Hypothesis testing and infection prevalence

Ross A. Alford, School of Marine and Tropical Biology, James Cook University, Townsville, Queensland 4811, Australia. Email ross.alford@jcu.edu.au

Presented at the 2004 meeting of the Research and Analysis Network for Neotropical Amphibians (RANA)

Updated September 2007

Where did Bd come from?

- **Two mutually exclusive hypotheses**
  - **The disease outbreak hypothesis**
  - **The novel pathogen hypothesis**

Ross A. Alford, ross.alford@jcu.edu.au

The disease outbreak hypothesis

- **Pathogen *is endemic* to the regions in which declines occurred**
- **Pathogen has undergone epidemic outbreaks, either**
  - **Natural or**
  - **Anthropogenic**
- **Transmission into new host species may have occurred as part of these outbreaks**

Ross A. Alford, ross.alford@jcu.edu.au

The novel pathogen hypothesis

- **Pathogen *is not endemic* to the regions in which declines occurred**
  - **Has not previously been encountered by any hosts in those regions**
  - **Has overwhelmed naïve host populations, causing epidemics**

Ross A. Alford, ross.alford@jcu.edu.au

Why these two hypotheses?

- **They correspond to the thinking of many biologists**
- **They form a genuine dichotomy with important implications**

Why care which is correct? Research priorities depend on it

Disease outbreak

- Must determine
  - what normal host-parasite relationships are
  - how they have been disrupted, e.g., ecological change

Novel pathogen

- Must determine
  - how new areas are invaded
  - how to slow or prevent these invasions
  - whether species can acquire resistance

Getting these priorities right may make life or death difference to other species

Ross A. Alford, ross.alford@jcu.edu.au

Discriminating between (testing) hypotheses

- **Normally involves two mutually exclusive hypotheses**
  - **One designated the null, H_0, the other the alternative, H_a**
- ***Hypotheses are tested by trying to disprove H_0***
- **Rejecting H_0 suggests (but does not prove) that H_a is correct**
- **Failing to reject H_0 suggests (but does not prove) that H_0 is correct**
  - **Failing to reject is stronger than succeeding at confirming**
  - **Reasons:**
    - **Possible bias**
    - **Statistical power**

Confirming *versus* testing the novel pathogen hypothesis

Confirming

- **Demonstrate that Bd is absent from all populations before declines**

Testing

- **Try to show that Bd was present long before declines took place in a region**

Which is more achievable?

Confirmation *versus* refutation: studies of prevalence

- **Much work to date has attempted to**
  - **Estimate infection rates when Bd present**
  - **Show that Bd was absent from regions before particular dates**
    - **i.e., to *Confirm* the novel pathogen hypothesis**
    - **Usual approach:**
      - **Examine frogs near dates of declines**
        - » **Confirm that Bd is present**
      - **Examine frogs from dates just before declines**
      - **Work backward until Bd not found**
        - » ***Confirm* date of invasion**
        - » **Assumption: failure to find Bd shows that it was absent**

Confirmation *versus* refutation: studies of prevalence

- **To show that *Bd* was absent from regions before particular dates**
  - **Test H_0: Each population in a region before some date has P(infection) = 0**

Ross A. Alford, ross.alford@jcu.edu.au

Demonstrating zero prevalence

- Test H_0: a population has P(infection) = 0
  - Examine frogs
  - Reject H_0 whenever an infected individual is found
- How many individuals need to be tested?
  - *All* individuals in population
  - This seems slightly impractical
    - Need to be able to test only a sample

Statistics of infection rates: binomial processes

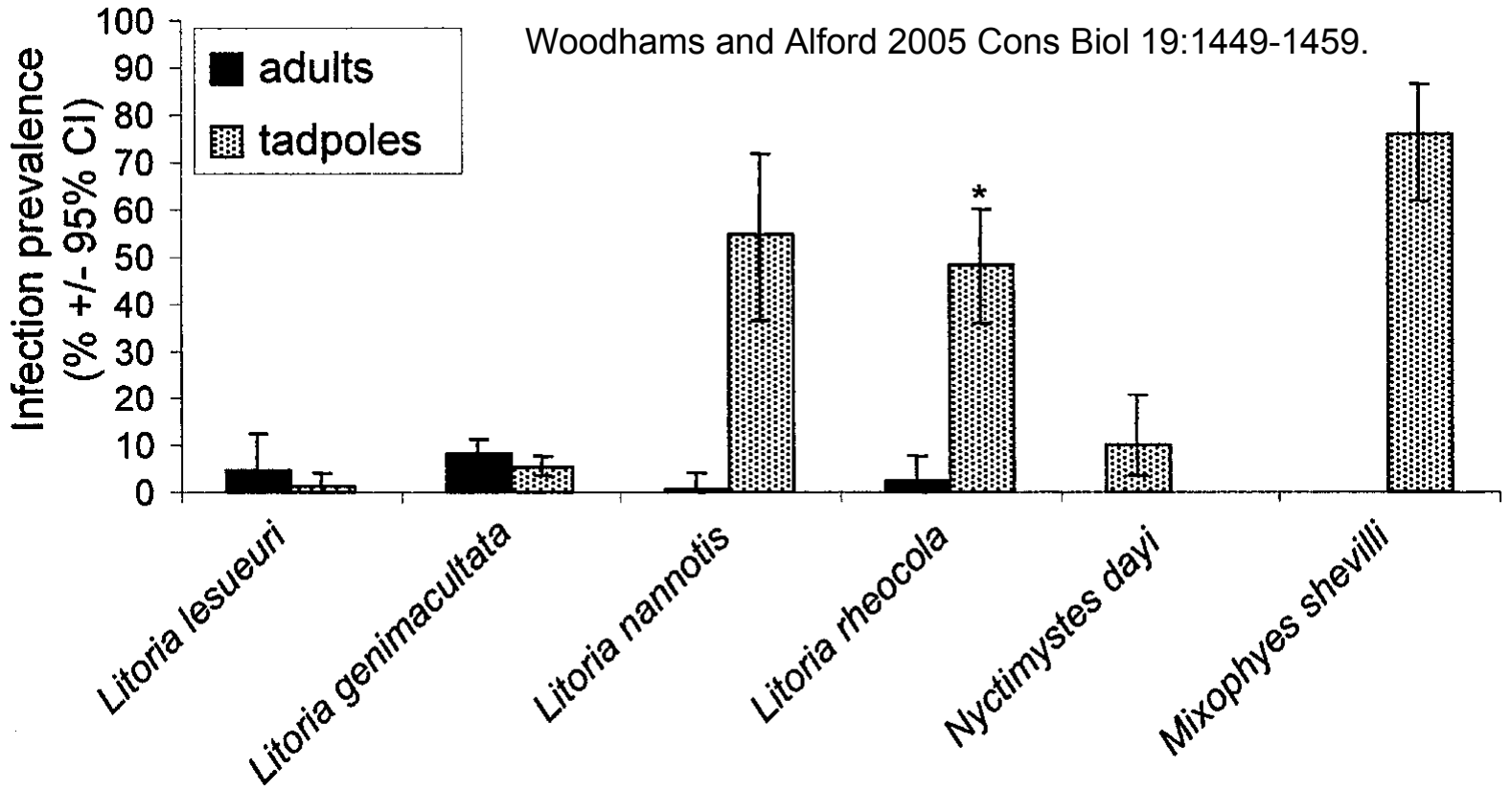
- **Testing for infection (in an ideal world, no false negatives or positives) has 2 possible outcomes**
  - **infected or uninfected**
- **Any population of animals at any moment has a true population prevalence or infection rate**
- **A series of determinations of individual infection status is a sample of observations on the population infection rate**
  - **Samples can be used to estimate population rates**
  - **larger samples give better estimates**

