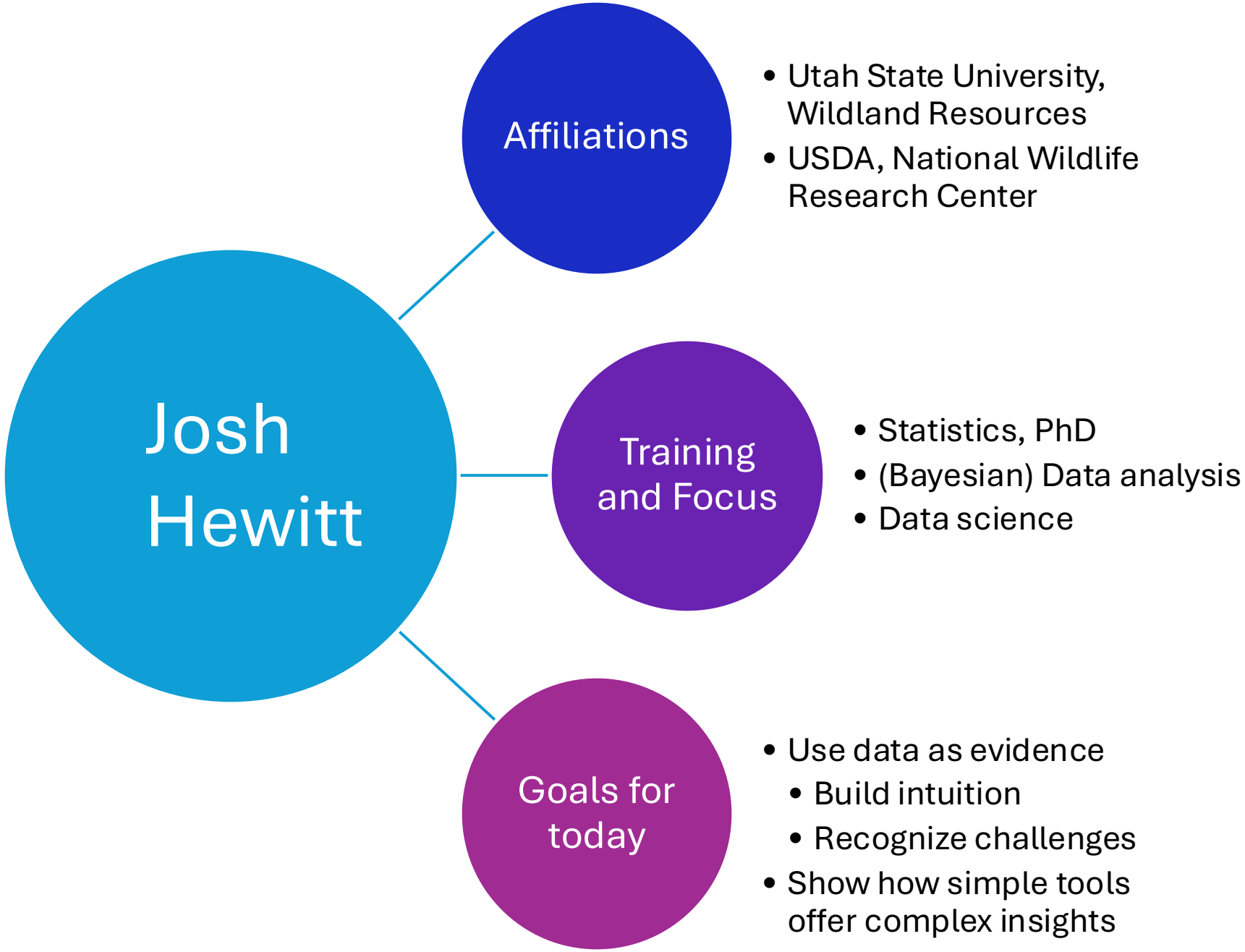
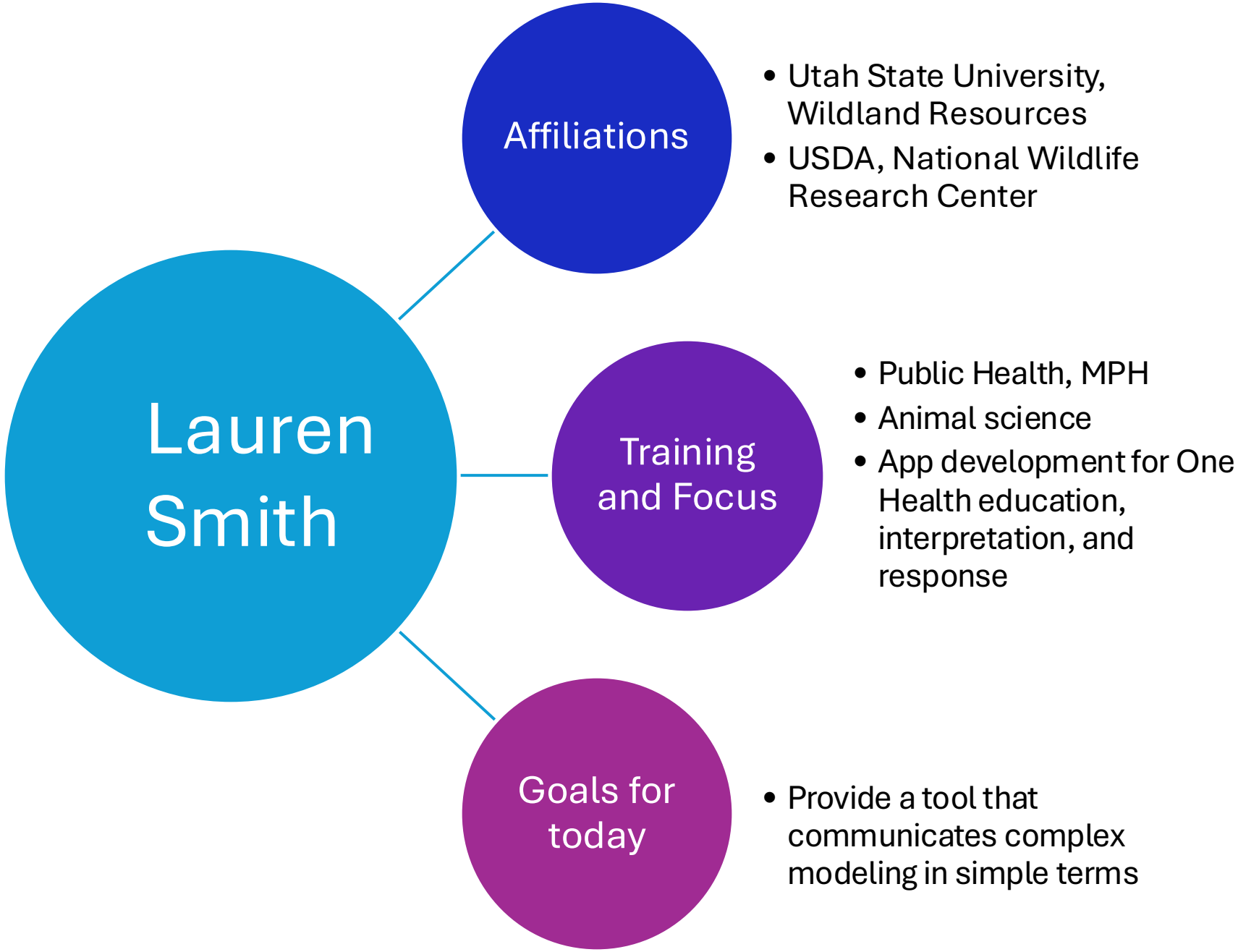
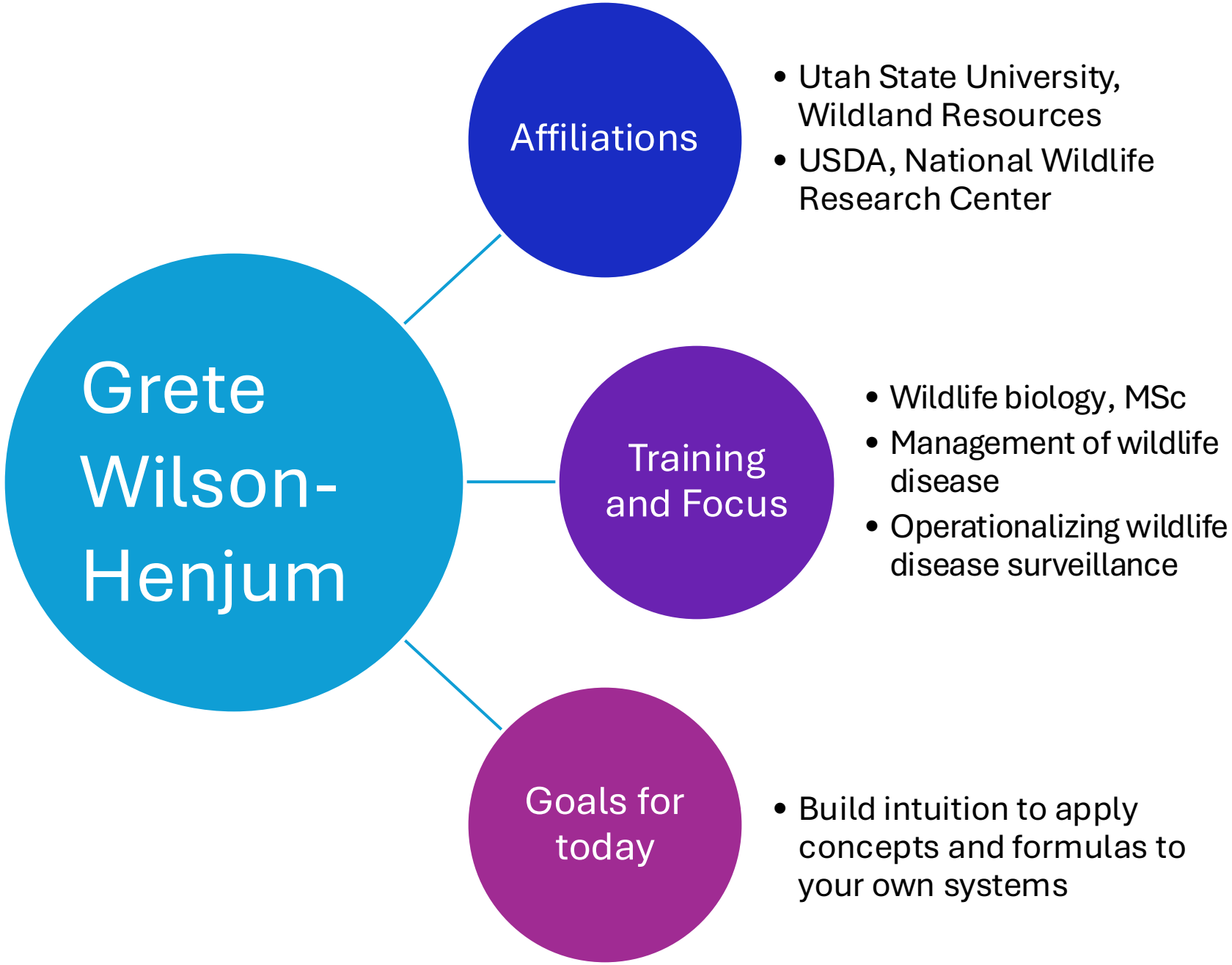
Wildlife disease surveillance design tools: trade-offs between design elements

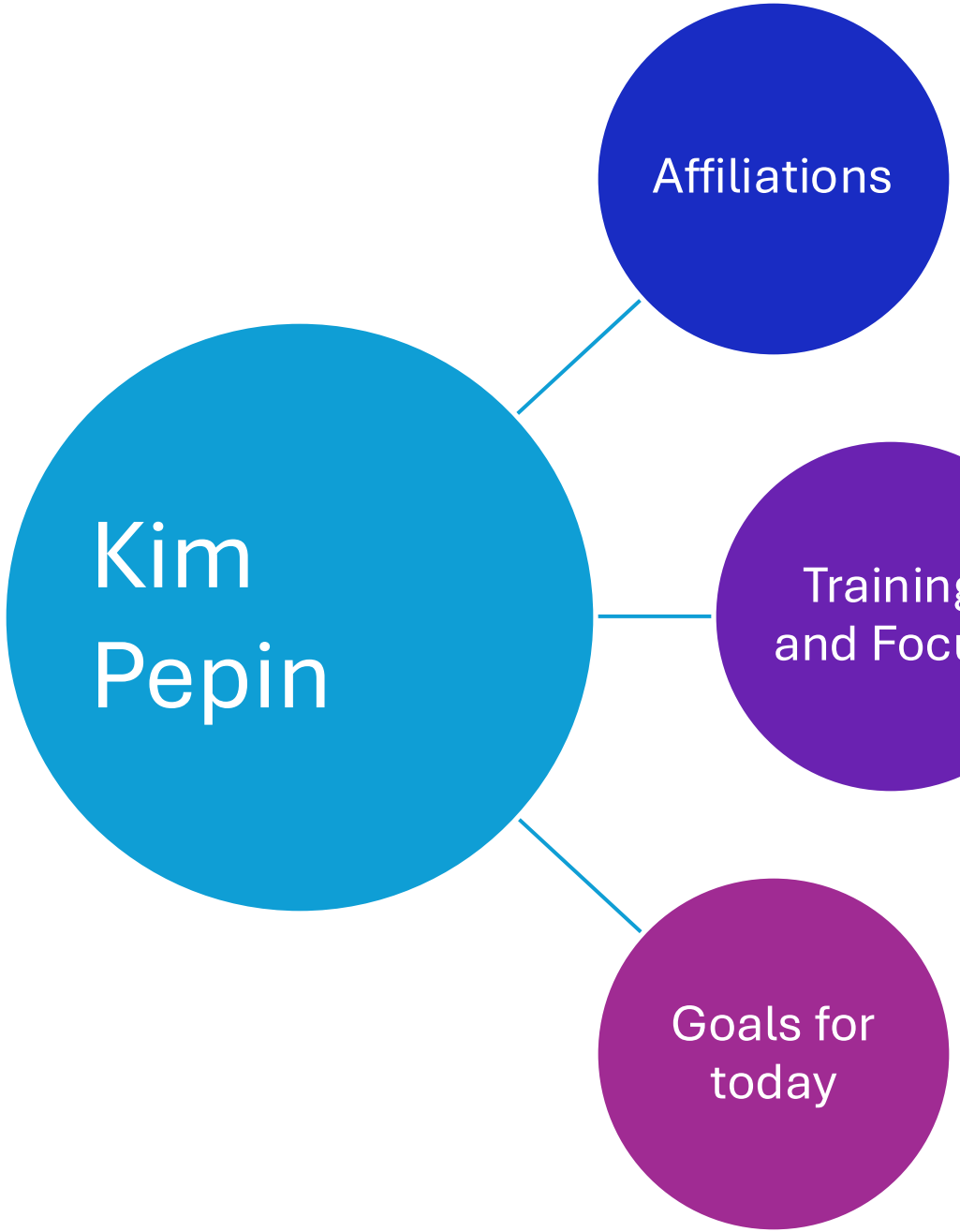
Josh Hewitt, Lauren Smith, Grete Wilson-Henjum, Kim Pepin

TWS Workshop, October 2024


Josh Hewitt
Affiliations
• Utah State University, Wildland Resources
• USDA, National Wildlife Research Center
Training and Focus
• Statistics, PhD
• (Bayesian) Data analysis
• Data science
Goals for today
• Use data as evidence
  • Build intuition
  • Recognize challenges
• Show how simple tools offer complex insights


Lauren Smith
Affiliations
Utah State University, Wildland Resources
USDA, National Wildlife Research Center
Training and Focus
Public Health, MPH
Animal science
App development for One Health education, interpretation, and response
Goals for today
Provide a tool that communicates complex modeling in simple terms


Grete Wilson-Henjum
Affiliations
• Utah State University, Wildland Resources
• USDA, National Wildlife Research Center
Training and Focus
• Wildlife biology, MSc
• Management of wildlife disease
• Operationalizing wildlife disease surveillance
Goals for today
• Build intuition to apply concepts and formulas to your own systems



- Research Biologist
- USDA, National Wildlife Research Center

- Disease ecology
- Epidemiological modeling
- Disease management: Control strategies, risk assessment, surveillance design
- Viral evolution/molecular biology, PhD

- How to apply simple guidance for surveillance design to complex ecological processes
- Make the guidance accessible to a variety of stakeholders

WHY ARE WE HERE?

- Researchers and wildlife professionals who need to plan a research study or surveillance system but may not be familiar with sample size statistics

- Those interested in transparent, easy-to-use tools for planning disease surveillance designs that fit the surveillance objectives and ecological processes in the system

WHY ARE YOU HERE?

Kickoff poll

Mentimeter

https://www.menti.com/alxy8ddaweaj

Logistics

- Schedule
  - Kickoff poll
  - Key sampling ideas
  - Discussion, app, and exercise topics
    - Disease detection
    - Prevalence estimation
    - Epidemiological dynamics
    - Seasonality
    - Spatial needs
  - Additional logistics
  - Wrap up

- Breaks in between sections
  - Bathrooms in hallway

- Get you laptop on wifi, eventually connect to https://deerdisease.shinyapps.io/Wildlife-surveillance-design-tools/

- Rules of behavior

  - Ask questions
    - This is a high-level view of a big topic
    - Everyone has different backgrounds and experiences

  - Be open to potentially new ideas
    - Relatively little standardization
    - Tools need training, even for familiar topics
    - Share your experiences too

  - Be respectful—tackle problems, not people

  - Be active participants, help each other

Learning goals: Trade-offs in surveillance design

- Connect surveillance objectives to appropriate surveillance designs

- Guide decision making with trade-offs:
  - Too little disease surveillance data can deliver unresolved answers for surveillance objectives
  - Too much data can inefficiently use resources
  - Too little data can inefficiently use resources

- Use shiny app to guide answers to the questions

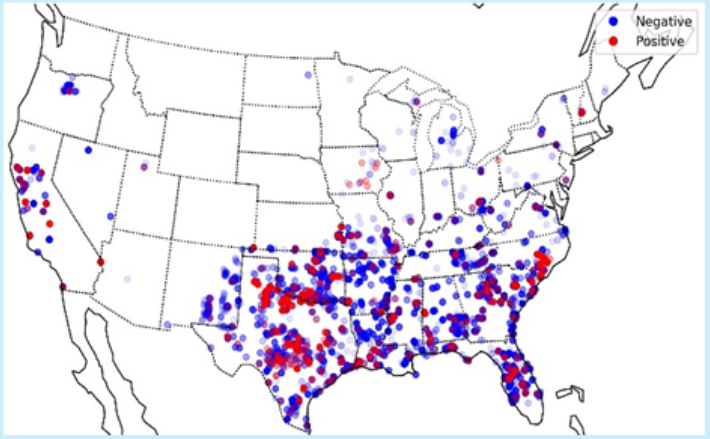
Workshop topics overview

Scale	Metric	Example surveillance objective
Individual Population	**Detection**	Which individuals are most at risk?
Population	**Prevalence**	What proportion of the population is affected?
Population	**Seasonality**	How frequent are outbreaks?
Population	**Epidemiological dynamics** (Incidence, R0, persistence, etc.)	What are the rates of infection and how do they change over time? Is the system predictable? What are the ecological drivers?
Landscape	**Spatial distribution**	Over what spatial area does the disease occur?

