

TESIS DOCTORAL

**Intervalos de confianza y contrastes de hipótesis
para parámetros de tests diagnósticos binarios**



Departamento de Estadística e Investigación Operativa
Universidad de Granada

Miguel Ángel Montero Alonso
Granada, 2010

Editor: Editorial de la Universidad de Granada
Autor: Miguel Ángel Montero Alonso
D.L.: en trámite
ISBN: en trámite



Universidad de Granada
Departamento de Estadística e Investigación Operativa

**INTERVALOS DE CONFIANZA Y CONTRASTES DE HIPÓTESIS
PARA PARÁMETROS DE TESTS DIAGNÓSTICOS BINARIOS**

Memoria de TESIS DOCTORAL realizada bajo la dirección de los Doctores D. Juan de Dios Luna del Castillo y D. José Antonio Roldán Nofuentes, profesores del Departamento de Estadística e Investigación Operativa de la Universidad de Granada, presentada por Miguel Ángel Montero Alonso para optar al grado de Doctor en Estadística por la Universidad de Granada.

Fdo.: Miguel Ángel Montero Alonso

Vº Bº Directores de tesis

Fdo. Dr. D. Juan de Dios Luna del Castillo Fdo. Dr. D. José Antonio Roldán Nofuentes

La presente tesis doctoral esta avalada hasta la fecha de su lectura por los artículos:

Roldán Nofuentes, J.A., Luna del Castillo, J.D., **Montero Alonso, M.A.**, (2009a). Determining sample size to evaluate and compare the accuracy of binary diagnostic tests in the presence of partial disease verification. *Computational Statistics and Data Analysis*, **53**(3): 742 – 755.

DOI: 10.1016/j.csda.2008.07.030

Roldán Nofuentes J.A., Luna del Castillo J.D., **Montero Alonso, M.A.**, (2009b). Confidence intervals of weighted kappa coefficient of a binary diagnostic test. *Communications in Statistics. Simulation and Computation*, **38**: 1562 – 1578.

DOI: 10.1080/03610910903039473

Agradecimientos

Quisiera expresar mi más sincero agradecimiento a los directores de esta memoria, Dr. D. Juan de Dios Luna del Castillo y Dr. D. José Antonio Roldán Nofuentes, por el interés y colaboración que desde el principio han mostrado en la realización de este trabajo.

También mi agradecimiento a aquellas personas que, de un modo u otro han ayudado en la realización de este trabajo ofreciendo todo su apoyo y ánimo, mis amigos y compañeros, y los no tan amigos, porque de todas las personas se aprende algo.

Por último, dar las gracias a mi familia, por su apoyo incondicional, en especial a mi madre, por decir tanto cuando no dice nada.

Gracias.

A mi padre,

In Memoriam

Índice

	Pág
Prólogo	VII
1. Medidas de un test de diagnóstico binario: verificación completa y verificación parcial.....	1
1.1. Medidas de la calidad de un test de diagnóstico binario.....	3
1.1.1. Sensibilidad y especificidad.....	4
1.1.2. Razones de verosimilitud.....	5
1.1.3. Índice de Youden.....	7
1.1.4. Valores predictivos.....	8
1.1.5. Coeficiente kappa ponderado.....	9
1.2. Estimación de los parámetros de un test diagnóstico binario con verificación completa.....	14
1.2.1. Estimación de sensibilidad y especificidad.....	15
1.2.1.1. Estimación bajo un muestreo transversal.....	16
1.2.1.2. Estimación bajo un muestreo caso-control.....	21
1.2.2. Estimación de las razones de verosimilitud	22
1.2.3. Estimación del índice de Youden.....	24
1.2.4. Estimación de los valores predictivos	25
1.2.4.1. Estimación bajo un muestreo transversal.....	25
1.2.4.2. Estimación bajo un muestreo caso-control.....	28
1.2.5. Estimación del coeficiente kappa ponderado (Aportación).....	30
1.2.5.1. Intervalos de confianza para el coeficiente kappa ponderado.....	30
1.2.5.2. Experimentos de simulación.....	33

	Pág
1.2.5.3. Aplicación.....	40
1.2.5.4. Discusión.....	42
1.3. La verificación parcial de la enfermedad y el sesgo de verificación ligado a ella.....	44
1.3.1. Estimación de la sensibilidad y la especificidad.....	46
1.3.1.1. Estimadores máximo verosímiles de la sensibilidad y de la especificidad sin covariables.....	47
1.3.1.2. Estimadores máximo verosímiles en presencia de covariables.....	57
1.3.1.3. Estimadores simples de la sensibilidad y de la especificidad.....	60
1.3.1.4. Intervalos de confianza para la sensibilidad y la especificidad.....	62
1.3.1.5. Comparación entre intervalos (Aportación).....	67
1.3.1.5.1. Comparación de los intervalos para la sensibilidad.....	106
1.3.1.5.2. Comparación de los intervalos para la especificidad.....	108
1.3.1.5.3. Conclusiones.....	109
1.3.1.6. Análisis de sensibilidad ante la no verificación de la hipótesis MAR.....	110
1.3.2. Estimación de los valores predictivos.....	114
1.3.2.1. Estimadores simples de los valores predictivos.....	114
1.3.2.2. Estimadores máximo verosímiles de los valores predictivos.....	118
1.3.3. Estimación del coeficiente kappa ponderado.....	121
1.3.3.1. Estimación del coeficiente kappa ponderado sin covariables.....	122

	Pág
1.3.3.2. Estimación del coeficiente kappa ponderado con covariables.....	123
2. Comparación de tests diagnósticos binarios.....	127
2.1. Verificación completa.....	128
2.1.1. Comparación de las sensibilidades y de las especificidades....	128
2.1.1.1. Comparación con muestras independientes.....	128
2.1.1.2. Comparación con muestras apareadas.....	130
2.1.1.3. Intervalos de confianza para la diferencias de las sensibilidades y de las especificidades con muestras apareadas.....	133
2.1.2. Comparación de razones de verosimilitud	134
2.1.2.1. Test de hipótesis global.....	138
2.1.3. Comparación de valores predictivos.....	138
2.1.4. Comparación de coeficientes kappa ponderados.....	141
2.1.5. Comparación de múltiples coeficientes kappa ponderados (Aportación).....	142
2.1.5.1. Test de hipótesis conjunto.....	143
2.1.5.2. Experimentos de simulación.....	147
2.1.5.2.1. Error tipo I.....	149
2.1.5.2.2. Potencia.....	155
2.1.5.3. Aplicación.....	158
2.1.5.4. Conclusiones.....	162
2.2. Verificación parcial.....	163
2.2.1. Comparación de dos tests binarios.....	164
2.2.1.1. Comparación de las sensibilidades y de las	

	Pág
especificidades.....	164
2.2.1.1.1. Comparación de sensibilidades y especificidades sin covariables.....	165
2.2.1.1.2. Comparación de sensibilidades y especificidades con covariables.....	168
2.2.1.2. Comparación de razones de verosimilitud.....	171
2.2.1.2.1. Comparación de las razones de verosimilitud sin covariables.....	171
2.2.1.2.2. Comparación de las razones de verosimilitud con covariables.....	173
2.2.1.3. Comparación de los valores predictivos.....	175
2.2.1.3.1. Comparación de valores predictivos sin covariables.....	175
2.2.1.3.2. Comparación de valores predictivos con covariables.....	176
2.2.1.3.3. Comparación de valores predictivos con los estimadores simples.....	177
2.2.1.4. Comparación de coeficientes kappa ponderados.....	179
2.2.1.5. Imputación múltiple para comparar dos tests.....	181
2.2.1.6. Algoritmos EM y SEM para comparar dos tests.....	183
2.2.2. Comparaciones de múltiples tests.....	188
2.2.2.1. Comparación de las sensibilidades y de las especificidades.....	189
2.2.2.2. Comparación de las razones de verosimilitud.....	194
2.2.2.2.1. Comparación de las razones de verosimilitud positivas.....	194
2.2.2.2.2. Comparación las de razones de verosimilitud negativas.....	196

	Pág
2.2.2.3. Comparación de los valores predictivos.....	197
2.2.2.4. Comparación de coeficientes kappa ponderados.....	198
3. Cálculo del tamaño de muestra para evaluar y comparar parámetros de tests diagnósticos binarios	201
3.1. Verificación completa.....	202
3.1.1. Tamaño de muestra para la estimación por intervalos de confianza.....	202
3.1.1.1. Tamaño de muestra para estimar la sensibilidad y la especificidad.....	203
3.1.1.2. Tamaño de muestra para estimar el coeficiente kappa ponderado (Aportación).....	205
3.1.1.2.1. Experimentos de simulación.....	207
3.1.1.2.2. Aplicación.....	213
3.1.1.2.3. Conclusión.....	214
3.1.2. Tamaño muestral para la comparación de dos tests binarios...	214
3.1.2.1. Tamaño muestral para comparar las sensibilidades y/o especificidades con dos muestras independientes.....	215
3.1.2.2. Tamaño muestral para comparar las sensibilidades y/o especificidades con dos muestras apareadas.....	217
3.2. Verificación parcial.....	219
3.2.1. Tamaño muestral para estimar la sensibilidad y la especificidad de un test binario en presencia de verificación parcial de la enfermedad (Aportación).....	219
3.2.1.1. Experiencia de simulación.....	223
3.2.1.2. Aplicación.....	230
3.2.2. Tamaño de muestra para la comparación de dos tests binarios en presencia de verificación parcial. (Aportación).....	231

	Pág
3.2.2.1. Experimentos de simulación.....	237
3.2.2.2. Aplicación.....	245
3.2.2.3. Conclusiones.....	246
Conclusiones.....	251
Bibliografía.....	255

Prólogo

La evaluación y comparación de la precisión de tests diagnósticos binarios es un importante tópico en el estudio de los métodos estadísticos para el diagnóstico de enfermedades. La precisión de un test diagnóstico binario se mide en términos de distintos parámetros, como son la sensibilidad y la especificidad, las razones de verosimilitud positiva y negativa, los valores predictivos positivo y negativo, y el coeficiente kappa ponderado. La evaluación y/o la comparación de tests diagnósticos binarios se realiza respecto a un gold estándar, que es un método que permite conocer el verdadero estado de enfermedad de un paciente. Por tanto, la estimación y comparación de los parámetros de tests diagnósticos binarios requiere conocer tanto los resultados de los tests diagnósticos binarios como el resultado del gold estándar en todos los pacientes de una muestra aleatoria. Sin embargo, en muchas situaciones prácticas es frecuente que el gold estándar no se aplique a todos los pacientes de la muestra, surgiendo el denominado problema de la verificación parcial de la enfermedad. Si en esta situación la estimación y/o comparación de algunos parámetros se realiza prescindiendo de los pacientes a los que no se les aplica el gold estándar, los estimadores obtenidos están afectados por el llamado sesgo de verificación. Por tanto, en esta situación la aplicación de los métodos clásicos de estimación y/o comparación de parámetros puede dar lugar a resultados erróneos.

Por otra parte, una importante cuestión en la evaluación y la comparación de la precisión de tests diagnósticos binarios es la determinación del tamaño de muestra necesario para llevar a cabo el estudio, tanto cuando el gold estándar se aplica a todos los pacientes de la muestra como en presencia de verificación parcial de la enfermedad. En el primer caso, el tamaño muestra se determina aplicando métodos en tablas 2×2 (cuando se evalúa un único test binario) o en tablas 2×4 (cuando se comparan dos tests binarios); sin embargo en presencia de verificación parcial de la enfermedad estos métodos no se pueden aplicar ya que para algunos pacientes el estado de enfermedad es desconocido.

La presente Tesis Doctoral, estructurada en tres capítulos, supone una ampliación en el estudio de los intervalos de confianza y contrastes de hipótesis para distintos parámetros de tests diagnósticos tanto cuando el gold estándar se aplica a todos los pacientes de una muestra aleatoria como en presencia de verificación parcial de la enfermedad.

En el primer Capítulo, estructurado en dos partes bien diferenciadas, se estudian dos cuestiones generales. En primer lugar se definen los parámetros de un test binario, se estudian sus propiedades y sus estimaciones cuando el gold estándar se aplica a todos los pacientes de una muestra aleatoria. En segundo lugar se estudia el problema de la verificación parcial de la enfermedad y el sesgo de verificación ligado a ella, y se estudia la estimación de algunos parámetros en presencia del sesgo de verificación. En este primer Capítulo se han realizado dos aportaciones, la primera consiste en la estimación por intervalos de confianza del coeficiente kappa ponderado cuando el gold estándar se aplica a todos los pacientes de una muestra, y la segunda aportación consiste en un nuevo intervalo de confianza para la sensibilidad y la especificidad de un test diagnóstico binario en presencia del sesgo de verificación.

En el segundo Capítulo se estudian distintos tests de hipótesis para comparar las sensibilidades y las especificidades, las razones de verosimilitud positiva y negativa, los valores predictivos positivo y negativo, y los coeficientes kappa ponderados, de dos tests binarios en presencia de verificación completa y en presencia de verificación parcial de la enfermedad. En el caso de verificación parcial, también se estudian los tests de hipótesis globales para comparar los parámetros de múltiples tests binarios. En

este segundo Capítulo se ha realizado una aportación, la comparación de múltiples coeficientes kappa ponderados cuando todos los pacientes son verificados con el gold estándar.

En el tercer Capítulo se estudia el problema de la determinación del tamaño muestral necesario para estimar parámetros de un test binario y para comparar parámetros de dos tests binarios, tanto cuando el gold estándar se aplica a todos los pacientes como en presencia del sesgo de verificación. En este Capítulo se han realizado dos aportaciones: la primera consiste en un método para calcular el tamaño muestral necesario para estimar el coeficiente kappa ponderado de un test binario cuando todos los pacientes son verificados con el gold estándar, y la segunda es un método para calcular el tamaño muestral necesario para estimar la sensibilidad (especificidad) de un test binario y para comparar las sensibilidades (especificidades) de dos tests binarios en presencia de verificación parcial de la enfermedad.

Granada, enero de 2010.

Capítulo 1

Medidas de un test diagnóstico binario: verificación completa y verificación parcial

La constante evolución de la Medicina en estas últimas décadas ha hecho necesario que la Estadística desarrolle nuevos métodos para resolver los diferentes problemas que se han ido planteando, destacando entre ellos los métodos estadísticos para el diagnóstico médico. El proceso para llegar a un diagnóstico médico puede ser imperfecto y se ha de ser consciente de que la presunción diagnóstica se hace en términos de probabilidades, y que con frecuencia no se tiene la certidumbre de un diagnóstico definitivo incluso después de múltiples pruebas médicas aplicadas al paciente. Para abordar este estudio, es necesario conocer algunos conceptos básicos, por lo que en este capítulo se describen diversos conceptos introductorios relacionados con las diferentes medidas de la calidad de un test de diagnóstico binario, su estimación y el problema que surge con la verificación parcial, así como una serie de aportaciones a dicho estudio.

