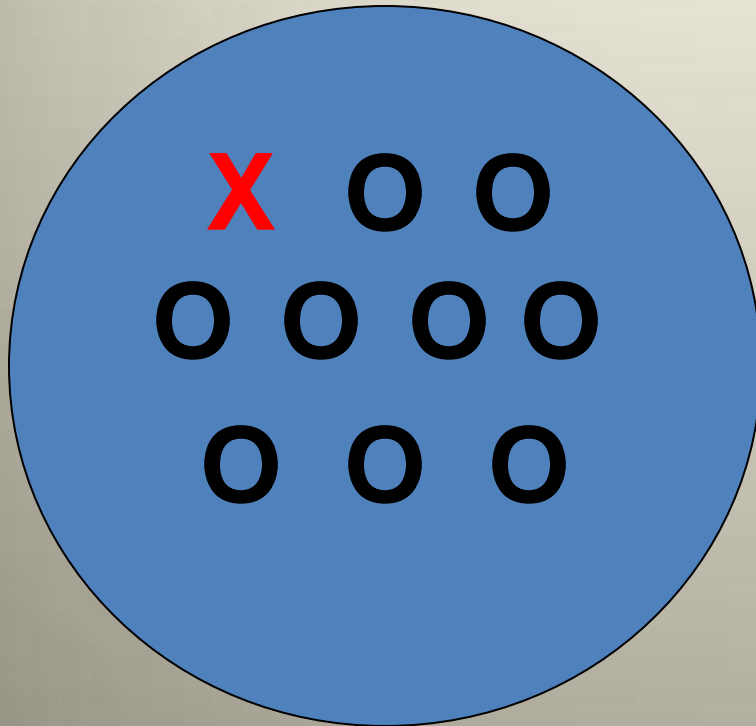


Sample size to estimate prevalence

Prevalence of disease X = 10%



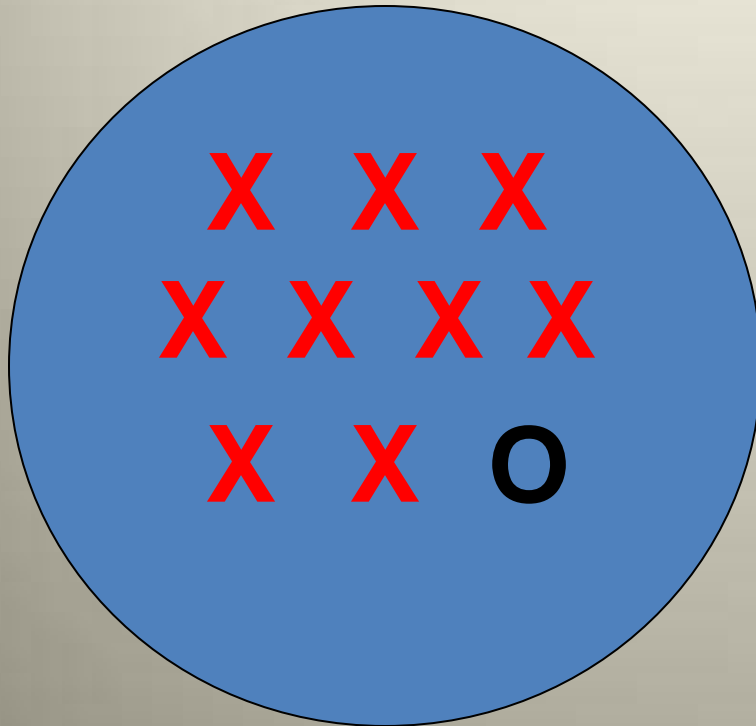
Possible results

0/5 (0%)

or 1/5 (20%)

If I sample 5 out of the population of 10, how accurately will I estimate prevalence?

