Monthly Research Presentation

Partial Verification Bias Correction in Diagnostic Accuracy Studies

Date: 21 September 2023

Dr Wan Nor Arifin Wan Mansor

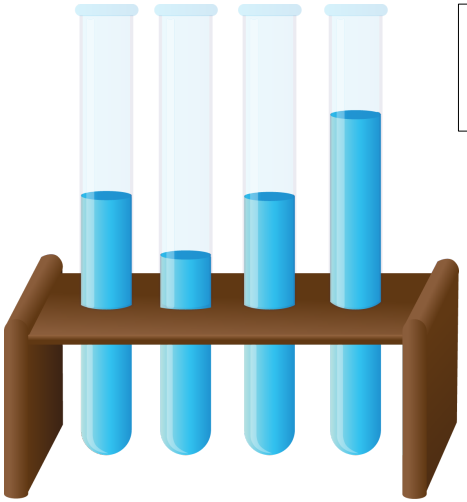
Biostatistics & Research Methodology Unit,
School of Medical Sciences, Universiti Sains Malaysia

Outlines

- Partial Verification Bias
- Bias Correction Methods
- PVBcorrect R Package
- Relevance to Medical Researchers
- Conclusion

Partial Verification Bias

Background

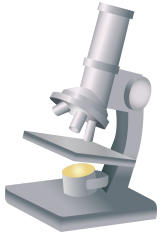


Extremely important role in medical care[1]

Objective assessment[2]

Disease vs No disease[3]

Diagnostic Tests



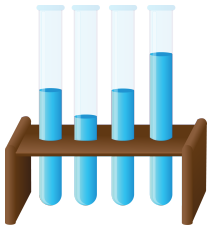
Evaluation[4]

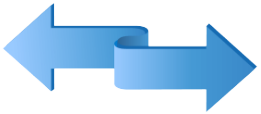
Diagnostic Accuracy Studies

Newly Developed Tests

[1]O'Sullivan et al. (2018). [2]Kosinski & Barnhart (2003). [3]Gotzche (2007). [4]Linnet et al. (2012).

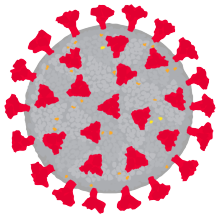
Diagnostic accuracy studies





Newly Developed Test | **Gold Standard Test**

Compare*

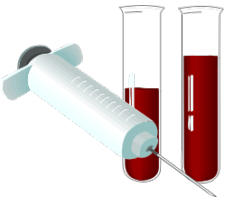


New Covid-19 Test Kit

Reverse Transcription Polymerase Chain Reaction (RT-PCR)

New Pregnancy Test Kit

Human Chorionic Gonadotropin (hCG)



Binary Accuracy Measures

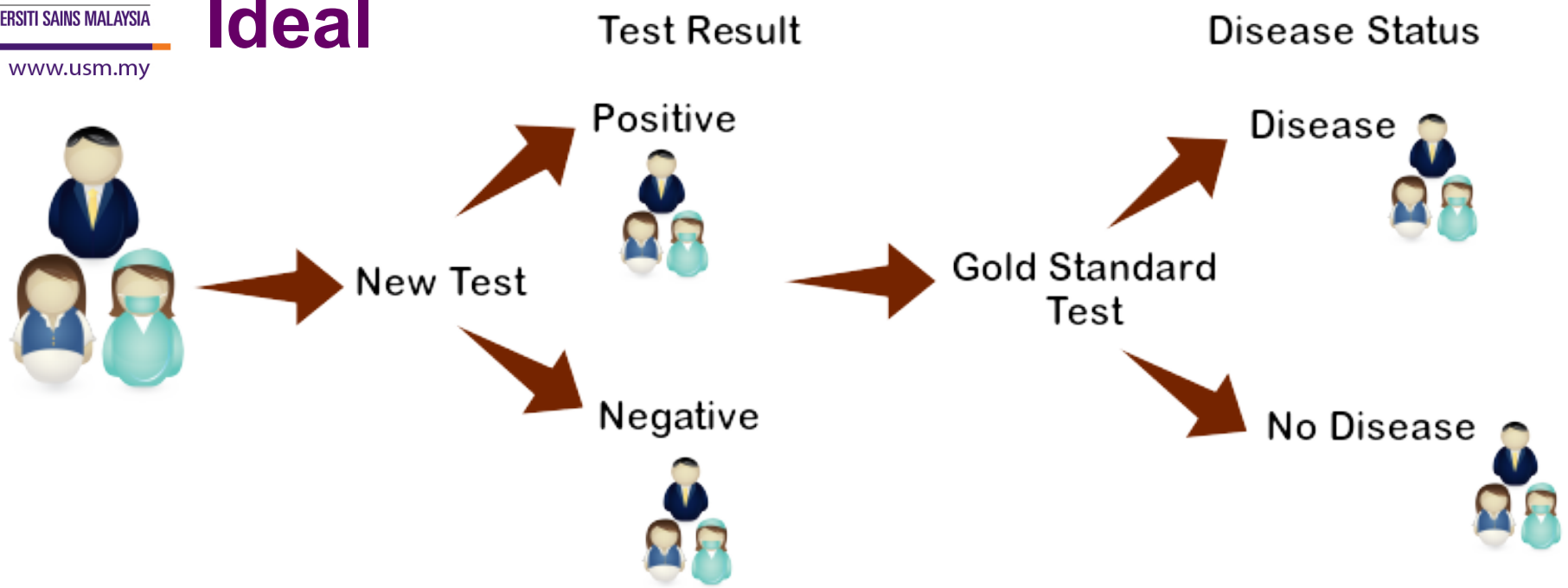
Sensitivity (True Positive Rate)

