Sample sizes for different epidemiologic purposes

Topics for discussion

- Why calculate sample size?
- How large a sample should I take?
- Sample sizes for 3 common scenarios

Why calculate sample size?

- Practical considerations
  - Study planning
    - Budgeting
    - Resource needs
    - Assessment of study feasibility
    - Give yourself a reasonable chance of detecting significant risk factors for a disease
  - Grant/research proposals require them!

How large a sample to take?

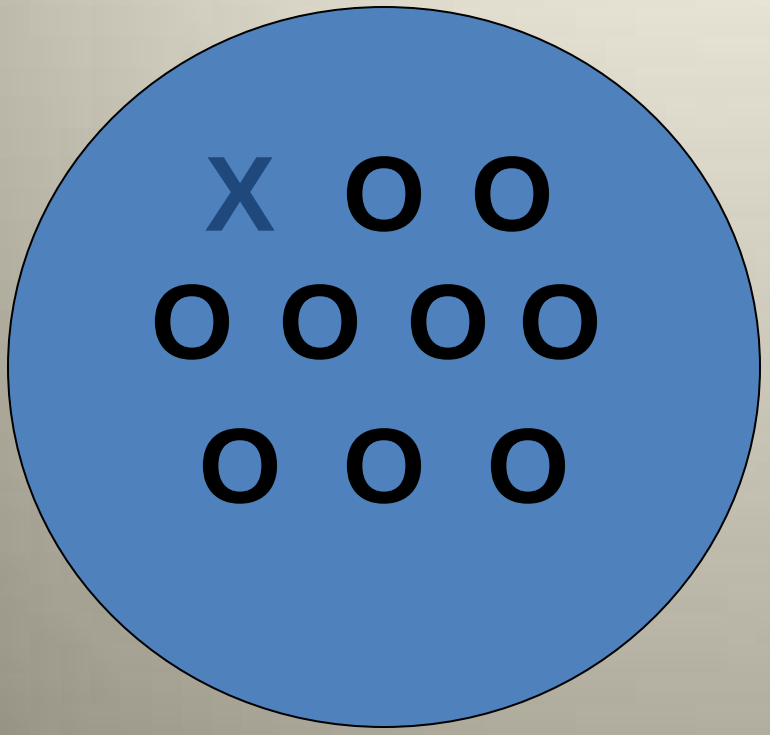
- Most important consideration is the objective of the study/diagnostic testing

  i.e. what question is being asked?
- Three common sampling situations:
  - to detect disease
    - outbreak investigations
    - flock diagnosis
    - disease freedom certification
  - to estimate prevalence
    - prevalence survey
  - to detect a difference in prevalence or incidence between groups
    - risk factor studies/clinical trials

Sample size to detect disease:

