An Assembly of Survey Sample Size Planning Parameters

Observed Values of Vaccination Coverage Intracluster Correlation Coefficients from Many Recent Studies

DALE RHODA, CAITLIN CLARY, BECCA ROBINSON, MIA YU, MARY KAY TRIMNER, SHIRAZ JOHNSON, JENNIFER BRUSTROM

BIOSTAT GLOBAL CONSULTING

MAY 12, 2022

Outline

Context – sample size for vaccination coverage cluster surveys

Elements of the design effect

Extract elements from recent datasets

Revise World Health Organization (WHO) guidance

Our software can also be helpful for other cluster survey outcomes

Photo credit: Dr. Pierre Claquin, Bangladesh 2014

Survey Planning
Sample Size Steps & Feedback Cycle

1. Quantify inferential goal
   (e.g., confidence interval no wider than +/- 5% if coverage is 85%)
2. Calculate effective sample size
3. Inflate by (guessed) design effect
4. Collect & analyze data
5. Review whether inferential goal was met. If not, why not?
   (e.g., Rhoda et al., 2020, supplemental materials)

Estimate proportion to within +/- X%

Solve equation for confidence intervals to calculate effective sample size (N_{eff})

For a complex sample, we inflate sample size by the design effect (DEFF)

$N_{target} = N_{eff} \times DEFF$

DEFF ≅ **[clustering term]** x **(heterogeneous weights term)**

DEFF ≅ **[1 + (m-1) x ICC]** x **$(1 + CV_w^2)$**

where m = avg # of respondents per cluster
ICC = intracluster correlation coefficient
CV_w= coefficient of variation of survey weights

We Guess at Parameter Values

All three parameters: m, ICC, and CVw are unknown when we do the sample size calculation

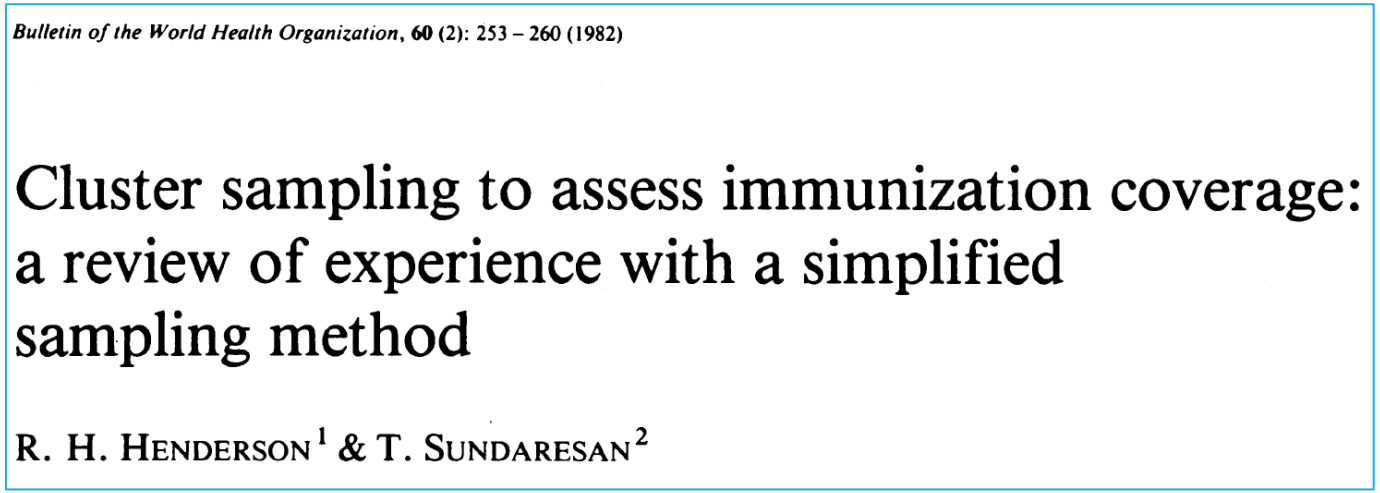
Select values that are conservative, so the CI will be narrow

But not so conservative that we waste $$$ on too large a sample

For Vaccination Coverage

Canonical cluster survey design comes from this 1982 paper for WHO Expanded Programme on Immunization (EPI)

- Visit 30 clusters
- Self-weighted, so CVw = 0
- Quota sample with m = 7
- In most cases, DEFF will be ≤ 2 (which implies ICC ≤ 1/6)


Bulletin of the World Health Organization, **60** (2): 253 – 260 (1982)

Cluster sampling to assess immunization coverage: a review of experience with a simplified sampling method

R. H. HENDERSON[1] & T. SUNDARESAN[2]

Many hundreds of EPI surveys have been conducted using 30 clusters x 7 respondents

For Vaccination Coverage

In 2015, WHO recommendations changed:

- Calculate survey weights based on probability of selection and post-stratify them using population estimates; now $CV_w > 0$
- Instead of a quota sample, select a fixed number of households; now m varies
- Note that in practice, sometimes ICC > 1/6

New manual suggested:

- For well run immunization campaigns, ICC of 1/6 may still be conservative
- For routine immunization surveys, suggest 1/3 as a conservative value