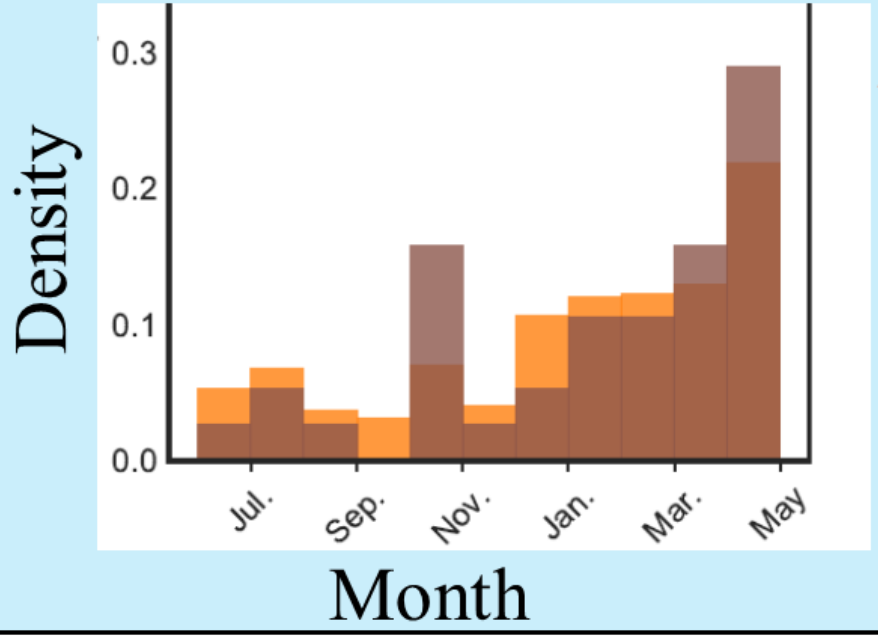
Importance of sampling design & interpretation

*Sampling design must be accounted for in inference

National-scale opportunistic sampling

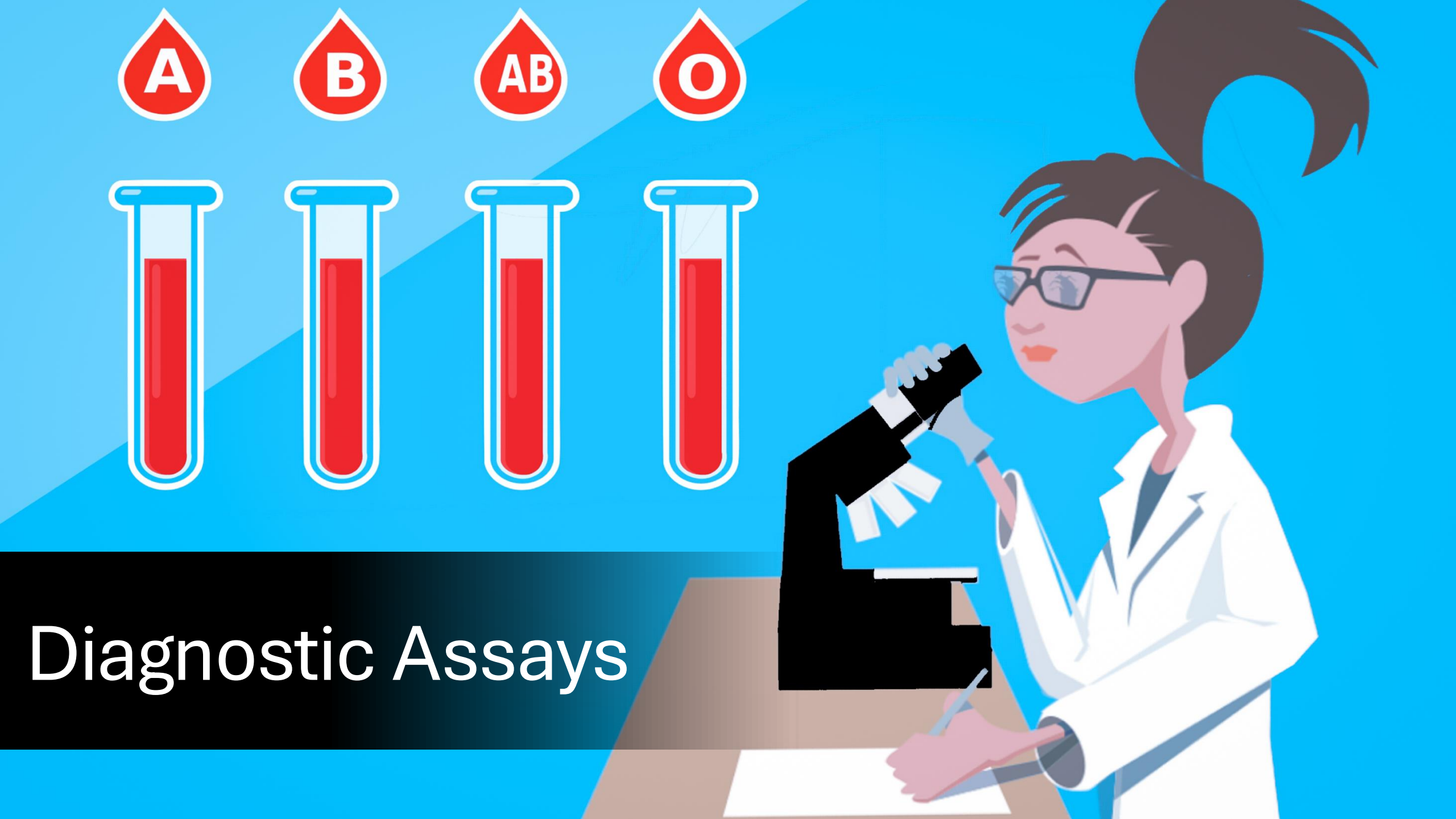


Apparent seroprevalence = sampling design



All samples

Positive samples (seroprevalence)



Diagnostic Assays

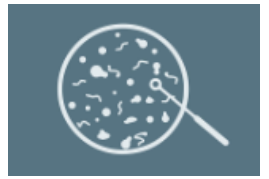
Typical Assays

Pathogen detection

Aims to isolate pathogen from biological sample, indicating active or recent infection

Examples:

- Cultures to detect live pathogens (bacterial and fungal infections)
- PCR to detect pathogen nucleic acid – pathogen can be live or dead (viral, bacterial, and fungal infections)

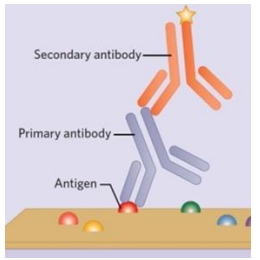


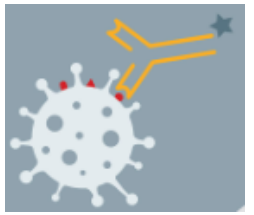
Immune response detection

Aims to detect specific antigen or antibody immune response, indicating pathogen exposure

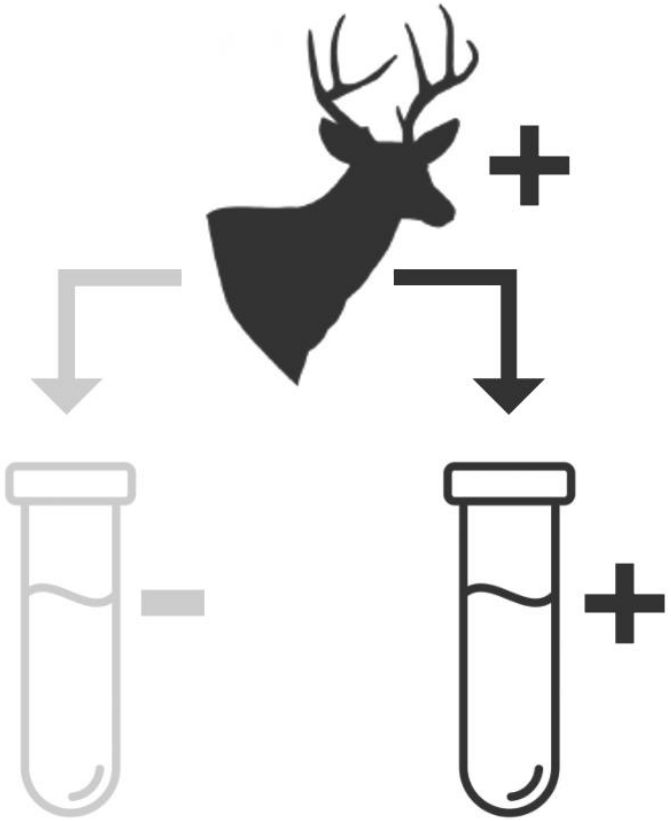
Examples:

- Immunohistochemistry to detect antigen markers (prion infections)
- Enzyme-linked immunosorbent assay (ELISA) to detect antigen-antibody interactions
- Virus neutralization to detect antibody response to virus
- Passive hemagglutination assay (PHA) to detect antibody responses to bacteria



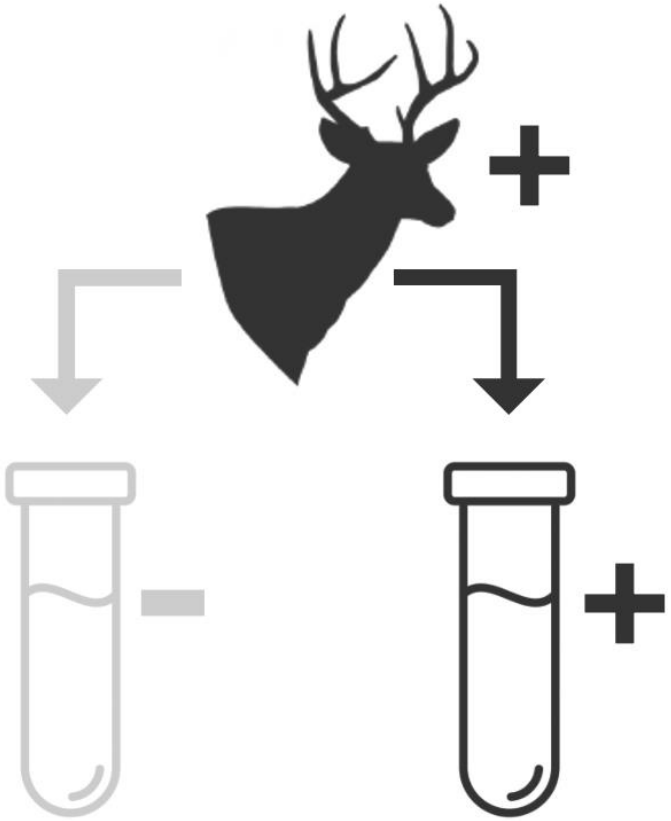


Sensitivity vs.

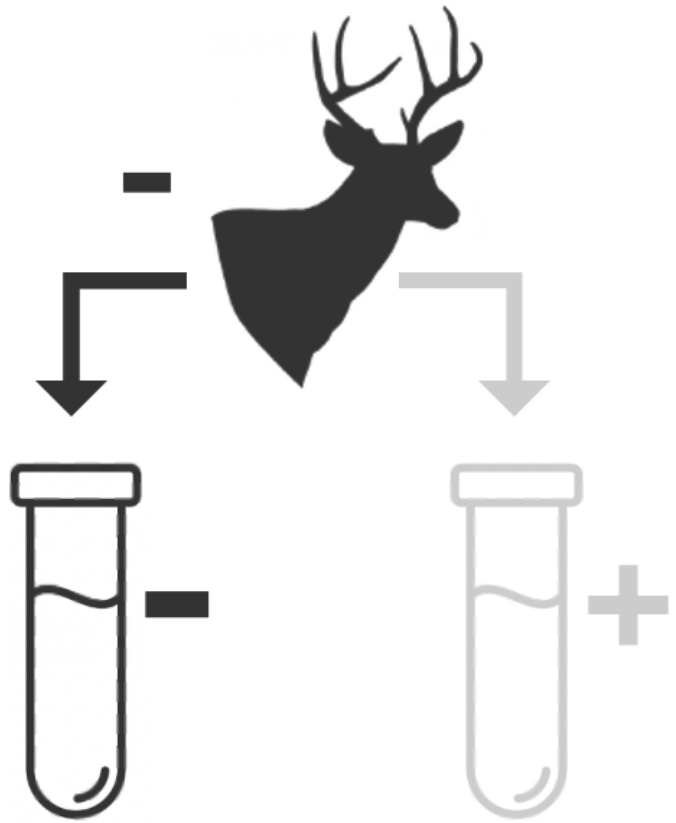


e.g., 95%

Sensitivity vs. Specificity



e.g., 95%



e.g., 95%

Detection

Design factors

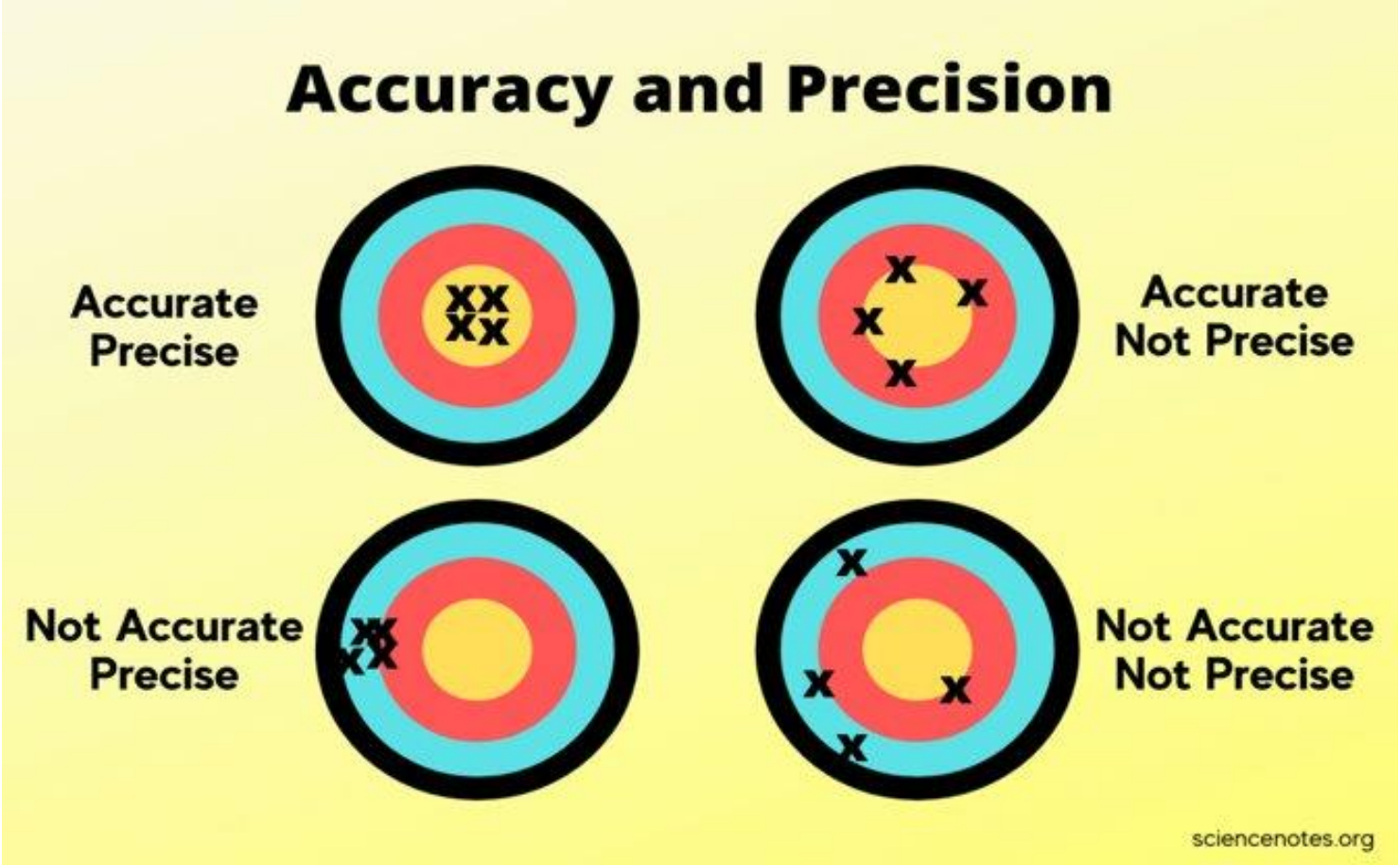
Learning goals

- Analyze data to evaluate disease detection
- Decide when, where, and how many samples to collect
- Collecting too much or too little data inefficiently uses resources
  - Evaluate ability to observe occurrence/presence
- Detection examples, and connections to surveillance design
- Design surveillance to meet potential detection objectives
  - Disease freedom declaration
  - Occurrence/presence

Planning is hard

Unlike STAT101, there is no single, simple sample size formula

Design can control accuracy and precision



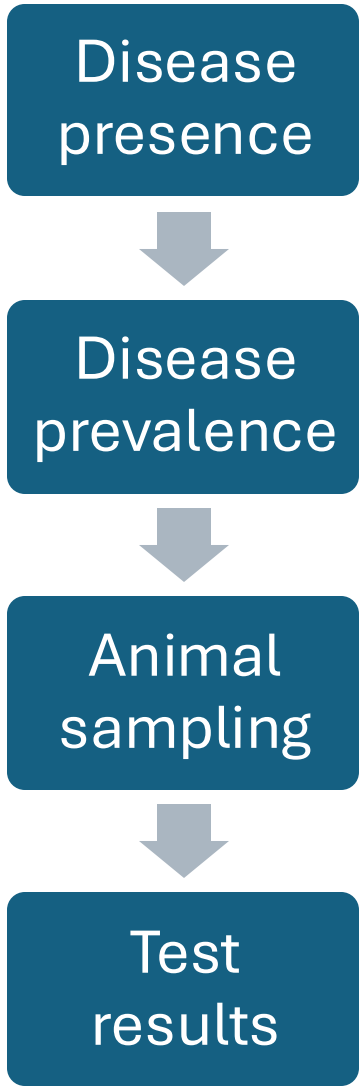
Sample size improves accuracy, precision at different rates

- Required sample size depends on goals
  - Precision: An individual study's confidence in it's findings
    - "Based on the data and assumptions provided, it is 99% probable there is no disease in the system"
  - Accuracy: The regularity with which studies will make conclusions that agree with the truth
    - "On average, 50% of studies with these sample sizes and assumptions will have enough precision to declare disease is/is not present when prevalence is really 5%"
  - Sample sizes may be different for "is" vs. "is not" accuracy statements
    - "Is" vs "is not" are not either/or study outcomes
    - An individual study may not be precise enough to support either conclusion

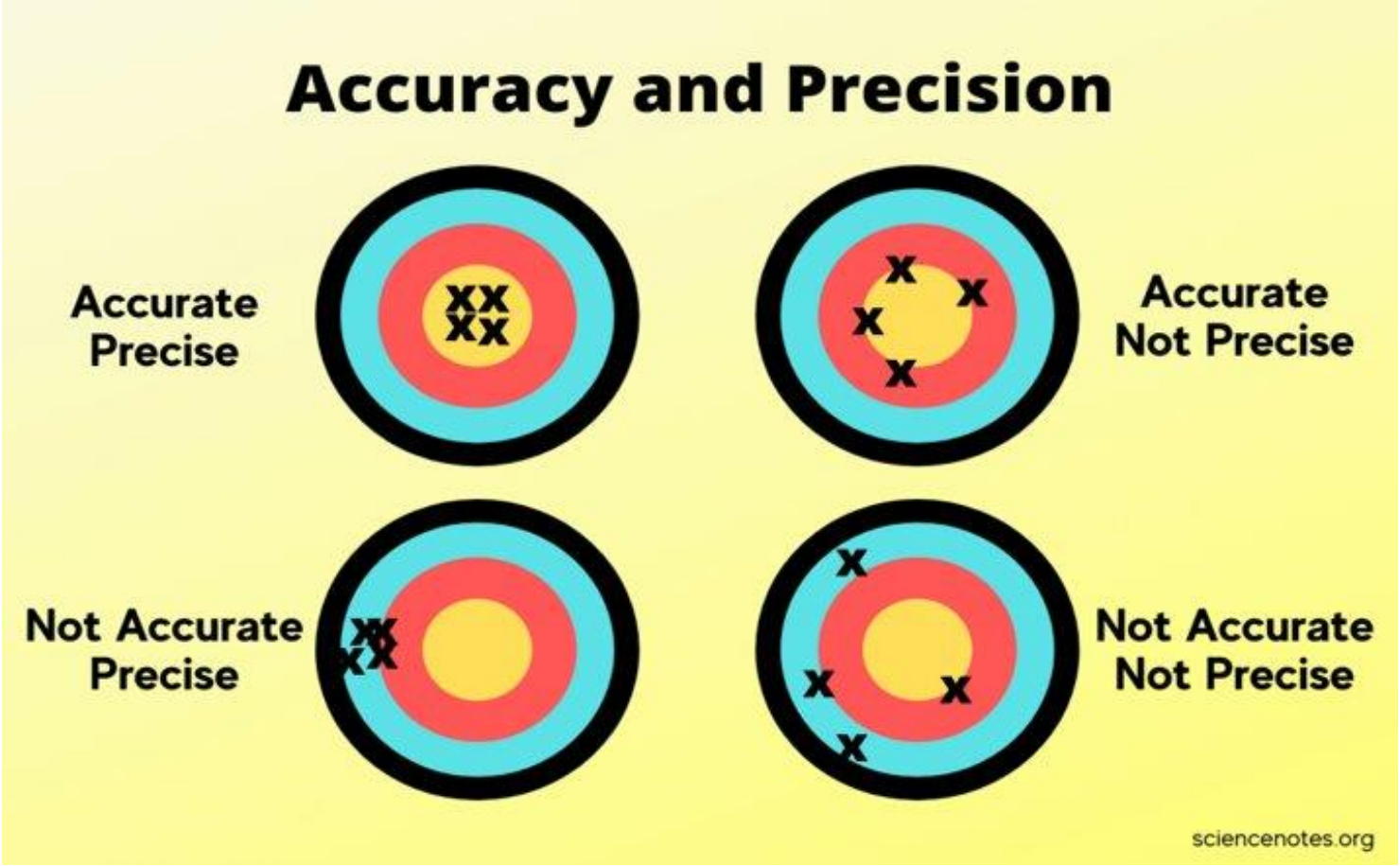
Statistics for Detection

- Model assumptions
  - Likely disease prevalence range
  - Constant prevalence during sampling
  - Sample size small relative to population
  - Equally likely chance of being sampled

