities with informative, naïve, and uninformative initial priors, respectively, and 0.138 for the time-sequential probability ratio test (Fig. 2a-c).

While the average number of sampling bouts required to reach a decision with the time-sequential probability ratio test across outbreak severities only varied in magnitude with the subsample size, for the sequential test of Bayesian posterior probabilities converged to a similar pattern for outbreak severities ≥ 0.5 (Fig. 2d-f). In general, the sequential test of Bayesian posterior probabilities required less sampling bouts with 2.18, 2.60, and 2.65 average bouts pooled over all subsample sizes and outbreak severities for informative, naïve, and uninformative initial priors, respectively, compared to the time-sequential probability ratio test with 3.23. The difference in required sample size between time-sequential probability ratio test and sequential

test of Bayesian posterior probabilities is especially noticeable when $\boldsymbol{\Omega} \geq \boldsymbol{U}_T(\xi = 0.5)$ (Fig. 2d-f), which is also when this test commits less classification mistakes (Fig. 2a-c).

Detecting rare species

In general, monitoring programs using sequential tests of Bayesian posterior probabilities were more efficient detecting rare species than a one-time, fixed-sample-size approach. Simulations showed the probability of not collecting any individuals when a species is present, β, decreases more rapidly with the mean density of the species when samples are analyzed sequentially (Fig. 3a). Approximately 30 samples are required with a fixed-sample-size approach to achieve a power of $1 - \beta = 0.95$ to infer that there are < 0.1 individuals per sampling unit in an area. Use of monitoring with a sequential test of Bayesian posterior probabilities offers at least two options to increase power with reduced sampling effort: (i) sampling an average of 2.28 bouts each with five samples increases power to $1 - \beta = 0.993$; or (ii) sampling an average of 1.49 bouts each with ten samples increases power to $1 - \beta > 0.999$ (Fig. 3). Roughly two bouts (mean = 1.94) of 10 samples each without collecting or observing any individuals allows to infer that the density of the species is < 0.05 individuals per sampling unit, with 0.027 risk of being wrong (Fig. 4). We also found that purely sequential monitoring programs (i.e., one-at-a-time) are too inaccurate to produce reliable conclusions about the presence of rare species, although performed better than a fixed-sample-size of ≤ 10 (Fig. 3).

As the sequential test of Bayesian posterior probabilities does not require to set a maximum sampling effort, we calculated the number of sampling bouts required to reach a decision when the species is absent ($\Omega = 0$). We found the required number of bouts with all zeros depended on the subsample size and the initial priors, except for purely sequential sampling, which required 20 bouts to reach a decision regardless of the initial priors (Fig. 4). As

expected, the numbers of bouts were smallest when initial priors had high credibility for H_0 :$\Omega = 0$, and required more bouts when they were low as more evidence was required to counteract initial beliefs about H_0. The most efficient subsample size was five, as it provided power levels comparable with the conventional 30 fixed-sample-size (0.99 vs 0.95 of the fixed-sample-size), while requiring a maximum of five bouts (25 samples total) when $\Omega = 0$ for $p_0 = 0.5$. In contrast, when the subsample size is ten, although power levels increase to close to 1, no less than three bouts are required when $\Omega = 0$, even if initial priors for H_0 are low, which does not translate in any gains in sampling effort compared to the conventional fixed-sample-size approach with a comparable power level (n = 30) (Fig. 4).

Discussion

We present a novel generalized approach to sequentially test complementary hypotheses about population size in ecology. Previous approaches are either based on probability ratios, which require the specification of two non-complementary hypotheses (Wald 1945, Cressie and Morgan 1993), or aimed to estimate population sizes at fixed precision levels at the cost of being labor and cost intensive (Kuno 1969, Green 1970). Our test, the sequential test of Bayesian posterior probabilities, is based on Bayes' theorem to update the credibility of the hypothesis as new data is collected. We demonstrated the test with three case studies that involve purely sequential sampling to decide if a population is above a predefined threshold, group sequential sampling to decide if a population is compatible with a hypothetical trajectory, and group sequential sampling to infer presence of a species in an area. The sequential test of Bayesian posterior probabilities shares some philosophical properties of previous probability ratio approaches, such as updating posterior probabilities of the variable-sample-size probability ratio

tests (Cressie and Morgan 1993) and the predefinition of stop thresholds based on tolerable type I and type II error rates of the sequential probability ratio test (Wald 1945).