International Collaboration to Update Insight

Contact colleagues with data from recent surveys

Assemble a large dataset of observed values of coverage parameters

Summarize with plots & percentiles

Survey planners can trade off cost of data collection against risk of having CIs that are too wide

We show data today from:

Bangladesh	Guinea Bissau	Liberia	Pakistan	Togo
Burkina Faso	Kenya	Madagascar	Swaziland	
Benin	Kyrgyzstan	Nepal	Tajikistan	
Ethiopia	Lao PDR	Nigeria	The Gambia	

Develop & Share Software

We've written generic code for Stata & R: *iccloop*

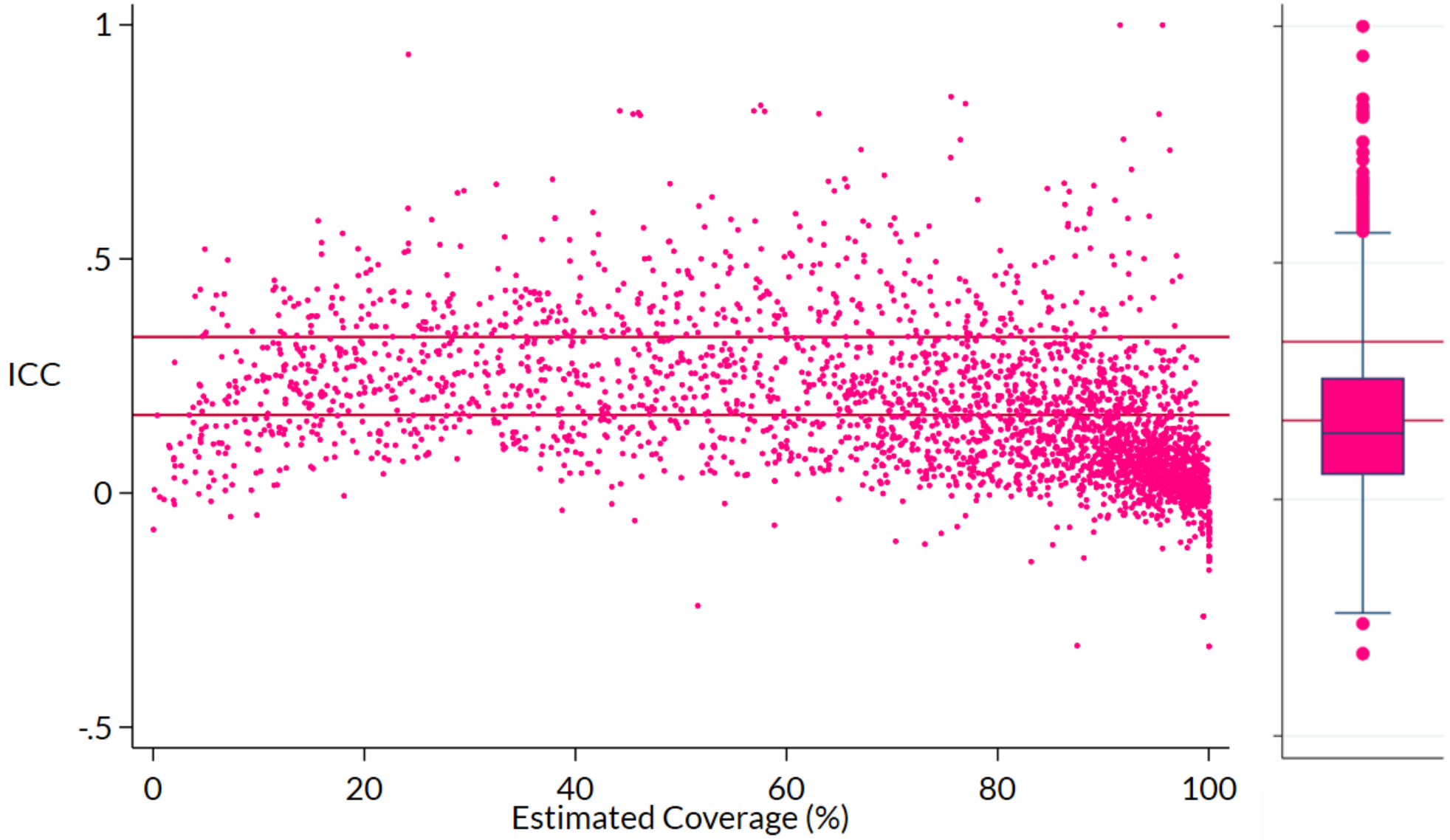
- Define survey sample design
- *iccloop* mines the values of m, ICC, CVw, DEFF, coverage, CI width for binary outcomes over all the survey strata

We developed the programs for vaccination coverage datasets, but the they are generic and could be used to mine parameters for other binary outcomes from complex sample surveys