Infection rate of a population--2nd approach

- **Refine the hypothesis**
  - **Choose an infection rate that is *acceptably close* to 0**
  - **Populations shown to be below this boundary considered “uninfected”**
  - **New H_0: P(infection) > boundary**

Infection rate of a population --2nd approach

- New H_0: P(infection) > boundary
  - Many known infected populations have prevalences between 1% and 10% as determined by histology



Ross A. Alford, ross.alford@jcu.edu.au

Demonstrating "zero" prevalence --2nd and more practical approach

- New H_0: P(infection) > some boundary
  - Many known infected populations have prevalences between 1% and 10% as determined by histology
    - To demonstrate absence, *need to show that prevalence is much lower than this*
    - 1% or 0.1% seem like the highest reasonable boundaries for demonstrating absence, 0.1% is more reasonable
- New approach:
  - Collect sample large enough to ensure that upper 95% confidence limit for prevalence is at or below the chosen boundary, ideally 0.1% or less.
  - Then declare population *Bd*-free

Estimating binomial confidence limits

- **Formulas in Zar, JH, *Biostatistical Analysis, fourth edition*. Prentice-hall International, New Jersey, USA. ISBN 0-13-082390-2, Pp. 527-529, and from many other sources**
- **Implemented in spreadsheet available for download**

Ross A. Alford, ross.alford@jcu.edu.au

Examples of upper binomial confidence limits

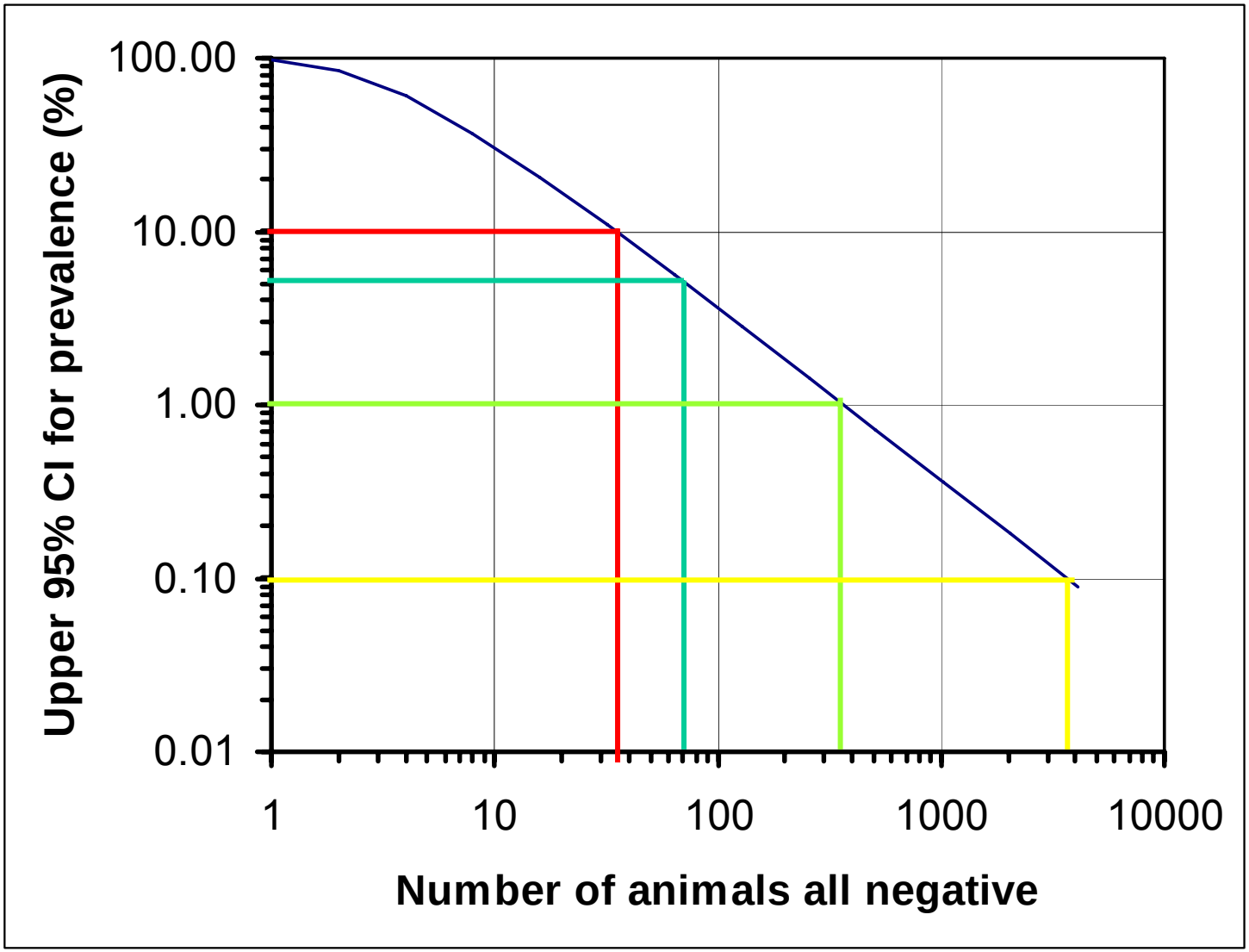
For samples of various numbers of animals, in which none are found to be infected, what is the upper 95% confidence limit for true prevalence in the sampled population?

