

Calculating diagnostic test characteristics in clustered data:

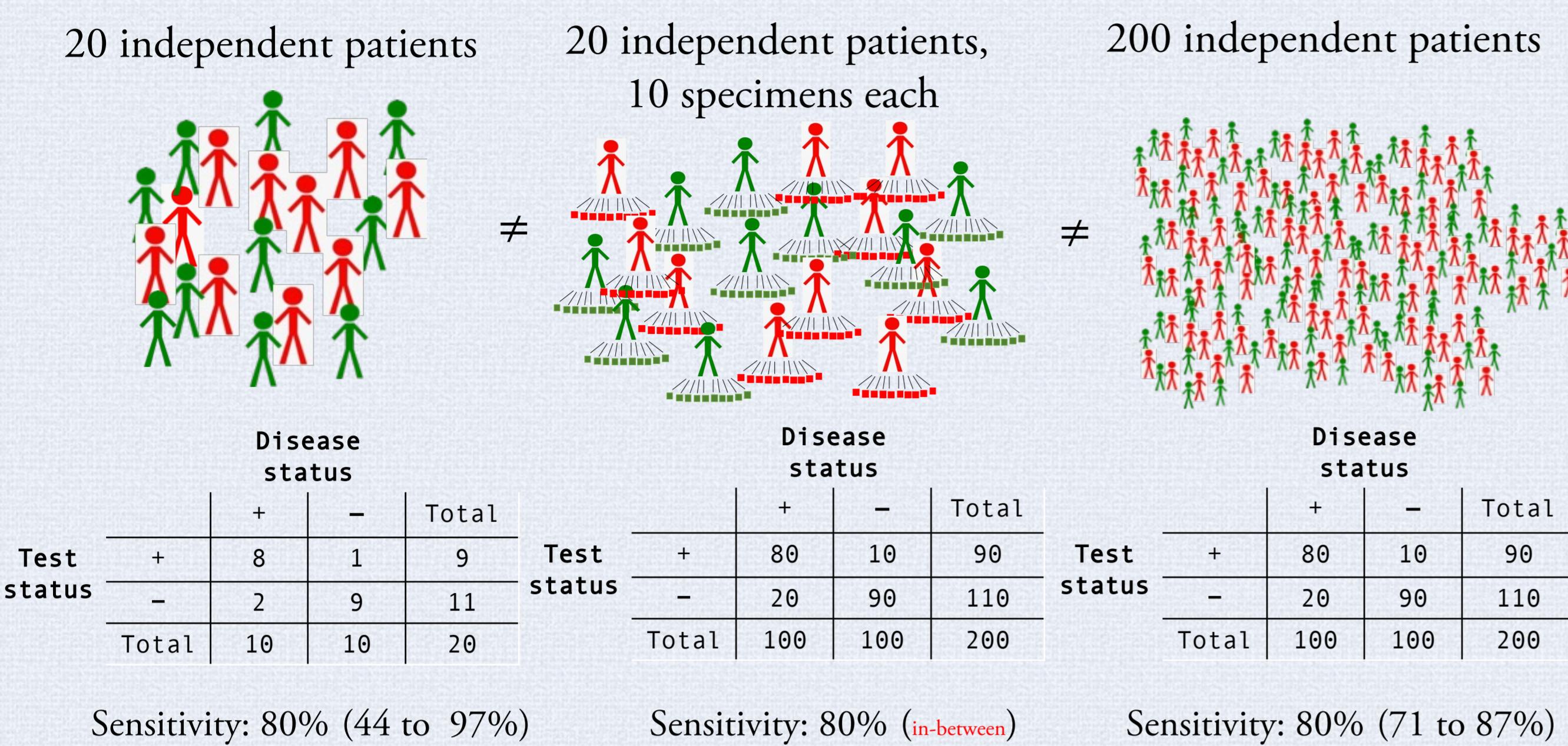
A simulation study



Katalin Tamási, PhD and Gerton Lunter, PhD
 Unit of Medical Statistics and Decision Making
 Department of Epidemiology, University Medical Center Groningen
tamasi.k@umcg.nl, g.a.lunter@umcg.nl



INTRODUCTION WHY DO WE NEED TO BE AWARE OF CLUSTERING?