Los métodos de diagnóstico son fundamentales en la práctica de la medicina moderna (Barratt et al., 1999) y consisten en aplicar a un individuo una determinada prueba médica para confirmar o no la presencia de una determinada enfermedad. La aplicación de éstos en la evaluación de enfermedades tiene diferentes propósitos (McNeil y Adelstein, 1976; Sox et al., 1989 y Roldán Nofuentes, 2003):

- Proporcionar información fiable sobre el estado de la enfermedad de un paciente.
- Influir en la planificación del tratamiento de un paciente.
- Comprender el mecanismo y la naturaleza de la enfermedad a través de la investigación.

La interpretación de un test de diagnóstico depende a su vez de varios factores:

- De la habilidad intrínseca del test diagnóstico para distinguir entre pacientes enfermos y sanos (exactitud discriminatoria)
- De las características particulares de cada individuo
- Del ambiente en el que se aplica el test diagnóstico.

Los tests diagnósticos se clasifican en:

- Binarios, por ejemplo, padecer o no una cierta enfermedad: si padece la Hepatitis C o no padece la Hepatitis C.
- Discretos, por ejemplo, realizar un test de glucosa donde el resultado puede ser test positivo, test negativo o test con resultado de incertidumbre.
- Ordinales, por ejemplo, una clasificación de la presencia de la enfermedad: definitivamente, probablemente, probablemente no, definitivamente no.
- Cuantitativos o continuos, por ejemplo la temperatura del cuerpo humano.

En la práctica clínica los tests diagnósticos más comunes son los binarios debido a su facilidad de manejo.

Para evaluar la exactitud de un test diagnóstico es necesario disponer de un estimador insesgado de la exactitud del test, para lo cual se necesita determinar el estado de enfermedad (enfermo o no) de cada paciente, independientemente del resultado del

test diagnóstico. El procedimiento mediante el cual se conoce el verdadero estado de enfermedad de cada paciente se denomina gold estándar (por ejemplo, una biopsia, una evaluación clínica, etc.). En la práctica, es frecuente que el gold estándar no se aplique a todos los pacientes de una muestra y por tanto, el estado de enfermedad de algunos pacientes es desconocido, surgiendo el llamado problema de la verificación parcial de la enfermedad.

En este primer Capítulo se diferencian tres partes: en la primera se analizan los distintos parámetros de un test diagnóstico binario; en la segunda se estiman dichos parámetros cuando todos los pacientes tienen verificado su estado de enfermedad (verificación completa de la enfermedad); y en la tercera se estudia el problema de la verificación parcial de la enfermedad y el sesgo de verificación ligado a ella, así como el método de corrección del sesgo, la obtención de los estimadores por máxima verosimilitud de la exactitud de un test binario y el efecto del sesgo en distintos parámetros de interés de un test binario. En este capítulo se realizan dos aportaciones: 1) estimación por intervalos de confianza del coeficiente kappa ponderado de un test diagnóstico binario cuando el test diagnóstico y el gold estándar se aplican a todos los pacientes de una muestra aleatoria, y 2) intervalo de confianza cuadrático para la sensibilidad y la especificidad de un test diagnóstico binario en presencia de verificación parcial de la enfermedad.

1.1.- Medidas de la calidad de un test diagnóstico binario

Considérese una determinada enfermedad tal que los individuos de una población la pueden presentar o no. Supóngase que se dispone de un gold estándar para el diagnóstico de dicha enfermedad. En este caso, se define D como la variable aleatoria que modeliza el resultado del gold estándar, de tal forma que $D = 1$ cuando el paciente tiene la enfermedad y $D = 0$ cuando el paciente no tiene la enfermedad. La probabilidad de que un sujeto tenga la enfermedad, denotada como $p = P(D = 1)$, se denomina prevalencia de la enfermedad. Si a un individuo de la población objeto de estudio se le aplica un test diagnóstico para diagnosticar dicha enfermedad, de tal forma que el resultado de dicho test diagnóstico puede ser positivo o negativo, se define la variable

aleatoria T como la variable que modeliza el resultado del test diagnóstico, de tal forma que $T = 1$ cuando el resultado del test es positivo (indicando la presencia provisional de la enfermedad) y $T = 0$ cuando el resultado del test es negativo (indicando la ausencia provisional de la enfermedad). El test diagnóstico se puede equivocar y por tanto se tienen probabilidades de acertar y de fallar en el diagnóstico de la enfermedad. De esta forma, si a un individuo que está enfermo (determinado con el gold estándar) se le aplica el test y éste da positivo se está ante un acierto que se denomina verdadero positivo. Cuando a un individuo no enfermo se le aplica el test y éste da negativo se está ante un acierto denominado verdadero negativo. Por el contrario, si el test da un resultado negativo estando el paciente enfermo se presenta un fallo denominado falso negativo. De igual forma si el test da un resultado positivo cuando el individuo no está enfermo se está cometiendo un fallo denominado falso positivo. Por tanto, un test diagnóstico intenta distinguir los pacientes con la enfermedad de los pacientes sin la enfermedad. Si el verdadero estado de enfermedad se puede conocer mediante la aplicación de un gold estándar, la exactitud del test diagnóstico se puede medir en términos de diferentes parámetros. A continuación se estudian las medidas más importantes del rendimiento de un test diagnóstico binario, haciendo referencia a las probabilidades de los sucesos mencionados anteriormente.

1.1.1. Sensibilidad y especificidad

En la literatura médica la exactitud de un test diagnóstico (Feinstein, 1975; Swets, 1979; Swets et al., 1979) es normalmente medida por las probabilidades condicionadas de clasificar correctamente un paciente enfermo o no enfermo: la sensibilidad y la especificidad. La sensibilidad (Se) es la probabilidad de que el resultado del test binario sea positivo cuando el paciente está enfermo, es decir, es la probabilidad de que un test sea positivo dado que el verdadero estado es positivo, es decir $Se = P(T = 1 / D = 1)$, y la especificidad (Sp) es la probabilidad de que el resultado del test binario sea negativo cuando el paciente no está enfermo, es decir $Sp = P(T = 0 / D = 0)$. Tanto la sensibilidad como la especificidad son probabilidades de acierto en el diagnóstico de la enfermedad. Al acierto que hace referencia la

sensibilidad se le denomina verdadero positivo (TP) y al acierto al que hace referencia la especificidad se le denomina verdadero negativo (TN). Se verifica que la suma de las probabilidades de un verdadero positivo y de un falso negativo es la unidad,

$$P(T = 1 | D = 1) + P(T = 0 | D = 1) = 1, \quad (1.1)$$

y de forma análoga, se tiene que

$$P(T = 0 | D = 0) + P(T = 1 | D = 0) = 1. \quad (1.2)$$

A la probabilidad $P(T = 0 | D = 1) = 1 - P(T = 1 | D = 1)$ se la denominada probabilidad de un falso negativo (FN) y a la probabilidad $P(T = 1 | D = 0) = 1 - P(T = 0 | D = 0)$ se la denomina probabilidad de un falso positivo (FP). La sensibilidad y la especificidad representan la exactitud del test diagnóstico y dependen solamente de la habilidad intrínseca del test diagnóstico para distinguir entre pacientes enfermos y pacientes no enfermos. Por consiguiente, ambos parámetros dependen de la base biológico, física, química, etc..., del test diagnóstico para distinguir entre ambos tipos de pacientes (enfermos y no enfermos). Un test diagnóstico con una alta sensibilidad es útil para descartar la enfermedad ya que la probabilidad de un falso negativo es pequeña, y un test con una alta especificidad es útil para confirmar la enfermedad ya que la probabilidad de un falso positivo es pequeña. De esta forma, como test de screening se requieren tests diagnósticos con una alta sensibilidad.

1.1.2. Razones de verosimilitud

Otras medidas del rendimiento de un test diagnóstico binario son las razones de verosimilitud (LRs). Para un resultado i del test diagnóstico, la razón de verosimilitud es un cociente de dos probabilidades definido como

$$LR = \frac{P(T = i | D = 1)}{P(T = i | D = 0)} \quad (1.3)$$

donde $i = 0, 1$. Si el resultado del test diagnóstico es positivo, la razón de verosimilitud, denominada razón de verosimilitud positiva, es el cociente entre la sensibilidad y uno menos la especificidad, esto es

$$LR^+ = \frac{P(T = 1 | D = 1)}{P(T = 1 | D = 0)} = \frac{Se}{1 - Sp}, \quad (1.4)$$

y representa el cociente entre la probabilidad de un resultado positivo del test diagnóstico en la población de individuos enfermos y la probabilidad de un resultado positivo del test diagnóstico en la población de individuos no enfermos. Cuando el resultado del test es negativo, la razón de verosimilitud, denominada razón de verosimilitud negativa, es

$$LR^- = \frac{P(T = 0 | D = 1)}{P(T = 0 | D = 0)} = \frac{1 - Se}{Sp}, \quad (1.5)$$

y representa el cociente entre la probabilidad de un resultado negativo del test diagnóstico en la población de individuos enfermos y la probabilidad de un resultado negativo del test diagnóstico en la población de individuos no enfermos. Las razones de verosimilitud positiva y negativa no dependen de la prevalencia de la enfermedad y solamente dependen de la sensibilidad y especificidad del test diagnóstico, y sus valores varían entre cero e infinito. Cuando el test diagnóstico y el gold estándar son independientes entonces $LR^+ = LR^- = 1$, y si el test diagnóstico clasifica correctamente a todos los pacientes (enfermos o no enfermos) entonces $LR^+ = \infty$ y $LR^- = 0$. Un valor $LR^+ > 1$ indica que un resultado positivo del test diagnóstico es más probable en un individuo enfermo que en un individuo no enfermo (como es de esperar), y un valor $LR^- < 1$ indica que un resultado negativo del test diagnóstico es más probable en un paciente no enfermo que en uno enfermo. Las razones de verosimilitud cuantifican el incremento sobre el conocimiento de la presencia de la enfermedad mediante la aplicación del test diagnóstico. Antes de aplicar el test diagnóstico, la odds de que un paciente tenga la enfermedad es

$$\text{pre-test odds} = \frac{p}{1 - p}, \quad (1.6)$$

y después de aplicar el test, la odds de que un paciente tenga la enfermedad es

$$\text{post-test odds} = \frac{P(D=1|T=i)}{P(D=0|T=i)}, \quad i=0,1. \quad (1.7)$$

Por consiguiente, las razones de verosimilitud relacionan la pre-test odds con las post-tests odds, esto es

$$\begin{aligned} \text{Post-test odds } (T=1) &= LR^+ \times \text{pre-test odds} \\ \text{Post-test odds } (T=0) &= LR^- \times \text{pre-test odds}, \end{aligned} \quad (1.8)$$

y por tanto las razones de verosimilitud cuantifican el cambio en la odds de la enfermedad obtenido por el conocimiento de la aplicación del test diagnóstico binario.

1.1.3. Índice de Youden

El índice de Youden (Youden, 1950) se define como la suma de la sensibilidad y la especificidad menos uno, es decir,

$$I = Se + Sp - 1. \quad (1.9)$$

Sus valores varían entre -1 y 1, y presenta las siguientes propiedades:

- Cuando el test es independiente de la enfermedad, la especificidad y sensibilidad son complementarias, es decir, si el test y el gold estándar son independientes

$$P(D=1|T=1) = P(D=1) \Rightarrow \quad (1.10)$$

$$\frac{P(D=1)P(T=1|D=1)}{P(D=1)P(T=1|D=1) + P(D=0)P(T=1|D=0)} = P(D=1),$$

y después de unos cálculos se obtiene $Se + Sp = 1$.

- El índice de Youden mide la discrepancia del test diagnóstico con la independencia del gold estándar, de manera que se obtiene la máxima

discrepancia cuando se verifica que $P(T = 1/D = 1) = P(T = 0/D = 0) = 1$.

En esta situación, se habla de dependencia total. La situación en la que ambas probabilidades son cero da lugar a una situación que no tiene sentido en la práctica.

- Cuando el test diagnóstico está relacionado con la enfermedad (aspecto deseable para un test diagnóstico), entonces $P(D = 1|T = 1) > P(D = 1)$ y $P(D = 0|T = 0) > P(D = 0)$ obteniéndose que $Se + Sp > 1$. Este resultado es muy importante, ya que si no se verifica, obliga a que uno de los dos sumandos tiene que ser inevitablemente inferior a 0.5, lo que implicaría que el test falla más que una asignación aleatoria al estado de la enfermedad. Por consiguiente, a todo test diagnóstico se le debe exigir que su índice de Youden sea mayor que cero ($Se + Sp - 1 > 0$).

El índice de Youden es por tanto una medida de la exactitud de un test diagnóstico binario que no depende de la prevalencia de la enfermedad, pero presenta el inconveniente de una difícil interpretación al ser la suma de dos probabilidades de distinto condicionante menos una constante.

1.1.4. Valores predictivos

Los valores predictivos positivo y negativo representan la exactitud clínica del test diagnóstico cuando este se aplica a una cohorte de individuos, y dependen tanto de la habilidad intrínseca del test diagnóstico para distinguir entre individuos enfermos e individuos no enfermos como de la prevalencia de la enfermedad. El valor predictivo positivo (vpp) es la probabilidad de que el paciente esté enfermo cuando el resultado del test es positivo, esto es $vpp = P(D = 1|T = 1)$; y el valor predictivo negativo (vpn) es la probabilidad de que el paciente no esté enfermo cuando el resultado del test es negativo, es decir $vpn = P(D = 0|T = 0)$. Aplicando el teorema de Bayes se obtiene que

$$vpp = P(D = 1|T = 1) = \frac{pSe}{pSe + (1 - p)(1 - Sp)} \quad (1.11)$$

y

$$vpn = P(D = 0 | T = 0) = \frac{(1-p)Sp}{(1-p)Sp + p(1-Se)}, \quad (1.12)$$

siendo p la prevalencia de la enfermedad. El problema que presentan estas medidas es que dependen de la prevalencia de la enfermedad en la población objeto de estudio y no sólo de la calidad del test diagnóstico. Cuanto mayor es la prevalencia mayor es el valor predictivo positivo, y cuanto más pequeña sea la prevalencia de la enfermedad mayor será el valor predictivo negativo. Un test diagnóstico es útil para confirmar la enfermedad si el valor predictivo positivo es alto, es útil para descartar la enfermedad si el valor predictivo negativo es alto, o también puede ser útil para ambos motivos si los dos valores predictivos son altos.

