Design and Analysis of Ranavirus Studies: Insights into Planning Surveillance, Modeling Host-Pathogen Dynamics, and Performing Risk Analyses

Matthew J. Gray, Jesse L. Brunner, Julia E. Earl, Wytamma Wirth, Angela Peace, and Ellen Ariel

1 Surveillance

Designing surveillance studies and analyzing data, whether epidemiological or any other sort, requires care and thought. While we might wish for a simple decision tree telling us what tests to use, these paths are often much murkier than we hope. Sometimes, perhaps often, methods that may "work" with our data produce answers that are not what we seek or expect. We should view experimental design and statistics as tools to help us understand rather than oracles that tell us what is or is not true or real. Thus, rather than introducing a blizzard of statistical tests and procedures,

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-64973-8_9.

M. J. Gray (✉)
Center for Wildlife Health, School of Natural Resources, University of Tennessee, Knoxville, TN, USA
e-mail: mgray11@utk.edu

J. L. Brunner
School of Biological Sciences, Washington State University, Pullman, WA, USA

J. E. Earl
School of Biological Sciences, Louisiana Tech University, Ruston, LA, USA

W. Wirth
Peter Doherty Institute for Infection and Immunity, University of Melbourne, Parkville, VIC, Australia

A. Peace
Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX, USA

E. Ariel
College of Public Health, Medical and Veterinary Sciences, James Cook University, Townsville, QLD, Australia

M. J. Gray, V. G. Chinchar (eds.), *Ranaviruses*,
https://doi.org/10.1007/978-3-031-64973-8_9

along with a litany of cases where one should or should not use them, we will present a few basic principles and tools and focus on interpreting the results. For this reason, we present a few equations and restrict ourselves to *R* functions that more closely correspond to those equations. There are many statistical packages that will do a lot of the hard work for you. In particular, we recommend the very complete epiR package, which has functions for almost every analysis. However, before proceeding with any analysis, you should be clear on what it is you are trying to learn from it and, we hope, have a little bit of intuition about what you expect to find and why.

1.1 What Are We Screening for During Surveillance?

Choose a handful of scientific studies that test samples for *Ranavirus*, and you are likely to find that they are actually screening for very different things. For some purposes, simply finding *Ranavirus* DNA is sufficient. Scientists commonly use DNA-based methods—often conventional or real-time qPCR, but increasingly isothermal methods like LAMP are being developed and used—to detect *Ranavirus* DNA in host tissues or swabs, but also from the environment itself (environmental DNA or eDNA). It is important to note that the mere presence of pathogen DNA does not by itself indicate that viable virus is present, nor that hosts are infected with the virus. To demonstrate infection, scientists rely on virus isolation to establish that viable virus is present, and histopathology and microscopy, sometimes with immunohistochemical or related methods, to establish that the virus is present and replicating in host tissues. To demonstrate disease, or ranavirosis, requires physical examination often paired with histopathology to demonstrate pathogenesis associated with infection (see Miller et al. 2024).

The statistics we present here can apply to any of these binary data—presence or absence of the pathogen's genome, the pathogen itself, viable pathogen, infections, or disease—with one exception: they do not apply when we are trying to detect events, such as epidemics or die-offs. In these cases, the samples (i.e., time points) are not interchangeable, and our probability of detecting an event changes relative to the timing of the unobserved or partially observed event. There are not yet any clear, universal guidance in these cases without making strong assumptions about how the events unfold, when they occur, etc.

1.2 Detecting Ranavirus or Establishing Freedom from Ranavirus

Frequently, the goal of surveillance is to establish whether ranavirus or ranavirosis is present or absent. Or, relatedly, we may wish to establish with a certain degree of confidence that a population is *free of* ranavirus. Of course, if the pathogen or

disease is found, we have a clear answer (assuming we trust our diagnostic test)—it is here!—but what can we make of their apparent absence? Your intuition is likely that the more samples we collect that are negative, the more confident we can be that the virus or disease is absent. Stated another way, we may not be able to say with complete certainty that the pathogen we are looking for is not present, but we can place upper limits on how common it could be given how many negative samples we have collected.

Let us formalize this intuition in two settings: (1) large populations where we can assume that our removing individuals to sample does not affect the prevalence in the rest of the population (e.g., most settings in the wild) and (2) small populations with known size where this assumption does not hold (e.g., most captive populations).

1.2.1 Large Populations

- *Approach*

We can describe the upper 95th percentile of the prevalence, p, in a large population from which we have tested n samples, all of which are negative, as:

$$p = 1 - 0.05^{1/n} \tag{1}$$

For instance, if we had a sample size of $n = 30$ and none tested positive, we could be 95% confident that the prevalence was at or below $p = 1 - 0.05^{(1/30)} = 0.095$ or 9.5%. Similarly, in *R*, you can use the *dbinom()* function to find the probability of observing $x = 0$ positives out of a sample of size = 30 with different prevalences or values of probability. For example, *dbinom*($x = 0$, size = 30, prob = 0.095) = 0.05.

There is a shortcut, an approximation, that works surprisingly well when n gets larger than about 30:

$$p \approx \frac{3}{n} \tag{2}$$

See that 3/30 = 0.1, which is a pretty good estimate of the value we just calculated.

We can also re-arrange the formula, above, to arrive at the sample size required to have a 95% chance of detecting the pathogen or disease, assuming the prevalence is p:

$$n = \frac{\log(0.05)}{\log(1-p)}, \tag{3}$$

which can be helpful in designing surveillance programs. So, if we want to have a 95% chance of detecting $p \leq 0.05$, we could calculate log(0.05)/ log (0.95)= 58.4, which would round up to 59. Or, again, we could use the shortcut approximation of $3/p = 3/0.05 = 60$, which is quite close.

- *Explanation*

Equations 1, 2, and 3 can be used on their own—simply plug in the right numbers—but understanding their derivation and assumptions can help us not just understand when they are most (or least) appropriate but provide us with some clearer intuition about what they mean. Let us begin by assuming that we are collecting samples from a population large enough that our sampling does not affect prevalence in it. That is, if we removed 30 individuals, none of which were positive, it does not appreciably enrich the fraction of the rest of the population that is positive. If this assumption is reasonable, then we can use the binomial distribution to describe the probability of observing k positives out of n samples, with the probability, p, any given sample is positive (i.e., assuming random sampling p = prevalence) as:

$$\Pr\left(k|\,p,n\right)=\binom{n}{k}p^{k}\left(1-p\right)^{(n-k)}, \tag{4}$$

where $\binom{n}{k}$ is the binomial coefficient:

$$\binom{n}{k}=\frac{n!}{k!\left(n-k\right)!}.$$

If $k = 0$, then this binomial term simplifies to 1 and the whole equation simplifies to:

$$\Pr\left(\,0|\,p,n\right)=\left(1-p\right)^{n} \tag{5}$$

As we can see, as n increases the probability of observing $k = 0$ decreases for any given value of $p > 0$. And, equivalently, as n increases, the observation of $k = 0$ is consistent only with smaller and smaller values of p.

If we set the probability of observing $k = 0$ to some predetermined value, often $\Pr(0|p,n) = 0.05$, or 5%, we can find the value of p that is consistent with our data (i.e., $k = 0$ of n samples positive).

$$\begin{aligned}\Pr\left(0|\,p,n\right)=0.05&=\left(1-p\right)^{n}\\0.05^{1/n}&=1-p\\p&=1-0.05^{1/n}\end{aligned}$$

This gives us the formula for prevalence, p, that would give us a 5% chance of getting zero positives out of a sample size of n (Eq. 1). Or, equivalently, this means we can be 95% confident that the prevalence in the population is equal to or less than p.

We can also rearrange this equation to determine the sample size required to detect the pathogen or disease with, for instance, 95% confidence, given an expected level of prevalence, p (Eq. 3).

$$0.05 = (1-p)^n$$
$$\ln(0.05) = n \times \ln(1-p)$$
$$n = \frac{\ln(0.05)}{\ln(1-p)}$$

For instance, if we wanted to have 95% confidence in detecting a pathogen that was at most 5% prevalent ($p = 0.05$), we would calculate:

$$n = \frac{\ln(0.05)}{\ln(1-0.05)} = \frac{\ln(0.05)}{\ln(0.95)} = \frac{-2.9957}{-0.0513} = 58.404$$

implying that we need a sample size of 59 or larger.

This derivation, taking the log of both sides of our equation, relates to a rule of thumb that provides a reasonable approximation of the upper bound of p, assuming we desired 95% confidence and $n \gtrapprox 30$. We start with the logarithm of both sides, as above:

$$\ln(0.05) = n \times \ln(1-p)$$

but then we insert two approximations. First since $\ln(0.05) = -2.9957 \approx -3$, we substitute -3 into the left-hand size of the equation. Second, because $\ln(1-p) \approx -p$ when p is small, we substitute $-p$ into the right-hand size of the equation. This yields:

$$-3 \approx -np$$
$$p \approx \frac{3}{n}$$

or the "rule of three" (Eq. 2). This approximation suggests that with $n = 30$ and no positives, we can be 95% confident that the prevalence is less than or equal to $p \approx 3/30 = 0.1$, which is quite close to $p = 1 - 0.05^{(1/30)} = 0.09503$. And, with $0.05 \approx 3/n \rightarrow n \approx 3/0.05 = 60$, we approximate the sample size needed to be 95% certain of detecting infection or disease if the prevalence were 5%.

1.2.2 Small Populations

- *Approach*

In small populations, it is more appropriate to consider the *number* of positive individuals rather than prevalence because prevalence is not continuous (i.e., 1 of 10 → 10% prevalence, 2 of 10 → 20% prevalence, etc.). In addition, prevalence is affected by our sampling in small populations; removing individuals to test changes the prevalence in the remaining population. We thus need a different distribution, the hypergeometric, where we consider the probability that there are K positive individuals in a population of size N, rather than the prevalence.

We can describe the probability of getting zero positives from the n samples we collected as:

$$\Pr(0) = \frac{\binom{N-K}{n}}{\binom{N}{n}}, \tag{6}$$

where the parentheses represent the binomial expansion (this is a special case of the hypergeometric distribution; see below). There is unfortunately no simple expression to find the upper 95th percentile on the number of positive individuals in the population, K, but we can let the computer do the hard work for us. For instance, with a population of $N = 15$, where we sample $n = 6$ individuals, we can find the probability of observing zero positives if $K = 1, 2, 3, \ldots, 10$ individuals in the population were positive using the *dhyper()* function in *R* and finding the value of K for which this probability is ≤ 0.05. In this example, the code is:

```
N <- 15
n <- 6
K <- 1:10

cbind(K, Pr = dhyper(x=0, m=K, n=N-K, k=n))
```

```
       K           Pr
 [1,]  1 0.6000000000
 [2,]  2 0.3428571429
 [3,]  3 0.1846153846
 [4,]  4 0.0923076923
 [5,]  5 0.0419580420
 [6,]  6 0.0167832168
 [7,]  7 0.0055944056
 [8,]  8 0.0013986014
 [9,]  9 0.0001998002
[10,] 10 0.0000000000
```

Since the fifth element, 0.0419580420 is just below our cutoff of 5%, we can have $\geq 95\%$ confidence that $K = 5$ or less.

We can similarly find the sample size, n, required to have a certain confidence of detecting at least one positive (i.e., not getting all negatives) if we set K to some value. For instance, with $K = 3$ positives in a population of size $N = 15$, we can use the following code:

```
N <- 15
K <-3
n <- 1:10

cbind(n, Pr = dhyper(x=0, m=K, n=N-K, k=n) )
```

```
       n          Pr
 [1,]  1 0.80000000
 [2,]  2 0.62857143
 [3,]  3 0.48351648
 [4,]  4 0.36263736
 [5,]  5 0.26373626
 [6,]  6 0.18461538
 [7,]  7 0.12307692
 [8,]  8 0.07692308
 [9,]  9 0.04395604
[10,] 10 0.02197802
```

and choose the row where Pr is ≤5% (i.e., when $n = 9$).

- *Explanation*

The large sample, binomial distribution approach assumes that every sample has the same probability of being positive as every other sample, which is reasonable if we were sampling from a very large population or if, for some reason, we were sampling with replacement, such as collecting non-lethal samples for ranavirus testing. Researchers are increasingly working with finite populations of known size, such as in captive collections or in trade, and lethal samples for ranavirus testing (from the liver or kidneys) are common. In smaller populations, sampling *with* removal, the probabilities of being positive are no longer interchangeable. Imagine sampling a population of $N = 10$ with $K = 1$ positive individual. The first sample would have a one in ten chance of being positive, but after we removed five negative individuals, the next sample would have a one in five chance of being positive; the probability has doubled. Or instead, if we had already sampled the one positive individual, every subsequent sample would have zero probability of being positive. In such finite situations where populations are sampled without replacement, the binomial assumption is no longer a reasonable simplification, and we instead use a hypergeometric distribution.

In a population of size N with K positive individuals, the probability of obtaining k positives from a sample of size n is:

$$\Pr\left(k|K,N,n\right)=\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \tag{7}$$

where we employ the binomial coefficient three times. When $k = 0$, this simplifies somewhat to:

$$\Pr(0|K,N,n) = \frac{\binom{N-K}{n}}{\binom{N}{n}},$$

which we saw in Eq. 6. While there is no simple formula for K given a sample size of n out of a population of size N and an assumed $\Pr(0|K,N,n)$, the logic is the same as above, and *R* functions can do the calculations for us. For instance, we can calculate the probability of observing $k = 0$ positive samples out of $n = 10$ samples tested from a population of size $N = 20$, as:

```
K <- 1:10
N <- 20
n <- 10
# Note that the R function is parameterized slightly differently
data.frame(K=K, Pr = round( dhyper(x=0, m=K, n=N-K, k=n), 3) )
```

```
    K    Pr
1   1 0.500
2   2 0.237
3   3 0.105
4   4 0.043
5   5 0.016
6   6 0.005
7   7 0.002
8   8 0.000
9   9 0.000
10 10 0.000
```

We could be approximately 95% confident that $K \leq 4$ out of the 20 individuals in the population.

1.3 Estimating Precision on Infection Prevalence Estimates

The inferences in the preceding sections about how common the pathogen or disease might be in a population given a certain number of negative samples are just a special case of understanding the precision with which we estimated prevalence. A

larger sample size increases the precision of our estimate and should thus yield narrower bounds on the values of prevalence that are consistent with our data.

There is a myriad of methods for approximating the confidence interval of a proportion, especially in the case of large (or infinite) populations using a binomial distribution. They range from fairly imprecise (i.e., the normal approximation that is common in introductory statistics textbooks, which often produces cutoffs <0 or >1) to quite good (e.g., Wilson's (1927) method). Rather than enter into this sometimes bewildering fray, we instead suggest a simple, coherent approach that produces intuitive results. In any case, most methods produce similar intervals, and we would caution against placing too much weight on any particular value or cutoff.

1.3.1 Large Populations

- *Approach*

While there is no simple closed-form formula for it, we can use an *R* function for the beta distribution (see below for more details), particularly the quantile function of the beta, to find an interval on the prevalence within a large population. For instance, if we collected 30 samples, 1 of which was positive and 29 that were negative, we could find the 95% confidence interval as:

```
round(
  qbeta(p=c(0.025, 0.975), # the quantiles of the 95th percentile
        shape1=1+1, # the number of positives + 1
        shape2=1+29), # the number of negatives + 1
  3)
```

```
[1] 0.008 0.167
```

Few researchers are familiar with the beta distribution, so it is worth some introduction. It is a continuous distribution bounded between 0 and 1 that is useful for describing the distribution of prevalence. It has two parameters, often called α and β, but sometimes referred to as shape1 and shape2. If we had no information about what prevalence might be, meaning that every possible value is equally likely, we could describe this as a beta distribution with $\alpha = \beta = 1$ (see the flat line gray line in the Fig. 1).

After collecting 30 samples, 6 of which are positive, our estimate of prevalence would be 6/30 = 0.2. To find the consistency of different values of prevalence with these data (see the curve in Fig. 1), we can simply add the positives to α and the negatives to β in the beta distribution.

Notice that our simple estimate of prevalence, 6/30 = 0.2, is most consistent with our data, but then the probability density tapers off rapidly with much lower or higher values of prevalence.

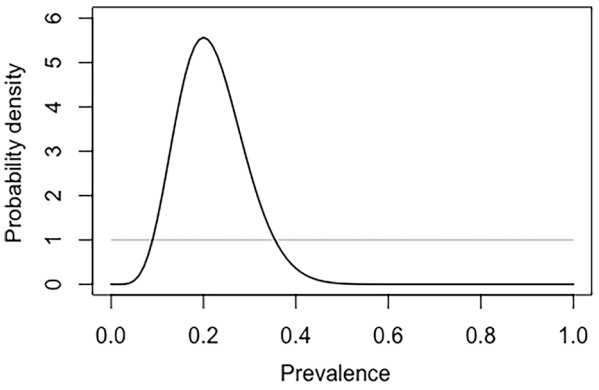

Fig. 1 Beta distributions describing the probability (really probability *density*, *y*-axis) that prevalence takes on any particular value (*x*-axis). The flat gray line is if we had no information on prevalence and thought any value was equally probable. This corresponds to the parameters $\alpha = \beta = 1$. The black curve represents the situation if we had collected 30 samples and found that 6 were positive and 24 were negative, corresponding to $\alpha = 1 + 6 = 7$ and $\beta = 1 + 24 = 25$

We can find a specific interval, such as a 95% interval, around our estimate of prevalence using the cumulative distribution function for the beta, which is qbeta() in *R*:

```
#95% CI on prevalence given 6 positive and 24 negative tests
qbeta(p=c(0.025, 0.975), shape1=1+6, shape2=1+24)
```

```
[1] 0.09594217 0.37473217
```

Note that this interval stems from a Bayesian perspective on our estimate of prevalence. It turns out that the beta distribution is the conjugate prior for a binomial likelihood, which means that the posterior is the same form as the prior, just with the parameters modified by the data. If that does not mean anything to you, do not fret; you can use the method without understanding priors! Moreover, do not get too worried about Bayesian versus frequentist approaches; all the intervals you calculate will be quite similar, even if they are conceptually different. However, one potential feature of this Bayesian approach is that if you had prior information about the prevalence (e.g., from an earlier survey), or were willing to assume that some values were much less likely than others (e.g., you feel confident that prevalence is less than 0.5), you could include that information in the prior values of α and β. This would produce more informative estimates of the prevalence and the interval that are consistent with your data and prior understanding of the system; however, using prior information is not necessary.