- Inputs
  - Disease presence chance (Prior)
  - Potential prevalence if present (Prior)
  - Sample size
  - Test sensitivity/specificity
  - Test results (i.e., num. pos, Data analysis)
    …or… True prevalence (Power study)



Outputs characterize



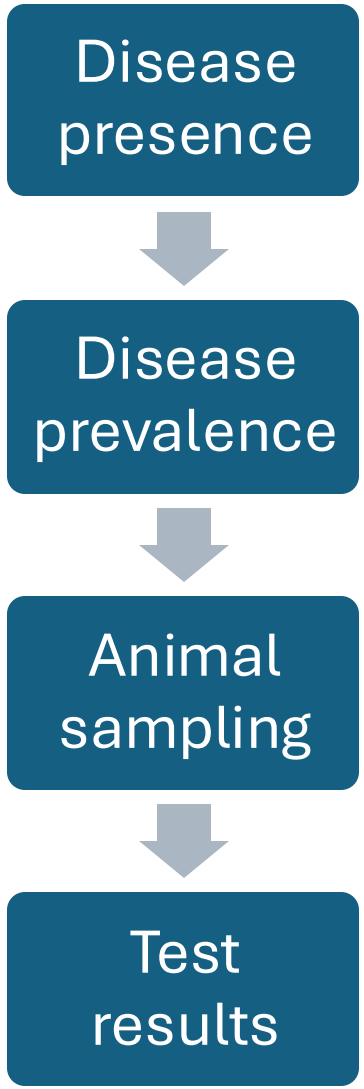
Statistics for Detection

- Outputs
  - Sample interpretation (Data analysis)
    - Probability for disease presence/absence (i.e., freedom, **precision**)
    - Upper bound for prevalence, if present but not observed (i.e., precision)
  - Surveillance design evaluation (Power study)
    - Sample size impact on probability design declares presence/absence **(i.e., accuracy)**
- Considerations for wildlife…

Statistics for Detection

- Model assumptions
  - Likely disease prevalence range
  - Constant prevalence during sampling
  - Sample size small relative to population
  - Equally likely chance of being sampled

- Inputs
  - Disease presence chance (Prior)
  - Potential prevalence if present (Prior)
  - Sample size
  - Test sensitivity/specificity
  - Test results (i.e., num. pos, Data analysis)
    ...or... True prevalence (Power study)



Equally likely: Ecological group (Key idea!)

- Target population
  - A group of individuals, who collectively reflect the surveillance program's goals

- vs. Ecological group (i.e., statistical sampling population)

  A group of individuals, any one of whom is equally representative of others (i.e., based on risk factors, strata, etc.)

- Impacts
  - ⚠ **Ignoring ecological groups can bias results** ⚠

- Definitions can depend on disease patterns
  - Age/Sex/Social structure/Behavior… e.g.,  
  - Season    
  - Climate and landcover    
  - Proximity to people    

- Solution: Stratified or probabilistic sampling

Example: CWD ecological groups defined by demographic and behavioral attributes

Demographic

- Male
- Female
- Adult
- Subadult

Behavioral

- Live animals presenting diseas e signs
- Road-kill
- Hunter harvest

Table 1. Estimated mean hazard ratios, associated standard deviations, Markov Chain Monte Carlo (MCMC) errors, and quantiles for various demographic groups of mule deer (*Odocoileus hemionus*) based on samples collected from chronic wasting disease-positive areas in Colorado during 2003–6.

Group	Mean	Standard deviation	Markov Chain Monte Carlo error	2.50 percent	Median	97.50 percent
Suspect female	14.13	2.390	0.0151	9.90	13.97	19.24
Suspect male	12.19	2.070	0.0135	8.50	12.06	16.64
Other	1.93	0.245	0.0018	1.49	1.92	2.44
Harvest—adult males	1.00	NA	NA	NA	NA	NA
Harvest—adult females	0.57	0.065	0.0005	0.46	0.57	0.71
Harvest—yearling females	0.44	0.150	0.0009	0.20	0.43	0.78
Harvest—yearling males	0.25	0.077	0.0004	0.12	0.24	0.43
Harvest—fawns	0.03	0.032	0.0002	0.001	0.02	0.12

Walsh, D. P. (2012). US Department of the Interior, US Geological Survey.

Example: CWD ecological groups defined by demographic and behavioral attributes

Demographic

- Male
- Female
- Adult
- Subadult

Behavioral

- Live animals presenting disease signs
- e.g., road-kill
- Hunter harvest

Table 1. Estimated mean hazard ratios, associated standard deviations, Markov Chain Monte Carlo (MCMC) errors, and quantiles for various demographic groups of mule deer (*Odocoileus hemionus*) based on samples collected from chronic wasting disease-positive areas in Colorado during 2003–6.

Group	Mean	Standard deviation	Markov Chain Monte Carlo error	2.50 percent	Median	97.50 percent
Suspect female	14.13	2.390	0.0151	9.90	13.97	19.24
Suspect male	12.19	2.070	0.0135	8.50	12.06	16.64
Other	1.93	0.245	0.0018	1.49	1.92	2.44
Harvest—adult males	1.00	NA	NA	NA	NA	NA
Harvest—adult females	0.57	0.065	0.0005	0.46	0.57	0.71
Harvest—yearling females	0.44	0.150	0.0009	0.20	0.43	0.78
Harvest—yearling males	0.25	0.077	0.0004	0.12	0.24	0.43
Harvest—fawns	0.03	0.032	0.0002	0.001	0.02	0.12

Walsh, D. P. (2012). US Department of the Interior, US Geological Survey.

Example: CWD ecological groups defined by demographic and behavioral attributes

Demographic

- Male
- Female
- Adult
- Subadult

Behavioral

- Live animals presenting disease signs
- Road-kill
- Hunter harvest

Table 1. Estimated mean hazard ratios, associated standard deviations, Markov Chain Monte Carlo (MCMC) errors, and quantiles for various demographic groups of mule deer (*Odocoileus hemionus*) based on samples collected from chronic wasting disease-positive areas in Colorado during 2003–6.

Group	Mean	Standard deviation	Markov Chain Monte Carlo error	2.50 percent	Median	97.50 percent
Suspect female	14.13	2.390	0.0151	9.90	13.97	19.24
Suspect male	12.19	2.070	0.0135	8.50	12.06	16.64
Other	1.93	0.245	0.0018	1.49	1.92	2.44
Harvest—adult males	1.00	NA	NA	NA	NA	NA
Harvest—adult females	0.57	0.065	0.0005	0.46	0.57	0.71
Harvest—yearling females	0.44	0.150	0.0009	0.20	0.43	0.78
Harvest—yearling males	0.25	0.077	0.0004	0.12	0.24	0.43
Harvest—fawns	0.03	0.032	0.0002	0.001	0.02	0.12

Walsh, D. P. (2012). US Department of the Interior, US Geological Survey.

Example: CWD ecological groups defined by demographic and behavioral attributes

Demographic

- Male
- Female
- Adult
- Subadult

Behavioral

- Live animals presenting diseas e signs
- Road-kill
- Hunter harvest

Table 1. Estimated mean hazard ratios, associated standard deviations, Markov Chain Monte Carlo (MCMC) errors, and quantiles for various demographic groups of mule deer (*Odocoileus hemionus*) based on samples collected from chronic wasting disease-positive areas in Colorado during 2003–6.

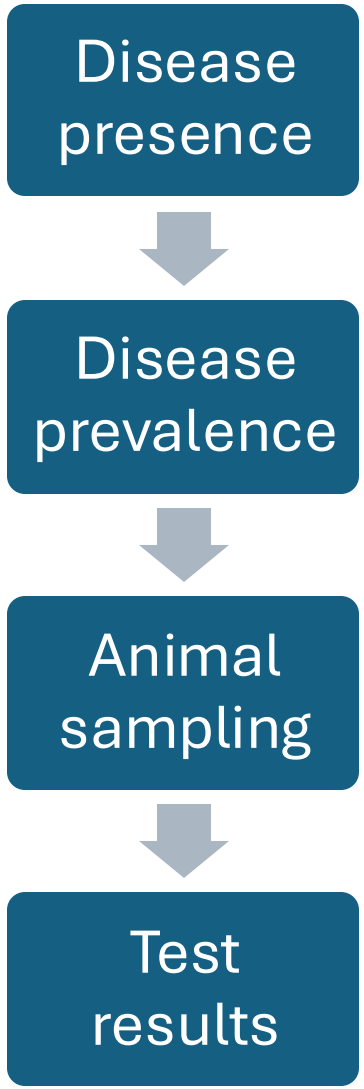
Group	Mean	Standard deviation	Markov Chain Monte Carlo error	2.50 percent	Median	97.50 percent
Suspect female	14.13	2.390	0.0151	9.90	13.97	19.24
Suspect male	12.19	2.070	0.0135	8.50	12.06	16.64
Other	1.93	0.245	0.0018	1.49	1.92	2.44
Harvest—adult males	1.00	NA	NA	NA	NA	NA
Harvest—adult females	0.57	0.065	0.0005	0.46	0.57	0.71
Harvest—yearling females	0.44	0.150	0.0009	0.20	0.43	0.78
Harvest—yearling males	0.25	0.077	0.0004	0.12	0.24	0.43
Harvest—fawns	0.03	0.032	0.0002	0.001	0.02	0.12

Walsh, D. P. (2012). US Department of the Interior, US Geological Survey.

Statistics for Detection

- Model assumptions
  - Likely disease prevalence range
  - Constant prevalence during sampling
  - Sample size small relative to population
  - Equally likely chance of being sampled

- Inputs
  - Disease presence chance (Prior)
  - Potential prevalence if present (Prior)
  - Sample size
  - Test sensitivity/specificity
  - Test results (i.e., num. pos, Data analysis)
    …or… True prevalence (Power study)

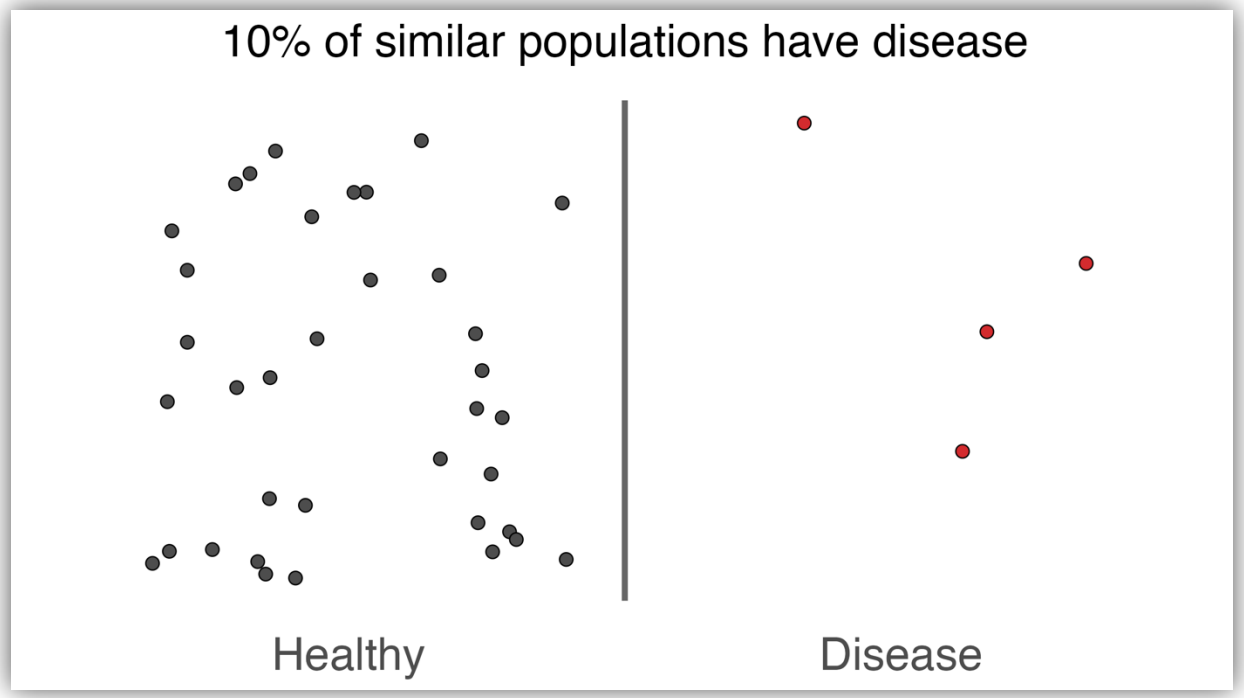


Risk as Disease prior

- Disease prior quantifies:

  Probability at least one individual in the population has infection

- Could compare to other pop's.
  - Location
  - Density
  - Age/sex structure
  - Spillover potential

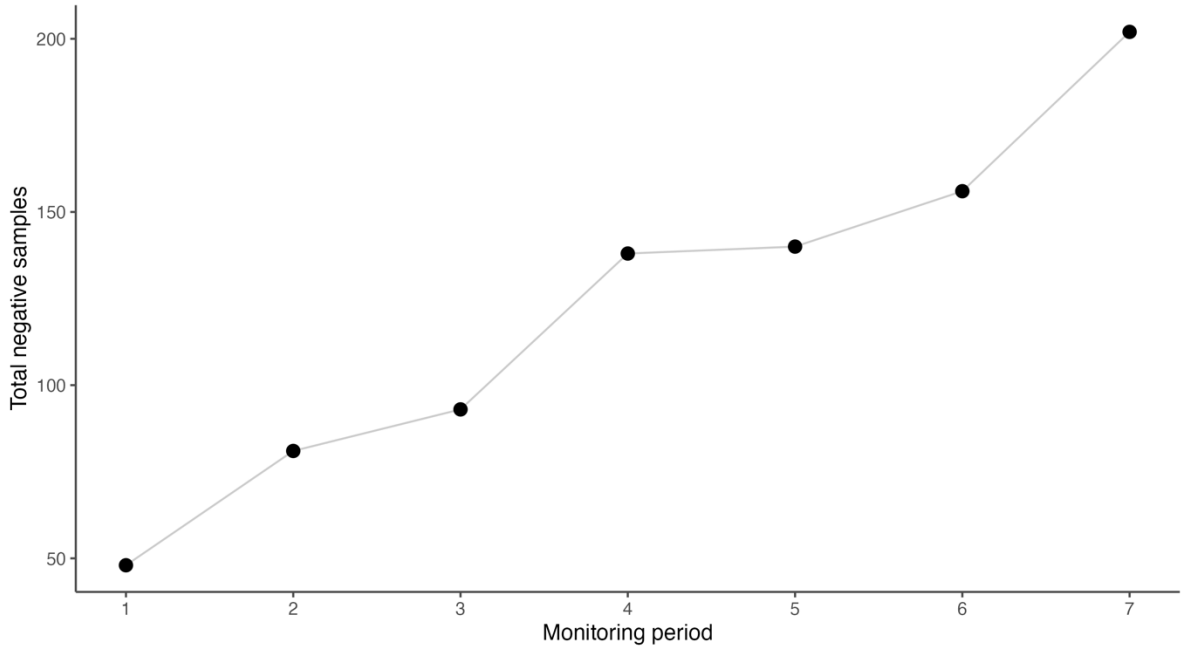


Evidence as Disease prior

- Disease prior quantifies:

  Probability evidence gathered so far indicates disease freedom

- Could use standard disease freedom methods to quantify
  - Long-term sampling increases confidence
  - Introduction risk downweights old data



Financial cost as Disease prior

- Disease prior quantifies:
  Potential impact of mistakes

- Compares cost estimates for
  - Surveillance
  - Management if disease found
  - Impact of not finding disease
  - Etc.

* May be hard!

Total $ =

Surveillance +

Mitigation x Detect prob. +

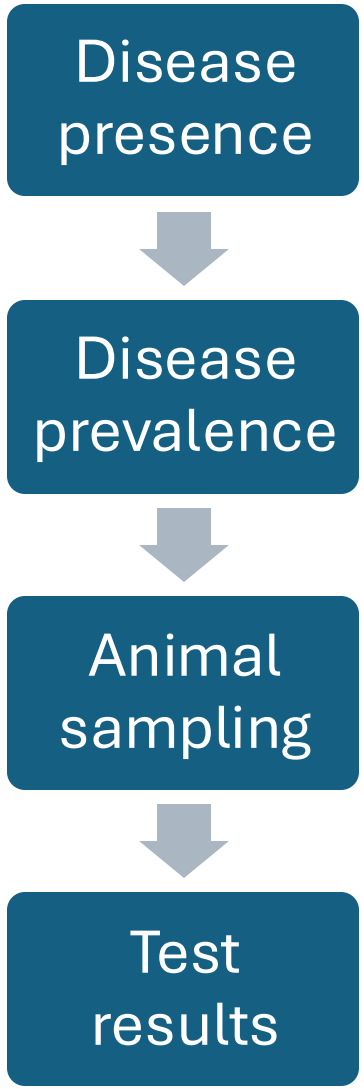
“Do nothing” x (1 – Detect prob.)

- Detect prob. depends on
  - Disease prior
  - Sampling
  - Truth (so, run sim’s.)