1.1.5. Coeficiente kappa ponderado

El coeficiente kappa ponderado (Kraemer, 1992; Kraemer et al., 2002) de un test diagnóstico binario es una medida del acuerdo clasificatorio entre el test diagnóstico y el gold estándar, y depende de la sensibilidad y especificidad del test diagnóstico, de la prevalencia de la enfermedad y de los costes relativos que se cometen al clasificar erróneamente a un paciente (enfermo o no enfermo) con el test diagnóstico. Sea L la pérdida o coste que se comete al clasificar erróneamente con el test diagnóstico a un paciente enfermo y L' la pérdida o coste que se comete al clasificar erróneamente a un paciente no enfermo. Se asume que las pérdidas son cero cuando un paciente (enfermo o no enfermo) es clasificado correctamente con el test diagnóstico. En la Tabla 1.1 se muestran las probabilidades asociadas a la evaluación de un test diagnóstico binario con respecto a un gold estándar y en la Tabla 1.2 se muestran las pérdidas respectivas, donde la variable T modeliza el resultado del test diagnóstico ($T = 1$ cuando el test es positivo y $T = 0$ es negativo) y la variable D modeliza el resultado del gold estándar ($D = 1$ cuando el paciente está enfermo y $D = 0$ cuando no está enfermo).

Tabla 1.1. Probabilidades obtenidas al aplicar un test binario a una muestra.

	$T = 1$	$T = 0$	Total
$D = 1$	pSe	$p(1 - Se)$	p
$D = 0$	$(1 - p)(1 - Sp)$	$(1 - p)Sp$	$1 - p$
Total	$Q = pSe + (1 - p)(1 - Sp)$	$1 - Q = p(1 - Se) + (1 - p)Sp$	1

Tabla 1.2. Pérdidas obtenidas al aplicar un test binario a una muestra.

	$T = 1$	$T = 0$	Total
$D = 1$	0	L	L
$D = 0$	L'	0	L'
Total	L'	L	$L + L'$

En términos de las probabilidades dadas en la Tabla 1.1 y de las pérdidas dadas en la Tabla 1.2, la pérdida esperada al aplicar el test diagnóstico (riesgo de error en la terminología de Bloch, 1997) es

$$p(1 - Se)L + q(1 - Sp)L', \tag{1.13}$$

siendo $q = 1 - p$, y la pérdida debida al azar es

$$p\{p(1 - Se) + qSp\}L + q\{pSe + q(1 - Sp)\}L'. \tag{1.14}$$

La pérdida esperada es la pérdida media que se comete al clasificar erróneamente con el test diagnóstico a un paciente enfermo o no enfermo, y su rango de valores varía entre cero e infinito. La pérdida aleatoria es la pérdida que se comete cuando el test diagnóstico y el gold estándar son independientes, es decir cuando

$P(T = i | D = j) = P(T = i)$. En la práctica, tanto la pérdida esperada como la pérdida debida al azar no se pueden estimar ya que las pérdidas L y L' asociadas a una clasificación errónea del test diagnóstico son valores desconocidos. En términos de la pérdida esperada y de la pérdida debida al azar se define el coeficiente kappa ponderado de un test diagnóstico binario como

$$\kappa = \frac{\text{Pérdida debida al azar} - \text{Pérdida esperada}}{\text{Pérdida debida al azar}}, \quad (1.15)$$

que es una medida de la discrepancia relativa entre la pérdida debida al azar y la pérdida esperada. Sustituyendo en la ecuación (1.15) cada pérdida por su fórmula se obtiene que el coeficiente kappa ponderado de un test diagnóstico binario viene dado por la expresión

$$\kappa(c) = \frac{pq(Se + Sp - 1)}{p\{p(1 - Se) + qSp\}c + q\{pSe + q(1 - Sp)\}(1 - c)}, \quad (1.16)$$

donde $q = 1 - p$ y $c = L/(L + L')$ es el índice de ponderación. El índice de ponderación c varía entre 0 y 1, y representa la pérdida relativa entre los falsos positivos y los falsos negativos. En la práctica el índice c es desconocido, pero sus valores se pueden determinar dependiendo del objetivo para el que se va a utilizar el test diagnóstico. Si el test diagnóstico se va a utilizar como un test de screening, hay una mayor preocupación por los falsos negativos y el índice c es mayor que 0.5; si el test diagnóstico se va a utilizar como paso previo a un tratamiento intensivo, hay más preocupación por los falsos positivos y el índice c es menor que 0.5; y el índice c vale 0.5 cuando el test diagnóstico se utiliza para un diagnóstico simple. Si en la ecuación (1.14) se verifica que $L = L'$, entonces $c = 0.5$ y $\kappa(0.5)$ se denomina coeficiente kappa de Cohen (Cohen, 1960), mientras que si $L > L'$ entonces $0.5 < c < 1$ y si $L' > L$ entonces $0 < c < 0.5$. Cuando la pérdida L es cero entonces $c = 0$ y el coeficiente kappa ponderado viene dado por la expresión

$$\kappa(0) = \frac{Sp - (1 - Q)}{Q} = \frac{vpp - p}{1 - p}, \quad (1.17)$$

donde $Q = pSe + (1-p)(1-Sp)$ y vpp es el valor predictivo positivo. Cuando la pérdida L' es cero entonces $c = 1$ y el coeficiente de la kappa ponderado es

$$\kappa(1) = \frac{Se - Q}{1 - Q} = \frac{vpp - (1 - p)}{p}, \quad (1.18)$$

donde vpp es el valor predictivo negativo. El coeficiente kappa ponderado $\kappa(c)$ se puede escribir en términos de p , Q , $\kappa(0)$ y $\kappa(1)$ como

$$\kappa(c) = \frac{p(1-Q)c\kappa(1) + (1-p)Q(1-c)\kappa(0)}{p(1-Q)c + (1-p)Q(1-c)}, \quad (1.19)$$

con $0 < c < 1$, por lo que el coeficiente kappa ponderado es una media ponderada de $\kappa(0)$ y $\kappa(1)$. El coeficiente kappa ponderado de un test binario presenta las siguientes propiedades:

- a) Si el acuerdo clasificatorio entre el test binario y el gold estándar es perfecto ($Se = Sp = 1$) entonces la pérdida esperada es 0 y $\kappa(c) = 1$.
- b) Si la sensibilidad y la especificidad son complementarias ($Se = 1 - Sp$) entonces $\kappa(c) = 0$.
- c) Si la pérdida debida al azar es mayor que la pérdida esperada entonces $\kappa(c) > 0$; y si la pérdida debida al azar es menor que la pérdida esperada entonces $\kappa(c) < 0$ y los resultados del test diagnóstico deben intercambiarse: $T = 1$ debe ser el resultado negativo y $T = 0$ debe ser el resultado positivo. Por tanto, el análisis se debe limitar a los valores positivos del coeficiente kappa ponderado, y sus valores se puede clasificar en la siguiente escala (Landis y Koch, 1977): de 0 a 0.20 la concordancia es mala, de 0.21 a 0.40 la concordancia es mediocre, de 0.41 a 0.60 la concordancia es moderada, de 0.61 a 0.80 la concordancia es buena y de 0.81 a 1 la concordancia es muy buena.
- d) El coeficiente kappa ponderado es una función del índice c que es creciente si $Q > p$, decreciente si $Q < p$ o una función constante igual a $Se + Sp - 1$ si

$Q = p$. En las Figuras 1.1 y 1.2 se muestra como varía el coeficiente kappa ponderado en función del índice de ponderación c y de la prevalencia de la enfermedad, para unos mismos valores de sensibilidad y especificidad. El coeficiente kappa ponderado es una función creciente o decreciente en el índice de ponderación c , dependiendo de la prevalencia de la enfermedad.

Figura 1.1. Coeficiente Kappa ponderado para $Se = 0.80$, $Sp = 0.75$ y diferentes valores para la prevalencia.

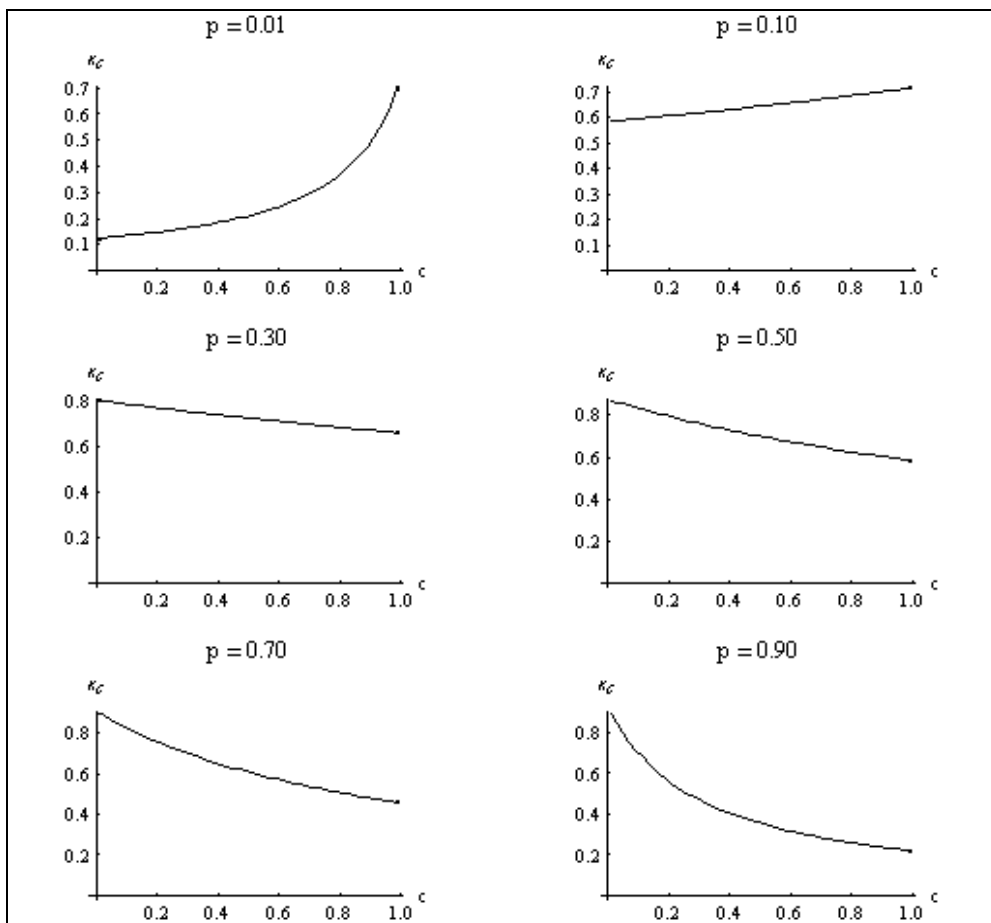
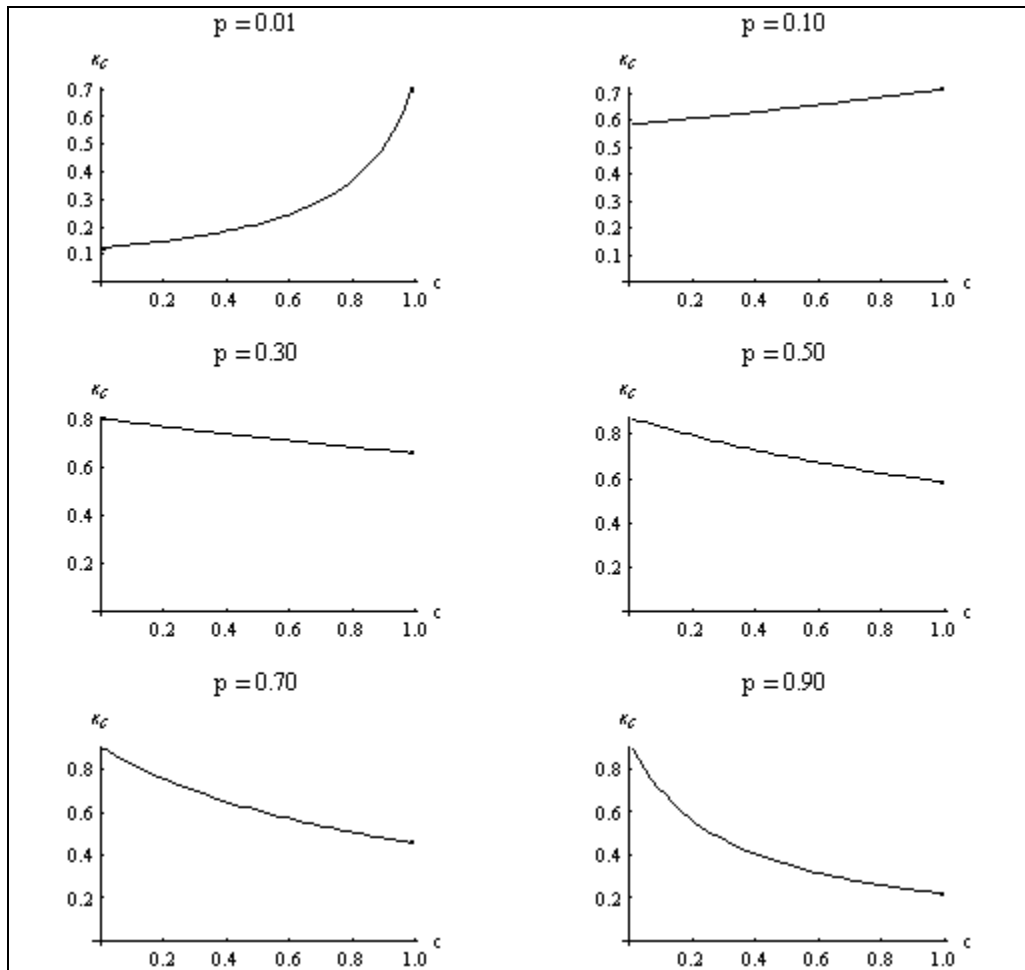


Figura 1.2. Coeficiente Kappa ponderado para $Se = 0.75$, $Sp = 0.95$ y diferentes valores para la prevalencia.



Por tanto, el coeficiente kappa ponderado de un test diagnóstico binario es una medida válida para evaluar y comparar el rendimiento de un test diagnóstico y mide el acuerdo clasificatorio o concordancia entre el test diagnóstico y el gold estándar cuando ambos se aplican a una misma cohorte de pacientes.

1.2.- Estimación de los parámetros de un test diagnóstico binario con verificación completa

La estimación de los parámetros estudiados en la Sección 1.1 se puede realizar mediante un estudio transversal o un estudio caso-control. En un estudio transversal tanto el test diagnóstico como el gold estándar se aplican a todos los individuos de una muestra

aleatoria de tamaño n ; y en un estudio caso-control el test diagnóstico se aplica a todos los individuos de dos muestras, una de n_1 individuos que están enfermos ($D = 1$) y otra de n_2 individuos que no están enfermos ($D = 0$). Cuando el muestreo es transversal, se pueden estimar todos los parámetros estudiados en la Sección 1.1. Cuando el muestreo es caso-control, la sensibilidad, la especificidad y las razones de verosimilitud se pueden estimar en este tipo de estudio, pero los valores predictivos y el coeficiente kappa ponderado solamente se pueden estimar si se conoce un estimador de la prevalencia de la enfermedad que sea independiente del propio estudio (por ejemplo, un estimador de la prevalencia de la enfermedad obtenido a partir de otro estudio diferente). Otro tipo de diseño diferente a los dos anteriores (transversal y caso-control) es el muestreo prospectivo, que consiste en aplicar el gold estándar a todos los individuos de dos muestras aleatorias, una formada por individuos en los que el test diagnóstico es positivo y otra muestra formada por individuos en los que el test diagnóstico es negativo. En este tipo de muestreo no se pueden estimar ni la sensibilidad ni la especificidad, por lo que no se utiliza en la práctica.

