Verification Bias in Studies on the Accuracy of the Clinical Diagnosis of Alzheimer's Disease

X.H. Andrew Zhou, Ph.D

Professor, Department of Biostatistics, University of Washington
Co-investigator and Senior Biostatistician, NACC, University of Washington

Gold standard on the diagnosis of AD

- Histopathology is the most commonly accepted gold standard on AD
- Three histopathological diagnoses

  1)The Khachaturian criteria

  2)The Consortium to Establish a Registry for Alzheimer Disease (CERAD) criteria.

  3)The National Institute on Aging (NIA)-Reagan Institute Consensus Conference.
- If a subject does not have postmortem data, the subject will not have a gold standard diagnosis on AD.

Analytic Goal

- Assessment of the accuracy of clinical diagnoses when some subjects do not have a gold standard.

NACC Minimum Data Set

- The National Alzheimer Coordinating Center (NACC) maintains a cumulative database on subjects from approximately 30 NIA-funded AD centers.
- Clinical data are available for nearly all subjects, but postmortem data are only available for those who died and underwent autopsy.

Our data set

- We selected a subset of all subjects with a clinical diagnosis of dementia from NACC Minimum Data Set
- We dichotomized the clinical diagnoses into AD and non-AD dementia.
- Those subjects with autopsy were also classified by neuropathology (NP) under the CERAD criteria, used as the gold standard, into AD and non-AD dementia.

Data

		Clinical Diagnosis		
	True Dx verified based on CERAD	AD	Non AD	Total
	AD	***1992***	***200***	***2192***
	No AD	***261***	***216***	***477***
	Unknown	***27521***	***6136***	***33657***
	Total	***29774***	***6552***	***36326***

Accuracy of Diagnostic Tests

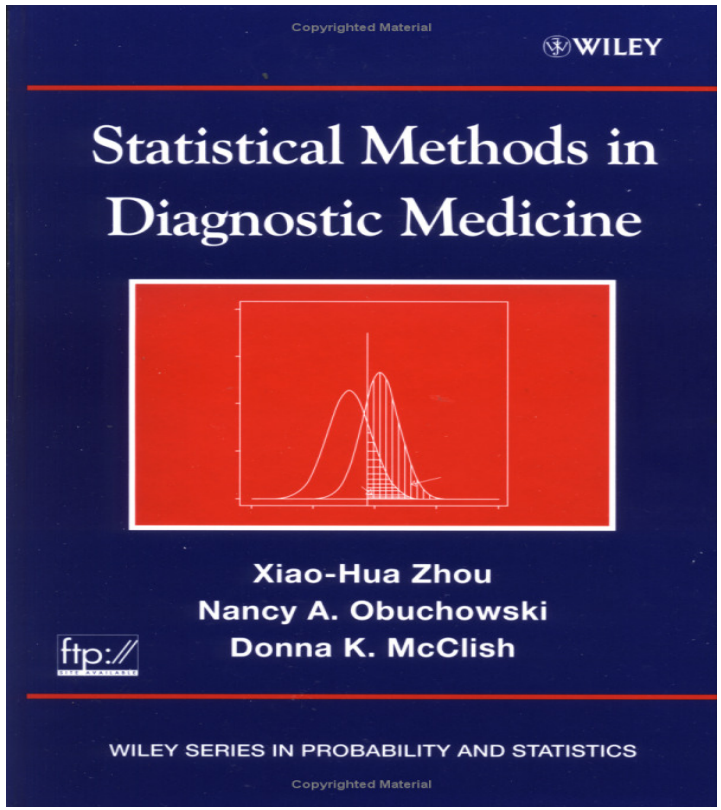
- Binary-scale: sensitivity, specificity, positive and negative predictive values.

- Sensitivity is the probability that the clinical diagnosis would be AD given the true status is AD

– Specificity is the probability that the clinical diagnosis is non-AD given the true status is Non-AD

- Ordinal-scale or continuous-scale: ROC curves.

Textbook on Statistical Methods in Diagnostic Medicine



Naïve approach

- The naïve approach for dealing with missing gold standards is to restrict the analysis to subjects with gold standard information, called “complete-data analysis”.
- Using only the autopsy-verified sample (naïve approach),

	Estimate	(95% CI)
Sensitivity	.907	(.895, .919)
Specificity	.449	(.402, .496)

The fallacy of “complete-data” analysis

- The “complete-data” analysis assumes one can extrapolate necropsy material to the whole study population.
- However, the clinical diagnoses in those having necropsy can differ substantially from those who did not.
- In addition, subjects who died have a different disease trajectory than subjects who survived, despite clinical diagnoses of Alzheimer's disease.