Specificity (True Negative Rate)

***Hall et al. (2019), O'Sullivan et al. (2018).**

Attributions: [1]https://commons.wikimedia.org/wiki/File:Coronavirus_SVG_Vector_Image.svg [2]https://commons.wikimedia.org/wiki/File:Pregnancy-test.svg [3]https://commons.wikimedia.org/wiki/File:201904_RNA.svg

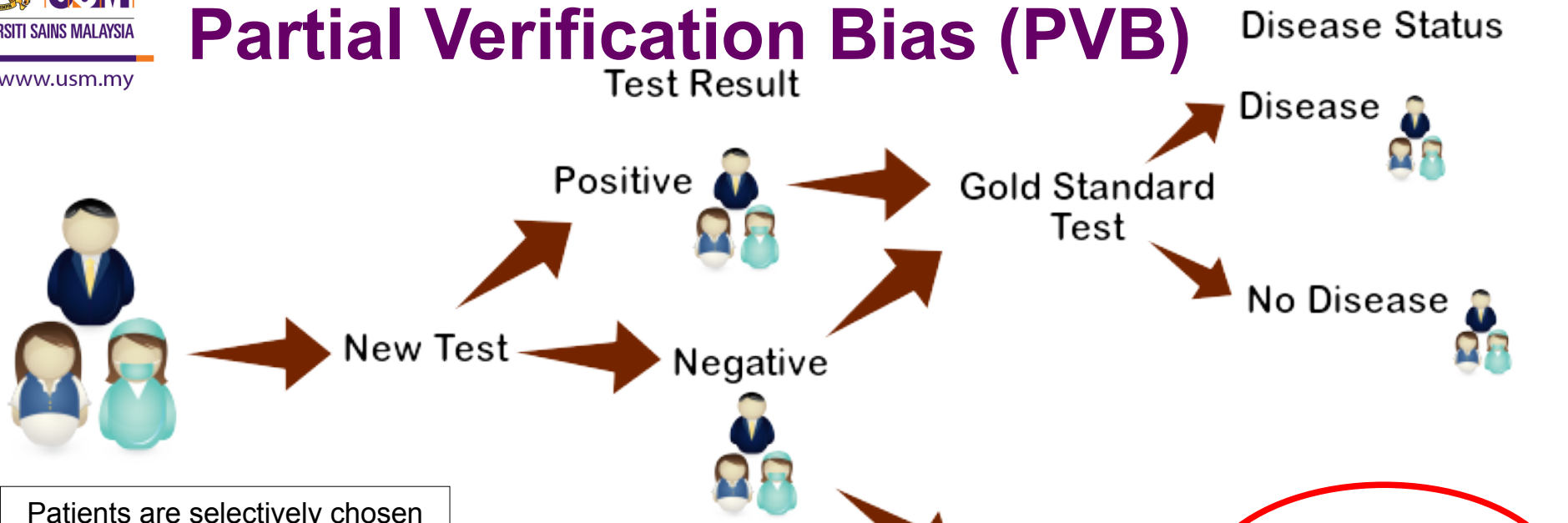
Ideal



	Disease Status	
Test Result	P	N
P	TP	FP
N	FN	TN
	Sensitivity= TP/(TP+FN)	Specificity= TN/(TN+FP)

Abbreviations: TP, true positive; TN, true negative; FP, false positive; FN, false negative.