Sample size	Upper 95% binomial confidence limit	
	As a proportion	As a percentage
1	0.9750	97.50
2	0.8419	84.19
4	0.6024	60.24
8	0.3694	36.94
16	0.2059	20.59
32	0.1089	10.89
64	0.0560	5.60
128	0.0284	2.84
256	0.0143	1.43
512	0.0072	0.72
1024	0.0036	0.36
2048	0.0018	0.18
4096	0.0009	0.09

2-tailed upper confidence limits of true prevalence when no infected animals detected

Upper 95% limit of:

- **10% -- requires sample size of 35**
- **5% -- requires 72**
- **1% -- requires 368**
- **0.1% -- requires 3686**
- **0.01% -- requires 36860, etc.**



Demonstrating zero infection rates requires very large samples

- **From 368 (1%) to 3685 (0.1%) animals must be tested to establish that infection rate is less than reasonable upper boundaries**
  - **Even if we chose 1% as a reasonable upper boundary, 368 animals would have to be tested**

- **Such large samples from any one population are very uncommon**
  - ***Bd* absence can be demonstrated in very few populations**

- **What about merging many small samples across populations?**

Ross A. Alford, ross.alford@jcu.edu.au

What I mean by “populations”

- **Samples or specimens of the *same* species, collected close enough together in time so that they are likely to come from a population with a constant prevalence**
- **Trying to estimate prevalence for anything except a population as defined above is meaningless**
  - **Species are known to vary in prevalence even when exposed to common environment, due to differences in susceptibility to infection**
    - **A “prevalence” estimated across such species reflects the species composition of the sample as much as the actual prevalence of the disease organism**
  - **Prevalence varies through time, within species and areas, so estimating “prevalence” across time suffers from the same problem**

Is it possible to combine samples?

- **Merging samples would allow a test of:**
  - **H_0: the mean infection rate across all of the tested populations is > boundary**

Ross A. Alford, ross.alford@jcu.edu.au

Is it possible to combine samples?

- **Merging samples would allow a test of:**
  - **H_0: the mean prevalence across all of the tested populations is > boundary**
    - **There is are 2 *big* problems with this:**
    - **As already mentioned, mean "prevalence" is a determined by the composition of the sample, not the actual dynamics of the host and pathogen**
    - **Also, rejecting this H_0 *does not mean that no populations have prevalences above the boundary,* only that *the average across all populations* is not above it.**
      - **Some of the populations in the sample could have infection rates above or even far above the boundary**
    - **So, this H_0 is not actually very interesting**

- **It is probably better to combine the probabilities, instead of the samples**

The combined probability test

- **Combining probabilities: use the combined probability test (useful for many other things, too)**
- **Reference, Sokal and Rohlf, *Biometrics***
- **Produces chi-squared test statistic:**

$$X^2 = -2\sum_{i=1}^{k} \ln(P_i)$$

- **Where**
  - **H_0: Across all sampled populations, tested null hypothesis is true**
  - ***k* = the number of independent tests of the same hypothesis**
  - **P_i = the probability of the outcome of each test**
  - **The test statistic is distributed as chi-squared with 2*k* degrees of freedom**

