# A Sampling Strategy Designed to Maximize the Efficiency of Data Collection of Food Web Relationships

Kristen E. Sauby[1] and Mary C. Christman[2]

[1]Dept. of Biology, Univ. of Florida, ksauby@ufl.edu

[2]Statistical Consulting LLC, Dept. of Biology and Dept.of Statistics, Univ. of Florida

ksauby@ufl.edu

@KristenSauby

9 August 2017, Ecological Society of America Conference, Portland, OR

# Empirical Food Web, NE Florida

**Objectives:**

- estimate insect occupancy
- estimate impact of herbivory to cactus demography

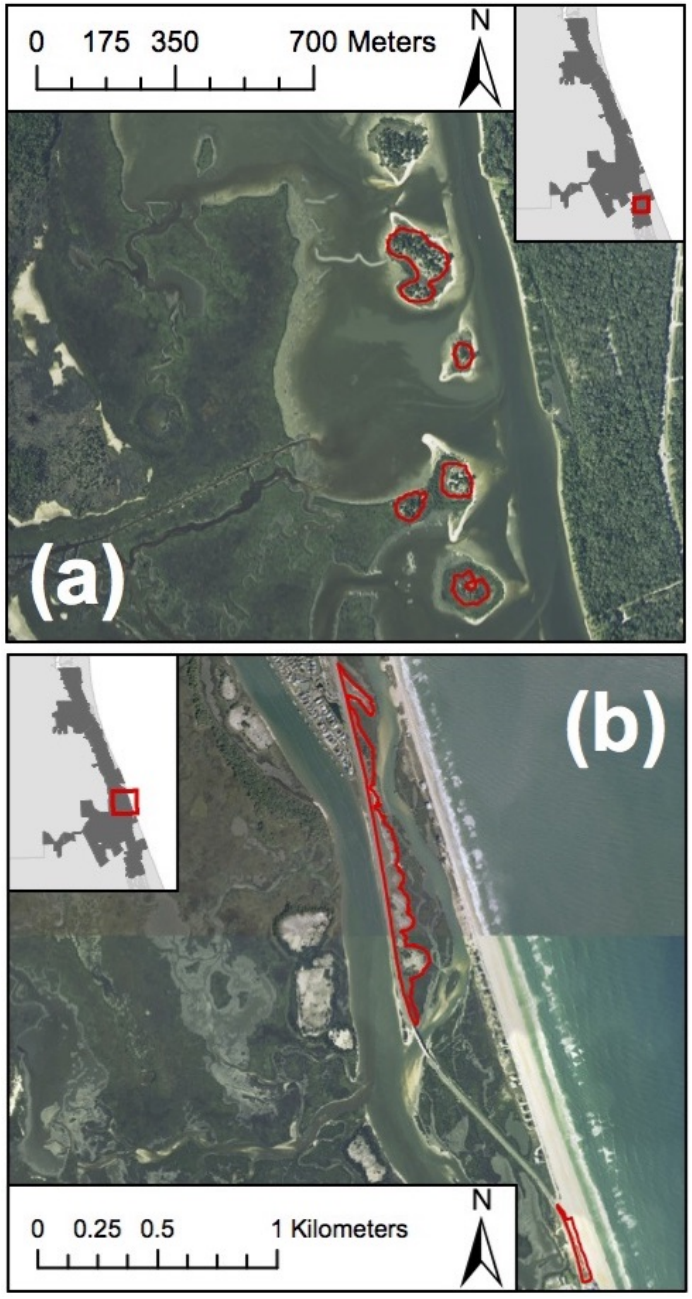
Set up 1346 1-$m^2$ plots at 7 habitat patches from May-July 2012 in Northeastern Florida