Partial Verification Bias (PVB)



Patients are selectively chosen for verification by gold standard – test positive, other criteria*

Partially Verified Sample

	Disease Status		
Test Result	P	N	Unverified
P	TP	FP	?TP/FP
N	FN	TN	?FN/TN

Sensitivity? Specificity?

*O'Sullivan et al. (2018). Abbreviations: TP, true positive; TN, true negative; FP, false positive; FN, false negative.

Partial Verification Bias (PVB)

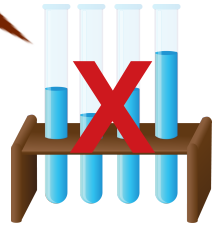
Study Design

Efficiency, cost, technical, ethical

Clinical Practice

Clinical likelihood

Invasive procedure



Missing Gold Standard*

Infeasibility

Technically impossible

*Naaktkeboren et al. (2016).

Impacts of Partial Verification Bias

Partial Verification Bias

Correction methods to account for PVB[8]

More accurate estimates of accuracy measures

Statistical Estimate

Biased estimates of accuracy measures[1]

Over/underestimation[2]

Often sensitivity and specificity[3]

Clinical Practice

Invalid diagnostic tests[4]

Premature implementation of new tests[5]

Unnecessary cost of testing[6]

Clinical errors[7]

[1]de Groot, Bossuyt et al. (2011), Naaktkeboren et al. (2016), Hall et al. (2019). [3]de Groot , Bossuytet al. (2011), Alonzo (2014). [3]Chikere et al. (2019). [5]Rutjes et al. (2007). [6]de Groot , Bossuyt et al. (2011). [7]Hall et al. (2019). [8]O'Sullivan (2018).

