Received: 11 March 2025 | Accepted: 14 April 2025

DOI: 10.1111/2041-210X.70053

Methods in Ecology and Evolution

RESEARCH ARTICLE

Sequential testing of complementary hypotheses about population density

Diego F. Rincon[1,2] | Izzy McCabe[1] | David W. Crowder[1]

[1]Department of Entomology, Washington State University, Pullman, Washington, USA

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

Correspondence
Diego F. Rincon
Email: diego.rincon@wsu.edu

Funding information
National Institute of Food and Agriculture, Grant/Award Number: 2022-67013-36422

Handling Editor: Robin Boyd

Abstract

1. Making inferences about population density is paramount in ecology and pest management for decision-makers who often seek to determine how a population compares to a pre-established static or dynamic threshold through sampling or monitoring.
2. Sequential data analysis is appealing for monitoring and decision making as it is more cost-efficient than fixed-sample-size approaches. However, a limitation of existing sequential testing procedures is that they require specification of two non-complementary competing hypotheses to allow for sequential calculation of probability ratios as sampling proceeds.
3. We overcame this limitation by using Bayes' theorem to sequentially update the posterior probability of a tested hypothesis against its complementary as data is collected. The new test can explicitly consider simple or composite hypotheses about static or dynamic population densities and process either purely sequential (one-at-a-time) or group sequential data to produce a trajectory of posterior probabilities related to the tested hypothesis. The efficiency of our new test is demonstrated with three case studies that involve inferences about static or dynamic pest populations and the detection of rare species through monitoring.
4. Our new test, the sequential test of Bayesian posterior probabilities, offers a more efficient and accurate approach to assess if a sampled population exceeds or is below a threshold than probability ratios or fixed-sample-size approaches. Although the test requires fewer samples, more incorrect decisions may be produced for purely sequential designs when densities are below the thresholds (type I error), compared with probability ratios. We show the new test is a powerful, easily implementable framework with applications in natural resource and pest management, whose outputs are easily interpretable for decision making.

KEYWORDS
Bayes' theorem, conditional probability, population management, probability theory, variable-sample-size sequential probability ratio test

1 | INTRODUCTION

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

Most sequential tests monitor evidence in favour of two competing models by updating some version of the likelihood ratio each time new data is collected. Models are often specified as simple hypotheses about density (i.e. single values for number of individuals per sampling unit) and sampling continues as long as the monitored likelihood ratio falls between predefined stopping criteria (Lai, 2001). The sequential probability ratio test is widely used due to its optimality in sampling costs vs. losses derived from an incorrect terminal decision (Schnuerch & Erdfelder, 2020; Wald, 1945; Wald & Wolfowitz, 1948). Yet, a key limitation of this approach is that it requires the specification of two competing non-complementary hypotheses, which is often not directly relevant when assessing if a population is above a threshold (Lai et al., 2012). Furthermore, the optimality properties of current sequential 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) (Lai, 2001; Morgan & Cressie, 1997).

Modifications of the sequential probability ratio test involve a test where the alternative hypothesis is implicitly derived from a predefined effect and maximum sample size (Pramanik et al., 2021). Other modifications account for group sequential sampling with constant sample size or include weighting parameters to account for composite hypotheses (Binns et al., 2000; Wald, 1945; Wald, 1947). 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 hypothesis relative to an alternative (Cressie & Morgan, 1993). Stopping criteria or sample sizes for this approach are not predefined but 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 & Cressie, 1997).

Here we use the variable-sample-size probability ratio test approach to handle explicit simple or composite complementary hypotheses (i.e. above or below a threshold) on dynamic or static populations, without requiring a pair of competing, non-complementary hypotheses. Our approach, the sequential test of Bayesian posterior probabilities, proposes that stopping decision criteria should be predefined based on the available sampling effort and desired statistical power. We assess the 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 criterion is reached.

2 | METHODS

2.1 | Model derivation

In a simple case, the aim is to test $H_1: \mu \geq \psi$ against its complementary $H_0: \mu < \psi$, where μ is the true density of a population per sampling unit and ψ is a density threshold above which an action is taken. By applying Bayes' theorem to determine credibility for H_1 under the condition of an independent and identically distributed instance $x \in X$ from the population, we get:

$$P(H_1 | X = x) = \frac{P(X = x | H_1) P(H_1)}{P(X = x | \neg H_1) P(\neg H_1) + P(X = x | H_1) P(H_1)}, \quad (1a)$$

where $\neg H_1$ denotes the negation of H_1. As H_1 and H_0 are complementary, $\neg H_1 \equiv H_0$, and Equation (1) may also be expressed including explicitly H_1 and H_0:

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

The conditional prior density function can be defined as:

$$\pi(\mu | \mu \geq \psi) \propto 1(\mu \geq \psi) \quad (2a)$$

and

$$\pi(\mu | \mu < \psi) \propto 1(0 < \mu < \psi), \quad (2b)$$

where $\pi(\mu)$ is an improper prior density. Improper priors are prior probability densities without a finite integral that are a useful tool for Bayesian inference when there is no prior information on a proper distribution, and are more appropriate than vague proper distributions that integrate to the unity (Berger, 2000; Robert, 2007; Taraldsen & Lindqvist, 2010). If we express Equation (1) in terms of probability densities on the parameter space of μ, we get:

$$P(\mu \geq \psi | X = x) = \frac{\int_{\psi}^{\sup \mu} f_X(x;\mu) \pi(\mu) \, d\mu}{\int_{\inf \mu}^{\psi} f_X(x;\mu) \pi(\mu) \, d\mu + \int_{\psi}^{\sup \mu} f_X(x;\mu) \pi(\mu) \, d\mu}, \quad (3a)$$

where $f_X(\bullet)$ is a probability density or mass function with mean μ and known dispersion. In each of the intervals of integration, $\pi(\mu)$ is a constant, either p or $1-p$, where $p = P(\mu \geq \psi)$. Thus, by substituting:

$$P(\mu \geq \psi \mid X = x) = \frac{\int_{\psi}^{\sup \mu} f_X(x;\mu)\, p\, d\mu}{\int_{\inf\mu}^{\psi} f_X(x;\mu)(1-p)\, d\mu + \int_{\psi}^{\sup \mu} f_X(x;\mu)\, p\, d\mu}. \quad (3b)$$