We note the sequential test of Bayesian posterior probabilities has some limitations dealing with type I error for purely sequential designs. Type I error can be ten times higher with a sequential test of Bayesian posterior probabilities compared with the sequential probability ratio test, as it tends to decide in favor of H_0:$\Omega \geq \psi$ when Ω is still about 30% below ψ. This trend persists even if values for p_L and $1 - p_U$ smaller than 0.01 are selected. For example, if both p_L and $1 - p_U$ are set to 0.0005, further gains in reduction of type II error are observed with average decision error rates of 0.002, 0.002 and 0.006, with only marginal improvement in type I error rates with 0.223, 0.251 and 0.291 using informative, naïve, and uninformative initial priors, respectively. Furthermore, the reduction of p_L and $1 - p_U$ to 0.0005 also entailed increases in sampling size close to 100%. Even if H_0 is swapped to H_0:$\Omega \leq \psi$ high counts obtained randomly from sampling when Ω is still below ψ tend to deflate p_i and maintain a pattern of (type II, in this case) error rates like that shown in Fig. 1a for H_0:$\Omega \geq \psi$ (Fig. S1). Thus, purely sequential sampling using a sequential test of Bayesian posterior probabilities is biased in favor of $\Omega \geq \psi$ when Ω is $\approx\leq$ 30% below a hypothesized population size regardless of selected values for p_L and $1 - p_U$ or the way H_0 is formulated. This is somewhat compensated for by the lower average error rate produced when $\Omega \geq \psi$ and the considerably smaller sampling effort required by a sequential test of Bayesian posterior probabilities. These properties can make the sequential test of Bayesian posterior probabilities suitable for pest management decision making because type II error (for H_0:$\Omega \geq \psi$) is more costly than type I error (i.e., not spraying when there was a pest problem vs spraying when perhaps it was not necessary) and quicker decisions that require less sampling effort will always benefit timely actions.

With group sequential sampling, the sequential test of Bayesian posterior probabilities outperforms the conventional time-sequential probability ratio test and the fixed-sample-size approach. This is especially true when the subsample size per bout is increased, since larger gains in precision are observed for the sequential test of Bayesian posterior probabilities compared with the marginal gains in both the time-sequential probability ratio test and the fixed-sample-size with proportional subsampling size increases. Group sequential designs can be useful to assist site-specific pest management decision making in crop systems that deploy monitoring networks and collect count data on pest abundance at regular intervals. For example, at least across the USA, most pome fruit producers invest in installing pheromone traps for the codling moth, *Cydia pomonella* (L.) (Lepidoptera: Tortricidae), from which counts of male captures are collected at weekly intervals (Calkins and Faust 2003). Phenology models are used to forecast proportion of codling moth individuals in a life stage based on heat accumulation (Jones et al. 2010), but can also produce population trajectories above which economic damage is expected (Rincon et al. 2024). The sequential test of Bayesian posterior probabilities could be applied to sequentially test the hypothesis of a field-collected codling moth trajectory being equal to or greater than a modeled damaging population trajectory and decide about the need and time for pest control tactics as new data is collected.