A Murmuration of Starlings

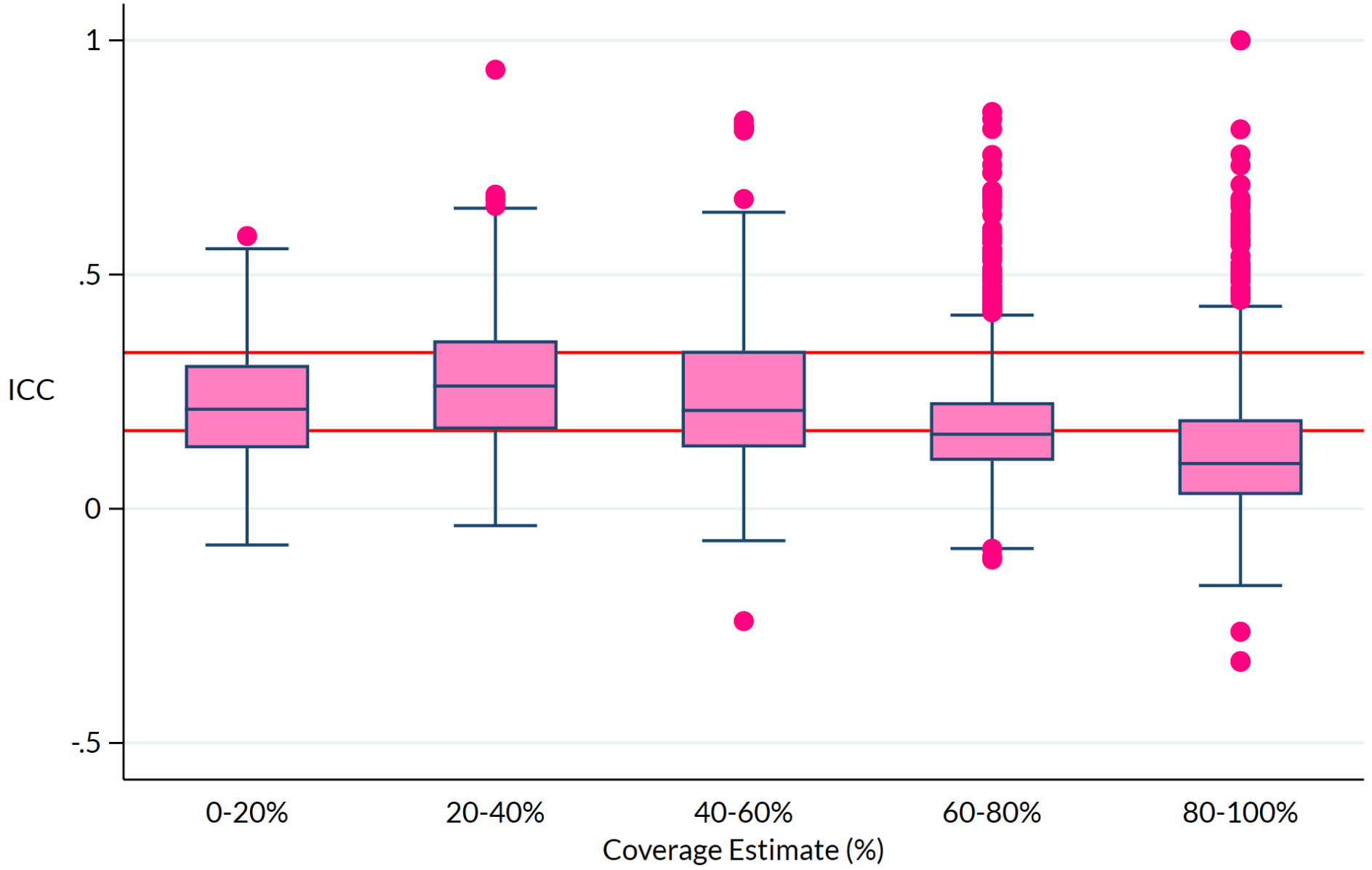
https://www.reuters.com/news/picture/starlings-in-the-sky-idINRTS2XJHH

