Welcome to the Sentinel Innovation and Methods Seminar Series

The webinar will begin momentarily

Please visit www.sentinelinitiative.org for recordings of past sessions and details on upcoming webinars.

Note: closed-captioning for today's webinar will be available on the recording posted at the link above.

Sentinel

Adaptive Validation Designs: premise and methods

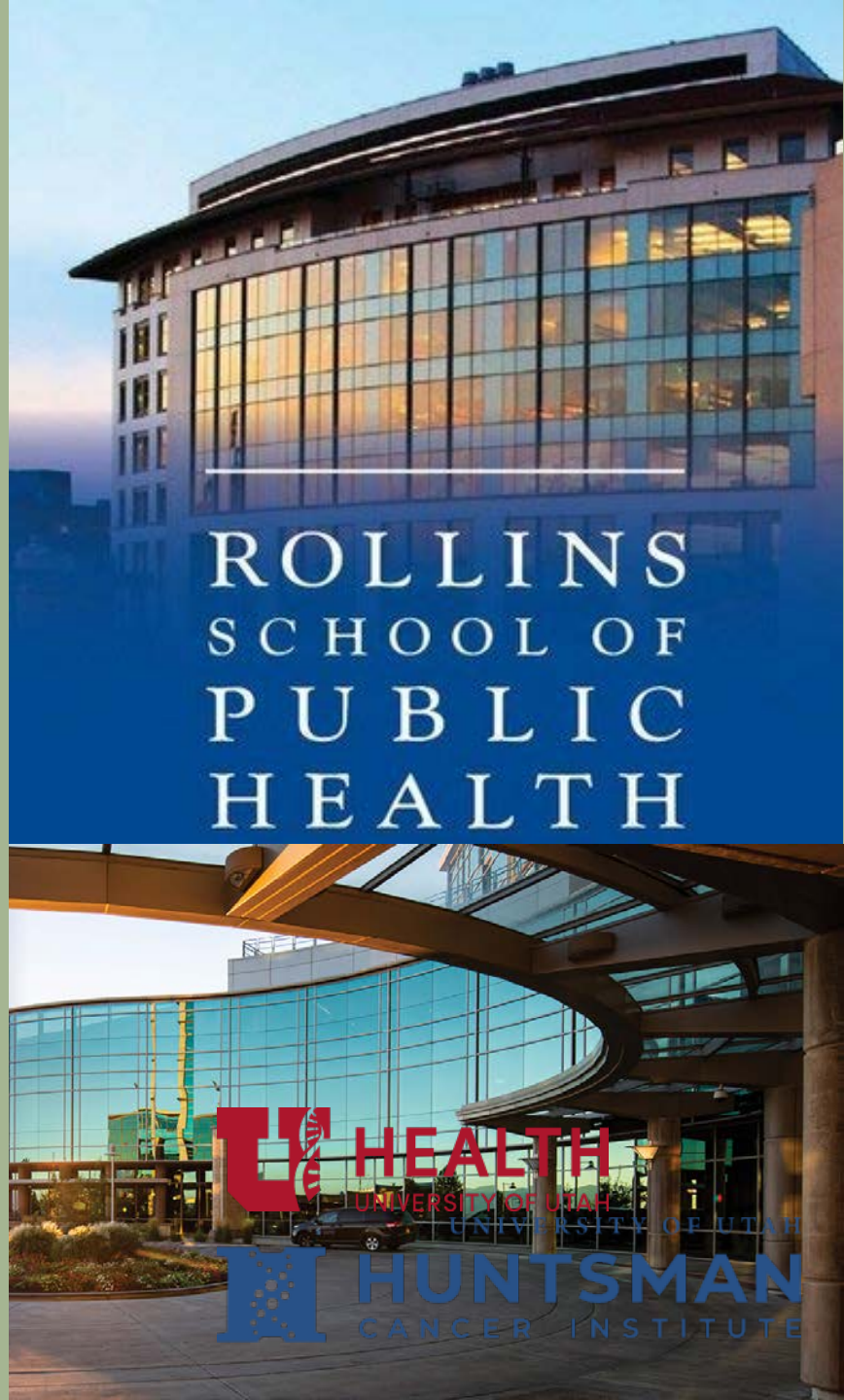
Timothy L. Lash

Emory University

&

Lindsay J. Collin

Huntsman Cancer Institute



Competing Interests

Timothy Lash serves on the methods advisory council for Amgen for which he receives consulting fees and travel support.

Timothy Lash is the author of two epidemiologic methods textbooks related to bias analysis for which he receives royalties.

Lindsay Collin declares no competing interest.

Outline

Tim
Validation study premise and utility to inform bias analysis
Brief illustration of balanced design
Cost-efficiency methods

Lindsay
Framework for adaptive validation design
Stopping criteria
Applied examples
Future directions

Fundamental terminology: information bias

Information bias

The estimate of disease occurrence or of effect is expected to be distorted by inaccurate measurement or classification of the exposure, outcome, or a covariate.

Bias models will usually require

Estimate of the sensitivity and specificity of classification (overall, or within categories of the differential variable). Sometimes also prevalence of the misclassified variable.

Estimate of the positive and negative predictive value of the misclassified variable within categories of (some) other variables in the analysis

Exposure Misclassification
Truth to expected observed

	Truth		Expected Observation	
	X=E+	X=E–	X=E+	X=E–
D+	A	B	$a=s_{D+}A + (1-t_{D+})B$	$b=t_{D+}B + (1-s_{D+})A$
D-	C	D	$c=s_{D-}C + (1 – t_{D-})D$	$d=t_{D-}D + (1 – s_{D-})C$
Total	N_{E+} (A+C)	N_{E-} (B+D)	n_{E+} (a+c)	n_{E-} (b+d)

Non-differential misclassification requires that

$s_{D+}=s_{D-}=s$ and $t_{D+}=t_{D-}=t$.

Exposure Misclassification
Observed to expected truth using PV

	Observation		Expected Truth	
	X=E+	X=E–	X=E+	X=E–
D+	a	b	$A=a*PPV_{D+}+ b*(1-NPV_{D+})$	B=D+ – A
D–	c	d	$C=c*PPV_{D-}+ d*(1-NPV_{D-})$	D=D– – C
Total	n_{E+} (a+c)	n_{E+} (b+d)	N_{E+} (A+C)	N_{E-} (B+D)

Because the margins are fixed (fixed number of cases and non-cases), we can calculate B & D by subtraction

Given an observation and estimates of positive (PPV) and negative (NPV) predictive values within diseased and undiseased, recalculate expected truth

Obtain estimates of PPV and NPV in a substudy with gold standard measurement

Exposure Misclassification
Observed to expected truth

	Observation		Expected Truth	
	X=E+	X=E–	X=E+	X=E–
D+	a	b	$[a–(1–t_{D+})\ D+] / [s_{D+}–(1–t_{D+})]$	B=D+ – A
D–	c	d	$[c–(1–t_{D–})\ D–] / [s_{D+}–(1–t_{D–})]$	D=D– – C
Total	n_{E+} (a+c)	$n_{E–}$ (b+d)	N_{E+} (a+c)	$N_{E–}$ (b+d)

Because the margins are fixed (fixed number of cases and non-cases), we can calculate B & D by subtraction

Given an observation and estimates of sensitivities and specificities, recalculate expected truth

Obtain estimates of s and t from literature, pilot studies, or substudy with gold standard measurement. s and t not necessarily non-differential

Where do we find values to assign?