Another potential application of the sequential test of Bayesian posterior probabilities is the use of data collected from monitoring programs for invasive species to declare areas free from a given species using formal statistical inference. Efforts have focused on identifying the optimal density of traps (Bogich et al. 2008) or search effort (Mehta et al. 2007) to maximize detection, but few address the interplay between sampling effort and ability to make inferences about the absence of an invasive species from data. For example, the Forest Inventory and

435 Analysis program of the US Department of Agriculture deploys over 350,000 plots across the
436 country to monitor invasive plants (Oswalt et al. 2021). While presence is reported from
437 detection, the declaration of an area being free from an invasive plant could be assisted by the
438 sequential test of Bayesian posterior probabilities based on the number of plots in that area, the
439 number of sampling bouts carried out without observing the target species, and the detection
440 probability (μ in case study 3).

441 Deployment of monitoring networks and data collection are becoming cheaper and more
442 reliable with the onset of new technologies, such as satellite imagery or artificial intelligence, but
443 harnessing the data to assist timely decision-making is still challenging. Sequential data analyses
444 offer a cost-efficient alternative to use data collected routinely and inform management tactics
445 and resource allocation. We show that the sequential test of Bayesian posterior probabilities
446 offers an equivalent or superior alternative to previous sequential analyses for making inferences
447 about population size in ecology, with applications in pest and natural resource management.

448 **Acknowledgements**

449 We thank X. Chen and B. Run for their assistance with model derivation and notation.
450 Research was supported by USDA-NIFA award 2002-67013-36422.

451 **Conflict of Interest Statement**

452 The authors declare no conflicts of interest.

453 **References**

454 Barnes, B., F. Giannini, M. Parsa, and D. Ramsey. 2021. Inferring species absence from zero-
455 sighting records using analytical Bayesian models with population growth. Methods in
456 Ecology and Evolution **12**:2208-2220.
457 Binns, M. R., and J. P. Nyrop. 1992. Sampling insect populations for the purpose of IPM
458 decision-making. Annual Review of Entomology **37**:427-453.
459 Binns, M. R., J. P. Nyrop, and W. v. d. Werf. 2000. Sampling and monitoring in crop protection:
460 the theoretical basis for developing practical decision guides. CABI Pub., Wallingford,
461 Oxon, UK; New York, N.Y.
462 Bogich, T. L., A. M. Liebhold, and K. Shea. 2008. To sample or eradicate? A cost minimization
463 model for monitoring and managing an invasive species. Journal of Applied Ecology
464 **45**:1134-1142.
465 Burr, I. 1976. Statistical Quality Control Methods. Taylor & Francis, Boca Raton, FL.
466 Calkins, C. O., and R. J. Faust. 2003. Overview of areawide programs and the program for
467 suppression of codling moth in the western USA directed by the United States
468 Department of Agriculture—Agricultural Research Service. Pest Management Science
469 **59**:601-604.
470 Cressie, N., and P. B. Morgan. 1993. The VPRT: A Sequential Testing Procedure Dominating
471 the SPRT. Econometric Theory **9**:431-450.
472 Green, R. H. 1970. On fixed precision level sequential sampling. Researches on Population
473 Ecology **12**:249-251.
474 Green, R. H., and R. C. Young. 1993. Sampling to Detect Rare Species. Ecological Applications
475 **3**:351-356.

476 Jones, V. P., J. F. Brunner, G. G. Grove, B. Petit, G. V. Tangren, and W. E. Jones. 2010. A web-
477 based decision support system to enhance IPM programs in Washington tree fruit. Pest
478 Management Science **66**:587-595.
479 Kuno, E. 1969. A new method of sequential sampling to obtain the population estimates with a
480 fixed level of precision. Researches on Population Ecology **11**:127-136.
481 Lai, T. L. 2001. Sequential analysis: Some classical problems and new challenges. Statistica
482 Sinica **11**:303-351.
483 Lai, T. L., P. W. Lavori, and M.-C. Shih. 2012. Adaptive Trial Designs. Annual Review of
484 Pharmacology and Toxicology **52**:101-110.
485 Ma, K. C. K., C. W. McKindsey, and L. E. Johnson. 2022. Detecting Rare Species With Passive
486 Sampling Tools: Optimizing the Duration and Frequency of Sampling for Benthic Taxa.
487 Frontiers in Marine Science **9**.
488 Mehta, S. V., R. G. Haight, F. R. Homans, S. Polasky, and R. C. Venette. 2007. Optimal
489 detection and control strategies for invasive species management. Ecological Economics
490 **61**:237-245.
491 Morgan, P. B., and N. Cressie. 1997. A Comparison of the Cost-efficiencies of the Sequential,
492 Group-sequential, and Variable-sample-size-sequential Probability Ratio Tests.
493 Scandinavian Journal of Statistics **24**:181-200.
494 Oswalt, S., C. Oswalt, A. Crall, R. Rabaglia, M. K. Schwartz, and B. K. Kerns. 2021. Inventory
495 and Monitoring of Invasive Species. Pages 231-242 Invasive Species in Forests and
496 Rangelands of the United States. Springer International Publishing.

