THE PROBABILITY OF DETECTING INFECTION IN FINITE POPULATIONS WITH INDIVIDUAL SAMPLES, POOLED SAMPLES, & ENVIRONMENTAL DNA

JESSE BRUNNER
SCHOOL OF BIOLOGICAL SCIENCES
WASHINGTON STATE UNIVERSITY
THE NEEDS TO DETECT PATHOGENS IN TRADE

Numerous important pathogens moved through trade

- **Viruses** (e.g., Infectious haematopoietic necrosis virus)
- **Bacteria** (e.g., *Aerococcus viridans*)
- **Fungi** (e.g., *Batrachochytrium dendrobatidis*; chytridiomycosis)
- **Metazoans** (*Myxobolus cerebralis*; whirling disease)
THE NEEDS TO DETECT PATHOGENS IN TRADE

Ranaviruses moved and spillover from live animal trade

- American bullfrogs for food, etc.
  - within US (via eBay)       Brunner et al. (in review) Dis Aquat Org
- Xenopus laevis in research   Robert et al. (2007) J Wildl Dis 43:645-652
- International pet trade (Hong Kong) Kolby et al. (2014) PLoS One 9:e90750
PROBLEM 1: THE MAGNITUDE

• Immense numbers of animals are involved
• $2.7 \times 10^7$ live amphibians imported into the U.S.A. (2006–2014)
• Individual shipments may include 100’s – 1000’s of animals
  Wombwell et al. (2016) Ecohealth 13:456-466
• Sampling a fraction of animals involved is implausible
PROBLEM 1: THE MAGNITUDE

- Immense numbers of animals are involved
  - $2.7 \times 10^7$ live amphibians imported into the U.S.A. (2006–2014)
  - Individual shipments may include 100’s – 1000’s of animals
    Wombwell et al. (2016) Ecohealth 13:456-466

- Sampling a fraction of animals involved is implausible
  - pool samples (e.g., swabs)
  - environmental DNA
PROBLEM 2: DIAGNOSTICS AND INFERENCE

Need to:

• determine probability of finding $x$ infections with sample size, $n$

• incorporate diagnostic performance (sensitivity & specificity)

<table>
<thead>
<tr>
<th></th>
<th>Individual samples</th>
<th>Pooled samples</th>
<th>eDNA in finite population</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perfect test</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infinite population (sampling w/ replacement)</td>
<td><strong>Binomial</strong> C&amp;B</td>
<td><strong>Pooled Hypergeometric</strong> T&amp;D</td>
<td>?</td>
</tr>
<tr>
<td>Finite population (sampling w/o replacement)</td>
<td><strong>Hypergeometric</strong> C&amp;B</td>
<td></td>
<td>?</td>
</tr>
<tr>
<td><strong>Imperfect test</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual samples</td>
<td><strong>Binomial</strong> C&amp;B</td>
<td><strong>Hypergeometric</strong> C&amp;B</td>
<td>?</td>
</tr>
<tr>
<td>Pooled samples</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

PROBLEM 2: DIAGNOSTICS AND INFERENCE

Need to:

- determine probability of finding $x$ infections with sample size, $n$
- incorporate diagnostic performance (sensitivity & specificity)

<table>
<thead>
<tr>
<th></th>
<th>Individual samples</th>
<th>Pooled samples</th>
<th>eDNA in finite population</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perfect test</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Binomial</td>
<td>Binomial</td>
<td>Pooled</td>
<td></td>
</tr>
<tr>
<td>C&amp;B</td>
<td>Hypergeometric</td>
<td>Hypergeometric</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>C&amp;B</td>
<td>C&amp;B</td>
<td>?</td>
</tr>
<tr>
<td><strong>Imperfect test</strong></td>
<td>Binomial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Binomial</td>
<td>Hypergeometric</td>
<td>Pooled</td>
<td></td>
</tr>
<tr>
<td>C&amp;B</td>
<td>Hypergeometric</td>
<td>C&amp;B</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>?</td>
</tr>
</tbody>
</table>

INDIVIDUAL SAMPLES

HYPERGEOMETRIC (C&B)

Probability of $x$ samples testing positive

$$P(T^+ = x) = \sum_{y=0}^{d} \frac{\binom{d}{y} \binom{N-d}{n-y}}{\binom{N}{n}} \times \sum_{j=0}^{\min(x,y)} \binom{y}{j} S_e^j (1-S_e)^{y-j} \times \binom{n-y}{x-j} (1-S_p)^{x-j} S_p^{n-x-y+j}$$

Probability of $y$ of $d$ diseased individuals in sample of size $n$

Takes care of false positives & false negatives
POOLED INDIVIDUAL SAMPLES

POOLED HYPERGEOMETRIC (MODIFIED FROM T&D)

