SUPPLEMENTAL MATERIAL: Using serorsurveys to optimize surveillance for zoonotic pathogens

APPENDIX 1: SUPPLEMENT TO THE MATHEMATICAL FOUNDATION

Model (1) in the main text is a system of ordinary differential equations (ODEs) describing the ecology of a reservoir species and epidemiology of a pathogen expressed in counts. We use model (1) as the mathematical foundation of our method and to generate synthetic epidemics in a reservoir host. To find the pathogen free equilibrium for the population, $\mathcal{R}_0$, and the pathogen present equilibria for each disease compartment using model (1), we assume a constant birth rate by substituting the average birth rate $\bar{b}$ for $b(t)$. The total population size at equilibrium when the pathogen is absent using this assumption is

$$N_{eq} = \frac{\bar{b} - \mu}{k}. \tag{S1}$$

We can now use this pathogen free equilibrium to calculate $\mathcal{R}_0$ following the next generation method [1] giving us

$$\mathcal{R}_0 = \frac{\beta(\bar{b} - \mu)}{k(\bar{b} + \gamma)}. \tag{S2}$$

Using equation S1, we can also calculate the pathogen present equilibria for each compartment assuming birth rate is constant:

$$S_{eq} = \frac{\bar{b} + \gamma}{\beta} \tag{S3a}$$

$$I_{eq} = -\frac{\left(\omega_A + \bar{b}\right)\left(\bar{b} + \omega_T\right)\left(\bar{b}(k - \beta) + \beta\mu + \gamma k\right)}{\beta k\left(\omega_A\left(\bar{b} + \gamma + \omega_T\right) + \left(\bar{b} + \gamma\right)\left(\bar{b} + \omega_T\right)\right)} \tag{S3b}$$

$$R_{Aeq} = -\frac{\gamma\left(\bar{b} + \omega_T\right)\left(\bar{b}(k - \beta) + \beta\mu + \gamma k\right)}{\beta k\left(\omega_A\left(\bar{b} + \gamma + \omega_T\right) + \left(\bar{b} + \gamma\right)\left(\bar{b} + \omega_T\right)\right)} \tag{S3c}$$

$$R_{Teq} = -\frac{\gamma\omega_A(\bar{b}(k - \beta) + \beta\mu + \gamma k)}{\beta k\left(\omega_A\left(\bar{b} + \gamma + \omega_T\right) + \left(\bar{b} + \gamma\right)\left(\bar{b} + \omega_T\right)\right)}. \tag{S3d}$$

We use the pathogen present equilibria as the initial values to simulate disease dynamics in our synthetic reservoir populations. Since we focus on a reservoir animal that experiences both density independent and density dependent mortality, we need to parameterize model (1) with a value for density dependent death rate, k, such that

the population size reaches a stable equilibrium given birth rate, $b(t)$, and density independent death rate, μ. To find k that meets this criteria, we assume effective lifespan (L) must be equal to $\frac{1}{b}$, and therefore

$$k = \frac{\frac{1}{L} - \mu}{N_{eq}} \tag{S4}$$

when the population is at equilibrium. Equation S4 was used to calculate the values for k used for simulating surveillance data (table S2).

We assume serosurveillance data will most often be given in terms of proportion seropositive animals obtained within a sample. To convert model (1) to proportions we make a change of variables directly to the system of ODEs using the quotient rule to get model (2) as it appears in main text:

$$\dot{s} = \frac{\dot{S}}{\dot{N}} = \frac{N\dot{S} - \dot{N}S}{N^2} = b(t) - s\big(b(t) + \iota\beta N\big) + \omega_T r_T \tag{S5a}$$

$$\dot{\iota} = \frac{\dot{I}}{\dot{N}} = \frac{N\dot{I} - \dot{N}I}{N^2} = \iota\big(s\beta N - b(t) - \gamma\big) \tag{S5b}$$

$$\dot{r}_A = \frac{\dot{R}_A}{\dot{N}} = \frac{N\dot{R}_A - \dot{N}R_A}{N^2} = \gamma\iota - r_A\big(b(t) + \omega_A\big) \tag{S5c}$$

$$\dot{r}_T = \frac{\dot{R}_T}{\dot{N}} = \frac{N\dot{R}_T - \dot{N}R_T}{N^2} = r_A\omega_A - r_T b(t) - \omega_T r_T. \tag{S5d}$$