En esta Sección se va a estudiar la estimación de los parámetros estudiados en la Sección 1.1 bajo un diseño transversal y/o caso-control, y en todos los casos se asume que el estado de enfermedad de todos los individuos objeto de estudio es conocido (verificación completa de la enfermedad), y por consiguiente no hay ningún individuo para el cual no se conozca si está o no enfermo. En esta Sección se realiza una aportación: la estimación por intervalos de confianza del coeficiente kappa ponderado bajo un muestreo transversal.

1.2.1. Estimación de la sensibilidad y especificidad

La sensibilidad y la especificidad de un test diagnóstico binario se pueden estimar a partir de una única muestra (muestreo transversal) y a partir de dos muestras independientes de individuos enfermos y de individuos no enfermos (muestreo caso-control).

1.2.1.1. Estimación bajo un muestreo transversal

Cuando el test diagnóstico y el gold estándar se aplican a todos los individuos de una muestra aleatoria de tamaño n (estudio transversal) se obtiene la Tabla 1.3. Los datos (s_1, s_0, r_1, r_0) de esta Tabla son la realización de una distribución multinomial cuyas probabilidades son las que se muestran en la Tabla 1.1.

Tabla 1.3. Frecuencias observadas al aplicar un test binario a una muestra de individuos.

	$T = 1$	$T = 0$	Total
$D = 1$	s_1	s_0	s
$D = 0$	r_1	r_0	r
Total	$s_1 + r_1$	$s_0 + r_0$	n

Condicionando en los totales de las filas se obtiene que $s_1 \rightarrow B(s, Se)$ y $r_0 \rightarrow B(r, Sp)$ y los estimadores máximo verosímiles de la sensibilidad y de la especificidad son los estimadores de proporciones binomiales, esto es

$$\hat{Se} = s_1/s \quad \text{y} \quad \hat{Sp} = r_0/r, \quad (1.20)$$

y las estimaciones de sus varianzas son

$$\widehat{Var}(\hat{Se}) = \frac{\hat{Se}(1-\hat{Se})}{s} \quad \text{y} \quad \widehat{Var}(\hat{Sp}) = \frac{\hat{Sp}(1-\hat{Sp})}{r}, \quad (1.21)$$

que son las varianzas estimadas de proporciones binomiales. Por consiguiente, los intervalos de confianza para la sensibilidad y la especificidad son los mismos intervalos de confianza para proporciones binomiales. A continuación se presentan algunos de los intervalos de confianza más comunes.

a. Intervalo de confianza exacto de Clopper-Pearson

El intervalo de confianza exacto para la sensibilidad es el intervalo (Se_1, Se_2) tal que

$P(\hat{Se}_1 \leq Se \leq \hat{Se}_2) = 1 - \alpha$ donde, por definición

$$\hat{Se}_1 \text{ es el valor de } Se \text{ tal que } \sum_{h=s_1}^s P(B = h | Se) \leq \frac{\alpha}{2} \quad (1.22)$$

$$\hat{Se}_2 \text{ es el valor de } Se \text{ tal que } \sum_{h=0}^{s_1} P(B = h | Se) \leq \frac{\alpha}{2}.$$

Aunque la solución de este problema es iterativa, se puede obtener el intervalo considerando la siguiente propiedad

$$P(B \leq x) = 1 - P\left(F(2(x+1); 2(n-x)) < \frac{n-x}{x+1} \frac{p}{1-p}\right). \quad (1.23)$$

De esta forma para Se_2 se tiene que

$$\frac{\alpha}{2} = P(B \leq s_1) = 1 - P\left(F(2(s_1+1), 2s_0) < \frac{s_0}{s_1+1} \frac{Se_2}{1-Se_2}\right), \quad (1.24)$$

y por tanto

$$P\left(F(2(s_1+1), 2s_0) < \frac{s_0}{s_1+1} \frac{Se_2}{1-Se_2}\right) = 1 - \frac{\alpha}{2}$$

$$F_{1-\frac{\alpha}{2}}(2(s_1+1), 2s_0) = F_2 = \frac{s_0}{s_1+1} \frac{Se_2}{1-Se_2} \quad (1.25)$$

$$F_2(s_1+1)(1-Se_2) = s_0 Se_2$$

$$F_2(s_1+1) = s_0 Se_2 + F_2(s_1+1) Se_2,$$

y despejando Se_2 de la última ecuación de (1.25) se obtiene que

$$\hat{Se}_2 = \frac{F_2(s_1+1)}{s_0 + F_2(s_1+1)}. \quad (1.26)$$

Para el cálculo de Se_1 se procede de forma similar,

$$\begin{aligned} \frac{\alpha}{2} &= P(B \geq s_1) = 1 - P(B < s_1) = 1 - P(B \leq s_1 - 1) = \\ &P\left(F(2s_1, 2(s_0 + 1)) < \frac{s_0 + 1}{s_1} \frac{Se_1}{1 - Se_1}\right), \end{aligned} \quad (1.27)$$

por lo que

$$F_{\frac{\alpha}{2}}(2s_1, 2(s_0 + 1)) = \frac{s_0 + 1}{s_1} \frac{Se_1}{1 - Se_1}, \quad (1.28)$$

y como

$$F_{\frac{\alpha}{2}}^{-1}(2(s_0 + 1), 2s_1) = F_1^{-1} = \frac{s_0 + 1}{s_1} \frac{Se_1}{1 - Se_1} \Rightarrow F_1(s_0 + 1) Se_1 = s_1(1 - Se_1), \quad (1.29)$$

despejando Se_1 de la última expresión de (1.29) se obtiene que

$$\hat{Se}_1 = \frac{s_1}{s_1 + F_1(s_0 + 1)}. \quad (1.30)$$

Finalmente, el intervalo de confianza exacto para la sensibilidad es

$$\left(\frac{s_1}{s_1 + F_1(s_0 + 1)}, \frac{F_2(s_1 + 1)}{s_0 + F_2(s_1 + 1)} \right). \quad (1.31)$$

De forma similar se deduce el intervalo de confianza exacto para la especificidad, siendo su expresión

$$\left(\frac{r_0}{r_0 + F_1(r_1 + 1)}, \frac{F_2(r_0 + 1)}{r_1 + F_2(r_0 + 1)} \right). \quad (1.32)$$

b. Intervalo de confianza cuadrático de Fleiss

El intervalo de confianza cuadrático para una proporción binomial se basa en la normalidad asintótica del estimador de la proporción. Para la sensibilidad del test diagnóstico se verifica que

$$\hat{Se} = \frac{s_1}{s} \rightarrow N\left(Se, \sqrt{Se(1 - Se)/s}\right), \quad (1.33)$$

por lo que un intervalo de confianza para la sensibilidad es

$$\frac{s_1}{s} \pm z_{1-\alpha/2} \sqrt{Se(1-Se)/s}, \quad (1.34)$$

donde $z_{1-\alpha/2}$ es el $100(1-\alpha/2)$ -ésimo percentil de la distribución normal estándar. Este intervalo de confianza no es útil ya que depende de Se . Igualando Se a cada uno de los extremos de este intervalo, añadiendo una corrección por continuidad y resolviendo las ecuaciones de segundo grado en Se , esto es,

$$\begin{aligned} Se - \frac{s_1}{s} - z_{1-\alpha/2} \sqrt{Se(1-Se)/s} - \frac{1}{2s} &= 0 \\ Se - \frac{s_1}{s} + z_{1-\alpha/2} \sqrt{Se(1-Se)/s} + \frac{1}{2s} &= 0, \end{aligned} \quad (1.35)$$

se obtiene el intervalo cuadrático a nivel $100(1-\alpha)\%$ para la sensibilidad, dado por la expresión

$$Se \in \frac{1}{s + z_{1-\alpha/2}^2} \left((s_1 \pm 0.5) + \frac{z_{1-\alpha/2}^2}{2} \pm z_{1-\alpha/2} \sqrt{\frac{z_{1-\alpha/2}^2}{4} + \frac{(s_1 \pm 0.5)(s_0 \pm 0.5)}{s}} \right). \quad (1.36)$$

Este intervalo de confianza es válido cuando s_1 y s_0 son mayores que 5. De forma similar se deduce el intervalo de confianza cuadrático para la especificidad, siendo su expresión

$$Sp \in \frac{1}{r + z_{1-\alpha/2}^2} \left((r_0 \pm 0.5) + \frac{z_{1-\alpha/2}^2}{2} \pm z_{1-\alpha/2} \sqrt{\frac{z_{1-\alpha/2}^2}{4} + \frac{(r_0 \pm 0.5)(r_1 \pm 0.5)}{r}} \right), \quad (1.37)$$

siendo válido cuando r_1 y r_0 son mayores que 5.

c. Intervalo de confianza de Wald con corrección por continuidad

Sustituyendo en la expresión (1.34) Se por su estimador máximo verosímil y añadiendo una corrección por continuidad, se obtiene que el intervalo de confianza de Wald con corrección por continuidad a nivel $100(1-\alpha)\%$ para la sensibilidad es

$$Se \in \hat{Se} \pm \left\{ z_{1-\alpha/2} \sqrt{\frac{\hat{Se}(1-\hat{Se})}{s}} + \frac{1}{2s} \right\}, \quad (1.38)$$

y es un intervalo de confianza válido cuando s_1 y s_0 son mayores que 20. De forma similar se obtiene el intervalo de confianza para la especificidad,

$$Sp \in \hat{Sp} \pm \left\{ z_{1-\alpha/2} \sqrt{\frac{\hat{Sp}(1-\hat{Sp})}{r}} + \frac{1}{2r} \right\}, \quad (1.39)$$

que es válido cuando r_1 y r_0 son mayores que 20.

d. Intervalo de confianza de Agresti y Coull

Agresti y Coull (1998) han propuesto un intervalo de confianza llamado intervalo de confianza score que no utiliza el cálculo estándar para la proporción binomial. De esta forma evitan inconvenientes típicos de los intervalos asintóticos, ya sea por la deficiente cobertura que éstos presentan o por dar resultados fuera de los límites de cualquier probabilidad. Así, el intervalo de confianza score para la sensibilidad viene dado por la expresión

$$Se \in \frac{\hat{Se} + \frac{z_{1-\alpha/2}^2}{2s} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{Se}(1-\hat{Se}) + \frac{z_{1-\alpha/2}^2}{4s}}{s}}}{1 + \frac{z_{1-\alpha/2}^2}{s}}, \quad (1.40)$$

y para la especificidad

$$Sp \in \frac{\hat{Sp} + \frac{z_{1-\alpha/2}^2}{2r} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{Sp}(1-\hat{Sp}) + \frac{z_{1-\alpha/2}^2}{4r}}{r}}}{1 + \frac{z_{1-\alpha/2}^2}{r}}. \quad (1.41)$$

e. Intervalo de confianza logit de Rubin y Schenker

Rubin y Schenker (1987) han propuesto un intervalo de confianza para una proporción binomial utilizando la transformación logit y argumentos bayesianos con una distribución a priori de Jeffreys. Adaptando este intervalo a los datos de la Tabla 1.3, el intervalo logit a nivel $100(1-\alpha)\%$ para la sensibilidad es

$$\text{logit}^{-1} \left\{ \text{logit} \left(\frac{s_1 + 0.5}{s + 1} \right) \pm \frac{z_{1-\alpha/2}}{\sqrt{(s+1) \left(\frac{s_1 + 0.5}{s+1} \right) \left(1 - \frac{s_1 + 0.5}{s+1} \right)}} \right\}, \quad (1.42)$$

y para la especificidad

$$\text{logit}^{-1} \left\{ \text{logit} \left(\frac{r_0 + 0.5}{r + 1} \right) \pm \frac{z_{1-\alpha/2}}{\sqrt{(r+1) \left(\frac{r_0 + 0.5}{r+1} \right) \left(1 - \frac{r_0 + 0.5}{r+1} \right)}} \right\}. \quad (1.43)$$

1.2.1.2. Estimación bajo un muestreo caso-control

La sensibilidad y la especificidad dependen de la habilidad intrínseca del test diagnóstico para distinguir entre enfermos y no enfermos, no dependiendo de la prevalencia de la enfermedad, por lo que estos parámetros se pueden estimar cuando el muestro es retrospectivo o caso-control. En esta situación el test diagnóstico se aplica a todos los individuos de dos muestras, una muestra de n_1 individuos enfermos y otra muestra de n_2 individuos no enfermos, dando lugar a la Tabla 1.4. En esta situación, todo lo comentado en la Sección 1.2.1.1 es válido, siendo necesario sustituir s por n_1 para la estimación de la sensibilidad y r por n_2 para la estimación de la especificidad.

Tabla 1.4. Frecuencias observadas al aplicar un test a dos muestras, una de individuos enfermos y otra de individuos sanos.

	$T = 1$	$T = 0$	Total
$D = 1$	s_1	s_0	n_1
$D = 0$	r_1	r_0	n_2
Total	$s_1 + r_1$	$s_0 + r_0$	n

1.2.2. Estimación de las razones de verosimilitud

Las razones de verosimilitud positiva y negativa dependen solamente de la sensibilidad y especificidad del test diagnóstico, por lo que su estimación se puede realizar bajo un muestreo transversal o un muestreo caso-control.

Cuando el test diagnóstico y el gold estándar se aplican a todos los individuos de una muestra aleatoria de tamaño n (Tabla 1.3), condicionando en los totales de las filas el estimador máximo verosímil de la razón de verosimilitud positiva es

$$\widehat{LR}^+ = \frac{\hat{Se}}{1 - \hat{Sp}} = \frac{s_1/s}{r_1/r} = \frac{s_1 r}{r_1 s}, \quad (1.44)$$

y el de la razón de verosimilitud negativa es

$$\widehat{LR}^- = \frac{1 - \hat{Se}}{\hat{Sp}} = \frac{s_0/s}{r_0/r} = \frac{s_0 r}{r_0 s}, \quad (1.45)$$

y aplicando el método delta (Serfling, 1980), sus varianzas estimadas son

$$\widehat{Var}(\widehat{LR}^+) = \frac{\hat{Se}(1 - \hat{Se})}{s(1 - \hat{Sp})^2} + \frac{\hat{Se}^2 \hat{Sp}(1 - \hat{Sp})}{r(1 - \hat{Sp})^4} \quad (1.46)$$

y

$$\widehat{Var}(\widehat{LR}^-) = \frac{\hat{Se}(1 - \hat{Se})}{s\hat{Sp}^2} + \frac{(1 - \hat{Se})^2 \hat{Sp}(1 - \hat{Sp})}{r\hat{Sp}^4}, \quad (1.47)$$

en cuyos cálculos se ha tenido en cuenta que $Cov(Se, Sp) = 0$ ya que la sensibilidad y la especificidad son parámetros independientes entre sí.