Ross A. Alford, ross.alford@jcu.edu.au

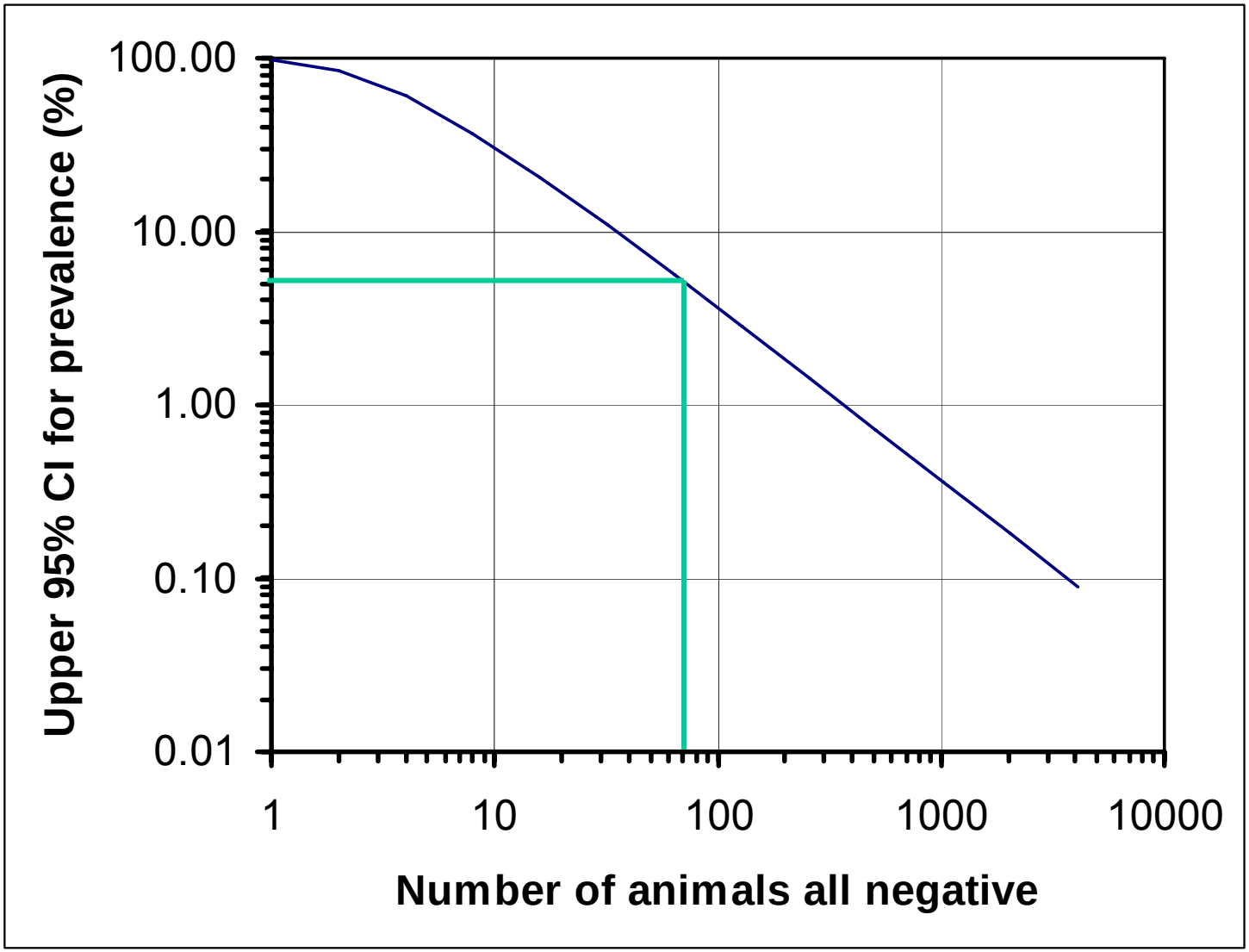
Combined probability test example

- **Seven samples from the same species, different populations or times or both**
  - **Samples contain 12, 5, 22, 8, 3, 17, and 5 individuals, total 72**
  - **All not infected**
  - **What not to do—consider all as one sample**
    - **Test H_0: mean prevalence across samples is less than 5%**
    - **Examine upper 95% CI for prevalence if 72 individuals are all negative**

2-tailed upper confidence limits of true prevalence when no infected animals detected

Upper 95% limit of:

- **5% -- requires 72**



Combined probability test example

- **Seven samples from the same species, different populations or times or both**
  - **Samples contain 12, 5, 22, 8, 3, 17, and 5 individuals, total 72**
  - **All not infected**
  - **What *not to do*—consider all as one sample**
    - **Test H_0: mean prevalence across samples is less than 5%**
    - **Examine upper 95% CI for prevalence if 72 individuals are all negative**
      - **upper 95% CI for prevalence is less than 5%, so you would reject H_0: mean prevalence $\geq$ 0.05**
    - ***This means nothing*, since it depends entirely on prevalences in individual populations and proportion of the sample made up by those populations**
    - **we are really interested in H_0: prevalences *in all sampled populations* were less than some boundary**

Ross A. Alford, ross.alford@jcu.edu.au

Combined probability test example

- **7 populations sampled, 72 total individuals negative**
- **Combined probability tests for three null hypotheses**
  - **True P in all sampled populations less than or equal to 0.1%, 1%, 5%**

Total sampled =	72					
Sample size	for true prev ≤ 0.001		for true prev ≤ 0.01		for true prev ≤ 0.05	
	P(0 positives)	ln(P)	P(0 positives)	ln(P)	P(0 positives)	ln(P)
12	0.988066	-0.01201	0.886385	-0.1206	0.54036	-0.61552
5	0.99501	-0.005	0.95099	-0.05025	0.773781	-0.25647
22	0.978229	-0.02201	0.801631	-0.22111	0.323534	-1.12845
8	0.992028	-0.008	0.922745	-0.0804	0.66342	-0.41035
3	0.997003	-0.003	0.970299	-0.03015	0.857375	-0.15388
17	0.983135	-0.01701	0.842943	-0.17086	0.41812	-0.87199
5	0.99501	-0.005	0.95099	-0.05025	0.773781	-0.25647
chi-squared=-2*sum of ln(P)		0.144072		1.447248		7.386234
degrees of freedom		7		7		7
P(chi-squared)		0.999992		0.98408		0.389802

∴ We do not even come close to rejecting H_0: prevalence in all populations is less than 5%

Combined probability test example 2

- **Add 5 more small samples**
- **12 samples, total is now 100 individuals, all negative**
- **Overall P has *increased,* we are further from rejecting H_0**
- **Small samples are really not even useful**

Total sampled =	100					
Sample size	for true prev ≤ 0.001		for true prev ≤ 0.01		for true prev ≤ 0.05	
	P(0 positives)	ln(P)	P(0 positives)	ln(P)	P(0 positives)	ln(P)
12	0.988066	-0.01201	0.886385	-0.1206	0.54036	-0.61552
5	0.99501	-0.005	0.95099	-0.05025	0.773781	-0.25647
22	0.978229	-0.02201	0.801631	-0.22111	0.323534	-1.12845
8	0.992028	-0.008	0.922745	-0.0804	0.66342	-0.41035
3	0.997003	-0.003	0.970299	-0.03015	0.857375	-0.15388
17	0.983135	-0.01701	0.842943	-0.17086	0.41812	-0.87199
5	0.99501	-0.005	0.95099	-0.05025	0.773781	-0.25647
3	0.997003	-0.003	0.970299	-0.03015	0.857375	-0.15388
6	0.994015	-0.006	0.94148	-0.0603	0.735092	-0.30776
8	0.992028	-0.008	0.922745	-0.0804	0.66342	-0.41035
4	0.996006	-0.004	0.960596	-0.0402	0.814506	-0.20517
7	0.993021	-0.007	0.932065	-0.07035	0.698337	-0.35905
chi-squared=-2*sum of ln(P)		0.2001		2.010067		10.25866
degrees of freedom		12		12		12
P(chi-squared)		1		0.99939		0.59328

Ross A. Alford, ross.alford@jcu.edu.au

Combined probability test example 3

- **5 populations sampled, all individuals negative**
- **Tests show that P for 0.01 prevalence < 0.05**
- **True prevalence in all sampled populations is probably less than 1%**