Pedigo, L. P., and J. W. van Schaik. 1984. Time-Sequential Sampling: A New Use of the Sequential Probability Ratio Test for Pest Management Decisions. Bulletin of the Entomological Society of America **30**:32-36.

Pramanik, S., V. E. Johnson, and A. Bhattacharya. 2021. A modified sequential probability ratio test. Journal of Mathematical Psychology **101**:102505.

Rincon, D. F., E. D. Esch, J. Gutierrez-Illan, M. Tesche, and D. W. Crowder. 2024. Predicting insect population dynamics by linking phenology models and monitoring data. Ecological Modelling **493**:110763.

Rincon, D. F., H. F. Rivera-Trujillo, L. Mojica-Ramos, and F. Borrero-Echeverry. 2021. Sampling plans promoting farmers' memory provide decision support in *Tuta absoluta* management. Agronomy for Sustainable Development **41**:33.

Schnuerch, M., and E. Erdfelder. 2020. Controlling decision errors with minimal costs: The sequential probability ratio t test. Psychological Methods **25**:206-226.

Schönbrodt, F. D., E.-J. Wagenmakers, M. Zehetleitner, and M. Perugini. 2017. Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. Psychological Methods **22**:322-339.

Wald, A. 1945. Sequential Tests of Statistical Hypotheses. The Annals of Mathematical Statistics **16**:117-186.

Wald, A. 1947. Sequential Analysis. Wiley, New York.

Wald, A., and J. Wolfowitz. 1948. Optimum Character of the Sequential Probability Ratio Test. The Annals of Mathematical Statistics **19**:326-339.

Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and Management of Animal Populations. Elsevier Science.

520 **Figure captions**

521 **Figure 1.** Operating characteristics from simulations of a sequential probability ratio test (black)
522 and sequential tests of Bayesian posterior probabilities (colored lines). Shown are (a) proportion
523 of correct decisions as a function of population size and (b) the average number of required
524 samples to reach a decision. The hypothesis tested is that population size is ≥ 9 (denoted by the
525 vertical line); line colors represent different initial priors: blue is informative, dark green is naïve,
526 and light green is uninformative.

527 **Figure 2.** Operating characteristics from simulations of a time-sequential probability ratio test
528 (black) and sequential tests of Bayesian posterior probabilities (colored lines). Shown are (a-c)
529 the proportions of correct decisions as a function of outbreak severity and (d-f) the average
530 numbers of required sampling bouts to reach a decision. Line patterns represent subsample size:
531 solid is $n = 5$, dashed is $n = 10$, dotted is $n = 20$ and dash-dotted is $n = 30$. Line colors represent
532 different initial priors: blue is informative, dark green is naïve, and light green is uninformative.
533 The hypothesis tested is that outbreak severity is ≥ 0.5 (denoted by the vertical line).

