Multiple imputation for the comparison of two screening tests in two-phase Alzheimer studies

Ofer Harel[1] and X.H. Zhou[2]

[1]Department of Statistics, University of Connecticut
[2]Department of Biostatistics, University of Washington

Prepared for The International Conference on Helath Policy Statistics January, 2008.

Outline

- ✓ Background
- ✓ Data set-up
- ✓ The use of MI
- ✓ Simulations
- ✓ Data Example
- ✓ Discussion

1. Background

- Two-phase designs are common in epidemiological studies of dementia
- In the First phase - all subjects are screened using common screening test(s)
- In the Second phase - only a subset of these subjects is tested using a more definitive verification assessment (gold standard)
- We are interested in comparing the accuracy of two screening tests in a two-phase study of dementia

Background

- √ Inferences are commonly made using only the verified sample
- √ It is well documented that in that case, there is a risk for bias, called verification bias.
- √ For a two-level tests we estimate the differences of sensitivities, specificities and their confidence intervals (CI)
- √ This is equivalent to estimating CI for the difference of binomial proportions of paired data
- √ This is not a trivial problem even in complete-data scenarios
- √ Wald-type test showed to have many disadvantages

Background

- √ Alzheimer studies are often done using two-phase designs
- √ In Alzheimer research the common test is a clinical test
- √ While the verification is done using autopsy

2. Data set-up

Let T_1 and T_2 be the two test indicators, D be the true status, which is known only for the subjects that have been verified V. All these variables are indicators. The data can be summarized into:

Table 1: Aggregated data

		$T_1 = 1$		$T_1 = 0$	
		$T_2 = 1$	$T_2 = 0$	$T_2 = 1$	$T_2 = 0$
$V = 1$	$D = 1$	x^A_{111}	x^A_{101}	x^A_{011}	x^A_{001}
	$D = 0$	x^A_{110}	x^A_{100}	x^A_{010}	x^A_{000}
$V = 0$		x^B_{11+}	x^B_{10+}	x^B_{01+}	x^B_{00+}
Total		n_{11+}	n_{10+}	n_{01+}	n_{00+}

Data set-up

Consider we have a complete data set. In this case the data can be summarized into:

Table 2: Complete data

	$T_1 = 1$		$T_1 = 0$	
	$T_2 = 1$	$T_2 = 0$	$T_2 = 1$	$T_2 = 0$
$D = 1$	x_{111}	x_{101}	x_{011}	x_{001}
$D = 0$	x_{110}	x_{100}	x_{010}	x_{000}
Total	n_{11+}	n_{10+}	n_{01+}	n_{00+}

where the sensitivity is the test ability to detect the condition when it is present. While, the specificity is the test ability to exclude the condition to those with out the condition.

Notations

- ✓ Let (X_{0k}, X_{1k}), $k = 1, 2, \ldots, n$, be an iid sample of pairs
- ✓ (X_0, X_1) are correlated Bernoulli RV with proportions p_0 and p_1
- ✓ We are interested in the difference $p = p_1 - p_0$

Existing methods for complete data

McNemar's interval:

$$\hat{p} \pm z_{1-\alpha/2} n^{-1/2} \sqrt{\hat{p}_1(1-\hat{p}_1) + \hat{p}_0(1-\hat{p}_0) + 2(\hat{p}_0\hat{p}_1 - \hat{p}_{11})},$$

where

- $\hat{p}_i = \frac{\Sigma_{k=1}^n X_{ik}}{n}$
- $\hat{p} = \hat{p}_1 - \hat{p}_0$
- $\hat{p}_{11} = \frac{\Sigma_{k=1}^n X_{0k} X_{1k}}{n}$,
- z_α is the α-th quantile of the standard normal distribution.

McNemar's interval with continuity correction:

$$\hat{p} \pm (z_{1-\alpha/2} SE + 1/n)$$

Existing methods for complete data

Newcombe Hybrid (NH) Interval: Newcombe (1998) reviewed and compared several existing intervals for the difference between two binomial proportions based on paired data.