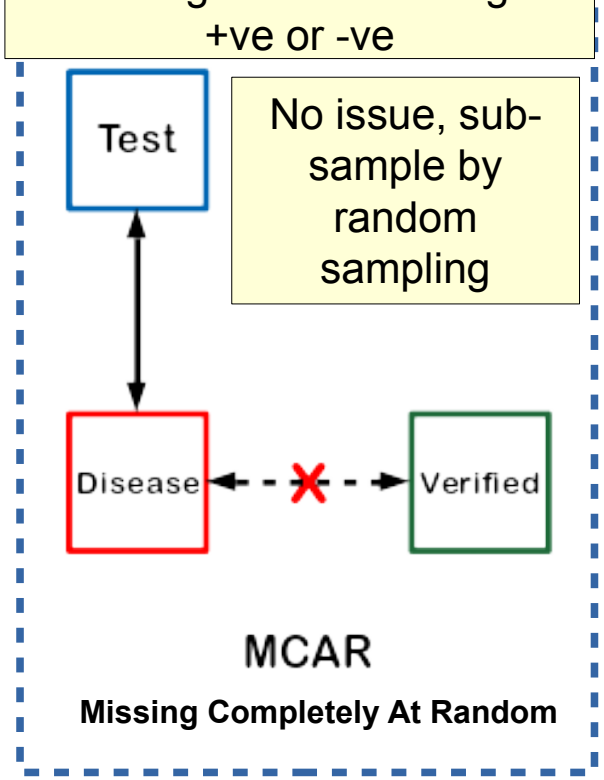
Bias Correction Methods

Missing Data Mechanisms in PVB

COVID-19 -- Proceed to RT-PCR regardless RTK-Ag is +ve or -ve

COVID-19 -- Proceed to RT-PCR when RTK-Ag is +ve

COVID-19 -- Proceed to RT-PCR when RTK-Ag is +ve AND patient looks sick



Cause of selection bias known & recorded

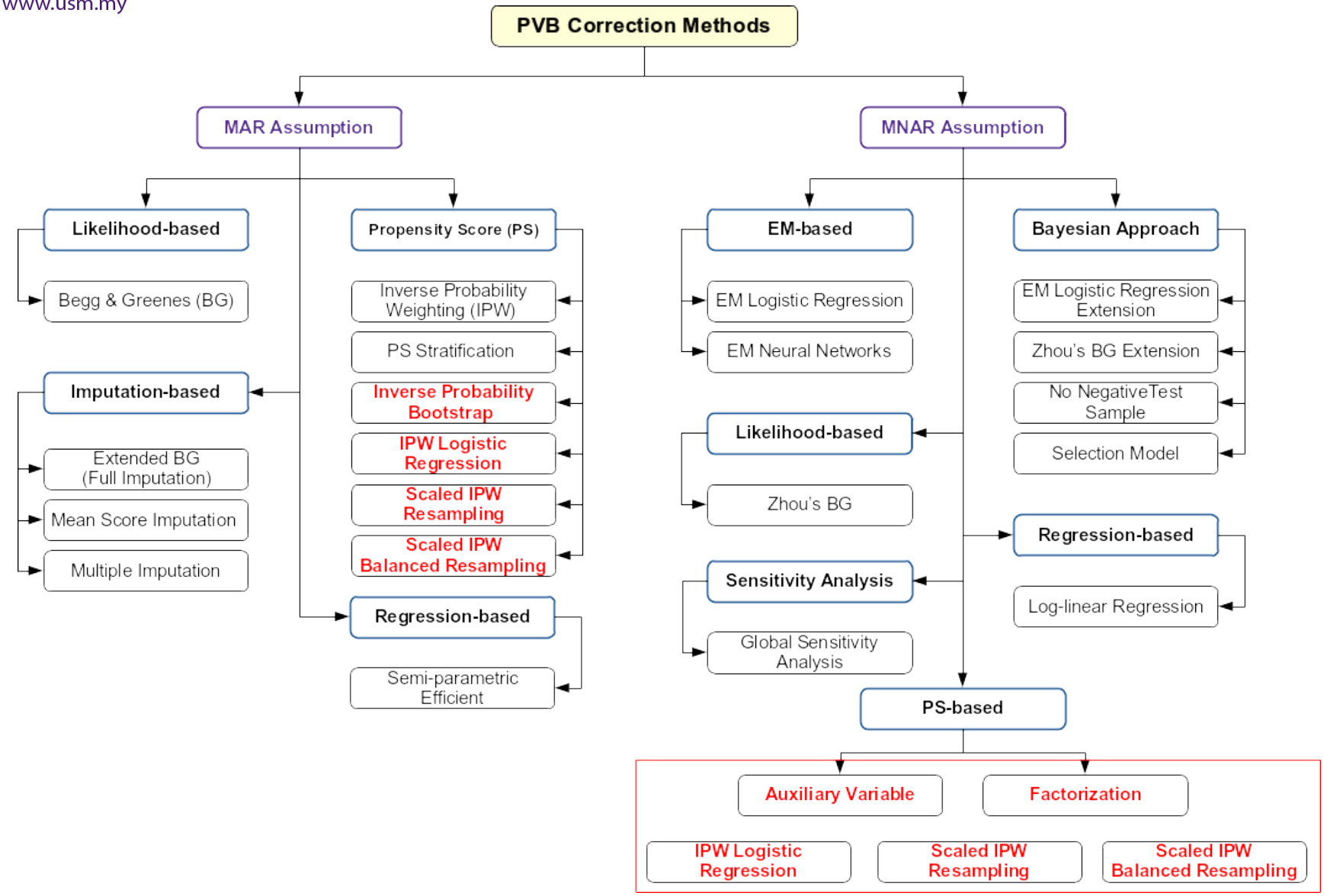
Cause of selection bias not known

No selection bias

With Partial Verification Bias

*Adapted from Figure 2 in Schafer & Graham (2002)

Existing PVB Correction Methods



EM, expectation maximization; **MAR**, missing at random; **MNAR**, missing not at random; **BG**, Begg & Greenes; **PVB**, partial verification bias

PVBcorrect R Package

About

- Available in GitHub

github.com/wnarifin/PVBcorrect

- Tutorial published in Statistics in Medicine

doi.org/10.1002/sim.9311

Features

- Version 0.1.1
- Available methods:
  - Begg and Greenes' methods
  - Multiple imputation method by logistic regression
  - EM-based logistic regression method
- Point estimates & CIs
- Other methods still in coding stage

Installation

In R command:

```
install.packages("devtools")
devtools::install_github("wnarifin/PVBcorrect")
```

Some example outputs

No correction

```
cca_out = acc_cca(data = cad_pvb, test = "T", disease = "D", ci = TRUE)
cca_out$acc_results
```

```
## Estimates of accuracy measures
## Uncorrected for PVB: Complete Case Analysis
##
##           Est         SE     LowCI     UppCI
## Sn  0.9750000 0.01103970 0.9533626 0.9966374
## Sp  0.1439114 0.02132173 0.1021216 0.1857013
## PPV 0.4566745 0.02410569 0.4094282 0.5039207
## NPV 0.8863636 0.04784519 0.7925888 0.9801385
```