- Internal sources: substudies or naturally occurring subpopulations
  - substudies: investigator designed to collect information on only a portion of the study / source population
  - naturally occurring: only a subpopulation of the study / source population have information available (not by investigator design)
- External sources
  - Similar studies in similar populations
  - Educated guesses
- Implicit values

Designing substudies: Resources

Study resources are:

- Time (investigator and staff)
- Money (funds available to collect data)

Options for allocation of study resources

- More information (more people, more follow-up)
- Better information (validation substudies)

Designing substudies: Sampling

Sampling of the study population is almost always required. Otherwise, “validation” data are available for the whole study / source population

Sampling has consequences for:

- Precision of estimated validation parameters
- Validation parameters that can be calculated

International Epidemiological Association

International Journal of Epidemiology, 2020, 1–5
doi: 10.1093/ije/dyaa090
Education Corner

Education Corner

Common misconceptions about validation studies

Matthew P. Fox,[1,2]* Timothy L. Lash[3] and Lisa M. Bodnar[4]

[1]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA,
[2]Department of Global Health, Boston University School of Public Health, Boston, MA, USA,
[3]Department of Epidemiology, Rollins School of Public Health, Emory University, Boston, MA, USA and
[4]Department of Epidemiology, University of Pittsburgh School of Public Health, Pittsburgh, PA, USA

International Journal of Epidemiology, 2020, 1–5
doi: 10.1093/ije/dyaa090
Education Corner

Education Corner

Common misconceptions about validation studies

Matthew P. Fox,[1,2]* Timothy L. Lash[3] and Lisa M. Bodnar[4]

[1]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA,
[2]Department of Global Health, Boston University School of Public Health, Boston, MA, USA,
[3]Department of Epidemiology, Rollins School of Public Health, Emory University, Boston, MA, USA and
[4]Department of Epidemiology, University of Pittsburgh School of Public Health, Pittsburgh, PA, USA

234 people completed survey
58% said they had completed training, 42% were current students
35% said they had ever designed or implemented a study to validate a measurement
36% said their coursework included no information about validation study design
55% said they were very or somewhat confident they could design and implement a validation study

Hypothetical Example: Information Bias

Truth	Exposed	Unexposed
Cases	90	270
Non-cases	9910	89730
Total	10000	90000
Risk	0.009	0.003
Difference	0.006	
Ratio	3	

Hypothetical Example: Information Bias

81=0.6·90+
(1–0.9)·270

0.6 sensitivity
0.9 specificity

Truth	Exposed	Unexposed
Cases	90	270
Non-cases	9910	89730
Total	10000	90000
Risk	0.009	0.003
Difference	0.006	
Ratio	3	

Observed	Exposed	Unexposed
Cases	81	279
Non-cases	14919	84721
Total	15000	85000
Risk	0.0054	0.0033
Difference	0.0021	
Ratio	1.65	

Hypothetical Example: Information Bias

Truth	Exposed	Unexposed
Cases	90	270
Non-cases	9910	89730
Total	10000	90000
Risk	0.009	0.003
Difference	0.006	
Ratio	3	

0.6 sensitivity
0.9 specificity

Observed	Exposed	Unexposed
Cases	81	279
Non-cases	14919	84721
Total	15000	85000
Risk	0.0054	0.0033
Difference	0.0021	
Ratio	1.65	

0.6*90/81

0.9*270/279

	cases	non-cases
PPV	0.667	0.399
NPV	0.871	0.953

0.6*9910/14919

0.9*89730/84721

Hypothetical Example: Information Bias

Cases

10% sample Observed	Truth Exposed	Unexposed	Total
Exposed	5	3	8
Unexposed	4	24	28
Total	9	27	
	sensitivity	0.556	0.24, 0.84
	specificity	0.889	0.73, 0.97
	PPV	0.625	0.28, 0.89
	NPV	0.857	0.69, 0.95

Non-cases

10% sample Observed	Truth Exposed	Unexposed	Total
Exposed	595	897	1492
Unexposed	396	8076	8472
Total	991	8973	
	sensitivity	0.600	0.57, 0.63
	specificity	0.900	0.89, 0.91
	PPV	0.399	0.37, 0.42
	NPV	0.953	0.95, 0.96

Simple Random Sample:
Get all the right answers, but statistically inefficient and costly

Hypothetical Example: Information Bias

Cases — balanced, observed

	Truth		
Observed	Exposed	Unexposed	Total
Exposed	54	27	81
Unexposed	10	71	81
Total	64	98	162
sensitivity	0.844	0.74, 0.92	
specificity	0.724	0.63, 0.81	
PPV	0.667	0.56, 0.76	
NPV	0.877	0.79, 0.94	

Non-cases — balanced, observed

	Truth		
Observed	Exposed	Unexposed	Total
Exposed	32	49	81
Unexposed	4	77	81
Total	36	126	162
sensitivity	0.889	0.75, 0.96	
specificity	0.611	0.52, 0.69	
PPV	0.395	0.29, 0.50	
NPV	0.951	0.89, 0.98	

less costly, cannot estimate sensitivity or specificity

Holcroft CA, Spiegelman D. Design of validation studies for estimating the odds ratio of exposure-disease relationships when exposure is misclassified. Biometrics. 1999 Dec;55(4):1193–201.

Hypothetical Example: Information Bias

Cases — balanced, truth

	Truth		
Observed	Exposed	Unexposed	Total
Exposed	54	9	63
Unexposed	36	81	117
Total	90	90	180
sensitivity	0.600	0.50, 0.70	
specificity	0.900	0.82, 0.95	
PPV	0.857	0.75, 0.93	
NPV	0.692	0.60, 0.77	

Non-cases — balanced, truth

	Truth		
Observed	Exposed	Unexposed	Total
Exposed	54	9	63
Unexposed	36	81	117
Total	90	90	180
sensitivity	0.600	0.50, 0.70	
specificity	0.900	0.82, 0.95	
PPV	0.857	0.75, 0.93	
NPV	0.692	0.60, 0.77	

specificity, much less costly, cannot estimate PPV or NPV

Holcroft CA, Spiegelman D. Design of validation studies for estimating the odds ratio of exposure-disease relationships when exposure is misclassified. Biometrics. 1999 Dec;55(4):1193–201.

Adaptive Design

Validation study design inherently involves sampling, and often involves expense

Can conduct validation study sampling until reaching a threshold

- Bias parameter measured well enough to stop
- Bias-adjusted estimate measured well enough to stop
- Resource allocation favors study size over validation study size

Cost-efficient validation designs

Typically assume

- Sample size constrained by fixed budget or budget proportion
- Fixed price per validated record

See, for examples

Spiegelman D, Gray R. Cost-Efficient Study Designs for Binary Response Data with Gaussian Covariate Measurement Error. Biometrics. 1991;47(3):851–69.

Stram DO, Longnecker MP, Shames L, Kolonel LN, Wilkens LR, Pike MC, et al. Cost-efficient design of a diet validation study. Am J Epidemiol. 1995 Aug 1;142(3):353–62.

Spiegelman D. Cost-efficient study designs for relative risk modeling with covariate measurement error. J Stat Plan Inference. 1994 Nov 1;42(1):187–208.

Adaptive Validation Design

Adaptive Design: background

Previous guidance on sampling of participants for validation studies apply to scenarios where the study population enrollment and follow-up have been completed.

Alternatively, researchers may want to collect validation data prospectively and identify at which point sufficient validation data have been collected, potentially saving time and resources.

Adaptive Design: background

This design uses the framework of Bayesian monitoring techniques, often used in clinical trials to monitor treatment response over time.

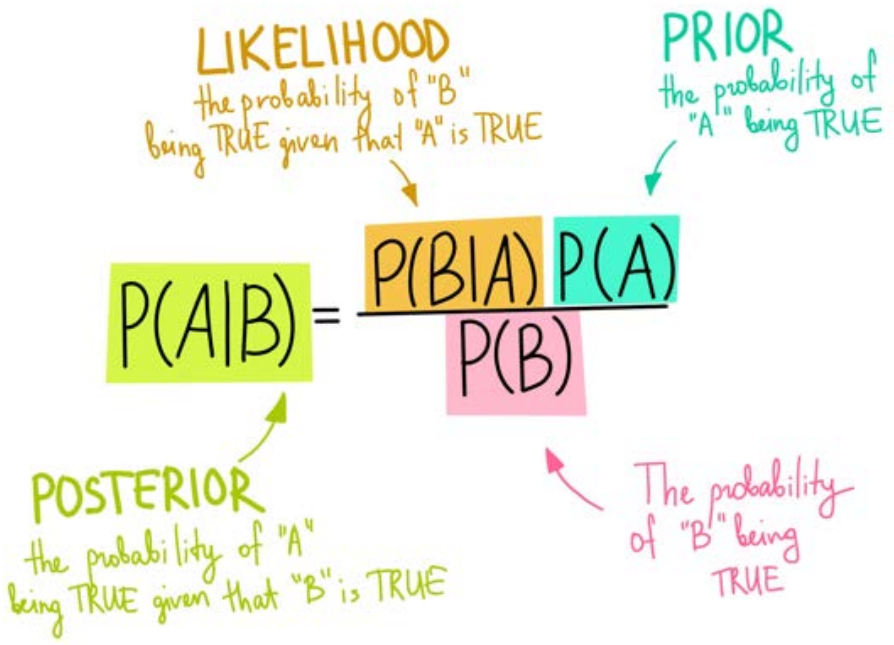
In our approach, we extend this framework to inform when sufficient validation data have been collected to meet the goals of validation.