As p and $1-p$ no longer depend on μ, they can be moved outside the integrals to get:

$$P(\mu \geq \psi \mid X = x) = \frac{p \int_{\psi}^{\sup \mu} f_X(x;\mu)\, d\mu}{(1-p) \int_{\inf\mu}^{\psi} f_X(x;\mu)\, d\mu + p \int_{\psi}^{\sup \mu} f_X(x;\mu)\, d\mu}. \quad (3c)$$

Notice that likelihood of x given $f_X(\bullet)$ is integrated over the parameter space of $\mu = \{\inf \mu, \ldots, \sup \mu\}$ according to H_1 and H_0, such that $0 \leq \inf \mu < \sup \mu$ and that $\sup \mu$ can be ∞. As long as $f_X(\bullet)$ (the kernel distribution) is a proper probability density function, the result of Equation (3) is always on the interval [0, 1].

If a sample x_i is collected from a total of I bouts and $i \in \{0, \ldots, I\}$, then $p_{i+1} = P(\mu \geq \psi \mid X = x_i)$ and a sequence of credibility levels for H_1 can be produced across sampling bouts by using the posterior value from the last bout as the prior for the next like so:

$$p_{i+1} = \frac{p_i \int_{\psi}^{\sup \mu} f_X(x_i;\mu)\, d\mu}{(1-p_i) \int_{\inf\mu}^{\psi} f_X(x_i;\mu)\, d\mu + p_i \int_{\psi}^{\sup \mu} f_X(x_i;\mu)\, d\mu} \quad (4)$$

which is similar to the decision maker's posterior probability (Cressie & Morgan, 1993), except that Equation (4) includes explicitly the parameter space associated with H_1 and its complementary H_0. Note that H_1 should be formulated so that p_i can be directly associated with the need for a population management action (e.g. control, quarantine, reintroduction, etc.).

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 stopping criteria to decide in favour or against H_1, but sampling may also be terminated before a predefined maximum sampling effort 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_1 with the data at hand and can be associated with the cost of triggering an action from accepting H_1 or otherwise. If one knows (i) L_{11}, the cost of deciding in favour of H_1 when H_1 is true ($\mu \geq \psi$); (ii) L_{00}, the cost of deciding in favour of H_0 (against H_1) when H_0 is true ($\mu < \psi$); (iii) L_{10}, the cost of deciding in favour of H_1 when H_0 is true ($\mu < \psi$); and (iv) L_{01}, the cost of deciding in favour of H_0 when H_1 is true ($\mu \geq \psi$), the optimal decision rule in favour of H_1 occurs when (Cressie & Morgan, 1993):

$$L_{11}p_I + L_{10}(1-p_I) < L_{01}p_I + L_{00}(1-p_I) \quad (5a)$$

or in favour of H_0 (against H_1) when:

$$L_{11}p_I + L_{10}(1-p_I) \geq L_{01}p_I + L_{00}(1-p_I), \quad (5b)$$

where p_I is the posterior of the final sampling bout from a total of I bouts.

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

Except for p_0, the entire set of values for p_i are estimated sequentially from the data. The chosen value for p_0 is based on prior knowledge of the system and expectations for μ before sampling begins. When there is no information before sampling starts, it is advised to use $p_0 = 0.5$, but when there is reliable information about the credibility for H_1, values different from 0.5 can be assigned to p_0, which may reduce the required sampling effort and increase accuracy.

We validated the sequential test of Bayesian posterior probabilities with three case studies that involve the management of biological populations. The first case addresses a common situation in agriculture where managers need to assess whether a pest population is above a predefined economic threshold at a single point in time. The second deals with population 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 the presence of rare species over time with a prespecified probability of detection.

2.2 | Case 1: Testing static population densities through purely sequential sampling

Prior research developed a sequential probability ratio test to assess if a given tomato leafminer, *Tuta absoluta* (Meyrick), population is above a static economic threshold at specific times (Rincon et al., 2021). The model used a threshold of 9 larvae per plant, and assumed a negative binomial distribution of tomato leafminer counts among tomato plants within greenhouses. Even though the aim is to test the hypothesis of $\mu \geq 9$ larvae per plant, a sequential probability ratio test requires specification of two non-complementary hypotheses. These were H_0: $\mu = 8$ larvae per plant, and H_1: $\mu = 10$ larvae per plant, and probabilities of committing type I and type II error, α and β, were both set to 0.1. The idea is to count the tomato leafminer larvae on one randomly selected tomato plant, 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 stop criteria is reached. The upper stopping criterion

to decide in favour of H_1 and recommend a control to prevent yield loss is given by $(1-\beta)/\alpha$ and the lower to decide in favour of H_0 and recommend doing nothing by $\beta/(1-\alpha)$. To facilitate sequential calculation of likelihood ratios and comparison with the stopping criterion values for sequential probability ratio tests, two parallel ascending stop lines 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 fit by Rincon et al. (2021) to estimate the dispersion parameter k as a function of μ at the threshold density as:

$$k(\mu) = \frac{\mu^2}{1.83\mu^{1.22} - \mu}. \quad (6)$$

The model for the sequential test of Bayesian posterior probabilities can be formulated to test explicitly whether the tomato leafminer population density is above the economic threshold of 9 larvae per plant. Thus, to test $H_1: \mu \geq 9$ larvae per plant, against its complementary $H_0: \mu < 9$ larvae per plant, posterior probabilities were estimated as:

$$p_{i+1} = \frac{p_i \int_9^{\infty} NB_X(x_i;\mu,k(\mu))\, d\mu}{(1-p_i)\int_0^{9} NB_X(x_i;\mu,k(\mu))\, d\mu + p_i \int_9^{\infty} NB_X(x_i;\mu,k(\mu))\, d\mu}, \quad (7)$$