Invasive Cactus Moth, *Cactoblastis cactorum*

Native Cactus Moth, *Melitara prodenialis*

*Opuntia stricta*

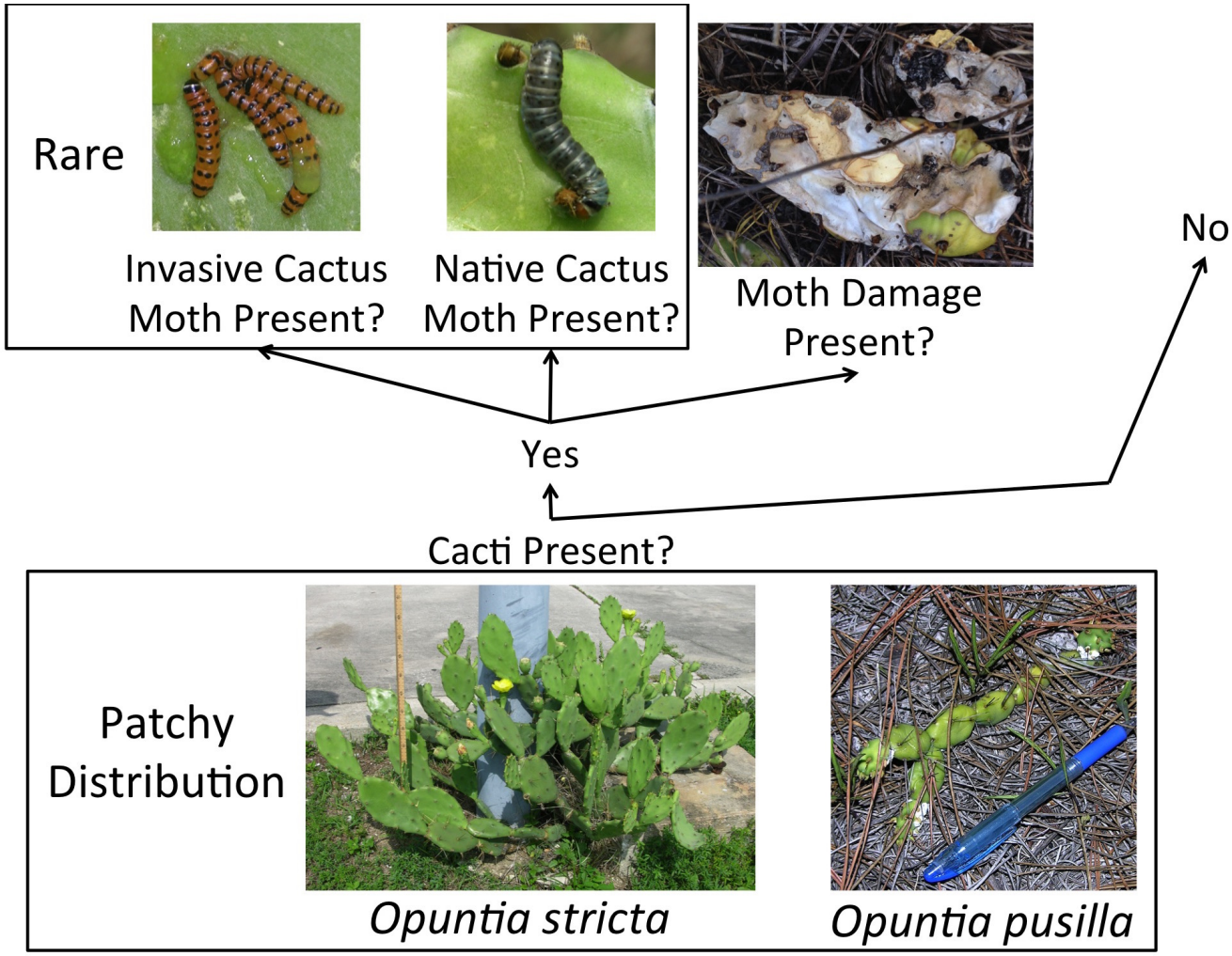
*Opuntia pusilla*



# Hierarchy of Sampling of Our Empirical Food Web

**Objectives:**

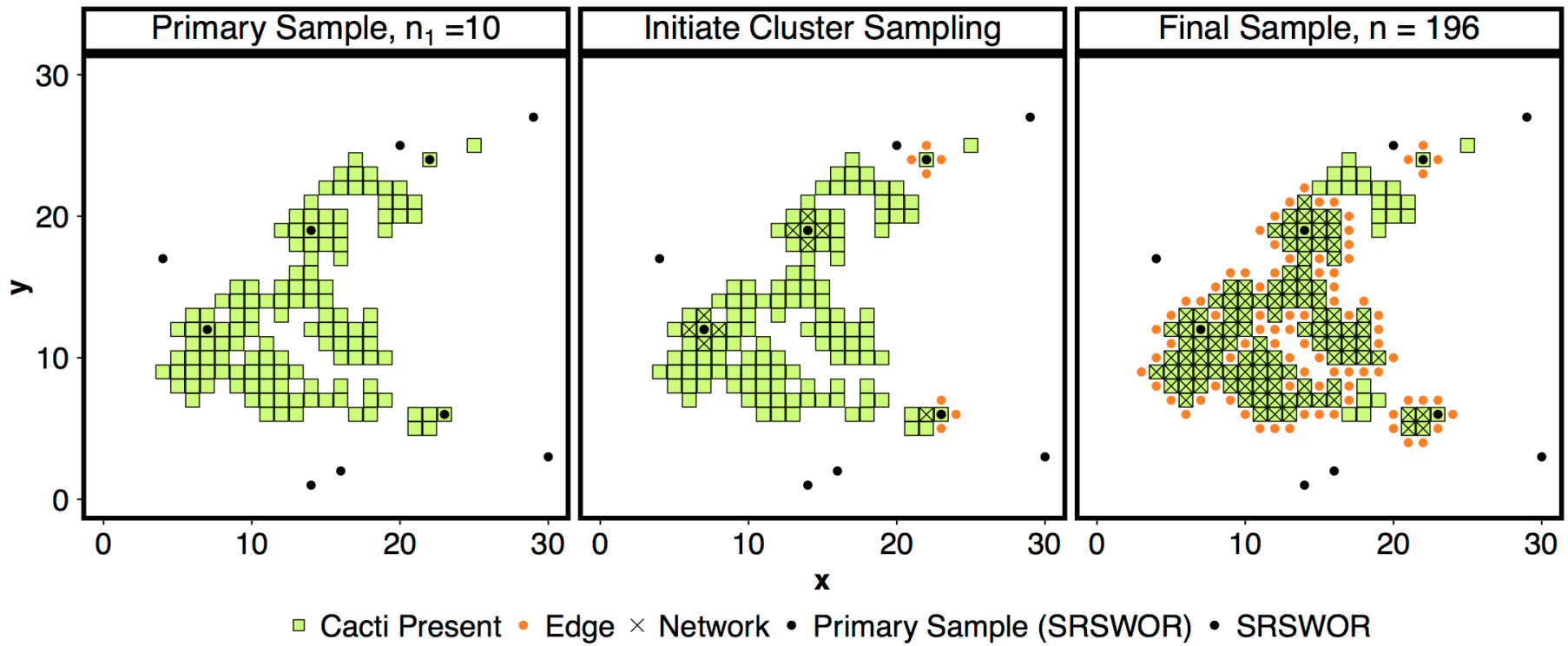
- estimate insect occupancy
- estimate impact of herbivory to cactus demography



# Adaptive Cluster Sampling (ACS)

Popular design for sampling of **rare, clustered populations**.

- Primary sample is drawn according to SRSWOR (simple random sampling without replacement)
- Do primary units contain species of interest? If yes, sample all neighboring units until entire patch has been sampled




Proposed in Thompson, S. (1990). Adaptive cluster sampling. Journal of the American Statistical Association, 85(412), 1050–1059.