Adaptive Design: approach

We iteratively update a prior (the classification parameter of interest) with validation data at specified time points until stopping criteria have been met.

Classification Parameters:

- Positive and Negative Predictive Values (PPV and NPV)
- Sensitivity and Specificity



Adaptive Design: approach

We start with a uniform prior, which assumes all values (0–1) of the classification parameter as being equally likely, $Beta\,(1,1)$.

We then update the parameters in blocks of participants.

Observed Exposure	Gold standard Exposure: E+	Gold standard Exposure: E−
E+	a	b
E−	c	d

Posterior: $\boldsymbol{\theta|y_j \sim Beta\,(\alpha_j, \beta_j)}$

We update the classification parameters in a beta-binomial Bayesian model.

Adaptive Design: approach

Classification parameters are updated at specified time intervals while the validation data accrue

$$\boldsymbol{Time\ 1}: \underbrace{p(\theta|y_1)}_{\text{Posterior}} \propto \underbrace{p(y_1|\theta)}_{\text{Likelihood}} * \underbrace{p(\theta)}_{\text{Prior}}$$

$$\boldsymbol{Time\ 2}: p(\theta|y_2, y_1) \propto p(y_2|\theta) * p(\theta|y_1)$$

$$\boldsymbol{Time\ j}: p(\theta|y_j, \dots y_2, y_1) \propto p(y_j|\theta) * p(\theta|y_{j-1}, \dots, y_2, y_1)$$

Adaptive Design: stopping criteria

Goals of validation often vary across study designs

Can include:

1) Precision of classification parameters
2) Efficacy/futility of validation
3) Precision of bias-adjusted estimate of association

Adaptive Validation Design

Stopping based on precision of classification parameters

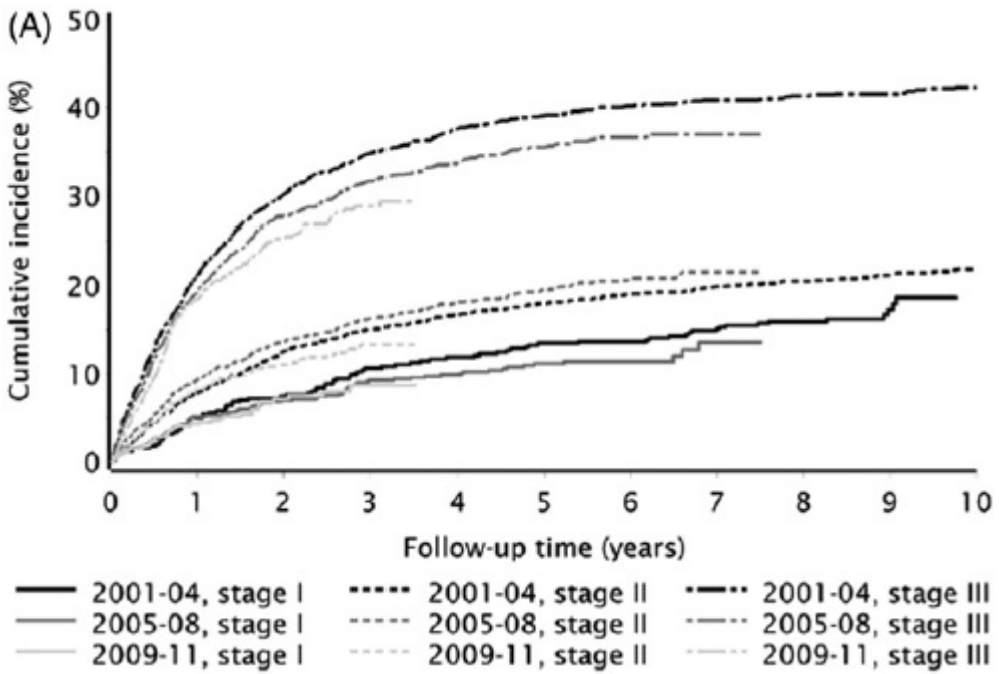
Adaptive Design: example

Colorectal cancer is the third most common malignancy and the third leading cause of cancer related mortality globally.

Despite improved survival, the 5-year risk of recurrence is estimated to be 20–50% depending on stage and other tumor characteristics at diagnosis.

Cancer recurrence is not routinely collected by most population-based registries.

Cumulative Incidence of CRC recurrence



Holmes, Riis, Erichsen et al. 2017, *Acta Oncologica*

Adaptive Design: example

Use the adaptive validation study design to model when sufficient validation data have been been collected to validate CRC recurrences identified through an algorithm.

Adaptive Design: study population

Danish CRC patients who underwent surgery and registered with the Danish Colorectal Cancer Group (DCCG) database.

Patients were enrolled (n=355) and actively followed biennially for colorectal cancer recurrence, 63 (18%) developed a recurrence over follow-up.

IJC
International Journal of Cancer

A validated algorithm to ascertain colorectal cancer recurrence using registry resources in Denmark

Timothy L. Lash[1,2], Anders H. Riis[2], Eva B. Ostenfeld[2,3], Rune Erichsen[2], Mogens Vyberg[4] and Ole Thorlacius-Ussing[3]

[1] Department of Epidemiology, Rollins School of Public Health, Atlanta, GA
[2] Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark
[3] Department of Gastrointestinal Surgery, Aalborg University Hospital, Aalborg, Denmark
[4] Institute for Pathology, Aalborg University Hospital, Aalborg, Denmark

Adaptive Design: approach

Two validation substudies

1. Sample based on algorithm defined recurrences and validate against gold standard of follow-up. This allows for estimation of PPV and NPV.

2. Sample based on actively followed cohort and compare with algorithm defined recurrences. This allows for estimation of Sensitivity and Specificity.

Adaptive Design: stopping criteria

Predefined threshold value and level of precision

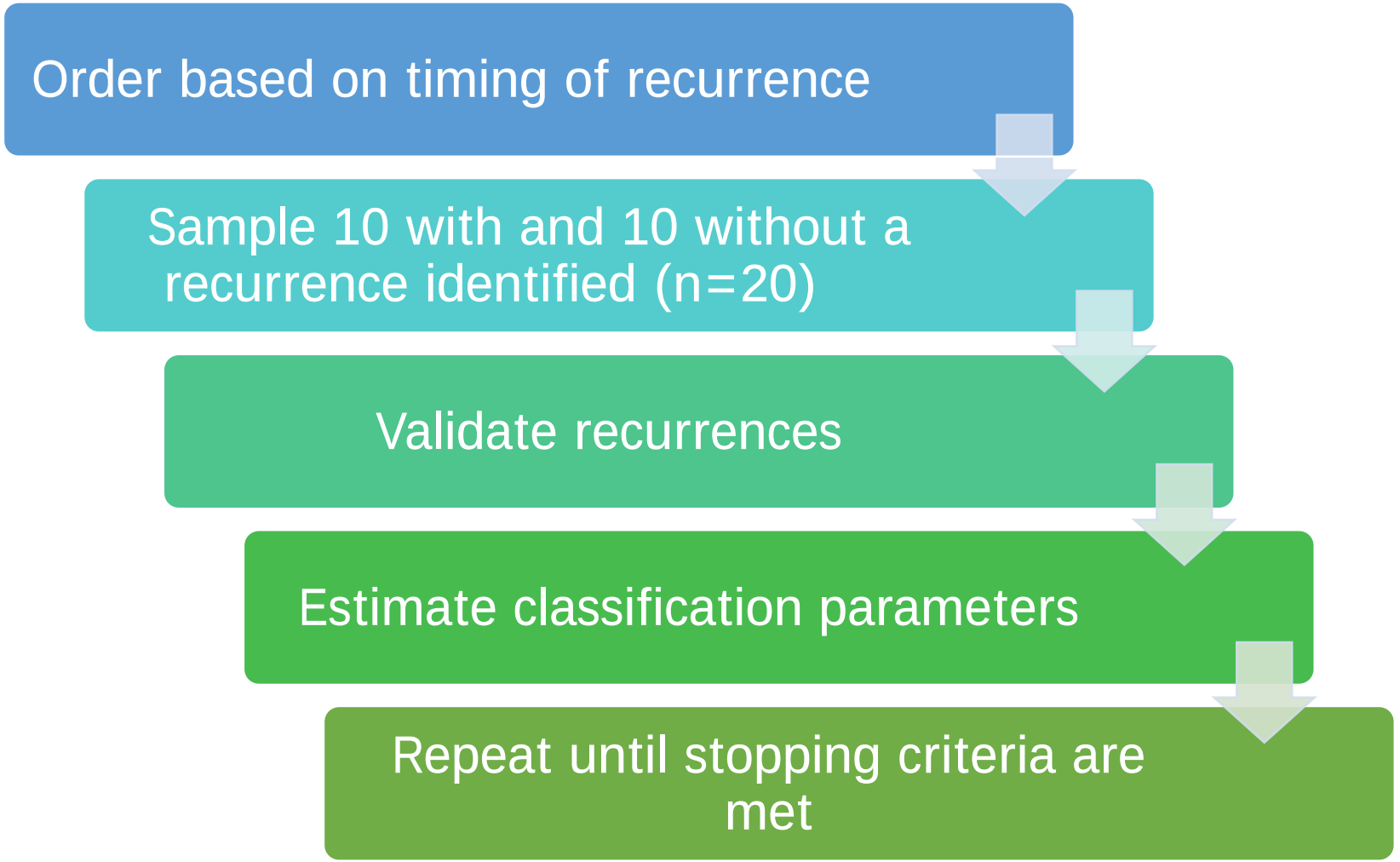
PPV and NPV: threshold= lower 95% credible bound for PPV and NPV >0.80