Statistics for Detection

- Model assumptions
  -  Likely disease prevalence range
  -  Constant prevalence during sampling
  -  Sample size small relative to population
  -  Equally likely chance of being sampled

- Inputs
  - Disease presence chance (Prior)
  - Potential prevalence if present (Prior)
  - Sample size
  - Test sensitivity/specificity
  - Test results (i.e., num. pos, Data analysis)
    ...or... True prevalence (Power study)

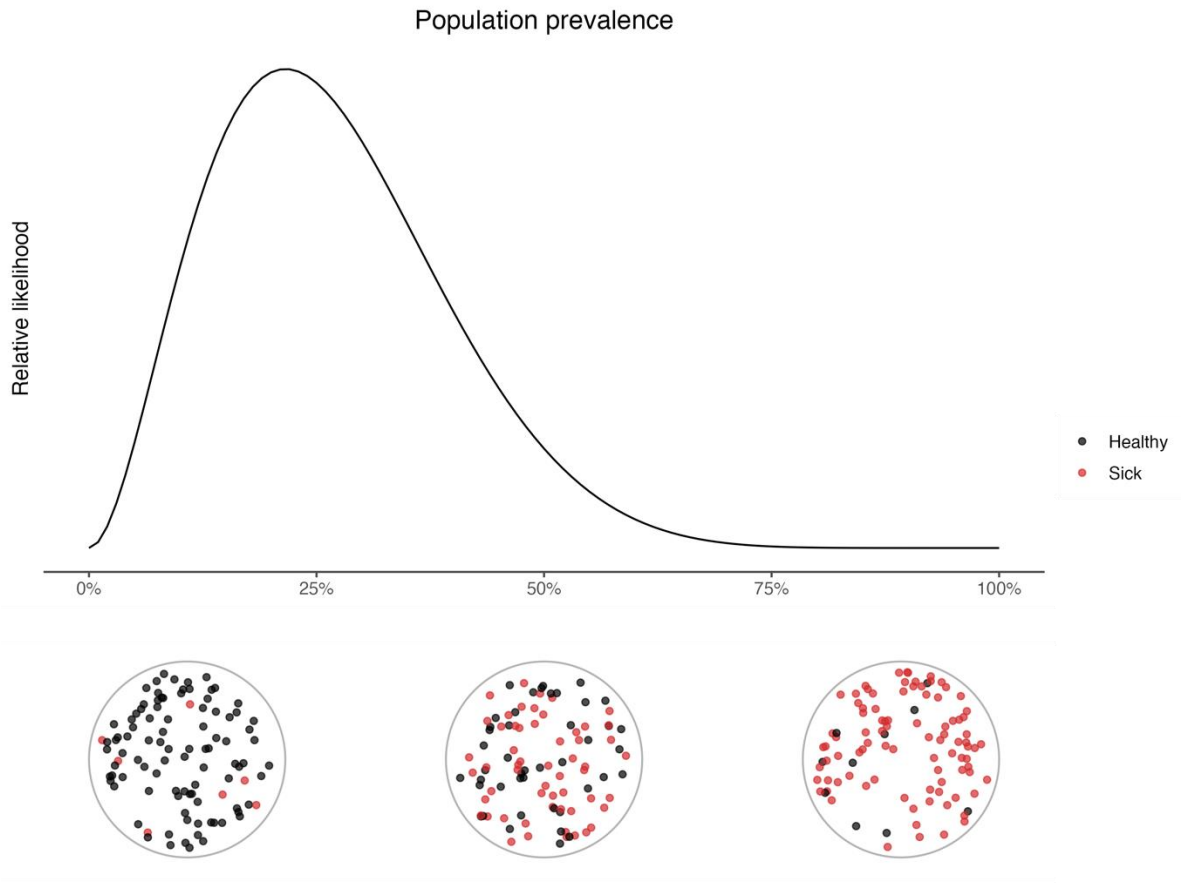


Prevalence prior

- Prevalence prior quantifies:
  Likely prevalence in population

- Could compare to other pop's.
  - Location
  - Density
  - Age/sex structure

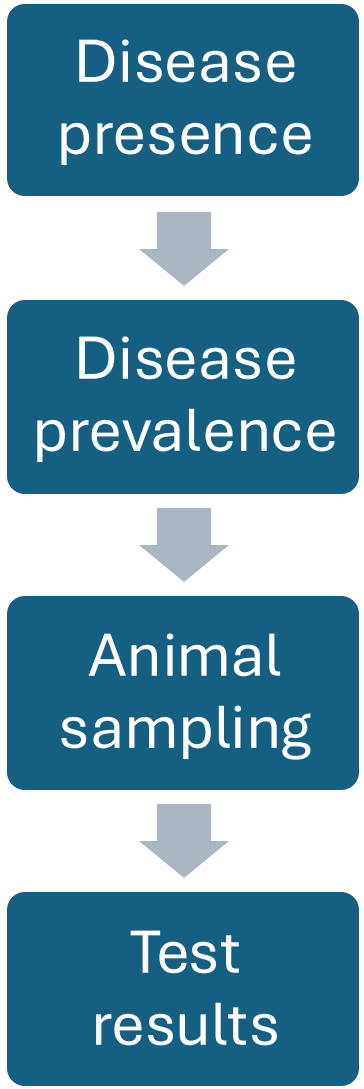
- Can impact analytic results
  - Sample more to reduce impacts!
  - Poor choices can bias analyses
  - Flat/uninformative prior has implications too



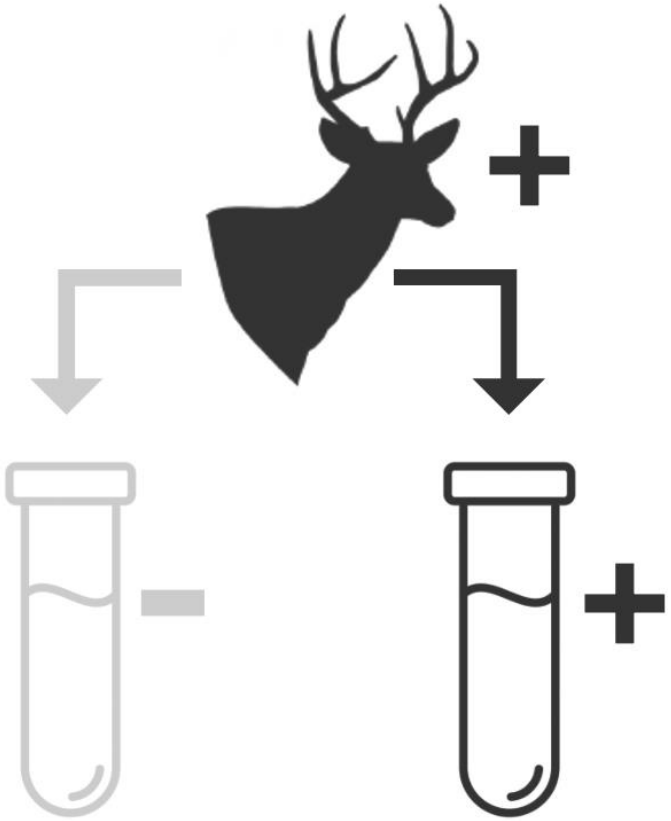
Statistics for Detection

- Model assumptions
  - Likely disease prevalence range
  - Constant prevalence during sampling
  - Sample size small relative to population
  - Equally likely chance of being sampled

- Inputs
  - Disease presence chance (Prior)
  - Potential prevalence if present (Prior)
  - Sample size
  - Test sensitivity/specificity
  - Test results (i.e., num. pos, Data analysis)
    …or… True prevalence (Power study)

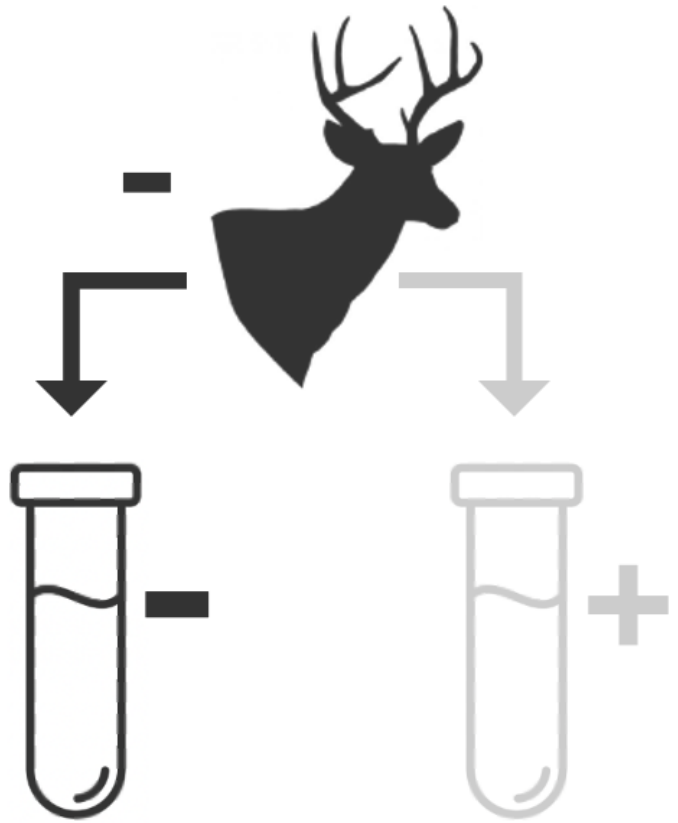


Sensitivity vs.



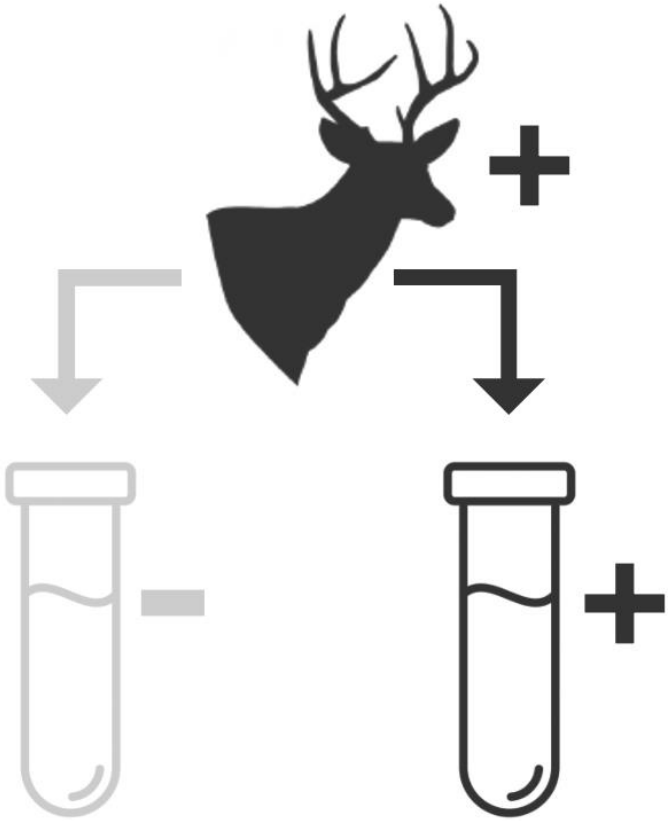
e.g., 95%

vs. Specificity

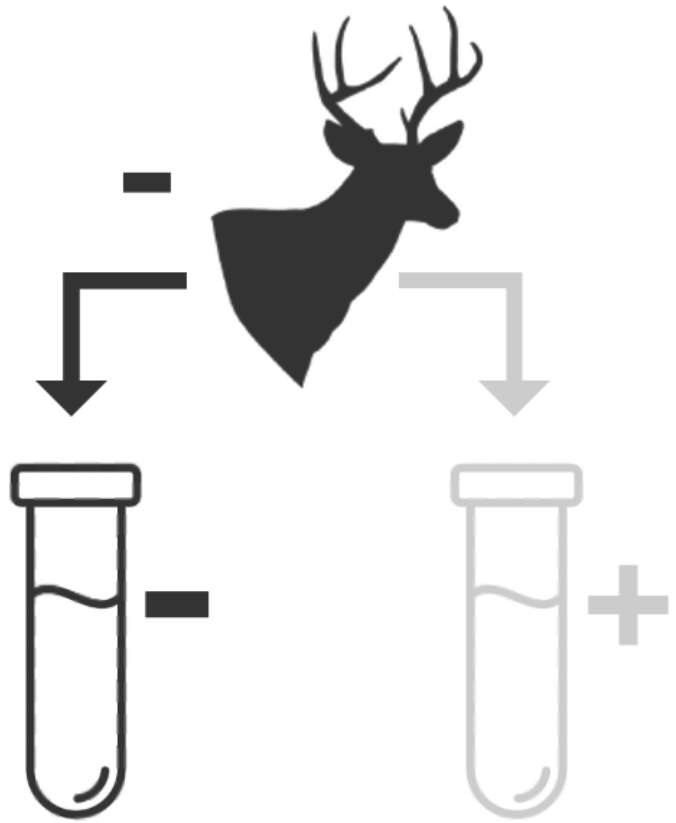


e.g., 95%

Sensitivity vs. Specificity



e.g., 95%

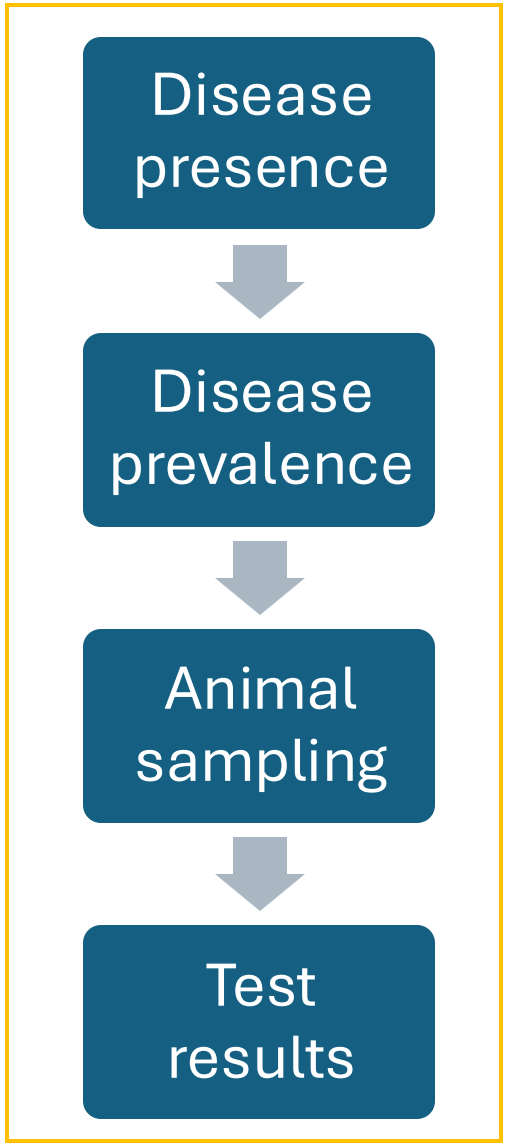


e.g., 95%

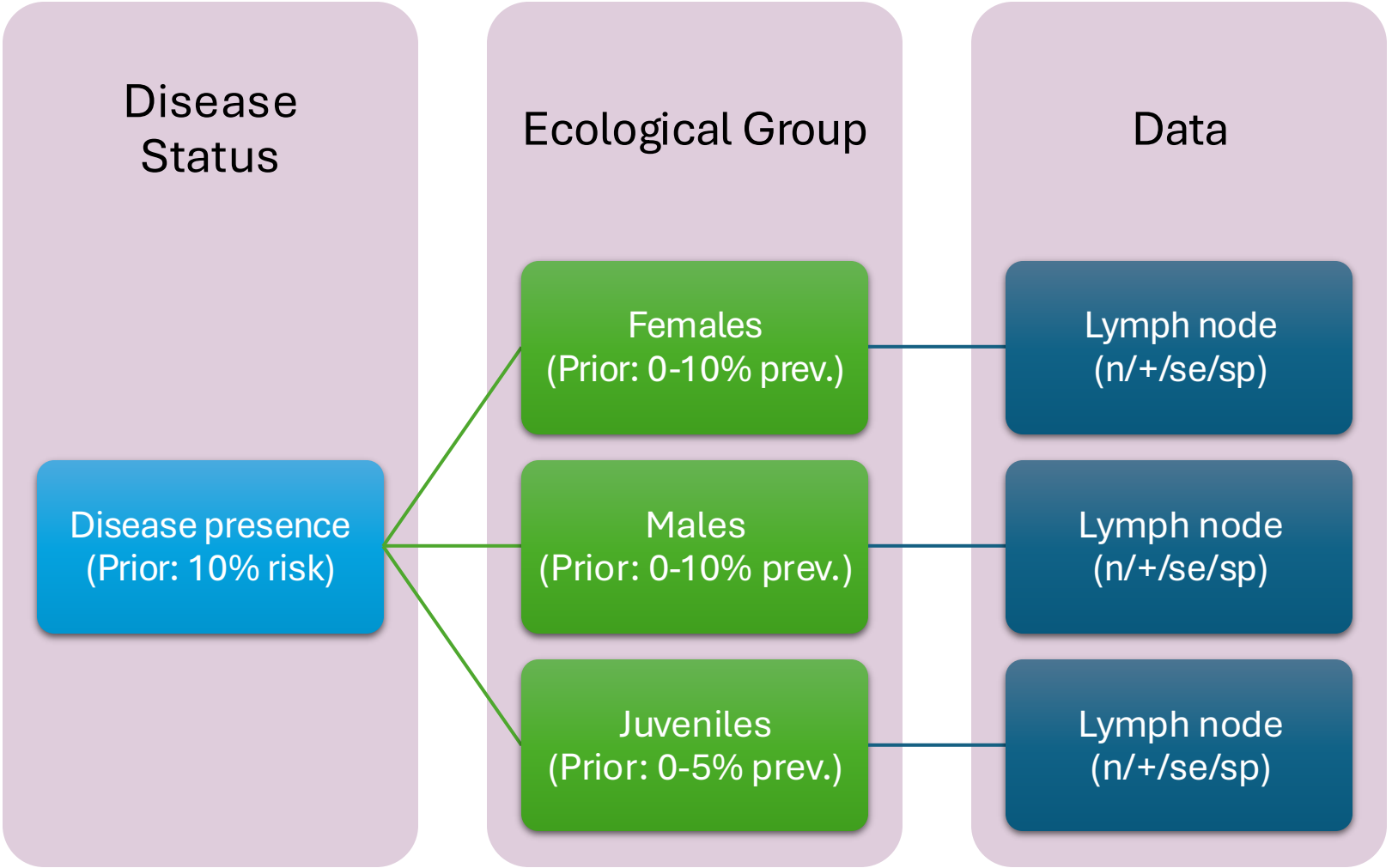
Statistics for Detection

- Model assumptions
  - Likely disease prevalence range
  - Constant prevalence during sampling
  - Sample size small relative to population
  - Equally likely chance of being sampled

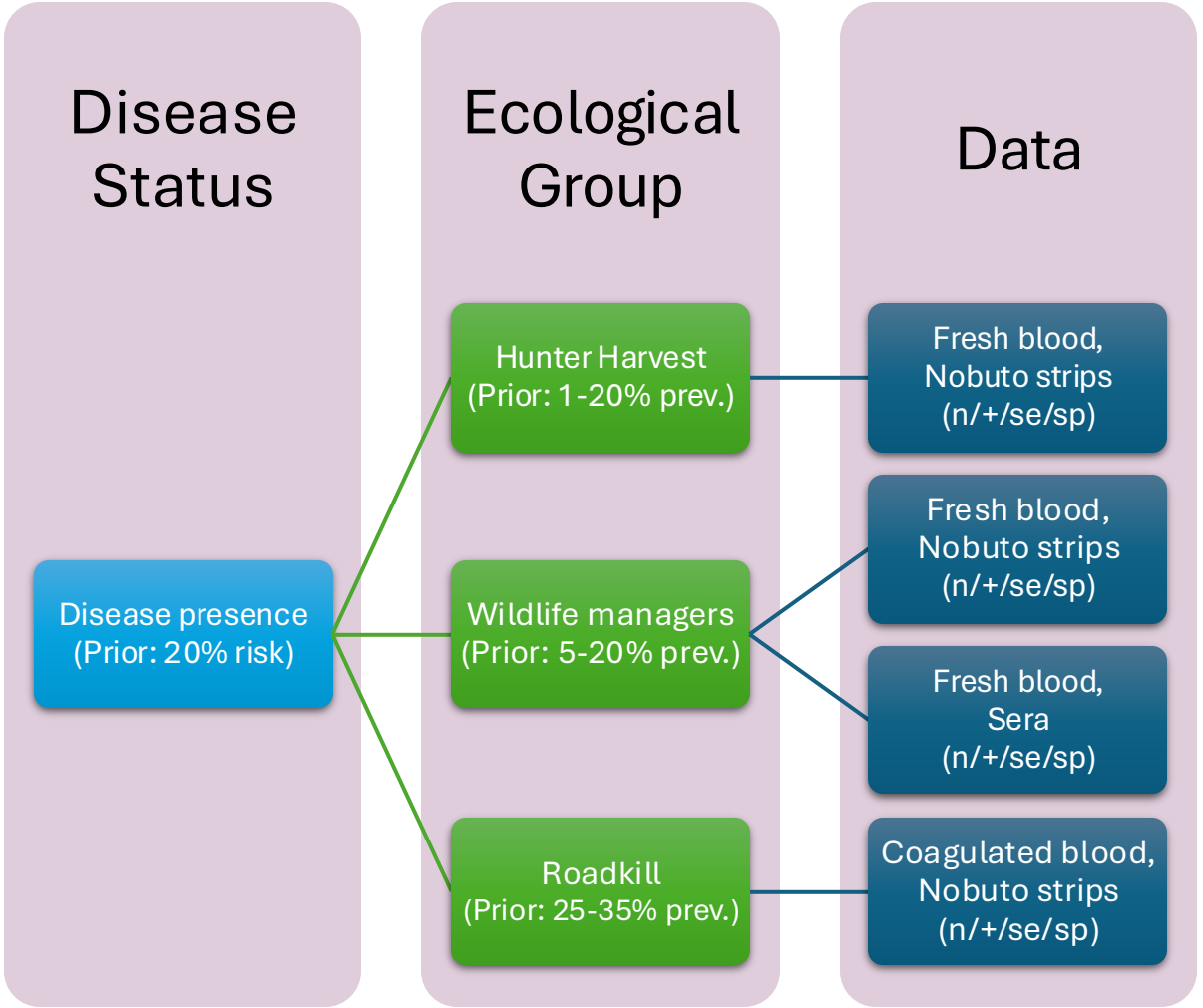
- Inputs
  - Disease presence chance (Prior)
  - Potential prevalence if present (Prior)
  - Sample size
  - Test sensitivity/specificity
  - Test results (i.e., num. pos, Data analysis)
    ...or... True prevalence (Power study)



Example Biological model structure



Example Sampling bias model structure



App link:

Surveillance Analysis and Sample Size Explorer (SASSE)

Use this tool to explore sample size needs and data analyses as they relate to common wildlife disease surveillance questions.