He recommended a score interval with continuity correction called Newcombe Hybrid (NH).

$$[\hat{p} - (\delta_1^2 - 2\hat{\phi}\delta_1\epsilon_2 + \epsilon_2^2)^{1/2}, \hat{p} - (\epsilon_1^2 - 2\hat{\phi}\epsilon_1\delta_2 + \delta_2^2)^{1/2}]$$

Function of some quadratic equations of x

Existing methods for complete data

May and Johnson (MJ) interval: This interval was studied by May and Johnson (1997)
The CI is

$$\left[Max\{0, \frac{-B-(B^2-4AC)^{1/2}}{2A}\}, \;\; Min\{1, \frac{-B+(B^2-4AC)^{1/2}}{2A}\}\right]$$

where $A = (1+\frac{z^2_{\alpha/2}}{n})$, $B = -2\frac{\Sigma_k(1-X_{0k})X_{1k}-\Sigma_k X_{0k}(1-X_{1k})}{n}$, and $C = (\frac{\Sigma_k(1-X_{0k})X_{1k}}{n} - \frac{\Sigma_k X_{0k}(1-X_{1k})}{n})^2 - z^2_{\alpha/2}\frac{\Sigma_k(1-X_{0k})X_{1k}+\Sigma_k X_{0k}(1-X_{1k})}{n^2}$.

Existing methods for complete data

Zhou and Qin (ZQ) interval: Zhou and Qin (2005) used the Edgeworth expansion to correct for the skewness of the Wald statistics

$$\left[Max(-1, \hat{p} - \frac{\hat{\sigma}}{\sqrt{n}} g^{-1}(z_1 - \alpha/2)), Min(1, \hat{p} - \frac{\hat{\sigma}}{\sqrt{n}} g^{-1}(z_1 - \alpha/2))\right],$$

where

$$g^{-1}(y) = \begin{cases} \frac{\sqrt{n}}{\hat{b}\hat{\sigma}} \left[(1 + 3(\hat{b}\hat{\sigma})(\frac{y}{\sqrt{n}} - \frac{\hat{a}\hat{\sigma}}{n}))^{1/3} - 1\right] & \hat{b}\hat{\sigma} \neq 0 \\ y - \frac{\hat{a}\hat{\sigma}}{n} & \hat{b}\hat{\sigma} = 0 \end{cases}$$

Existing methods for incomplete data

Maximum Likelihood interval: The only method available to date to deal with verification bias in paired comparisons of sensitivity and specificity was introduced by Zhou (1998).

- ✓ The estimators for the sensitivities of two tests are as follows:

$$\hat{\pi}_{1ML} = \frac{\sum_{j=0}^{1}(x_{1j1}^{A} n_{1j+})/(x_{1j1}^{A} + x_{1j0}^{A})}{\sum_{i=0}^{1}\sum_{j=0}^{1}(x_{ij1}^{A} n_{ij+})(x_{ij1}^{A} + x_{ij0}^{A})},$$

and

$$\hat{\pi}_{2ML} = \frac{\sum_{i=0}^{1}(x_{i11}^{A} n_{i1+})/(x_{i11}^{A} + x_{i10}^{A})}{\sum_{i=0}^{1}\sum_{j=0}^{1}(x_{ij1}^{A} n_{ij+})(x_{ij1}^{A} + x_{ij0}^{A})}.$$

- ✓ Similarly there are estimates for the specificity

Existing methods for incomplete data

Maximum Likelihood interval: Zhou (1998) also provided estimates for the covariance matrices

- ✓ The CI for the sensitivities will be

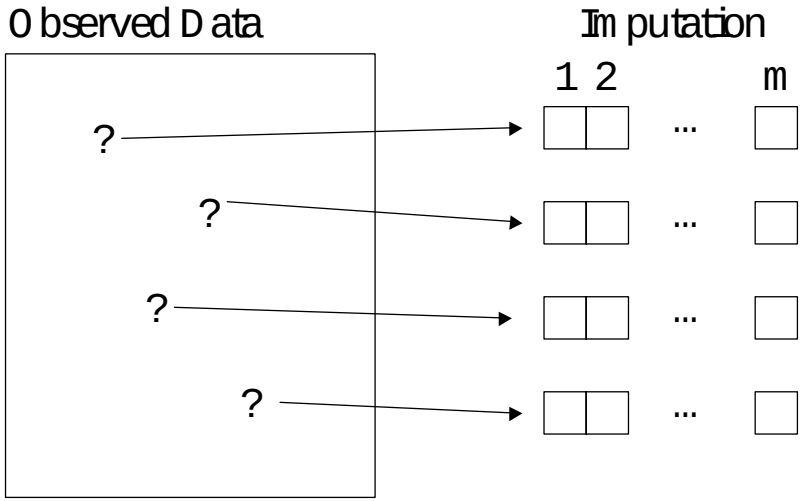
$$\hat{\pi}_2 - \hat{\pi}_1 \pm z_{1-\alpha/2} \sqrt{\{\hat{Var}(\hat{\pi}_1) + \hat{Var}(\hat{\pi}_2) - 2\hat{Cov}(\hat{\pi}_1, \hat{\pi}_2)\}}$$

and for the specificities

$$\hat{\tau}_2 - \hat{\tau}_1 \pm z_{1-\alpha/2} \sqrt{\{\hat{Var}(\hat{\tau}_1) + \hat{Var}(\hat{\tau}_2) - 2\hat{Cov}(\hat{\tau}_1, \hat{\tau}_2)\}}$$

3. The use of MI

MI is separated into 3 stages: Imputation stage, Analysis stage, and Combining Results stage.