1.3.2 Small Populations

The approach to estimating the distribution of prevalence values consistent with our data in small, closed populations (e.g., in captive settings or trade) is similar to that for large populations; only we use a hypergeometric distribution. This will require a slightly more overtly Bayesian approach, but the code is still simple. We will take this in steps.

First, let us again imagine that we collected a sample of 30 individuals from a population of size 50. Prior to examining our results, we think that each possible value of *K*, the actual number of infected individuals in the population, is equally likely.

```
n <- 30 # sample size
N <- 50 # population size
Ks <- 0:N # possible numbers of positives in population

prior <- 1/length(Ks) # N+1 equally probable values of K
```

A plot of the prior against the possible values of *K* would yield a flat line or, actually, because *K* can only take on discrete values, a bar plot with bars of equal heights (Fig. 2).

In a Bayesian analysis, the *posterior* probability of any given value of *K* given our prior expectation and data (i.e., $x = 6$ positives out of n samples from a population of size N) is the *likelihood* of observing the data, given a particular value of *K*, times the prior probability that *K* took on this particular value:

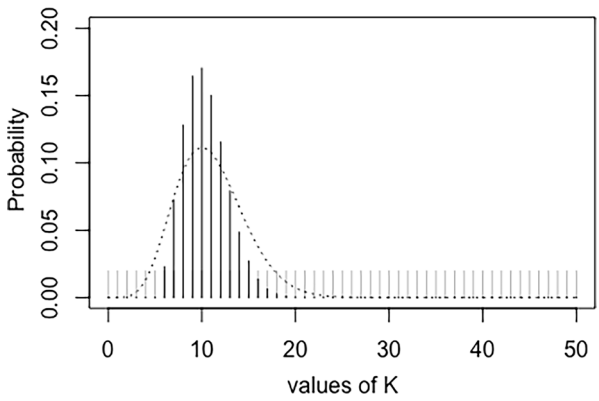

Fig. 2 The probability of different numbers of infected or diseased individuals, *K*, in a population of size $N = 50$. Note that because we are considering individuals rather than proportions of the population, the distribution is discrete. The light gray bars represent the case where we have no prior information or expectation about the commonness of infection in the population. The darker bars represent the case where we have collected 30 samples from this population without replacement and found that 6 were positive. The dotted line represents the equivalent beta distribution for the same data if we had instead been sampling from a very large population

$$\Pr(K|x,n,N) = \frac{\Pr(x|K,n,N) \times \Pr(K)}{\text{normalizing term}}$$

The likelihood is the term on the left of the numerator, and the prior is the term on the right. Note that we need to divide by some normalizing term so that this is a proper probability distribution. In discrete cases this normalizing term is simply the summation of the numerator for $K = 0, 1, 2, \ldots, N$.

$$\Pr(K|x,n,N) = \frac{\Pr(x|K,n,N) \times \Pr(K)}{\sum_{K=0}^{N} \Pr(x|K,n,N) \times \Pr(K)}$$

The likelihood for each possible value of K, in terms of code, is just the probability of observing six positives with a sample size of n from a hypergeometric distribution.

```
likelihood <- dhyper(x=6, # The number of sample positive
                     m=Ks, # possible numbers of positives
                     n=N-Ks, # number of negatives, given Ks
                     k=n) # sample size

cbind(Ks, likelihood)[1:21,] # likelihoods for first 21
values of K
```

```
      Ks   likelihood
 [1,]  0 0.0000000000
 [2,]  1 0.0000000000
 [3,]  2 0.0000000000
 [4,]  3 0.0000000000
 [5,]  4 0.0000000000
 [6,]  5 0.0000000000
 [7,]  6 0.0373661953
 [8,]  7 0.1188924397
 [9,]  8 0.2101354747
[10,]  9 0.2701741818
[11,] 10 0.2800586031
[12,] 11 0.2464515707
[13,] 12 0.1895781313
[14,] 13 0.1297113530
[15,] 14 0.0797549535
[16,] 15 0.0443083075
[17,] 16 0.0222807489
[18,] 17 0.0101276131
[19,] 18 0.0041431145
[20,] 19 0.0015138303
[21,] 20 0.0004883324
```

Notice that the likelihood of observing six positives is zero if K is less six, which we hope makes some intuitive sense. Likewise, there is a low or zero probability of observing six positives if K is very large.

Next, we can calculate the posterior, the results of which are plotted in Fig. 2.

```
posterior <- likelihood*prior/sum(likelihood*prior)

cbind(Ks, posterior)[1:21,] # posterior prob for 1st 21
values of K
```

```
       Ks    posterior
 [1,]  0 0.0000000000
 [2,]  1 0.0000000000
 [3,]  2 0.0000000000
 [4,]  3 0.0000000000
 [5,]  4 0.0000000000
 [6,]  5 0.0000000000
 [7,]  6 0.0227127854
 [8,]  7 0.0722679535
 [9,]  8 0.1277294062
[10,]  9 0.1642235223
[11,] 10 0.1702316999
[12,] 11 0.1498038959
[13,] 12 0.1152337661
[14,] 13 0.0788441558
[15,] 14 0.0484785012
[16,] 15 0.0269325007
[17,] 16 0.0135432003
[18,] 17 0.0061560001
[19,] 18 0.0025183637
[20,] 19 0.0009201714
[21,] 20 0.0002968295
```

Notice that the most probable value of K is 10, which corresponds to a prevalence of 10/50 = 20%, which is the same prevalence as in the sample (6/30 = 20%).

1.4 Assumptions About Detecting Ranavirus and Ranavirosis

All our estimates of prevalence or confidence that ranavirus is absent are built on the critical assumptions that we sample randomly—in essence that positive and negative samples are equally likely to be sampled—and that our diagnostic test is essentially perfect. Let us delve into these separately.

1.4.1 Sampling Biases

All the statistical methods presented assume that we are sampling randomly (or at least haphazardly), without biases in who is sampled. If we were more likely to sample infected individuals than uninfected individuals, we would, of course, have a bias in our sample. If we were to ignore this bias, then our inferences about infection could be quite off. For example, imagine we were sampling during a die-off and captured primarily or solely dead or moribund individuals. Assuming that ranavirus were the cause of this event, we would expect that essentially all our samples would test positive for ranavirus. The problem is extrapolating this biased die-off sample to the population, and inferring that prevalence was near 100%. Note that such a bias could be in the opposite direction, where uninfected individuals were more likely to be sampled, perhaps because infected individuals are more cryptic or dead individuals decompose or are scavenged rapidly.

It is sometimes possible to correct for such biases if we have some independent estimates of them (e.g., from prior data, capture-mark-reencounter studies; see Cooch et al. 2012 for a discussion), but more often we simply must acknowledge their presence and interpret our estimates with caution. More often the best solution is to identify and correct for such biases *before* collecting samples.

1.4.2 Detection Biases

Our diagnostic tests are rarely perfect. They may have false negatives, meaning less than perfect sensitivity (*Se*), and false positives, meaning less than perfect specificity (*Sp*). Depending on the goals of your research, false negatives or false positives may be more concerning. While there are no universal solutions or optima, it can be helpful to consider the source of these spurious results. False-positive qPCR results might come from sample contamination in the field or laboratory with *Ranavirus* DNA. For example, it is possible to contaminate samples if gloves are not changed between infected and uninfected individuals (Gray et al. 2018). Contamination of an assay also could occur with qPCR standards without proper well labeling or pipetting technique. Field blanks, extraction controls, and no-template PCR controls are meant to help a researcher detect when contamination occurs. In these cases, the false positives come from outside of the diagnostic test itself; the primer and probe specificity for carefully designed qPCR assays usually precludes amplification of non-target DNA. Immunological methods as well as clinical diagnosis of disease also suffer from false positives. Here, the background signal (e.g., from cross-reactive antibodies or nonspecific signs of infection) from uninfected individuals can look very much like, and quantitatively overlap with, positive individuals. Training and experience can reduce these errors, but never to zero.

False negatives can also stem from error—hence the need for positive controls in most assays—but also because the probability of detection is inherently quantitative; it takes a certain amount of pathogen DNA, viable pathogen, or pathology to ensure detection (see Brunner 2020 for discussion). Beyond developing good

protocols, researchers may also be able to design their studies to ensure better sensitivity. For instance, researchers often hold ranavirus-exposed animals for several days or weeks to let infections develop to detectable levels. Moreover, lethal samples (i.e., organ tissue) will likely result in greater detection of ranavirus compared to nonlethal samples (i.e., swabs, tail-clips; Gray et al. 2012).

It may be useful as well to revisit the question of what it is you are trying to detect. If you are primarily interested in ranavirosis, then a test that has a non-trivial chance of missing a low-level infection may not be much of an issue. However, if instead you are working to detect sublethal, inapparent infections, you will need very sensitive methods of detection and might consider using a two-step process: an initial test that is very sensitive, but prone to false positives, followed by an independent and more specific test to verify the initial results (Miller et al. 2024).

In principle, it is possible to establish the diagnostic sensitivity and specificity of a test or assay and then correct results with these values, but in practice there is no good reason to think that these values are constant among species, between life history stages, or even over time. We thus caution against using estimates of sensitivity and specificity, or the corrected estimates of prevalence or other quantities, as if they are well established or even constant from setting to setting or study to study. Rather, it is worth bearing in mind that poor data on the presence or absence of ranavirus or disease (or any other response) are never really improved by clever statistics.

All that said, there are several useful frameworks for adjusting estimates or probabilities of observing k positives with outside estimates of Se and Sp. We point the reader to Cameron and Baldock (1998) for the formulae (or the Electronic Supplement 1, which includes *R* code) and to the epiR package in *R* for the calculations.

1.5 Interpreting Prevalence

While prevalence is more commonly estimated than incidence during surveillance studies, it is simply a "snap shot" of the proportion of a population that are infected or diseased at a given time. It is difficult to interpret these snap shots in the absence of biological context. Indeed, interpreting *any* statistical result necessarily invokes some implicit or explicit model of how the system is thought to work. In the case of prevalence, we can think of it as the ratio of the positive individuals relative to the population as a whole. The number (or density or fraction) of positive individuals is a product of both the gain and loss of those positives. The population presumably gains positives through the process of transmission and infection, which is governed by the amount of direct or indirect exposure to the ranavirus as well as by the probability that individuals are infected given an exposure (one meaning for the term "susceptibility"). Various factors influence the probability of successful ranavirus transmission, and those factors are rarely static (Brunner et al. 2024). Additionally, a population loses positive individuals as they recover or die. We expect prevalence to change throughout a year and during the course of an epidemic, thus the timing of when we collect samples can be very relevant to interpreting prevalence estimates.

Prior data and context can be very helpful in interpreting prevalence. For example, if experimental exposures show that a species dies rapidly following ranavirus exposure, then high prevalence would be most consistent with sampling during (or soon after) a period of high transmission or during a ranavirus-induced die-off. Several studies have reported species-level susceptibility under controlled conditions (e.g., Hoverman et al. 2011; Brenes et al. 2014b; Brunner et al. 2024). Biological context also can be gleaned from the density of the population and the timing of the survey relative to the phenology of the organism. For example, observing low prevalence and a dense population of amphibian larvae early in spring would be more consistent with the virus recently being introduced rather than an outbreak already occurring.

Lastly, it is important to recognize that infected individuals may be more or less likely to be detected or captured than uninfected individuals, which can bias prevalence estimates (Cooch et al. 2012). For instance, moribund fish and tadpoles are often found near the surface, making them more easily detected, which would inflate prevalence estimates, whereas sick turtles may move less and have lower detection probabilities resulting in underestimates of prevalence. Variation in detection probabilities through time (e.g., developmental stages) and among locations also can lead to apparent differences in prevalence that do not reflect actual differences in the proportion infected. See Mosher et al. (2019) for an example of how accounting for detection probabilities can substantially improve estimates of prevalence.

1.6 Comparing Prevalence Estimates and Other Logistic Regressions

There are many ways of comparing counts or contingencies such as prevalence data. Many classic approaches, such as χ^2-tests, make strong assumptions about the numbers in each cell or whether sums across rows are fixed or not. Logistic regression is an equivalent, but more flexible approach, when we know the number of "successes" (i.e., number positive) out of the number of "trials" (i.e., number tested). These analyses model the probability of success as a linear function of the variables of interest, whether they are discrete (e.g., two samples in different times or places) or continuous (e.g., $\log_{10}$ (dose) of ranavirus exposure). The one potentially tricky bit is that they use a logit "link" function to map the linear model into probability space (see next section).

1.6.1 Approach

The basic form of the logistic regression is:

$$\text{logit}\left(p\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots, \tag{8}$$

where p is the probability of being infected and the β s are the effects of the predictors of interest, designated x_1, x_2, etc. For any linear model that you might construct in a typical regression, you can also use in a logistic regression.

Let us use two examples to illustrate how this can work. First, imagine our data from before, where we found 6 of 30 samples were positive from wetland A, and then imagine we collected 25 samples from a neighboring wetland B where 12 samples were positive. Our question is whether there is strong evidence that these two sites have different prevalence; after all, sometimes, by chance, you might get more or fewer positives from a sample.

```
( df <- data.frame(wetland = c("A","B"),
                   Pos = c(6, 12),
                   Neg = c(24, 13)) )
```

```
  wetland Pos  Neg
1       A   6   24
2       B  12   13
```

```
m1 <- glm(cbind(Pos, Neg) ~ wetland,
          data=df,
          family = "binomial")
```

```
summary(m1)
```

```
Call:
glm(formula = cbind(Pos, Neg) ~ wetland, family = "binomial",
    data = df)

Deviance Residuals:
[1]  0  0

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.3863     0.4564  -3.037  0.00239 **
wetlandB      1.3063     0.6071   2.152  0.03143 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 4.9038e+00  on 1  degrees of freedom
Residual deviance: 1.0658e-14  on 0  degrees of freedom
AIC: 11.125

Number of Fisher Scoring iterations: 3
```

Running the model is relatively straightforward. We can give the generalized linear model function, glm(), two columns (the cbind() function binds columns together), the first with the number of positive samples and the second the number of negatives corresponding with the wetlands (rows).

We can interpret the coefficients more easily if we exponentiate them, yielding estimates of the *odds* of being positive. In this case the odds of being positive in wetland A are exp(−1.3863) = 0.25, which is essentially 6:24 or the ratio of the number of positives to the number of negatives. The expected odds of being positive in wetland B is exp(−1.3863 + 1.3063) = 0.923, which is 12:13.

We can also interpret the exponentiated coefficient for wetland B as being the increased odds of being positive if the sample comes from wetland B *relative* to wetland A. Thus exp(1.3063) = 3.692 is roughly (12/13)/(6/24). That is, in this example wetland B has 3.69-fold higher odds of a sample being positive compared to wetland A. If we had included a continuous variable in our model, for wetland salinity, the exponentiated parameter would be interpreted as the change in odds for a one-unit change in the predictor.

As a second example, let us consider our laboratory dose-response experiment with a ranavirus. We individually exposed larval wood frogs (*Lithobates sylvaticus*) to one of five doses of the virus, from ~25 to 250,000 plaque-forming units (pfu) per mL, and then recorded whether each individual (n = 11–12 per dose) died (Warne et al. 2011). We expected that the probability of death would tend to increase with increasing exposure to ranavirus, but our specific questions were as follows: (a) how rapidly (i.e., what is the slope of this relationship?) and (b) at what dose would we predict a 50% chance of death (the LD_{50})? These are the data and plotted in Fig. 3.

```
( df2 <- data.frame( Dose = c(1.40, 2.40, 3.40, 4.40, 5.40),
                     Died = c(   2,    3,   11,   11, 12),
                     N =    c(  12,   11,   12,   12, 12)) )
```

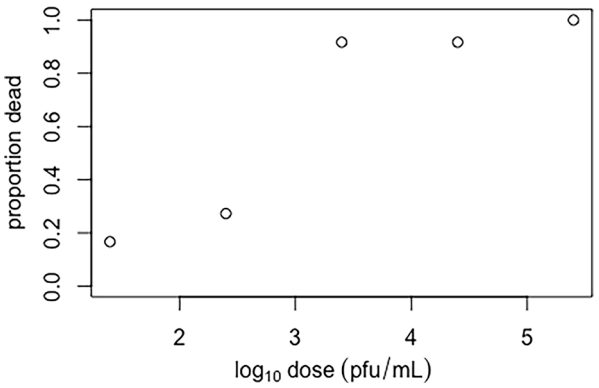

Fig. 3 The proportion of wood frog larvae that died when water bath exposed to one of five doses of a ranavirus. (Data from Warne et al. 2011)

```
  Dose Died  N
1  1.4    2 12
2  2.4    3 11
3  3.4   11 12
4  4.4   11 12
5  5.4   12 12
```

```
plot(Died/N ~ Dose, data=df2,
     xlab=expression( log[10]~dose~(pfu/mL) ),
     ylab="proportion dead",
     ylim=c(0,1))
```

Constructing a logistic regression is virtually the same as above. Note how we provide the function two columns, representing number of successes (=Died) and failures (=survived = N-Died).

```
m2 <- glm(cbind(Died, N-Died) ~ Dose,
          data=df2,
          family = "binomial")
summary(m2)
```

```
Call:
glm(formula = cbind(Died, N - Died) ~ Dose, family =
"binomial",
    data = df2)

Deviance Residuals:
      1        2        3        4        5
 0.5533  -1.0053   1.0944  -0.6384   0.4291

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.4848     1.2446  -3.604 0.000314 ***
Dose          1.7317     0.4363   3.969  7.2e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 38.0898  on 4  degrees of freedom
Residual deviance:  3.1061  on 3  degrees of freedom
AIC: 16.048

Number of Fisher Scoring iterations: 5
```

In this case the exponentiated coefficient, exp(1.7317) = 5.7, is interpreted as the increased odds of death with every one-unit increase in the predictor, which is an order of magnitude in dose. This also implies that with an increase in exposure dose of *two* orders of magnitude, the odds of death increase exp(1.7317 × 2) = 31.9-fold.

