Tests for Equivalence or Non-inferiority for Paired Binary Data

James J. Chen
Division of Biometry and Risk Assessment, National Center for Toxicological Research
3900 NCTR Drive
Jefferson, Arkansas 72079, USA
jchen@nctr.fda.gov

Huey-miin Hsueh and Jen-pei Liu
Division of Biostatistics and Bioinformatics, National Health Research Institutes
128 Yen-Chiu-Yua Rd, Section 2
Taipei, 11529, Taiwan
jpliu@nhri.org.tw; hsueh@nhri.org.tw

1. INTRODUCTION

Sensitivity and specificity have traditionally been used to assess the performance of a diagnostic procedure. Diagnostic procedures with both high sensitivity and high specificity are desirable, but these procedures are frequently too expensive, hazardous, and/or difficult to operate. A less sophisticated procedure may be preferred, if the loss of the sensitivity or specificity is determined to be acceptable. Assessment of equivalence or non-inferiority between two diagnostic procedures involves comparisons of the response rates between paired binary endpoints. Equivalence between an alternative procedure and a reference procedure is formulated by the null hypothesis that the difference between the two procedures is no smaller than a specific equivalence limit against the alternative hypothesis that the difference is between the limits. The test can be performed by simultaneously testing two sets of one-sided hypothesis. A non-inferiority test is similarly formulated by an one-sided equivalent hypothesis.

2. METHOD

Let variable T_1 be the binary diagnosis by a new test procedure, T_2 be the diagnosis by a reference procedure and Y be the diagnosis of a gold standard. Let x_{ijk} denote the number of the observed outcomes $(T_1 = i, T_2 = j, Y = k)$ and p_{ijk} be the probability associated with the respective outcome x_{ijk}. The data can be summarized by a 4×4 table:

		Y = 1		Y = 0	
Standard	Test Outcome	$T_2 = 1$	$T_2 = 0$	$T_2 = 1$	$T_2 = 0$
Y = 1	$T_1 = 1$	x_{111} (p_{111})	x_{101} (p_{101})	0	0
	$T_1 = 0$	x_{011} (p_{011})	x_{001} (p_{001})	0	0
Y = 0	$T_1 = 1$	0	0	x_{110} (p_{110})	x_{100} (p_{100})
	$T_1 = 0$	0	0	x_{010} (p_{010})	x_{000} (p_{000})

Define the probabilities $p_{+k} = P(Y = k) = (p_{11k} + p_{10k} + p_{01k} + p_{00k})$ and $P(T_1 = i, T_2 = j|Y = k) = p_{ij|k} = p_{ijk}/p_{+k}$. The sensitivities of T_1 and T_2, respectively, are given as $\theta_{e1} = p_{11|1} + p_{10|1}$ and $\theta_{e2} = p_{11|1} + p_{01|1}$. The difference of the sensitivities is $\theta_e = p_{10|1} - p_{01|1}$. Similarly, the specificities of T_1 and T_2 are given as $\theta_{p1} = p_{01|0} + p_{00|0}$ and $\theta_{p2} = p_{10|0} + p_{00|0}$. The difference of the specificities is $\theta_p = p_{01|0} - p_{10|0}$.

Alternatively, The 4 x 4 Table can be collapsed into the following 2×2 table representing the correct/incorrect diagnosis classifications,

		T_2	
	Result	Correct	Incorrect
T_1	Correct	$x_{111} + x_{000}(p_{111} + p_{000})$	$x_{101} + x_{010}(p_{101} + p_{010})$
	Incorrect	$x_{011} + x_{110}(p_{011} + p_{110})$	$x_{001} + x_{100}(p_{001} + p_{100})$

This paper addresses the problem of comparing sensitivity and specificity between an alternative procedure and a reference procedure. We will present asymptotic test and exact procedures to establish equivalence or non-inferiority.

RESUME

James J. Chen: Mathematical Statistician, Division of Biometry and Risk Assessment, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson AR 72079, USA. Ph.D. in Statistics, Iowa State University (1978); MA in Mathematics, University of Pittsburgh (1974); BS in Mathematics, Taiwan Tsing-Hua University (1970).