The use of MI - Cont.

Imputation: $P(Y_{mis}|Y_{obs})$ in our case is

$$(x^B_{ij0}, x^B_{ij1})|Y_{obs}, \theta \sim M(x^B_{ij+}, (\theta_{ij1}/\theta_{ij+}, \theta_{ij0}/\theta_{ij+})).$$

Hence, We can use the following distributions in order to draw the missing values

$$\begin{aligned} x|\theta &\sim M(n, \theta) && \text{likelihood} \\ \theta &\sim D(\alpha) && \text{prior} \\ \theta|Y &\sim D(\alpha + x) && \text{posterior} \end{aligned}$$

Analysis: Given the complete data sets we can use any of the complete-data procedures mentioned before.

The use of MI - Cont.

Combining rules: Calculate and store the estimates $\hat{Q}^{(j)}$ and SE $U^{(j)}$ for $j = 1, \ldots, m$ and combine:

$$\bar{Q} = m^{-1} \sum_{j=1}^{m} \hat{Q}^{(j)}$$

$$\bar{U} = m^{-1} \sum_{j=1}^{m} U^{(j)}$$

$$B = (m-1)^{-1} \sum_{j=1}^{m} \left(\hat{Q}^{(j)} - \bar{Q} \right)^2$$

$$T = \bar{U} + (1 + m^{-1}) B$$

The use of MI - Cont.

Which leads to an approximate 95% interval for Q

$$\bar{Q} \pm t_\nu \sqrt{T},$$

where the degrees of freedom are

$$\nu = (m-1)\left[\frac{(1+m^{-1})B}{T}\right]^{-2}.$$

4. Simulations

we simulate data as follows:

- ✓ sample size $(588, 1000)$
- ✓ sensitivities $Se_k = P(T_k = 1|D = 1)$
- ✓ specificity $Sp_l = P(T_l = 0|D = 0)$
- ✓ verification probabilities $\lambda_{ij} = P(V = 1|T_1 = i, T_2 = j)$
- ✓ $\lambda_{11} = \lambda_{10} = 0.7$, $\lambda_{01} = 0.25$ and $\lambda_{00} = 0.14$
- ✓ prevalence is 0.35
- ✓ we run simulations 10000 times
- ✓ for methods required complete data we used $m = 10$ imputations.

Results

(a) $N = 588$ $Se_1 = Se_2 = 0.9$, $Sp_1 = Sp_2 = 0.95$, and 95% coverage

	Sensitivity					
	Est	MSE	lower	upper	length	Coverage
MI	-0.0012	0.0016	-0.0792	0.0767	0.1559	93.0
MI+Correction	-0.0012	0.0016	-0.0840	0.0816	0.1656	95.0
NH	-0.0012	0.0016	-0.3342	0.3293	0.6635	100
MJ	-0.0012	0.0016	0.0252	0.0841	0.0589	0
ZQ	-0.0012	0.0016	-0.0561	0.0536	0.1096	81.3
ML	-0.0013	0.0018	-0.0818	0.0769	0.1587	49.7
	Specificity					
	Est	MSE	lower	upper	length	Coverage
MI	0.0004	0.0005	-0.0421	0.0429	0.0850	93.6
MI+Correction	0.0004	0.0005	-0.0447	0.0456	0.0903	95.0
NH	0.0004	0.0005	-0.1584	0.1605	0.3189	100
MJ	0.0004	0.0005	0.0130	0.0456	0.0325	0
ZQ	0.0004	0.0005	-0.0299	0.0307	0.0606	85.5
ML	0.0000	0.0005	-0.0449	0.0408	0.0857	48.3