This seems like an enormous increase in the odds of death, and it is, but note that this is *relative* to where we started before increasing the dose by two orders of magnitude. If, for instance, we started at a dose of 1 ($=10^0$ = 1 pfu/mL) and then added two orders of magnitude ($=10^{0+2}$ = 100 pfu/mL), the odds of death would be:

```
exp(-4.4848 + 1.7317*(0+2))
```

```
[1] 0.3600905
```

which is still less than equal odds of dying. The expected *probability* of death at this dose (10^2 pfu/mL) is as follows:

```
exp(-4.4848 + 1.7317*(0+2))/( 1 + exp(-4.4848 + 1.7317*(0+2)) )
```

```
[1] 0.2647548
```

Equivalently, we can use the predict() function for this estimate.

```
predict(m2,
        newdata=data.frame(Dose = 2), # <- predict at a dose of 2
        type = "response") # default is "link"
```

```
        1
0.2647569
```

The predict() function is also useful to draw a graph showing the predicted line (Fig. 4).

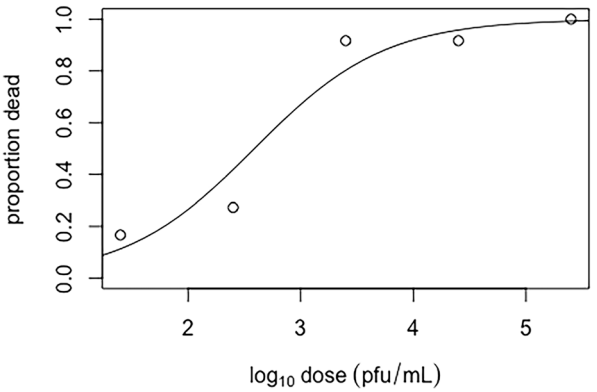

Fig. 4 The proportion of wood frog larvae that died when water bath exposed to one of five doses of a ranavirus (circles) as well as the best-fit line from a logistic regression fit to these data. (From Warne et al. 2011)

```
# our raw data
plot(Died/N ~ Dose, data=df2,
     xlab=expression( log[10]~dose~(pfu/mL) ),
     ylab="proportion dead",
     ylim=c(0,1))
# generate predictions
preds <- data.frame(Dose = seq(from=1, to=6, length.out=50))
preds$Prob <- predict(m2,
                      newdata=preds,
                      type="response")
# plot predicted line
lines(x=preds$Dose, y=preds$Prob, type = "l")
```

At what dose do we get an expected 50% chance of death? This ends up being estimated as:

$$\mathrm{LD}_{50} = -\frac{\beta_0}{\beta_1} \tag{9}$$

or in our case $-(-4.4848)/1.7317 = 2.59$. Recall that this is on the $\log_{10}$-scale, so this, the estimated LD_{50}, is $10^{2.59}$ or 3.89×10^2 pfu/mL.

Note that the *slope* of the dose-response relationship also tells us something about the underlying distribution of susceptibility (here, susceptibility to death from exposure). In general, a steeper slope implies less variability in the population (Ben-Ami et al. 2010). See Electronic Supplement 2 for a visual explanation of the relationship between the slope of a dose-response study and the underlying distribution of susceptibility in the population.

1.6.2 Explanation

Logistic regression connects binomial data (i.e., the number of "successes" out of a number of "trials") to a linear model using the logit link. This is important because a typical linear regression is unbounded, meaning it can produce values well below zero and well above one, but the binomial likelihood part of the model requires a bounded probability. The logit link transforms these unbounded values from the linear model into bounded probability estimates.

The logit transform is simply the log-odds of success. So, if p is the probability of success, then the odds of success are $p/(1-p)$, and the log-odds are $\log[p/(1-p)]$. In a logistic regression, we are saying that the log-odds are a linear function of our predictor(s):

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \ldots \tag{10}$$

From this equation, we can see that exponentiating the linear model (or a part of it) gives the predicted odds (or change in odds), which is useful for interpreting the coefficients of a logistic regression.

To convert the log-odds (or logit) of p or the results of our linear model back to probability space, we can use the inverse of the logit transform, which is also known as the logistic transform:

$$\begin{aligned} \log\left(\frac{p}{1-p}\right) &= \beta_0 + \beta_1 x_1 + \cdots \\ \frac{p}{1-p} &= \exp\left(\beta_0 + \beta_1 x_1 + \cdots\right) \\ p = \frac{\exp\left(\beta_0 + \beta_1 x_1 + \cdots\right)}{1-\exp\left(\beta_0 + \beta_1 x_1 + \cdots\right)} &= \frac{1}{1+\exp\left(-\left[\beta_0 + \beta_1 x_1 + \cdots\right]\right)} \end{aligned} \tag{11}$$

If the linear model produces some very large value, then this equates to a probability very close to one, and a linear model that produces very small values equates to probabilities very close to zero.

We can find the value of the predictor(s), such as the dose, expected to produce a 50% chance of death (or other outcomes) by inserting $p = 0.5$ (or any other value of interest such as $p = 0.9$ if we were interested in the LD_{90}) in the left-hand side of the logit equation and re-arranging we get Eq. 9:

$$\begin{aligned} \log\left(\frac{0.5}{1-0.5}\right) &= \beta_0 + \beta_1 x_1 \\ \log(1) = 0 &= \beta_0 + \beta_1 x_1 \\ x &= \frac{\beta_0}{\beta_1} \end{aligned}$$

So, with our parameters, we get $LD_{50} = -(-4.4848/1.7317) = 2.59$ pfu/mL.

1.7 Incidence

Infection prevalence is useful when describing the distribution of ranaviruses among regions and host species, but it does not convey information about risk or rates of infection. Infection incidence is the rate at which individuals become infected with a pathogen (i.e., the number of new cases that occur in a specified time period; Wobeser 2006). In small captive populations, it may be possible to determine how many individual animals become infected over short intervals. Consider, for example, a study with two surveys. The initial survey found that 2 of 50 animals were infected, and a second survey at the end of the week found that 10 individuals were

infected. Often incidence is presented as the number of new cases (here, 8) per time period *out of the whole population*, which in our example would be 8 out of 50 or 16% per week (assuming infected individuals did not die or recover in between). However, some sources (e.g., veterinary epidemiologists; Dohoo et al. 2003; Wobeser 2006) will present incidence as the number of new cases per unit time *out of those at risk*, in which case incidence would be 16.7% (=8 new cases/48 at risk). This second version is synonymous with the force of infection, or per capita risk of infection, which is commonly considered (and distinguished from "incidence") in epidemic models (Hens et al. 2012). In any case, be sure to clearly present how you derive incidence as the definition is not universal. Also, if populations are not closed (i.e., immigration and birth or emigration and death occur), calculations of incidence rates need to be adjusted for the time at risk (see Dohoo et al. 2003 for details). Estimating incidence in wild populations can be difficult because we generally cannot track the fate or infection status of individuals.

Two approaches can help us estimate incidence in wild populations. First, one might use sentinels, uninfected individuals that are introduced into the environment (e.g., tadpoles in cages placed in a wetland) and regularly screened for infection to estimate force of infection. Sentinel species should be highly susceptible to ranavirus, so as to have a high chance of exhibiting an infection if given a reasonable exposure. One must also consider the assumption that the risk of exposure is the same for the sentinels as the free-living individuals (McCallum 2008).

A second approach is capture-mark-reencounter (CMR) studies where individuals are given unique marks and released. During subsequent encounter periods (e.g., trapping, netting), the researcher records the new and recaptured individuals and determines their infection status. Given that individuals are released, infection status must be determined using nonlethal methods (St-Amour and Lesbarrères 2007; Gray et al. 2012). Ultimately, CMR models estimate the probability of individuals changing infection status while accounting for imperfect detection of the pathogen and imperfect recapture probability of the host. We are not aware of any studies with ranavirus infections that have taken either approach.

1.8 Infection Intensity (Ranavirus Load)

1.8.1 Approach

Infections occur along a gradation of intensities, which may be relevant in our understanding of the dynamics or outcome of infections or epidemics. For instance, viral titers (from plaque-assays or $TCID_{50}$ assays) or viral gene quantities (from quantitative real-time PCR) are likely to be low soon after infection or carrier species that maintain persistent sublethal infections. So how can we analyze such quantitative information?

The first step is to recognize that viral quantities, Q, are usually lognormally distributed. We can think of the amount of virus at a given time in a growing virus

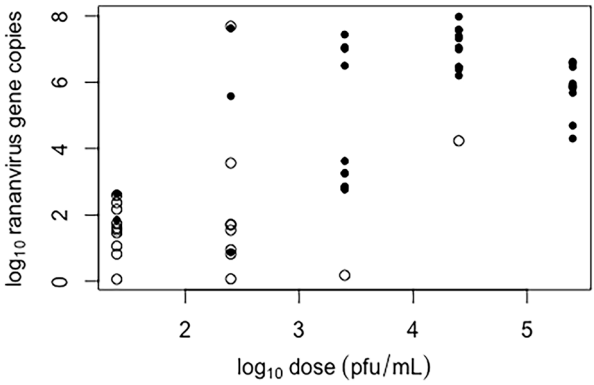

Fig. 5 The intensity of infections, described as the number of ranavirus gene copies per sample, detected with a qPCR reaction in animals that survived (open circles) or died (filled circles) in the dose-response study by Warne et al. (2011)

infection as the realization of a number of independent, *multiplicative* steps. Thus, the log of viral quantities, log(Q), can be thought of as the sum of the log of those multiplicative effects and so should be approximately normally distributed. Note that the base of the logarithm is not important, except for interpretation. Most researchers use $\log_{10}$ so that a change in one unit corresponds to a change in an order of magnitude. Thus, most statistical models that assume normally distributed responses may be appropriate for analyzing viral quantities.

Consider the real-time quantitative PCR (qPCR) data from the dose-response study with *L. sylvaticus* larvae introduced above (Fig. 5).

```
df3 <- data.frame(
  Dose=c(1.4, 1.4, 1.4, 1.4, 1.4, 1.4, 1.4, 1.4, 1.4, 1.4, 1.4, 1.4,
         2.4, 2.4, 2.4, 2.4, 2.4, 2.4, 2.4, 2.4, 2.4, 2.4, 2.4,
         3.4, 3.4, 3.4, 3.4, 3.4, 3.4, 3.4, 3.4, 3.4, 3.4, 3.4, 3.4,
         4.4, 4.4, 4.4, 4.4, 4.4, 4.4, 4.4, 4.4, 4.4, 4.4, 4.4, 4.4,
         5.4, 5.4, 5.4, 5.4, 5.4, 5.4, 5.4, 5.4, 5.4, 5.4, 5.4, 5.4),
  Q=c(0.06, 0.82, 1.06, 1.46, 1.55, 1.61, 1.74, 1.85, 2.17, 2.37,
      2.59, 2.62,
      0.07, 0.82, 0.87, 0.95, 1.54, 1.7, 1.72, 3.56, 5.58, 7.63, 7.69,
      0.18, 2.76, 2.79, 2.84, 2.85, 3.24, 3.26, 3.62, 6.5, 7.01,
7.06, 7.44,
      4.23, 6.2, 6.38, 6.46, 6.46, 6.99, 7.05, 7.32, 7.4, 7.57,
7.59, 7.98,
      4.3, 4.69, 5.68, 5.83, 5.84, 5.88, 5.91, 5.96, 6.46, 6.55,
6.6, 6.62),
  Died=c(0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1,
         0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0,
         0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
         0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
         1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)
)
```

We are interested in whether the $\log_{10}(Q)$ in the larvae that died (filled circles in Fig. 5) varies with the dose to which the larvae were exposed. If there were a quantity of ranavirus in the host that was lethal and hosts die when that quantity is exceeded—a simple null model of virus-induced death we have considered (Brunner et al. 2005)—then there should be no clear relationship between the dose of exposure and the quantity in the larvae upon death. We might use a simple linear regression of the $\log_{10}(Q)$ against the $\log_{10}$(Dose):

```
m3 <- lm(Q ~ Dose,
         data=df3[df3$Died==1,]) # <- just those that died
summary(m3)
```

```
Call:
lm(formula = Q ~ Dose, data = df3[df3$Died == 1, ])

Residuals:
    Min      1Q  Median      3Q     Max
-3.2751 -1.4101 -0.0122  1.4921  3.4849

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.2115     1.0140   2.181  0.03561 *
Dose          0.8057     0.2374   3.393  0.00166 **
---
Signif. Codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1

Residual standard error: 1.677 on 37 degrees of freedom
Multiple R-squared:  0.2373,    Adjusted R-squared:  0.2167
F-statistic: 11.51 on 1 and 37 DF,  p-value: 0.001659
```

From these results it seems clear that there is an effect of dose on at least the $\log_{10}$ quantity of virus found in those larvae that died. It is a good idea to try to visualize what your model is telling you, especially relative to the original data.

```
# plot raw data
plot(Q ~ Dose,
     data=df3[df3$Died == 1,], # <- just those that died
     xlab=expression(log[10]~dose~(pfu/mL)),
     ylab=expression(log[10]~rananvirus~gene~copies)
)
```

```
# generate predicted values
preds <- data.frame(Dose = seq(from=1, to=6, length.out=50))
preds$Q <- predict(m3,
                   newdata=preds,
                   type="response")
# plot the predicted line
lines(x=preds$Dose, y=preds$Q, type = "l")
```

So, it appears that we are capturing the general pattern of increasing viral quantities with increasing exposure dose (Fig. 6), though there is a good deal of deviation from the model's predictions. That is, there is a pattern, but we are certainly not capturing the general features of the data.

1.8.2 Complications and Interpretation

This is a perfectly reasonable model, but we need to be aware of two important considerations. The first issue is how to interpret our model. That is, what do the coefficients for the (Intercept) and ($\log_{10}$) dose mean, biologically? It can be useful to convert from the log-scale to the typical linear scale. Our model says that the expectation of the $\log_{10}$(Dose) is:

$$\mathbb{E}\left(\log_{10}\left(Q_i\right)\right) = \beta_0 + \beta_{\text{dose}} \log_{10}\left(Dose_i\right), \tag{12}$$

where (Intercept) in the *R* output corresponds to β_0 and Dose to β_{dose}. If we exponentiate both sides, we get:

$$\mathbb{E}\left(Q_i\right) = 10^{\beta_0 + \beta_{\text{dose}} \log_{10}\left(Dose_i\right)} = 10^{\beta_0} Dose_i^{\beta_{\text{dose}}}.$$

On this scale, our model does not seem so simple and linear. First, note that 10^{β_0} is the expected pathogen load at death in larvae exposed to a dose of zero (we will come back to this in a moment). We will call this the baseline viral load, which in our example is $10^{2.2115} = 162.7$ copies. Second, as we increase the dose of inoculum,

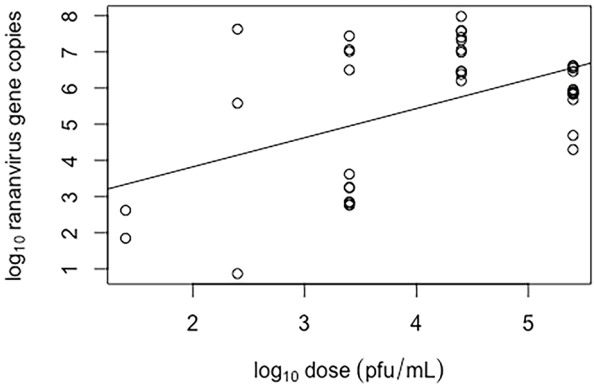

Fig. 6 The intensity of infections in animals that died in the dose-response study by Warne et al. (2011), as well as the best-fit linear regression line fit to these data

we raise it to the *power* of β_{dose}. If β_{dose} were zero, then this second term would always be equal to one, meaning that the expected ranavirus load in a larva is unchanged by the dose of inoculum—a scenario that is clearly inconsistent with these data! If $\beta_{dose} = 1$, then the expected load increases virion-per-virion with the exposure dose. So, we have two values of β_{dose} with clear biological meaning, and the expected relationship between dose and ranavirus load in dead larvae is only linear under these two cases; elsewhere it is a nonlinear power relationship.