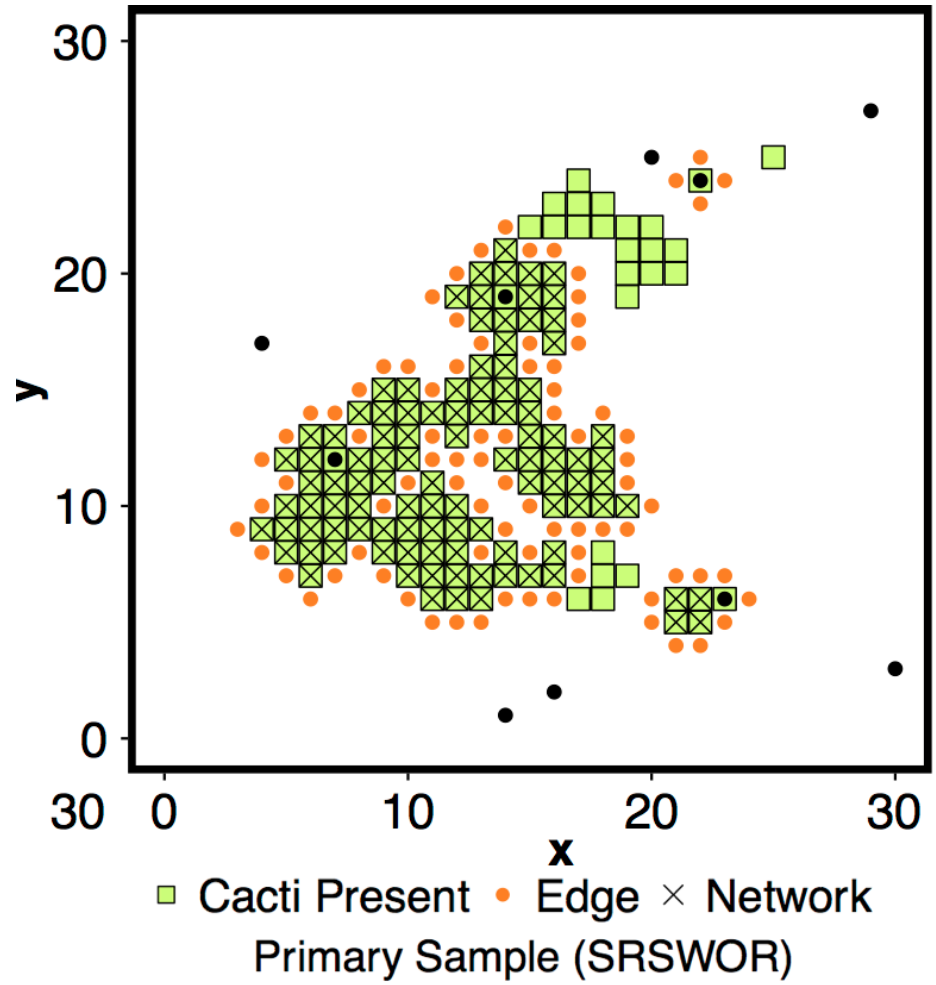
# Estimators for ACS

We can calculate **unbiased estimates** of the mean, total, and variance using Horvitz-Thompson estimators

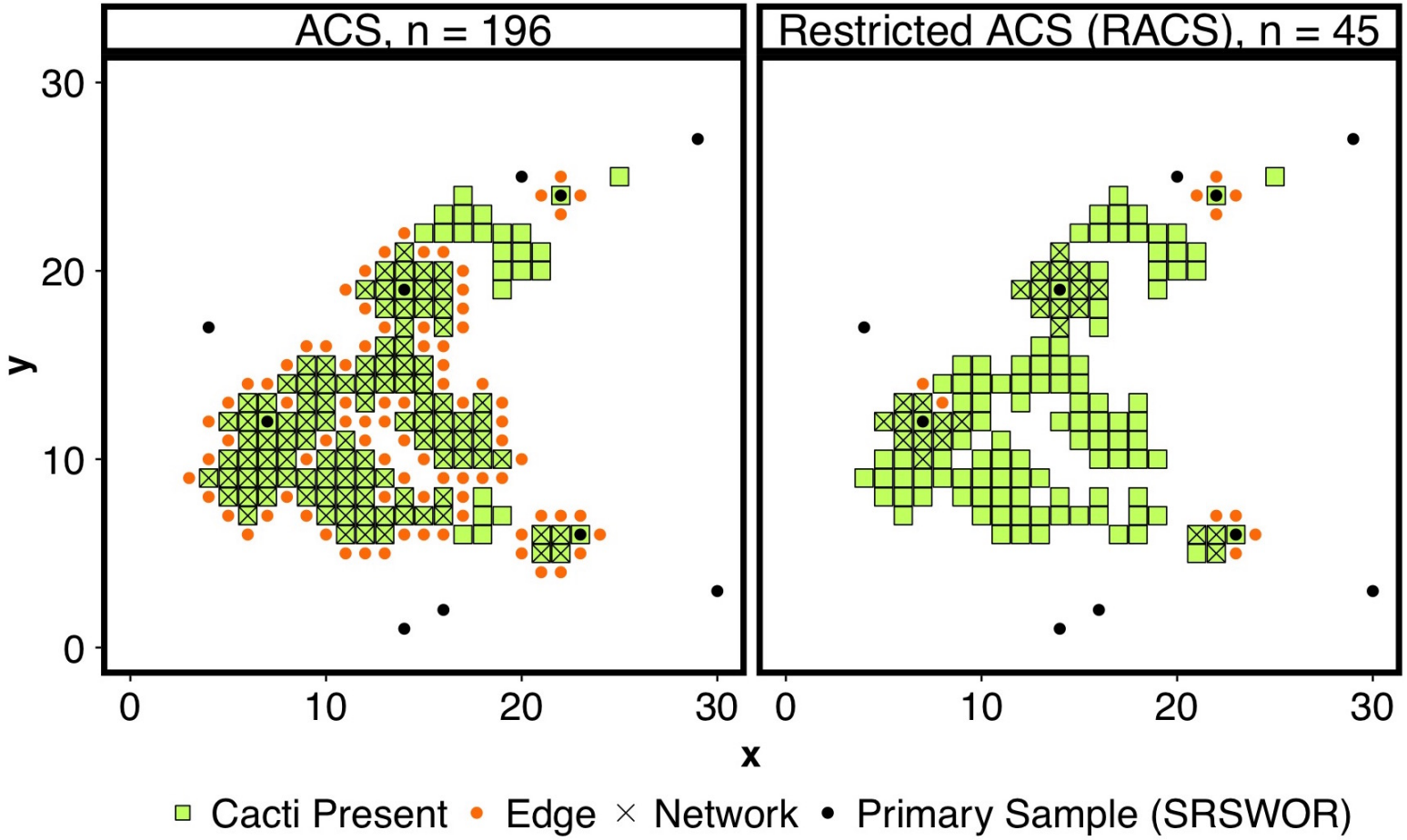
These estimators weight observations from networks by their **inclusion probabilities**

- **network**: a set of neighboring units containing the species of interest
  - 3 out of the 6 "patches" are sampled in example (each surrounded by edge units)
- **inclusion probability**: probability of inclusion in the sample

The inclusion probability of a network is proportional to its size



# Restricted ACS (RACS)




**Pros**

- Compared to SRSWOR, increased sampling efficiency and reduced variance <u>for rare and clustered populations</u>

**Cons**

- Final sample size is unknown at the start of sampling

**Pros**

- same procedure as ACS, except that we restrict the number of neighboring units to q (in our example, q = 12)
- <u>Smaller final sample size compared to ACS</u>