Logical reasoning

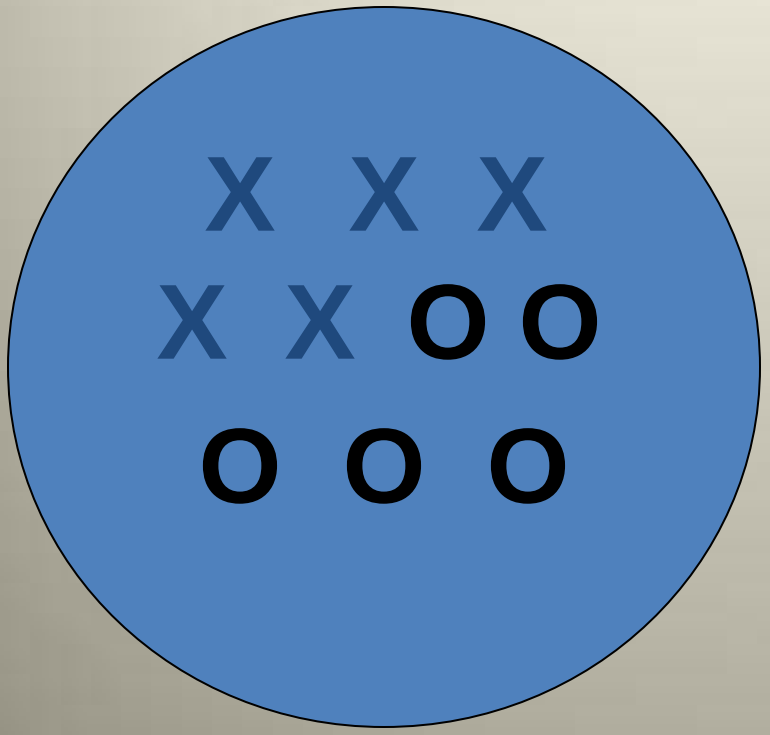
Prevalence of disease X = 10%



Confidence level	
100% (perfect)	95% (v. high)
n =	n =

How many (n) do I need to sample to detect disease (at least one X) with high confidence?

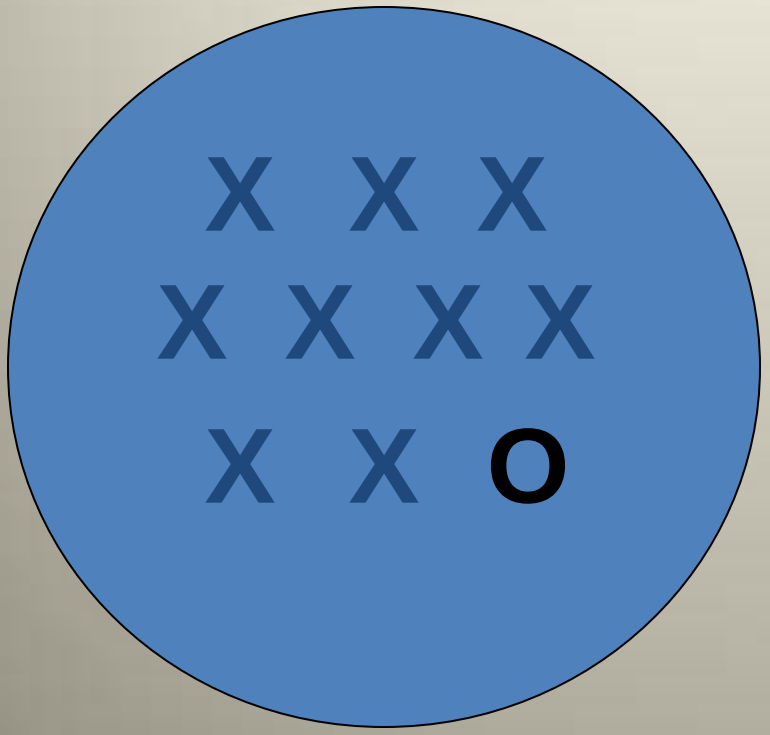
Prevalence of disease X = 50%



Confidence level	
100%	95%
n =	n =

How many (n) do I need to sample to detect disease (at least one X) with high confidence?

Prevalence of disease X = 90%



Confidence level	
100%	95%
n =	n =

How many (n) do I need to sample to detect disease (at least one X) with high confidence?

Sample size to detect disease

- Important in outbreak investigations and disease surveillance– laboratory sample submission
- Don’t have to use random sampling if a prevalence estimate is not necessary
- We assume small sample from population (binomial sampling) – formulas more complicated for hypergeometric sampling (entire population or a large part thereof is sampled)

Sample size to detect AI

- Goal is to detect ≥ 1 positive bird (in an infected flock) when we sample "n" birds from flock with unknown prevalence (P)

Probability that all "n" are disease free = $(1-P)^n$

Probability of ≥ 1 positive = $1 - (1-P)^n$

Set C (confidence) = $1 - (1-P)^n$ and solve for n

$$n = \log (1-C) / \log (1-P)$$

Sample size to detect AI

- The choice of P for an infected flock (or population) is a critical factor for this calculation, can be
  - Minimum expected prevalence
  - Prevalence that > 95% flocks would exceed
  - Typical (average) prevalence

Sample size calculations for disease detection

- Published tables
- Excel worksheet that performs calculations readily for you
- Your friendly statistician!!!