Total sampled = 565

Sample size	for true prev ≤ 0.001		for true prev ≤ 0.01		for true prev ≤ 0.05	
	P(0 positive)	ln(P)	P(0 positive)	ln(P)	P(0 positive)	ln(P)
200	0.8186	-0.2001	0.1340	-2.0101	0.0000	-10.2587
100	0.9048	-0.1001	0.3660	-1.0050	0.0059	-5.1293
75	0.9277	-0.0750	0.4706	-0.7538	0.0213	-3.8470
80	0.9231	-0.0800	0.4475	-0.8040	0.0165	-4.1035
110	0.8958	-0.1101	0.3310	-1.1055	0.0035	-5.6423
chi-squared=-2*sum of ln(P)		1.1306		11.3569		57.9614
P(chi-squared)		0.9514		0.0447		0.0000

Ross A. Alford, ross.alford@jcu.edu.au

Combining probabilities across species

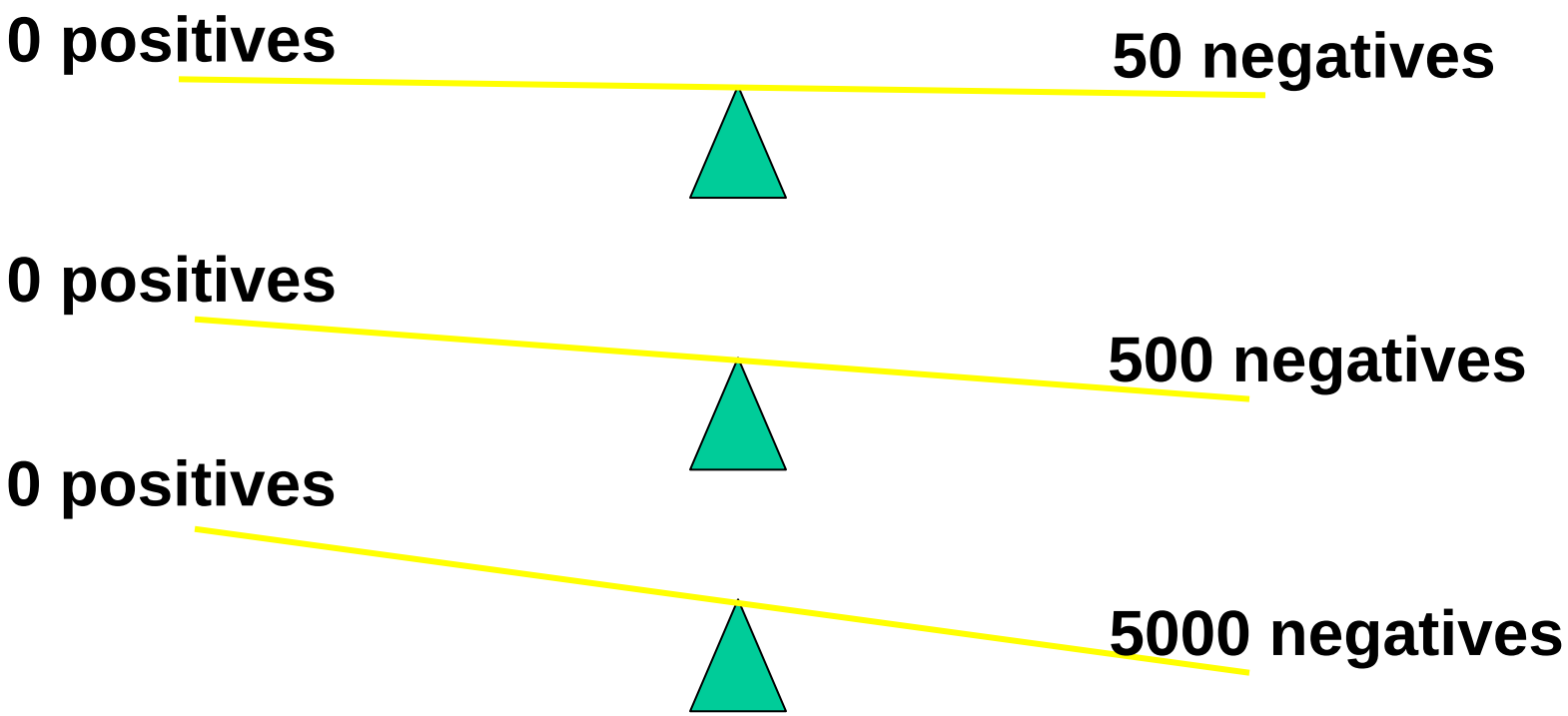
- **Only seems reasonable if all species known to have reasonably similar susceptibility**
  - **If they don't, or it is not known, result of hypothesis test could still depend simply on species composition of the sample**
    - **If it includes many/large samples from species with low or zero susceptibilities, biased towards rejecting H_0**

Demonstrating zero prevalence requires large samples *of all populations, all the same species or at least species with same susceptibilities*

- **Effectively means *it is very difficult and rarely practical to prove that Bd is absent from any place or time***
- **So why bother sampling at all?**
- **To *test* the novel pathogen hypothesis rather than attempting to confirm it**
  - **We *can easily prove* that *Bd* was *present***
  - **This takes only one confirmed positive finding**
  - **i.e., we cannot prove that our hypothesis is true, but we can try (and possibly fail) to prove that it is false.**

“Leverage”

- **How much data points are worth**
  - **Attempting to demonstrate *Bd* absence**

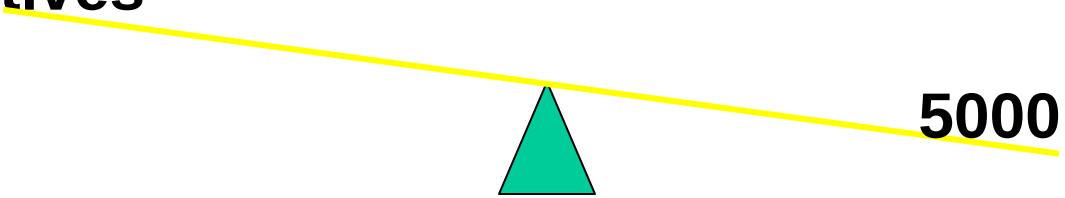


Ross A. Alford, ross.alford@jcu.edu.au

“Leverage”