534 **Figure 3.** Operating characteristics from simulations of a fixed-sample-size sampling plan
535 (black) and a sequential test of Bayesian posterior probabilities (dark green) designed to detect
536 rare species in an area. Shown are (a) the probability of not collecting or observing individuals
537 when the species is present, *Beta*, as a function of the mean number of individuals per sampling
538 unit and (b) the average number of required sampling bouts to reach a decision. The hypothesis
539 tested is that population size = 0 with naïve priors. Line patterns represent subsample size: solid
540 is $n = 1$, dashed is $n = 3$, dotted is $n = 5$ and dash-dotted is $n = 10$; sample size for the fixed-
541 sample-size approach (black), dashed is $n = 10$, dotted is $n = 20$ and dash-dotted is $n = 30$. The

542 solid curve in (a) represents *Beta* as a function of the mean number of individuals per sampling
543 unit for $n = 30$.

544 **Figure 4.** Number of sampling bouts required to reach a decision for a sequential test of
545 Bayesian posterior probabilities designed to detect rare species in an area when the species is
546 absent as a function of the subsample size. Bar colors represent different initial priors used for
547 H_0: population size = 0: light green is low ($p_0 = 0.1$), dark green is naïve ($p_0 = 0.5$), and blue is
548 high ($p_0 = 0.9$).

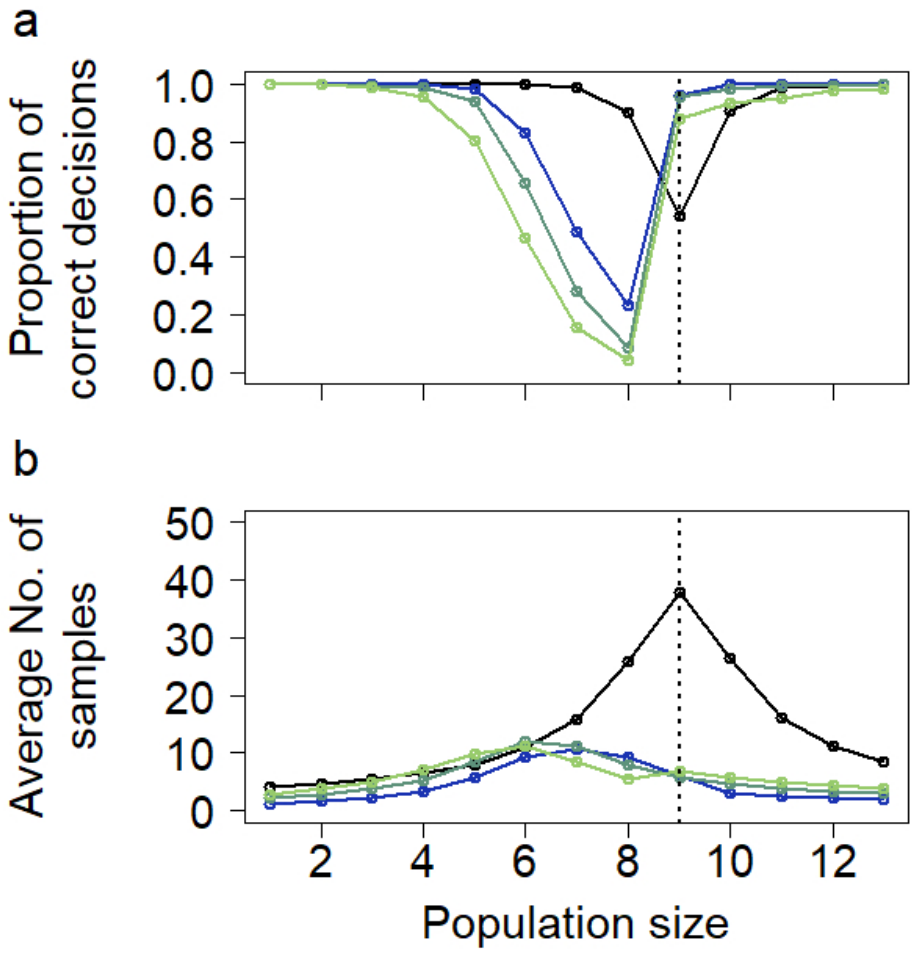

Figure 1. Operating characteristics from simulations of a sequential probability ratio test (black) and sequential tests of Bayesian posterior probabilities (colored lines). Shown are (a) proportion of correct decisions as a function of population size and (b) the average number of required samples to reach a decision. The hypothesis tested is that population size is ≥ 9 (denoted by the vertical line); line colors represent different initial priors: blue is informative, dark green is naïve, and light green is uninformative.