APPENDIX 2: SUPPLEMENT FOR THE ESTIMATION OF SEASONAL BIRTH PULSES

A function for $b(t)$

If the information is available and adding information about fluctuating birth pulses to the model is needed, below we offer a method to estimate $b(t)$. Similar to serosurveillance data, we assume a sample of reservoir animals is captured at multiple times each year. The data that can be used to estimate seasonal patterns of reproduction are diverse and may include, for example, the proportion of pregnant females, the proportion of juvenile animals, or the proportion of lactating females over time. Thus, data will consist of a discrete sampling date (i), a sample size (n_i), and the number of animals within the sample that were birthed (or a metric thereof) (x_i). Given any of these

data examples, we use the cyclic Gaussian function described and tested by [2]:

$$b(t) = g \cdot e^{-s\cos^2(\pi f t - \psi)}, \tag{S6}$$

where g controls the amplitude, s controls the shape of the cycle, f is the frequency, ψ is the phase shift, and t is continuous time.

To fit equation (S6) to data, we must estimate the unknown parameters g, s, and ψ. We assume f to be a known parameter based on natural history of a species reproductive cycle (e.g. species are known to reproduce once or twice a year).

A noticeable problem arises when fitting data on proportion captured animals that are pregnant or lactating instead of data on birth rates directly. When the estimation is performed on, for example, proportion lactating females, this specifies a function that is proportional to $b(t)$, preserving the shape and timing, but not the amplitude in units of per capita birth rates. Once the parameters are fitted, this problem can be rectified because we can adjust the value of g if we know or can estimate the average birth rate of the population $\bar{b}$ with the following equation:

$$g = \frac{\bar{b} \cdot \frac{1}{f}}{\int_0^{\frac{1}{f}} e^{-s\cos^2(\pi f t - \psi)} dt}. \tag{S7}$$

Now, with the adjusted g, the parameterized equation (S6) can be plugged into equation (3) in the main text to calculate $\hat{\iota}$ if this is needed.

Fitting $b(t)$ to data

To estimate the unknown parameters in equation (S6), we assume the likelihood of observing a temporal sequence of seroprevalence values can be specified using a binomial:

$$\mathcal{L}(\theta) = \prod_{i=1}^{\tau} \binom{n_i}{x_i} b(t)^{x_i} \big(1 - b(t)\big)^{n_i - x_i}, \tag{S8}$$

where the product is carried over τ total sampling time points, which occur over continuous time, t, and $\theta = \{g, s, \psi\}'$. For each sampling time point i, n_i defines the number of animals sampled and x_i defines the number of sampled animals found to be

"born" or a metric thereof (e.g. lactating, pregnant, juvenile, etc.).

To parameterize the birth function $b(t)$ for *E. helvum*, we fit the time series of proportion lactating individuals in the sample collected from December 2018 and November 2019 in Cameroon. We performed Bayesian estimation in rstan, which implements the Hamiltonian Monte Carlo algorithm [3]. We allowed the algorithm to run for 1,000 iterations for the simulated data and 10,000 iterations to obtain estimates on the empirical data using the likelihood function in equation (S8). Prior distributions and estimates for model parameters and are given below in table (S1). We assumed *E. helvum* has one birth pulse per year [4, 5]. Once the posterior modes were obtained, we had to rescale the parameter g to obtain the final function $b(t)$. To do this we assumed that $\bar{b}$ is equal to L^{-1}. For *E. helvum* we assume longevity in the wild is 6 years [6]. We used this value and the parameter values in table (S1) to calculate g with equation (S7). The results of the estimation that were used fit the proportion lactation data to equation (S6) are shown in figure (S1) (note: figure (S1) shows the birth pulse in units of proportion lactating where $g = 0.74$ to show the fit to the data).

Table S1: Prior distributions and parameter estimates for Bayesian inference of the birth function $b(t)$ in equation S6. Parameter estimates are the modes of the marginal posterior distributions. All units are in weeks.