precision= 0.15

Se and Sp: threshold= lower 95% credible bound for Se and Sp >0.90

precision= 0.08

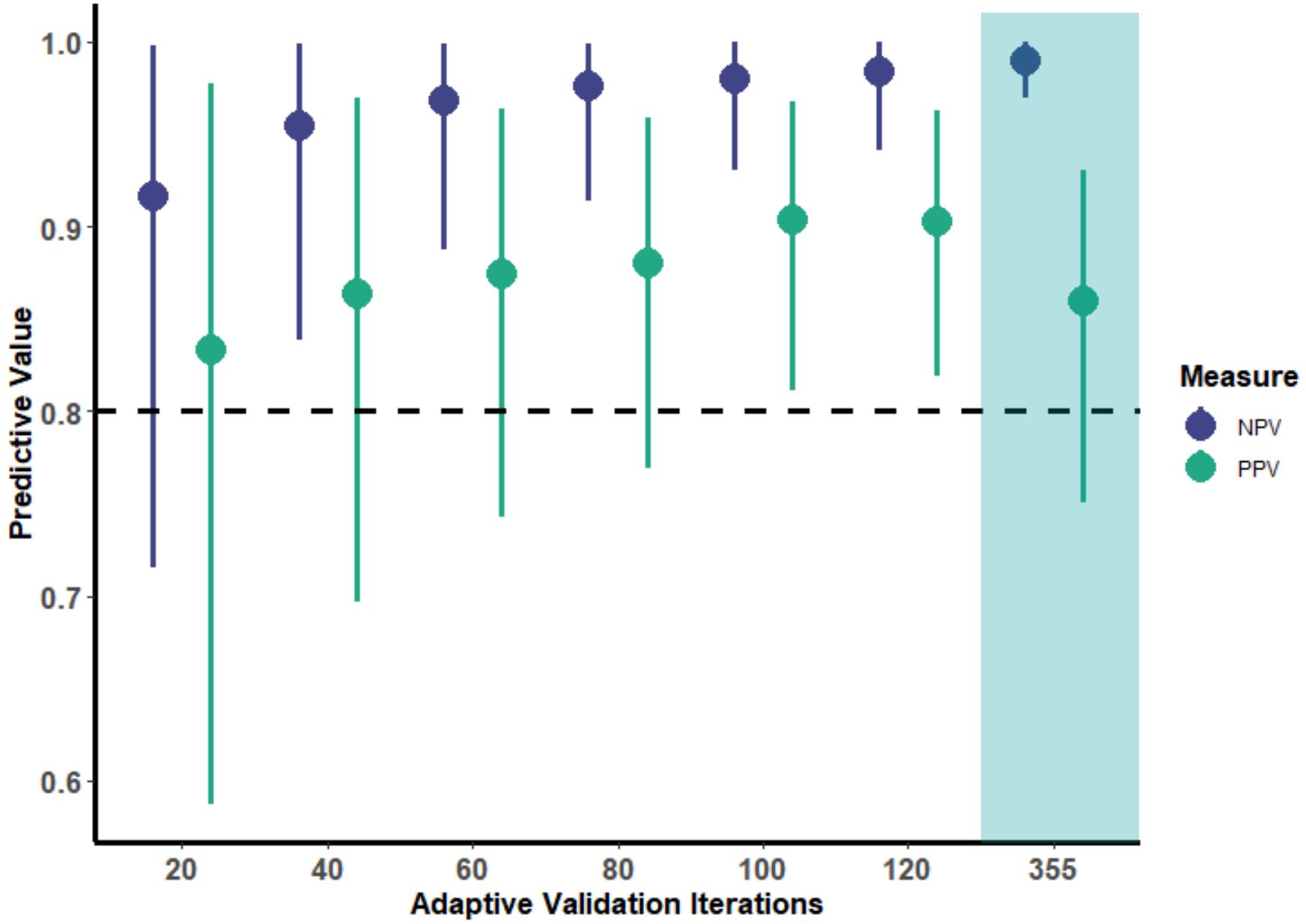
Adaptive Design: sampling strategy



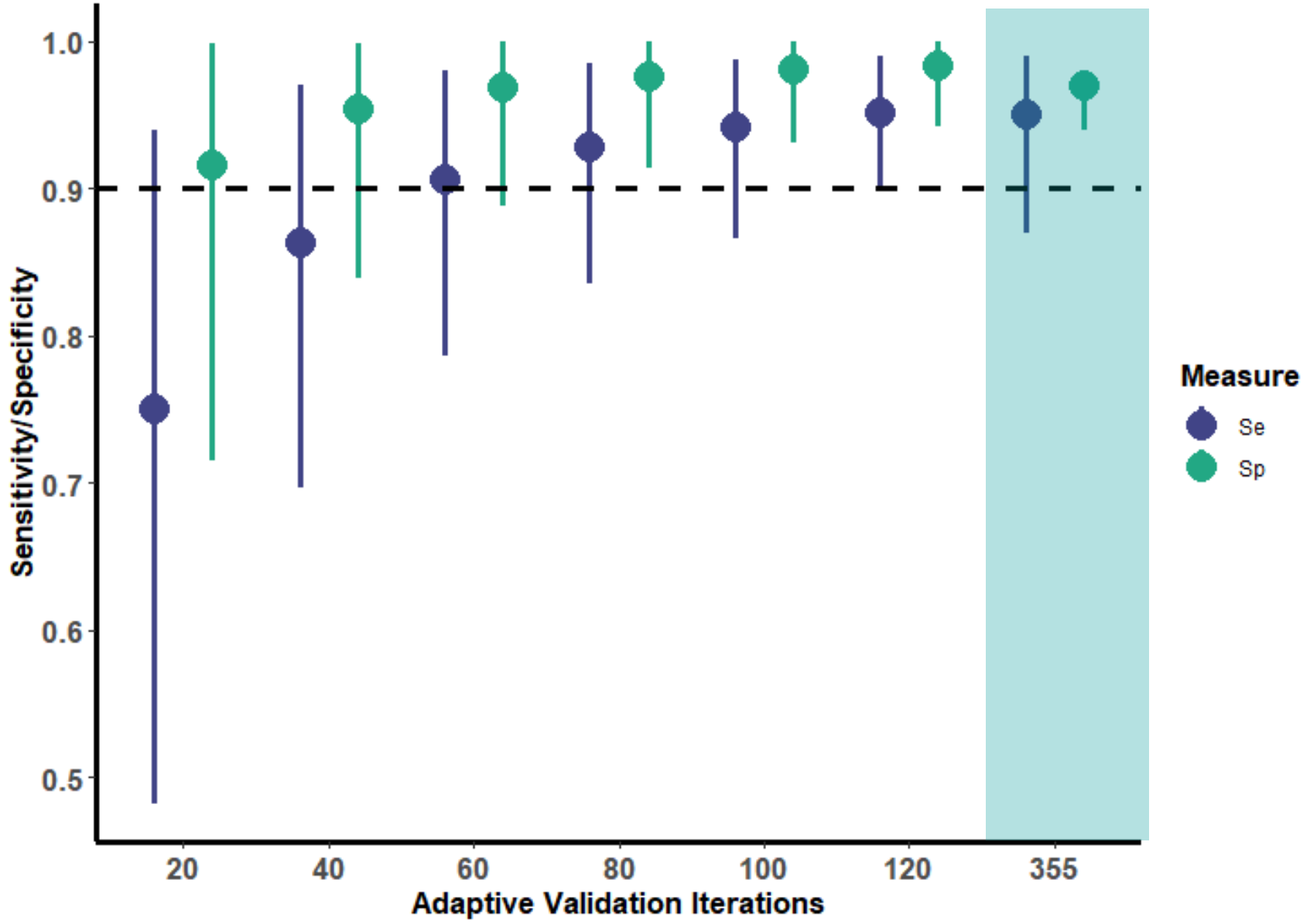
Adaptive Design: results

Method	PPV	NPV	Number Validated
Complete data	0.86 (0.75, 0.93)	0.99 (0.97, 1.00)	355
Adaptive Validation –20	0.90 (0.82, 0.96)	0.98 (0.94, 1.00)	120
	Sensitivity	Specificity	
Complete data	0.95 (0.87, 0.99)	0.97 (0.94, 0.98)	355
Adaptive Validation –20	0.95 (0.90, 0.99)	0.98 (0.94, 1.00)	120

Positive and Negative Predictive Values



Sensitivity and Specificity



Adaptive Design: summary

Classification parameters estimated using the adaptive validation approach were similar to those computed using complete validation.

Considerations:

Stopping criteria should be informed based on subject matter knowledge and goals for validation.

PPV/NPV within strata of exposure/outcome if using classification parameters in bias analysis.

Sensitivity/Specificity require subset of population to have gold standard available, or random sample (which can be inefficient).

N included at each iteration of validation.

Adaptive Validation Design

Stopping based on efficacy/futility of validation

Adaptive Design: example 2

Study of Transition, Outcomes and Gender (STRONG) cohort was established to understand long-term effects of hormone therapy and surgery on gender dysphoria, mental health, and chronic illnesses.

Electronic health record-based study of individuals identifed from Kaiser Permanente health plans in Georgia, Northern California, and Southern California between 2006 and 2014.

Study populations included transgender and gender nonconforming children and adolescents (n=1,331) and adults (n=4,725).

Adaptive Design: exposure validation

The exposure of interest was transmasculine and transfeminine status, which can be determined from knowledge of the sex recorded at birth and current gender identity.

The misclassified sex recorded at birth was based on the recent electronic health record (EHR) data and known to be misclassified because it could either represent sex recorded at birth or concurrent gender identity.

The misclassified sex recorded at birth variable was validated by medical record review for all members who were ≥18 years old as of January 1, 2015 (n=535; 40% of the youth subcohort and 100% adult subcohort).

Adaptive Design: exposure validation

We estimate classification parameters (PPV and NPV)

Misclassified sex recorded at bith	Gold standard sex recorded at birth: TM	Gold standard sex recorded at birth: TF	
TM	a	b	PPV=a/(a+b)
TF	c	d	NPV=d/(c+d)