Una vez obtenidos los estimadores y sus varianzas, la obtención de un intervalo de confianza para las razones de verosimilitud es un resultado menos directo que para la sensibilidad y la especificidad, ya que las razones de verosimilitud son el cociente de dos probabilidades. Basándose en la transformación logarítmica de las razones de verosimilitud, el intervalo de confianza a nivel $100(1-\alpha)\%$ para las razones de verosimilitud es (Simel et al., 1991)

$$LR^+ \in \exp \left\{ \ln \left(\frac{\hat{Se}}{1-\hat{Sp}} \right) \pm z_{1-\alpha/2} \sqrt{\frac{1-\hat{Se}}{s\hat{Se}} + \frac{\hat{Sp}}{r(1-\hat{Sp})}} \right\} \quad (1.48)$$

y

$$LR^- \in \exp \left\{ \ln \left(\frac{1-\hat{Se}}{\hat{Sp}} \right) \pm z_{1-\alpha/2} \sqrt{\frac{\hat{Se}}{s(1-\hat{Se})} + \frac{1-\hat{Sp}}{r\hat{Sp}}} \right\}. \quad (1.49)$$

Estos intervalos de confianza, basados en la transformación logarítmica de las razones de verosimilitud, tienen un mejor comportamiento que los intervalos de confianza sin transformar $\left(\widehat{LR} \pm z_{1-\alpha/2} \sqrt{\widehat{Var}(\widehat{LR})} \right)$ debido a que la distribución de las razones de verosimilitud (positiva y negativa) es asimétrica (Simel et al., 1991; Roldán Nofuentes y Luna del Castillo, 2007) y se simetriza con las transformaciones referidas.

Gart y Nam (1988) han estudiado un intervalo de confianza para el cociente de dos proporciones independientes. Como las razones de verosimilitud son el cociente de dos proporciones binomiales independientes, aplicando los resultados de Gart y Nam (1988), el intervalo de confianza a nivel $100(1-\alpha)\%$ para la razón de verosimilitud positiva se obtiene resolviendo la ecuación

$$\frac{\left\{ \hat{Se} - LR^+ (1 - \hat{Sp}) \right\}^2}{\frac{\hat{Se}(1 - \hat{Se})}{s} + \frac{(LR^+)^2 (1 - \hat{Sp}) \hat{Sp}}{r}} = z_{1-\alpha/2}^2, \quad (1.50)$$

y para la razón de verosimilitud negativa

$$\frac{\{(1-\hat{S}e) - LR^- \hat{S}p\}^2}{\frac{\hat{S}e(1-\hat{S}e)}{s} + \frac{(LR^-)^2(1-\hat{S}p)\hat{S}p}{r}} = z_{1-\alpha/2}^2. \quad (1.51)$$

Cuando el diseño muestral es caso-control (Tabla 1.4), todo lo comentado anteriormente para las razones de verosimilitud es válido sustituyendo s por n_1 y r por n_2 .

1.2.3. Estimación del índice de Youden

De forma similar a como se han obtenido los estimadores anteriores, bajo un muestreo transversal, condicionando en los totales de las filas de la Tabla 1.3, el estimador máximo verosímil del índice de Youden es

$$\hat{I} = \frac{s_1 r_0 - s_0 r_1}{sr} \quad (1.52)$$

y su varianza estimada es

$$\widehat{Var}(\hat{I}) = \frac{\hat{S}e(1-\hat{S}e)}{s} + \frac{\hat{S}p(1-\hat{S}p)}{r}, \quad (1.53)$$

ya que la $Cov(S_e, S_p) = 0$. Finalmente, el intervalo de confianza a nivel $100(1-\alpha)\%$ para el índice de Youden viene dado por

$$I \in \hat{I} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{S}e(1-\hat{S}e)}{s} + \frac{\hat{S}p(1-\hat{S}p)}{r}}. \quad (1.54)$$

Cuando el muestreo es caso-control, las ecuaciones anteriores son válidas sustituyendo s por n_1 y r por n_2 .

1.2.4. Estimación de los valores predictivos

Los valores predictivos positivo y negativo dependen de la sensibilidad, de la especificidad del test diagnóstico y de la prevalencia de la enfermedad. Cuando el muestreo es transversal, los valores predictivos se pueden estimar sin ningún problema ya que la sensibilidad, la especificidad y la prevalencia se pueden estimar en este tipo de estudios; pero cuando el muestreo es caso-control, para poder estimar los valores predictivos es necesario conocer una estimación de la prevalencia de la enfermedad, ya que en un muestreo caso-control ésta no se puede estimar (los valores n_1 y n_2 son fijados por el investigador y $n_1/(n_1 + n_2)$ no es un estimador de la prevalencia).

1.2.4.1. Estimación bajo un muestreo transversal

Cuando se evalúa el rendimiento de un test diagnóstico con respecto a un gold estándar mediante un estudio transversal, sustituyendo en las ecuaciones (1.11) y (1.12) cada parámetro por su estimador máximo verosímil, esto es Se por s_1/s , Sp por r_0/r y p por s/n , los estimadores máximo verosímiles de los valores predictivos positivo y negativo son

$$\widehat{vpp} = \frac{s_1}{s_1 + r_1} \quad (1.55)$$

y

$$\widehat{vpn} = \frac{r_0}{s_0 + r_0}, \quad (1.56)$$

respectivamente. Condicionando en los totales de las columnas, las estimaciones de sus varianzas son

$$\widehat{Var}(\widehat{vpp}) = \frac{\widehat{vpp}(1 - \widehat{vpp})}{s_1 + r_1} \quad \text{y} \quad \widehat{Var}(\widehat{vpn}) = \frac{\widehat{vpn}(1 - \widehat{vpn})}{s_0 + r_0}, \quad (1.57)$$

que son, al igual que para la sensibilidad y la especificidad, las varianzas estimadas de proporciones binominales. Por consiguiente, en un estudio transversal los intervalos de confianza que se pueden calcular para los valores predictivos son los mismos que para la sensibilidad y la especificidad. Condicionando en los totales de las columnas, los

intervalos de confianza más comunes para los valores predictivos positivo y negativo son los se muestran a continuación.

a. Intervalo de confianza exacto de Clopper-Pearson

El intervalo de confianza exacto a nivel $100(1-\alpha)\%$ para el valor predictivo positivo es

$$\left(\frac{s_1}{s_1 + F_1(r_1 + 1)}, \frac{F_2(s_1 + 1)}{r_1 + F_2(s_1 + 1)} \right), \quad (1.58)$$

y para el valor predictivo negativo

$$\left(\frac{r_0}{r_0 + F_1(s_0 + 1)}, \frac{F_2(r_0 + 1)}{s_0 + F_2(r_0 + 1)} \right). \quad (1.59)$$

b. Intervalo de confianza cuadrático de Fleiss

El intervalo de confianza cuadrático para el valor predictivo positivo es

$$vpp \in \frac{1}{s_1 + r_1 + z_{1-\alpha/2}^2} \left\{ (s_1 \pm 0.5) + \frac{z_{1-\alpha/2}^2}{2} \pm z_{1-\alpha/2} \sqrt{\frac{z_{1-\alpha/2}^2}{4} + \frac{(s_1 \pm 0.5)(r_1 \pm 0.5)}{s_1 + r_1}} \right\}, \quad (1.60)$$

y es un intervalo válido cuando s_1 y r_1 son mayores que 5. El intervalo de confianza cuadrático para el valor predictivo negativo es

$$vpn \in \frac{1}{s_0 + r_0 + z_{1-\alpha/2}^2} \left\{ (r_0 \pm 0.5) + \frac{z_{1-\alpha/2}^2}{2} \pm z_{1-\alpha/2} \sqrt{\frac{z_{1-\alpha/2}^2}{4} + \frac{(r_0 \pm 0.5)(s_0 \pm 0.5)}{s_0 + r_0}} \right\}, \quad (1.61)$$

que es válido para s_0 y r_0 mayores que 5.

c. Intervalo de confianza de Wald con corrección por continuidad

Cuando s_1 y r_1 son mayores que 20, el intervalo de confianza de Wald con corrección por continuidad para el valor predictivo positivo es

$$vpp \in \widehat{vpp} \pm \left\{ z_{1-\alpha/2} \sqrt{\frac{\widehat{vpp}(1-\widehat{vpp})}{s_1+r_1}} + \frac{1}{2(s_1+r_1)} \right\}. \quad (1.62)$$

Para el valor predictivo negativo, el intervalo de confianza es

$$vpn \in \widehat{vpn} \pm \left\{ z_{1-\alpha/2} \sqrt{\frac{\widehat{vpn}(1-\widehat{vpn})}{s_0+r_0}} + \frac{1}{2(s_0+r_0)} \right\}, \quad (1.63)$$

y que es válido para s_0 y r_0 mayores que 20.

d. Intervalo de confianza de Agresti y Coull

Los intervalos de confianza de Agresti y Coull para los valores predictivos positivo y negativo son

$$vpp \in \frac{\widehat{vpp} + \frac{z_{1-\alpha/2}^2}{2(s_1+r_1)} \pm z_{1-\alpha/2} \sqrt{\frac{\widehat{vpp}(1-\widehat{vpp}) + \frac{z_{1-\alpha/2}^2}{4(s_1+r_1)}}{s_1+r_1}}}{1 + \frac{z_{1-\alpha/2}^2}{s_1+r_1}} \quad (1.64)$$

y

$$vpn \in \frac{\widehat{vpn} + \frac{z_{1-\alpha/2}^2}{2(s_0+r_0)} \pm z_{1-\alpha/2} \sqrt{\frac{\widehat{vpn}(1-\widehat{vpn}) + \frac{z_{1-\alpha/2}^2}{4(s_0+r_0)}}{s_0+r_0}}}{1 + \frac{z_{1-\alpha/2}^2}{s_0+r_0}}, \quad (1.65)$$

respectivamente.

e. Intervalo de confianza logit de Rubin y Schenker

El intervalo logit a nivel $100(1-\alpha)\%$ para el valor predictivo positivo es

$$\text{logit}^{-1} \left\{ \text{logit} \left(\frac{s_1 + 0.5}{s_1 + r_1 + 1} \right) \pm \frac{z_{1-\alpha/2}}{\sqrt{(s_1 + r_1 + 1) \left(\frac{s_1 + 0.5}{s_1 + r_1 + 1} \right) \left(1 - \frac{s_1 + 0.5}{s_1 + r_1 + 1} \right)}} \right\}, \quad (1.66)$$

y para el valor predictivo negativo

$$\text{logit}^{-1} \left\{ \text{logit} \left(\frac{r_0 + 0.5}{s_0 + r_0 + 1} \right) \pm \frac{z_{1-\alpha/2}}{\sqrt{(s_0 + r_0 + 1) \left(\frac{r_0 + 0.5}{s_0 + r_0 + 1} \right) \left(1 - \frac{r_0 + 0.5}{s_0 + r_0 + 1} \right)}} \right\}. \quad (1.67)$$

1.2.4.2. Estimación bajo un muestreo caso-control

Cuando el muestreo es caso-control, los totales de las filas (número total de individuos enfermos y número de total de individuos no enfermos) son fijados por el investigador, por lo que a partir de este diseño se pueden estimar la sensibilidad y la especificidad del test diagnóstico (Sección 1.2.1.2) pero no se puede estimar la prevalencia de la enfermedad, y por consiguiente tampoco se pueden estimar los valores predictivos directamente a partir del propio estudio. Sin embargo, si se conoce una estimación de la prevalencia de la enfermedad a partir de otro estudio diferente, bajo un diseño caso-control (Tabla 1.4) los estimadores máximo verosímiles de los valores predictivos son

$$\widehat{vpp} = \frac{\hat{p} \frac{s_1}{n_1}}{\hat{p} \frac{s_1}{n_1} + (1 - \hat{p}) \frac{r_1}{n_2}} \quad (1.68)$$

y

$$\widehat{vpn} = \frac{(1-\hat{p})\frac{r_0}{n_2}}{\hat{p}\frac{s_0}{n_1} + (1-\hat{p})\frac{r_0}{n_2}}, \quad (1.69)$$

respectivamente, y siendo \hat{p} el estimador de la prevalencia de la enfermedad. En esta situación, la obtención de los intervalos de confianza de los valores predictivos no es tan inmediata como en el caso de un muestreo transversal. Teniendo en cuenta que las ecuaciones (1.68) y (1.69) se pueden escribir como

$$\widehat{vpp} = \frac{1}{1 + \frac{(1-\hat{p})(1-\hat{S}p)}{\hat{p}\hat{S}e}} = \frac{1}{1 + \frac{1-\hat{p}}{\hat{p}\widehat{LR}^+}} = \left(1 + \frac{1-\hat{p}}{\hat{p}\widehat{LR}^+}\right)^{-1} \quad (1.70)$$

y

$$\widehat{vpn} = \frac{1}{1 + \frac{\hat{p}(1-\hat{S}e)}{(1-\hat{p})\hat{S}p}} = \frac{1}{1 + \frac{\hat{p}}{1-\hat{p}}\widehat{LR}^-} = \left(1 + \frac{\hat{p}}{1-\hat{p}}\widehat{LR}^-\right)^{-1}, \quad (1.71)$$

siendo \widehat{LR}^+ y \widehat{LR}^- los estimadores máximo verosímiles de las razones de verosimilitud, la obtención de los intervalos de confianza de los valores predictivos se basa en la obtención de los intervalos de confianza de las razones de verosimilitud. Así, una vez conocido un estimador de la prevalencia de la enfermedad y calculado el intervalo de confianza para la razón de verosimilitud positiva, sustituyendo cada límite de este

intervalo en la expresión $\left(1 + \frac{1-\hat{p}}{\hat{p}\widehat{LR}^+}\right)^{-1}$ se calcula el intervalo de confianza para el valor

predictivo positivo. Para el intervalo de confianza del valor predictivo negativo, los límites del intervalo de confianza de la razón de verosimilitud negativa se sustituyen en

la expresión $\left(1 + \frac{\hat{p}}{1-\hat{p}}\widehat{LR}^-\right)^{-1}$.

1.2.5. Estimación del coeficiente kappa ponderado (Aportación)

El coeficiente kappa ponderado de un test binario es una medida del acuerdo clasificatorio entre el test diagnóstico y el gold estándar, y depende de la sensibilidad y especificidad del test diagnóstico, de la prevalencia de la enfermedad y de las pérdidas relativas entre los falsos positivos y los falsos negativos. Roldán Nofuentes, Luna del Castillo y Montero Alonso (2009b) han estudiado varios intervalos de confianza para el coeficiente kappa ponderado bajo un estudio transversal.