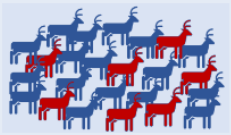
Detection

Is there disease in a population?

Enter module

Prevalence

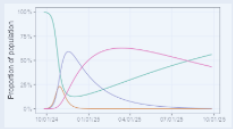
How much disease is in a population?



Enter module

Epidemiological Dynamics

How fast does disease spread through a population? How long does disease persist in a population?



Enter module

Spatial Distribution

How should we divide sampling effort across sites?

Enter module

Detection

Exercises

Click (Detection analysis tool)

- What is the probability the population is free from disease if
  - 30% of populations tend to have disease
  - Infected groups tend to have 1-15% prevalence
  - You have 0 positive test results from 30 samples from this population
  - The diagnostic test has 70% sensitivity and 100% specificity

Click (Detection analysis tool)

- What is the probability the population is free from disease if
  - 30% of populations tend to have disease
  - Infected groups tend to have 1-15% prevalence
  - You have 0 positive test results from 30 samples from this population
  - The diagnostic test has 70% sensitivity and 100% specificity

- Answer:
  - 92%, with prevalence not likely to exceed 5% otherwise
  - Data show strong evidence that there is either no disease, or very low disease

Change (Detection analysis tool)

- How many samples do you need to get 99% probability the population is free from disease if
  - 30% of populations tend to have disease
  - Infected groups tend to have 1-15% prevalence
  - You have 0 positive test results
  - The diagnostic test has 70% sensitivity and 100% specificity

Change (Detection analysis tool)

- How many samples do you need to get 99% probability the population is free from disease if
  - 30% of populations tend to have disease
  - Infected groups tend to have 1-15% prevalence
  - You have 0 positive test results
  - The diagnostic test has 70% sensitivity and 100% specificity

- Answer:
  - At least 79 samples

Change (Detection analysis tool)

- How does the probability change if you have 1 positive sample out of 30 total?

Change (Detection analysis tool)

- How does the probability change if you have 1 positive sample out of 30 total?

- Answer:
  - 0%

Reflect (Detection analysis tool)

- How strongly do the answers depend on sensitivity, specificity when a) there is 1 positive test result, b) 0 positive test results.

Reflect (Detection analysis tool)

- How strongly do the answers depend on sensitivity, specificity when a) there is 1 positive test result, b) 0 positive test results.

- Answer:
  - The most important observation is that specificity is highly important when there are positive test results.
    - Disease freedom probability with 1 positive increases from 0% to 87% when specificity decreases from 100% to 90%. Disease freedom probability is not substantially impacted by decreasing sensitivity.
    - With 0 positive samples, disease freedom probability is more impacted by changes to sensitivity than specificity.

Apply (Detection analysis tool)

- How strongly do the answers depend on sample size, prior for disease, and prevalence of disease? What does this mean for surveillance design?

Apply (Detection analysis tool)

- How strongly do the answers depend on sample size, prior for disease, and prevalence with disease? What does this mean for surveillance design?

- Answers:
  - Influences
    - Sample size has a big impact on disease freedom probability when specificity is less than 100%
    - Prior for disease impacts disease freedom probability more when sample size is small
  - For surveillance designs
    - Not enough samples or sampling at the right biological time mean higher probability of not finding disease when present on a landscape
    - Don’t necessarily need too many samples at any given site if detection is the key goal, based on theoretical curves
    - Test recommendations for “X samples to detect disease at Y prevalence”

Click (Detection power study tool)

- How often will a study a) declare a population to be free from disease, or b) declare a population to have disease if
  - Infected populations tend to have 0-30% prevalence
  - 30% of populations tend to have disease
  - The diagnostic test has 70% sensitivity and 100% specificity
  - Prevalence was actually 5%
  - 50 samples were collected from this population

Click (Detection power study tool)

- How often will a study a) declare a population to be free from disease, or b) declare a population to have disease if
  - Infected populations tend to have 0-30% prevalence
  - 30% of populations tend to have disease
  - The diagnostic test has 70% sensitivity and 100% specificity
  - Prevalence was actually 5%
  - 50 samples were collected from this population
- Answer:
  - a) About 12% (roughly 1 in 8 studies/populations)
  - b) About 78% (almost every study/population)

Change (Detection power study tool)

- How does the probability change with sample size, sensitivity, specificity, etc.?

Change (Detection power study tool)

- How does the probability change with sample size, sensitivity, specificity, etc.?

- Answer:
  - Some important observations
    i. Probability decreases with sample size, with diminishing returns
    ii. Probability depends on true prevalence, with higher probability when true prevalence is small
    iii. Increasing specificity can greatly reduce probability
    iv. Larger sample sizes required to offset low sensitivity

Reflect (Detection power study tool)

- What strategies can surveillance programs use to minimize probability of failing to detect disease?

Reflect (Detection power study tool)

- What strategies can surveillance programs use to minimize probability of failing to detect disease?

- Answer:
  - Use sample sizes that are appropriate for the potential disease burden and diagnostic assays (i.e., relative to sensitivity and specificity).
  - Can also explore potential tradeoffs in sample size by considering samples, assays, or confirmation procedures that are potentially more expensive to collect but have substantially higher sensitivity and specificity.

Within-site prevalence

Design factors

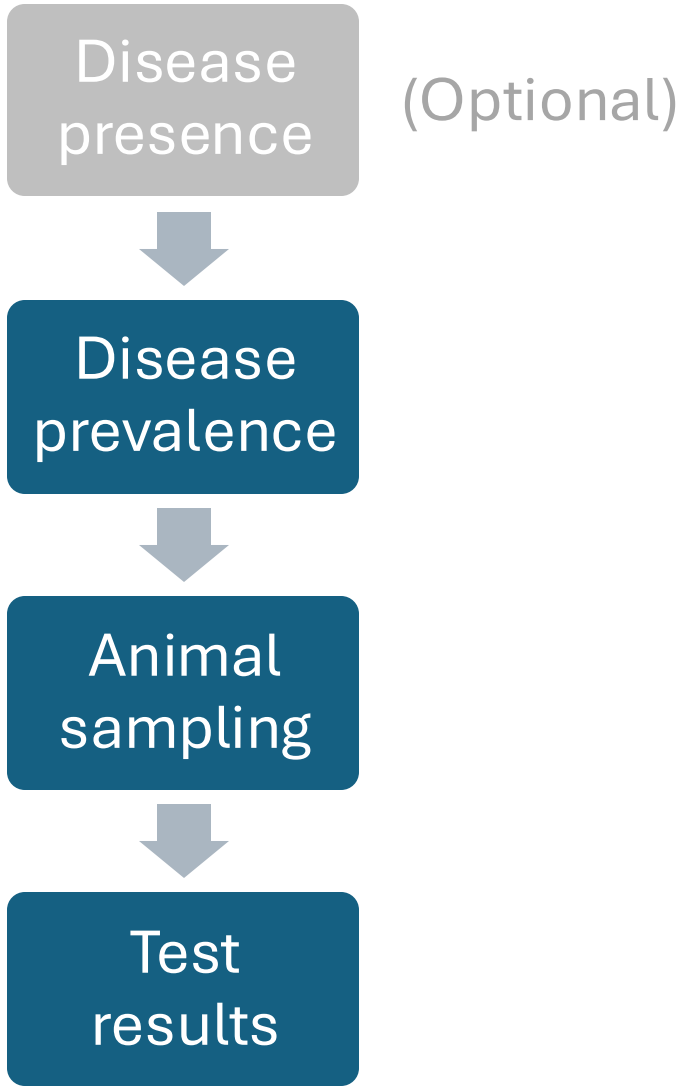
Learning goals

- Analyze data to estimate prevalence
- Decide when, where, and how many samples to collect
- Collecting too much or too little data inefficiently uses resources
  - Estimate sample sizes needed to meet precision goals
- Estimate true prevalence from observations, sample size, and test detection errors (i.e., diagnostic sensitivity and specificity)
- Design surveillance to meet potential prevalence objectives
  - Population or threshold monitoring

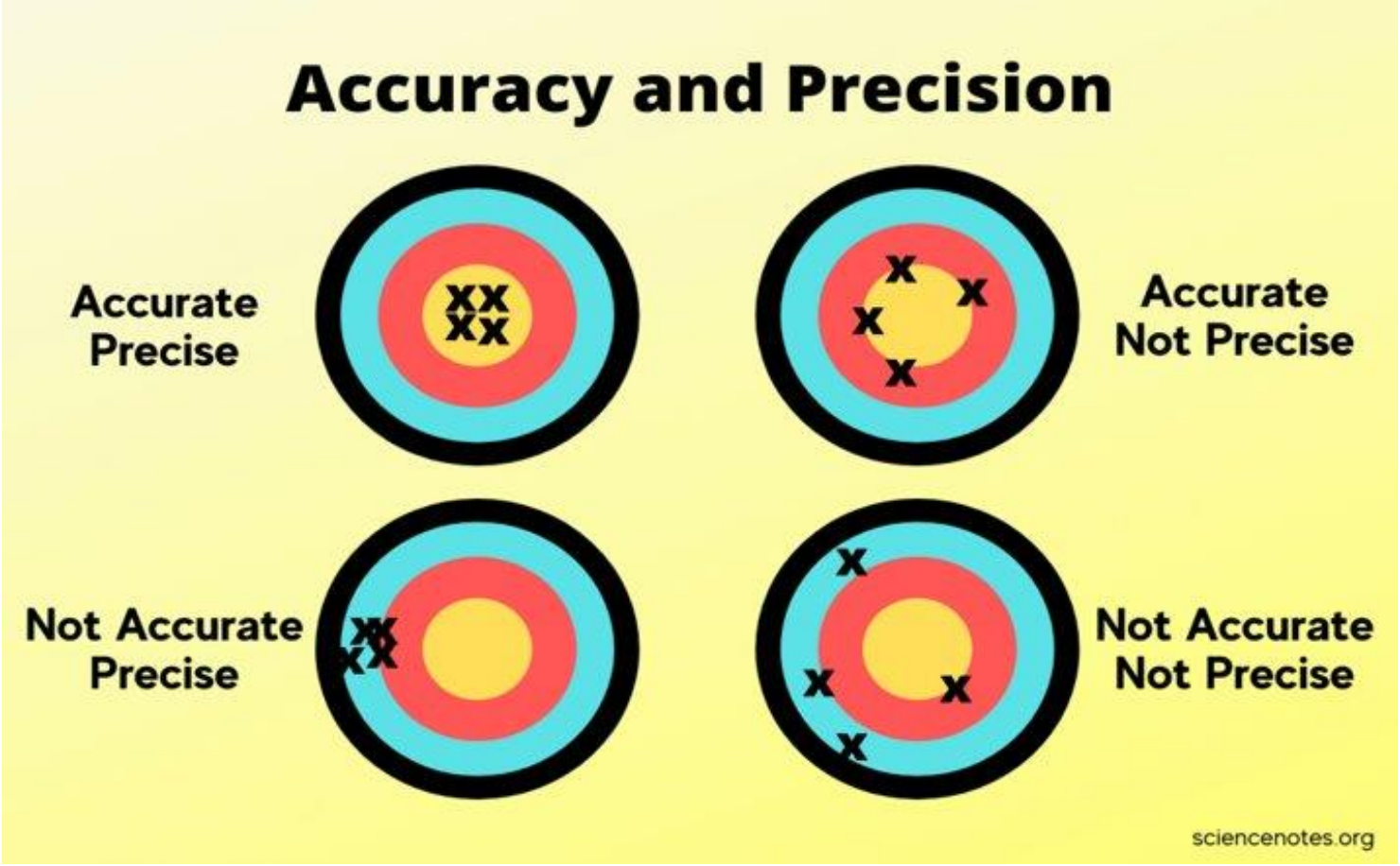
Statistics for Prevalence

- Model assumptions
  - Likely disease prevalence range
  - Constant prevalence during sampling
  - Sample size small relative to population
  - Equally likely chance of being sampled

- Inputs
  - Disease presence chance (Prior, optional)
  - Potential prevalence if present (Prior)
  - Sample size
  - Test sensitivity/specificity
  - Test results (i.e., num. pos, Data analysis)
    …or… True prevalence (Power study)



Outputs characterize



Statistics for Prevalence

- Outputs
  - Sample interpretation (Data analysis)
    - Prevalence estimate
      - 95% posterior credible interval for prevalence estimate ("i.e.," Bayesian confidence interval) **(i.e., precision)**
    - Probability for disease presence/absence (i.e., freedom, **precision**)
  - Surveillance design evaluation (Power study)
    - Sample size recommendations by studying
      - Probability credible interval "covers" true prevalence **(i.e., accuracy)**
      - Average credible interval width **(i.e., precision)**
      - (Optional) Probability design will declare presence/absence

Considerations for wildlife

- Low host density can make sampling challenging

- Ecological factors may challenge equally-likely sampling assumption
  - Non-random mixing from social structure and space use
  - Disease-induced behavior change
  - Sex or age biased capture probabilities

- Surveillance may require complex testing procedures
  - Multiple rounds of testing
    - May screen samples for additional testing
    - Positive samples may be subject to confirmation testing
  - Multiple assays per animal
    - Several tissues collected from each animal

Example: Differing CWD prevalences across ecological groups (e.g., demographic and behavioral attributes)

Demographic

- Male
- Female
- Adult
- Subadult

Behavioral

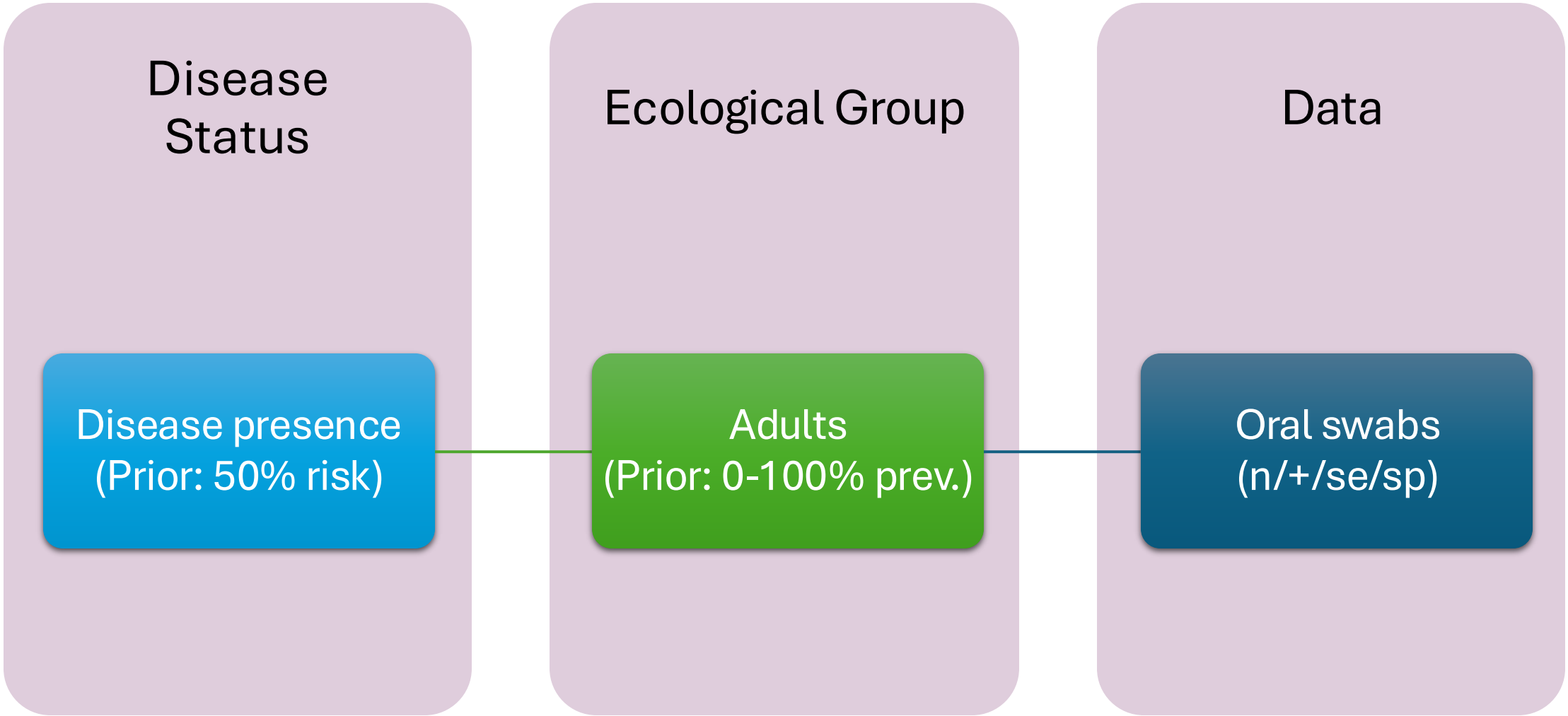
- Live animals presenting diseas e signs
- Road-kill
- Hunter harvest

Table 1. Estimated mean hazard ratios, associated standard deviations, Markov Chain Monte Carlo (MCMC) errors, and quantiles for various demographic groups of mule deer (*Odocoileus hemionus*) based on samples collected from chronic wasting disease-positive areas in Colorado during 2003–6.

Group	Mean	Standard deviation	Markov Chain Monte Carlo error	2.50 percent	Median	97.50 percent
Suspect female	14.13	2.390	0.0151	9.90	13.97	19.24
Suspect male	12.19	2.070	0.0135	8.50	12.06	16.64
Other	1.93	0.245	0.0018	1.49	1.92	2.44
Harvest—adult males	1.00	NA	NA	NA	NA	NA
Harvest—adult females	0.57	0.065	0.0005	0.46	0.57	0.71
Harvest—yearling females	0.44	0.150	0.0009	0.20	0.43	0.78
Harvest—yearling males	0.25	0.077	0.0004	0.12	0.24	0.43
Harvest—fawns	0.03	0.032	0.0002	0.001	0.02	0.12

Walsh, D. P. (2012). US Department of the Interior, US Geological Survey.

Example "Uninformative" model structure



App link:

Surveillance Analysis and Sample Size Explorer (SASSE)

Use this tool to explore sample size needs and data analyses as they relate to common wildlife disease surveillance questions.

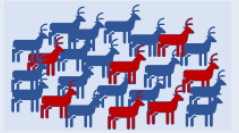
Detection

Is there disease in a population?

Enter module

Prevalence

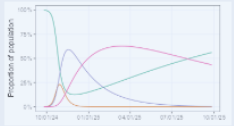
How much disease is in a population?



Enter module

Epidemiological Dynamics

How fast does disease spread through a population? How long does disease persist in a population?



Enter module

Spatial Distribution

How should we divide sampling effort across sites?