where $NB_X(\bullet)$ denotes a negative binomial density function with mean μ that describes tomato leafminer counts $X = \{x_1, \ldots, x_I\}$, and $k(\mu)$ is the dispersion parameter estimated as a function of μ from Equation (6). The values for p_L and $1 - p_U$ were both set to 0.01, as it 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 negative binomial distribution with a randomly varying dispersion parameter k to 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 to determine the proportion of correct decisions, and the average required number of samples, for each test across tomato leafminer population densities. We 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 actual values of μ, (ii) when the initial prior is informative but skewed away from the correct conclusion (incorrect), so that $p_0 = 0.9$ when $\mu < 9$ and $p_0 = 0.1$ when $\mu \geq 9$, and (iii) when the initial prior is informative and skewed towards the correct conclusion (correct), so that and $p_0 = 0.1$ when $\mu < 9$, $p_0 = 0.5$ when $\mu = 9$ and $p_0 = 0.9$ when $\mu > 9$. We ran 1000 simulations for each combination of sequential analysis, tomato leafminer density, and initial prior (only for the sequential test of Bayesian posterior probabilities).

2.3 | Case 2: Testing dynamic population densities through group sequential sampling

Another study used a time-sequential probability ratio test to assess composite hypotheses on dynamic populations with group sequential sampling and predict outbreaks of green cloverworm, *Hypena scabra* (F.), on soybeans (Pedigo & van Schaik, 1984). The idea was to monitor green cloverworm adults at 2 to 3-day intervals to determine early if the sampled population is consistent with an outbreak that requires a control. In this case, data is collected in groups (sampling bouts, I) made of n traps recorded at regular intervals, and the goal is to determine after few sampling bouts if the population is consistent with an outbreak. Even though the aim is to test whether a population is equal or above an outbreak trajectory, the time-sequential probability ratio test needs two non-complementary competing hypotheses. Pedigo and van Schaik (1984) described an outbreak, $U_O = \{u_1, \ldots, u_I\}$, and an "endemic", $U_E = \{v_1, \ldots, v_I\}$, population trajectory model, each made of nine sample periods ($I = 9$) of moths captured in pheromone traps per 0.1 ha, so $H_0: \mathrm{M} = U_E$ and $H_1: \mathrm{M} = U_O$ (Table S1), where $\mathrm{M} = \{\mu_1, \ldots, \mu_I\}$ is the true population trajectory in moths per trap, and α and β were both set to 0.01. Here pheromone trap counts are $\mathbf{X} = \{X_1, \ldots, X_I\}$, where each X_i is a collection of counts x_{ij} each from the jth trap out of a total n, and differ from X in case 1 in that sampling bouts, denoted as i, are collected over time, while the latter refers to samples collected at a single point in time.

The time-sequential probability ratio test is similar to the sequential probability ratio test, except that stop lines and cumulative counts are weighted according to relative differences between u_i and v_i. The upper stopping criterion for counts described with a negative binomial distribution, such as those for green cloverworm captured in pheromone traps, is then given by:

$$\log\left(\frac{1-\beta}{\alpha}\right) + k\sum_{i=1}^{I} n \log\left(\frac{k+u_i}{k+v_i}\right) \quad (8a)$$

the lower by:

$$\log\left(\frac{\beta}{1-\alpha}\right) + k\sum_{i=1}^{I} n \log\left(\frac{k+u_i}{k+v_i}\right) \quad (8b)$$

and the sampled counts are adjusted as:

$$\sum\left[\log\left(\frac{u_i}{v_i}\right) - \log\left(\frac{k+u_i}{k+v_i}\right)\right] \times \sum_{j=1}^{n} x_{ij} \quad (9)$$

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 specification of two hypothetical models, to make a fair comparison with the time-sequential probability ratio test, the sequential test of Bayesian posterior probabilities was set to test whether **M** was above or below a trajectory that is in between U_O and U_E. Such trajectory can be defined as:

$$U_T(\xi) = \left\{ \left(u_1^{\xi} v_1^{(1-\xi)} \right), \ldots, \left(u_l^{\xi} v_l^{(1-\xi)} \right) \right\}, \tag{10}$$

where ξ is the outbreak severity, ranges from 0 to 1, and indicates the logarithmic distance from the resulting trajectory, $U_T(\xi)$, to U_E and U_O, such that $U_T(\xi = 0) = U_E$ and $U_T(\xi = 1) = U_O$. In fact, the implicit goal of time-sequential probability ratio test is to classify a given trajectory **X** as above or below the logarithmic midpoint between U_E and U_O, $U_T(\xi = 0.5)$, by sequentially updating the probability ratios for $H_0: \mathrm{M} = U_E$ and $H_1: \mathrm{M} = U_O$. Thus, the sequential test of Bayesian posterior probabilities was derived from $H_1: \mathrm{M} \geq U_T(\xi = 0.5)$, against its complementary $H_0: \mathrm{M} < U_T(\xi = 0.5)$, so posterior probabilities were estimated as:

$$p_{i+1} = \frac{p_i \int_{r_i}^{\infty} \prod_{j=1}^{n} NB_X(x_{ij}; \mu, k)\, d\mu}{(1-p_i) \int_{0}^{r_i} \prod_{j=1}^{n} (x_{ij}; \mu, k)\, d\mu + p_i \int_{r_i}^{\infty} \prod_{j=1}^{n} NB_X(x_{ij}; \mu, k)\, d\mu}, \tag{11}$$

where $U_T(\xi = 0.5) = \{r_1, \ldots, r_l\}$, $\mu \in \mathrm{M}$, and $k = 1.16$ (Pedigo & van Schaik, 1984). The values for p_L and $1 - p_U$ were both set to 0.05, as this resulted in similar power levels as with α and β set to 0.01 for the time-sequential probability ratio test. Notice Equation (11) explicitly incorporates every observation, x_{ij}, and that sample size is allowed to vary across sampling bouts.