1.2.5.1. Intervalos de confianza para el coeficiente kappa ponderado

Sea un test diagnóstico binario cuyo rendimiento es evaluado con respecto a un gold estándar bajo un muestreo transversal, dando lugar a la Tabla 1.3. El coeficiente kappa ponderado del test diagnóstico, dado por la ecuación (1.16), es

$$\kappa(c) = \frac{pq(Se + Sp - 1)}{p\{p(1 - Se) + qSp\}c + q\{pSe + q(1 - Sp)\}(1 - c)},$$

con $q = 1 - p$ y $0 < c < 1$. Condicionando en los totales de las filas y sustituyendo en la ecuación anterior cada parámetro por su estimador máximo verosímil, el estimador máximo verosímil del coeficiente kappa ponderado es

$$\hat{\kappa}(c) = \frac{s_1 r + r_0 s - sr}{n_0 s c + n_1 r (1 - c)}, \quad (1.72)$$

con $0 < c < 1$. Para poder obtener una estimación del coeficiente de la kappa ponderado es necesario imponer que todas las frecuencias marginales $(s, r, s_1 + r_1, s_0 + r_0)$ de la Tabla 1.3 sean mayores que cero. Por otra parte, los datos (s_1, s_0, r_1, r_0) de la Tabla 1.3 son la realización de una distribución multinomial con probabilidades $\{pSe, p(1 - Se), (1 - p)(1 - Sp), (1 - p)Sp\}$. En esta tabla de frecuencias el único valor fijo es n , así que condicionando en los totales de las filas, s_1 y r_0 siguen una distribución binomial con parámetros s y Se y r y Sp , respectivamente. Aplicando las propiedades de la distribución binomial, los datos (s_1, s_0) y (r_1, r_0) pueden ser tratados

como muestras independientes y por tanto $Cov(\hat{Se}, \hat{Sp}) = 0$. Asimismo, la sensibilidad y la especificidad del test diagnóstico son parámetros independientes de la prevalencia de la enfermedad, por lo que $Cov(Se, p) = Cov(Sp, p) = 0$. Por tanto, condicionando en los totales de las filas, la sensibilidad, la especificidad y la prevalencia de la enfermedad son proporciones binomiales, siendo sus estimadores máximo verosímiles $\hat{Se} = \frac{s_1}{s}$,

$$\hat{Sp} = \frac{r_0}{r} \quad \text{y} \quad \hat{p} = \frac{s}{n} \quad \text{y} \quad \text{sus varianzas estimadas} \quad \widehat{Var}(\hat{Se}) = \frac{\hat{Se}(1-\hat{Se})}{s},$$

$$\widehat{Var}(\hat{Sp}) = \frac{\hat{Sp}(1-\hat{Sp})}{r} \quad \text{y} \quad \widehat{Var}(\hat{p}) = \frac{\hat{p}\hat{q}}{n} \quad \text{respectivamente, con } s_1 \text{ y } r_0 \text{ mayores de cero.}$$

Aplicando el método delta (Serfling, 1980), la varianza del coeficiente kappa ponderado es

$$Var(\hat{\kappa}(c)) = \left(\frac{\partial \kappa(c)}{\partial Se}\right)^2 Var(\hat{Se}) + \left(\frac{\partial \kappa(c)}{\partial Sp}\right)^2 Var(\hat{Sp}) + \left(\frac{\partial \kappa(c)}{\partial p}\right)^2 Var(\hat{p}),$$

por lo que sustituyendo en la ecuación anterior cada parámetro por su estimador y realizando las operaciones algebraicas, la varianza estimada de $\hat{\kappa}(c)$ es

$$\widehat{Var}(\hat{\kappa}(c)) = \frac{nr}{s \left[n^2 (1-c)r_1 + n(cr_0 - 2(1-c)r_1)s_0 + n\{r_0 - (1-c)r_1\}s_1 + s(s_0r_1 - s_1r_0) \right]^4} \times$$

$$\left\{ (s_0r_1 - s_1r_0)^2 \left[2(1-c)r_1ns - (1-c)r_1n^2 + s(c(s_0r_0 + 2s_0r_1 + s_1r_1) - r_1s) \right]^2 + \right. \quad (1.73)$$

$$\left. s_1s_0nr^3 \left[(1-c)r_1n + s(cr - r_1) \right]^2 + r_1r_0nsr^2 \left[s_1r + c(s^2 - s_1n) \right]^2 \right\}.$$

A continuación se estudian cuatro intervalos de confianza para el coeficiente kappa ponderado de un test diagnóstico binario cuando el test diagnóstico y el gold estándar se aplican a todos los pacientes de una muestra aleatoria, de los cuales tres intervalos están basados en la normalidad asintótica del estimador del coeficiente kappa ponderado (intervalo de Wald sin y con corrección por continuidad e intervalo logit) y un cuarto intervalo basado en el método Bootstrap.

a. Intervalo de confianza tipo Wald

Asumiendo la normalidad asintótica de $\hat{\kappa}(c)$, el intervalo de confianza de Wald a nivel $100(1-\alpha)\%$ para el coeficiente kappa ponderado es

$$\hat{\kappa}(c) \pm z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{\kappa}(c))},$$

donde $z_{1-\alpha/2}$ es el $100(1-\alpha/2)$ -ésimo percentil de la distribución normal estándar y $\widehat{\text{Var}}(\hat{\kappa}(c))$ viene dada por la ecuación (1.73). Este intervalo de confianza también se puede obtener con una corrección por continuidad, siendo el intervalo

$$\hat{\kappa}(c) \pm \left(z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{\kappa}(c)) + \frac{1}{2n}} \right). \quad (1.74)$$

b. Intervalo de confianza logit

En el análisis de tablas 2×2 es frecuente que el estudio de un parámetro no se realice directamente sobre el propio parámetro sino sobre transformaciones de este parámetro, como es la transformación logit. Por ejemplo, para una proporción binomial es posible obtener un intervalo de confianza basado en la transformación logit de la proporción (Rubin y Schenker, 1987). Para el coeficiente kappa ponderado, su análisis se limita a valores entre cero y uno, que son los mismos valores de una proporción binomial, por lo que la transformación logit es válida para estudiar el coeficiente kappa ponderado.

Asumiendo la normalidad asintótica de $\hat{\kappa}(c)$, la transformación logit de $\hat{\kappa}(c)$ es $\log\{\hat{\kappa}(c)/(1-\hat{\kappa}(c))\}$, que sigue una distribución normal de media $\log\{\kappa(c)/(1-\kappa(c))\}$. De esta forma el intervalo de confianza a nivel $100(1-\alpha)\%$ para $\text{logit}(\kappa(c))$ es

$$\text{logit}(\hat{\kappa}(c)) \pm z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\text{logit}(\hat{\kappa}(c)))}, \quad (1.75)$$

donde el estimador de la varianza del logit de $\hat{\kappa}(c)$, obtenido aplicando el método delta (Serfling, 1980), es

$$\widehat{Var}(\text{logit}(\hat{\kappa}(c))) = \frac{1}{[r_1s - (1-c)r_1n - c(s_0r_0 + 2s_0r_1 + s_1r_1)]^2} \times \left\{ \frac{s_1s_0r^2 [(1-c)r_1n + (cr - r_1)s]^2 + r_1r_0sr [s_1n - s_1s + c(s^2 - s_1n)]^2}{s(s_0r_1 - s_1r_0)^2} + \frac{[2(1-c)r_1ns - (1-c)r_1n^2 + s(c(s_0r_0 + 2s_0r_1 + s_1r_1) - r_1s)]^2}{nsr} \right\}. \quad (1.76)$$

Finalmente, el intervalo de confianza logit para el coeficiente kappa ponderado viene dado por la expresión

$$\left(\frac{\exp\{\text{logit}(\hat{\kappa}(c)) - z_{1-\alpha/2}\hat{\sigma}_{\text{logit}}\}}{1 + \exp\{\text{logit}(\hat{\kappa}(c)) - z_{1-\alpha/2}\hat{\sigma}_{\text{logit}}\}}, \frac{\exp\{\text{logit}(\hat{\kappa}(c)) + z_{1-\alpha/2}\hat{\sigma}_{\text{logit}}\}}{1 + \exp\{\text{logit}(\hat{\kappa}(c)) + z_{1-\alpha/2}\hat{\sigma}_{\text{logit}}\}} \right), \quad (1.77)$$

donde $\hat{\sigma}_{\text{logit}} = \sqrt{\widehat{Var}(\text{logit}(\hat{\kappa}(c)))}$.

c. Intervalo de confianza bootstrap

El intervalo de confianza bootstrap se calcula eligiendo un número B de replicas de manera independiente y con reemplazamiento de la muestra n de la que se dispone, calculando en cada una de ellas un estimador del coeficiente kappa ponderado dado por la ecuación (1.72). Como estimador del coeficiente kappa ponderado se propone la media de las B réplicas de tales estimadores, y el intervalo de confianza se calcula empleando el intervalo de confianza corregido por el sesgo (Efron y Tibshirani, 1993), dado por la expresión

$$\hat{\kappa}_B(c) \pm z_{1-\alpha/2} \sqrt{\widehat{Var}_B(\hat{\kappa}_B(c))}. \quad (1.78)$$

1.2.5.2. Experimentos de simulación

Para estudiar el rendimiento, medido en términos de la cobertura y la longitud media, de los intervalos de confianza propuestos en la Sección anterior, se han realizado unos experimentos de simulación Monte Carlo. Estos experimentos han consistido en generar

10000 muestras aleatorias de distribuciones multinomiales de tamaños 100, 200, 300, 400, 500 y 1000, y cuyas probabilidades se han calculado a partir de las probabilidades de la Tabla 1.1. Estos experimentos se han diseñado a partir de la ecuación (1.19) de la siguiente forma. Como prevalencia de la enfermedad se han tomado los valores 0.10 y 0.50, como valores de $\kappa(0)$ y $\kappa(1)$ se han tomado los valores 0.1, 0.2,...,0.9, y como valores del índice de ponderación c se han tomado 0.1, 0.5 y 0.9. Una vez fijados los valores de p , $\kappa(0)$ y $\kappa(1)$, los valores de la sensibilidad y especificidad se han calculado a partir de las ecuaciones (1.17) y (1.18), y el valor de $\kappa(c)$ se ha calculado a partir de la ecuación (1.19). Por tanto, las probabilidades de las distribuciones multinomiales se han calculado a partir de valores de coeficientes kappa ponderados y sin fijar los valores de la sensibilidad y especificidad. Para cada una de las 10000 muestras aleatorias de una misma distribución multinomial se han calculado los intervalos de confianza tipo Wald (sin y con corrección por continuidad) y logit, y para cada 10000 muestras aleatorias se han calculado las coberturas y longitudes promedio de cada uno de estos tres intervalos. Con respecto al intervalo bootstrap, para cada una de las 10000 muestras se han generado a su vez 1000 muestras aleatorias bootstrap, y a partir de éstas se ha calculado el intervalo bootstrap, y posteriormente se han calculado las coberturas y las longitudes promedio. Los experimentos de simulación se han diseñado de tal forma que en todas las muestras los totales marginales $(s, r, s_1 + r_1, s_0 + r_0)$ son todos mayores que cero (ya que en caso contrario no se puede estimar el coeficiente kappa ponderado). Para todo el estudio de simulación el nivel de confianza se ha fijado en el 95%.

En las Tablas 1.5, 1.6, 1.7 y 1.8 se muestran los resultados obtenidos para $\kappa(c)$ igual a 0.2, 0.4, 0.6 y 0.8 respectivamente. De los resultados de los experimentos de simulación se obtienen en términos generales las siguientes conclusiones. Para muestras de al menos 200 individuos la cobertura del intervalo logit tiene un mejor comportamiento que el resto de intervalos: una vez alcanzada la cobertura del 95%, la cobertura del intervalo tiene menores fluctuaciones en torno a este valor que el resto de intervalos. Con respecto al rendimiento de los intervalos con muestras pequeñas (tamaño 100), en general el rendimiento de los intervalos depende del valor del coeficiente kappa ponderado: para valores muy bajos ($\kappa(c)=0.2$) el intervalo logit tiene una buena cobertura, aunque en algunos casos su longitud promedio es un poco

mayor que la del resto de intervalos; para valores de $\kappa(c)$ iguales a 0.4 y 0.6, el intervalo tipo Wald sin corrección por continuidad y el intervalo bootstrap tienen un buen rendimiento; y para valores muy altos del coeficiente kappa ponderado ($\kappa(c) = 0.8$), el intervalo bootstrap presenta una mejor cobertura que el resto de intervalos, aunque su longitud promedio es un poco mayor que la del resto de intervalos. En cuanto al intervalo de confianza tipo Wald con corrección por continuidad, la corrección por continuidad no tiene un importante efecto sobre el rendimiento del intervalo, siendo el rendimiento de este intervalo muy similar al intervalo de Wald sin corrección. Por tanto, en términos generales, y para valores de coeficientes kappa ponderados comunes en la práctica clínica, cuando las muestras son pequeñas ($n = 100$) el intervalo tipo Wald sin corrección por continuidad y el intervalo bootstrap tienen un buen rendimiento; mientras que para muestras mayores ($n \geq 200$) el intervalo logit presenta un mejor rendimiento que el resto de intervalos.

Tabla 1.5. Cobertura y amplitud promedio de los I.C. para el coeficiente kappa ponderado ($\kappa(c) = 0.2$).