In our example, a dose of $\log_{10}(100)$, or 2, would increase the viral load by $2^{\beta dose}$, or $2^{0.8057} = 1.75$ times the baseline, up to $162.7 \times 1.75 = 284.7$ copies. A dose of log(1000), or 3, would then increase the viral load $3^{0.8057} = 2.42$ times the baseline, up to $162.7 \times 2.42 = 393.7$ copies. The effect of dose increases with the order of magnitude of the dose, but less than linearly. In other words, the effect of exposure dose seems to get relatively weaker and weaker as the dose increases in our system. We may be satisfied with this formulation, but it is important to recognize it does *not* produce a coefficient that can be interpreted as the effect *per virion* on the viral load as one might have expected. It can be helpful to think about what you hope your coefficient will mean, and checking the math, before blindly running analyses.

Before moving on, what about the strange meaning of the intercept of the model? We can fix this easily by centering our data on some particular value (e.g., the mean or some useful point of reference). For instance, our lowest dose of exposure was $10^{1.4}$ pfu/mL, so we could subtract 1.4 from all of the doses so that the lowest dose now equates to zero.

```
summary( lm(Q ~ I(Dose-1.4), # <- surround by the "AsIs" function
            data=df3[df3$Died==1,])
)
Call:
lm(formula = Q ~ I(Dose - 1.4), data = df3[df3$Died == 1, ])

Residuals:
    Min      1Q  Median      3Q     Max
-3.2751 -1.4101 -0.0122  1.4921  3.4849

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     3.3394     0.6990   4.777  2.8e-05 ***
I(Dose - 1.4)   0.8057     0.2374   3.393  0.00166 **
---
Signif. Codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1

Residual standard error: 1.677 on 37 degrees of freedom
Multiple R-squared:  0.2373,    Adjusted R-squared:  0.2167
F-statistic: 11.51 on 1 and 37 DF,  p-value: 0.001659
```

Note that the slope is unchanged, but the intercept has changed. Its meaning is still expected virus load in larvae when the predictor, dose, is zero, but now this occurs at the lowest dose (rather than no exposure). These models are mathematically equivalent, but this version is perhaps easier to interpret.

1.8.3 What to Do with Zeroes in Pathogen Load Data Set?

The second consideration when working with $\log_{10}$ quantities is what to do with the zeros. On the one hand, they may represent true negatives, and so they *should* be zero, but on the other hand, they might have come from a positive sample that simply did not test positive for one of a variety of reasons (e.g., they were below the limit of detection; assay failed). How can we discern which is which? The answer is that we cannot, at least not without outside information.

A more proximate, though related, problem is that $\log(0) = -\infty$. Values of negative infinity in our data tend to cause problems with our analyses; if we are lucky, it simply will not run, but if we are not, they might be excluded without us even knowing. So, what to do?

The answer largely depends on what you think these zeros represent, but there are no universally "right" answers. You might exclude the negative test results because you assume they represent real negatives, and you are only interested in values from infected animals. If this assumption is incorrect, then you will tend to bias your load estimates high. Alternatively, you might assume that they were simply low values that, because of the random nature of virus particles or gene targets getting into DNA extraction tubes, falsely tested negative, in which case you might be justified in setting those zeros to a low value, like the limit of detection. You also could add one to all values, since $\log_{10}(1) = 0$, which is a common approach to ensure that all values are ≥ 0. All of these options can be justified, but none are without consequence for your inference or estimates; even the value you decide to add to all samples can affect your results!

Alternatively, you can use zero-inflated or hurdle models. These models account for the probability that an individual is infected using the equivalent of a logistic model and then, given that an animal is infected, predicts the number of virions with another distribution, typically a Poisson or negative binomial, but continuous distributions are also possible to accommodate ranavirus titers. These models can also be applied to other surveillance data, such as the number of infected animals in a population, where there may be zeros because there is no infection in the population or because infected animals were missed during sampling. We direct the readers to Dohoo et al. (2003) and Zuur et al. (2012) for additional guidance on zero-inflated models.

1.9 Economics of Surveillance

The simplest way to reduce the cost of surveillance is not to do it. While it may seem drastic, there are situations where the outcomes of a surveillance program will not alter the management actions (see Value of Information analysis, Sect. 3.1).

However, there are scenarios where surveillance is necessary. In these cases, it's crucial to strike a balance between minimizing costs and maintaining the effectiveness of surveillance.

As discussed in Brunner et al. (2024), some hosts function as reservoirs for the virus and maintain subclinical infections resulting in low population prevalence, while other species serve as amplification hosts and initiate outbreaks. If funds are limited, a viable strategy would be to test amplification hosts, because these species tend to have lower resistance to ranavirus, and detection probabilities are therefore greater. Please see Halliday et al. (2007) for a discussion of how to select species as sentinels for community-wide risk.

A simple method for reducing project costs while maintaining effectiveness is sample pooling—also called batch sampling (Bilder 2019, Fig. 7). If n samples must be collected, pooling them into batches of size k means that if all batches are negative, only n/k batches need to be processed (Brunner 2020; Sabino-Pinto et al. 2019). If a batch is positive, the individual samples in that batch can be retested to find the positives. Effective batch sampling can reduce the number of diagnostics tests required, thus reducing the cost and time surveillance (Brunner 2020).

The economic efficiencies of batch sampling were first investigated by Dorfman (1943) and have since been used extensively to increase sample throughput and reduce costs, such as during the COVID-19 pandemic (Abdalhamid et al. 2020; Brunner 2020; Hogan et al. 2020). The expected savings from pooling samples can be calculated by comparing the cost of testing all individual samples to the cost of testing k pooled batches (Dorfman 1943). It is important to note that if a batch tests positive and data on individual samples is desired, the samples should be DNA

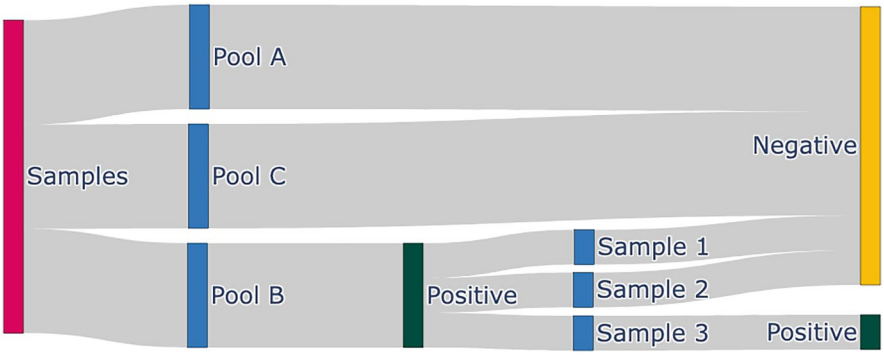

Fig. 7 Sankey diagram of the flow of samples through a two-level hierarchical pooled testing strategy. Starting with nine samples on the left, the samples are pooled into groups of three and tested. Pools A and C are negative, and so all the samples in those pools (six total) are recorded as negative (green). Pool B tests positive (red) and is deconvoluted into individual samples (1, 2, 3). The individual samples are tested again with the same assay, and sample 3 is found to be positive. By pooling the samples, only six out of nine potential tests were required (yellow boxes), resulting in a saving of 33% when using this testing strategy. Optimal pooling size should be determined for each project as savings will depend on factors such as prevalence, pool size, and assay sensitivity and specificity

extracted separately and only pooled during qPCR, which will increase the cost. Most diagnostic laboratories can pool samples and can provide comparative estimates (no pooling, qPCR pooling only, DNA extraction, and qPCR pooling).

Sample pooling has been utilized in ranaviral studies, although formal evaluation of pooling ranaviral samples is lacking (WOAH 2021). Brunner (2020) examined the utility of sample pooling and eDNA to facilitate pathogen surveillance in aquatic animal trade. While eDNA methods can be useful for screening large groups (e.g., in an aquatic trade setting where the goal is to find any infections), it cannot identify infected individuals or estimate prevalence precisely.

Sample pooling to determine individual infections is only effective when prevalence is low (e.g., <30%), although the optimal pooling strategy depends on many factors (Bilder 2019; Brunner 2020; Hitt et al. 2019). Hitt et al. (2019) provide an interactive shiny web app (https://bilder.shinyapps.io/PooledTesting) that can be used to determine optimal pooling size while considering factors such as testing algorithms, prevalence, and assay sensitivity and specificity. Natural groups (e.g., samples from the same locations) can be used to improve sample selection for pools. Positive samples are likely to be correlated due to local transmission chains; thus pooling related samples can reduce the number of pools that must be deconvoluted (Augenblick et al. 2022).

Pooling samples reduces the sensitivity of the assay, i.e., the more samples added to a pool, the more dilute the target will become (Brunner 2020). For this reason, assays utilizing pooled samples should be validated by comparing pooled results to individual samples. Usually, a maximum of three to five samples is recommended per batch.

Box 1 The Value of Simulating Data

It is difficult to provide broadly useful or appropriate guidance about the design and analysis of ranavirus studies. Most every study is different, and even those that follow a common pattern (e.g., experimental exposures, active surveillance among populations) have their idiosyncrasies. However, there is one piece of advice that we think is universal: simulating data from a study, ideally before conducting the study, is immensely helpful. See Electronic Supplement 3 for an example.

What do we mean by simulating data? We mean making up data as if it came from the planned study with samples from the treatments or across the gradient or among the populations you plan to use, assuming the effect sizes you presume will be at work and the sampling (and other) variability you expect. The process of simulating is often iterative. For instance, you might start by simply sketching graphs of the data you hope (or hope not) to see on a white board, then formalizing the relationship(s) you expect to see between the predictor(s) and response, and finally drawing values from distributions based on expected values (e.g., using *R*), revising and reconsidering your

(continued)

Box 1 (continued)

goals and approaches along the way. The further down this iterative process you go, the more you have to gain.

1. At a basic level, even sketching graphs can help you clarify your study, ensuring that your study is likely to answer the question(s) you hope to address. Many times, this process has helped us realize we needed to add treatment levels, use a stratified random sampling approach, or see similar improvements. Specifying a mathematical relationship between the predictor and response (e.g., linear, saturating) and choosing effect sizes (i.e., differences in means, slopes, etc.) then force us to consider whether we are looking for a strong or weak signal in what might be noisy data, whether we are comparing the right groups or levels (e.g., are our treatments distinct enough? Is there sufficient variation between wetlands in the predictor?), and whether our design might be able to distinguish between alternative explanations. We can then consider the variability we are likely to observe to give us a sense of whether the signal might be swamped by the noise in our data. Variability in the response variable comes, at a minimum, from sampling or measurement error. For instance, we do not expect repeated estimates of viral quantities would be identical; small (or large) differences in sample collection, DNA extraction, and qPCR assays will lead to at least some degree of sampling error. Add to this variability in underlying process (e.g., viral replication within a host), and we might reasonably expect estimates of viral quantities to vary by an order of magnitude or more between seemingly identical samples. We can simulate (e.g., with *R* or even Excel formulas) random observations from a distribution with the expected mean (from step 2) and this expected amount of variability to gain a sense of what our data might look like, even if everything is working as we expect. This can be especially helpful in settings where we lack a well-developed intuition, such as when working with binomial data. For instance, many would consider a sample size of $n = 10$ to be sufficient for estimating the probability of success (e.g., infection, death). However, with $p = 0.8$ observations from 5 of 10 to 10 of 10 are all quite likely to be seen from random chance alone.
2. If we are able to simulate data sets, then we are also able to apply our statistical analyses to those data. This is useful from the perspective of building a work flow—from the statistical models fit, statistics calculated, and graphs constructed—which is helpful to have ready for when the real data are available. However, this is even more useful in the sense that, unlike with real data, *you know the truth*. You know the slope or effect size used to simulate the data, so you can test whether your planned statistical analyses can consistently recover the true values of those parameters. With real data you will never know if everything is working right, but with simulated

(continued)

Box 1 (continued)

data you can! This gives us the opportunity to identify problems such as biased estimates or insufficient power to detect differences you know are there.

In summary, the process of constructing and working with simulated data can help us do better science. It helps us ensure that we are embarking on experiments or observational studies that have the best chances of answering our questions. It also gives us confidence that we are doing our analyses well. While this approach places a lot of the hard work *before* the study, that makes the rest of the process much more tractable, powerful, and transparent.

2 Use of Dynamic Models

Dynamic models can be very useful in studying host-pathogen interactions. Within-host models can elucidate physiological mechanisms that lead to host infection and disease (e.g., Mideo et al. 2008, 2011; Woodhams et al. 2008; Mihaljevic et al. 2019). In comparison, between-host models focus on the fate of individuals and populations when a pathogen is introduced or circulating (Hastings 1997). In this section, we will focus on the latter because of their usefulness in predicting the effects of pathogens on populations. The main types of between-host models are Susceptible-Infected-Recovered (SIR) models based on differential equations, Individual-Based Models (IBMs) that track the fate of individuals, and population models that track the transitions between and abundances of life stages or age classes. Deciding which model to use will depend on the questions and the data available. SIR models are best if data relate to prevalence and questions are about the dominant mode of transmission or short-term disease dynamics during a single outbreak. IBMs are helpful when explicitly considering behavior or variation among individuals and can relate to various time scales. Population models are useful when questions relate to host population declines and population persistence, and data include survival rates of different life stages or age classes.

Each model type can provide different and unique insights. It is important to note that models may not necessarily answer all (or any of) your questions directly. However, used carefully, they can guide the next steps, highlight important data needs, and differentiate between plausible and improbable scenarios related to patterns seen in real systems. Model output is always reliant on the quality of information input to the model. Many times, we do not have data for every parameter necessary or even knowledge of when it is reasonable to make certain assumptions (DiRenzo and Grant 2019; Earl 2019). In this case, a useful technique is to consider a range of scenarios and a range of plausible parameter values, i.e., make educated guesses. If changing that parameter or assumption greatly changes model output, then that should guide future data collection.

In this section, we will discuss each type of model, examples that use ranavirus, and other novel examples from other pathogens that illustrate potential future directions for dynamic models of ranavirus-host interactions. We will also provide some limited guidance on model building and testing. Learning to build and test dynamic models requires a lot of time and effort and is beyond the scope of this chapter. However, we encourage laboratory and field scientists to collaborate with modelers. Having a better understanding of modeling fundamentals will help you with establishing collaborations and understanding papers that contain ranavirus dynamics modeling results.

2.1 SI/SIR Models: Transmission

SI/SIR models typically explore transmission dynamics using a system of ordinary differential equations that model and predict one of three outcomes: pathogen extinction, host extinction, or pathogen-host persistence (Allen 2006). In many simple cases, the total population of hosts is divided into three subpopulations: individuals susceptible to infection (S), infected individuals (I), and individuals that have recovered R from infection and cannot be re-infected or at least have temporary immunity. R can also be the individuals removed from the population (e.g., through mortality). A simpler version of the model is where individuals cannot become immune and do not die (i.e., they stay infected for their lifetime), the susceptible-infected (SI) model (Allen 2006). In another scenario, the SIS model, if individuals clear the infection, they become susceptible again. Here, we describe the basic SIR model.

In the simplest SIR model, the total population size (N) can be assumed constant:

$$N = S + I + R \tag{13}$$

where S, I, and R represent the number of individuals in each respective subpopulation (Hastings 1997). The rate of change of each subpopulation at time t can be modeled as

$$\frac{dS}{dt} = -\beta SI \tag{14a}$$

$$\frac{dI}{dt} = \beta SI - \gamma I \tag{14b}$$

$$\frac{dR}{dt} = \gamma I \tag{14c}$$

where β is the rate at which hosts contact and transmit the infection to each other and γ is the host recovery rate (or removal rate). Here, transmission is assumed to be

density-dependent, as transmission is represented as βSI. Some evidence exists that transmission of ranavirus may be density-independent (Harp and Petranka 2006) and can be modeled as $\beta I/N$. McCallum et al. (2001) provide other forms of transmission functions, including nonlinear functions of host density. Because demography (birth, death, immigration, or emigration) is not included in this model, the only equilibrium occurs when all individuals are in the susceptible class (with $I = 0$). For an epidemic to occur, the number of infected individuals must increase, which corresponds with a basic reproductive number greater than one.

The basic reproductive number, R_0, is a key measure for estimating transmission potential and is fundamental in studying how diseases spread. R_0 is defined as the average number of secondary cases that arise from one primary case in a fully susceptible population. It is a measure of the spread of a disease at the beginning of an outbreak. For an epidemic to occur, the number of infected individuals must be initially increasing,

$$\left.\frac{dI}{dt}\right|_{t=0} > 0 \tag{15}$$

given a single infected cases $I(0) = 1$ in an otherwise susceptible population $S(0) = N - 1$. For large populations, we approximate $S(0) = N$ when we compute this initial spread of infection. If the model is simple enough, we can evaluate this initial spread directly from our equation. For example, consider the infection Eq. (14b), evaluating it at $t = 0$ assuming $I(0) = 1$ and approximating $S(0) = N$ yields

$$\left.\frac{dI}{dt}\right|_{t=0} = \beta N - \gamma. \tag{16}$$

The sign of this equation tells us whether the disease will spread or die out, and we formulate this as the basic reproductive number

$$R_0 = \frac{\beta N}{\gamma} \tag{17}$$

If $R_0 > 1$, the number of infected cases will increase, and there will be an outbreak. If $R_0 < 1$, the disease will die out. The expression for R_0 can be interpreted nicely into the product of biological components: the transmission rate β, the duration of infection $1/\gamma$, and the susceptible population density N. The expression for R_0 is model specific and will depend on model assumptions, for example, R_0 *given* in Eq. (17) is specific for model (14).