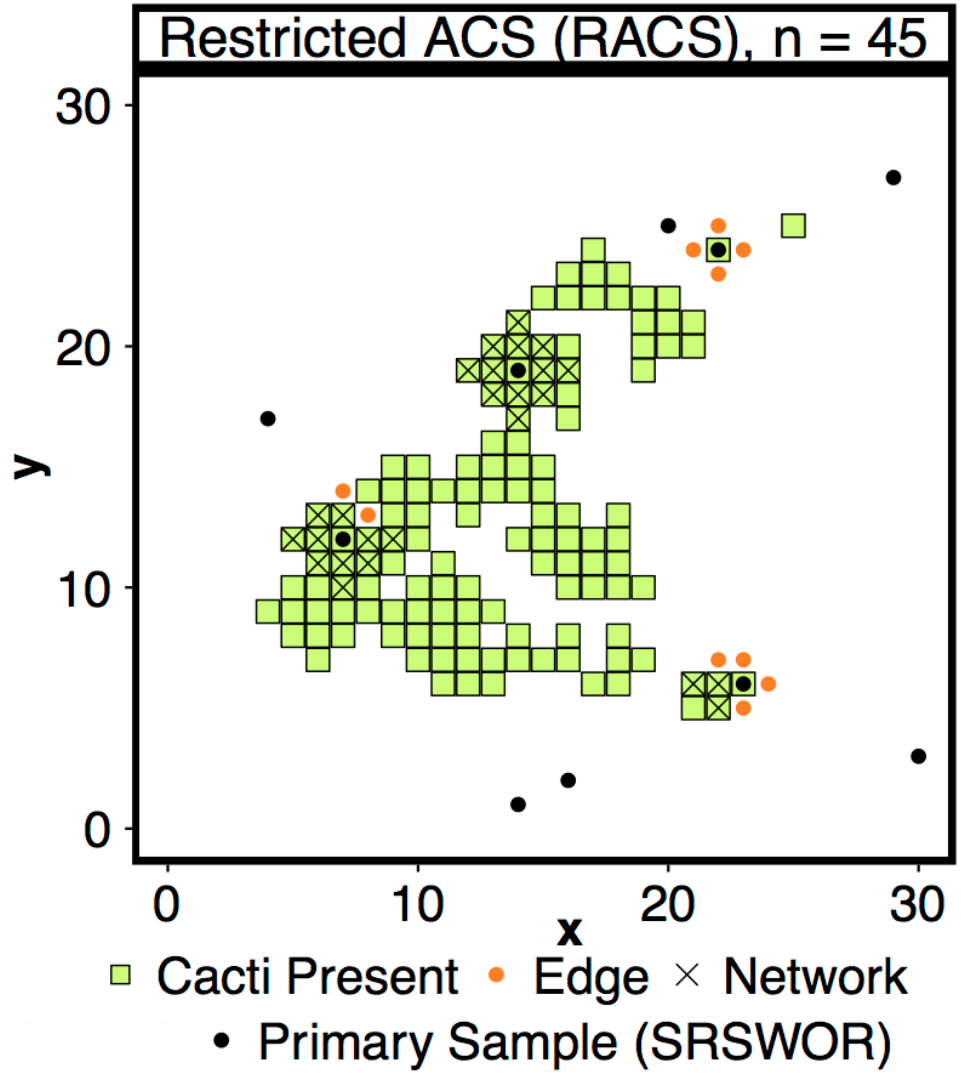
**Cons?**

# Estimators for RACS

However, for RACS we do **not** necessarily know the inclusion probabilities for all units

- the inclusion probability of a network is **not** necessarily proportional to its size
- networks may not have been sampled completely

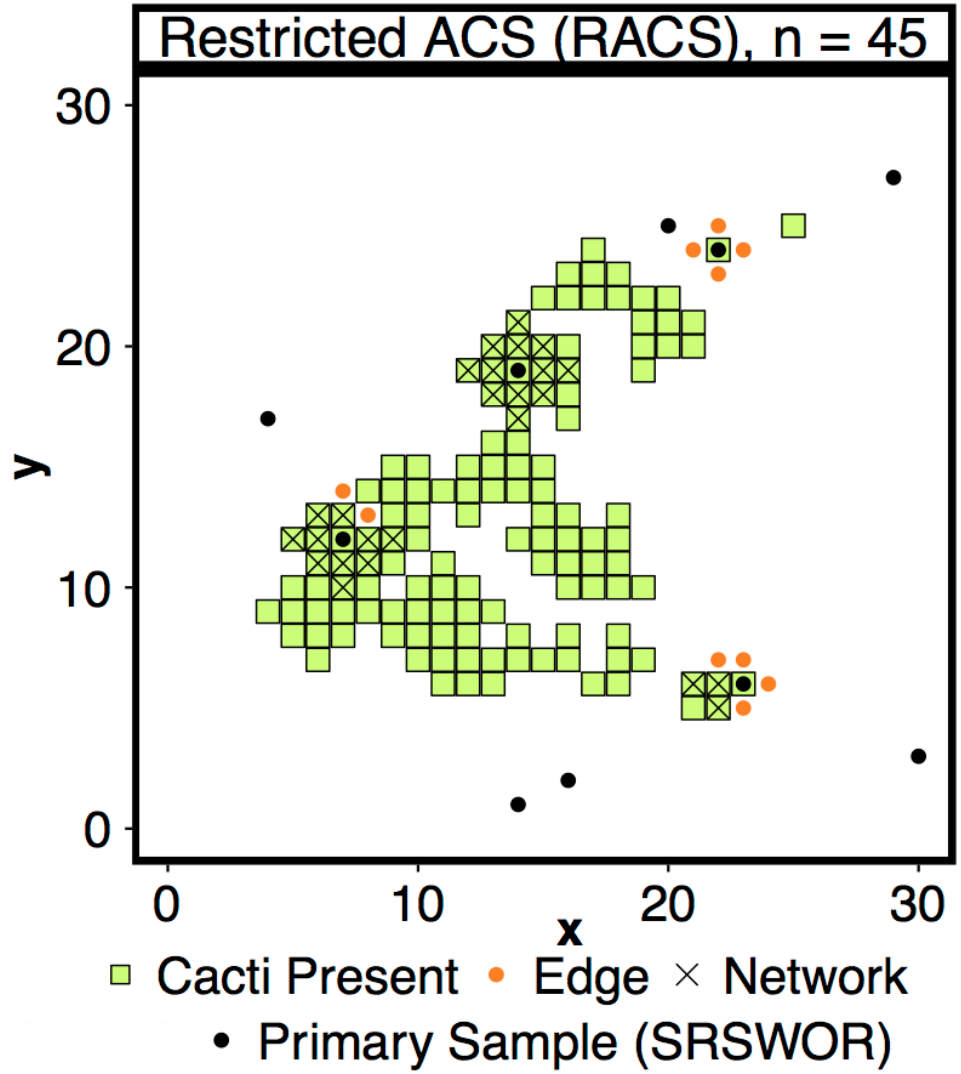
Thus, the Horvitz-Thompson estimators may be **biased** when we include improperly calculated inclusion probabilities



# RACS-Corrected Inclusion Probabilities

We can estimate the inclusion probabilities utilizing the following information about a sample:

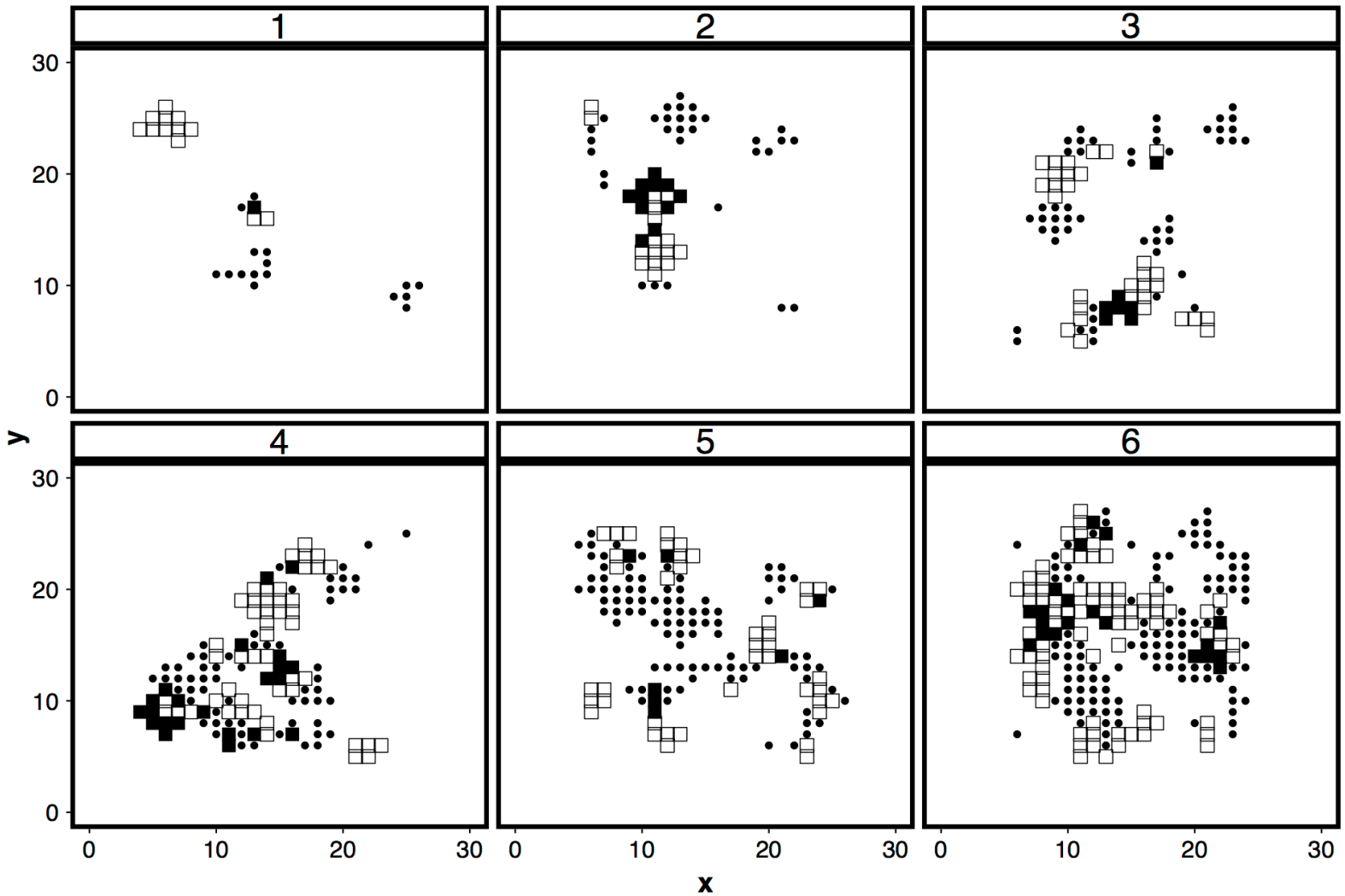
- spatial clustering (estimated using Moran's I)
- the primary sample size



# Simulations

**Objective**: Explore properties of the RACS design by repeatedly sampling from 6 artificial populations

**Artificial Populations**: randomly sampled from a dataset of 2 cactus species and 2 insect species.

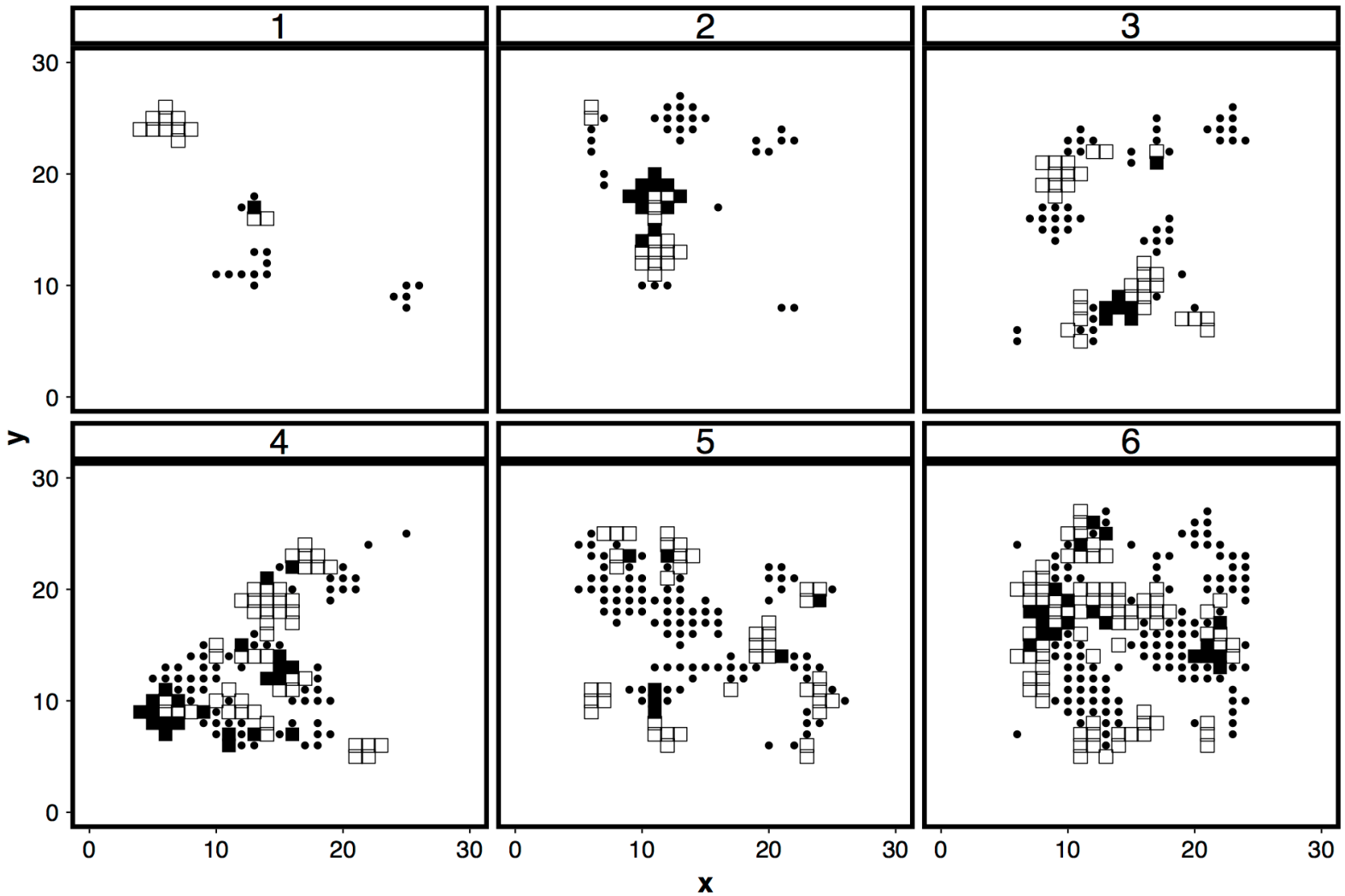



• *O. stricta* present ▫ *O. pusilla* present ▪ *O. stricta* and *O. pusilla* present