TM=transmasculine; TF=transfeminine

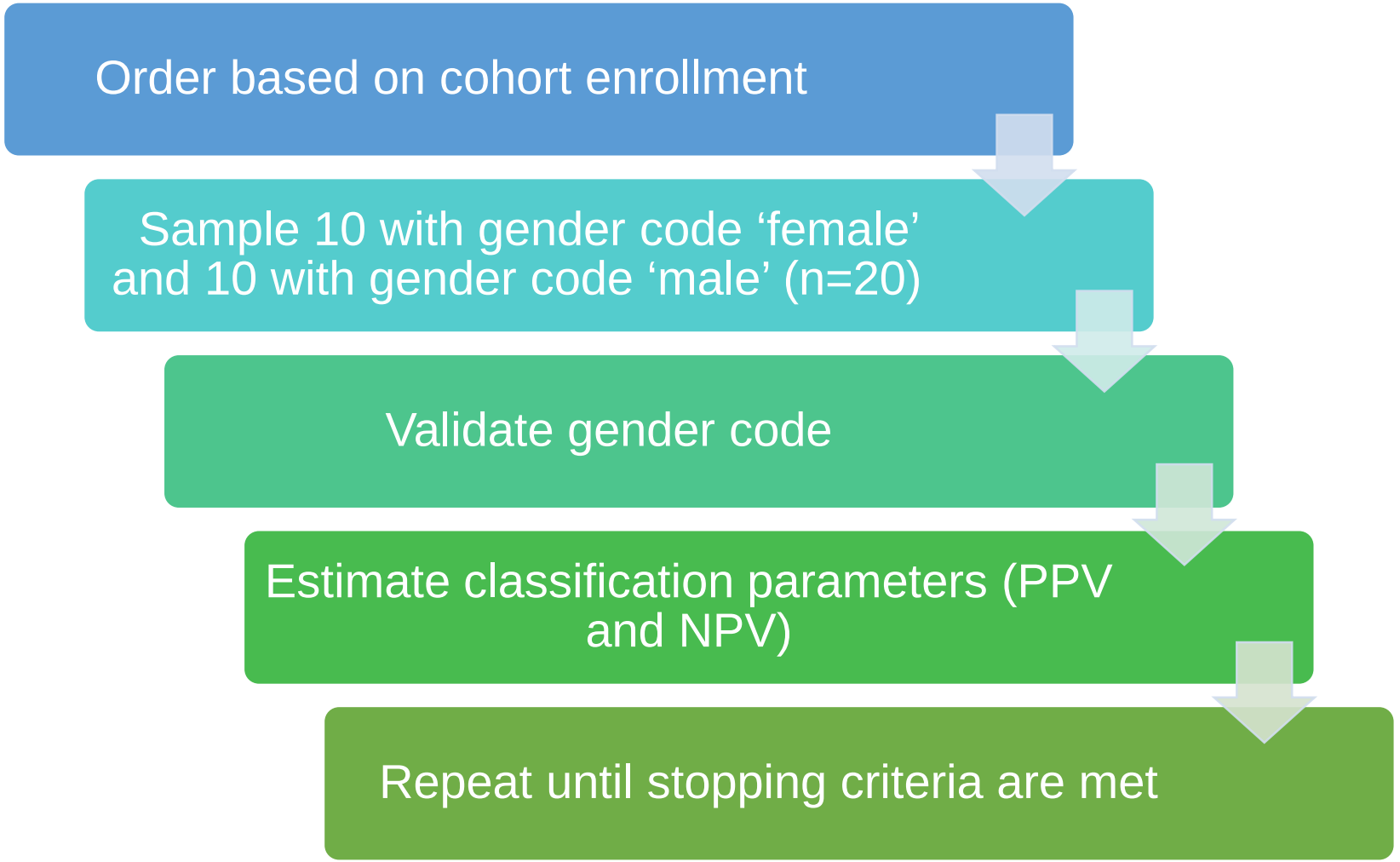
Adaptive Design: stopping criterion

Predefined threshold value

Efficacy: threshold= lower 95% credible bound for PPV and NPV >0.60

Futility: threshold= upper 95% credible bound for PPV and NPV <0.60

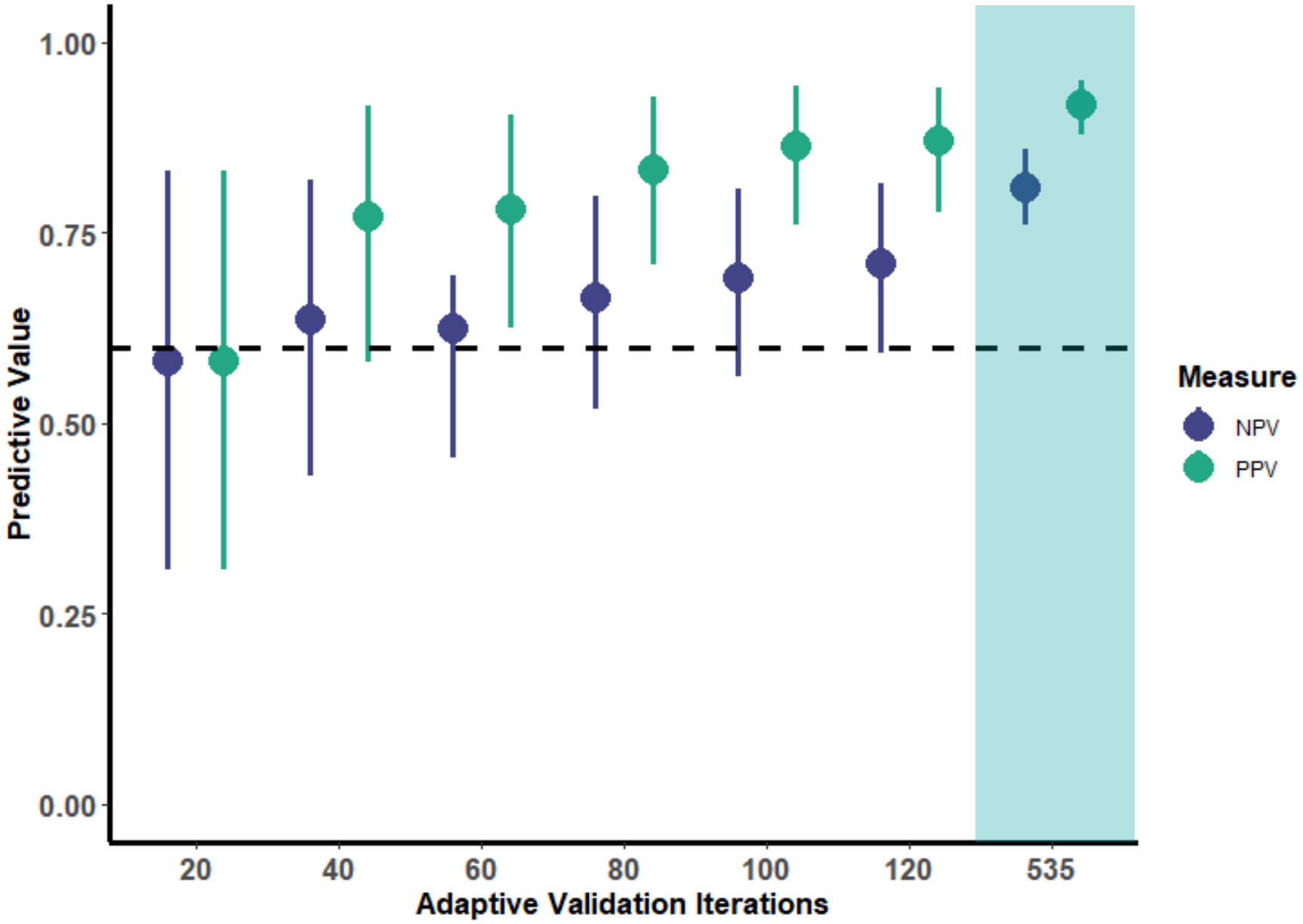
Adaptive Design: sampling strategy



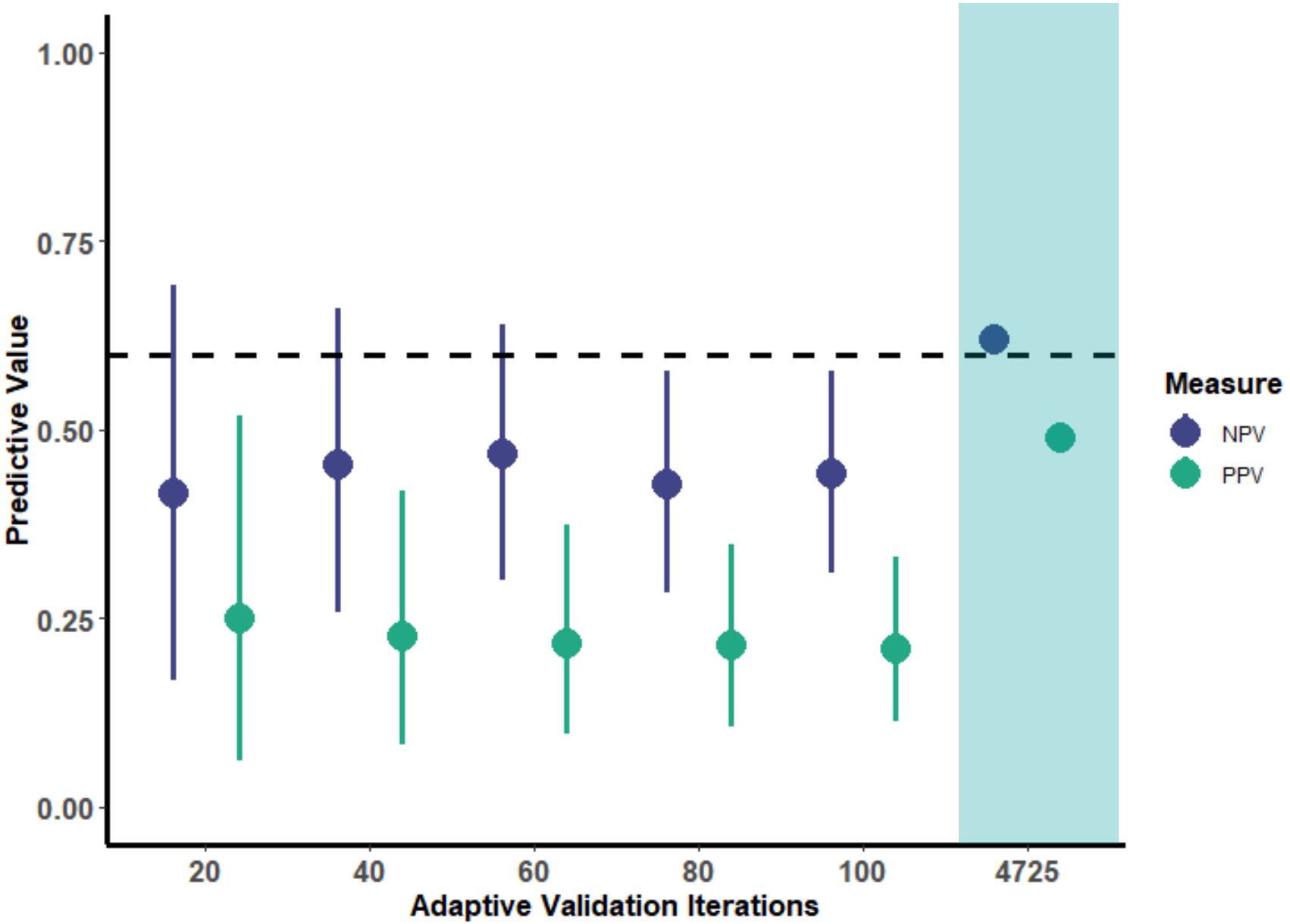
Adaptive Design: results

	Method	PPV	NPV	Number Validated	Days into Study Period
Youth	**Complete data**	0.92 (0.88, 0.95)	0.81 (0.76, 0.86)	535	2897
	Adaptive Validation –20	0.87 (0.79, 0.95)	0.71 (0.60, 0.82)	120	711
Adult	**Method**	**PPV**	**NPV**	**Number Validated**	**Days into Study Period**
	Complete data	0.49 (0.47, 0.51)	0.62 (0.60, 0.64)	4725	3921
	Adaptive Validation –20	0.25 (0.04, 0.48)	0.42 (0.16, 0.68)	20	3