Parameter	Prior Distribution	Estimate	Rescaled
f	fixed	1/52	NA
g	$\mathcal{U}[0,1]$	0.74	0.018
s	$\mathcal{U}[0,\infty]$	10.55	NA
ψ	$\mathcal{U}[0,\pi], [\frac{-\pi}{2}, \frac{\pi}{2}]$	-0.77	NA

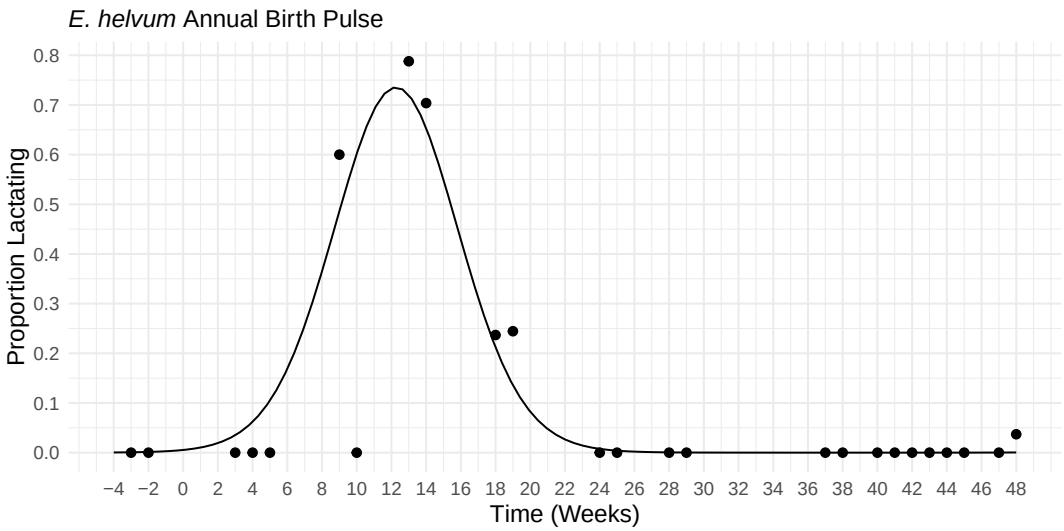

Figure S1: Estimated annual birth pulse in *E. helvum* from proportion lactating animals. Black dots represent the sampling data and the curve represents to estimated seasonal birth function.

APPENDIX 3: DETAILS ON THE SIMULATED SURVEILLANCE DATA

We generated 300 synthetic reservoir populations with seasonal forcing of a circulating pathogen over a period of 4 years. Parameter values for population simulations and $\mathcal{R}_0$ are given in table (S2). We ran 100 simulations per parameter combination. An example of the simulations generated from each parameter combination is shown in figure (S2).

Table S2: Parameter values used to generate the 300 simulated outbreaks in a synthetic wildlife population. Values that vary were used to generate datasets simulated from three different parameter combinations. All rates are given in days.

Parameter	Definition/Reference	Value in Simulations
N_0	Initial pop. size	10,000
m	Migration rate	2/365
s	Equation S6	4.00,4.00,8.00
f	Equation S6	1/365
ψ	Equation S6	2.00
g	Equation S7	0.0044, 0.0089, 0.013
μ	Table 1	1/1095, 1/730, 1/730
L	Equation S4	730, 365, 365
k	Table 1, Equation S4	$(L^{-1} - \mu)N_0^{-1}$
β	Table 1	6×10^{-5}
γ	Table 1	1/10
ω_A	Table 1	1/90
ω_T	Table 1	1/730, 1/365, 1/365
$\mathcal{R}_0$	Equation S2	5.65, 5.50, 5.16