Effects of Verification Bias

- *To learn how verification bias works, let us look at a hypothetical example with 200 patients with dementia.*
- *All patients have a clinical diagnosis, either AD or not AD, but some of them do not have a gold standard.*
- *We want to estimate sensitivity and specificity of the clinical assessment test on AD.*

Target population

- We assume that the true sensitivity and specificity of the clinical assessment are 80% and 90%, respectively, to be estimated.
- If all study subjects had died and had had autopsy results during the study, we would obtain the following table:
- ---

	Clinical Dx		
	AD	Not AD	Total
True Dx			
AD	80	20	100
Not AD	10	90	100
Total	90	110	200

Partial verification

- *The chance that a patient with the clinical diagnosis of AD would die and undergo autopsy during the study is 50%.*
- *The chance that a patient with a clinical diagnosis of non-AD would die and undergo autopsy during the study is 20%.*

Partial verification

- The following table contains the data with partial verification:

	Clinical Dx		
	AD	Not AD	Total
True Dx			
AD	**40**	**4**	**44**
Not AD	**5**	**18**	**23**
Unknown	**45**	**88**	**133**
Total	**90**	**110**	**200**

- Using only verified cases,
  Sens=40/(40+4)=0.91 and spec=18/(18+5)=0.78.
- True sensitivity was overestimated.
- Specificity was underestimated.

A special missing-data problem

- Verification Bias is a special type of missing data
- Without having a gold standard = missing value of the gold standard.

Missing-data mechanism

- Missing completely at random (MCAR): the reason for having missing data is not related to all variables in the study.
- This assumption would be violated if subjects who did not have autopsy were younger, on average, than subjects who did.

Missing-data mechanism, cont

- Missing at random (MAR): the reason for having missing data is only related to variables in the study that are completely observed. (This assumption would be violated if subjects with missing data on a particular variable tend to have lower (or higher) values on that variable than those with data present, after adjusting for other known characteristics).
- Non-MAR.

Taxonomy of missing-data methods

- “Complete-data” analysis: Procedures based on completely recorded units
- Single Imputation: substitute some reasonable values for each missing value and then proceed to do the analysis as if there were no missing data.

Single Imputation Methods

- Marginal Mean imputation: for each missing value on a given variable, substitute the mean for those cases with data present on that variable
- Conditional Mean imputation (regression imputation):

  replace missing values by predicted values from a regression of the missing item on items observed for the unit, usually calculated from units with both observed and missing variables present.
- Stochastic regression imputation: replace missing values by a value predicted by regression imputation plus a residual, drawn to reflect uncertainty in the predicted value.

Problem of Single Imputation

- The single imputation methods suffers from a fundamental problem: analyzing imputed data as through it were complete data produces standard errors that are underestimated.

Multiple Imputation (MI) Method

- MI is a simulation technique to deal with missing data.
- The MI method imputes **several** probable results of neuropathologic examination to each un-verified subject, based in part on the distribution of observed results among verified subjects with similar characteristics.
- Then the method uses both complete and imputed data on all subjects with clinical assessments to obtain unbiased estimates of the sensitivity and specificity.

MI, cont

- MI is performed following three steps:
  - Imputation - replace each missing value by m>1 plausible values. This step yields m complete data sets that differ in the imputed values.
  - Analysis - Analyze each complete data set by a complete-data method. This step results in m sets of estimates and their standard errors.
  - Combination - Combine the results by simple arithmetic rules. This step provides final estimates and standard errors taking into account the uncertainty involved in replacing missing with imputed data.

MI results

- The MI method using all subjects yields the following estimates:
- An estimated sensitivity of 0.884 with 95% confidence interval of (0.869,0.899) and
- An estimated specificity of 0.516 with 95% confidence interval of (0.459,0.571).

Conclusions

- The autopsy-only method consistently overestimated sensitivity and underestimated specificity of the clinical diagnosis of AD as compared to the MI method.
- The MI estimates show greater precision as a result of using all subjects

Current research

- In our research, we treated living AD and dead non-autopsy AD patients as the same.
- However, it is likely that the accuracy of clinical assessments is substantially greater among those who have died (because of longer f/u and potentially more clinical information on which to make the dx).
- We are working on new methods that can distinguish these two groups.

Acknowledgement

- NACC and University of Washington:
  Drs. Tom Koepsell and Bud Kukull
- University of Connecticut:
  Dr. Ofer Harel