Many amphibian pathogens, like ranavirus or chytrid fungi, can be environmentally transmitted by virus particles or zoospores. It can be necessary to explicitly model the environmental pathogen pool with a differential equation coupled to the compartmental model of the host population. Letting P denote the concentration of pathogens in the environment, we can expand the basic SIR model (14) into the following SIRP four-dimensional system of equations:

$$\frac{dS}{dt} = -\beta_d SI - \beta_e SP \tag{18a}$$

$$\frac{dI}{dt} = \beta_d SI + \beta_e SP - \gamma I \tag{18b}$$

$$\frac{dR}{dt} = \gamma I \tag{18c}$$

$$\frac{dP}{dt} = \omega I - \delta P \tag{18d}$$

Here, β_d is the transmission rate via direct contact with infected individuals, and β_e is transmission rate via contact with environmental pathogens. Infected individuals shed pathogens at rate ω, which persist in the environment for duration δ^{-1}.

For the simple SIR model (14), we were able to consider whether the infected population is initially increasing to calculate R_0 (Eq. 17). This technique will not be possible for more complicated models; however there is rich mathematical theory using next-generation matrices (NGM) that can help (Diekmann et al. 1990; van den Driessche 2017). For disease transmission models, the NGM method is particularly useful when models have multiple infected compartments. For example, the classical SEIR model divides the infected population into two compartments: exposed (E) and infectious (I) individuals (see Blackwood and Childs 2018 for examples). Below, we show how to use the NGM to calculate the R_0 for model (18).

The NGM approach considers a sub-model of all the "disease" compartments. For model (18) the disease compartments are I and P. We consider a two-dimensional sub-model:

$$\frac{dI}{dt} = \beta_d SI + \beta_e SP - \gamma I \tag{19a}$$

$$\frac{dP}{dt} = \omega I - \delta P \tag{19b}$$

which can be written in a general vector form:

$$\frac{d\vec{x}}{dt} = \mathcal{F}(\vec{x}) - \mathcal{V}(\vec{x}) \tag{20}$$

where $\vec{x}$ is the vector of all disease compartments $[I, P]$. $\mathcal{F}(\vec{x})$ contains terms that yield new infections,[1] and $\mathcal{V}(\vec{x})$ contains all other terms or the transition terms into and out of these infected compartments.

[1] Here we assume that shedding is not a new infection and place the shedding term ωI into the transition vector $\mathcal{V}$. Alternatively, you can consider shed pathogen particles as new infection and get a different expression for the next-generation matrix, which should result in the same threshold for $R_0 = 1$, but may be harder to interpret biologically. See van den Driessche (2017) for details using a cholera model example.

$$\frac{d\vec{x}}{dt} = \begin{bmatrix} dI/dt \\ dP/dt \end{bmatrix} = \left(\begin{bmatrix} \beta_d SI + \beta_e SP \\ 0 \end{bmatrix} - \begin{bmatrix} \gamma I \\ -\omega I + \delta P \end{bmatrix} \right) = \mathcal{F}(\vec{x}) - \mathcal{V}(\vec{x}) \tag{21}$$

Now, from vectors $\mathcal{F}(\vec{x})$ and $\mathcal{V}(\vec{x})$, we define matrices of partial derivatives evaluated at $(S, I, R, P) = (N, 0, 0, 0)$, the disease-free equilibrium:

$$F(\vec{x}) = \left[\frac{\partial \mathcal{F}_i}{\partial x_i} \right] = \begin{bmatrix} \frac{\partial \mathcal{F}_1}{\partial I} & \frac{\partial \mathcal{F}_1}{\partial P} \\ \frac{\partial \mathcal{F}_2}{\partial I} & \frac{\partial \mathcal{F}_2}{\partial P} \end{bmatrix} = \begin{bmatrix} \beta_d S & \beta_e S \\ 0 & 0 \end{bmatrix} \Bigg|_{(N,0,0,0)} = \begin{bmatrix} \beta_d N & \beta_e N \\ 0 & 0 \end{bmatrix} \tag{22}$$

$$V(\vec{x}) = \left[\frac{\partial \mathcal{V}_i}{\partial x_i} \right] = \begin{bmatrix} \frac{\partial \mathcal{V}_1}{\partial I} & \frac{\partial \mathcal{V}_1}{\partial P} \\ \frac{\partial \mathcal{V}_2}{\partial I} & \frac{\partial \mathcal{V}_2}{\partial P} \end{bmatrix} = \begin{bmatrix} \gamma & 0 \\ -\omega & \delta \end{bmatrix} \Bigg|_{(N,0,0,0)} = \begin{bmatrix} \gamma & 0 \\ -\omega & \delta \end{bmatrix} \tag{23}$$

These two matrices are used to define the next-generation matrix FV^{-1}. Here V^{-1} is the inverse of matrix V. R_0 is the spectral radius (dominant eigenvalue) of the next-generation matrix:

$$R_0 = \rho\left(FV^{-1}\right). \tag{24}$$

For our example this expression takes the following form:

$$R_0 = \rho\left(\begin{bmatrix} \beta_d N & \beta_e N \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \gamma & 0 \\ -\omega & \delta \end{bmatrix}^{-1} \right) = \left(\frac{\beta_d}{\gamma} + \frac{\beta_e \omega}{\gamma \delta} \right) N \tag{25}$$

Our R_0 has two components that relate to the two routes of transmission in the model: direct transmission from contact with infected individuals $\left(\frac{\beta_d N}{\gamma} \right)$ and environmental transmission from contact with pathogen particles $\left(\frac{\beta_e \omega N}{\gamma \delta} \right)$. Each term in the R_0 can be nicely translated biologically:

$$\left[\overbrace{\underbrace{(\beta_d)}_{\text{trans. rate}} \underbrace{\left(\frac{1}{\gamma}\right)}_{\text{infection duration}}}^{\text{Direct transmission}} + \overbrace{\underbrace{(\beta_e)}_{\text{trans. rate}} \underbrace{\left(\frac{1}{\gamma}\right)}_{\text{infection duration}} \underbrace{(\omega)}_{\text{shed rate}} \underbrace{\left(\frac{1}{\delta}\right)}_{\text{pathogen persistence}}}^{\text{Environmental transmission}} \right] \overbrace{N}^{\text{Host density}}$$

For ranavirus and most natural populations, the basic SIR model (17) and SIRP model (18) are likely too simplistic, and many published SIR type models involving

ranavirus expand on this model in different ways dependent on the host species and ranavirus strain. Duffus et al. (2019) used a discrete-time SI model to show that ranavirus could be maintained in a population of common frogs (*Rana temporaria*) in the UK with only transmission between adults. They also demonstrated that transmission between adults could maintain two syndromes of ranavirus (the ulcerative and hemorrhagic forms) in a single population. Farrell et al. (2018) modeled ranavirus transmission in tiger salamanders. They examined effects of the incidence function and the function describing life expectancy relative to time since infection on host persistence and population size after an epidemic.

Brunner and Yarber (2018) and Peace et al. (2019) modelled transmission in wood frogs (*Lithobates sylvaticus*) to determine the most likely route(s) of transmission. Both models compared the relative importance of environmental transmission, transmission from scavenging dead conspecifics (necrophagy), and from direct contact. Both models included terms to track the presence of carcasses available for necrophagy and the concentration of virions shed in the water available for environmental transmission. Peace et al. (2019) also included multiple stages of infection with different probabilities of mortality and virus shedding rates. Both models found that direct contact was the most important mode of transmission, but Peace et al. (2019) also found that environmental transmission could be more important with higher tadpole densities and later in an epidemic.

Other model expansions could be useful for predicting and/or gaining a better understanding of ranavirus dynamics in natural populations. For example, most ranavirus host species exist in communities where they are likely to interact with other susceptible species, possibly from different ectothermic vertebrate classes (Gray et al. 2009). Brenes et al. (2014a) demonstrated that interclass transmission of ranavirus through water was possible. He also showed that ranaviral disease outcomes depended on species composition in the amphibian community and which species was initially infected with ranavirus (Brenes 2013). These studies could serve as a starting point for determining transmission probabilities in aquatic communities with multiple species. In other disease systems, the addition of multiple species to transmission models had an effect on the focal host population, but depended on the host's competency as a reservoir and its dominance within the community (Keesing et al. 2006). The addition of multiple host species can make the analysis of SIR models challenging. To date, most models have included only two species and the pathogen (Keesing et al. 2006), which may be unrealistic for some ranavirus-host systems. Dobson (2004) dealt with the large number of parameters in multi-species models by scaling the parameters as allometric functions of host body size, although it is unclear such a relationship exists with transmission of ranavirus. Lélu et al. (2013) provide an example of a model including trophic transfer of a parasite (*Toxoplasma gondii*) from rats to cats and vertical transmission in cats. Similar complex interactions certainly occur among ranavirus hosts species, such as predation or necrophagy, and mechanical transmission by mosquitoes has been hypothesized (Allender et al. 2006; Johnson et al. 2007; Kimble et al. 2015). Despite the large number of possible interactions in a ranavirus-host system, several

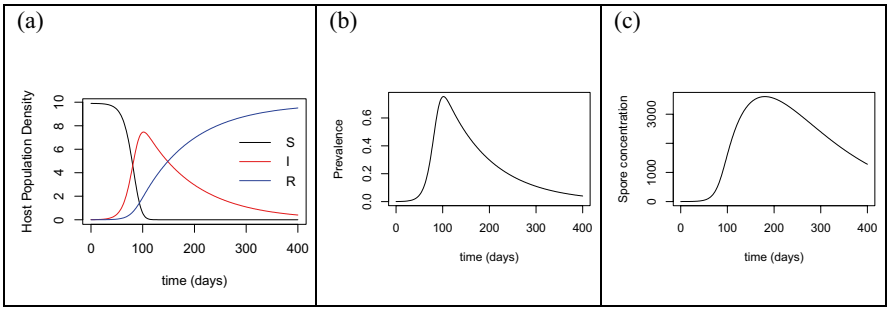

Fig. 8 Solutions for the base SIRP model (18) showing population densities (**a**), prevalence (**b**), and environmental spore loads (**c**) changing over time. Simulations were generated using ODE solvers in R's deSolve. *R* code for solving this model and generating these figures is at the end of Sect. 2.1

interactions are likely unimportant to its epidemiology. One strategy would be to create several competing models and fit them to data on dynamics in natural populations or in mesocosm studies to identify the most important mechanisms for transmission.

For researchers interested in using SIR models to examine ranavirus, we recommend Otto and Day's (2007) book, *A Biologist's Guide to Mathematical Modeling in Ecology and Evolution*, which reviews the mathematics and describes the process necessary for constructing and analyzing models primarily with ordinary differential equations. An understanding of computer programming and use of software (e.g., Matlab, Maple, Mathematica, *R*) will be necessary to construct models and perform simulations for most analyses. Sample *R* code for solving the base model (18) and producing Fig. 8 is below.

```
#R script to solve SIRP Model (18)

#required package for ODE solver:
install.packages("deSolve")
#load packages so we can use the ode function:
library(deSolve)

#Model Equations:
Model_SIRP <- function(t, variables, parameters)
{with( as.list(c(variables,parameters)),  #lets us access
variables and parameters by name
       { #System of ODES:
         dS = -betad*S*I-betae*S*P
         dI = betad*S*I+betae*S*P -gamma*I;
         dR = gamma*I;
         dP = omega*I-delta*P;
         #End of ODes
```

```
          return(list(c(dS,dI,dR,dP)))
        })
} #End Model Equations

#Initialization: load parameter values:
betad=0.0003  #Direct transmission rate
betae=.0001  #Environmental transmission rate
gamma=0.01 #Recovery rate
omega=10  #spore shedding rate
delta=0.01 #Spore degradation rate
#set parameter list:
parameter_values<-c(betad=betad, betae=betae, gamma=gamma,
omega=omega, delta=delta)

#Initialization: Set time:
tend=400 #Days for simulation
time_values<-seq(0,tend)

#Initialization: Set initial conditions:
N0=10; #%Initial population density

initial_values <- c(
  S=N0*0.999,
  I=N0*0.001, #assume 0.1% of the population is initially
infected
  R=0,
  P=0
)

#Solve Model
Model_solutions <- ode(
  y = initial_values,
  times = time_values,
  func = Model_SIRP,
  parms = parameter_values
)

Model_solutions<-as.data.frame(Model_solutions) #convert output
to dataframe so we can use with()

#plotting
with(Model_solutions, {
  plot(time, S, type = "l", col = "black",
       xlab = "time (days)", ylab = "Host Population Density")
  lines(time, I, col = "red")
  lines(time, R, col = "blue")
```

```
  legend("right", c("S", "I", "R"),
         col = c("black", "red", "blue"), lty = 1, bty = "n")
})

#Plot Prevalence
with(Model_solutions, {
  plot(time, I/(S+I+R), type = "l", col = "black",
       xlab = "time (days)", ylab = "Prevalence")
})

#Plot Pathogen load
with(Model_solutions, {
  plot(time, P, type = "l", col = "black",
       xlab = "time (days)", ylab = "Spore concentration")
})
```

2.2 Individual-Based Models/Pattern-Oriented Modeling

Individual-based models (IBMs), sometimes called agent-based models (ABMs), are also very useful for examining disease dynamics. IBMs are simulation-based, and during each time step, a set of rules or probabilistic events occur involving each individual. IBMs are often easier for biologists to construct than SIR models, because they do not require solving differential equations. However, IBMs can be complex and require computer programming skills. These models often operate on a set schedule of events that are implemented using sequential equations, a series of for-loops, and if-then statements that determine an individual's actions or fate. For disease IBMs, each individual's disease state is recorded, and their risk of infection can depend on their interaction with other individuals or the environment.

IBMs for ranavirus have been applied to Bosca's newt (Rosa et al. 2019) and the marbled newt (Rosa et al. 2022) in Portugal. The Bosca's newt example examined the length of time that outbreaks occurred and the effect of sex-biased mortality from ranavirus on population persistence. Researchers found that female-biased ranavirus mortality did not cause larger population declines but did slow down population recovery once outbreaks ceased (Rosa et al. 2019). In this model, mortality was applied as a percentage. In the Marbled Newt example, an individual-based ranavirus transmission model was linked to an individual-based population model and matched to existing population data covering 40 years (Rosa et al. 2022). Scenarios examined effects of ranavirus outbreaks and an invasive fish species separately and, when combined, found that both stressors were needed to generate patterns similar to those seen in the population data.

One attractive aspect of IBMs is that they can explicitly incorporate animal behavior. For ranavirus, researchers might be interested in how different behaviors,

such as schooling or necrophagy, affect host populations and persistence with the pathogen. Further, many researchers use IBMs to better understand spatial dynamics by explicitly modeling animal movement. Incorporating animal movement can allow research to estimate the spread of pathogens to new areas. Animal movement can affect epidemiological patterns when space is heterogenous (Fofana and Hurford 2017).

There are also other types of IBMs that use differential equations. For example, Briggs et al. (2010) developed an IBM with differential equations that explicitly incorporated individual chytrid fungus load and examined how a pathogen reservoir and a long-lived tadpole stage affected whether frog populations could persist with chytrid infections or would experience local extinction. Similar models could be developed for ranavirus that include viral load and shedding to better understand how the virus might inter act with the host and factors that contribute to die-offs.

A useful technique for creating IBMs and determining plausible interactions is called pattern-oriented modeling (POM). In POM, data are used to determine several salient patterns seen in a natural system of interest that form the basis of model evaluation. Multiple possible forms of an IBM are created, representing different hypotheses about host-pathogen interactions. The different IBMs are evaluated based on their ability to recreate the patterns (Grimm et al. 2005; Grimm and Railsback 2012). When a model matches multiple patterns, it is more likely to be structurally realistic (Wiegand et al. 2003) and capable of producing testable predictions. In using POM, researchers can also contrast different hypotheses, determine a useful model structure, and reduce parameter uncertainty.

For researchers interested in developing IBMs, we recommend two books: Grimm and Railsback's (2005) *Individual Based Modeling and Ecology* and Railsback and Grimm's (2011) *Agent-Based and Individual-Based Modeling*: *A Practical Introduction*. Both titles describe a "best model practice" called object-oriented design and description (ODD), which is a standard format to describe various aspects of an IBM. The latter title goes through the process of building IBMs with examples and code for a relatively user-friendly and free program called NetLogo (http://ccl.northwestern.edu/netlogo/index.shtml). NetLogo includes a library of preconstructed models, including AIDS, Disease Solo, and Virus, which could form the basis for the development of models for ranavirus. Further, NetLogo's website includes a Modeling Commons, where NetLogo users can share their models to help others in their own model development. Another free software is Vortex (https://scti.tools/vortex/), which is an individual based population modelling framework. Vortex can be linked to Outbreak (https://scti.tools/outbreak/), which models transmission dynamics and can incorporate management scenarios, and Spatial (https://scti.tools/spatial/; tracks movements) through MetaModelManager (https://scti.tools/metamodelmanager/). All of the extensions are also free. Rosa et al. (2022) model ranavirus by linking Vortex and Outbreak through MetaModelManager. Other software, such as Matlab and *R*, can be used to develop and analyze IBMs.