To compare the time-sequential probability ratio test against the sequential test of Bayesian posterior probabilities, we used Equation (10) to generate test trajectories $\mathrm{M} = \{\mu_1, \ldots, \mu_l\}$ for different outbreak severities from $\xi = 0.1$ to $\xi = 1$ at 0.1 intervals. Count data were generated for each test trajectory across the nine values of μ_i using a negative binomial distribution that describes green cloverworm counts with a common $k = 1.16$ (Pedigo & van Schaik, 1984). We then sampled randomly from computer-generated counts using both tests to determine the proportion of correct decisions and the average required sampling bouts for each trajectory generated across the range of ξ. The process was replicated using different subsamples, 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 in three scenarios: (i) when the initial prior is naïve so $p_0 = 0.5$ for all ξ, (ii) when the initial prior is informative but skewed away from the correct conclusion (incorrect), so $p_0 = 1 - \xi$, and (iii) when the initial prior is informative and skewed towards the correct conclusion (correct), so $p_0 = \xi$. We ran 1000 simulations for each combination of sequential analysis, subsample size, values for ξ, and initial priors.

2.4 | Case 3: Detecting rare species through monitoring

One application of the sequential test of Bayesian posterior probabilities may be to test hypotheses about the presence or absence of a species in an area. Current standards are based on fixed-sample-size approaches, in which the required sample size, n, is established as a function of a minimum population density per sampling unit and the probability of no collecting (or observing) individuals when the population density is actually >0 (type II error, β) (Green & Young, 1993). However, evidence indicates that detecting rare species requires sampling over time, which is not efficiently accounted for by fixed-sample-size approaches (Anderson et al., 2017; Barnes et al., 2021; Ma et al., 2022).

In the conventional approach, the sample size, n, required to detect a rare species whose population counts among samples follow a Poisson distribution is (Green & Young, 1993):

$$n = -\frac{1}{\mu} \log \beta, \tag{12a}$$

where μ is the mean density per sampling unit of the target species and β is the tolerable type II error. Only type II error is of concern, because this is the failure to detect the rare species when it is actually present. When testing for existence, type I error is always 0, as even a single detection proves that there exists the target species in the sampled area. Rearranging Equation (12a), we obtain the probability β of not collecting any individuals as a function of μ for a given sample size by:

$$\beta = \exp(-n\mu). \tag{12b}$$

The sequential test of Bayesian posterior probabilities to detect rare species with a Poisson distribution was derived as a simple, static hypothesis of $H_1: \mu = 0$ individuals per sampling unit, against its complementary $H_0: \mu > 0$ individuals, so posterior probabilities were estimated as:

$$p_{i+1} = \frac{p_i \prod_{j=1}^{n} \left(0^{x_{ij}} / x_{ij}!\right)}{(1-p_i) \int_{0}^{\infty} \prod_{j=1}^{n} \left(\frac{\mu^{x_{ij}} \exp(-\mu)}{x_{ij}!} \right) d\mu + p_i \prod_{j=1}^{n} \left(0^{x_{ij}} / x_{ij}!\right)}, \tag{13}$$

where x_{ij} is an observation collected from the jth subsample in the ith 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 Equation (13) can only take two values: 0 if there are any $x_{ij} > 0$ or 1 otherwise.

To compare the sequential test of Bayesian posterior probabilities against the conventional fixed-sample-size approach, we sampled Poisson computer-generated count data with μ varying from 0.01 to 0.2 at 0.05 intervals. We then computed the probability β of not collecting any individuals as a function of μ for different sample sizes, n:

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 starting with $p_0 = 0.5$. The probability β across values of μ was calculated from Equation (12b) for the fixed-sample-size approach. To determine the sample size required by the sequential test of Bayesian posterior probabilities to decide in favour of the absence of a species, sampling using this sequential analysis was carried out on all-zero counts with $p_0 = 0.1$, $p_0 = 0.5$ and $p_0 = 0.9$. We ran 1000 simulations for each combination of analysis approach and values for μ.

3 | RESULTS

3.1 | Purely sequential sampling

The sequential probability ratio test and sequential test of Bayesian posterior probabilities produced incorrect decisions at different ranges around the threshold, ψ (Figure 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 correct decisions around ψ becomes narrower as the values tested in the probability ratios are closer, with a pay toll in 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_1 (Figure 1a). The sequential probability ratio test produced fewer incorrect decisions, with an average of 0.052 across the range of tested leafminer densities, compared to 0.074, 0.11 and 0.19 from sequential tests of Bayesian posterior probabilities with correct, naïve, and incorrect initial priors, respectively. The average proportion of incorrect decisions for the sequential probability ratio test when the leafminer population was $< \psi = 9$ (type I error) was 0.012, and for sequential tests of Bayesian posterior probabilities was 0.10, 0.16, and 0.24 with correct, naïve, and incorrect initial priors, respectively. Yet, type II error (incorrectly deciding in favour of $\mu < 9$) was more common for the sequential probability ratio test with average incorrect decisions of 0.12 when the population was ≥9, compared with 0.025, 0.030, and 0.093 with the sequential test of Bayesian posterior probabilities with correct, naïve, and incorrect initial priors, respectively (Figure 1a).

Like the proportion of correct decisions, the average number of samples required to reach a decision for both tests peaked at different ranges around $\psi = 9$ (Figure 1b). While the numbers of required samples for the sequential probability ratio test peaked symmetrically around ψ, those for the sequential test of Bayesian posterior probabilities peaked earlier and remained relatively constant across the range of population densities. The sequential test of Bayesian posterior probabilities required 4.40, 5.52, and 6.34 samples to reach a decision with correct, naïve, and incorrect initial priors, respectively, which were all considerably lower than the average number of samples required by the sequential probability ratio test (13.85) (Figure 1b).