**Probability of** $x$ **groups testing positive**

$$P(GT^+ = x) = \sum_{y=0}^{\min(m,d)} \left( \binom{m}{y} \sum_{i=0}^{y} (-1)^i \binom{y}{i} \frac{N - d}{(m - y + i)k} \frac{N}{(m - y + i)k} \right)$$

**Probability** $y$ **of** $m$ **groups include $\geq 1$ diseased individuals**

$$\times \sum_{j=0}^{\min(x,y)} \left( \binom{y}{j} Se^j (1 - Se)^{y-j} \times \binom{m - y}{x-j} (1 - Sp)^{x-j} Sp^{m-x-y+j} \right)$$

**Takes care of false positives & false negatives**
POOLED INDIVIDUAL SAMPLES

POOLED HYPERGEOMETRIC (MODIFIED FROM T&D)

Probability of $x$ groups testing positive

$$P(GT^+ = x) = \sum_{y=0}^{\min(m,d)} \binom{m}{y} \sum_{i=0}^{y} (-1)^i \binom{y}{i} \frac{N-d}{(m-y+i)k} \frac{N}{(m-y+i)k}$$

Probability $y$ of $m$ groups include $\geq 1$ diseased individuals

$$\times \sum_{j=0}^{\min(x,y)} \binom{y}{j} Se^j (1-Se)^{y-j} \times \binom{m-y}{x-j} (1-Sp)^{x-j} Sp^{m-x-y+j}$$

n SAMPLES DIVIDED INTO $m$ GROUPS OF SIZE $k=m/n$

ASSUMES $Se$ AND $Sp$ NOT AFFECTED BY GROUP SIZE ($k$)

- $d=1$ is not swamped in $k$ samples
- $Se$ is same with $d=1$ or $d=k$
- false positives independent of $k$

POOLED HYPERGEOMETRIC (MODIFIED FROM T&D)

Probability of $x$ groups testing positive

$$P(GT^+ = x) = \sum_{y=0}^{\min(m,d)} \binom{m}{y} \sum_{i=0}^{y} (-1)^i \binom{y}{i} \frac{N-d}{(m-y+i)k} \frac{N}{(m-y+i)k}$$

Probability $y$ of $m$ groups include $\geq 1$ diseased individuals

$$\times \sum_{j=0}^{\min(x,y)} \binom{y}{j} Se^j (1-Se)^{y-j} \times \binom{m-y}{x-j} (1-Sp)^{x-j} Sp^{m-x-y+j}$$

Takes care of false positives & false negatives
POOLED INDIVIDUAL SAMPLES

POOLED HYPERGEOMETRIC (MODIFIED FROM T&D)

Probability of $x$ groups testing positive

$$P(GT^+ = x) = \sum_{y=0}^{\min(m,d)} \binom{m}{y} \sum_{i=0}^{y} (-1)^i \binom{y}{i} \frac{(N-d)}{(m-y+i)k} \frac{N}{(m-y+i)k}$$

Probability of $y$ of $m$ groups include $\geq 1$ diseased individuals

$$\times \sum_{j=0}^{\min(x,y)} \binom{y}{j} Se^j(1-Se)^{y-j} \times \binom{m-y}{x-j} (1-Sp)^{x-j} Sp^{m-x-y+j}$$

$\frac{n}{n}$ SAMPLES DIVIDED INTO $m$ GROUPS OF SIZE $k=m/n$

ASSUMES $Se$ AND $Sp$ NOT AFFECTED BY GROUP SIZE ($k$)

$\Rightarrow d=1$ is not swamped in $k$ samples

$\Rightarrow Se$ is same with $d=1$ or $d=k$

$\Rightarrow$ false positives independent of $k$

Takes care of false positives & false negatives
ENVIRONMENTAL DNA
BINOMIAL-ISH

Probability of $x$ eDNA samples testing positive

$P(eDNA^+ = x) = \binom{n}{x} \left[ 1 - (1 - Se)^d + (1 - Se)^d(1 - Sp) \right]^x \left[ (1 - Se)^d Sp \right]^{n-x}$

ASSUMES
- Target eDNA is well-mixed
- $Se$ is independent of $N$
- $Se$ defined so it is independent of $d$
**ENVIRONMENTAL DNA**

**BINOMIAL-ISH**

Probability of $x$ eDNA samples testing positive

$$P(eDNA^+ = x) = \binom{n}{x} \left[ (1 - (1 - Se)^d)^{x} \right] \left[ (1 - S_p) + (1 - Se)^d(1 - S_p) \right]^{n-x}$$

**ASSUMES**

- Target eDNA is well-mixed
- $Se$ is independent of $N$
- $Se$ defined so it is independent of $d$