3,187 Observed ICCs



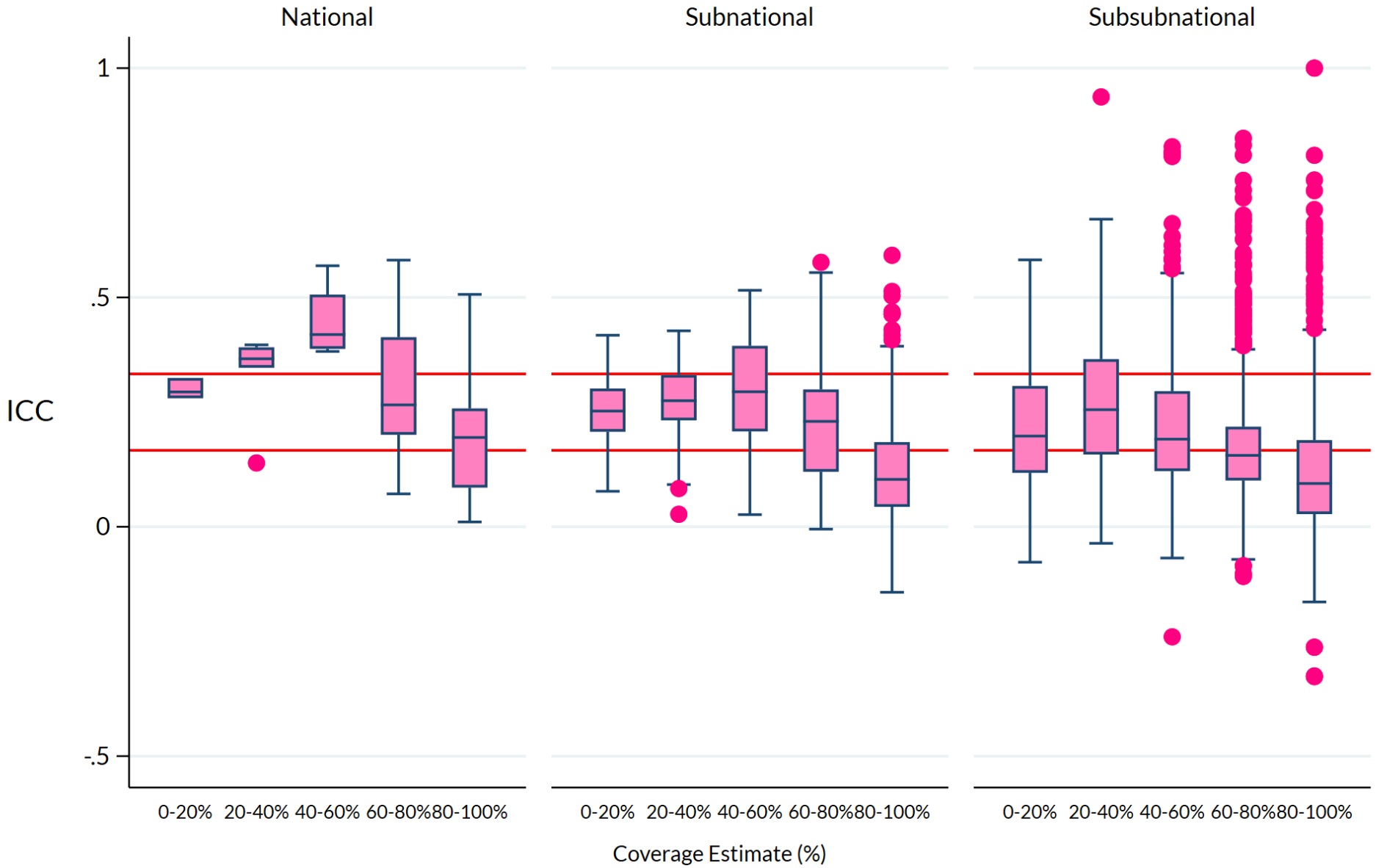
Red lines appear at 1/6 and 1/3

Compiled ICCs



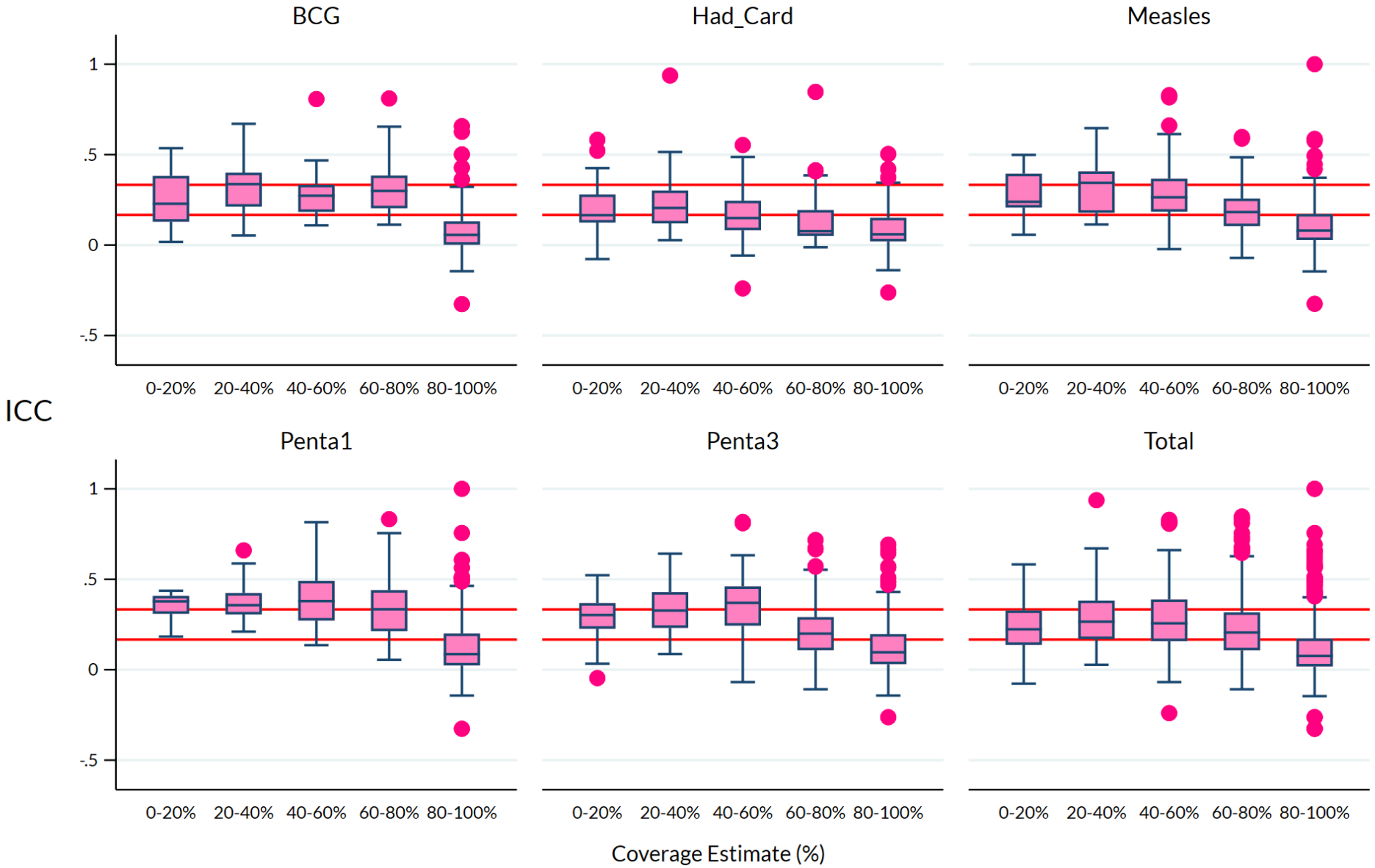
Red lines appear at 1/6 and 1/3

ICCs by Stratum Level



Red lines appear at 1/6 and 1/3

RI ICCs by Outcome



Red lines appear at 1/6 and 1/3

Results

The classic idea that ICC will be ≤ 1/6 is often true when coverage is > 80%

It is often not true when coverage falls between 20% and 80%.

Recent guidance that 1/3 may be conservative is not bad, but too broad

Some outcomes have ICC > 1/3 when coverage is near 50%

There are interesting high ICC outliers even when coverage is > 80%

We would be happy to discuss this further or have others use our software to explore these parameters for other types of surveys

Software Resources

Github repository:

https://github.com/BiostatGlobalConsulting/ExtractSurveyICC-DEFF-CVw

Youtube demo of Stata iccloop resources
https://www.youtube.com/watch?v=Lhl8asVdpf4

Youtube demo of R iccloop resources