- How much data points are worth
  - Attempting to demonstrate *Bd* absence

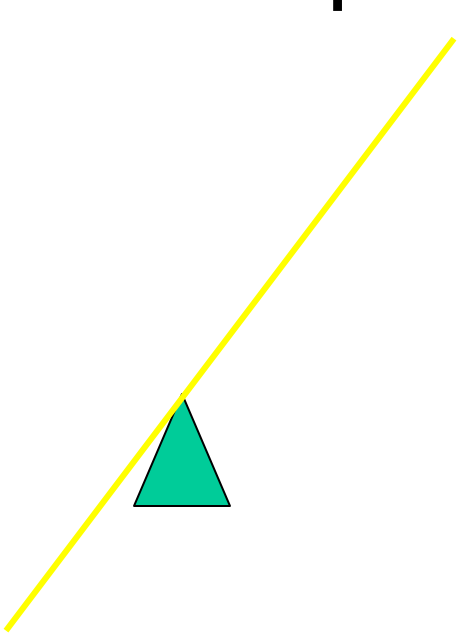
0 positives



5000 negatives

  - Attempting to test for Bd presence

5000 negatives



1 positive

Designing historical/museum surveys

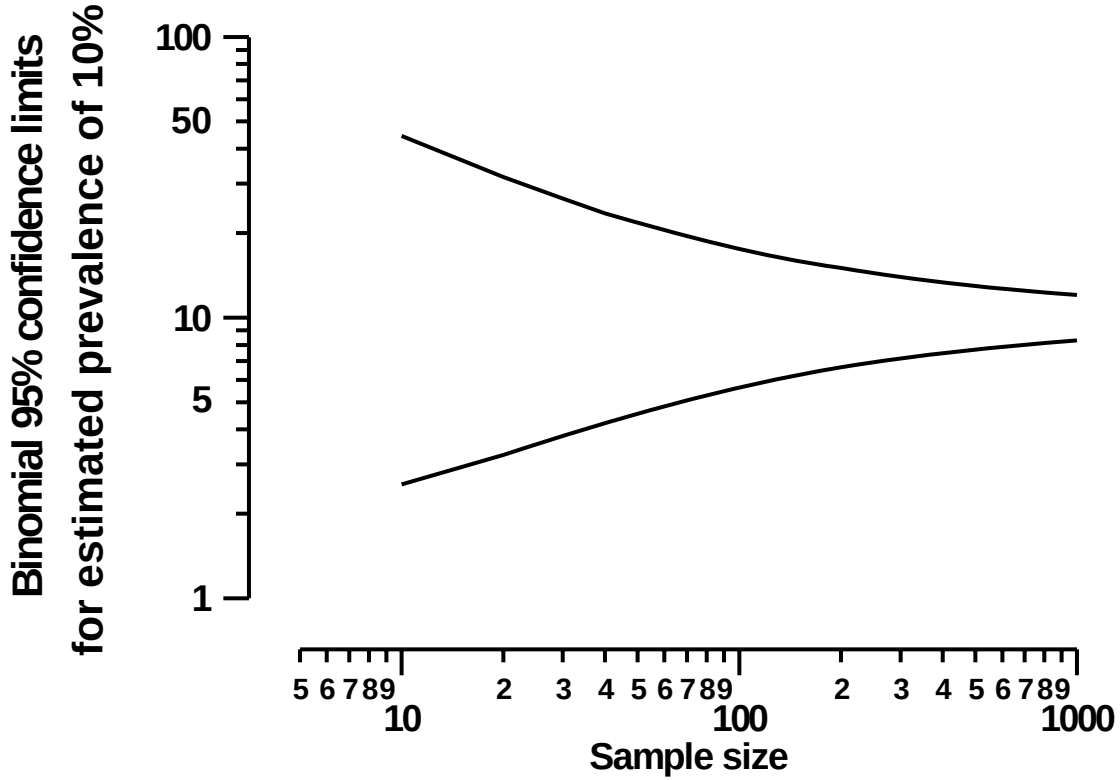
- **Start with the specimens or data points with the highest potential leverage, and work towards ones with less leverage**
  - **Highest leverage points are the oldest ones from areas where *Bd* is now present**
  - **There are usually relatively few of these, so relatively little work is required**
  - **If even a single one from long ago is positive, it sets the earliest date for *Bd* presence**

Examining prevalences from field data

- **Estimates of prevalence are meaningless without information on numbers examined, estimates of confidence limits**
- **Confidence limits are often quite broad when sample sizes are small**