# Simulations

We used cactus presence as our criterion to initiate cluster sampling

We ran 5000 simulations of the RACS and ACS designs




• *O. stricta* present ▫ *O. pusilla* present ▪ *O. stricta* and *O. pusilla* present

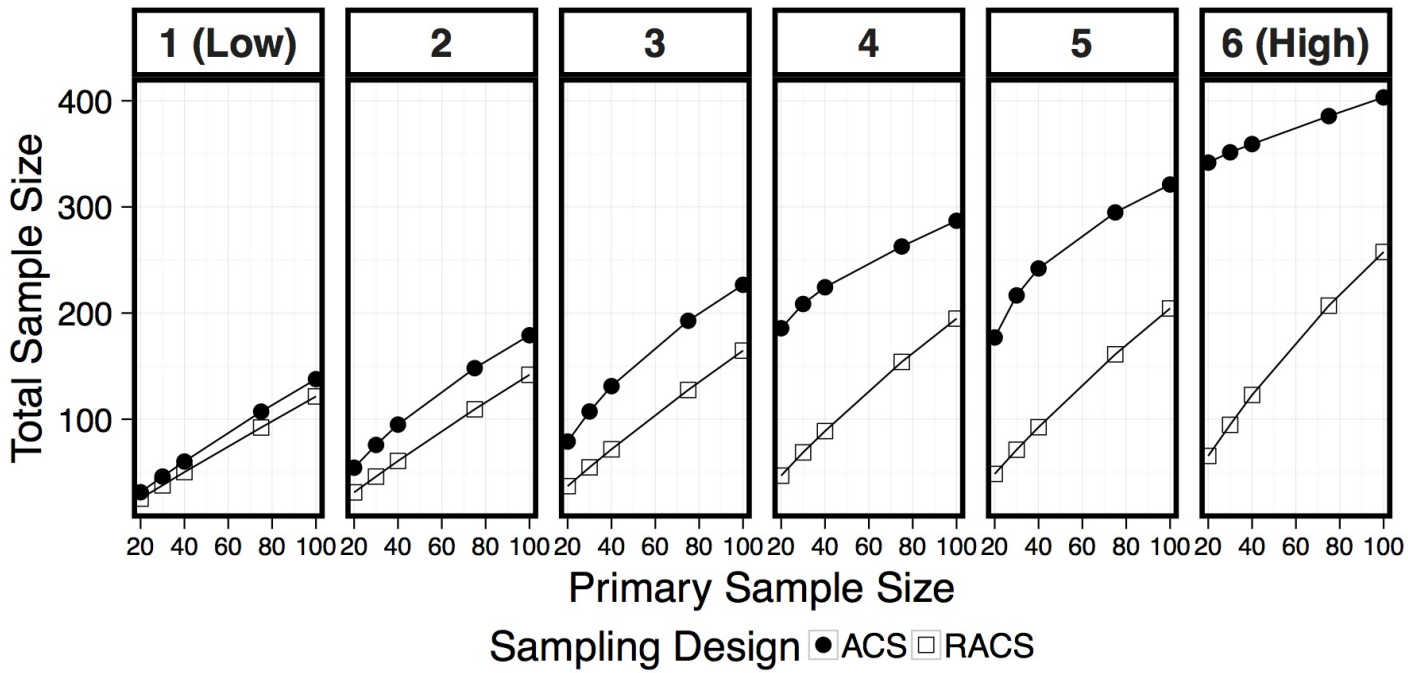
# Hypotheses About RACS

**Using improperly calculated inclusion probabilities**, estimators will be

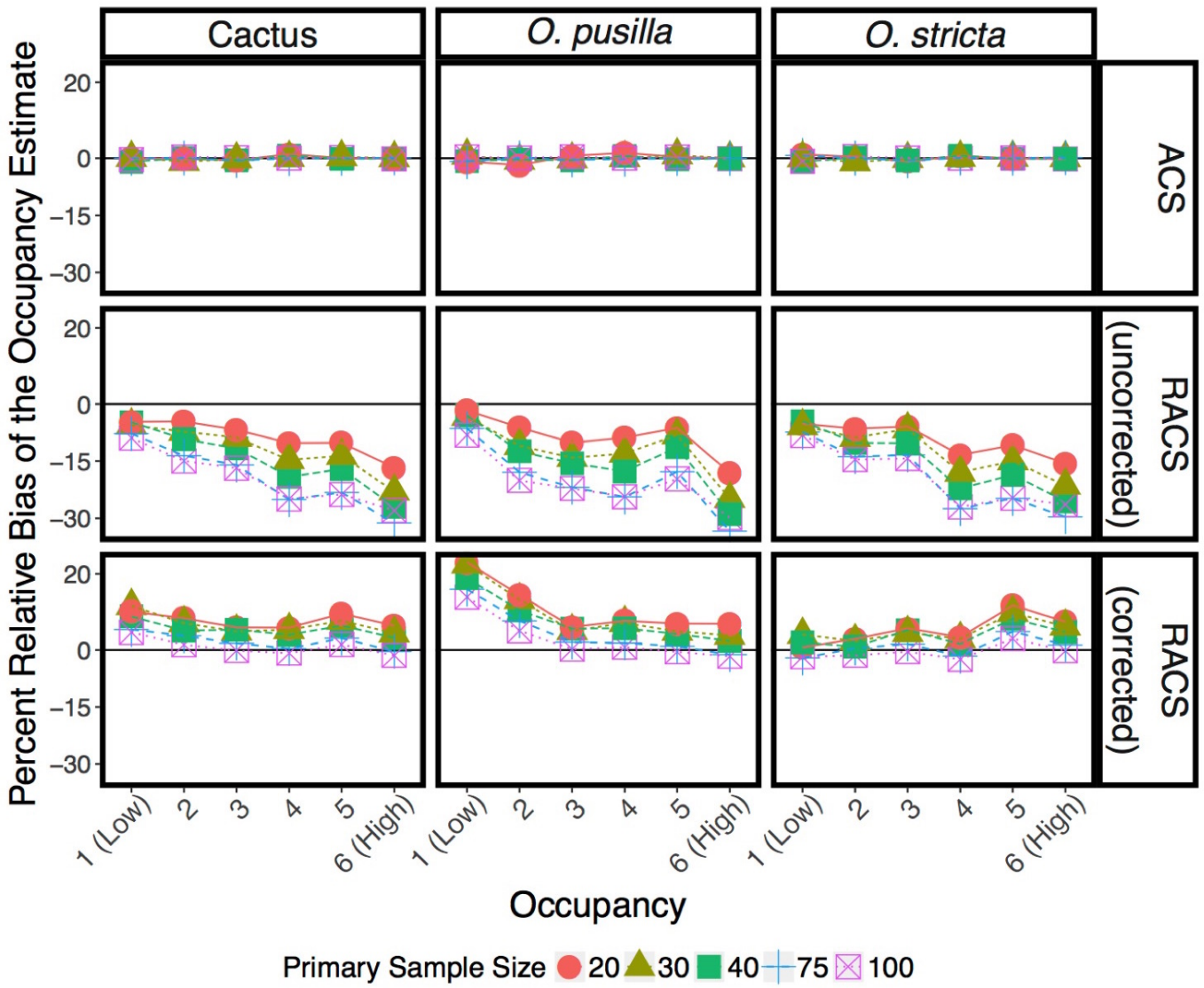
1. **biased** for the variable on which the RACS criterion is based
   - *i.e.*, cactus presence
2. **unbiased** for variables that
   - are not used to generate the RACS criterion
   - are uncorrelated within networks
   - *i.e.*, insect presence, evidence of insect damage