169x178mm (96 x 96 DPI)

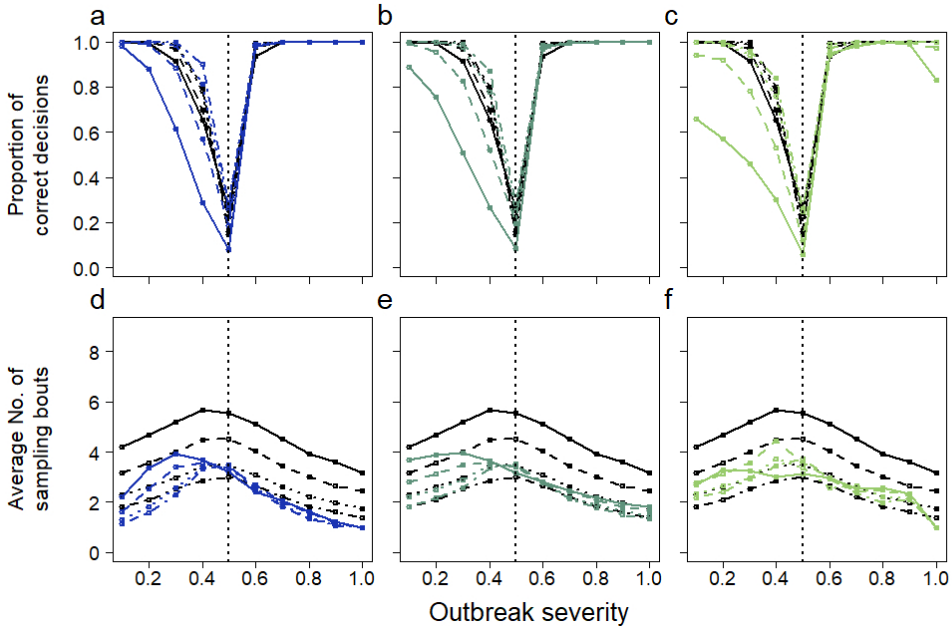

Figure 2. Operating characteristics from simulations of a time-sequential probability ratio test (black) and sequential tests of Bayesian posterior probabilities (colored lines). Shown are (a-c) the proportions of correct decisions as a function of outbreak severity and (d-f) the average numbers of required sampling bouts to reach a decision. Line patterns represent subsample size: solid is n = 5, dashed is n = 10, dotted is n = 20 and dash-dotted is n = 30. Line colors represent different initial priors: blue is informative, dark green is naïve, and light green is uninformative. The hypothesis tested is that outbreak severity is ≥ 0.5 (denoted by the vertical line).

265x171mm (96 x 96 DPI)