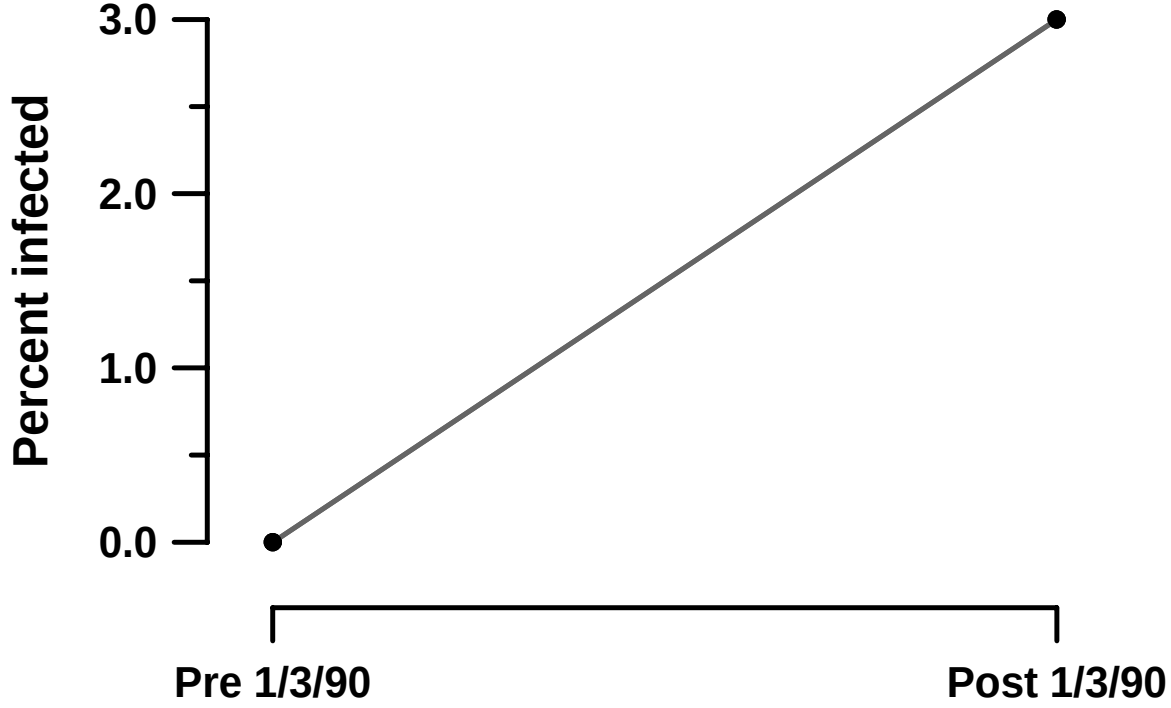
Meaning of 10% prevalence

Infection rate (%)	Number examined	95% CL for infection rate	
		lower	upper
10.00	10	2.50	44.50
10.00	50	3.33	21.81
10.00	100	4.90	17.62
10.00	500	7.51	12.97



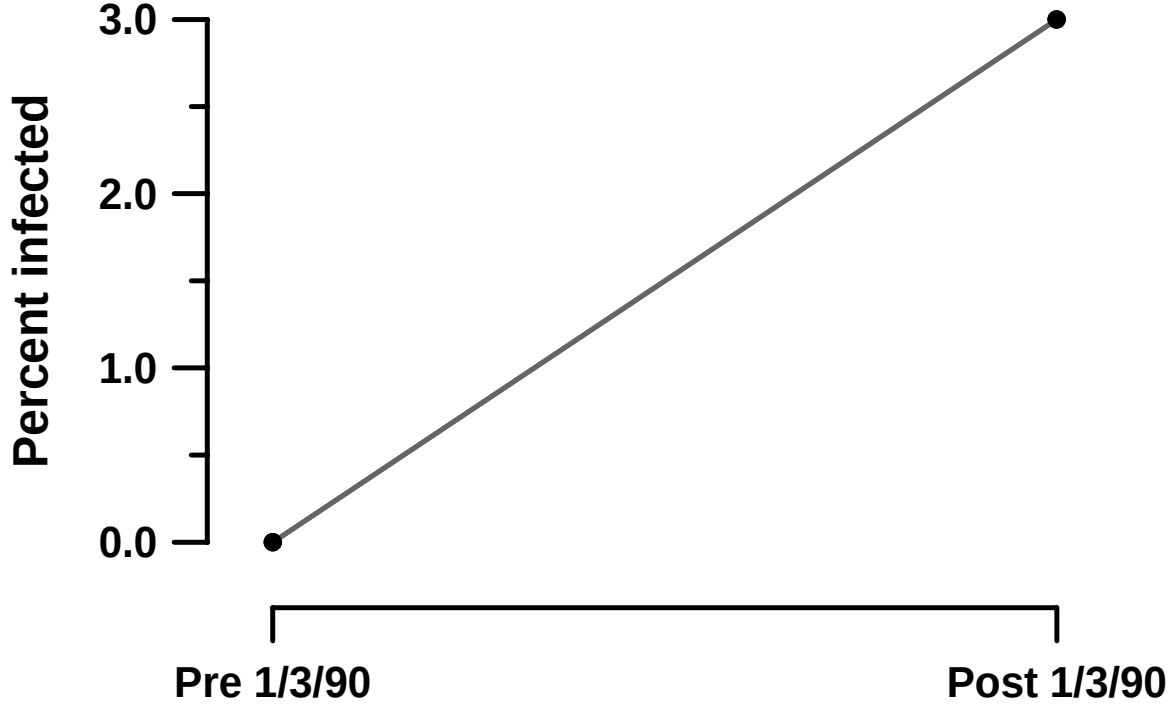
Ross A. Alford, ross.alford@jcu.edu.au

Prevalences without confidence limits can be deceptive: data on *Bd* in *Litoria nannotis* at Birthday Creek



Prevalence was 0% before 1/3/90, 3% after 1/3/90 – a dramatic difference, from this figure

Prevalences without confidence limits can be deceptive: data on *Bd* in *Litoria nannotis* at Birthday Creek