**Using RACS-corrected inclusion probabilities**, estimators will be **unbiased or have reduced bias**.

# Results - Final Sample Sizes



# Sampling Cactus Occupancy w/RACS

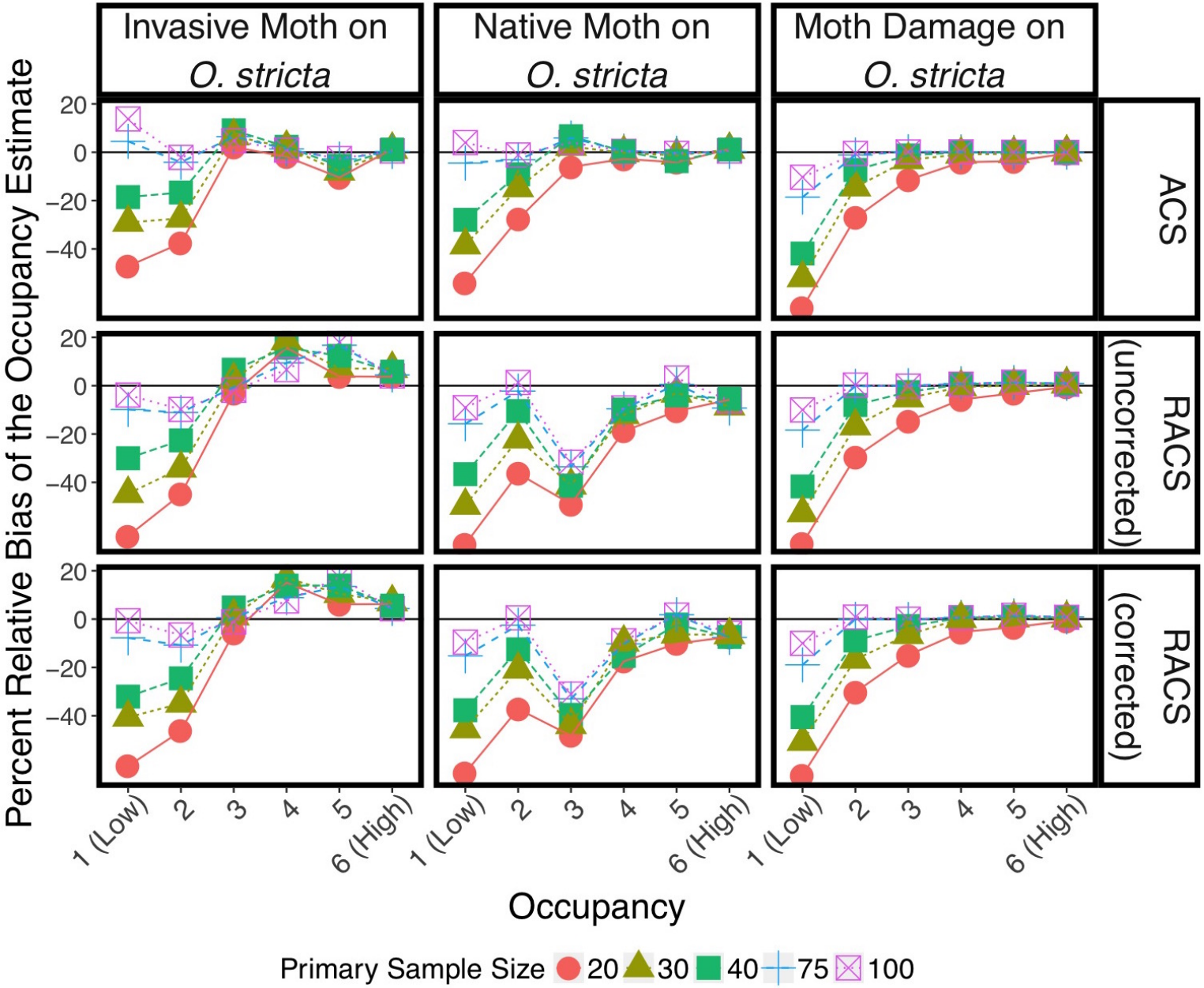



- Unbiased

- Underestimate
- Worsens as primary sample size ↑

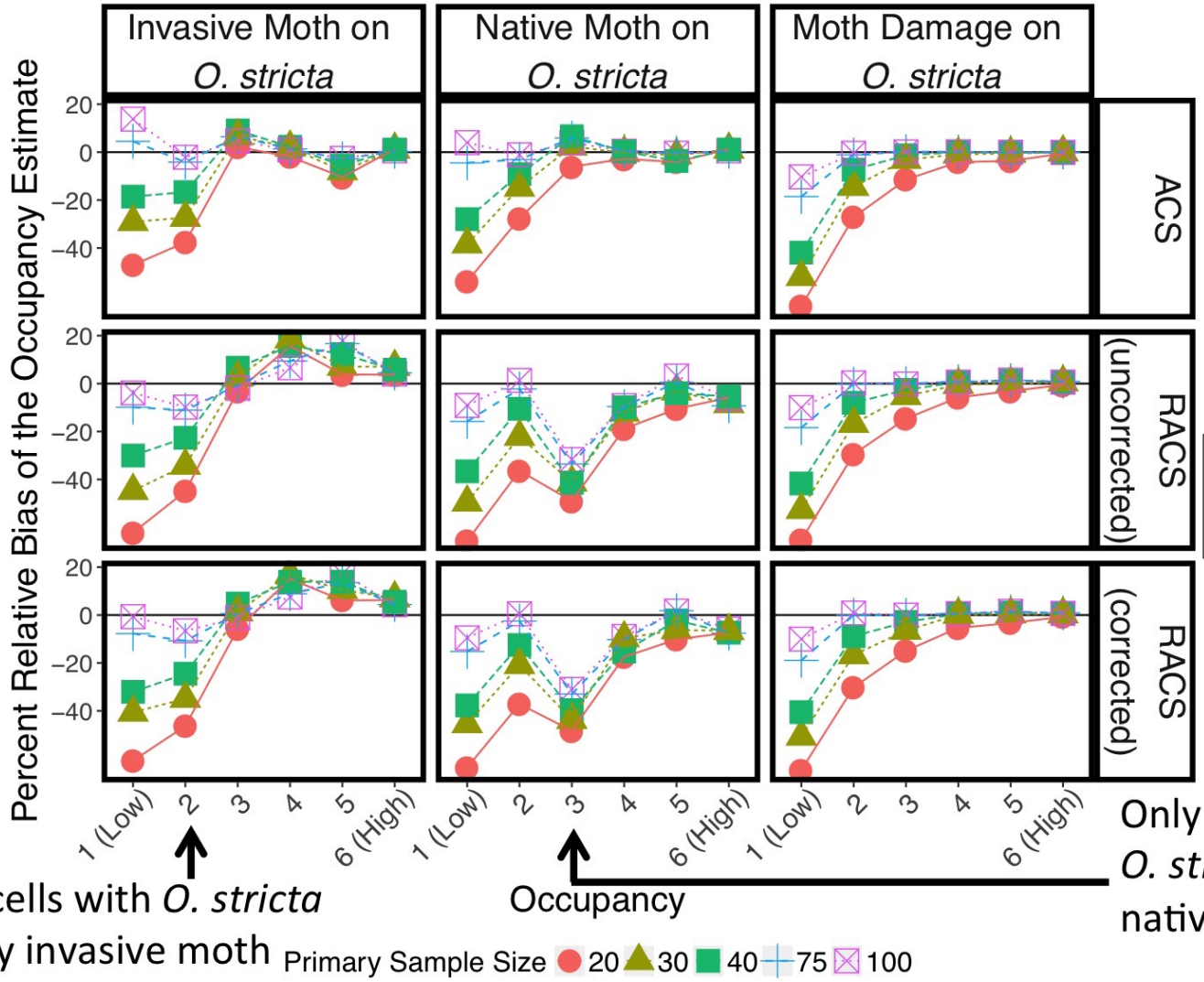
- Slight overestimate
- Improves as primary sample size ↑

# Sampling Insect Occupancy w/RACS




similar bias,
<u>less effort</u>

# Sampling Insect Occupancy w/RACS




similar bias, <u>less effort</u>

Only 2/42 cells with *O. stricta* occupied by invasive moth

Only 1/57 cells with *O. stricta* occupied by native moth

# Conclusions About RACS (So Far)

- It requires less effort than ACS (smaller final sample sizes)
- Inclusion probability estimation for RACS helps reduce estimation bias
  - Slight bias in estimation of the mean for the criterion variable (in our example, cactus occupancy)
  - No real penalty in terms of estimation of the mean of insect variables