Larger Sn
Lower Sp

Corrected by Extended B&G

```
ebg_out = acc_ebg(data = cad_pvb, test = "T", disease = "D", ci = TRUE,
ebg_out$acc_results
```

```
## Estimates of accuracy measures
## Corrected for PVB: Extended Begg and Greenes' Method
##
##           Est         SE     LowCI     UppCI
## Sn  0.8188629 0.06438696 0.6860317 0.9360441
## Sp  0.5918754 0.01759285 0.5587741 0.6285913
## PPV 0.4566745 0.02421467 0.4063396 0.5059415
## NPV 0.8863636 0.04925696 0.8005051 0.9949495
```

Corrected Sn & Sp

Corrected by Multiple Imputation

```
mi_out = acc_mi(data = cad_pvb, test = "T", disease = "D", ci = TRUE, seednum = 12345)
mi_out$acc_results
```

```
## Estimates of accuracy measures
## Corrected for PVB: Multiple Imputation Method
##
##           Est         SE     LowCI     UppCI
## Sn  0.8031807 0.06451495 0.6753082 0.9310532
## Sp  0.5874592 0.02199350 0.5440683 0.6308500
## PPV 0.4578988 0.02400632 0.4105655 0.5052321
## NPV 0.8699605 0.05498024 0.7609444 0.9789765
```

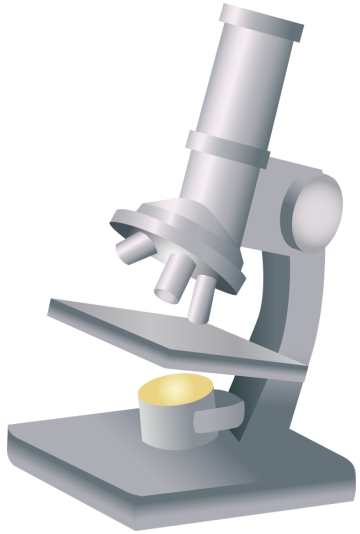
Corrected Sn & Sp

https://wnarifin.github.io/workshop/conf_r2022/PVBinR_Demo.html

Relevance to Medical Researchers

Scenario 1

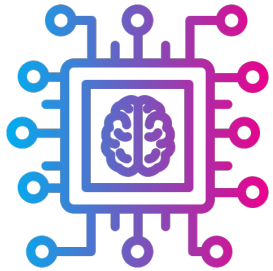
- Already-conducted diagnostic accuracy study – might be biased – PVB
- Gold standard test – only done to selected patients
- Identify reasons (variables) why patients are not selected
- Reanalyze / Reevaluate – Perform PVB correction for the diagnostic test performance results and compare w/out correction

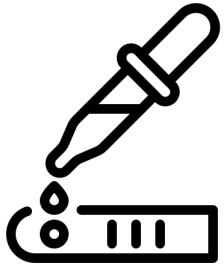


Scenario 2

- Limitation with gold standard test – invasive, costly, infeasible
- With planning – in view of selection bias – document reasons why patients are not selected
- Prioritize patients with higher clinical possibility of having disease (i.e. positive gold standard test)
- Ethical – avoid unnecessary procedure on healthy patients
- Save cost

Scenario 3





- Assume Machine Learning = Diagnostic Test
- Machine learning – the issue with disease labeling / confirmation
- Selective labeling – save clinicians' time
- May consider PVB correction as well – select only cases with high probability of disease for labeling

images: Flaticon.com

Conclusion

Take-home messages

- Although gold standard test is invasive/expensive/not feasible, with proper research planning in view of selection bias, diagnostic test performance can be accurately evaluated
- Bias correction is easy when potential sources of selection bias are identified and documented, even if we cannot avoid the bias
- PVBcorrect R Package helps researchers to perform the correction – GUI is planned

Publications

Arifin, W. N., & Yusof, U. K. (2022). Correcting for partial verification bias in diagnostic accuracy studies: A tutorial using R. *Statistics in Medicine, 41*(9), 1709–1727. (Published. Indexed by WoS, Impact Factor: 2.497, Q1 in Statistics & Probability)

Arifin, W. N., & Yusof, U. K. (2022). Correcting for partial verification bias in diagnostic accuracy measures using inverse probability bootstrap sampling. *Diagnostics, 12*(11), 2839. (Published. Indexed by WoS, Impact Factor: 3.992, Q2 Medicine, General & Internal)

Acknowledgment

- Universiti Sains Malaysia for the paid study leave for the PhD
- Kementerian Pendidikan Tinggi for the SLAI scholarship

Thank You