https://www.youtube.com/watch?v=cKJTpU98Mrw

Forthcoming R Shiny tool

URL will be in the forthcoming manuscript

References

Henderson RH, Sundaresan T. Cluster sampling to assess immunization coverage: a review of experience with a simplified sampling method. Bulletin of the World Health Organization 1982;60:253. (Link)

World Health Organization. Vaccination Coverage Cluster Surveys: Reference Manual. March 2019; (WHO/IVB/18.09). License: CC BY-NC-SA 3.0 IGO. Geneva: World Health Organization; 2018.

Rhoda DA, Wagai JN, Beshanski-Pedersen BR, Yusafari Y, Sequeira J, Hayford K, et al. Combining cluster surveys to estimate vaccination coverage: Experiences from Nigeria's Multiple Indicator Cluster Survey / National Immunization Coverage Survey (MICS/NICS), 2016–17. Vaccine 2020;38:6174–83. https://doi.org/10.1016/j.vaccine.2020.05.058.

We Are Grateful For

Financial support from Bill & Melinda Gates Foundation

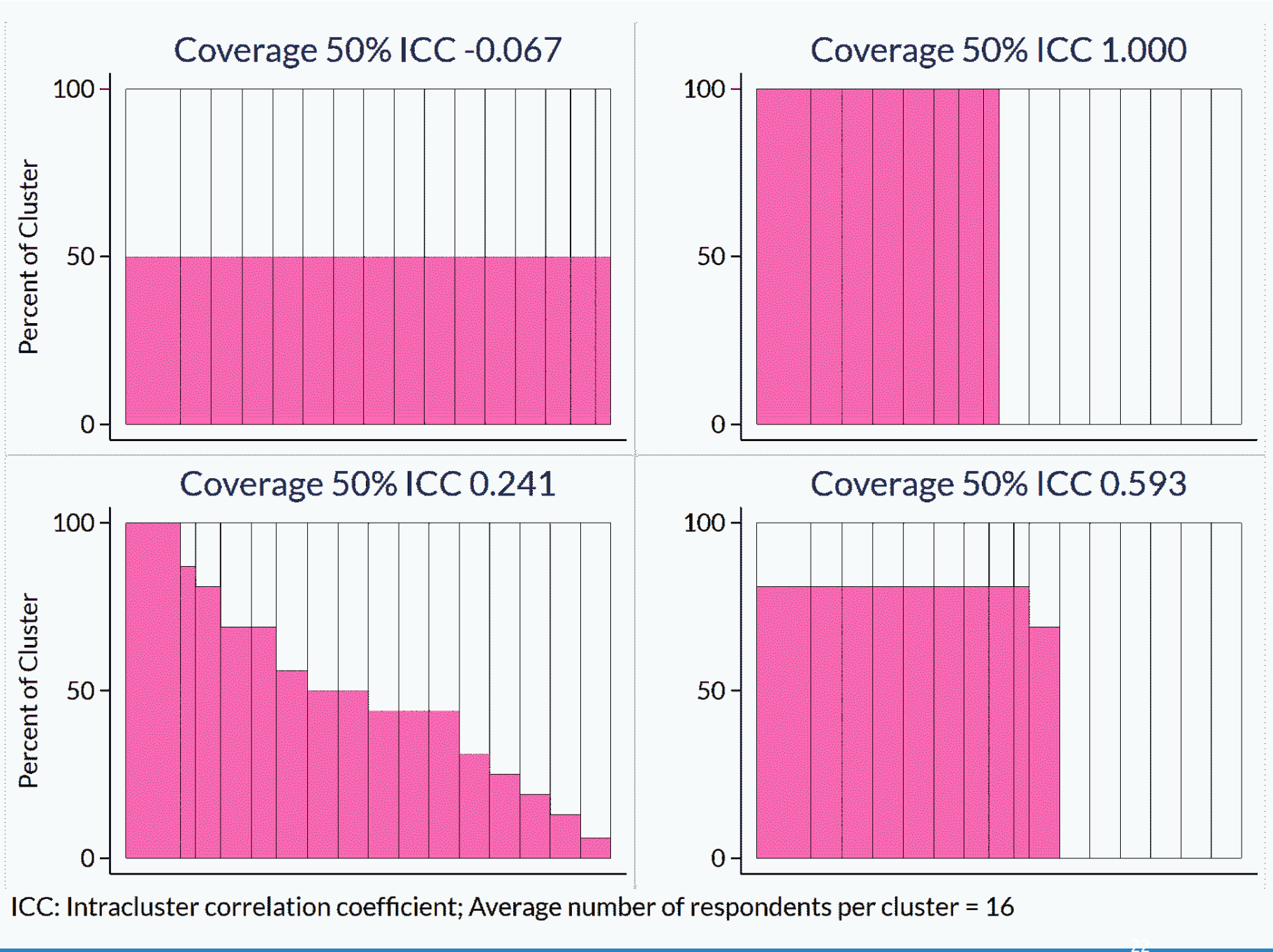
Survey data curation & parameter extraction from colleagues:

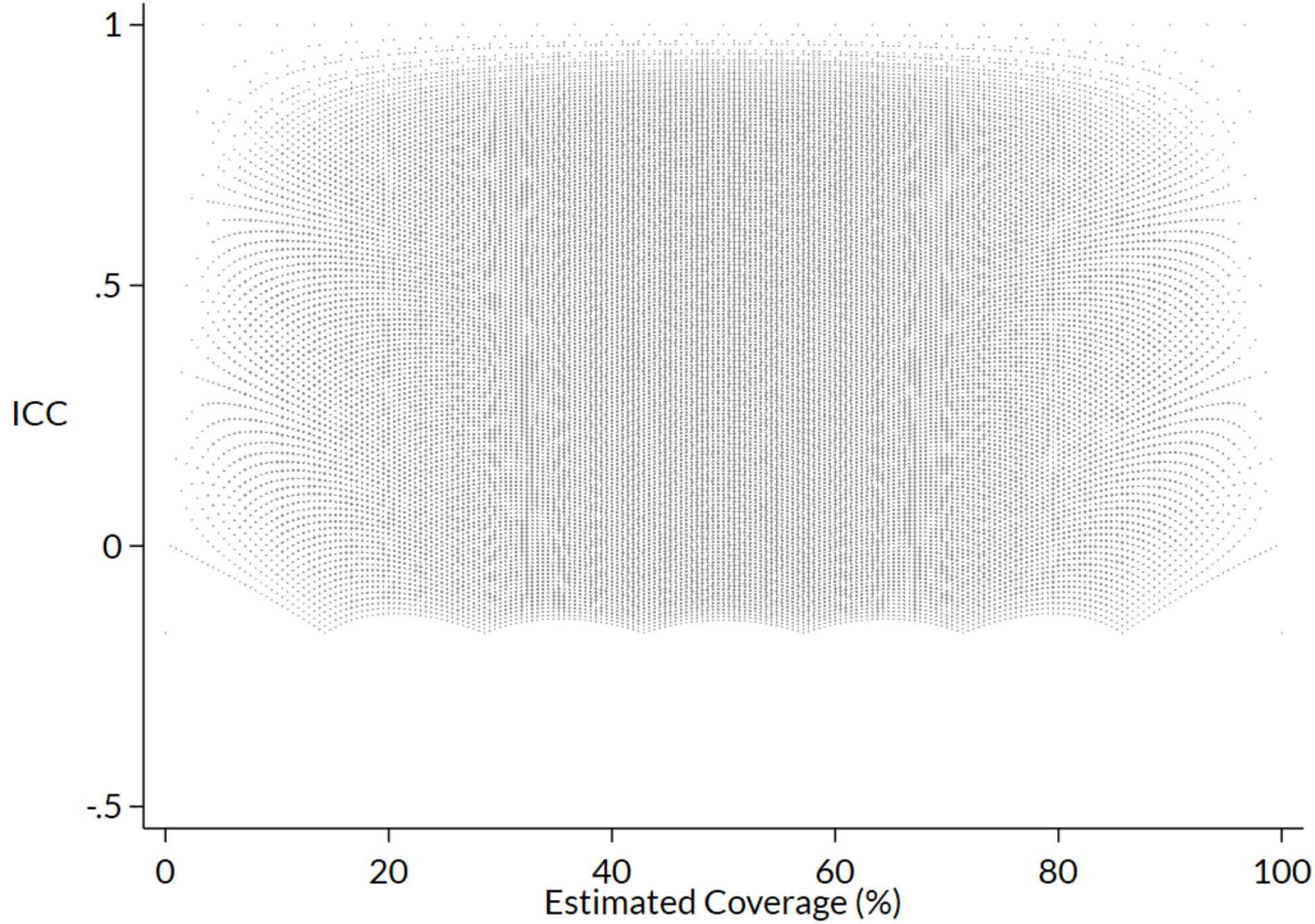
Yemane Berhane
Joseph Nsiari-muzeyi Biey
Carolina Danovaro
Siddhartha Sankar Datta
Theodros Getachew
Wenfeng Gong
Lassane Kabore
Ibrahim Kana
David Koffi
Aliya Kosbayeva
Mona Lacoul
Fortune Themba Mhlanga
Francisco Nogareda
Lorenzo Pezzoli
Jaurès Rabemanantena
Heather Scobie
Sajid Bashir Soofi
Alieu Sowe
Habtamu Teklie
John Wagai