Results

(a) $N = 1000$ $Se_1 = Se_2 = 0.9$, $Sp_1 = Sp_2 = 0.95$, and 95% coverage

	Sensitivity					
	Est	MSE	lower	upper	length	Coverage
MI	-0.0016	0.0010	-0.0626	0.0594	0.1220	94.1
MI+Correction	-0.0016	0.0010	-0.0654	0.0623	0.1277	95.1
NH	-0.0016	0.0010	-0.3194	0.3145	0.6339	100
MJ	-0.0015	0.0010	0.0243	0.0654	0.0412	0
ZQ	-0.0016	0.0010	-0.0444	0.0412	0.0856	83.1
ML	-0.0014	0.0010	-0.0605	0.0571	0.1176	63.7
	Specificity					
	Est	MSE	lower	upper	length	Coverage
MI	0.0006	0.0003	-0.0324	0.0337	0.0661	94.9
MI+Correction	0.0006	0.0003	-0.0340	0.0353	0.0692	95.8
NH	0.0006	0.0003	-0.1512	0.1535	0.3047	100
MJ	0.0006	0.0003	0.0116	0.0333	0.0217	0
ZQ	0.0006	0.0003	-0.0227	0.0240	0.0468	84.2
ML	0.0006	0.0003	-0.0323	0.0322	0.0645	64.3

5. Data Example

- An epidemiological study of dementia which investigated the role of environmental risk factors in the development of Alzheimer's disease
- The goal is to compare the existing (standard) screening test to a new one.
- The new test is based on information from a cognitive test given to a person and from a relative test given to someone who knows the subject.
- The standard test uses only the information from the subject's test
- Postmortem data available only for those who died and underwent autopsy

Data Example

Table 3: Data from an Alzheimer study $(Age > 75)$

T_1	positive		negative	
T_2	positive	negative	positive	negative
Verified				
Disease	31	5	3	1
Non-disease	25	10	19	55
Not verified	22	6	65	346
Total	78	21	87	402

Imputation

Imputation: $P(Y_{mis}|Y_{obs})$ in our case is

$$
\begin{aligned}
(x^B_{110}, x^B_{111})|Y_{obs}, \theta &\sim M(x^B_{11+}, (\theta_{111}/\theta_{11+}, \theta_{110}/\theta_{11+})) \\
(x^B_{100}, x^B_{101})|Y_{obs}, \theta &\sim M(x^B_{10+}, (\theta_{101}/\theta_{10+}, \theta_{100}/\theta_{10+})) \\
(x^B_{010}, x^B_{011})|Y_{obs}, \theta &\sim M(x^B_{01+}, (\theta_{011}/\theta_{01+}, \theta_{010}/\theta_{01+})) \\
(x^B_{000}, x^B_{001})|Y_{obs}, \theta &\sim M(x^B_{00+}, (\theta_{001}/\theta_{00+}, \theta_{000}/\theta_{00+})).
\end{aligned}
$$

Hence, We can use the following distributions in order to draw the missing values

$$
\begin{aligned}
x|\theta &\sim M(n, \theta) && \text{likelihood} \\
\theta &\sim D(\alpha) && \text{prior} \\
\theta|Y &\sim D(\alpha + x) && \text{posterior}
\end{aligned}
$$

Imputation

Analysis: Given the complete data sets we can use any of the complete-data procedures mentioned before.

Combinations: After having m sets of estimates and their standards errors (for each method), we would need to use Rubin's ruled to combine these estimates.

Results

		Multiple Imputation					
		NHCI	MJCI	ZQCI	MI	MI+Correction	ML
Sensitivity	est	0.1082	0.1054	0.1082	0.1082	0.1082	0.0703
	SE				0.0928	0.0928	0.0929
	up	0.9542	0.1870	0.2195	0.2982	0.3121	0.2524
	low	-0.6600	0.0433	-0.0032	-0.0819	-0.0957	-0.1118
Specificity	est	-0.1122	-0.1118	-0.1122	-0.1122	-0.1122	-0.1178
	SE				0.0201	0.0201	0.0201
	up	0.3325		-0.0804	-0.0725	-0.0706	-0.0785
	low	-0.5993		-0.1440	-0.1519	-0.1538	-0.1572

Results

- Sensitivity:
  - All the MI estimates are quite close to each other
  - The ML estimate is smaller, but still very close to the other estimates (less then a half of SE)
  - The confidence intervals are similar (except MJCI)
  - There is no difference between the sensitivities of the two tests
- Specificity:
  - The specificity estimates are all close to each other
  - It seems that all tests find significant differences between the two tests (except the NHCI)

6. Discussion

- ✓ We use a well accepted missing data procedure to correct for verification bias
- ✓ We showed that MI performs much better then the existing procedures
- ✓ For the scenario in which the missing values are not ignorable, it is simple to still use the MI procedure, with only slight modifications
- ✓ Using MI allows us to do sensitivity analysis by using several different analyses