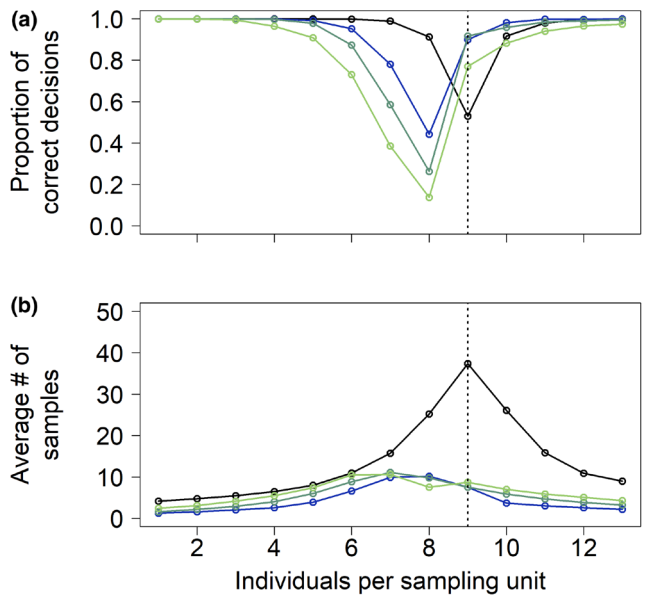

FIGURE 1 Operating characteristics from simulations of a sequential probability ratio test (black) and sequential tests of Bayesian posterior probabilities (coloured lines). Shown are (a) proportion of correct decisions as a function of population density and (b) the average number of required samples to reach a decision. The hypothesis tested is that population density is ≥9 individuals per sampling unit (denoted by the vertical line); line colours represent different initial priors: Blue for correct, dark green for naïve, and light green for incorrect.

3.2 | Group sequential sampling

Both the time-sequential probability ratio test and the sequential test of Bayesian posterior probabilities produced similar patterns of misclassification around the midpoint between a full outbreak and non-outbreak population. One exception is the sequential test of Bayesian posterior probabilities with incorrect initial priors, which produced large numbers of incorrect decisions at small and large outbreak severities (Figure 2c). Regardless of initial priors, the sequential test of Bayesian posterior probabilities is more sensitive to the number of subsamples (Figure 2a–c). For example, the average proportion of incorrect decisions across the range of outbreak severities for the time-sequential probability ratio test varied with the subsample size between 0.11 and 0.12, and from 0.10 to 0.27 for the sequential test of Bayesian posterior probabilities with naïve initial priors. The time-sequential probability ratio test produced fewer incorrect decisions than the sequential test of Bayesian posterior probabilities, with an average of 0.11 across all subsample sizes and outbreak severities, while the sequential test of Bayesian posterior probabilities produced 0.15, 0.17, and 0.30 with correct, naïve, and incorrect initial priors, respectively. The time-sequential probability ratio test also had lower type I error with an overall rate of incorrect decisions of 0.069 when the outbreak severity is <0.5, compared to 0.15, 0.20, and 0.31 obtained for the sequential test of Bayesian posterior probabilities using correct,

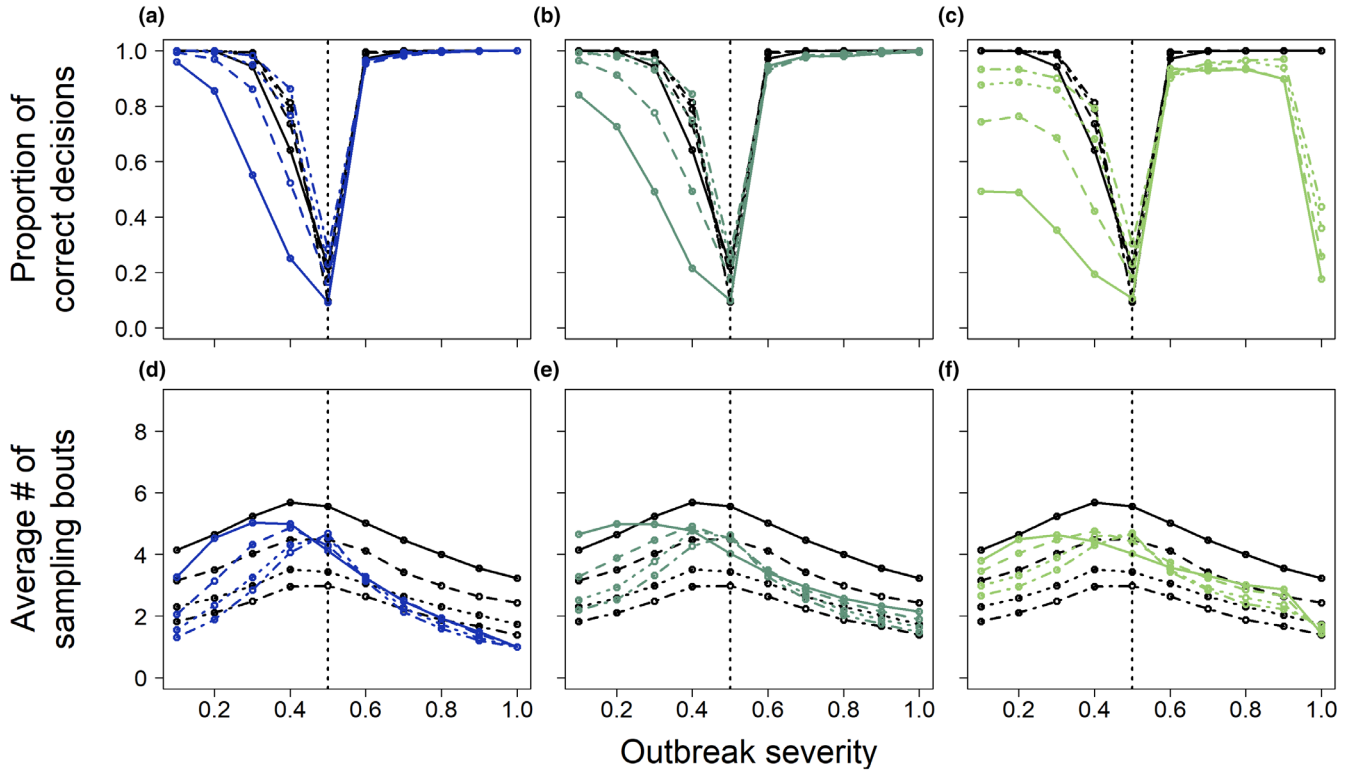

FIGURE 2 Operating characteristics from simulations of a time-sequential probability ratio test (black) and sequential tests of Bayesian posterior probabilities (coloured lines). Shown are (a–c) the proportions of correct decisions as a function of outbreak severity and (d–f) the average numbers of required 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 colours represent different initial priors: Blue for correct, dark green for naïve, and light green for incorrect. The hypothesis being tested is that outbreak severity is ≥0.5 (denoted by the vertical line).