Thank You
Asante
Shukria
Dhanyavadagalu
Vinaka
köszönöm
Kiitos
Maake
Manana
Dankon
Kam Sah Hammida
Matondo
Biyan
Maururu
Dank Je
Dankscheen
спасибо
Blagodaram
Ngiyabonga
Dziekuje
Chokrane
Tack
Taiku
Terima Kasih
Diolch i Chi
Juspaxar
Arigato
Grazie
Gracias
Mochchakkeram
Tingki
Paldies
Khap
Bedankt
Dakujem
Grazas
cảm ơn bạn
Gratias Tibi
Ua Tsaug Rau Koj
Obrigado
Kia Ora
Kop Khun
Děkuji
Nirringrazzjak
Di Ou Mèsi
Suksama
Hvala
Djiere Dieuf
Welalin
Danke
Go Raibh Maith Agat
Eskerrik Asko
Merci
Salamat
Najis Tuke
Matur Nuwun
Rahmat
Misaotra
ขอขอบคุณคุณ

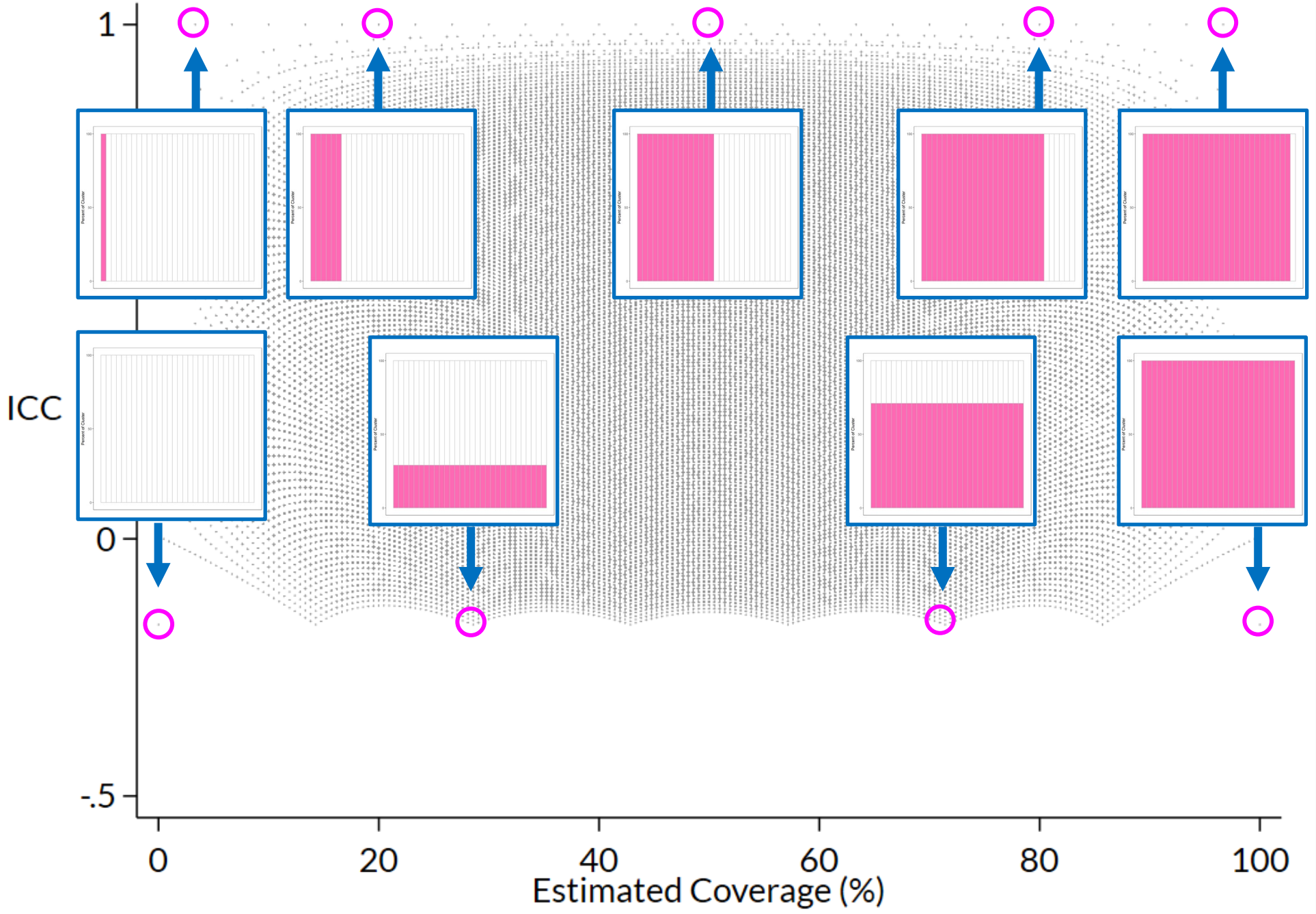
Backup Slides

Organ Pipe Plots

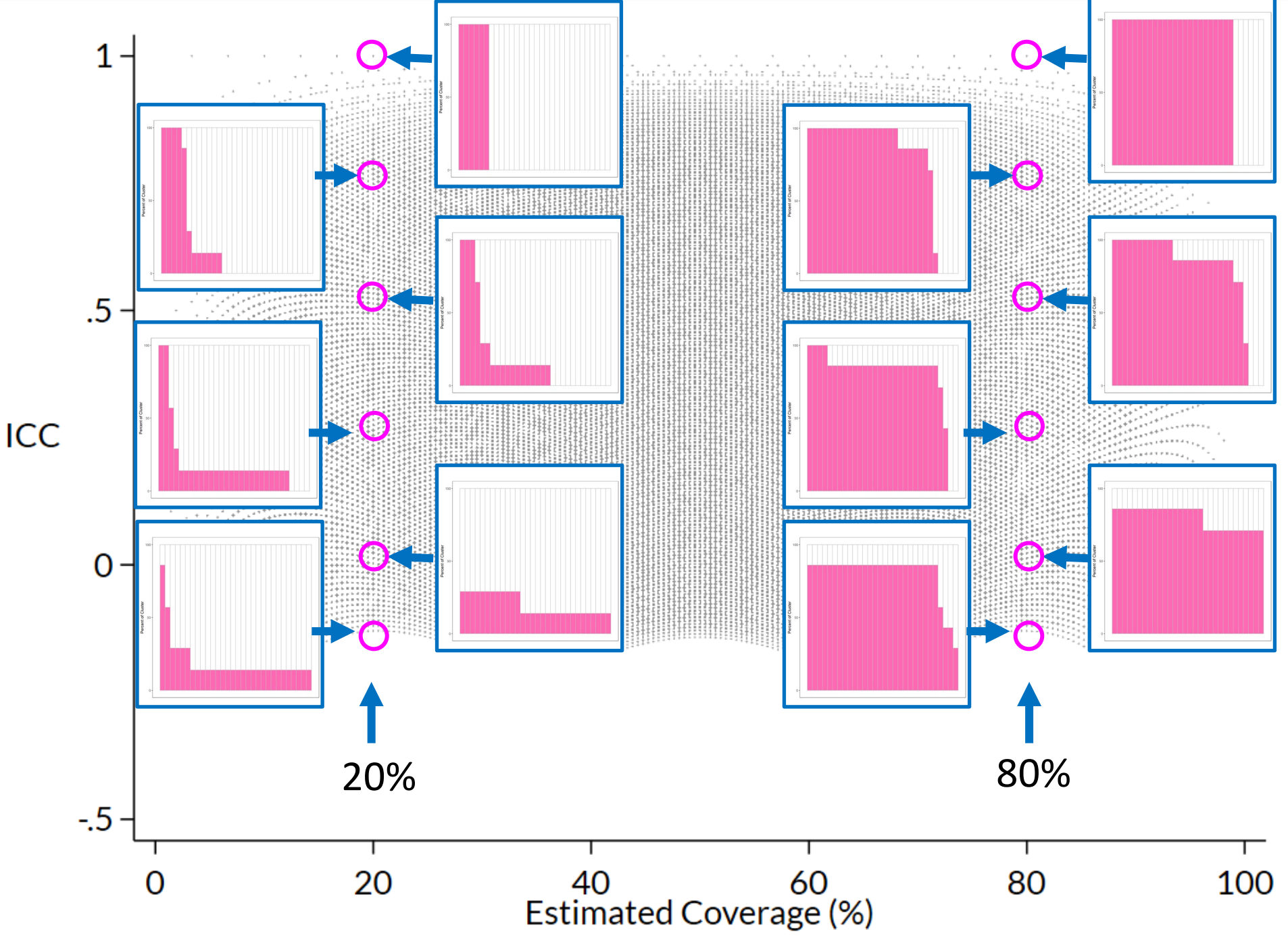



1
.5
0
-.5
ICC
0
20
40
60
80
100
Estimated Coverage (%)

All Possible Combinations of ICC & $\hat{p}$ for a 30 x 7 Cluster Survey Design



All Possible Combinations of ICC & $\hat{p}$ for a 30 x 7 Cluster Survey Design



A Murmuration of ICCs

N=3,187 ICCs from various vaccination coverage survey datasets.