$p = 0.10 \quad \kappa(0) = 0.2 \quad \kappa(1) = 0.2 \quad c = 0.1 \quad Se = 0.28 \quad Sp = 0.92$								
n	Wald		Waldcc		Logit		Bootstrap	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.950	0.559	0.954	0.569	0.959	0.615	0.941	0.532
200	0.938	0.396	0.942	0.401	0.955	0.415	0.924	0.386
300	0.937	0.324	0.939	0.327	0.958	0.345	0.926	0.320
400	0.936	0.282	0.940	0.284	0.967	0.292	0.932	0.280
500	0.935	0.251	0.935	0.253	0.969	0.257	0.930	0.251
1000	0.946	0.179	0.947	0.180	0.958	0.180	0.946	0.179
$p = 0.10 \quad \kappa(0) = 0.2 \quad \kappa(1) = 0.2 \quad c = 0.5 \quad Se = 0.28 \quad Sp = 0.92$								
n	Wald		Waldcc		Logit		Bootstrap	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.960	0.533	0.971	0.543	0.954	0.600	0.943	0.487
200	0.940	0.381	0.944	0.386	0.960	0.410	0.928	0.364
300	0.932	0.312	0.934	0.316	0.963	0.336	0.922	0.305
400	0.940	0.272	0.943	0.274	0.963	0.284	0.935	0.268
500	0.936	0.244	0.937	0.246	0.961	0.251	0.927	0.242
1000	0.944	0.174	0.945	0.175	0.956	0.175	0.941	0.174
$p = 0.10 \quad \kappa(0) = 0.2 \quad \kappa(1) = 0.2 \quad c = 0.9 \quad Se = 0.28 \quad Sp = 0.92$								
n	Wald		Waldcc		Logit		Bootstrap	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.960	0.560	0.964	0.570	0.960	0.615	0.961	0.517
200	0.947	0.395	0.953	0.400	0.967	0.416	0.931	0.378
300	0.943	0.325	0.945	0.328	0.965	0.343	0.932	0.319
400	0.930	0.282	0.934	0.284	0.962	0.294	0.923	0.279
500	0.945	0.251	0.946	0.253	0.963	0.257	0.938	0.250
1000	0.948	0.179	0.950	0.180	0.957	0.179	0.947	0.179
$p = 0.50 \quad \kappa(0) = 0.2 \quad \kappa(1) = 0.2 \quad c = 0.1 \quad Se = 0.60 \quad Sp = 0.60$								
n	Wald		Waldcc		Logit		Bootstrap	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.970	0.392	0.974	0.402	0.953	0.435	0.963	0.364
200	0.951	0.277	0.954	0.282	0.963	0.281	0.938	0.270
300	0.943	0.226	0.945	0.230	0.956	0.232	0.936	0.224
400	0.944	0.196	0.947	0.199	0.961	0.198	0.943	0.196
500	0.946	0.176	0.948	0.178	0.965	0.177	0.945	0.175
1000	0.947	0.124	0.950	0.125	0.954	0.124	0.945	0.124
$p = 0.50 \quad \kappa(0) = 0.2 \quad \kappa(1) = 0.2 \quad c = 0.5 \quad Se = 0.60 \quad Sp = 0.60$								
n	Wald		Waldcc		Logit		Bootstrap	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.960	0.379	0.968	0.389	0.940	0.430	0.947	0.345
200	0.944	0.270	0.955	0.275	0.952	0.275	0.935	0.262
300	0.947	0.221	0.951	0.224	0.962	0.226	0.940	0.218
400	0.942	0.192	0.944	0.194	0.964	0.194	0.941	0.190
500	0.937	0.171	0.942	0.173	0.951	0.172	0.937	0.171
1000	0.948	0.121	0.950	0.122	0.959	0.122	0.945	0.121
$p = 0.50 \quad \kappa(0) = 0.2 \quad \kappa(1) = 0.2 \quad c = 0.1 \quad Se = 0.60 \quad Sp = 0.60$								
n	Wald		Waldcc		Logit		Bootstrap	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.970	0.392	0.974	0.402	0.955	0.436	0.964	0.363
200	0.950	0.278	0.956	0.283	0.959	0.282	0.939	0.271
300	0.950	0.226	0.955	0.230	0.954	0.231	0.945	0.225
400	0.951	0.196	0.953	0.199	0.961	0.198	0.947	0.196
500	0.947	0.175	0.950	0.177	0.957	0.177	0.944	0.175
1000	0.958	0.124	0.959	0.125	0.959	0.124	0.958	0.124

Wald: Intervalo de confianza tipo Wald, Waldcc: intervalo de confianza tipo Wald con corrección por continuidad, Logit: intervalo de confianza logit, Bootstrap: intervalo de confianza Bootstrap, Cob.: cobertura, Long.: longitud media.

Tabla 1.6. Cobertura y amplitud promedio de los I.C. para el coeficiente kappa ponderado ($\kappa(c) = 0.4$).

$p = 0.10 \quad \kappa(0) = 0.4 \quad \kappa(1) = 0.4 \quad c = 0.1 \quad Se = 0.46 \quad Sp = 0.94$								
n	Wald		Waldcc		Logit		Bootstrap	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.914	0.609	0.920	0.619	0.977	0.561	0.915	0.617
200	0.933	0.437	0.935	0.442	0.967	0.415	0.934	0.443
300	0.923	0.358	0.928	0.362	0.953	0.345	0.927	0.361
400	0.936	0.311	0.939	0.313	0.950	0.302	0.936	0.313
500	0.940	0.278	0.941	0.280	0.954	0.271	0.942	0.279
1000	0.944	0.197	0.946	0.198	0.951	0.195	0.944	0.197
$p = 0.10 \quad \kappa(0) = 0.6 \quad \kappa(1) = 0.3 \quad c = 0.5 \quad Se = 0.34 \quad Sp = 0.98$								
n	Wald		Waldcc		Logit		Bootstrap	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.926	0.618	0.927	0.628	0.980	0.567	0.926	0.625
200	0.935	0.444	0.938	0.449	0.968	0.423	0.939	0.448
300	0.944	0.367	0.946	0.371	0.969	0.354	0.945	0.371
400	0.941	0.318	0.943	0.321	0.962	0.309	0.943	0.320
500	0.940	0.285	0.940	0.287	0.955	0.278	0.940	0.287
1000	0.950	0.202	0.952	0.203	0.957	0.200	0.947	0.203
$p = 0.10 \quad \kappa(0) = 0.3 \quad \kappa(1) = 0.6 \quad c = 0.9 \quad Se = 0.67 \quad Sp = 0.87$								
n	Wald		Waldcc		Logit		Bootstrap	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.937	0.479	0.942	0.489	0.977	0.455	0.933	0.479
200	0.946	0.341	0.947	0.346	0.967	0.330	0.945	0.343
300	0.946	0.279	0.948	0.283	0.960	0.273	0.946	0.280
400	0.941	0.243	0.944	0.245	0.951	0.238	0.939	0.243
500	0.946	0.217	0.948	0.219	0.951	0.214	0.948	0.217
1000	0.944	0.154	0.945	0.155	0.949	0.152	0.943	0.154
$p = 0.50 \quad \kappa(0) = 0.4 \quad \kappa(1) = 0.4 \quad c = 0.1 \quad Se = 0.70 \quad Sp = 0.70$								
n	Wald		Waldcc		Logit		Bootstrap	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.947	0.382	0.951	0.392	0.975	0.367	0.945	0.382
200	0.944	0.271	0.947	0.276	0.952	0.265	0.943	0.271
300	0.945	0.221	0.948	0.225	0.956	0.218	0.945	0.221
400	0.945	0.192	0.947	0.194	0.951	0.190	0.944	0.192
500	0.957	0.172	0.960	0.174	0.958	0.170	0.955	0.172
1000	0.955	0.122	0.959	0.123	0.957	0.121	0.956	0.122
$p = 0.50 \quad \kappa(0) = 0.3 \quad \kappa(1) = 0.6 \quad c = 0.5 \quad Se = 0.87 \quad Sp = 0.53$								
n	Wald		Waldcc		Logit		Bootstrap	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.951	0.335	0.955	0.345	0.975	0.325	0.950	0.336
200	0.942	0.239	0.947	0.244	0.956	0.234	0.943	0.239
300	0.950	0.195	0.954	0.198	0.956	0.193	0.950	0.195
400	0.944	0.169	0.948	0.171	0.949	0.167	0.942	0.169
500	0.943	0.151	0.946	0.153	0.948	0.150	0.944	0.151
1000	0.947	0.107	0.950	0.108	0.948	0.107	0.949	0.107
$p = 0.50 \quad \kappa(0) = 0.1 \quad \kappa(1) = 0.6 \quad c = 0.9 \quad Se = 0.94 \quad Sp = 0.23$								
n	Wald		Waldcc		Logit		Bootstrap	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.959	0.530	0.961	0.540	0.976	0.501	0.945	0.508
200	0.937	0.377	0.938	0.382	0.956	0.364	0.931	0.373
300	0.938	0.310	0.941	0.313	0.957	0.301	0.937	0.308
400	0.931	0.268	0.932	0.271	0.949	0.263	0.931	0.268
500	0.945	0.241	0.947	0.243	0.953	0.237	0.946	0.240
1000	0.956	0.171	0.957	0.172	0.957	0.169	0.954	0.171

Wald: Intervalo de confianza tipo Wald, Waldcc: intervalo de confianza tipo Wald con corrección por continuidad, Logit: intervalo de confianza logit, Bootstrap: intervalo de confianza Bootstrap, Cob.: cobertura, Long.: longitud media.

Tabla 1.7. Cobertura y amplitud promedio de los I.C. para el coeficiente kappa ponderado ($\kappa(c) = 0.6$).

$p = 0.10 \quad \kappa(0) = 0.6 \quad \kappa(1) = 0.6 \quad c = 0.1 \quad Se = 0.64 \quad Sp = 0.96$									
n	Wald		Waldcc		Logit		Bootstrap		Long.
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	
100	0.925	0.590	0.927	0.600	0.986	0.535	0.937	0.612	
200	0.926	0.417	0.930	0.422	0.959	0.396	0.931	0.424	
300	0.930	0.342	0.932	0.346	0.954	0.330	0.933	0.345	
400	0.944	0.298	0.945	0.301	0.961	0.290	0.945	0.301	
500	0.940	0.266	0.941	0.268	0.948	0.261	0.940	0.268	
1000	0.941	0.189	0.943	0.190	0.949	0.187	0.943	0.190	
$p = 0.10 \quad \kappa(0) = 0.6 \quad \kappa(1) = 0.6 \quad c = 0.5 \quad Se = 0.64 \quad Sp = 0.96$									
n	Wald		Waldcc		Logit		Bootstrap		Long.
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	
100	0.932	0.530	0.941	0.540	0.984	0.488	0.939	0.553	
200	0.942	0.376	0.945	0.381	0.963	0.360	0.947	0.384	
300	0.933	0.306	0.937	0.309	0.949	0.297	0.938	0.310	
400	0.952	0.265	0.955	0.268	0.964	0.259	0.955	0.268	
500	0.956	0.238	0.958	0.240	0.965	0.234	0.958	0.240	
1000	0.954	0.168	0.957	0.169	0.956	0.167	0.954	0.169	
$p = 0.10 \quad \kappa(0) = 0.6 \quad \kappa(1) = 0.6 \quad c = 0.9 \quad Se = 0.64 \quad Sp = 0.96$									
n	Wald		Waldcc		Logit		Bootstrap		Long.
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	
100	0.926	0.594	0.927	0.604	0.985	0.538	0.931	0.615	
200	0.923	0.419	0.931	0.424	0.959	0.398	0.929	0.426	
300	0.936	0.344	0.937	0.347	0.959	0.331	0.937	0.347	
400	0.947	0.298	0.948	0.300	0.963	0.290	0.948	0.300	
500	0.945	0.268	0.947	0.270	0.954	0.262	0.950	0.269	
1000	0.945	0.189	0.946	0.190	0.952	0.187	0.944	0.190	
$p = 0.50 \quad \kappa(0) = 0.6 \quad \kappa(1) = 0.6 \quad c = 0.1 \quad Se = 0.80 \quad Sp = 0.80$									
n	Wald		Waldcc		Logit		Bootstrap		Long.
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	
100	0.930	0.350	0.934	0.360	0.953	0.338	0.929	0.350	
200	0.936	0.250	0.941	0.255	0.952	0.245	0.936	0.249	
300	0.941	0.205	0.945	0.208	0.945	0.202	0.941	0.204	
400	0.955	0.178	0.958	0.180	0.957	0.176	0.953	0.177	
500	0.941	0.159	0.946	0.161	0.948	0.157	0.945	0.158	
1000	0.950	0.112	0.953	0.113	0.953	0.112	0.954	0.112	
$p = 0.50 \quad \kappa(0) = 0.6 \quad \kappa(1) = 0.6 \quad c = 0.5 \quad Se = 0.80 \quad Sp = 0.80$									
n	Wald		Waldcc		Logit		Bootstrap		Long.
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	
100	0.926	0.311	0.938	0.321	0.955	0.302	0.932	0.312	
200	0.945	0.221	0.955	0.226	0.952	0.217	0.948	0.221	
300	0.953	0.181	0.954	0.184	0.961	0.179	0.955	0.181	
400	0.955	0.157	0.957	0.159	0.960	0.155	0.956	0.157	
500	0.953	0.140	0.954	0.142	0.957	0.139	0.953	0.140	
1000	0.947	0.099	0.950	0.100	0.948	0.099	0.948	0.099	
$p = 0.50 \quad \kappa(0) = 0.6 \quad \kappa(1) = 0.6 \quad c = 0.9 \quad Se = 0.80 \quad Sp = 0.80$									
n	Wald		Waldcc		Logit		Bootstrap		Long.
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	
100	0.941	0.351	0.946	0.361	0.968	0.339	0.940	0.351	
200	0.943	0.250	0.948	0.255	0.951	0.245	0.943	0.250	
300	0.948	0.204	0.952	0.208	0.955	0.202	0.948	0.204	
400	0.947	0.177	0.951	0.180	0.953	0.176	0.946	0.177	
500	0.949	0.159	0.953	0.161	0.953	0.157	0.949	0.158	
1000	0.950	0.112	0.954	0.113	0.952	0.112	0.950	0.112	

Wald: Intervalo de confianza tipo Wald, Waldcc: intervalo de confianza tipo Wald con corrección por continuidad, Logit: intervalo de confianza logit, Bootstrap: intervalo de confianza Bootstrap, Cob.: cobertura, Long.: longitud media.

Tabla 1.8. Cobertura y amplitud promedio de los I.C. para el coeficiente kappa ponderado ($\kappa(c) = 0.8$).