naïve, and incorrect initial priors, respectively. However, in both cases, 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 those observed for the time-sequential probability ratio test were obtained after excluding this sample size. Both tests performed similarly with type II error, with 0.14, 0.15, and 0.29 for the sequential test of Bayesian posterior probabilities with correct, naïve, and incorrect initial priors, respectively, and 0.14 for the time-sequential probability ratio test (Figure 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 for outbreak severities ≥0.5 (Figure 2d–f). In general, the sequential test of Bayesian posterior probabilities required fewer samples with 2.75, 3.23, and 3.34 average bouts pooled over all subsample sizes and outbreak severities for correct, naïve, and incorrect initial priors, respectively, compared to the time-sequential probability ratio test with 3.24. These differences in required samples were especially noticeable when $M \geq U_T(\xi = 0.5)$ (Figure 2d–f), which is also where this test commits fewer classification mistakes (Figure 2a–c).

3.3 | Detecting rare species

In general, monitoring using sequential tests of Bayesian posterior probabilities was more efficient at detecting rare species than a fixed-sample-size approach. Simulations showed that 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 analysed sequentially (Figure 3a). From Equation (12a), 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 (solid curve in Figure 3a). 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 2.42 bouts each with five samples increases power to $1 - \beta = 0.962$; or (ii) sampling an average of 1.56 bouts each with 10 samples increases power to $1 - \beta = 0.979$ (Figure 3). Roughly two bouts (mean = 2.21) 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 0.136 risk of being wrong (Figure 3). We also found that purely sequential monitoring programmes (i.e. one-at-a-time) are too inaccurate to produce reliable conclusions about the presence of rare species, although

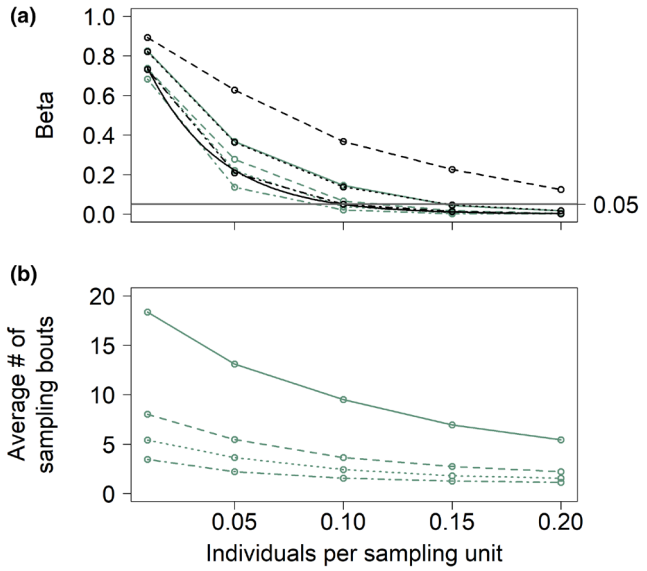

FIGURE 3 Operating characteristics from simulations of a fixed-sample-size sampling plan (black) and a sequential test of Bayesian posterior probabilities (dark green) designed to detect rare species in an area. Shown are (a) the probability of not collecting or observing individuals when the species 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 initial 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 from Equation (12b) for a fixed sample size of $n = 30$.

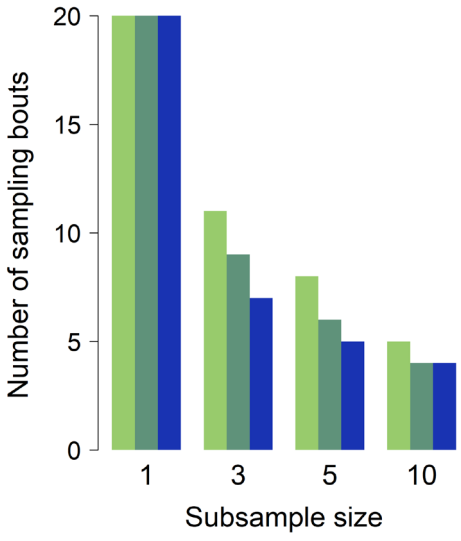

FIGURE 4 Number of sampling bouts required to reach a decision for a sequential test of Bayesian posterior probabilities designed to detect rare species in an area when the species is absent as a function of the subsample size. Bar colours represent different initial priors for the rare species being absent in the sampled area (H_1: Population density = 0): Light green is low ($p_0 = 0.1$), dark green is naïve ($p_0 = 0.5$), and blue is high ($p_0 = 0.9$).

performed similar to fixed sample sizes of $n = 20$ and better than fixed sample sizes of $n \leq 10$ (Figure 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 ($\mu = 0$). We found the required number of bouts with all zeros depends on both the subsample size and the initial priors, except for purely sequential sampling, which reached the predefined maximum of 20 bouts regardless of the initial priors (Figure 4). As expected, the numbers of bouts were smallest when initial priors had high credibility for H_1: $\mu = 0$, and required more bouts when they were low as more evidence was required to counteract initial beliefs about H_1. The most efficient subsample size was five, as it provided power levels comparable with the conventional 30 fixed-sample-size (0.96 vs. 0.95 of the fixed-sample-size), while requiring a maximum of six bouts (30 samples total) when $\mu = 0$ for $p_0 = 0.5$, but it could only take up to five bouts (25 samples total) if there are reasons to increase p_0 in favour of H_1. In contrast, when the subsample size is 10, although power levels increase to close to 1, no less than four bouts are required when $\mu = 0$, even if initial priors for H_1 are high, which does not translate in any gains in sampling effort compared to the conventional fixed-sample-size approach with a comparable power level ($n = 30$).