2.3 Population Matrix Models

Population matrix models examine changes in population size and age structure over time. These models include parameters for the transition probability between each age class. To incorporate disease, the survival following exposure to ranavirus can be incorporated for each age class. Several population models have been developed for amphibian populations affected by ranavirus (Campbell et al. 2018; Earl and Gray 2014; Earl et al. 2016). These include either age- or stage-structured matrix models to predict the effects of ranavirus exposure during different life stages. These combine population models with data on mortality after ranavirus exposure (e.g., Haislip et al. 2011) to predict population outcomes. Earl and Gray (2014) and Earl et al. (2016) modeled populations of three different North American anuran species and demonstrated that local extinction of closed populations is possible. Earl et al. (2016) further assessed the impacts of immigration from nearby populations on population persistence. Campbell et al. (2018) modeled common frogs (*Rana temporaria*) and showed that ranavirus outbreaks could truncate age distributions, which made populations more susceptible to environmental stochasticity.

Population matrix models can also be combined with transmission models to more realistically model both dynamics simultaneously. For example, Briggs et al. (2005) merged a population model of yellow-legged frogs (*Rana muscosa*) and an SIR model of the infection dynamics of Bd based on the current knowledge of transmission and mortality rates. This model combined discrete-time between-year population dynamics with a continuous time transmission dynamics within each year. By running the model with different parameter values, Briggs et al. (2005) were able to determine which conditions resulted in extinction of the frog population, non-persistence of the pathogen, and persistence of the frog population and the pathogen.

Population models can also be scaled up to take into account metapopulation processes. A metapopulation is a set of spatially structured local populations that periodically interact via dispersal (Marsh and Trenham 2001; Smith and Green 2005). Several ranavirus host species are likely structured as metapopulations. Metapopulation models incorporate parameters for dispersal probability between local populations and demographic parameters in each local population. Metapopulation models are useful to understand the spatial spread of pathogens among populations and examine the effectiveness of disease intervention strategies (Hess 1996). For researchers interested in population matrix models, we recommend Caswell's (2000) *Matrix Population Models*: *Construction*, *Analysis*, *and Interpretation*. Hanski's (1999) *Metapopulation Ecology* will be useful for those interested in investigating ranavirus effects on metapopulation dynamics.

2.4 Modeling Disease Intervention Strategies

One goal of modeling host-pathogen dynamics is to identify intervention strategies that thwart disease outbreaks. Currently, there are few proposed control options for ranavirus, but vaccine development is possible in the future (Miller et al. 2011;

Chen et al. 2018). Other options include quarantining individuals or populations, culling, and creating captive populations for reinforcement or reintroduction if disease is likely to cause extremely high mortality to populations of conservation concern. Reinforcement efforts could easily be modeled with immigration parameters. Models also can be used to identify vulnerable points in the host-pathogen cycle that can be interrupted with intervention strategies. For example, if outbreaks are a consequence of density, emergent vegetation in wetlands can reduce the probability of transmission among amphibian larvae (Greer and Collins 2008; Malagon et al. 2020). If stressors in the aquatic environment (e.g., high nitrogen levels) are resulting in reoccurring outbreaks, strategies that improve those stressors (e.g., water quality) can be used. Rosa et al. (2022) modeled the removal of an invasive fish species that was affecting newt populations simultaneously with ranavirus outbreaks. The model showed that fish removal would likely prevent local extinction.

A thorough understanding of the factors responsible for outbreaks and the ranavirus-host system is essential to identifying plausible intervention strategies. In some cases, possible intervention strategies might be infeasible to implement, excessively costly, or undesirable in natural populations. However, if strategies are feasible, models can be used to determine when and how often the strategy should be employed for the best results. SIR models and their variants can be used to explore vaccination strategies (Hethcote 2000) and other control techniques such as culling (Lloyd-Smith et al. 2005). Cost of disease control can be incorporated into models to determine the best strategies given financial constraints (Fenichel et al. 2010). Woodhams et al. (2011) discussed possible intervention strategies for *Batrachochytrium dendrobatidis* and presented model results of their efficacy on individuals with and without an adaptive immunity. They also went on to show that reducing the host population size (i.e., decreasing transmission probability) could prevent extinction. For researchers interested in implementing optimal control models, we recommend Lenhart and Workman's (2007) *Optimal Control Applied to Biological Models*, which focuses on control of continuous ordinary differential equation models and includes sample code for the computer program Matlab. Optimal control can also be applied to IBMs, but effective techniques are still being developed (Federico et al. 2013).

2.5 Model Parameterization and "Evaludation"

There are a number of ways to parameterize models and integrate them with data. Frequently, modelers choose parameter values by searching the literature, but often not all parameter values are available. Another method is to construct a model and fit the output to an existing data sequence (e.g., Mihaljevic et al. 2019; Rosa et al. 2022), which unfortunately is not very common in amphibian population models (Earl 2019). In the case of ranavirus modeling, predictions could be fit to surveillance data that include abundances of infected and uninfected individuals or the magnitude and timing of a die-off. After the model is fit to the data, the parameter values that give the best fit or that match multiple patterns (as in POM) are then used.

If some parameters are known and researchers have a good idea of the possible range of other parameters, these ranges of values can be explored to determine how they change the model output. Assessing the effects of changes in parameter values is called sensitivity analysis (Cariboni et al. 2007). If the model is especially sensitive to a certain parameter, it suggests that better parameter estimation would be a valuable research direction (Biek et al. 2002; Cariboni et al. 2007), especially if the parameter estimate is not based on robust data (e.g., low sample sizes). Cariboni et al. (2007) suggest best practices for sensitivity analysis. An excellent review of parameter estimation for disease modeling of natural populations can be found in Cooch et al. (2012).

The aim of model evaluation is to determine if models typify natural systems well enough to represent the intended dynamics. This often involves determining whether or not they can be used to make accurate predictions. Frequently, the terms model evaluation, model validation, and model testing are used interchangeably. Because models are built on assumptions and simplifications, they are never truly "valid" or "correct." Augusiak et al. (2014) have suggested the term "evaludation" to represent the process of assessing the model's quality and reliability and discuss six elements for proper "evaludation" of a model:

- Assessing the quality of the data used to build the model
- Evaluating the simplifying assumptions structuring the model
- Verifying that the model is correctly implemented
- Verifying that the output matches the data used to design the model
- Exploring model sensitivity to changes in parameter values
- Assessing whether the model can fit an independent data set not used in original model formulation

It is recommended that model formation and "evaludation" follow a documentation procedure called TRACE (TRAnsparent and Comprehensive Ecological documentation) that is designed to ensure reliability of models and link the science to application (Grimm et al. 2014).

3 Risk Analysis for Introduction of Ranavirus into an Uninfected Area

The movement of animals inherently carries the risk of transmitting their parasites and diseases. In an ideal world, disease spread via animal trade and translocation would be prevented through testing of all animals for potential pathogens. However, real-world decisions must be made based on restricted information and finite resources. Risk analysis plays a crucial role in these decisions. This section delves into the multifaceted process of decision-making and risk analysis for the introduction of ranaviruses into uninfected areas.

3.1 Value of Information Analysis

There is an assumption that obtaining more information will always lead to better decision-making and management outcomes. Value of Information (VoI) analysis serves as a critical tool in the realm of decision-making (Schlai and Raiffa 1961). This methodology quantifies the potential benefit of obtaining additional information to enhance decision outcomes (Canessa et al. 2015). The key principle underlying VoI is that the acquisition of further data is deemed valuable only if it possesses the capacity to meaningfully influence or alter the preferred strategy currently considered by the decision-maker (Williams and Johnson 2015). When evaluating the merit of pursuing more information, one must account for the "cost" of this pursuit, which is typically characterized by the financial expenditure necessary to obtain the additional data. Thus, VoI analysis allows decision-makers to weigh the expected benefits of enhanced decision-making against these costs, facilitating more informed and effective decisions.

Consider the following scenario: Over time, a decline in a local frog population has been noted, and it is determined that it is necessary to bolster the population by translocating individuals from an external source population. Currently the population is free from ranaviral disease; however, there is a concern that ranaviruses could be introduced with the translocated frogs. Environmental DNA (eDNA)-based screening methods will be used to assess potential source populations for ranaviruses; however, there is limited funding available, and so unnecessary screening should be reduced where possible. Three potential source populations (*Sites 1*, *2*, and *3*) have been identified. The goal is to determine which, if any, of these three potential source populations should undergo eDNA screening for ranavirus given the desire to minimize screening expenses.

VoI analysis can be used to help decide which potential source populations should be prioritized for eDNA screening based on the potential impact of decision outcomes. Canessa et al. (2015) provide a comprehensive primer on VoI analysis, and here we adapt their methods to the scenario above.

The prior probability of ranavirus in each of the three populations is 0.1, 0.5, and 0.8 based on the ranaviral infection rate at surrounding sites. The eDNA screening method has a sensitivity (true-positive rate) and specificity (true negative rate) of 0.4 and 1, respectively. For simplicity, the population size is 100 frogs, and 35 will be introduced from a single source population. If ranavirus is introduced with the translocated frogs, then modeling predicts that the local population will be reduced to 55 frogs. If ranavirus is not introduced, no frogs will die leaving a population of 135 frogs. If no action is taken, the population remains stable (100 frogs total) regardless of ranaviral status at source sites. This information can be used to construct a consequence table (Table 1).

The expected value for translocating from each site with uncertainty, i.e., without testing (EVTU), represents the expected number of frogs if the decision is made to translocate regardless of the outcome. The EVTU (Table 2) can be calculated with the following:

Table 1 Consequence table of translocating or doing nothing under the hypothesis that ranavirus is present or not in the source population

	Ranavirus (rana+)	No ranavirus (rana−)
Translocate (T)	55	135
Do nothing (D)	100	100

Table 2 Components of EVPI calculations

Site	P(rana+)	EVTU	EVU	Decision uncertainty	EVC	EVPI	Testing priority
1	0.1	127	127	Translocate	131.5	4.5	3
2	0.5	95	100	Do nothing	117.5	17.5	1
3	0.8	71	100	Do nothing	107	7	2

Refer to Canessa et al. (2015) for equation details

$$V\left[T,\text{rana}+\right]*P\left(\text{rana}+\right)+V\left[T,\text{rana}\right]*P\left(\text{rana}-\right)$$

where $V[T$, rana+] and $V[T$, rana−] are the predicted population sizes if frogs are translocated under the hypothesis of ranavirus at the source site or not (55 and 135). P(rana+) and P(rana−) are the probabilities of ranavirus (0.1, 0.5, or 0.8) or no ranavirus (0.9, 0.5, or 0.2) at the source site. The EVTU represents the expected number of frogs if the decision is made to translocate regardless of the outcome.

Translocating is not the optimal decision under uncertainty at *Site 3* as the EVTU is 71 which is less than the expected value of doing nothing (100). The expected value (number of frogs) from making the optimal decision under uncertainty (EVU) is 100 in the case of *Site 3* as doing nothing has more expected value.

In each case, there is the potential to increase the expected value by gathering more information; however, this is dependent on whether the presence of ranavirus at the source site can be accurately determined. The expected value of making a decision under certainty (EVC) is the maximum value that could be gained by testing the potential source populations. By assuming a perfect screening method, i.e., no false positives or negatives EVC can be calculated as:

$$\begin{aligned}&\left[\max\left(V\left[T,\text{rana}+\right],V\left[D,\text{rana}+\right]\right)*P\left(\text{rana}+\right)\right]\\&+\left[\max\left(V\left[T,\text{rana}-\right],V\left[D,\text{rana}-\right]\right)*P\left(\text{rana}-\right)\right],\end{aligned}$$

where $V[D$, rana+] and $V[D$, rana−] are the predicted population sizes if no frogs are translocated (100 and 100). The expected value of perfect information (EVPI), that is, the expected value gained by eliminating uncertainty entirely, can be calculated by subtracting the expected values under certainty and uncertainty (Raiffa 1968).

Intuitively, a negative result from a perfect screening method at *Site 3* would indicate that this site is suitable as a source population. However, there is a slim chance of a negative result (as the prior suggests), and so the EVPI is low (Table 2). Similarly, at *Site 1* the prior (0.1) suggests a low chance of finding ranavirus. The

EVPI is low at *Site 1* because the results of a perfect test will likely be negative, so little value is gained by testing. At *Site 2*, the prior (0.5) indicates uncertainty around the ranavirus status. Screening the population at *Site 2* would relieve a lot of the uncertainty, and so the EVPI is high.

EVPI provides a useful measure of the maximum benefit from perfect information; however, like most real-world scenarios, the eDNA screening method is not perfect (sensitivity = 0.4). Bayesian pre-posterior analysis (Berger 1985) and the methods described in Canessa et al. (2015) can be used to calculate the expected value of sample information (EVSI), that is, the value expected to be gained by performing the imperfect eDNA screening at each site.

EVSI is calculated by considering how obtaining a positive (EV+) or negative (EV−) results would change beliefs (prior) and decisions (Decision+ and Decision−). Canessa et al. (2015) provide a useful spreadsheet as supplemental information that can be used to calculate the values in Table 3.

If the eDNA screening method is positive at any site, then it would not be chosen as source populations (Decision+) as the eDNA screening method does not produce false positives (100% specificity).

The probability of ranavirus at *Site 3* is high (0.8) and test sensitivity is low; therefore the negative predictive value (NPV) is also low. As a result of the low NPV, a negative screening of *Site 3* would likely be a false negative. Therefore, Site 3 can be definitively ruled out as a source population without any testing required.

In the case of *Site 1*, a lot of value is gained if the eDNA screening method is negative (higher NPV); however, the EVSI is low as there is a strong prior belief (0.1) that the test will be negative; therefore not much is gained by performing the test. A negative eDNA screening result at *Site 2* would change the decision from doing nothing (decision uncertainty, Table 2) to translocating (Decision−, Table 3), emphasizing the utility of VoI analysis to identify areas of focus.

In cases where the expected value of sample information is greater than zero, the cost of performing the eDNA screening must be weighed against the EVSI, i.e., 1.8 frogs at *Site 1* and 4 at *Site 2*. For example, if funds are only available for one site, then screening *Site 2* would give twice as much value. Following screening (assuming *Site 2* is negative), other factors could be used to select between *Site 1* and *2*, such as the absolute expected value, cost/difficulty of translocation, source population health, other pathogens, etc. (see Sect. 3.2).

When performing VoI analysis, it is important to keep in mind that Bayesian updating, EVPI, and EVSI calculations are just tools to inform decisions. They are based on estimates and models, which may not represent real-world complexities.

Table 3 Components of EVSI calculations

Site	P(rana+)	EV+	Decision+	EV−	Decision−	EVSI	Testing priority
1	0.1	100	Do nothing	130	Translocate	1.8	2
2	0.5	100	Do nothing	105	Translocate	4.0	1
3	0.8	100	Do nothing	100	Do nothing	0.0	No test

Refer to Canessa et al. (2015) for equation details

Therefore, while VoI analysis can provide useful insights, it should be used in conjunction with other analyses, knowledge, expertise, and judgment.

3.2 Risk Analysis

Risk analysis is a procedure that can be used to determine the threat of a pathogen entering a system. The consequences of pathogen introduction can be monitored directly (Sect. 1) or simulated using models (Sect. 2). Several standards exist for performing risk analysis of animal diseases, including the World Organization for Animal Health (WOAH, founded as OIE) risk analysis model and the OIE-IUCN Wildlife Disease Risk Analysis (DRA) guidelines (Brückner et al. 2010; Jakob-Hoff et al. 2014; Travis and Smith 2019). Both standards can be used to assess the potential threats of pathogens entering a system and the risks associated with their spread, although they differ in their focus, scope, and application (Fig. 9).

The OIE risk analysis model was primarily developed from a trade perspective as a tool for Import Risk Analysis (IRA) between two countries or regions. IRA is used to transparently assess the disease risk associated with the importation of animals or their products (Brückner et al. 2010). The OIE-IUCN Wildlife DRA guidelines were specifically developed with wildlife disease in mind. The OIE risk analysis model was adapted to encompass the special features associated with disease risk analysis as it is applied to wildlife and biodiversity conservation resulting

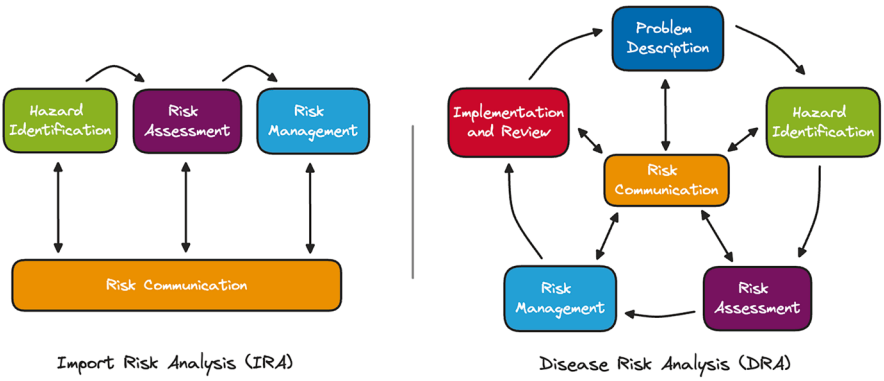

Fig. 9 Comparison of IRA and DRA frameworks. Boxes represent major components as identified by the respective guidelines. Both frameworks emphasize risk communication between interested stakeholders and experts throughout the analysis. The DRA framework identifies problem description (blue) and implementation and review (red) as distinct components (nested in other components in IRA). Starting and ending with the problem description, the circular nature of the DRA framework emphasizes the requirement of dynamic and ongoing analyses. The IRA is more linear and focused on producing an objective and defensible method of assessing disease import risks, ultimately resulting in sanitary measure recommendations

in the production of the Manual and Guidelines for Wildlife DRA (Jakob-Hoff et al. 2014).