Enter module

Within-site prevalence

Exercises

Click (Prevalence analysis tool)

- What is prevalence estimated to be if
  - It is known that disease is present in the population
  - We are unsure if prevalence is currently large or small
  - You have 15 positive test results from 30 total samples collected from this population
  - The diagnostic test has 95% sensitivity and 95% specificity

Click (Prevalence analysis tool)

- What is prevalence estimated to be if
  - It is known that disease is present in the population
  - We are unsure if prevalence is currently large or small
  - You have 15 positive test results from 30 total samples collected from this population
  - The diagnostic test has 95% sensitivity and 95% specificity
- Answer:
  - Using a uniform prior for prevalence (i.e., prior prevalence range is from 0 to 1) and setting prior for disease presence equal to 1, prevalence is estimated to be 50% (95% HPD: 31-69%)

Change (Prevalence analysis tool)

- What is prevalence estimated to be if
  - It is known that disease is present in the population
  - Instead, you have strong initial beliefs or expectations that prevalence is small, say between 0% and 15%
  - You have 15 positive test results from 30 total samples collected from this population
  - The diagnostic test has 95% sensitivity and 95% specificity

Change (Prevalence analysis tool)

- What is prevalence estimated to be if
  - The population is known to have disease
  - Instead, you have strong initial beliefs or expectations that prevalence is small, say between 0% and 15%
  - You have 15 positive test results from 30 total samples collected from this population
  - The diagnostic test has 95% sensitivity and 95% specificity

- Answer:
  - Prevalence is estimated to be 21% (95% HPD: 11-30%)

Reflect (Prevalence analysis tool)

- How do your estimates depend on sample size and prior beliefs?
  - Hint: Guess and check

Reflect (Prevalence analysis tool)

- How do your estimates depend on sample size and prior beliefs?
  - Hint: Guess and check

- Answers:
  - Prevalence estimates are heavily influenced by strong prior assumptions when sample size is small

Apply (Prevalence analysis tool)

- What type of bias can prior beliefs expose surveillance designs to? How can surveillance programs mitigate this bias?

Apply (Prevalence analysis tool)

- What type of bias can prior beliefs expose surveillance designs to? How can surveillance programs mitigate this bias?

- Answers:
  - Confirmation bias. Collecting too little data doesn't give practitioners the ability to adequately confirm or refute their prior beliefs or expectations.
  - Surveillance programs should explore potential consequences of confirmation bias for their needs before adopting prior beliefs.

Click (Prevalence power study tool)

- What is the coverage probability (i.e., accuracy) and average credible interval width (i.e., precision) for studies in which
  - We are unsure if prevalence is currently large or small
  - The population is known to have disease
  - The diagnostic test has 95% sensitivity and 95% specificity
  - Prevalence is actually 10%
  - 100 samples were collected

Click (Prevalence power study tool)

- What is the coverage probability (i.e., accuracy) and average credible interval width (i.e., precision) for studies in which
  - We are unsure if prevalence is currently large or small
  - The population is known to have disease
  - The diagnostic test has 95% sensitivity and 95% specificity
  - Prevalence is actually 10%
  - 100 samples were collected

- Answers:
  - Credible interval coverage probability: 94%
  - Average credible interval width: 13% (i.e., +/- 6.5%)

Change (Prevalence power study tool)

- What is the coverage probability (i.e., accuracy) and average credible interval width (i.e., precision) for studies in which
  - We are unsure if prevalence is currently large or small
  - The population is known to have disease
  - The diagnostic test has 95% sensitivity and 95% specificity
  - Prevalence is actually 10%
  - More than 100 samples are collected

Change (Prevalence power study tool)

- What is the coverage probability (i.e., accuracy) and average credible interval width (i.e., precision) for studies in which
  - We are unsure if prevalence is currently large or small
  - The population is known to have disease
  - The diagnostic test has 95% sensitivity and 95% specificity
  - Prevalence is actually 10%
  - More than 100 samples are collected

- Answers:
  - Credible interval coverage probability goes toward 95% (i.e., more accurate)
  - Average credible interval width decreases (i.e., more precise)

Reflect (Prevalence power study tool)

- What strategies can surveillance programs use to maximize accuracy and precision for prevalence estimation?

Reflect (Prevalence power study tool)

- What strategies can surveillance programs use to maximize accuracy and precision for prevalence estimation?

- Answers:
  - Use sample sizes that are appropriate for the potential disease burden and diagnostic assays (i.e., sensitivity and specificity)
  - Use sample sizes large enough to achieve goals (i.e., monitor to make sure prevalence is below X%)
  - Make careful decisions about using informative priors since they can bias estimates at small sample sizes

Epidemiological Dynamics

Design factors

Learning goals

- Run a basic epidemiological model

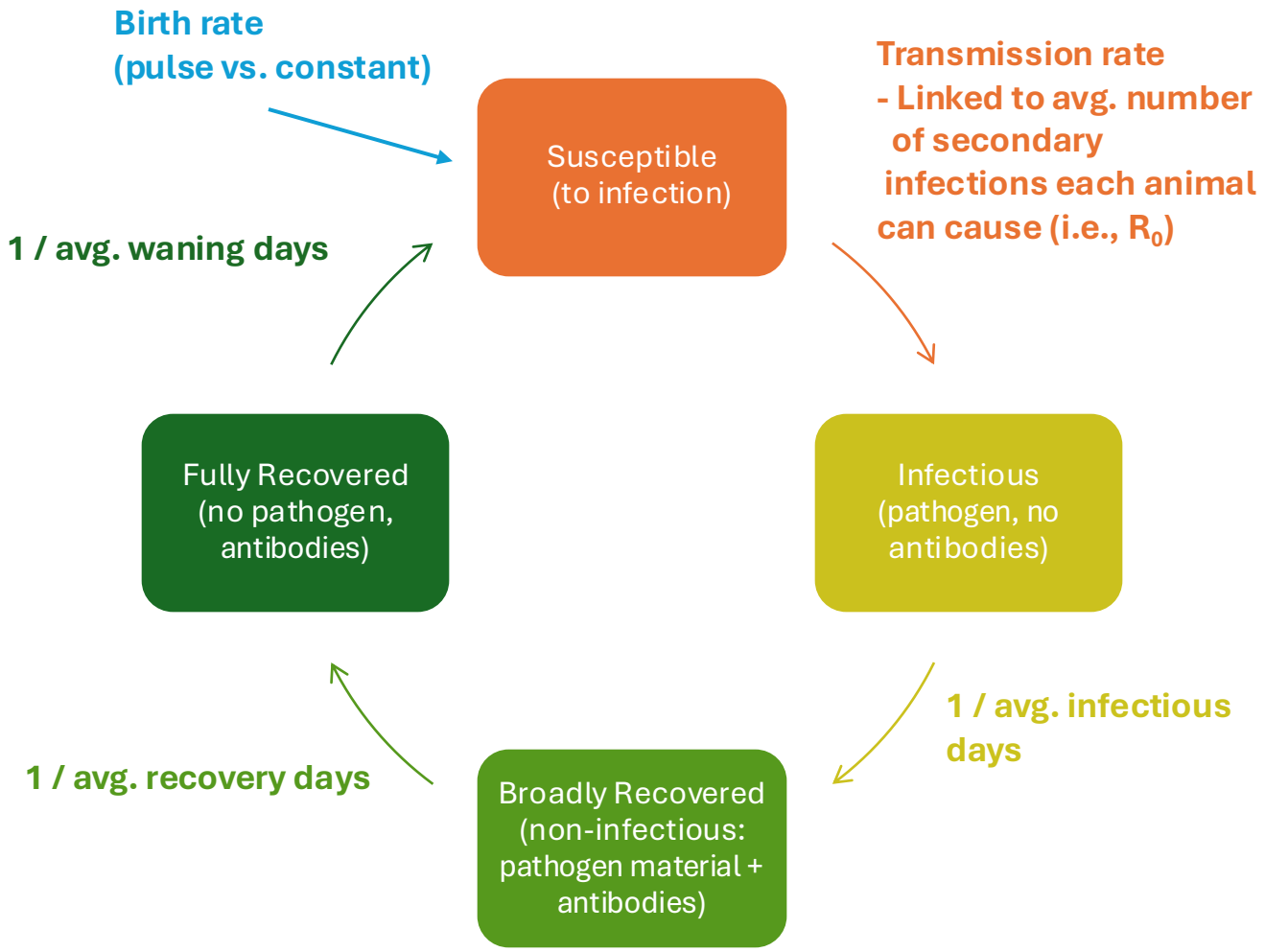
- Decide when to sample

Epidemiological dynamics

- Understand how the disease spreads in a population
- New epidemic vs endemic equilibrium
- SEIR dynamics (state transitions)
- Transmission mechanisms: direct, vector-borne, vertical, environmental, etc
- Parameters of inference: force of infection, incidence (distinguish from prevalence), R0, Re, herd immunity, infectious period, incubation period, waning immunity, timing of peak
- Assay type (paired sampling – pathogen and serology)
- Design elements: repeated cross-sectional, cohort study

Mathematics for Epidemiological dynamics

- Model assumptions
  - All animals belong to one of 4 "compartments"
  - Well-mixed population
  - Differential equation model controls time changes
  - Recruitment of new susceptible animals allowed at arbitrary times, as pulses (i.e., births)
- Inputs
  - Transition rates between compartments
  - Birth pulse sizes and timings
  - Initial compartment sizes
  - Sensitivity, specificity for pathogen and antibody tests
- Outputs
  - Proportion of population in compartments over time
  - Expected positivity rates for pathogen, antibody tests
  - No parameter estimation today
- Considerations for wildlife…



SIBR reference: Hewitt, et al. (2024) *MEE* https://www.doi.org/10.1111/2041-210X.14370

Considerations for wildlife

- Different disease systems require different compartment models, e.g.,
  - Susceptible, infectious, susceptible (SIS)
    - No immunity upon recovery
  - Susceptible, infectious, recovered (SIR)
    - Lifelong immunity, immune response immediately removes all pathogenic material
  - Susceptible, exposed, infectious, recovered (SEIR)
    - Latent, non-infectious incubation period

- Compartment models may be unimportant for endemic disease
  - Steady-state, equilibrium behavior often interpreted as endemic disease

- Can be hard to estimate parameters from surveillance data
  - Requires early, consistent longitudinal sampling
  - Parameters may vary between populations due to landscape and human factors

App link:

Surveillance Analysis and Sample Size Explorer (SASSE)

Use this tool to explore sample size needs and data analyses as they relate to common wildlife disease surveillance questions.

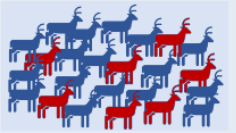
Detection

Is there disease in a population?

Enter module

Prevalence

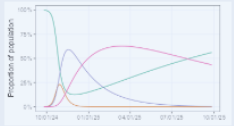
How much disease is in a population?



Enter module

Epidemiological Dynamics

How fast does disease spread through a population? How long does disease persist in a population?



Enter module

Spatial Distribution

How should we divide sampling effort across sites?

Enter module

Epidemiological Dynamics

Exercises

Click (Epidemiological Dynamics tool)

- What is the maximum proportion of the population that will test positive for a pathogen in the first three years, and when do the maximums occur if
  - Animals remain infectious for 7 days, on average
  - Immune systems remove all pathogenic material in 90 days, on average
  - Individual immunity lasts one year, on average
  - Each infectious animal infects 2.5 additional animals, on average
  - Initially, 999 individuals in the population are susceptible and 1 is infectious
  - The first infection occurred on 3/1/2021
  - 40% of total population is new, susceptible animals after spring births (i.e., 6/1)
  - Sensitivity for the pathogen test is 95%, and specificity is 99%

Click (Epidemiological Dynamics tool)

- What is the maximum proportion of the population that will test positive for a pathogen in the first three years, and when do the maximums occur if
  - Animals remain infectious for 7 days, on average
  - Immune systems remove all pathogenic material in 90 days, on average
  - Individual immunity lasts one year, on average
  - Each infectious animal infects 2.5 additional animals, on average
  - Initially, 999 individuals in the population are susceptible and 1 is infectious
  - The first infection occurred on 3/1/2021
  - 40% of total population is new, susceptible animals after spring births (i.e., 6/1)
  - Sensitivity for the pathogen test is 95%, and specificity is 99%

- Answers:
  - 69% on 4/16/2021
  - 38% on 7/25/2022
  - 8% on 1/1/2023

Change (Epidemiological Dynamics tool)

- What is the maximum proportion of the population that will test positive for a pathogen in the first three years, and when do the maximums occur if
  - Animals remain infectious for 7 days, on average
  - Immune systems remove all pathogenic material in 90 days, on average
  - Instead, individual immunity lasts 18 months, on average
  - Instead, each infectious animal infects 1.5 additional animals, on average
  - Initially, 999 individuals in the population are susceptible and 1 is infectious
  - The first infection occurred on 3/1/2021
  - 40% of total population is new, susceptible animals after spring births (i.e., 6/1)
  - Sensitivity for the pathogen test is 95%, and specificity is 99%

Change (Epidemiological Dynamics tool)

- What is the maximum proportion of the population that will test positive for a pathogen in the first three years, and when do the maximums occur if
  - Animals remain infectious for 7 days, on average
  - Immune systems remove all pathogenic material in 90 days, on average
  - Instead, individual immunity lasts 18 months, on average
  - Instead, each infectious animal infects 1.5 additional animals, on average
  - Initially, 999 individuals in the population are susceptible and 1 is infectious
  - The first infection occurred on 3/1/2021
  - 40% of total population is new, susceptible animals after spring births (i.e., 6/1)
  - Sensitivity for the pathogen test is 95%, and specificity is 99%

- Answers:
  - 33% on 5/30/2021
  - 28% on 7/12/2021
  - 6% on 1/1/2022
  - 1% on 2/8/2023

Reflect (Epidemiological Dynamics tool)

- How can epidemiological dynamics potentially impact surveillance plans?

Reflect (Epidemiological Dynamics tool)

- How can epidemiological dynamics potentially impact surveillance plans?

- Answers:
  - Seemingly small changes to disease parameters can greatly change disease dynamics. Instead of a regular annual peak, multi-year patterns can be more dispersed in timing and scale.
  - Estimating epidemiological parameters is important, but challenging. Small changes to parameter estimates (or model structure) can have similar impacts.
  - Surveillance plans may need to adapt timing and effort from year to year
  - Maximum test positivity may not coincide with maximum infectiousness

Apply (Epidemiological Dynamics tool)

- What parameters have the greatest impact on peak disease timing and consistency? How can you design a surveillance system to accommodate the impacts?

Apply (Epidemiological Dynamics tool)

- What parameters have the greatest impact on peak disease timing and consistency? How can you design a surveillance system to accommodate the impacts?

- Answers:
  - Immunity duration, transmission rates, and birth pulse timings most impact peak disease timing and consistency
  - Surveillance plans that ensure coverage at a range of potential minimum and maximum times should help ensure adequate monitoring at all epidemic stages

Seasonality

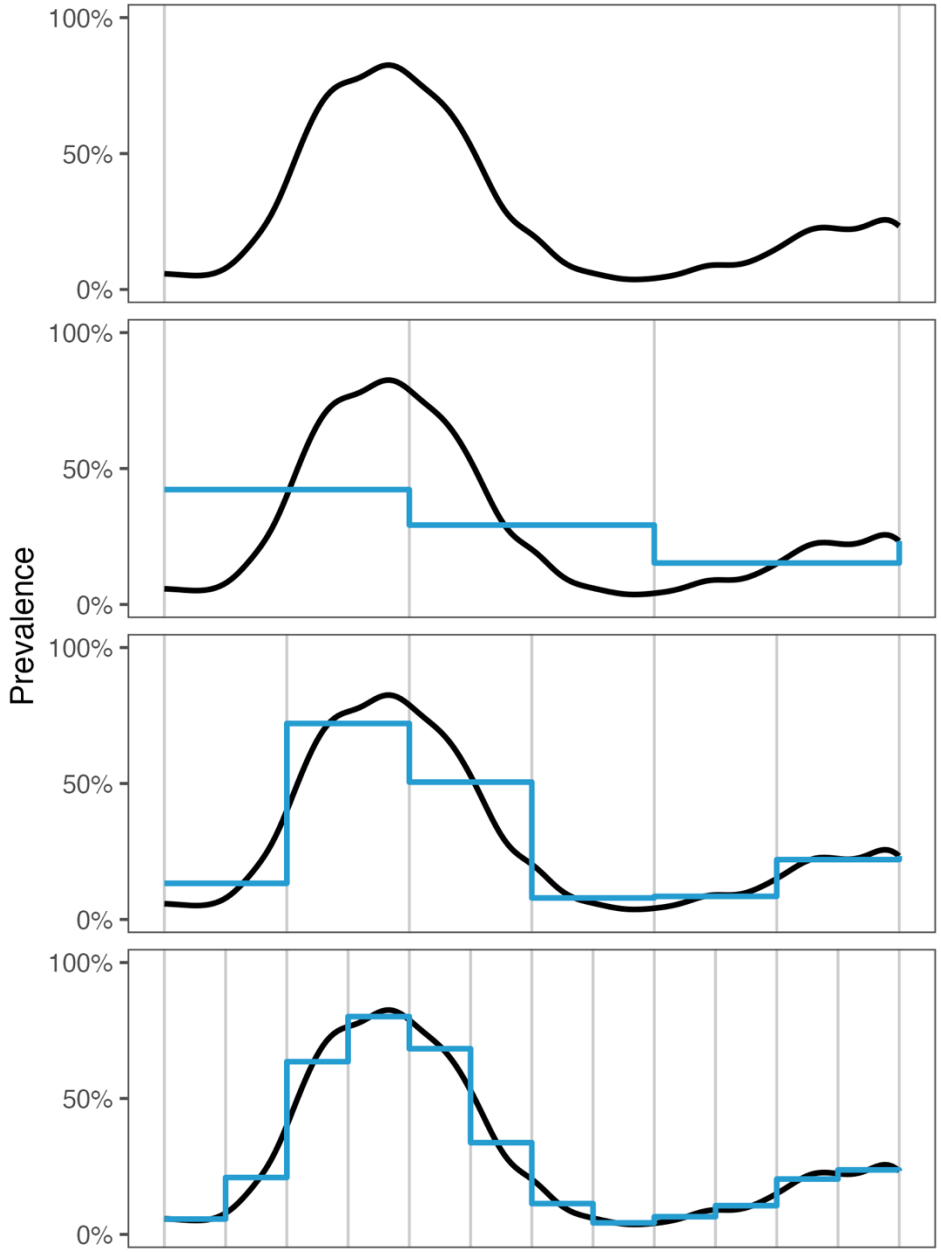
Design factors

Statistics for Seasonality

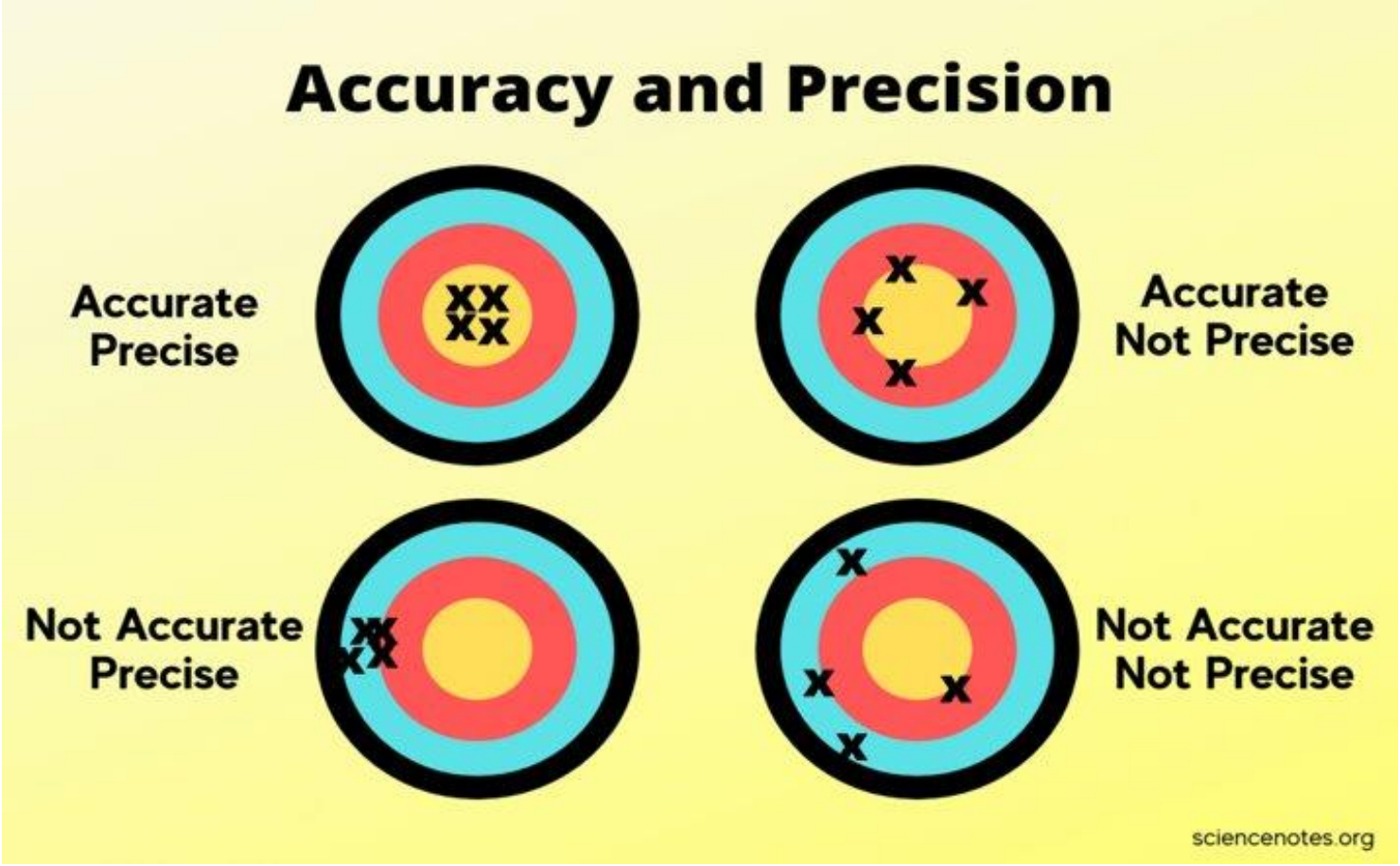
- Time can change ecological groups
  - “Seasons” can approximate fine-scale change

- Model assumptions (For each season/period)
  - Likely disease prevalence range
  - Constant prevalence during sampling
  - Sample size small relative to population
  - Equally likely chance of being sampled

- Inputs (For each season/period)
  - Disease presence chance (Prior, optional)
  - Potential prevalence if present (Prior)
  - Sensitivity, Specificity, Sample size, etc.