    - similar bias as ACS with smaller sample sizes
    - not very sensitive to mis-calculations of inclusion probabilities
- the bias is lowest for larger primary sample sizes

# Future Work

- Estimate variance for RACS
- compare efficiency of RACS, ACS, and SRSWOR
  - e.g., would it be better to do simple random sampling without replacement (n = 100) or RACS (final sample size = 100)?
- release R package, "ACSampling" (currently private on Github)

# Acknowledgements

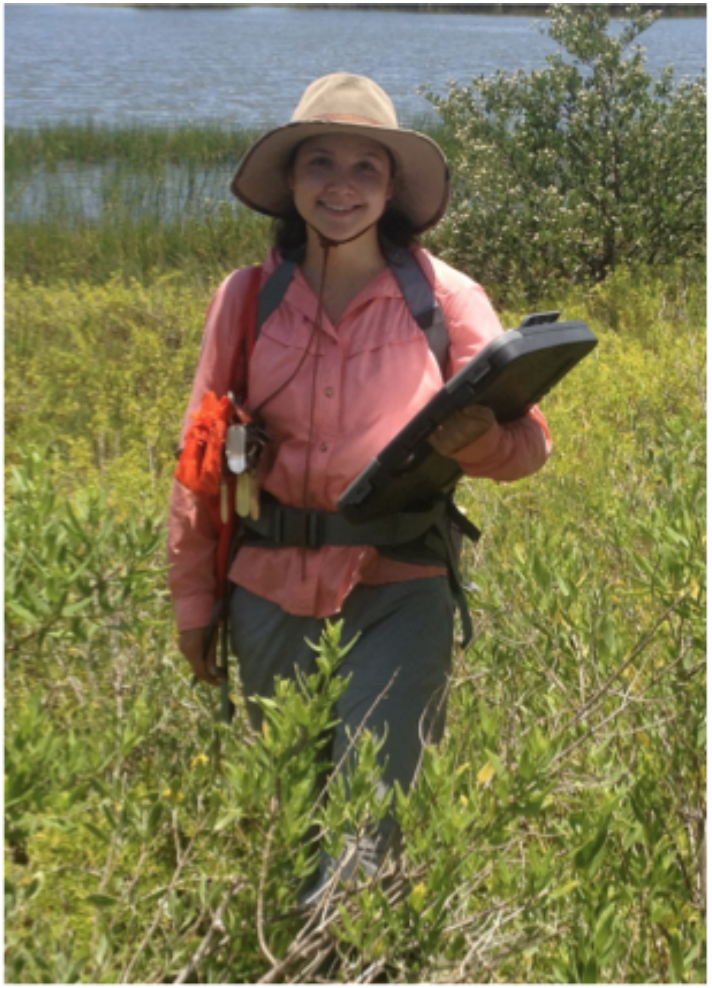
co-advisor, R.D. Holt

**Data Collection**

- Guana Tolomato Matanzas National Estuarine Research Reserve
  - Katie Petrinec
  - Matt Welsh
- Polly Harding, undergrad. volunteer, now at Texas A&M (pictured)

**Funding**

- QSE3 IGERT (NSF Award # 0801544)
- Cactus and Succulent Society of America
- Garden Club of America/Center for Plant Conservation
- University of Florida Graduate Student Council
- Department of Biology
- Botanical Society of America
- Florida Native Plant Society



# Questions?

ksauby@ufl.edu
@KristenSauby