Sample size to detect disease

Confidence level (C)

Est. P	0.9	0.95	0.99
0.05	45	59	90
0.1	22	29	44
0.15	15	19	29
0.2	11	14	21
0.25	9	11	17
0.3	7	9	13
0.4	5	6	10
0.5	4	5	7
0.6	3	4	6
0.7	2	3	4
0.8	2	2	3
0.9	1	1	2

“Samp size detect” worksheet

POPULATION SIZE	**1,000**
CONFIDENCE LEVEL	**95%**
SENSITIVITY	**99%**
EXPECTED PREVALENCE	**60%**
SAMPLE SIZE (n)	**4**

$$n \cong \frac{(1-(1-\alpha)^{1/D})(N-\frac{1}{2}(SeD-1))}{Se}$$

Sample size to detect disease

- Sample size increases
  - As prevalence tends to zero
  - Desired confidence increases
  - Sensitivity of detection decreases (not shown)

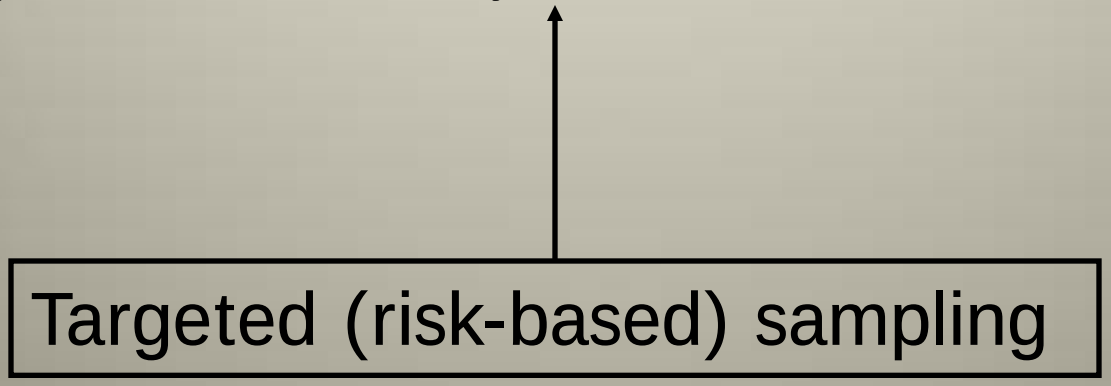
- Remember that it is sensible to focus on high-risk groups (e.g. clinically affected individuals or groups where prevalence is likely to be greater) to reduce sample sizes

Sample size to detect disease

- If infection clusters in some flocks/age groups but not in others,
  - then sampling a single unit or building may no longer be adequate,
  - and thus the sample size represents the *numbers for each flock*.

Sample size to detect disease

- If the goal is detection of infection and not estimation of prevalence, then animals with clinical signs consistent with disease can be sampled selectively



Sample size to detect disease

Confidence level (C)

Est. P	0.9	0.95	0.99
0.05	45	59	90
0.1	22	29	44
0.15	15	19	29
0.2	11	14	21
0.25	9	11	17
0.3	7	9	13
0.4	5	6	10
0.5	4	5	7
0.6	3	4	6
0.7	2	3	4
0.8	2	2	3
0.9	1	1	2

Interpretation of 0 positive in the number of selected samples

- The table can also be used once results are obtained

- Assume 6 samples were selected and the results are all negative, what can we conclude assuming that our testing procedure (test and sample collection and handling) were perfect?

Interpretation of 0 positive in the number of selected samples

- Prevalence is zero?

  or

- Prevalence is <0.6 (60%) with 99% confidence