- Clustering introduces multiple levels of analysis (patient-level, specimen-level, etc.)
- Higher degree of clustering → Less information encoded → Less certainty in the estimates
- Clustering is ubiquitous ⇔ Routinely ignored in medical literature, possibly because there is no straightforward way to adjust for it

→ WE DEVELOPED AN R SCRIPT TO PROVIDE A ONE-STOP-SHOP SOLUTION

TAKE-HOME MESSAGE

WHAT LEVEL TO ANALYZE YOUR CLUSTERED DATA?

- Patient-level analyses:
 - Appropriate if intervention / clinical consequence is patient-level (e.g., chemo- vs. immunotherapy?)
 - Clustering is not a concern if patients are independent
- Specimen-level analyses:
 - Appropriate if intervention / clinical consequence is specimen-level (e.g., stent placement vs. angioplasty?)
 - If clustering is present, adjusting for it yields more valid (= wider) CIs vs. naïve methods

WHICH METHOD(S) DEAL BEST WITH CLUSTERING?

No across-the-board winner but several sensible approaches emerged:

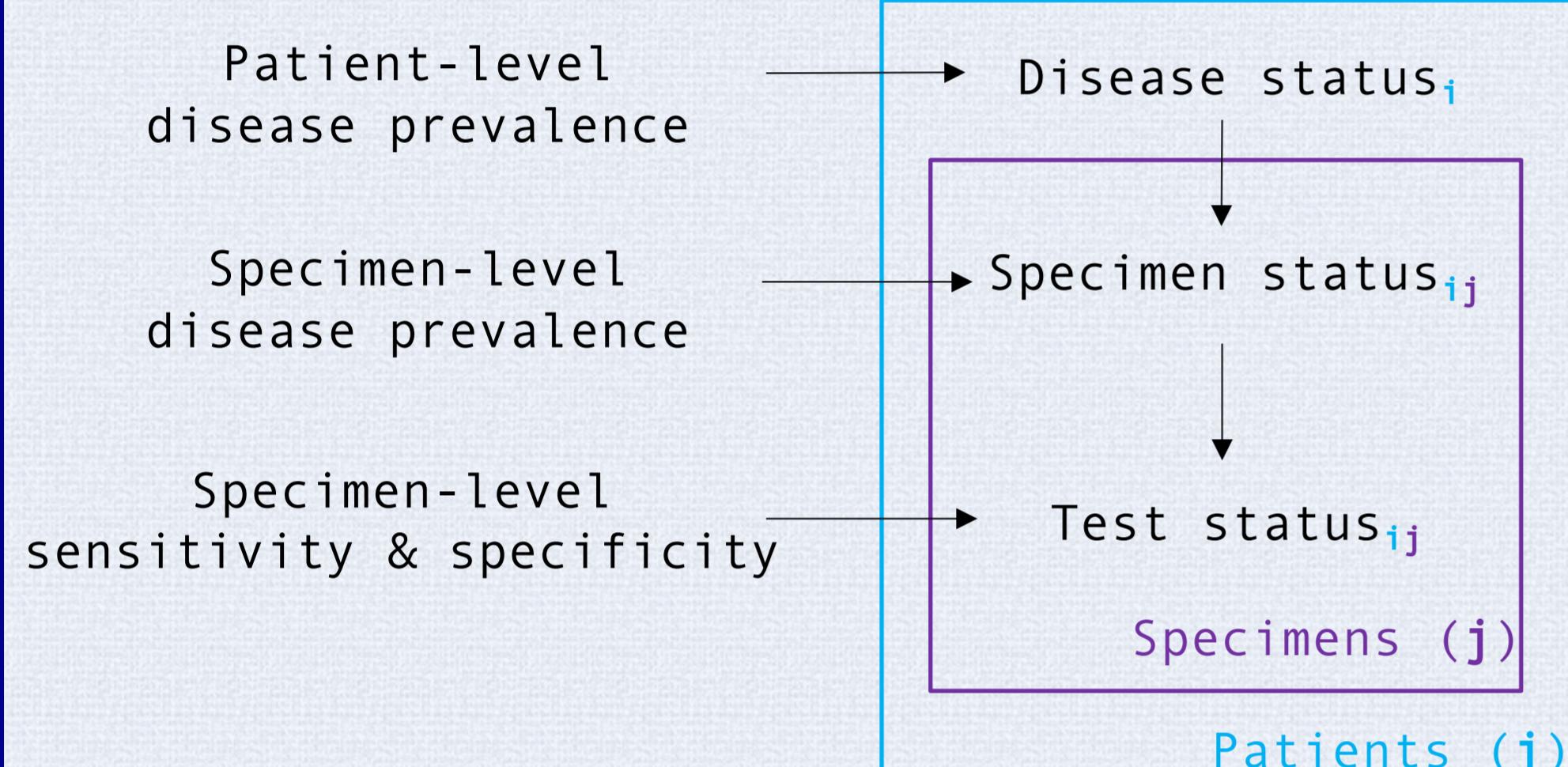
- Cluster bootstrap performs consistently well in diverse configurations, works off-the-shelf
- Mixed effects models may need tweaking to perform decently
- Logistic regression models perform adequately only with robust SEs and when diseased vs. non-diseased populations are separated
- GEEs are only valid with large samples