4 | DISCUSSION

We present a novel generalized approach to sequentially test complementary hypotheses and make inferences about population density in ecology. Previous approaches are either based on probability ratios, which require the specification of two non-complementary hypotheses (Cressie & Morgan, 1993; Wald, 1945), or aim to estimate population densities at fixed precision levels at the cost of being labor intensive (Green, 1970; Kuno, 1969). 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 static threshold, group sequential sampling to decide if a population is above a hypothetical trajectory, and group sequential sampling to infer the absence of a species in an area. The sequential test of Bayesian posterior probabilities shares some properties of previous probability ratio approaches, such as updating posterior probabilities of the variable-sample-size probability ratio tests (Cressie & Morgan, 1993) and the predefinition of stop decision criteria based on tolerable type I and type II error rates of the sequential probability ratio test (Wald, 1945).

When population densities vary widely (cases 1 and 2), discrete probability distributions that allow overdispersion are often

required to capture the distribution of the data (Binns et al., 2000). Sometimes variance–mean models are available to obtain dispersion parameters for a range of population densities (case 1), but oftentimes a single value must be used (case 2). For approaches based on probability ratios, fixed sample sizes or fixed precision levels, only one or a few values of the dispersion parameter are used at target or threshold densities, and the impact of using a common value is small (Binns et al., 2000; Madden et al., 2007). However, for the sequential test of Bayesian posterior probabilities, likelihoods are integrated over the whole range of possible values of the mean and, ideally, the associated variation of the dispersion parameter should be captured. Empirical variance–mean relationships that capture the variation of dispersion parameters across population densities may increase the efficiency of the sequential test of Bayesian posterior probabilities. For example, if a constant dispersion parameter k is used in Equation (7) for case study 1, the overall average error rate increases by 4%. However, the effect of using a common k for all population densities may vary among species (Taylor et al., 1979).

The sequential test of Bayesian posterior probabilities applied to purely sequential designs had some limitations associated with type I error (case 1), as it tends to decide in favour of $H_1: \mu \geq \psi$ when μ is ≈≤30% below ψ. This trend persists even if values for p_L and $1 - p_U$ smaller than 0.01 are used. For example, if both p_L and $1 - p_U$ are set to 0.0005, further reductions of type II error are observed with average decision error rates of 0.013, 0.012 and 0.025, with only marginal improvement in type I error rates with 0.106, 0.159 and 0.189 using correct, naïve, and incorrect initial priors, respectively. Further, the reduction of p_L and $1 - p_U$ to 0.0005 also entailed increases in sampling size close to 100%. Even if H_1 is swapped to $H_1: \mu \leq \psi$, high random counts from sampling when μ is still below ψ tend to deflate p_i and maintain a pattern of error rates like that shown in Figure 1a (Figure S1). This is somewhat compensated for by lower average error rates when $\mu \geq \psi$ (about 5 times lower with correct initial priors) and the considerably smaller sampling effort required by the sequential test of Bayesian posterior probabilities. These properties make the sequential test of Bayesian posterior probabilities suitable for pest management decision making because type II error (for $H_1: \mu \geq \psi$) is more costly than type I error (i.e. not spraying when there was a pest problem vs. spraying when it was not necessary) and because it allows for decisions with less sampling. This consistent asymmetry in error rates about threshold densities may be a result of sequential test of Bayesian posterior probabilities relying in likelihoods to obtain p_i. As maximum likelihoods differ from means in asymmetric distributions, unrealistic high p_i values are produced about the mode (not the mean) with purely sequential sampling, but the effect is ameliorated as sample size per bout increases.

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 initial priors are correct (or at least naïve), and when the subsample size per bout is increased, since larger gains in precision occur 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 approach 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, many pome fruit producers invest in installing pheromone traps for the codling moth, *Cydia pomonella* (L.), from which counts of male captures are collected at weekly intervals (Calkins & Faust, 2003). Phenology models are used to forecast the 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 trajectory being greater than a 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 programmes 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. One important exception is the approach presented by Anderson et al. (2017) who use the Bayesian logic to update the posterior probability of an area being free from a target species with each new sampling (surveillance) bout. The main difference between the method by Anderson et al. (2017) and the sequential test of Bayesian posterior probabilities is that the former calculates the likelihood of false negatives (first term in the denominator of Equation 13), $P(x = 0 | \mu > 0)(1 - p)$, from a custom detection probability while in the later it is explicitly obtained from a sampling distribution. Both approaches can be integrated to incorporate both sampling error and additional sources of missing target individuals or disease cases by adding appropriate constant or varying detection probabilities. For example, the Forest Inventory and Analysis programme of the US Department of Agriculture deploys over 350,000 surveillance plots across the country to monitor invasive plants (Oswalt et al., 2021). While presence is reported from detection, the declaration of an area being free from an invasive plant could be assisted by the sequential test of Bayesian posterior probabilities based on the number of surveillance plots in that area, the number of sampling bouts carried out without observing the target species, sampling error, and any additional sources of analytical or observational detection error.

Deployment of monitoring networks and data collection are becoming cheaper and more reliable with the onset of new technologies, such as satellite imagery or artificial intelligence, but harnessing the data to assist timely decision making is still challenging. Sequential data analyses offer a cost-efficient alternative to using data collected routinely and informing management tactics and resource allocation. We show that the sequential test of Bayesian posterior probabilities offers an equivalent or superior alternative

to previous sequential analyses for making inferences about population size in ecology, with applications in pest and natural resource management.

AUTHOR CONTRIBUTIONS

Diego F. Rincon developed the model with Izzy McCabe in consultation with David W. Crowder. Diego F. Rincon ran simulations and generated the results with input from Izzy McCabe. All authors drafted, proofread, and edited the manuscript.

ACKNOWLEDGEMENTS

We thank X. Chen and B. Run for their assistance with model derivation and notation. Research was supported by USDA-NIFA award 2022-67013-36422.

CONFLICT OF INTEREST STATEMENT

No authors have conflicts of interest to declare.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.70053.

DATA AVAILABILITY STATEMENT

All the code required to run case studies and produce figures is available as a compressed file at https://doi.org/10.5281/zenodo.15243328 or alternatively as a GitHub repository at https://github.com/rincondf/STBP (Rincon et al., 2025).