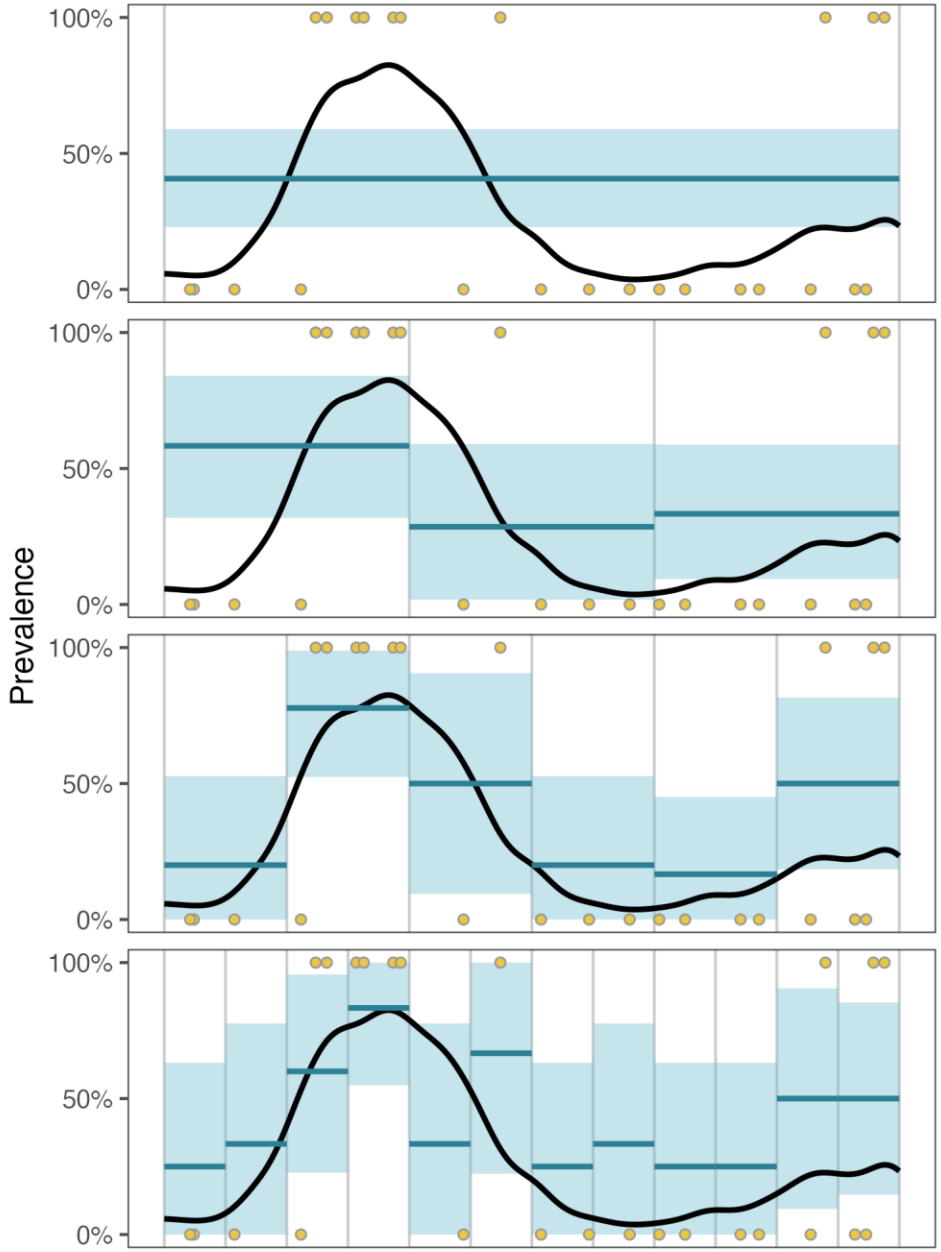


Outputs characterize



Statistics for Seasonality

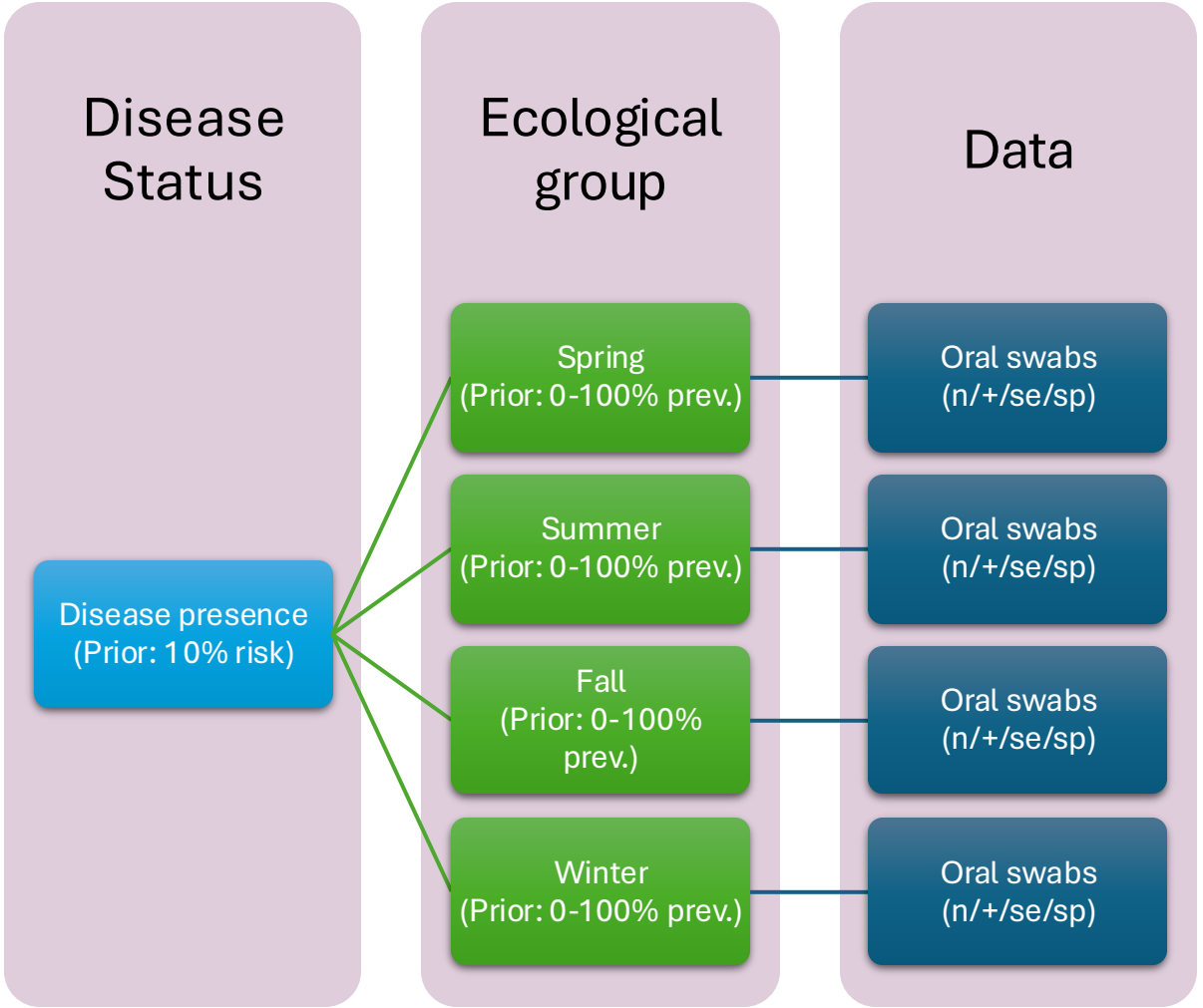
- Outputs (For each season/period)
  - Sample interpretation (Data analysis)
    - Prevalence estimate
    - 95% posterior credible interval for prevalence estimate
      ("i.e.," Bayesian confidence interval)
    - (Optional) Probability for disease presence/absence (i.e., freedom)
  - Surveillance design evaluation (Power study)
    - Probability credible interval "covers" true prevalence
    - Average credible interval width
    - (Optional) Probability design will declare presence/absence



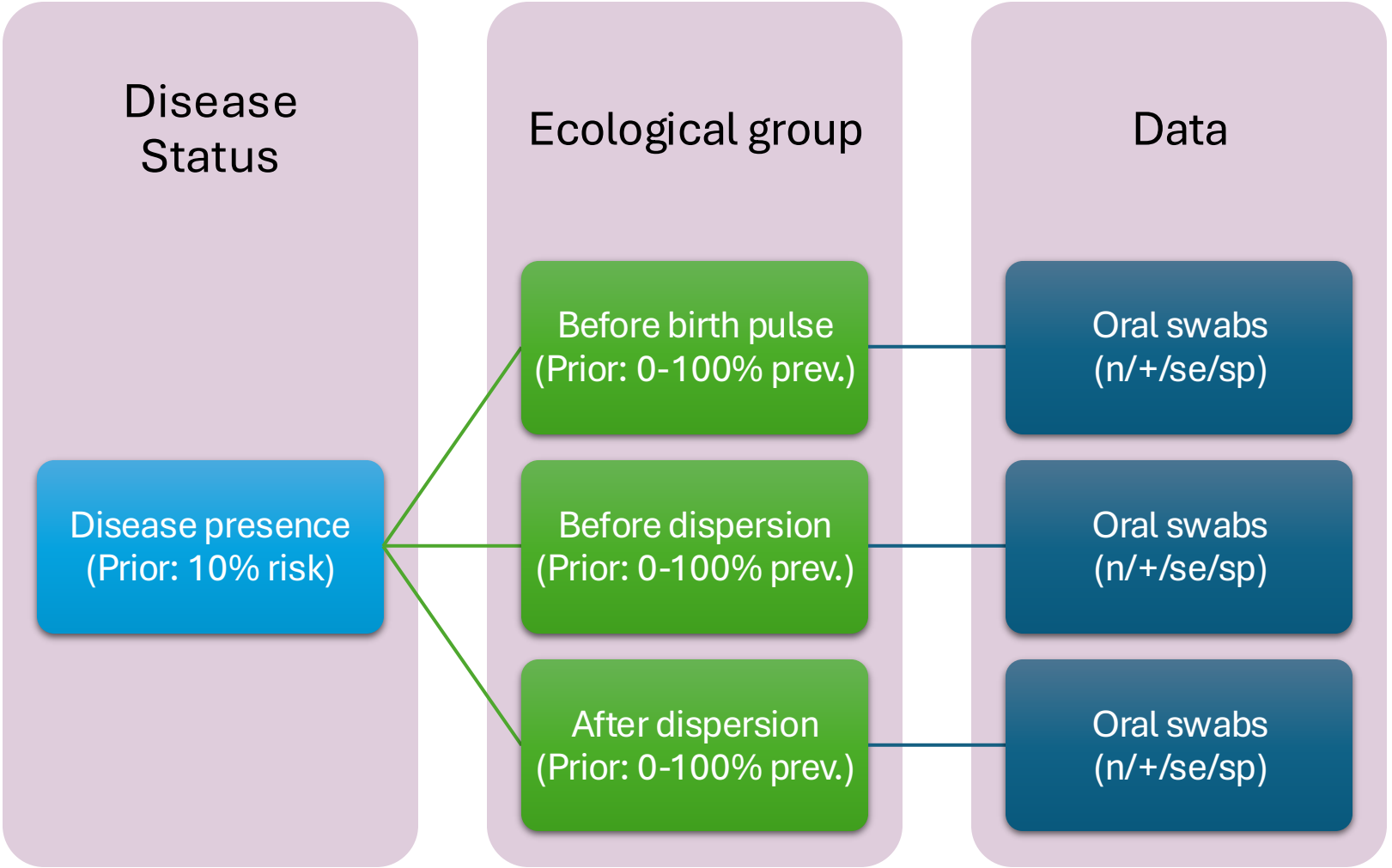
Statistics for Seasonality

- Considerations for wildlife
  - Time between sampling periods
    - Wait longer than incubation + infectious period to see population-level changes
    - Sample different life-history/seasonal stages of the host
      - E.g., before and after birth pulse times, before and after dispersal
  - The “copy/paste within site ideas” approach presented here is a starting point
    - Can do more advanced seasonal analysis with quantitative collaborators
    - Advanced analyses have similar data needs
    - Advanced analyses “smooth” the individual seasonal estimates
  - Tradeoff between number of seasons and sampling effort/precision

Example Seasonal model structure



Example Social/Biological model structure



Seasonality

Exercises

Challenge

Opportunistic sampling of deer in your area reveals that SARS-CoV-2 is starting to infect deer. No cases of SARS-CoV-2 in deer have previously been detected. Your first detection comes from a deer sampled on 3/1/2021. You know from experimental infection studies that the SARS-CoV-2 **infectious period in deer is about 7 days**, **non-infectious RNA fragments can persist 60 days on average, immunity waves over 365 days on average**, **and R0 is estimated to be 1.92**. Data from other states suggests that different demographic groups are at different risks for infection: **prevalences tend to be higher in adults and males**, compared to subadults and females. Deer in this area undergo rut from late September to early December, and **parturition from mid May to late June, so 40% of the population is new and susceptible after 7/1**. Based on your budget, you can collect **1000 samples** in your state over the next year. You have unlimited access to harvested, culled, or live deer at any point in the year.

- How would you design a sampling program in your state to learn about temporal changes in disease prevalence in deer in your state?
- How would the differing risk among demographic groups factor into your sampling design?
- Hint: Use the Epidemiological Dynamics tool to explore simulated peaks in infection over time, followed by the Prevalence Power Study tool to assess sample size needs

Challenge

- Answers:
  - Epidemiological Dynamics tool can help explore when prevalence spikes following an outbreak and following the birth pulse
    - Might want to consider pre- vs. post-birth pulse ecological groups
    - Might also want to consider where peaks are simulated to occur throughout the year, consider when they are occurring and how long they occur to plan sampling
  - Prevalence Power Study tool shows prevalence can be estimated to within approx. +/- 9% with a sample size of 100 when assuming disease is present, using uninformative priors and a test with 95% sensitivity and 100% specificity when true prevalence is 50% or below, or +/- 6.5% when true prevalence is 15%
    - Could potentially use 1,000 samples to study 10 different ecological groups
    - Ecological groups cover sampling periods, may need to chose what you prioritize samples to describe temporal vs demographic trends
  - Detection power study tool shows presence can likely be established with 25-50 samples even if true prevalence is as low as 5%
    - Could potentially quickly survey for outbreak areas, then commit more resources to areas of active outbreak

Spatial distribution

Design factors

Learning goals

- How to distribute samples across space when the objective is to identify where the disease occurs

- Collecting too much or too little data can inefficiently use resources

Statistics for Spatial distribution

- Space can change ecological groups
  - Each site may represent different animals
  - Re-use within site ideas for detection, prevalence

- Model assumptions (For number of sites)
  - Constant detection probability across sites
  - Presence/absence at each site
  - Site populations are independent (e.g., don't mix)

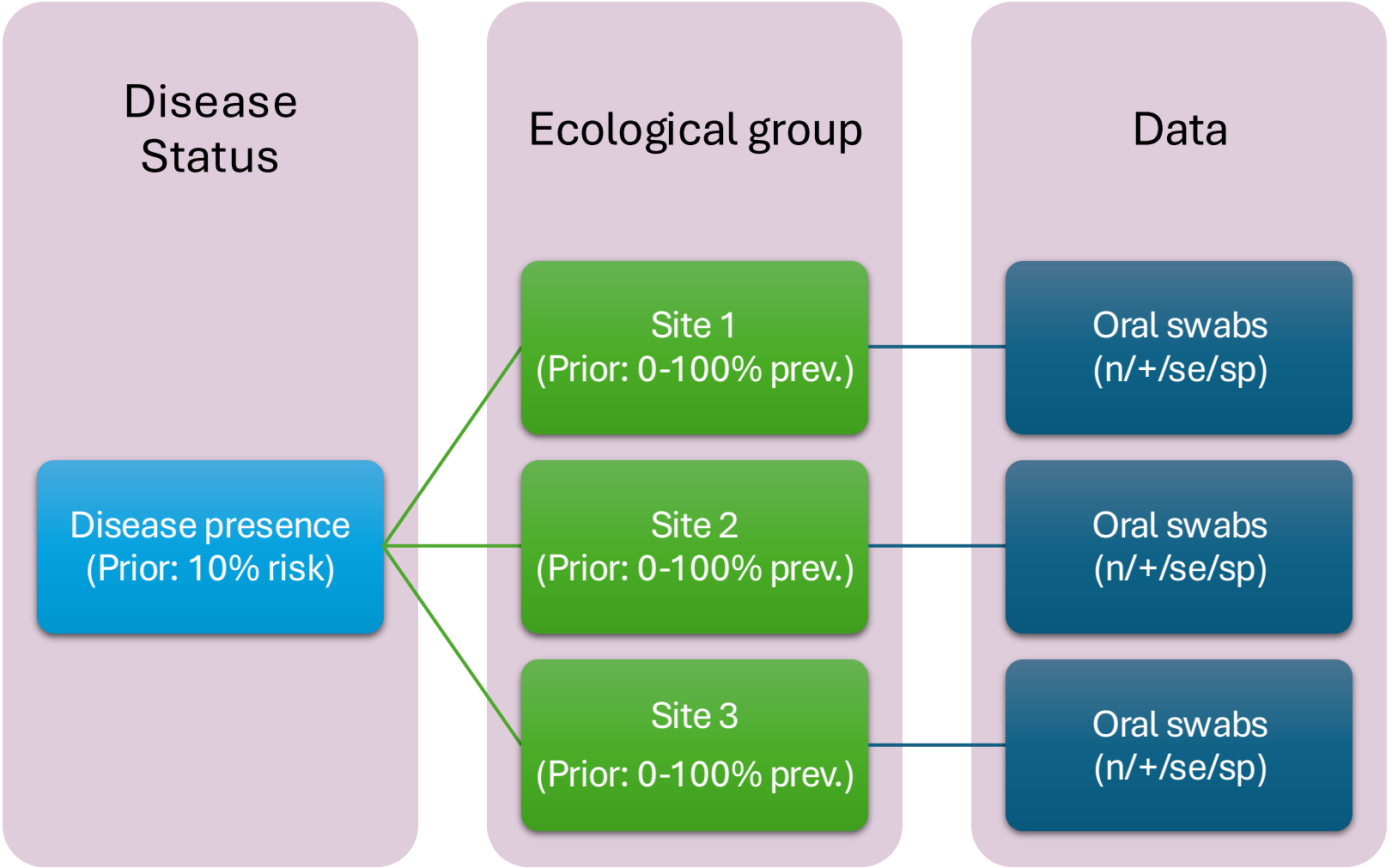
- Inputs
  - Number of sites
  - Detection probability for sites

- Outputs
  - Probability that no sites will have detections
  - Probabilities that at least X sites will have detections

Statistics for Spatial distribution

- Considerations for wildlife
  - Space between sampling locations
    - Sites should be spaced far enough apart that populations generally don't mix (i.e., counter spatial autocorrelation to get independent samples)
    - Sample sites with different risk factors
      - E.g., urban/rural human populations, forested vs. open, wet vs. dry
  - The "copy/paste within site ideas" approach presented here is a starting point
    - Can do more advanced spatial analysis with quantitative collaborators
    - Advanced analyses have similar data needs
    - Advanced analyses interpolate between the individual location estimates
  - Tradeoff between number of sites and sampling effort/precision
    - Also, surveillance methods might differ between site (i.e., sample type, roadkill, hunters, managers)
    - Sample size/effort should be tailored to each site, following detection probability

Example "Spatial" model structure



App link:

Surveillance Analysis and Sample Size Explorer (SASSE)

Use this tool to explore sample size needs and data analyses as they relate to common wildlife disease surveillance questions.