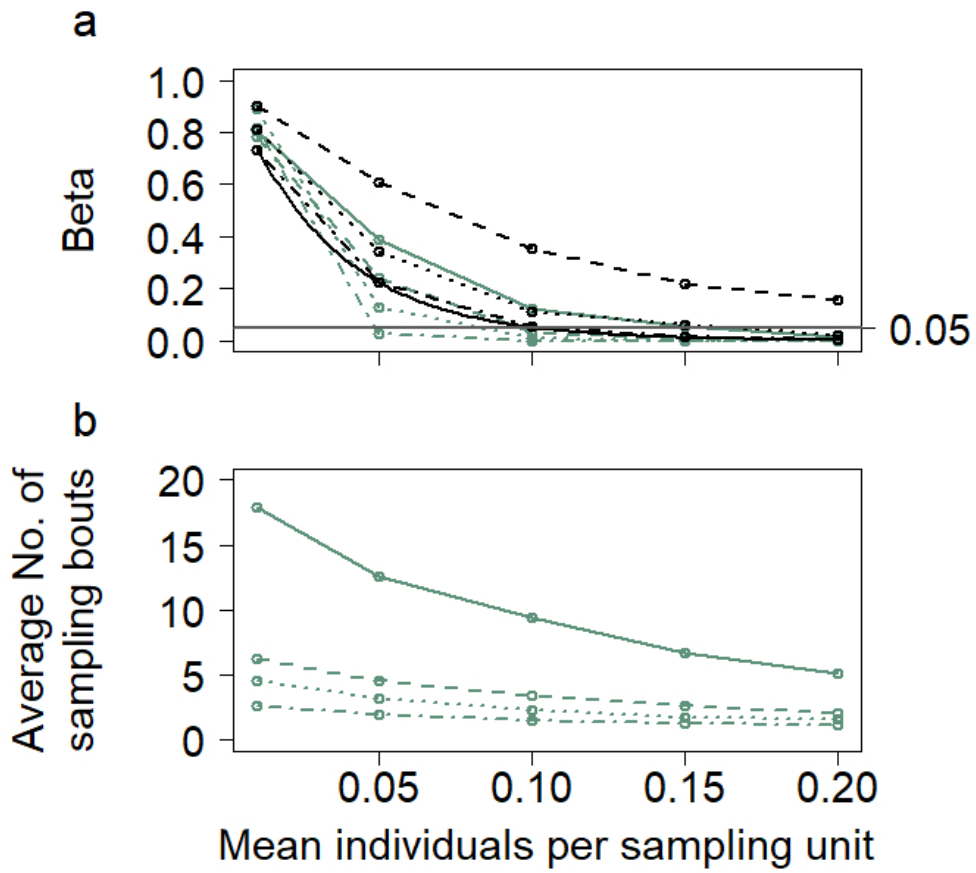

Figure 3. Operating characteristics from simulations of a fixed-sample-size sampling plan (black) and a sequential test of Bayesian posterior probabilities (dark green) designed to detect rare species in an area. Shown are (a) the probability of not collecting or observing individuals when the species is present, *Beta*, as a function of the mean number of individuals per sampling unit and (b) the average number of required sampling bouts to reach a decision. The hypothesis tested is that population size = 0 with naïve priors. Line patterns represent subsample size: solid is n = 1, dashed is n = 3, dotted is n = 5 and dash-dotted is n = 10; sample size for the fixed-sample-size approach (black), dashed is n = 10, dotted is n = 20 and dash-dotted is n = 30. The solid curve in (a) represents *Beta* as a function of the mean number of individuals per sampling unit for n = 30.

179x174mm (96 x 96 DPI)