ORCID

Diego F. Rincon https://orcid.org/0000-0001-6869-4430
Izzy McCabe https://orcid.org/0009-0004-5179-0345
David W. Crowder https://orcid.org/0000-0002-3720-1581

REFERENCES

Anderson, D. P., Gormley, A. M., Ramsey, D. S. L., Nugent, G., Martin, P. A. J., Bosson, M., Livingstone, P., & Byrom, A. E. (2017). Bio-economic optimisation of surveillance to confirm broadscale eradications of invasive pests and diseases. *Biological Invasions*, *19*, 2869–2884.

Barnes, B., Giannini, F., Parsa, M., & Ramsey, D. (2021). Inferring species absence from zero-sighting records using analytical Bayesian models with population growth. *Methods in Ecology and Evolution*, *12*, 2208–2220.

Berger, J. O. (2000). Bayesian analysis: A look at today and thoughts of tomorrow. *Journal of the American Statistical Association*, *95*, 1269–1276.

Binns, M. R., & Nyrop, J. P. (1992). Sampling insect populations for the purpose of IPM decision-making. *Annual Review of Entomology*, *37*, 427–453.

Binns, M. R., Nyrop, J. P., & Werf, W.v.d. (2000). *Sampling and monitoring in crop protection: The theoretical basis for developing practical decision guides*. CABI Pub.

Bogich, T. L., Liebhold, A. M., & Shea, K. (2008). To sample or eradicate? A cost minimization model for monitoring and managing an invasive species. *Journal of Applied Ecology*, *45*, 1134–1142.

Burr, I. (1976). *Statistical quality control methods*. Taylor & Francis.

Calkins, C. O., & Faust, R. J. (2003). Overview of areawide programs and the program for suppression of codling moth in the western USA directed by the United States Department of Agriculture—Agricultural Research Service. *Pest Management Science*, *59*, 601–604.

Cressie, N., & Morgan, P. B. (1993). The VPRT: A sequential testing procedure dominating the SPRT. *Econometric Theory*, *9*, 431–450.

Green, R. H. (1970). On fixed precision level sequential sampling. *Researches on Population Ecology*, *12*, 249–251.

Green, R. H., & Young, R. C. (1993). Sampling to detect rare species. *Ecological Applications*, *3*, 351–356.

Jones, V. P., Brunner, J. F., Grove, G. G., Petit, B., Tangren, G. V., & Jones, W. E. (2010). A web-based decision support system to enhance IPM programs in Washington tree fruit. *Pest Management Science*, *66*, 587–595.

Kuno, E. (1969). A new method of sequential sampling to obtain the population estimates with a fixed level of precision. *Researches on Population Ecology*, *11*, 127–136.

Lai, T. L. (2001). Sequential analysis: Some classical problems and new challenges. *Statistica Sinica*, *11*, 303–351.

Lai, T. L., Lavori, P. W., & Shih, M.-C. (2012). Adaptive trial designs. *Annual Review of Pharmacology and Toxicology*, *52*, 101–110.

Ma, K. C. K., McKindsey, C. W., & Johnson, L. E. (2022). Detecting rare species with passive sampling tools: Optimizing the duration and frequency of sampling for benthic taxa. *Frontiers in Marine Science*, *9*, 809327. https://doi.org/10.3389/fmars.2022.809327

Madden, L. V., Hughes, G., & Bosch, F.v.d. (2007). *The study of plant disease epidemics*. American Phytopathological Society.

Mehta, S. V., Haight, R. G., Homans, F. R., Polasky, S., & Venette, R. C. (2007). Optimal detection and control strategies for invasive species management. *Ecological Economics*, *61*, 237–245.

Morgan, P. B., & Cressie, N. (1997). A comparison of the cost-efficiencies of the sequential, group-sequential, and variable-sample-size-sequential probability ratio tests. *Scandinavian Journal of Statistics*, *24*, 181–200.

Oswalt, S., Oswalt, C., Crall, A., Rabaglia, R., Schwartz, M. K., & Kerns, B. K. (2021). Inventory and monitoring of invasive species. In T. M. Poland, T. Patel-Weynand, D. M. Finch, C. F. Miniat, D. C. Hayes, & V. M. Lopez (Eds.), *Invasive species in forests and rangelands of the United States* (pp. 231–242). Springer International Publishing.

Pedigo, L. P., & van Schaik, J. W. (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., Johnson, V. E., & Bhattacharya, A. (2021). A modified sequential probability ratio test. *Journal of Mathematical Psychology*, *101*, 102505.

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

Rincon, D. F., McCabe, I., & Crowder, D. W. (2025). rincondf/STBP: Sequential testing of complementary hypotheses about population density. *Zenodo*, https://doi.org/10.5281/zenodo.15243715

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

Robert, C. (2007). *The Bayesian choice: From decision-theoretic foundations to computational implementation*. Springer.

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

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

Taraldsen, G., & Lindqvist, B. H. (2010). Improper priors are not improper. *The American Statistician*, *64*, 154–158.

Taylor, L. R., Woiwod, I. P., & Perry, J. N. (1979). Negative binomial as a dynamic ecological model for aggregation, and the density dependence of *k*. *Journal of Animal Ecology*, *48*, 289–304.
Wald, A. (1945). Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, *16*, 117–186.
Wald, A. (1947). *Sequential analysis*. Wiley.
Wald, A., & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, *19*, 326–339.
Williams, B. K., Nichols, J. D., & Conroy, M. J. (2002). *Analysis and management of animal populations*. Elsevier Science.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

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

Figure S1: Operating characteristics from simulations of a sequential probability ratio test (black lines) and sequential tests of Bayesian posterior probabilities (various coloured lines). Shown are (a) proportion of correct decisions as a function of population density and (b) average number of required samples to reach a decision. The hypothesis tested is that population density is ≤9 individuals per sampling unit (denoted by the vertical line); line colours represent different initial priors: blue for correct, dark green for naïve, and light green for incorrect.

How to cite this article: Rincon, D. F., McCabe, I., & Crowder, D. W. (2025). Sequential testing of complementary hypotheses about population density. *Methods in Ecology and Evolution*, *16*, 1228–1238. https://doi.org/10.1111/2041-210X.70053