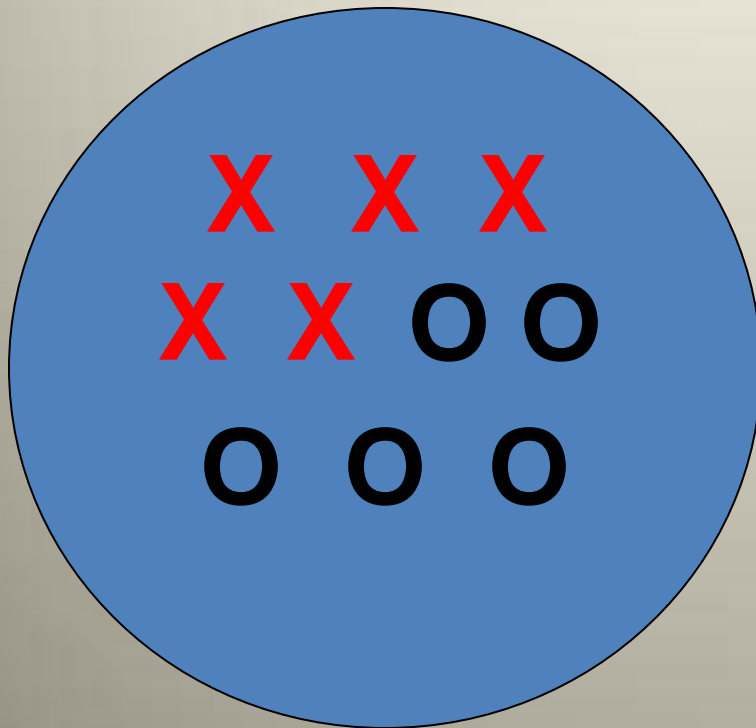
Prevalence of disease X = 90%



Possible results
 $3/5$ (60%)
or $4/5$ (80%)

If I sample 5 out of the population of 10, how accurately will I estimate prevalence?

Prevalence of disease X = 50%



Possible results

If I sample 5 out of the population of 10, how accurately will I estimate prevalence?

Sample size formula for number to estimate prevalence in a survey

$$n = \frac{Z^2_{1-\alpha/2} P(1-P)}{e^2}$$

where n = number to sample

$Z^2 = (1.96)^2$ for 95% confidence (i.e. $\alpha = 0.05$)

P = “best guess” for prevalence (e.g. ± 0.50)

e = maximum tolerable error for the prevalence estimate (e.g. ± 0.05)

Calculations

- Use published tables (as in next slide)
- Enter formula in Excel or use a hand calculator
- Use EpiInfo – Statcalc module

Sample size to estimate prevalence

Maximum tolerable error in estimated prevalence

Est. P	± 0.05	± 0.10	± 0.15	± 0.2
0.01	16	<10	<10	<10
0.02	31	<10	<10	<10
0.05	73	19	<10	<10
0.1	139	35	16	<10
0.15	196	49	22	13
0.2	246	62	28	16
0.25	288	72	32	18
0.3	323	81	36	21
0.35	350	88	39	22
0.4	369	93	41	24
0.45	381	96	43	24
0.5	384	96	43	24

95% confidence; for prevalence > 0.5, use 1 – estimated prevalence

Example

- Assume that you would like to estimate the prevalence (P) of avian influenza antibodies in a poultry flock with an error margin of ± 0.05 and the “best guess” of P is 0.5.

$$n = \frac{Z^2_{1-\alpha/2} P(1-P)}{e^2} = \frac{1.96^2 * 0.5 * (1 - 0.5)}{0.05 * 0.05} \approx 384$$

If we wanted a narrower interval: $e = 0.01$ then ≈ 9604

If p were lower, say 0.1 then ≈ 139

Example

```
Command Prompt - epi6
EpiInfo Version 6          Statcalc          November 1993
Population Survey or Descriptive Study Using Random <Not Cluster> Sampling

Size of population from which
the sample will be selected :          999,999

Expected frequency of the factor
under study (err toward 50%) :          50.00 %

If 50.00 % is the true rate in the population, what is the result
farthest from the rate that you would accept in your sample (higher
or lower)?

Worst acceptable result :          45.00 %

F1-Help          F4-Calc          F6-Open File          F10-Done
```

Example

```
Command Prompt - epi6
EpiInfo Version 6          Statcalc          November 1993
Population Survey or Descriptive Study Using Random <Not Cluster> Sampling

Population Size      :      999,999
Expected Frequency   :      50.00 %
Worst Acceptable     :      45.00 %

Confidence Level     :      Sample Size
-----
      80 %            :      164
      90 %            :      270
      95 %            :      384
      99 %            :      663
      99.9 %          :     1,082
      99.99 %         :     1,511

Change value of Population, Frequency, or Worst Acceptable to recalculate.

F1-Help          F5-Print          F6-Open File          F10-Done
```

Summary

- Sample size to estimate prevalence increases as
 - Prevalence tends towards 50% from either 0% or 100%
 - Allowable “error margin” decreases
 - Confidence increases from 95% to 99% to 100%