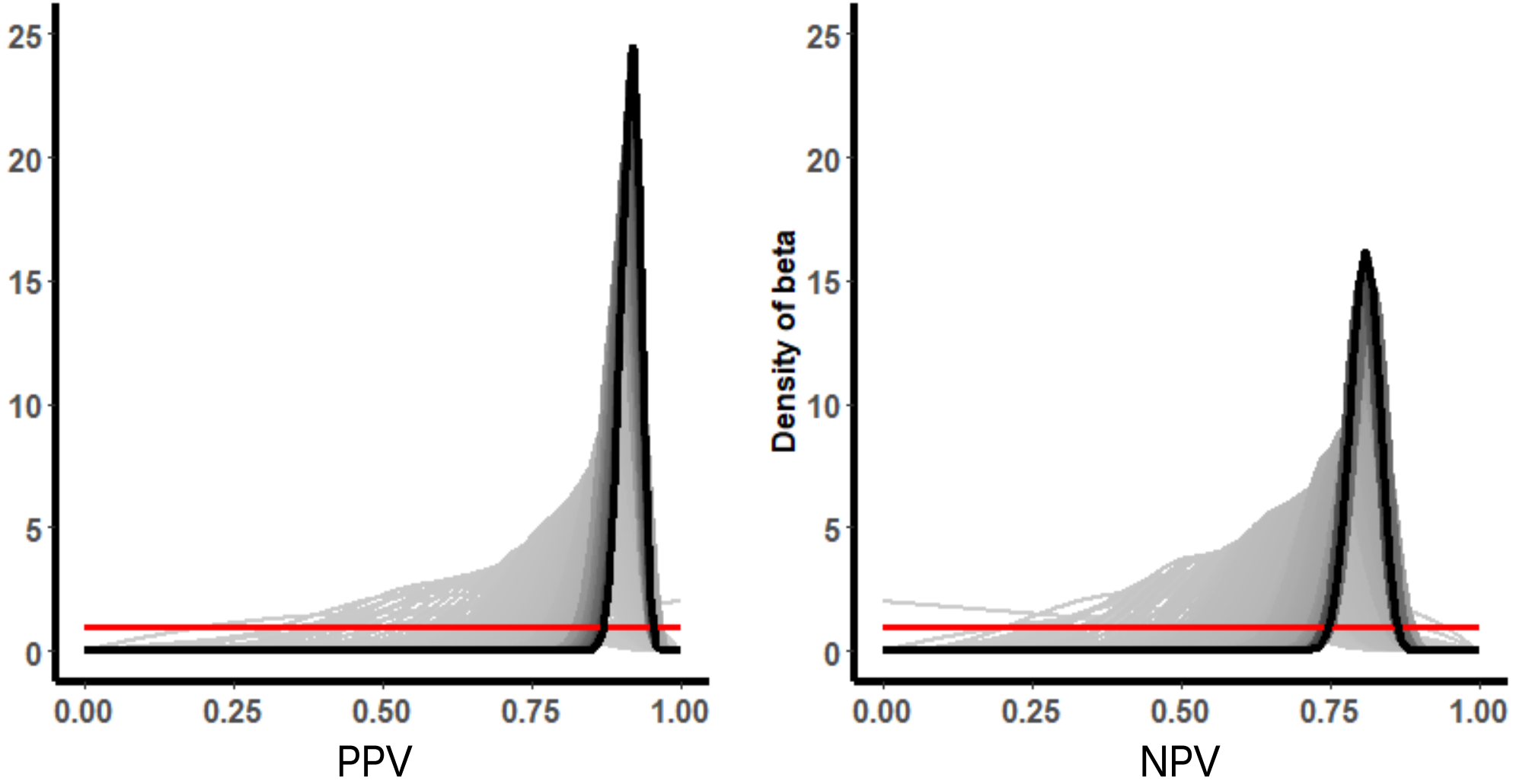
Youth Subcohort



Adult Subcohort



Adaptive Design: single person validation, youth subcohort



Adaptive Design: summary

Demonstrate how the method can be used to determine efficacy/futility of validation, optimizing study resources concurrent with cohort enrollment and follow-up.

Considerations:

Efficacy/futility should be consistent with goals for validation.

Prospective monitoring of validation data allows for detection of time trend in classification parameters, which is not possible in conventional validation study designs.

Detection of a time trend may change the approach to validation or use of classification parameters in quantitative bias analysis.

Adaptive Validation Design

Stopping based on precision of bias-adjusted estimate of association

Adaptive Design: example 3

Sample validation data are collected until the bias-adjusted estimate of effect reaches a prespecified level of precision.

Example of this approach in the association between transmasculine/transfeminine status and self-inflicted injury, adjusting for the possible misclassification of transmasculine/transfeminine status.

Adaptive Design: exposure validation

We estimate PPV and NPV within strata of self-inflicted injury (yes/no) to be used in the bias analysis

Misclassified sex recorded at bith	Gold standard sex recorded at birth: TM	Gold standard sex recorded at birth: TF	
TM	a	b	PPV=a/(a+b)
TF	c	d	NPV=d/(c+d)

TM=transmasculine; TF=transfeminine

Adaptive Design: stopping criterion

Precision of the bias-adjusted OR no more than 80% wider than the precision of the conventional estimate.

Self-inflicted Injury	Transmasculine	Transfeminine
Yes	113	54
No	597	567

OR=1.99 (95%CI: 1.41, 2.80)

Conventional: precision = 2.80/1.41 = 1.99

Bias-adjusted: precision = 1.99*1.80 = 3.58

Adaptive Design: stopping criterion

Precision of the bias-adjusted OR requires computation of the variance of the bias-adjusted OR.

J Clin Epidemiol Vol. 43, No. 9, pp. 941–947, 1990
Printed in Great Britain.

0895-4356/90 $3.00 + 0.00

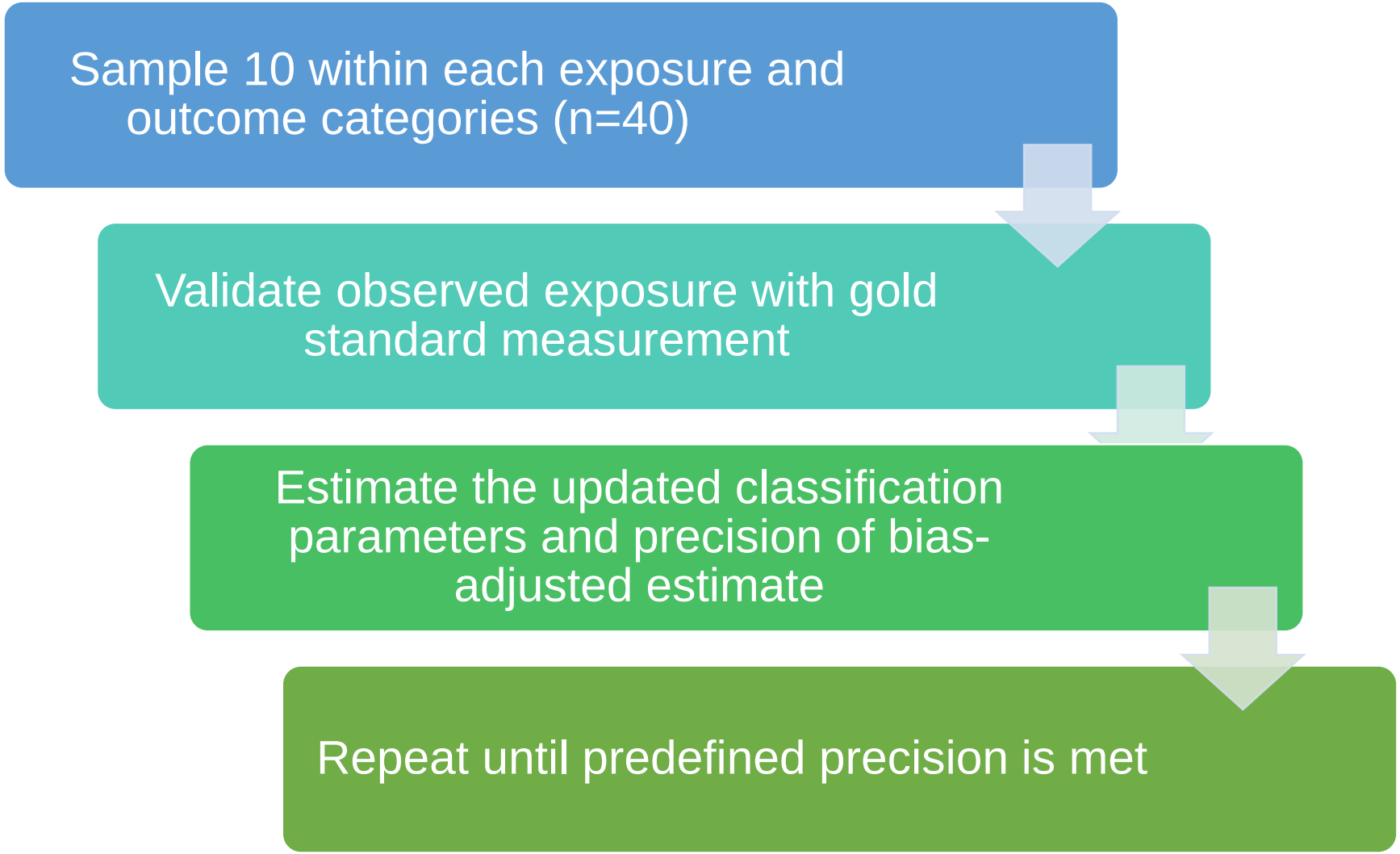
VALIDATION STUDY METHODS FOR ESTIMATING EXPOSURE PROPORTIONS AND ODDS RATIOS WITH MISCLASSIFIED DATA