GIVEN YOUR DATA, OUR OPEN-ACCESS R SCRIPT OUTPUTS

- Contingency tables at each level of analysis
- Estimates of the degree of clustering (ICC) in diseased and healthy populations
- Summary tables of diagnostic test value and CIs using every method
- Forest plots of each diagnostic test value and CIs using every method

SO YOU CAN MAKE AN INFORMED CHOICE BASED ON YOUR UNIQUE CIRCUMSTANCES

METHOD



Parameter space: plausible & at cases extreme values to assess robustness

Parameters	Config 1	Config 2	Config 3	Config 4
Population size	30	60	100	200
Patient-level disease prevalence	0.1	0.2	0.4	0.5
Number of specimens per patient	3	5	10	20
Specimen-level disease prevalence	0.9	0.6	0.3	0.1
Specimen-level sensitivity	0.6	0.8	0.9	0.95
Specimen-level specificity	0.95	0.9	0.85	0.7
Number of clusters	20	10	5	3
Degree of clustering manipulated by Between-cluster variance	0 → 1	0 → 1	0 → 1	0 → 1

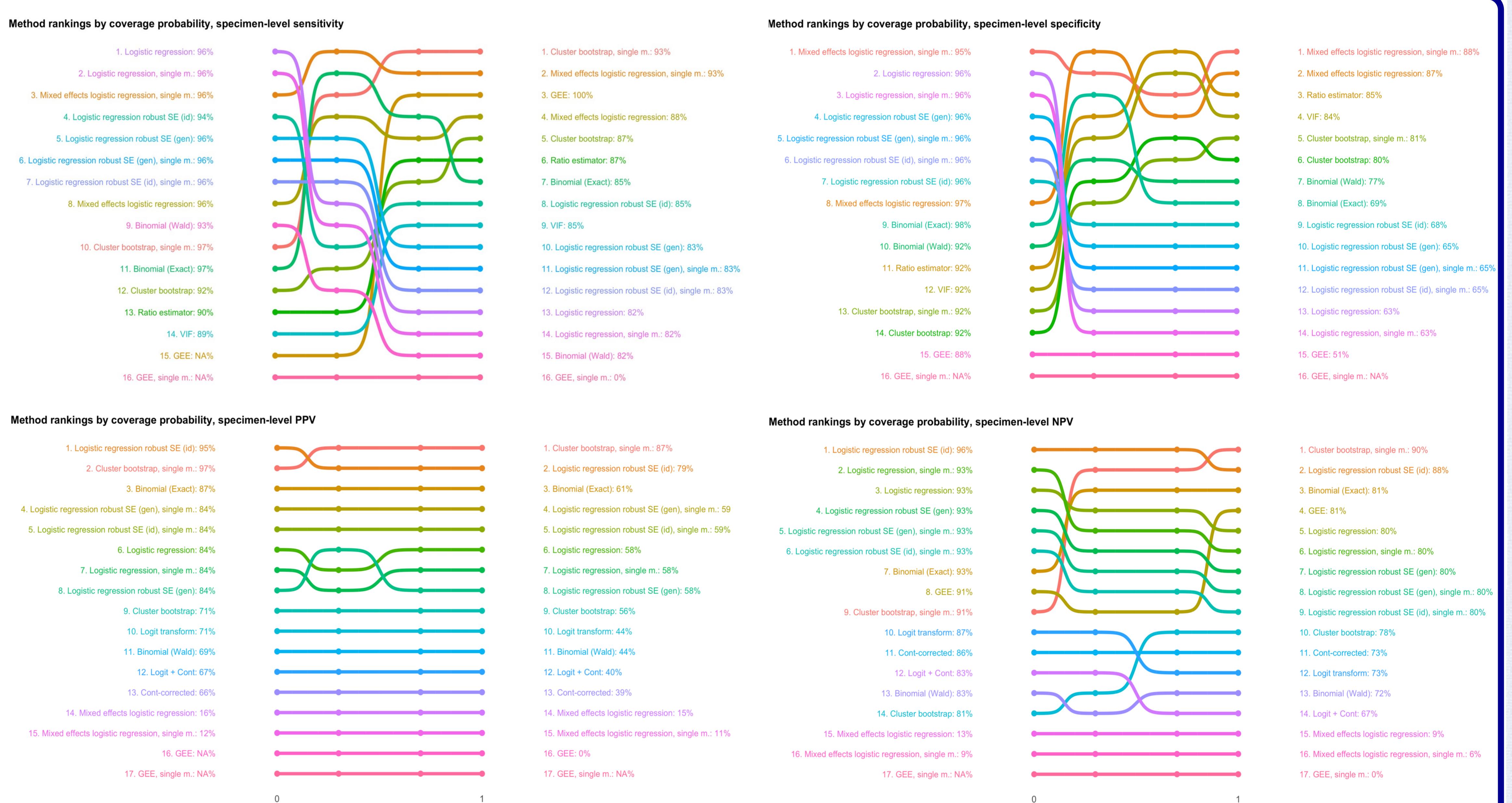
SETUP

- We simulated a large number (5000) of patient samples given each of the 4 configurations
- Within each sample: Calculated each diagnostic test value + 95% CI using each method
- Evaluated methods based on
 - Coverage probability: the closer to the nominal 95% the better
 - CI length: the narrower the better
- Assessed robustness of performance as degree of clustering ↑

RESULTS

As degree of clustering ↑ (x-axis):

- Coverage probabilities systematically decline
- Adjusting methods outperform naïve ones
- Relative rankings (but not coverage probabilities) remain stable across configurations (only Config 2 shown →)
- Comparable coverage probabilities with only 1000 samples (max difference: ± 3 %)
- Performance remains closest to nominal across diagnostic test values with single model of cluster bootstrap
- Mixed effects models only perform well when the approximation of the integral over random effects is decent → May need to change no. of quadrature points / optimizer to obtain better approximation
- Logistic regression models only reliable when robust SEs are adjusted by cluster and diseased vs. healthy populations separated (→ separate ICCs)
- GEEs don't reliably converge in small samples (<5 patients per any cell in contingency table)
- Other variance adjustment methods decent but not superior to cluster bootstrap or regression methods