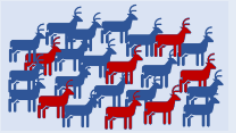
Detection

Is there disease in a population?

Enter module

Prevalence

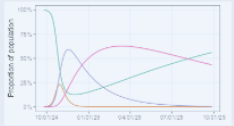
How much disease is in a population?



Enter module

Epidemiological Dynamics

How fast does disease spread through a population? How long does disease persist in a population?



Enter module

Spatial Distribution

How should we divide sampling effort across sites?

Enter module

Spatial distribution

Exercises

Click (Spatial distribution activity)

- You want to know where disease is occurring across the landscape. As you consider how many sites to include in your study, you decide to explore how your probability of detecting disease at all sites changes as you increase the number of sites you sample. A) How does the probability of detecting disease at all sites change, and B) what is the probability of detecting disease at none of your sites, if:
  - You consider sampling up to 10 sites
  - Your probability of detecting this disease is 50% (given your sampling and diagnostic methods)

Click (Spatial distribution activity)

- You want to know where disease is occurring across the landscape. As you consider how many sites to include in your study, you decide to explore how your probability of detecting disease at all sites changes as you increase the number of sites you sample. A) How does the probability of detecting disease at all sites change, and B) what is the probability of detecting disease at none of your sites, if:
  - You consider sampling up to 10 sites
  - Your probability of detecting this disease is 50% (given your sampling and diagnostic methods)
- Answers:
  - The probability of detecting disease at all sites decreases non-linearly as you increase the number of sites
  - The probability of detecting disease at none of the sites 0.1%

Click (Spatial distribution activity)

- How does the probability of detecting disease at all sites change if you hold 10 as your maximum number of possible sites, but:
  - Your detection probability is less than 0.5?
  - Your detection probability is greater than 0.5?

Click (Spatial distribution activity)

- How does the probability of detecting disease at all sites change if you hold 10 as your maximum number of possible sites, but:
  - Your detection probability is less than 0.5?
  - Your detection probability is greater than 0.5?
- Answers:
  - As detection probability decreases, your probability of detecting disease at all sites decreases more quickly, changing the shape of your curve. Your probability of detecting no disease at any site also increases
  - As detection probability increases, your probability of detecting disease at more sites increases, while your probability of detecting disease at no sites decreases

Click (Spatial distribution activity)

- How does the probability of detecting disease at all sites change if you hold detection probability at 0.5, but:
  - You reduce the maximum number of sites possible to 5?
  - You increase the maximum number of sites possible to 20?

Click (Spatial distribution activity)

- How does the probability of detecting disease at all sites change if you hold detection probability at 0.5, but:
  - You reduce the maximum number of sites possible to 5?
  - You increase the maximum number of sites possible to 20?
- Answers:
  - As you reduce the number of sites, your probability of detecting disease at all sites decreases more quickly, changing the shape of your curve. Your probability of detecting no disease at any site also increases
  - As you increase the number of sites, your probability of detecting disease at more sites increases, while your probability of detecting disease at no sites decreases

Click (Spatial distribution activity)

- There are 20 parks in your area where you are considering sampling deer for disease. You know from past GPS studies that deer populations do not mix across these parks. Using the Prevalence Power Study activity and part 2 of the Spatial Distribution activity, explore how many samples you can collect per site during a particular sampling period, and what site-specific credible interval width (i.e., uncertainty) surrounding your prevalence estimates if:
  - Prevalence of this disease in other populations usually falls between 0 and 30%
  - You are sure disease is present in these populations
  - Diagnostic assay sensitivity and specificity are both 95%
  - You explore your data assuming a true prevalence of 5%
  - You are curious about including all 20 potential sites
  - Detection probability of this disease is 70%
  - Your resources allow you to collect 500 samples across all sites

Click (Spatial distribution activity)

- There are 20 parks in your area where you are considering sampling deer for disease. You know from past GPS studies that deer populations do not mix across these parks. Using the Prevalence Power Study activity and part 2 of the Spatial Distribution activity, explore how many samples you can collect per site during a particular sampling period, and how large the credible interval width (i.e., uncertainty) surrounding site-specific prevalence estimates is if:
  - Prevalence of this disease in other populations usually falls between 0 and 30%
  - You are sure disease is present in these populations
  - Diagnostic assay sensitivity and specificity are both 95%
  - You explore your data assuming a true prevalence of 5%
  - You are curious about including all 20 potential sites
  - Detection probability of this disease is 70%
  - Your resources allow you to collect 500 samples across all sites

Answers:

You can distribute 25 samples per site during this sampling period, which generates an average credible interval width of 18% surrounding prevalence estimates

Click (Spatial distribution activity)

- How does your credible interval width (i.e., uncertainty) surrounding your prevalence estimates change if:
  - You reduce the number of sites to 10, but maintain a total of 500 samples

  OR
  - You maintain all 20 sites, but increase your total number of samples to 1000

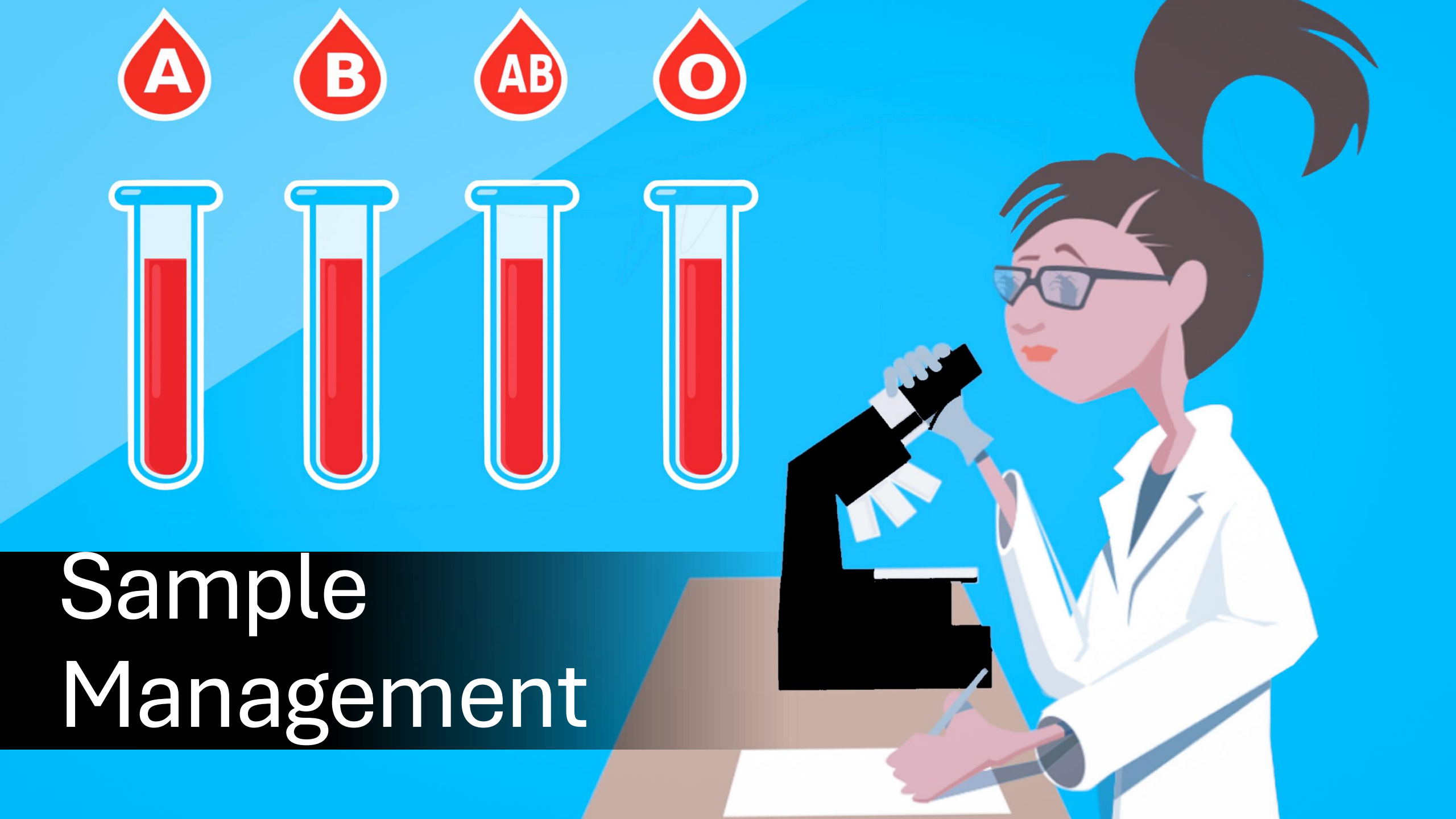
  - How do these changes influence study design decisions?

Click (Spatial distribution activity)

- How does your credible interval width (i.e., uncertainty) surrounding your prevalence estimates change if:
  - You reduce the number of sites to 10, but maintain a total of 500 samples

  OR
  - You maintain all 20 sites, but increase your total number of samples to 1000

  - How do these changes influence study design decisions?
- Answers:
  - Reducing to 10 sites to distribute 500 samples across decreases the credible interval width to 14%
  - Increasing the number of samples to1000 across 20 sites also reduces the credible interval width to 14%
  - Different sampling designs trade off spatial coverage for sampling intensity, leading to the same site-specific credible interval width surrounding prevalence estimates. When designing sampling, using a target credible interval width can help guide decisions surrounding trade offs between spatial coverage goals across sites and sampling intensity within sites



Sample Management

Sample Management Across Lifecycle to Ensure Diagnostic Quality

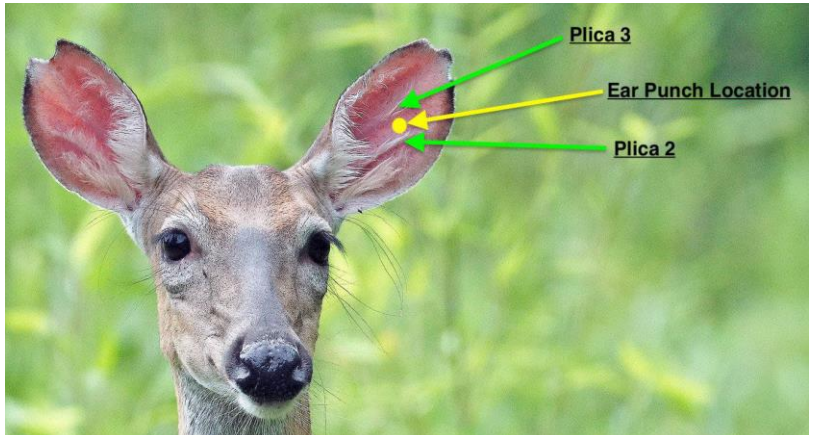
Before Collection

- Gather required materials for sampling
- Ensure capacity for proper material storage across lab, field, and transport

Sample Management Across Lifecycle to Ensure Diagnostic Quality

Field Collection

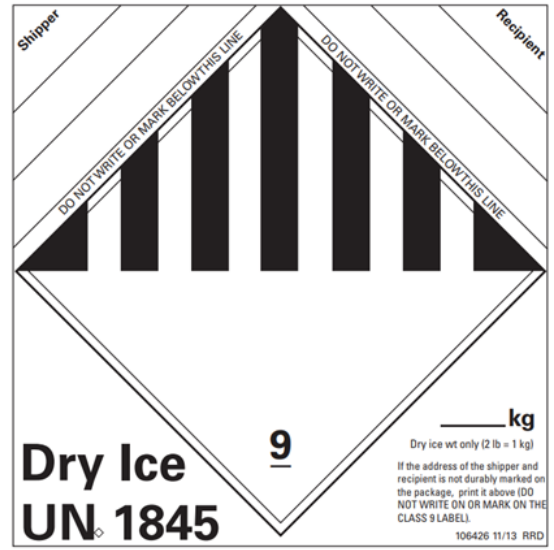
- Ensure sample collection targets the correct organ or body part
- Ensure sampling is “clean”
- Ensure proper post-collection conditions are possible in the field



Sample Management Across Lifecycle to Ensure Diagnostic Quality

Transport

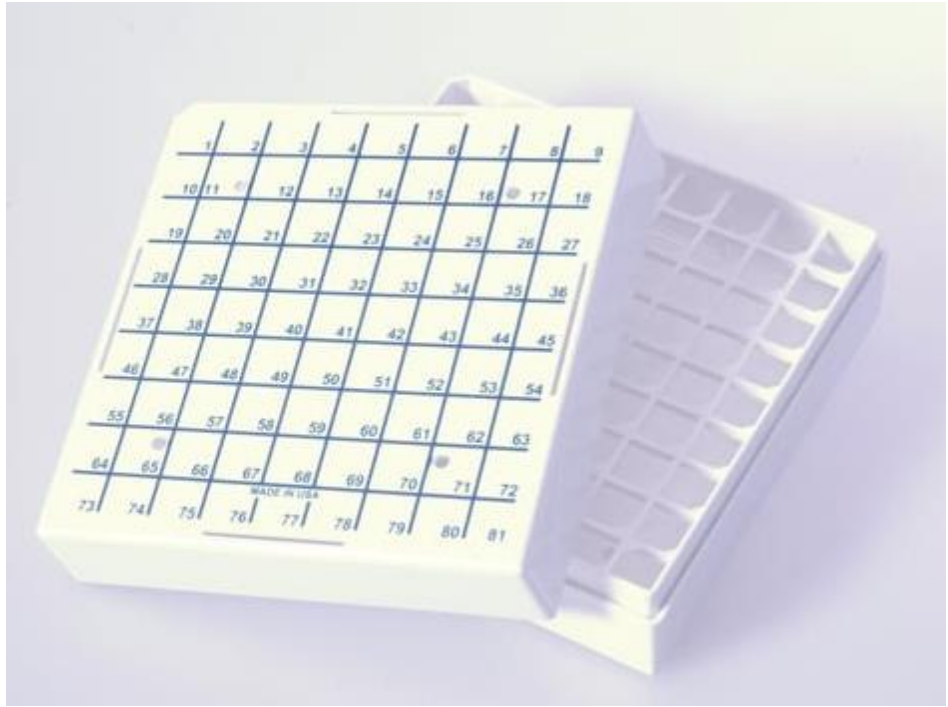
- Ensure post-collection conditions are possible during sample transport from field to laboratory
- Avoid shaking blood samples – cell rupture
- Surveillance objectives determine timeline required for sample submission



Sample Management Across Lifecycle to Ensure Diagnostic Quality

Diagnostic Testing

- Communicate with laboratory about scheduling receiving
- Provide laboratory information about incoming accession and sample type
- Determine results communication pipeline

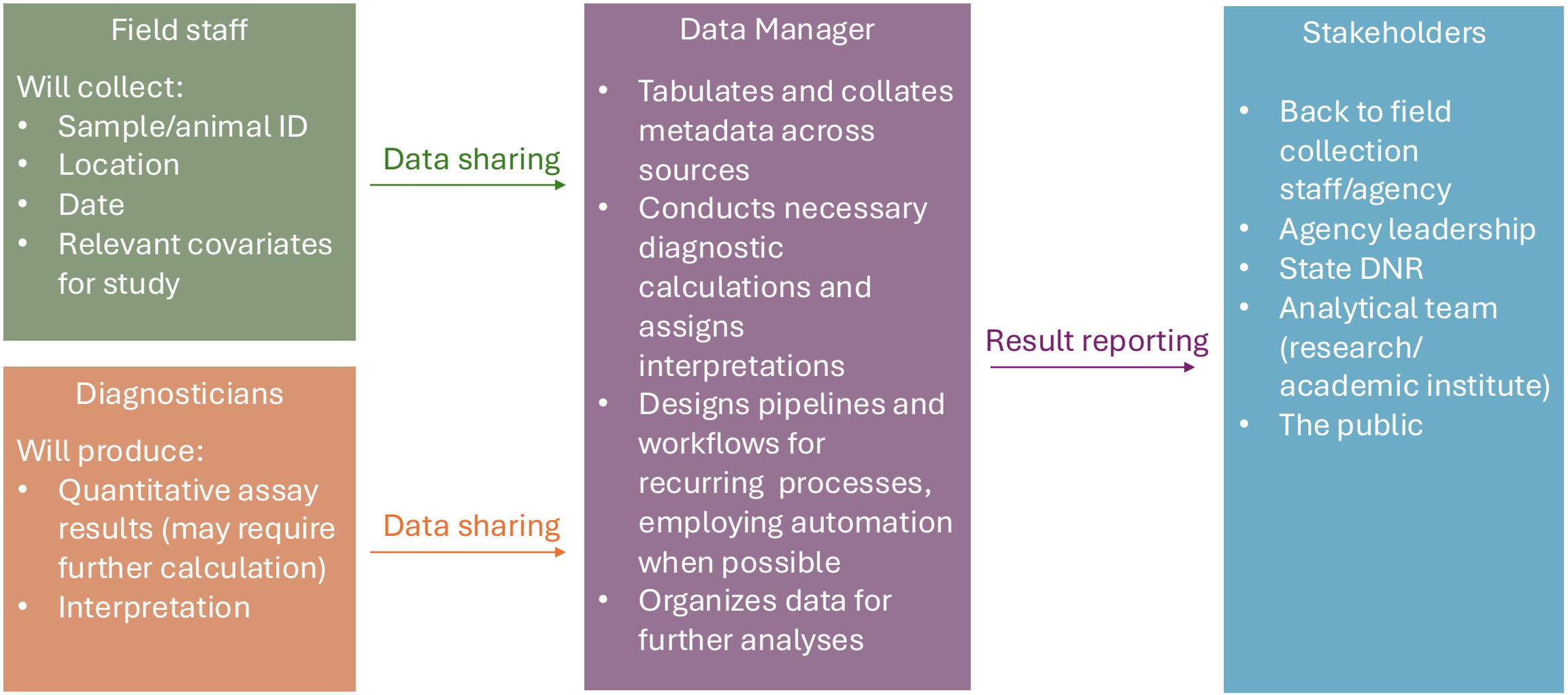


Sample Management Across Lifecycle to Ensure Diagnostic Quality

Archiving

- Establish with cooperating institutions who is responsible for sample archiving
- Managing institution might have archiving requirements
- Space and conditions need to be considered

Surveillance System Data Management Example Roles



Wrap up, Discussion