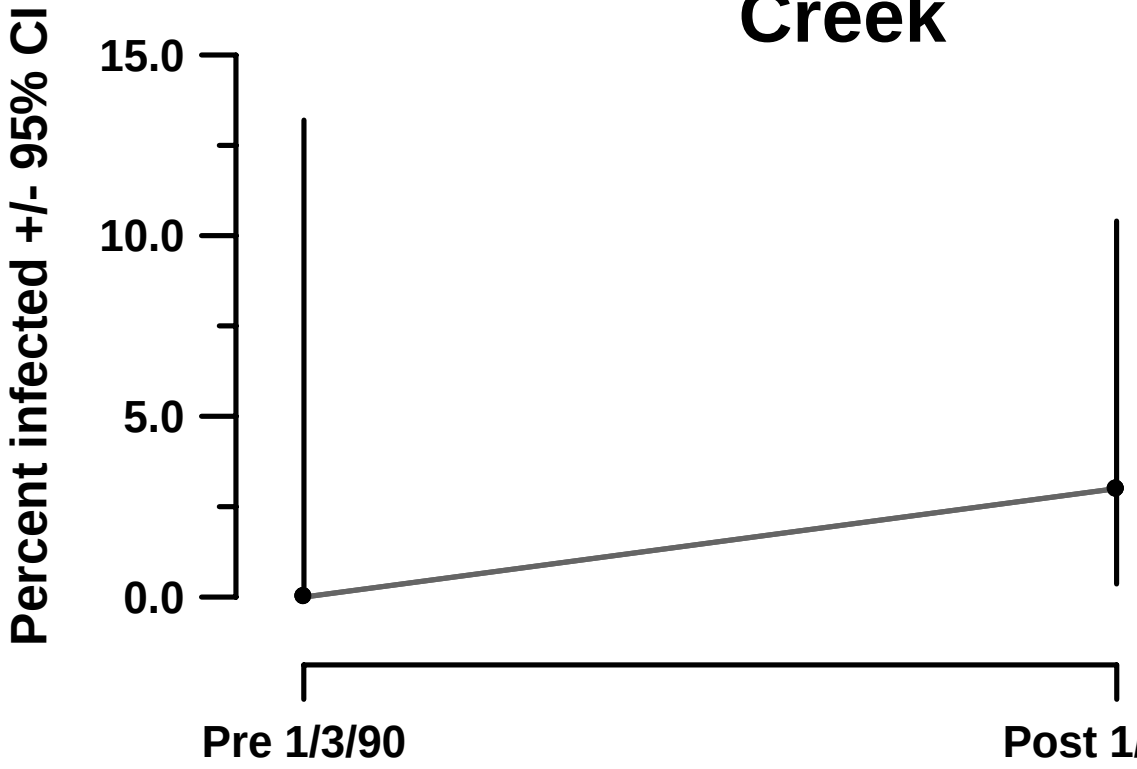


Prevalence was 0% before 1/3/90, 3% after 1/3/90 – a dramatic difference, from this figure

Time period	Number		Chytrid prevalence (%)
	total	positive for chytrids	
Pre 1/3/90	26	0	0.0
Post 1/3/90	67	2	3.0

Seeing actual numbers reduces the drama somewhat

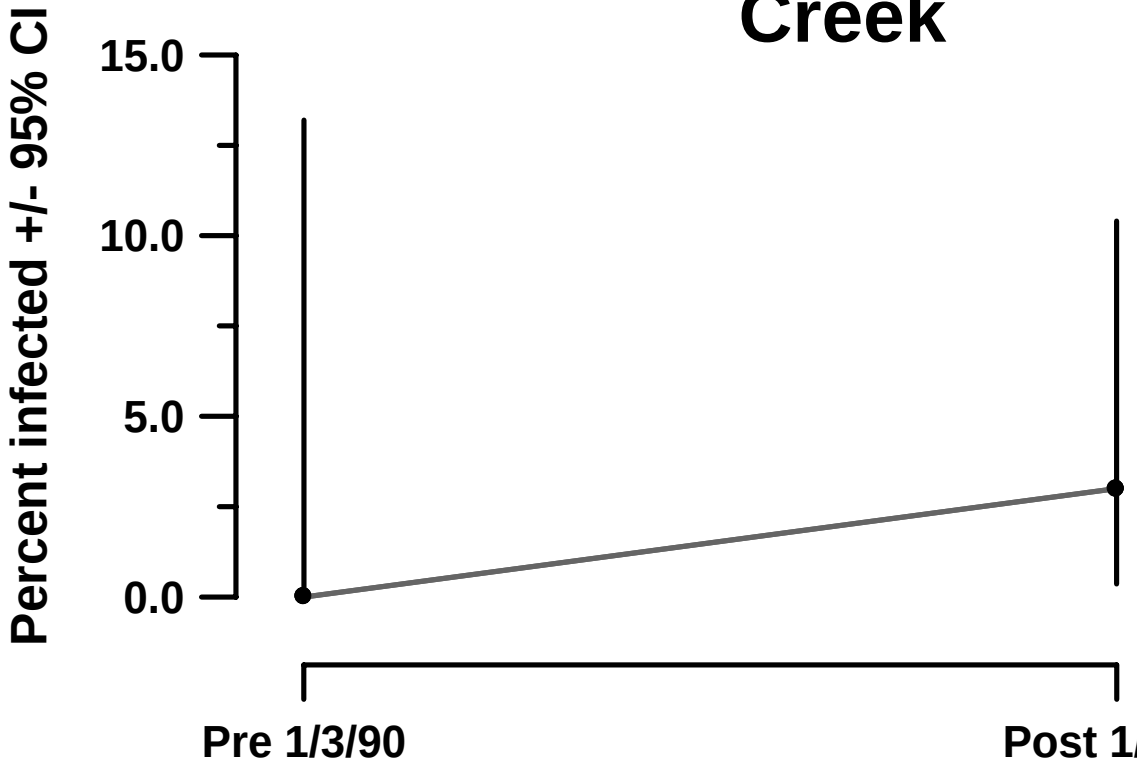
Prevalences without confidence limits can be deceptive: data on *Bd* in *Litoria nannotis* at Birthday Creek



Add confidence limits: prevalence could have been 0%-13.2% before 1/3/90, 0.36%-10.4% after 1/3/90

Time period	Number		Chytrid prevalence (%)	95% CL for chytrid prevalence (%)	
	total	positive for chytrids		lower	upper
Pre 1/3/90	26	0	0.0	0.0	13.2
Post 1/3/90	67	2	3.0	0.4	10.4

Prevalences without confidence limits can be deceptive: data on *Bd* in *Litoria nannotis* at Birthday Creek



In fact, If true prevalence is 3%, 45% of samples of 26 individuals will have 0 infected animals in them

Time period	Number		Chytrid prevalence (%)	95% CL for chytrid prevalence (%)	
	total	positive for chytrids		lower	upper
Pre 1/3/90	26	0	0.0	0.0	13.2
Post 1/3/90	67	2	3.0	0.4	10.4

Conclusion

- **Need to differentiate between the novel pathogen and disease outbreak hypotheses so that future research can be focused effectively**
- **Should test the novel pathogen hypothesis instead of attempting to confirm it**
- **If looking at multiple samples, should combine probabilities rather than samples**
- **All infection rates need confidence limits**

Strategy for testing the novel pathogen hypothesis

- **Examine samples from times and places of declines for pathogen(s) to show that they were present--this is already being done**
- **Examine all oldest samples first for pathogen**
  - **attempt to demonstrate the pathogen was historically present, thus reject the novel pathogen hypothesis**
- **work *forward* in time, increasing number of specimens**
  - **if novel pathogen hypothesis is ultimately rejected, this minimises necessary effort**
  - **If novel pathogen hypothesis ultimately is not rejected, effort is same as would be required by confirmatory approach**