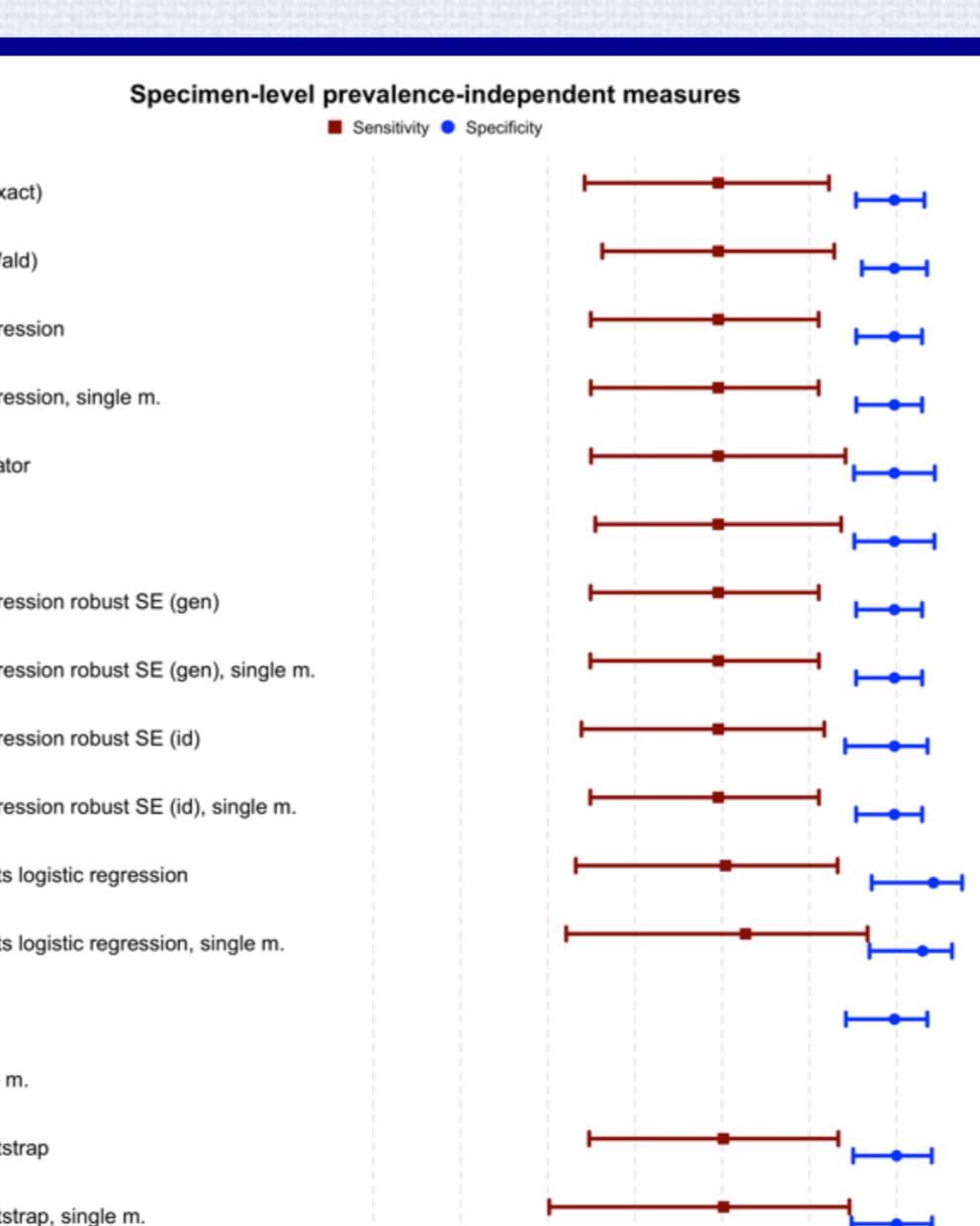
DISCUSSION

WE AIM TO GIVE CLINICIANS THE TOOLS TO:

- Determine if clustering is an issue
 - Specimen-level analysis is indicated & ICC > 0
- Choose the optimal approach given needs & constraints
 - Cluster bootstrap: most versatile choice
 - Mixed effects models: default parameters may be suboptimal
 - Logistic regression models: robust SE separate model outperforms the others
 - GEEs: only work with large samples (≥ 5 in contingency table)
- Minimize error & bias when designing diagnostic studies
- Properly interpret merits and limitations of published findings

R SCRIPT OUTPUT

- Degree of clustering (ICC) in diseased and healthy populations
- Summary tables of each diagnostic test value & CIs
- Forest plots of each diagnostic test value & CIs (example →)
- Contingency tables at each level of analysis



ACKNOWLEDGMENTS

We thank our department members and Dieuwke Broekstra for discussions and feedback and the Center for Information Technology of the University of Groningen for providing access to the Peregrine high performance computing cluster.
 More information including the complete R code, graphs, and references can be found at <http://github.com/KataTam/DiagTestCluster>.