**Could be enforced (e.g., homogenized)**
**ENVIRONMENTAL DNA**

**BINOMIAL-ISH**

Probability of \( x \) eDNA samples testing positive

\[
P(\text{eDNA}^+ = x) = \binom{n}{x} \left[ 1 - (1 - Se)^d \right] + (1 - Se)^d (1 - Sp)^x \left[ 1 - Se \right]^{n-x}
\]

**ASSUMES**
- Target eDNA is well-mixed
- \( Se \) is independent of \( N \)
- \( Se \) defined so it is independent of \( d \)

Could be enforced (e.g., homogenized)

Target not swamped by other DNA

True positives

False positives

False negatives

True negatives
COMPARING PERFORMANCE

Number of samples screened

Probability of detecting at least one infection

N = 10
N = 50
N = 100

ρ = 1
ρ = 2
ρ = 5

Sensitivity

- 1
- 0.75

Individual
COMPARING PERFORMANCE

ROUGHLY 1/k AS MANY GROUPS NEED TO BE SCREENED
BUT STILL NEED TO COLLECT k×n SAMPLES

Sensitivity

- Pooled (k=3)
- Individual

Number of samples screened

Probability of detecting at least one infection
COMPARING PERFORMANCE

eDNA IS NOT AFFECTED BY POPULATION SIZE

EVEN A VERY POOR eDNA TEST CAN BE MORE EFFICIENT THAN HIGHLY SENSITIVE INDIVIDUAL-BASED TEST

Sensitivity

1
0.75
0.5
0.25
DIGRESSION ON QUANTITATIVE NATURE OF DETECTION PROBABILITY

Probability of positive result can be described by a “single-hit” model:

\[ P(T^+) = 1 - e^{-\phi C} \]

C influenced by:

- infection (intensity, shedding amount, etc.)
- process of collecting (e.g., site, method) & processing samples (e.g., extraction efficiency, volume in reaction)
Simple model of $C$:

$$\frac{dC}{dt} = \psi d\alpha - \delta C$$

- # infected shedding
- proportion that ends up in eDNA sample
- degradation
DIGRESSION ON QUANTITATIVE NATURE OF DETECTION PROBABILITY

Simple model of $C$:

$$\frac{dC}{dt} = \psi d\alpha - \delta C$$

- # infected shedding
- proportion that ends up in eDNA sample
- degradation

![Graph showing the relationship between $\phi C$ and the probability of a positive test. The graph includes a curve that rises from left to right, indicating an increase in probability with increasing $\phi C$.](Graph.png)
SCREENING SHIPMENTS INTO U.S.A.

HYPOTHETICAL EXAMPLE

- Used $Se$ from study with American bullfrog tadpoles
- Assume
  - $3 \times 10^6$ individuals/yr
  - $N=100$ per shipment (i.e., bag)
  ----> $3 \times 10^4$ shipments
- Want to detect $d=1$

<table>
<thead>
<tr>
<th></th>
<th>Individual ($Se=0.831$)</th>
<th>Pooled, $k=3$ ($Se=0.831$)</th>
<th>eDNA ($Se_{effect}=0.13$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥80% Confidence</td>
<td>$96 \times 3 \times 10^4$ = $2,880,000$</td>
<td>$32 \times 3 \times 10^4$ = $960,000$</td>
<td>$12 \times 3 \times 10^4$ = $360,000$</td>
</tr>
</tbody>
</table>

Brunner et al. (accepted) Dis Aquat Org
HYPOTHETICAL EXAMPLE
SCREENING SHIPMENTS INTO U.S.A.

- Used $Se$ from study with American bullfrog tadpoles
- Assume
  - $3 \times 10^6$ individuals/yr
  - $N=100$ per shipment (i.e., bag)
  $\Rightarrow 3 \times 10^4$ shipments
- Want to detect $d=1$

<table>
<thead>
<tr>
<th></th>
<th>Individual ($Se=0.831$)</th>
<th>Pooled, $k=3$ ($Se=0.831$)</th>
<th>eDNA ($Se_{effect}=0.13$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥80% Confidence</td>
<td>$96 \times 3 \times 10^4 = 2,880,000$</td>
<td>$32 \times 3 \times 10^4 = 960,000$</td>
<td>$12 \times 3 \times 10^4 = 360,000$</td>
</tr>
</tbody>
</table>

Brunner et al. (accepted) Dis Aquat Org

CONSERVATIVE:
- Scaled to effective $Se$ $= Se \times 0.25L/5L \times 3d$

Could improve $Se$ by:
- filter more
- hold in water longer
SUMMARY

• Have theoretical basis to
  • evaluate & design sampling schemes
  • interpret test results
• By side-stepping problem of including infecteds in samples, eDNA is often very powerful
• Best approach depends on details (e.g., volume, time, N)
• Need to test in real-world settings

This work supported by:
Association of Zoo & Aquariums
American Association of Zoo Veterinarians
Morris Animal Foundation

MHMOLINARO/FLICKR (CC BY-NC-ND 2.0)