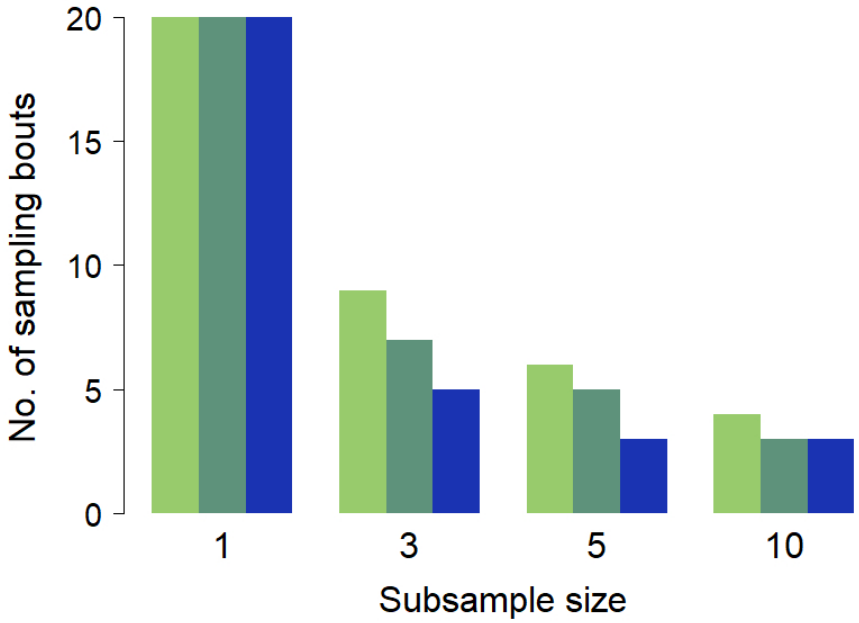

Figure 4. Number of sampling bouts required to reach a decision for a sequential test of Bayesian posterior probabilities designed to detect rare species in an area when the species is absent as a function of the subsample size. Bar colors represent different initial priors used for H_0: population size = 0: light green is low (p_0 = 0.1), dark green is naïve (p_0 = 0.5), and blue is high (p_0 = 0.9).

235x154mm (96 x 96 DPI)

Table S1. Green cloverworm, *Hypena scabra* (F.), population trajectory endemic and outbreak models on soybean in number of moths captured in pheromone traps per 0.1 ha (from Pedigo and Schaik 1984)

Sample period (i)	**Population trajectory model**	
	Endemic ($\boldsymbol{U}_E$)	Outbreak ($\boldsymbol{U}_O$)
1	2	4
2	3	5
3	4	16
4	7	18
5	8	23
6	6	38
7	3	34
8	2	26
9	1	25

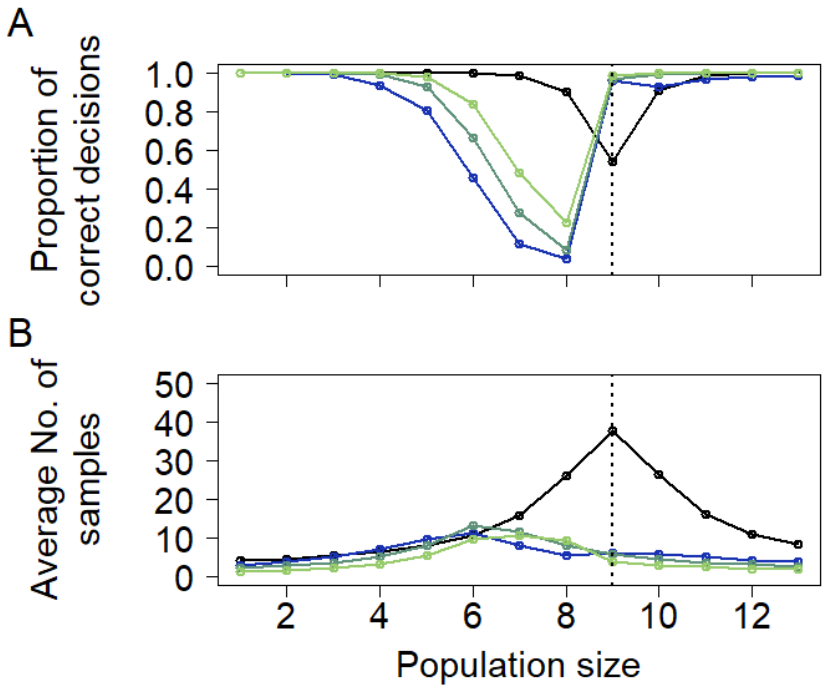

Figure S1. Operating characteristics from simulations of a sequential probability ratio test (black) and sequential tests of Bayesian posterior probabilities (colored lines). Shown are (a) the proportion of correct decisions as a function of population size and (b) the average number of required samples to reach a decision. The hypothesis tested is that population size is ≤ 9 (denoted by the vertical line), and line colors represent different initial priors: blue is informative, dark green is naïve, and light green is uninformative.