In general IRA focuses on the potential infection of one species or several species within the same taxonomic class, whereas DRA can involve a broader range of species and ecosystems. As discussed in Duffus et al. (2015), many ranaviruses are multi-species pathogens that have the capability of infecting three vertebrate classes, which makes IRA more suited for the analysis of ranaviruses that infect a single class such as ENHV in fish. IRAs can be used to establish or revise trade or translocation guidelines for wildlife that could be subclinically infected with a pathogen (Smith et al. 2009). The WOAH lists ranaviruses that infect amphibians as notifiable pathogens, meaning that a subsample of amphibians that are involved in international trade should be verified ranavirus negative prior to shipment (Schloegel et al. 2010), although the degree to which these regulations are enforced globally is not clear (Kolby et al. 2014).

The procedures we outline below are based on principles and recommendations of the WOAH and IUNC (Brückner et al. 2010; Jakob-Hoff et al. 2014; Vose 2000); we provide examples of how they can be applied to parts of an IRA for the introduction of a ranavirus into an uninfected area. While this section focuses on IRA, many of the principles and recommendations can be applied to a wildlife DRA, which shares much of the terminology and concepts (Fig. 9).

3.2.1 Defining the Hazard

The first step in an IRA is defining an area of interest. The area could be a population of interest, such as one that contains an uncommon species that is susceptible to ranavirus, or it could be a geographic region or country (Rödder et al. 2009; OIE 2014). Generally, areas are defined based on artificial or natural barriers to animal movement or pathogen translocation (OIE 2014). For example, ranavirus virions can flow downstream in tributaries, and associated floodplains are often corridor for animal movement; thus, areas should be defined by watershed for lotic systems. In lentic systems, depressional wetlands or lakes containing possible ranavirus hosts could be defined as the area of interest, if it is hydrologically closed and surrounded by a terrestrial landscape. In zoological settings, the area of interest typically is the captive facility (OIE 2014).

The next step is determining the presence of ranavirus in the area of interest. Section 1 discussed surveillance studies, and additional guidelines are provided by OIE (2014). Minimum sample size to detect ranavirus depends on several factors (Sect. 1). Additionally, infrequent sampling can result in lack of detection. Todd-Thompson (2010) showed that ranavirus in Gourley Pond of the Great Smoky Mountains National Park appeared nonexistent except for a 3-week period in late spring when an outbreak occurred resulting in widespread mortality across multiple species. Thus, sampling sites every 2 weeks when hosts are present with a large sample size ($n > 30$) should result in a high detection probability. If resources are

limited, sampling at least four periods per year while hosts are present may be sufficient. Using this sampling frequency, Hoverman et al. (2012) detected ranavirus at 33 of 40 sites. Given that ranavirus could have been present at all sites in this study, a ballpark estimate of detection probability was 82.5–100% with their sampling frequency. Sampling should be performed over several years to verify that a site is ranavirus negative. Value of information analysis (Sect. 3.1) may be useful for determining the cost benefit tradeoff of additional surveillance. For large areas of interest, multiple sites spaced no less than the average dispersal distance of hosts should be sampled, which for amphibians is about 1 km (Wells 2007). Thus, distinct populations should be sampled without leaving large gaps between them. Environmental DNA methods may also be useful for determining ranavirus status without extensive effort (Brunner 2020). If ranavirus is detected, there is no reason to conduct an IRA, unless there is concern of a foreign strain of ranavirus being introduced.

Although the primary interest in the introduction of ranavirus to an area typically is for a certain species of concern, it is important that all ranavirus hosts are considered in an IRA. Marschang et al. (2024) provide a list of known ranavirus hosts, and several challenge studies (e.g., Hoverman et al. 2011; Brenes et al. 2014b) can provide insight into relative difference in susceptibility between species.

3.2.2 Risk Assessment

Risk assessment involves three primary steps: identifying routes of introduction, identifying the consequence of introduction, and estimating risk. It is often useful to develop flow diagrams that illustrate each step of assessment (Figs. 9 and 10). To describe this process, below we provide an example of assessing risk to wild amphibians via import of aquacultured fish that are infected with ranavirus.

- *Routes of Introduction*

Routes of introduction could include dispersal paths of hosts or translocation of the virus on fomites attached to non-hosts (i.e., birds and mammals, Gray et al. 2009). Humans can play a large role in the possible introduction of ranavirus by moving between contaminated and uncontaminated sites. The environmental persistence of ranavirus in unsterile water and soil is probably at least 1 week (Nazir et al. 2012). Thus, recreationists that move among watersheds without decontaminating footwear or gear could be a major source of ranavirus introduction (Gray et al. 2009). Fish hatcheries are known sites of ranavirus outbreaks (Waltzek et al. 2014); thus, the release of clinically or subclinically infected fish or their effluent from the hatchery could be another major source of ranavirus introduction. For a particular area of interest, it is important to identify the most likely routes of introduction. It can be useful to divide routes of introduction into three stages: import, release, and exposure (Fig. 10). In the case of imported aquacultured fish, the following steps define the import stage:

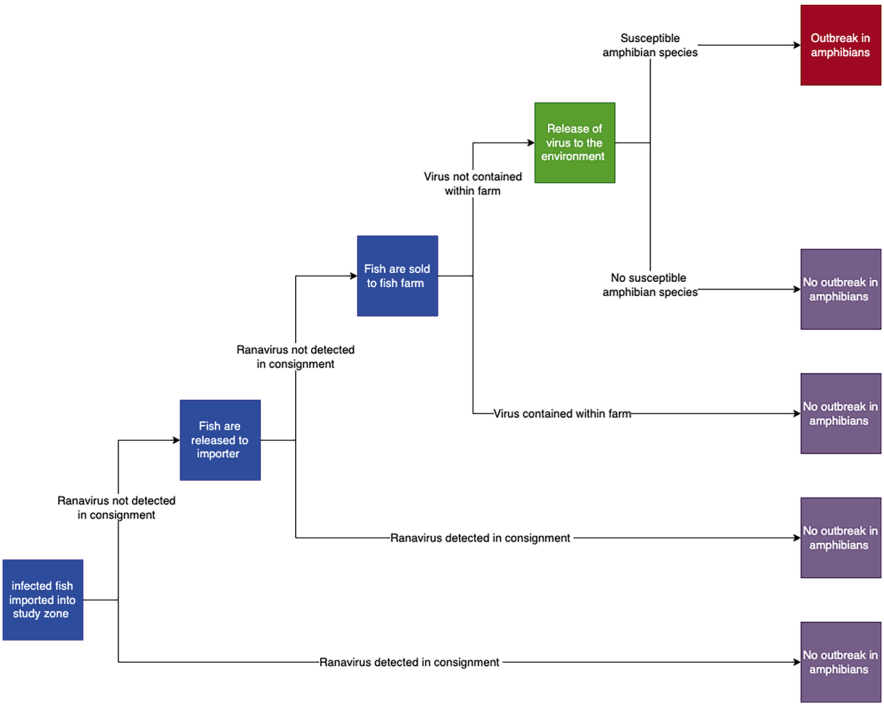

Fig. 10 Flow diagram for possible routes of transmission of ranavirus into a naïve susceptible population of amphibians in the wild. The routes of introduction are divided into three stages: import (blue), release (green), and exposure (red)

- Imported fish from an infected zone are infected with ranavirus.
- The infection passes undetected through border control.
- The infected fish are released to the retailer.
- The infected fish are sold to an aquaculture facility in the study zone.

Assuming that fish are contained in aquaculture ponds, ranavirus could be released into adjacent aquatic environments via several pathways:

- Virus contaminated effluent is released.
- Infected fish escape.
- Avian or mammalian predators could transport live or dead fish.
- Ranavirus hosts, such as amphibians or reptiles, could enter the pond, become infected, and disperse.
- Mechanical vectors, such as pets or humans, could transport the virus on fomites.

Finally, exposure to the virus could occur via several direct and indirect routes (Gray et al. 2009). Host species could be exposed to the virus in water, which is an efficient transmission medium, or the virus could be transmitted by direct contact or consumption of infected hosts (Miller et al. 2011). There is some evidence that

ranavirus transmission can be density independent, which can increase extinction probabilities (Brunner et al. 2024).

- *Consequence Assessment*

The outcome of ranavirus infection in a species can be described qualitatively or quantitatively in terms of direct or indirect consequences. Direct consequences are the effect that ranavirus has on the species of interest, which typically includes estimating the likelihood of population declines and extinction (Sect. 2). Highly susceptible species that are rare have the greatest probability of extinction (Earl and Gray 2014), especially if these species co-occur with other ranavirus hosts. Indirect consequences are costs associated with pathogen surveillance (i.e., field and diagnostic expenses) and possible repatriation of populations following extinction.

- *Risk Estimation*

The assumption is that the virus will travel along the routes identified from an infected animal to a susceptible animal. In cases where it is determined that the consequence of ranavirus introduction is unacceptable, a series of critical control points (CCPs) should be established along the routes of introduction identified above, where the virus could be intercepted and the transmission terminated. The probability of the infection passing unnoticed through a CCP is estimated for each CCP by addressing several questions. This process can be summarized in a scenario tree, where each CCP has a "yes" and a "no" branch, and a likelihood of detection is assigned (Fig. 11). In Fig. 11, CCP 1 and 4 are predetermined for each border control post, while CCP 3 will depend on the training and experience of the inspectors. CCP 2 can be affected by viral load, water temperature, and animal health.

Detecting a pathogen in a laboratory test in CCP 5 is a function of two processes: sample size (Sect. 1) and performance of molecular tests (i.e., the sensitivity and specificity of PCR, Miller et al. 2024). The sensitivity and specificity of PCR for ranavirus are an ongoing research direction (Miller et al. 2024) and can be affected by sample type (i.e., lethal vs. nonlethal collection, Gray et al. 2012). In general, it is believed that the liver and kidney tissue provide the most reliable estimate of detection followed by tail, toe clips, and blood (Miller et al. 2024). Assuming perfect sensitivity and specificity of PCR, the probability of detecting ranavirus is approximately 95% (Sect. 1). Risk of not detecting ranavirus in an imported consignment is calculated as $-1-$ the product of the detection probabilities at all CCPs (Fig. 11).

3.3 Risk Management and Communication

To manage the risk of ranavirus introduction, it is useful to perform a risk- consequence assessment. If risk is low but the consequence to the target species is high, risk management priority would be high. If, however, the risk of introducing ranavirus is high but the consequences are low, risk management priority would be low. If

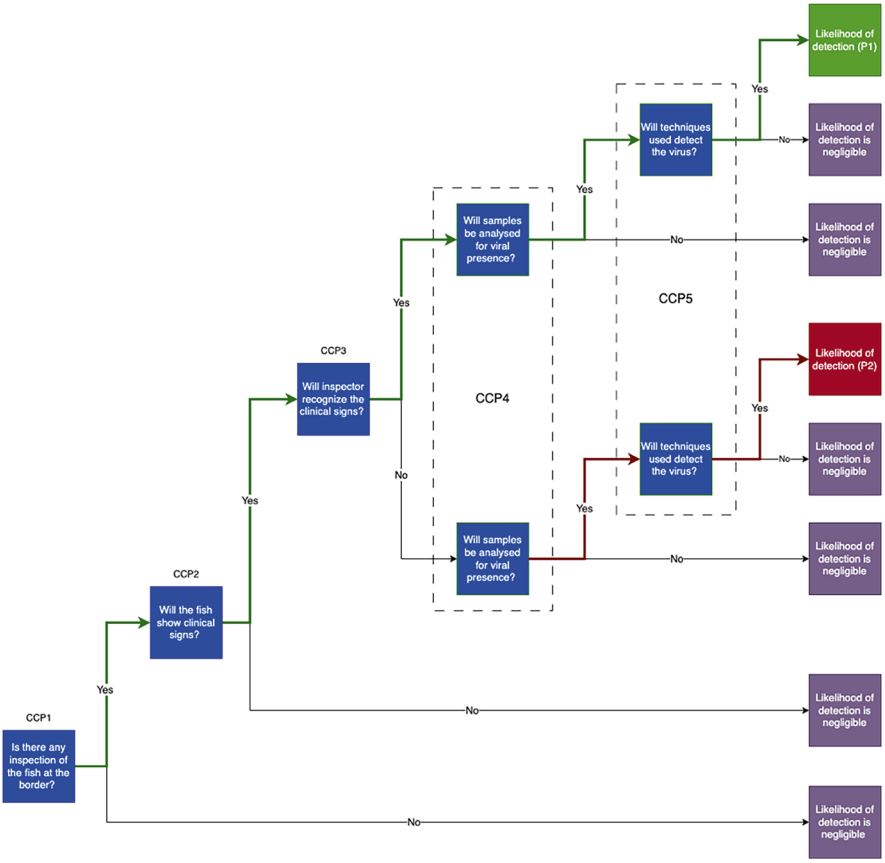

Fig. 11 Scenario tree for detection of ranavirus in a consignment of infected fish at the border inspection. Critical control points (CCP) 1–5 (blue) are opportunities identified where the virus could be detected and future transmission terminated. P1 (green) is the product of the "Yes" answer probabilities in the left branch of the tree. P2 (red) is the product of the two "yes" answer probabilities in the right branch of the scenario tree. The probability of ranavirus not being detected at the border is $-1 - (P1 + P2)$

the risk analysis indicates that the consequences are high, then the recommendations to management would focus on the CCPs and how to increase the likelihood of detecting and eliminating an infected consignment in a cost-effective manner.

Risk communication is an interactive and iterative process involving a two-way dialogue between decision-makers, experts, and stakeholders. Risk communication is used to gather information and opinions regarding hazards and risks and engage with relevant experts and stakeholders to maximize the quality of the analysis and the probability of recommendation implementation. It is essential for determining levels of acceptable risk. Risk communication is encouraged throughout risk analysis (Fig. 9) with stakeholders involved in the analysis from the outset. Jacobson (2009) is a useful resource of effective communication skills for conservation and

resource management problems. Several tools and templates for effective risk communication are available in the Manual of Procedures for Wildlife Disease Risk Analysis (Jakob-Hoff et al. 2014).

Effective communication is required among stakeholders, both when collecting information to feed into risk analyses and in terms of informing end users of the findings, management options, and their implementation. Risk communication is often centered at government level, but individual organizations such as fish farmers or herpetological societies can investigate and implement their own quarantine and surveillance guidelines with qualified diagnostic support. Cooperation and awareness at all levels will reduce the risk of introducing ranavirus into an uninfected area.

Many of the facts needed to carry out a comprehensive risk analysis may already be available in the published scientific literature and should be used to substantiate the recommendation. It is important to consider the applicability and quality of the published literature before it is used in risk analyses. Published data might be from a different species, time of year, or continent. If published data do not exist for your species or region, a pilot study can be performed to generate data. Alternatively, obtaining expert opinion following the Delphi method can be an approach to secure preliminary estimates for use in the risk analysis (Helmer 1967; Vose 2000). We recommend that all organizations that are interested in performing an IRA (or DRA) consult experts that study ranaviruses. The GRC is a collection of scientists, veterinarians, and practitioners that can provide guidance with setting up risk analyses (https://www.ranavirus.org/). Each continent has a regional GRC representative that can assist or make necessary connections with experts in your region.

Acknowledgments We graciously thank Mark Wilber and Joe Mihaljevic for their numerous helpful comments on an earlier draft, especially MW's corrections to precision of prevalence in the small populations (Sect.). MJG was partially supported by NSF Division of Environmental Biology grants 1814520 and 2207922. JLB was partially supported by NSF DEB grant 1754474. We thank the following organizations for providing funds to support Open Access publishing of the second edition: University of Tennessee's (UT) Open Publishing Support Fund, UT School of Natural Resources, UT Center for Wildlife Health, Washington State University Libraries, University of Mississippi Medical Center, Gordon State College, Association of Reptile and Amphibian Veterinarians, and the Global Ranavirus Consortium.