ROGER J. MARSHALL

Department of Community Health, University of Auckland, Auckland, New Zealand

(*Received in revised form* 8 *February* 1990)

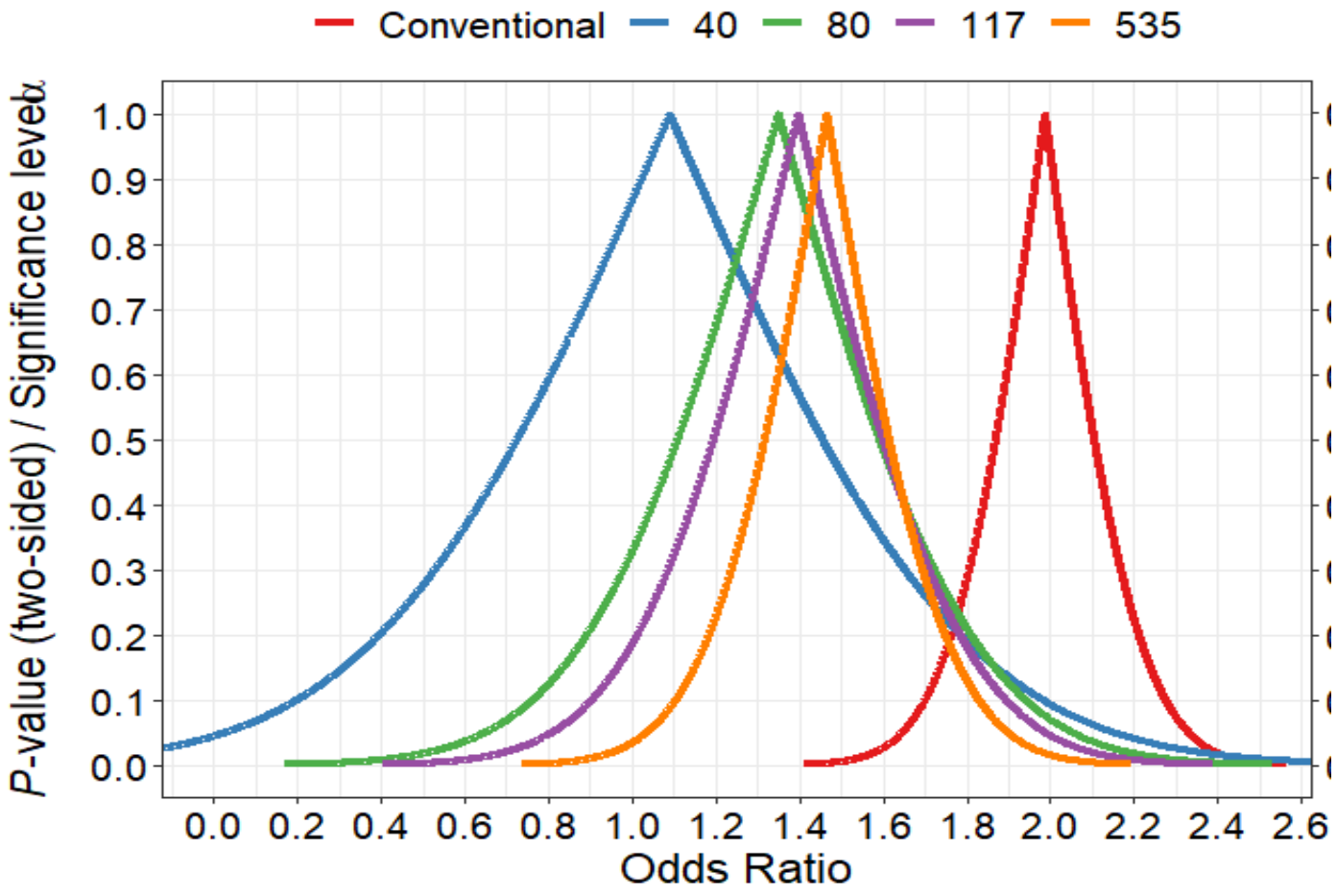
Adaptive Design: sampling strategy



Adaptive Design: results

Conventional and bias-adjusted effect estimates between transmasculine/transfeminine and self-inflicted injury in the STRONG cohort.							
	Self-inflicted injury		No self-inflicted injury				
Method	PPV (95% CI)	NPV (95% CI)	PPV (95% CI)	NPV (95% CI)	Number Validated	OR (95% CI)	Precision (UCL95/LCL95)
Conventional	100%	100%	100%	100%	0	1.99 (1.41, 2.80)	1.99

Adaptive Design: results



Adaptive Design: summary

The bias-adjusted OR computed with the adaptive validation approach was comparable to the bias-adjusted OR obtained using the complete validation data.

Considerations:

Define acceptable level of precision that will yield substantively meaningful results.

Probabilistic bias analysis is often preferable to account for uncertainty in classification parameters, confounders, and random error.

Adaptive Validation Design

Conclusions and future directions

Adaptive Design: conclusions

Example 1: Illustrate how the adaptive validation design can be used simultaneous with cohort enrollment and follow-up to collect validation data until a desired threshold and precision is met.

Adaptive Design: conclusions

Example 1: Illustrate how the adaptive validation design can be used simultaneous with cohort enrollment and follow-up to collect validation data until a desired threshold and precision is met.

Example 2: Illustrate how the method can be used to determine efficacy/futility of validation, optimizing study resources concurrent with cohort enrollment and follow-up.

Adaptive Design: conclusions

Example 1: Illustrate how the adaptive validation design can be used simultaneous with cohort enrollment and follow-up to collect validation data until a desired threshold and precision is met.

Example 2: Illustrate how the method can be used to determine efficacy/futility of validation, optimizing study resources concurrent with cohort enrollment and follow-up.

Example 3: Illustrate an approach to effective and efficient estimation of classification parameters as validation data accrue, with emphasis on the precision of the bias-adjusted estimate.

Adaptive Design: advantages

The approach offers a validation substudy design suitable for scenarios in which validation data are collected in real time and applicable to any parent epidemiologic study.

Allows researchers to carefully allocate fixed study resources when implementing validation studies, amenable to different objectives for validation.

Our method also outlines an approach to study design based on precision that can account for both random and systematic errors.

Adaptive Design: future directions

1) Time trend in classification parameters.

2) Assess probability that inference will change if stop validation efforts.

3) Methods to determine ‘next best’ subgroup for validation.

Continued work to guide effective and efficient validation substudy design is an important consideration in epidemiology

Acknowledgements

Funding:
US National Cancer Institute (F31CA239566)
US National Center for Advancing Translational Sciences (TL1TR002540)
US National Library of Medicine (R01LM013049)
US National Institute of General Medical Sciences (P20 GM103644)
US National Institute of Mental Health from the National Institutes of health (R01MH109507)
STRONG cohort data were collected with support from Contract AD-12-11-4532 from the Patient Centered Outcome Research Institute and by the NICHD R21HD076387

Collaborators:
Rich F MacLehose, Tom P Ahern, Michael Goodman, Darios Getahun, Michael Silverberg, Rune Erichsen, Ole Thorlacius-Ussing, Jaimie Gradus

Questions?

Contact Information

Tim Lash: timothy.lee.lash@emory.edu

Lindsay Collin: lindsay.collin@hci.utah.edu