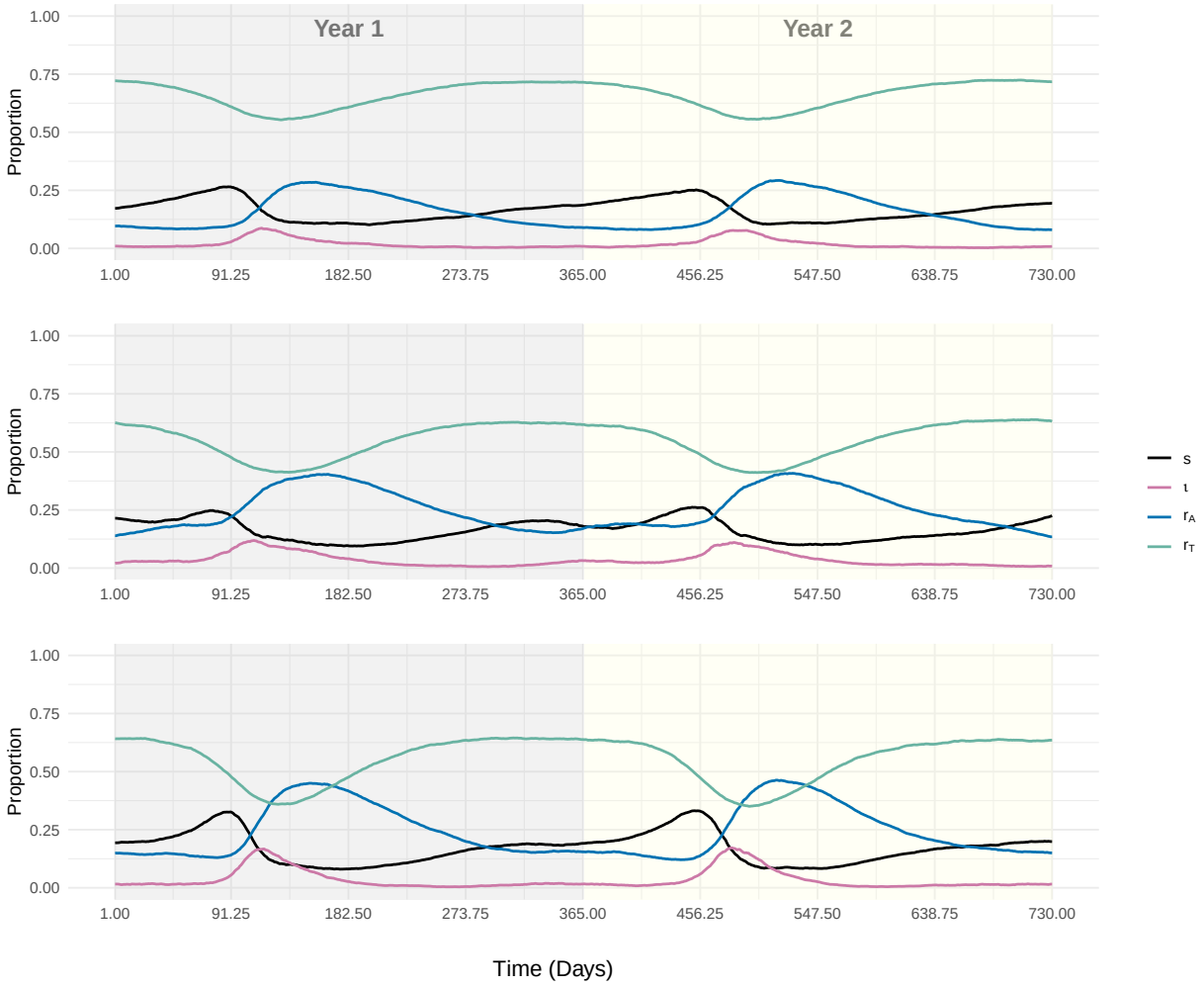

Figure S2: Examples of epidemiological dynamics in three simulated populations using the parameter values in Table (S2) over a two year time span.

REFERENCES

[1] Odo Diekmann, JAP Heesterbeek, and Michael G Roberts. The construction of next-generation matrices for compartmental epidemic models. *Journal of the royal society interface*, 7(47):873–885, 2010.

[2] A. J. Peel, J. R. C. Pulliam, A. D. Luis, R. K. Plowright, T. J. O'Shea, D. T. S. Hayman, J. L. N. Wood, C. T. Webb, and O. Restif. The effect of seasonal birth pulses on pathogen persistence in wild mammal populations. *Proceedings of the Royal Society B: Biological Sciences*, 281(1786):20132962, July 2014.

[3] Stan Development Team. *RStan: the R interface to Stan*, 2023. R package version 2.32.3.

[4] Carmen D Soria, Michela Pacifici, Moreno Di Marco, Sarah M Stephen, and Carlo Rondinini. COMBINE: a coalesced mammal database of intrinsic and extrinsic traits. *Ecology*, 102(6):e03344, 2021.

[5] David TS Hayman, Meng Yu, Gary Crameri, Lin-Fa Wang, Richard Suu-Ire, James LN Wood, and Andrew A Cunningham. Ebola virus antibodies in fruit bats, ghana, west africa. *Emerging infectious diseases*, 18(7):1207, 2012.

[6] David TS Hayman and Alison J Peel. Can survival analyses detect hunting pressure in a highly connected species? lessons from straw-coloured fruit bats. *Biological conservation*, 200:131–139, 2016.