References

Abdalhamid B, Bilder CR, McCutchen EL, Hinrichs SH, Koepsell SA, Iwen PC (2020) Assessment of specimen pooling to conserve SARS CoV-2 testing resources. Am J Clin Pathol 153(6):715–718

Allen LJS (2006) An introduction to mathematical biology. Pearson, New York

Allender MC, Fry MM, Irizarry AR et al (2006) Intracytoplasmic inclusions in circulating leukocytes from an eastern box turtle (*Terrapene carolina carolina*) with iridoviral infection. J Wildl Dis 42:677–684

Augenblick N, Kolstad J, Obermeyer Z, Wang A (2022) Pooled testing efficiency increases with test frequency. Proceedings of the National Academy of Sciences of the United States of America 119(2). https://doi.org/10.1073/pnas.2105180119

Augusiak J, Van Den Brink PJ, Grimm V (2014) Merging validation and evaluation of ecological models to "evaludation": a review of terminology and a practical approach. Ecol Model 280:117–128

Ben-Ami F, Ebert D, Regoes RR (2010) Pathogen dose infectivity curves as a method to analyze the distribution of host susceptibility: a quantitative assessment of maternal effects after food stress and pathogen exposure. Am Nat 175:106–115

Berger JO (1985) Statistical decision theory and Bayesian analysis. Springer-Verlag, New York

Biek R, Funk WC, Maxwell BA et al (2002) What is missing in amphibian decline research: insights from ecological sensitivity analysis. Conserv Biol 16:728–734

Bilder CR, Tebbs JM, McMahan CS (2019) Informative group testing for multiplex assays. Biometrics 75(1):278–288. https://doi.org/10.1111/biom.12988

Blackwood JC, Childs LM (2018) An introduction to compartmental modeling for the budding infectious disease modeler. Lett Biomath 5(1):195–221

Brenes R (2013) Mechanisms contributing to the emergence of ranavirus in ectothermic vertebrate communities. PhD dissertation, University of Tennessee

Brenes R, Gray MJ, Waltzek TB et al (2014a) Transmission of ranavirus between ectothermic vertebrate hosts. PLoS One 9:e92476

Brenes R, Miller DL, Waltzek TB et al (2014b) Susceptibility of fish and turtles to three ranaviruses isolated from different ectothermic vertebrate classes. J Aquat Anim Health 26(2):118–126

Briggs CJ, Vredenburg VT, Knapp RA et al (2005) Investigating the population-level effects of chytridiomycosis: an emerging infectious disease of amphibians. Ecology 86:3149–3159

Briggs CJ, Knapp RA, Vredenburg VT (2010) Enzootic and epizootic dynamics of the chytrid fungal pathogen of amphibians. Proc Natl Acad Sci USA 107:9695–9700

Brückner G, MacDiarmid SC, Murray N, Berthe F, Müller-Graf C, Sugiura K, Zepeda C, Kahn S, Mylrea G (2010) Handbook on import risk analysis for animal and animal products. In: Introduction and qualitative risk analysis, vol I, 2nd edn. World Organisation for Animal Health, Paris

Brunner JL (2020) Pooled samples and eDNA-based detection can facilitate the "clean trade" of aquatic animals. Sci Rep 10(1):10280. https://doi.org/10.1038/s41598-020-66280-7

Brunner JL, Yarber CM (2018) Evaluating the importance of environmental persistence for *Ranavirus* transmission and epidemiology. Adv Virus Res 101:129–148

Brunner JL, Richards K, Collins JP (2005) Dose and host characteristics influence virulence of ranavirus infections. Oecologia 144:399–406

Brunner JL et al (2024) Ranavirus ecology: from individual infections to population epidemiology to community impacts. In: Ranaviruses: Emerging pathogens of ectothermic vertebrates. Springer, Cham

Cameron AR, Baldock FC (1998) A new probability formula for surveys to substantiate freedom from disease. Prev Vet Med 34:1–17

Campbell LJ, Garner TWJ, Tessa G et al (2018) An emerging viral pathogen truncates population age structure in a European amphibian and may reduce population viability. PeerJ 6:e5949

Canessa S, Guillera-Arroita G, Lahoz-Monfort JJ, Southwell DM, Armstrong DP, Chadès I, Lacy RC, Converse SJ (2015) When do we need more data? A primer on calculating the value of information for applied ecologists. Methods Ecol Evol 6(10):1219–1228. https://doi.org/10.1111/2041-210X.12423

Cariboni J, Gatelli D, Liska R et al (2007) The role of sensitivity analysis in ecological modelling. Ecol Model 203:167–182

Caswell H (2000) Matrix population models: construction, analysis, and interpretation, 2nd edn. Sinauer, Sunderland

Chen Z, Li T, Gao X, Wang C, Zang Q (2018) Protective immunity induced by DNA vaccination against ranavirus infection in Chinese Giant salamander *Andrias davidianus*. Viruses 10(2):52

Cooch EG, Conn PB, Ellner SP et al (2012) Disease dynamics in wild populations: modeling and estimation: a review. J Ornithol 152(2):485–509

Diekmann O, Heesterbeek JA, Metz JA (1990) On the definition and the computation of the basic reproduction ratio R 0 in models for infectious diseases in heterogeneous populations. J Math Biol 28:365–382

DiRenzo GV, Grant EHC (2019) Overview of emerging amphibian pathogens and modeling advances for conservation-related decisions. Biol Conserv 236:474–483

Dobson A (2004) Population dynamics of pathogens with multiple host species. Am Nat 164:S64–S78

Dohoo I, Martin W, Stryhn H (2003) Veterinary epidemiologic research. AVC, Charlottetown

Dorfman R (1943) The detection of defective members of large populations. Ann Math Stat 14:436–440

Duffus ALJ, Waltzek TB, Stöhr AC, Allender MC, Gotesman M, Whittington RJ, Hick P, Hines MK, Marschang RE (2015) Distribution and host range of ranaviruses. In: Gray MJ, Chinchar VG (eds) Ranaviruses: lethal pathogens of ectothermic vertebrates. Springer, New York

Duffus ALJ, Garner TWJ, Nichols RA et al (2019) Modelling ranavirus transmission in populations of common frogs (*Rana temporaria*) in the United Kingdom. Viruses 11:556

Earl JE (2019) Evaluating the assumptions of population projection models used for conservation. Biol Conserv 237:145–154

Earl JE, Gray MJ (2014) Introduction of ranavirus to isolated wood frog population could cause local extinction. EcoHealth 11:581–592

Earl JE, Chaney JC, Sutton WB et al (2016) Ranavirus could facilitate local extinction of rare amphibian species. Oecologia 182:611–623

Farrell AP, Collins JP, Greer AL, Thieme HR (2018) Times from infection to disease-induced death and their influence on final population sizes after epidemic outbreaks. Bull Math Biol 80:1937–1961

Federico P, Gross LJ, Lenhart S et al (2013) Optimal control in individual-based models: implications from aggregated methods. Am Nat 181:64–77

Fenichel EP, Horan RD, Hickling GJ (2010) Management of infectious wildlife diseases: bridging conventional and bioeconomic approaches. Ecol Appl 20:903–914

Fofana AM, Hurford A (2017) Mechanistic movement models to understand epidemic spread. Philos Trans R Soc B 372:20160086

Gray MJ, Miller DL, Hoverman JT (2009) Ecology and pathology of amphibian ranaviruses. Dis Aquat Org 87:243–266

Gray MJ, Miller DL, Hoverman JT (2012) Reliability of non-lethal surveillance methods for detecting ranavirus infection. Dis Aquat Org 99:1–6

Gray MJ, Spatz JA, Carter ED, Yarber CM, Wilkes RP, Miller DL (2018) Poor biosecurity could lead to disease outbreaks in animal populations. PLoS One 13(3):e0193243

Greer AL, Collins JP (2008) Habitat fragmentation as a result of biotic and abiotic factors controls pathogen transmission throughout a host population. J Anim Ecol 77:364–369

Grimm V, Railsback SF (2005) Individual-based modeling and ecology. Princeton University Press, Princeton

Grimm V, Railsback SF (2012) Pattern-oriented modelling: a "mulit-scope" for predictive systems ecology. Philos Trans R Soc Lond Ser B Biol Sci 367:298–310

Grimm V, Revilla E, Berger U et al (2005) Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310:987–991

Grimm V, Augusiak J, Focks A et al (2014) Towards better modelling and decision support: documenting model development, testing, and analysis using TRACE. Ecol Model 280:129–139

Haislip NA, Gray MJ, Hoverman JT et al (2011) Development and disease: how susceptibility to an emerging pathogen changes through anuran development. PLoS One 6:e22307

Halliday JE, Meredith AL, Knobel DL, Shaw DJ, Bronsvoort BM, Cleaveland S (2007) A framework for evaluating animals as sentinels for infectious disease surveillance. J R Soc Interface 4:973–984

Hanski I (1999) Metapopulation ecology. Oxford University Press, New York

Harp EM, Petranka JW (2006) Ranavirus in wood frogs (*Rana sylvatica*): potential sources of transmission within and between ponds. J Wildl Dis 42:307–318

Hastings A (1997) Population biology: concepts and models. Springer, New York

Helmer O (1967) Analysis of the future: the Delphi method. RAND, Santa Monica

Hens N, Shkedy Z, Aerts M, Faes C, Van Damme P, Beutels P (2012) In: Hens N, Shkedy Z, Aerts M, Faes C, Van Damme P, Beutels P (eds) Estimating the force of infection from incidence and prevalence modeling infectious disease parameters based on serological and social contact data: a modern statistical perspective. Springer, New York

Hess G (1996) Disease in metapopulation models: implications for conservation. Ecology 77:1617–1632

Hethcote HW (2000) The mathematics of infectious disease. SIAM Rev 42:599–653

Hitt B, Bilder C, Tebbs J, McMahan C (2019) The objective function controversy for group testing: much ado about nothing? Stat Med 38:4912–4923

Hogan CA, Sahoo MK, Pinsky BA (2020) Sample pooling as a strategy to detect community transmission of SARS-CoV-2. JAMA 323(19):1967–1969

Hoverman JT, Gray MJ, Haislip NA et al (2011) Phylogeny, life history, and ecology contribute to differences in amphibian susceptibility to ranaviruses. EcoHealth 8:301–319

Hoverman JT, Gray MJ, Miller DL et al (2012) Widespread occurrence of ranavirus in pond-breeding amphibian populations. EcoHealth 9:36–48

Jacobson SK (2009) Communication skills for conservation professionals, 2nd Ed. Island Press, Washington, District of Columbia

Jakob-Hoff RM, MacDiarmid SC, Lees C, Miller PS, Travis D, Kock R (2014) Manual of Procedures for Wildlife Disease Risk Analysis. World Organisation for Animal Health, Paris, 160 pp. Published in association with the International Union for Conservation of Nature and the Species Survival Commission

Johnson AJ, Pessier AP, Jacobson ER (2007) Experimental transmission and induction of ranaviral disease in western ornate box turtles (*Terrapene ornata ornata*) and red-eared sliders (*Trachemys scripta elegans*). Vet Pathol 44:285–297

Keesing F, Holt RD, Ostfeld RS (2006) Effects of species diversity on disease risk. Ecol Lett 9:485–498

Kimble SJA, Karna AK, Johnson AJ, Hoverman JT, Williams RN (2015) Mosquitoes as a potential vector of ranavirus transmission in terrestrial turtles. EcoHealth 12:334–338

Kolby JE, Smith KM, Berger L et al (2014) First evidence of amphibian chytrid fungus (*Batrachochytrium dendrobatidis*) and ranavirus in Hong Kong amphibian trade. PLoS One 9:e90750

Lélu M, Langlais M, Poulle M et al (2013) When should a trophically and vertically transmitted parasite manipulate its intermediate host? The case of *Toxoplasma gondii*. Proc Biol Sci 280:20131143

Lenhart S, Workman JT (2007) Optimal control applied to biological models. Chapman & Hall/CRC, Boca Raton

Lloyd-Smith JO, Cross PC, Briggs CJ et al (2005) Should we expect population thresholds for wildlife disease? Trends Ecol Evol 20:511–519

Malagon DA, Melara LA, Prosper OF, Lenhart S, Carter ED, Fordyce JA, Peterson AC, Miller DA, Gray MJ (2020) Host density and habitat structure influence host contact rates and *Batrachochytrium salamandrivorans* transmission. Sci Rep 10:5584

Marschang RE et al (2024) Ranavirus distribution and host range. In: Ranaviruses: Emerging pathogens of ectothermic vertebrates. Springer, Cham

Marsh DM, Trenham PC (2001) Metapopulation dynamics and amphibian conservation. Conserv Biol 15:40–49

McCallum H (2008) Host–pathogen and host–parasite models. In: Population parameters: estimation for ecological models. John Wiley & Sons, New York, pp 284–312

McCallum H, Barlow N, Hone J (2001) How should pathogen transmission be modeled. Trends Ecol Evol 16:295–300

Mideo N, Barclay VC, Chan BHK et al (2008) Understanding and predicting strain-specific patterns of pathogenesis in the rodent malaria, *Plasmodium chabaudi*. Am Nat 172:E214–E238

Mideo N, Savill NJ, Chadwick W et al (2011) Causes of variation in malaria infection dynamics: insights from theory and data. Am Nat 178:174–188

Mijaljevic JR, Greer AL, Brunner JL (2019) Evaluating the within-host dynamics of *Ranavirus* infection with mechanistic disease models and experimental data. Viruses 11:396

Miller DL, Gray MJ, Storfer A (2011) Ecopathology of ranaviruses infecting amphibians. Viruses 3:2351–2373
Miller DL et al (2024) Pathology and diagnostics. In: Ranaviruses: Emerging pathogens of ectothermic vertebrates. Springer, Cham
Mosher BA, Brand AB, Wiewel ANM, Miller DAW, Gray MJ, Miller DL, Grant EHC (2019) Estimating occurrence, prevalence, and detection of amphibian pathogens: insights from occupancy models. J Wildl Dis 53:563–575
Nazir J, Spengler M, Marschang RE (2012) Environmental persistence of amphibian and reptilian ranaviruses. Dis Aquat Org 98:177–184
OIE (2014) Aquatic animal health code (online access). Office International des Epizooties, Paris. Accessed 21 May 2014
Otto SP, Day T (2007) A biologist's guide to mathematical modeling in ecology and evolution. Princeton University Press, Princeton
Peace A, O'Regan SM, Spatz JA et al (2019) A highly invasive chimeric ranavirus can decimate tadpole populations rapidly through multiple transmission pathways. Ecol Model 410:108777
Raiffa H (1968) Decision analysis: introductory lectures on choices under uncertainty. Addison-Wesley, Oxford, England
Railsback SF, Grimm V (2011) Agent-based and individual-based modeling: a practical introduction. Princeton University Press, Princeton
Rödder D, Kielgast J, Bielby J et al (2009) Global amphibian extinction risk assessment for the panzootic chytrid fungus. Diversity 1:52–66
Rosa GM, Bosch J, Martel A et al (2019) Sex-biased disease dynamics increase extinction risk by impairing population recovery. Anim Conserv 22:579–588
Rosa GM, Botto GA, Mitra AT et al (2022) Invasive host-pathogen dynamics leading to amphibian declines. Bio Conserv 276:109785
Sabino-Pinto J, Krause ET, Bletz MC, Martel A, Pasmans F, Steinfartz S, Vences M (2019) Detectability vs. Time and costs in pooled DNA extraction of cutaneous swabs: a study on the amphibian chytrid fungi. Amphib Reptil 40(1):29–39. https://doi.org/10.1163/15685381-20181011
Schlaifer R, Raiffa H (1961) Applied statistical decision theory. ClintonPress Inc, Boston
Schloegel LM, Ferreira CM, James TY et al (2010) The North American bullfrog as a reservoir for the spread of *Batrachochytrium dendrobatidis* in Brazil. Anim Conserv 14:53–61
Smith MA, Green DM (2005) Dispersal and the metapopulation paradigm in amphibian ecology and conservation: are all amphibian populations metapopulations? Ecography 28:110–128
Smith KF, Behrens M, Schoegel LM et al (2009) Reducing the risk of wildlife trade. Science 324:594–595
St-Amour V, Lesbarrères D (2007) Detecting ranavirus in toe clips: an alternative to lethal sampling methods. Conserv Genet 8:1247–1250
Todd-Thompson M (2010) Seasonality, variation in species prevalence, and localized disease for ranavirus in cades cove (Great Smoky Mountains National Park) amphibians. M.S., University of Tennessee. http://trace.tennessee.edu/utk_gradthes/665/. Accessed 21 May 2014
Travis DA, Smith K (2019) Risk analysis framework guidance for wildlife health professionals. In: Fowler's zoo and wild animal. Elsevier, St. Louis
Van den Driessche P (2017) Reproduction numbers of infectious disease models. Infect Dis Model 2:288–303
Vose D (2000) Risk analysis—a quantitative guide, 2nd edn. Wiley, Chichester
Waltzek TB, Miller DL, Gray MJ et al (2014) New disease records for hatchery-reared sturgeon. I. Expansion of frog virus 3 into *Scaphirhynchus albus*. Dis Aquat Org 111:219–227
Warne RW, Crespi EJ, Brunner JL (2011) Escape from the pond: stress and developmental responses to ranavirus infection in wood frog tadpoles. Funct Ecol 25:139–146
Wells KD (2007) The ecology and behavior of amphibians. University of Chicago Press, Chicago
Wiegand T, Jeltsch F, Hanski I et al (2003) Using pattern-oriented modeling for revealing hidden information: a key for reconciling ecological theory and application. Oikos 100:209–222

Williams BK, Johnson FA (2015) Value of information and natural resources decision-making. Wildl Soc Bull 39(3):488–496. https://doi.org/10.1002/wsb.575

Wilson EB (1927) Probable inference, the law of succession, and statistical inference. J Am Stat Assoc 22:209–212

WOAH (2021) Aquatic animal health code (online access). Office International des Epizooties, Paris. https://www.woah.org/en/what-we-do/standards/codes-and-manuals/aquatic-code-online-access

Wobeser GA (2006) Essentials of disease in wild animals. Blackwell, Ames

Woodhams DC, Alford RA, Briggs CJ et al (2008) Life-history trade-offs influence disease in changing climates: strategies of an amphibian pathogen. Ecology 89:1627–1639

Woodhams DC, Bosch J, Briggs CJ et al (2011) Mitigating amphibian disease: strategies to maintain wild populations and control chytridiomycosis. Front Zool 8:8

World Organisation for Animal Health (WOAH) (2021) Manual of diagnostic tests for aquatic animals. OIE, Paris. Chapter 2.1.2

Zuur AF, Savaliev AA, Ieno EN (2012) Zero inflated models and generalized linear mixed models with R. Highland Statistics, Newburgh