$p = 0.10 \quad \kappa(0) = 0.9 \quad \kappa(1) = 0.4 \quad c = 0.1 \quad Se = 0.43 \quad Sp = 0.995$								
n	Wald		Waldcc		Logit		Bootstrap	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.977	0.699	0.983	0.709	0.986	0.607	0.976	0.754
200	0.986	0.441	0.987	0.446	0.970	0.422	0.985	0.466
300	0.986	0.334	0.987	0.338	0.972	0.329	0.989	0.348
400	0.971	0.275	0.973	0.278	0.971	0.273	0.975	0.284
500	0.964	0.241	0.966	0.243	0.966	0.240	0.971	0.246
1000	0.935	0.164	0.936	0.165	0.946	0.164	0.940	0.166
$p = 0.10 \quad \kappa(0) = 0.8 \quad \kappa(1) = 0.8 \quad c = 0.5 \quad Se = 0.82 \quad Sp = 0.98$								
n	Wald		Waldcc		Logit		Bootstrap	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.976	0.434	0.978	0.444	0.972	0.422	0.977	0.468
200	0.932	0.286	0.938	0.291	0.960	0.286	0.937	0.296
300	0.940	0.232	0.947	0.236	0.961	0.233	0.944	0.237
400	0.932	0.200	0.934	0.203	0.954	0.201	0.933	0.204
500	0.938	0.179	0.939	0.181	0.952	0.179	0.938	0.181
1000	0.935	0.126	0.937	0.127	0.948	0.126	0.935	0.127
$p = 0.10 \quad \kappa(0) = 0.4 \quad \kappa(1) = 0.9 \quad c = 0.9 \quad Se = 0.92 \quad Sp = 0.88$								
n	Wald		Waldcc		Logit		Bootstrap	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.979	0.454	0.990	0.464	0.971	0.434	0.978	0.459
200	0.977	0.284	0.979	0.289	0.968	0.282	0.976	0.290
300	0.963	0.221	0.965	0.224	0.966	0.220	0.968	0.224
400	0.947	0.188	0.951	0.191	0.962	0.188	0.949	0.190
500	0.947	0.168	0.949	0.170	0.955	0.168	0.951	0.169
1000	0.941	0.119	0.944	0.120	0.945	0.119	0.943	0.119
$p = 0.50 \quad \kappa(0) = 0.9 \quad \kappa(1) = 0.4 \quad c = 0.1 \quad Se = 0.58 \quad Sp = 0.97$								
n	Wald		Waldcc		Logit		Bootstrap	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.976	0.286	0.981	0.296	0.977	0.283	0.975	0.286
200	0.951	0.187	0.959	0.192	0.960	0.187	0.952	0.188
300	0.929	0.152	0.936	0.156	0.949	0.153	0.933	0.152
400	0.930	0.132	0.936	0.135	0.938	0.132	0.933	0.132
500	0.943	0.119	0.947	0.121	0.954	0.119	0.944	0.119
1000	0.943	0.084	0.944	0.085	0.945	0.084	0.942	0.084
$p = 0.50 \quad \kappa(0) = 0.8 \quad \kappa(1) = 0.8 \quad c = 0.5 \quad Se = 0.90 \quad Sp = 0.90$								
n	Wald		Waldcc		Logit		Bootstrap	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.933	0.275	0.926	0.285	0.958	0.278	0.935	0.275
200	0.923	0.197	0.928	0.202	0.952	0.198	0.923	0.197
300	0.933	0.162	0.940	0.165	0.957	0.162	0.932	0.161
400	0.940	0.141	0.944	0.143	0.949	0.141	0.939	0.140
500	0.940	0.126	0.943	0.128	0.951	0.126	0.940	0.126
1000	0.943	0.090	0.946	0.091	0.948	0.090	0.941	0.089
$p = 0.50 \quad \kappa(0) = 0.4 \quad \kappa(1) = 0.9 \quad c = 0.9 \quad Se = 0.97 \quad Sp = 0.58$								
n	Wald		Waldcc		Logit		Bootstrap	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.978	0.288	0.982	0.298	0.976	0.288	0.977	0.290
200	0.951	0.188	0.954	0.193	0.959	0.188	0.953	0.189
300	0.942	0.153	0.948	0.156	0.952	0.153	0.944	0.153
400	0.937	0.133	0.941	0.135	0.942	0.133	0.941	0.133
500	0.939	0.119	0.942	0.121	0.941	0.119	0.938	0.119
1000	0.949	0.084	0.950	0.085	0.952	0.084	0.949	0.084

Wald: Intervalo de confianza tipo Wald, Waldcc: intervalo de confianza tipo Wald con corrección por continuidad, Logit: intervalo de confianza logit, Bootstrap: intervalo de confianza Bootstrap, Cob.: cobertura, Long.: longitud media.

1.2.5.3. Aplicación

Los resultados obtenidos se han aplicado al estudio de Weiner et al. (1979) sobre el diagnóstico de la enfermedad coronaria, utilizando como test diagnóstico una prueba de ejercicio y como gold estándar una arteriografía coronaria. En la Tabla 1.9 se muestran los resultados obtenidos por Weiner et al. (1979) para los varones según tres síntomas diferentes (angina definitiva, angina probable y dolor no isquémico), y donde la variable T modeliza el resultado del test de ejercicio y la variable D el resultado de la arteriografía coronaria; y en la Tabla 1.10 se muestran los valores estimados y los intervalos de confianza al 95% para el coeficiente kappa ponderado para diferentes valores del índice de ponderación.

Tabla 1.9. Datos del estudio de Weiner et al. (1979).

Angina definitiva			
	$T = 1$	$T = 0$	Total
$D = 1$	473	81	554
$D = 0$	22	44	66
Total	495	125	620
Angina probable			
	$T = 1$	$T = 0$	Total
$D = 1$	313	102	415
$D = 0$	47	132	179
Total	360	234	594
Dolor no isquémico			
	$T = 1$	$T = 0$	Total
$D = 1$	29	25	54
$D = 0$	46	151	197
Total	75	176	251

Tabla 1.10. Resultados del estudio de Weiner et al. (1979).

Angina definitiva				
c	$\hat{\kappa}(c)$	<i>Wald</i>	<i>Logit</i>	<i>Bootstrap</i>
0.1	0.527	(0.405 , 0.643)	(0.408 , 0.639)	(0.404 , 0.640)
0.2	0.462	(0.367 , 0.584)	(0.370 , 0.584)	(0.366 , 0.581)
0.3	0.412	(0.334 , 0.538)	(0.338 , 0.539)	(0.333 , 0.535)
0.4	0.371	(0.305 , 0.500)	(0.310 , 0.502)	(0.304 , 0.497)
0.5	0.338	(0.280 , 0.467)	(0.285 , 0.471)	(0.279 , 0.465)
0.6	0.310	(0.257 , 0.439)	(0.264 , 0.444)	(0.257 , 0.438)
0.7	0.287	(0.238 , 0.415)	(0.245 , 0.420)	(0.238 , 0.414)
0.8	0.267	(0.221 , 0.393)	(0.229 , 0.399)	(0.221 , 0.393)
0.9	0.249	(0.207 , 0.374)	(0.214 , 0.380)	(0.206 , 0.374)
Angina probable				
c	$\hat{\kappa}(c)$	<i>Wald</i>	<i>Logit</i>	<i>Bootstrap</i>
0.1	0.543	(0.453 , 0.626)	(0.452 , 0.624)	(0.452 , 0.624)
0.2	0.523	(0.434 , 0.596)	(0.434 , 0.594)	(0.433 , 0.593)
0.3	0.504	(0.415 , 0.569)	(0.416 , 0.569)	(0.414 , 0.566)
0.4	0.486	(0.396 , 0.546)	(0.398 , 0.546)	(0.395 , 0.543)
0.5	0.470	(0.379 , 0.526)	(0.380 , 0.526)	(0.377 , 0.523)
0.6	0.454	(0.362 , 0.507)	(0.364 , 0.508)	(0.360 , 0.505)
0.7	0.440	(0.346 , 0.491)	(0.348 , 0.492)	(0.344 , 0.489)
0.8	0.426	(0.331 , 0.476)	(0.333 , 0.477)	(0.329 , 0.474)
0.9	0.413	(0.317 , 0.462)	(0.319 , 0.464)	(0.315 , 0.460)
Dolor no isquémico				
c	$\hat{\kappa}(c)$	<i>Wald</i>	<i>Logit</i>	<i>Bootstrap</i>
0.1	0.340	(0.110 , 0.343)	(0.131 , 0.363)	(0.106 , 0.346)
0.2	0.351	(0.116 , 0.354)	(0.137 , 0.374)	(0.112 , 0.356)
0.3	0.354	(0.123 , 0.367)	(0.144 , 0.385)	(0.119 , 0.368)
0.4	0.377	(0.130 , 0.380)	(0.150 , 0.398)	(0.126 , 0.380)
0.5	0.391	(0.137 , 0.395)	(0.158 , 0.412)	(0.133 , 0.394)
0.6	0.406	(0.144 , 0.412)	(0.165 , 0.428)	(0.140 , 0.410)
0.7	0.423	(0.152 , 0.431)	(0.173 , 0.447)	(0.147 , 0.429)
0.8	0.441	(0.159 , 0.453)	(0.181 , 0.468)	(0.155 , 0.451)
0.9	0.460	(0.165 , 0.478)	(0.188 , 0.493)	(0.162 , 0.477)

Para los varones con angina definitiva o angina probable se obtiene que si el test de ejercicio se utiliza previo a un tratamiento intensivo ($0 < c < 0.5$), entonces el coeficiente de kappa ponderado toma un valor intermedio, por lo que la concordancia

entre el test de ejercicio y la arteriografía es moderada. Si el test de ejercicio se utiliza como un test de screening ($0.5 < c < 1$), el coeficiente kappa ponderado toma valores bajos-intermedios dependiendo del valor del índice de ponderación, por lo que la concordancia entre el test de ejercicio y la arteriografía varía entre mediocre y moderada. En el caso de varones con dolor no isquémico, si el test diagnóstico se realiza antes de un tratamiento intensivo ($0 < c < 0.5$), entonces la concordancia entre el test de ejercicio y la arteriografía varía entre mala y mediocre, dependiendo del valor del índice de ponderación. Si el test se realiza con un test de screening ($0.5 < c < 1$), la concordancia el test de ejercicio y la arteriografía varía entre mala, mediocre y moderada.

1.2.5.4. Discusión

El coeficiente kappa ponderado de un test diagnóstico es una medida del acuerdo clasificatorio o concordancia entre un test binario y el gold estándar, y es un parámetro que depende de la sensibilidad y la especificidad del test binario, de la prevalencia de la enfermedad y de las pérdidas originadas por una clasificación errónea de un paciente enfermo o no enfermo con el test binario. Dependiendo de la prevalencia de la enfermedad, éste parámetro no puede estimarse si el muestreo es caso-control (a menos que el valor de la prevalencia de la enfermedad sea conocido a partir de otro estudio). Con respecto a las pérdidas médicas, éstas no se pueden estimar en la mayoría de las ocasiones, y por consiguiente la elección del valor del índice de ponderación depende del uso clínico del test diagnóstico: si el test binario se utiliza previo a un tratamiento intensivo entonces $0 < c < 0.5$, si el test binario se utiliza como un test de screening entonces $0.5 < c < 1$, y si el test diagnóstico se utiliza para el diagnóstico simple de una enfermedad entonces $c = 0.5$.

Roldán Nofuentes, Luna del Castillo y Montero Alonso (2009b) han estudiado cuatro intervalos de confianza para el coeficiente kappa ponderado de un test diagnóstico binario cuando el test diagnóstico y el gold estándar se aplican a todos los pacientes de una muestra aleatoria. Tres intervalos están basados en la normalidad asintótica del estimador del coeficiente kappa ponderado (intervalo de Wald sin y con corrección por continuidad e intervalo logit) y un cuarto intervalo está basado en el

método bootstrap. Se han realizado unos experimentos de simulación Monte Carlo para estudiar el rendimiento, medido en términos de cobertura y longitud promedio, de estos intervalos de confianza al 95% de confianza. De los resultados de los experimentos de simulación se ha obtenido que, en términos generales, y para valores de coeficientes kappa ponderados comunes en la práctica clínica, cuando las muestras son pequeñas ($n = 100$) el intervalo tipo Wald sin corrección por continuidad y el intervalo bootstrap tienen un buen rendimiento; mientras que para muestras mayores ($n \geq 200$) el intervalo logit presenta un mejor rendimiento que el resto de intervalos. Este mejor comportamiento del intervalo logit en muestras grandes ($n \geq 200$) es debido a que la distribución del logit de $\hat{\kappa}(c)$ es menos asimétrica que la distribución de $\hat{\kappa}(c)$ y de $\hat{\kappa}_B(c)$.

Para estimar el coeficiente kappa ponderado es necesario que todos los totales marginales ($s, r, s_1 + r_1, s_0 + r_0$) de la tabla de frecuencias observadas (Tabla 1.3) sean mayores que cero. Cuando alguno de los totales marginales es cero, el problema se puede resolver de distintas formas:

1. Tomar una muestra suficientemente grande para que no haya ceros en los totales marginales.
2. Realizar un diseño caso-control en dos etapas. En la primera etapa se seleccionan dos muestras, una de n_1 individuos enfermos y otra de n_2 individuos no enfermos, y a continuación se estima la prevalencia p de forma independiente al diseño mediante el método adecuado. En la segunda etapa, se estima la sensibilidad a partir de la muestra de pacientes enfermos y la especificidad a partir de la muestra de pacientes no enfermos. Finalmente se estima el coeficiente kappa ponderado a partir de los valores estimados de la prevalencia, sensibilidad y especificidad.
3. No fijar ningún tamaño muestral a priori y tomar pacientes hasta que el tamaño de muestra sea mayor que un valor prefijado y todos los totales marginales sean mayores que cero.

Aunque el método propuesto para resolver el problema de estimación del coeficiente kappa ponderado se basa en el condicionamiento en los totales de las filas de una muestra de tamaño n (fijado a priori), y no en ninguna de las tres alternativas anteriores, el problema de las muestras con algún total marginal igual a cero no es un problema de gran importancia, ya que estas muestras son poco frecuentes. En los experimentos de simulación realizados se ha contado el número de muestras en las que algún total marginal era cero, obteniéndose que para muestras de tamaño 100 el porcentaje de muestras “defectuosas” (con al menos un total marginal igual a cero) ha sido en promedio aproximadamente el 5%, para muestras de tamaño 200 el porcentaje promedio ha sido inferior al 1% y para muestras de tamaño 300 ha sido inferior al 0.5%, siendo mucho más pequeño para muestras mayores.

La solución del problema de estimación del coeficiente kappa ponderado se basa en el condicionamiento en los totales de las filas. Otra forma de resolver este problema es del punto de vista incondicional. En esta situación, el estimador de sensibilidad y especificidad y, por consiguiente del coeficiente kappa ponderado, es sesgado (Kraemer, 1992, pp 34-40), y el sesgo de cada uno de ellos depende del tamaño de la muestra. Cuando el tamaño de la muestra es bastante grande, los estimadores son aproximadamente insesgados y sus varianzas estimadas son aproximadamente iguales a las dadas en la Sección 1.2.5.1, verificándose también que $Cov(\hat{S}_e, \hat{S}_p) \approx 0$ (Bennett, 1983). Por consiguiente, desde un punto de vista incondicional, cuando el tamaño de la muestra es grande, el problema de estimación del coeficiente kappa ponderado es análogo al basado en el condicionamiento en los totales de las filas.

1.3.- La verificación parcial de la enfermedad y el sesgo de verificación ligado a ella

En la práctica clínica cuando se evalúa un test diagnóstico es frecuente que el gold estándar no se aplique a todos los individuos de una muestra, surgiendo el llamado problema de la verificación parcial de la enfermedad. En esta situación, si el test binario se aplica a todos los sujetos de una muestra de tamaño n y el gold estándar se aplica solamente a un subconjunto de ellos, se obtiene la Tabla 1.11.

Tabla 1.11. Frecuencias observadas en presencia de verificación parcial.

	Test positivo	Test negativo
Verificados		
Enfermos	s_1	s_0
No enfermos	r_1	r_0
No verificados	u_1	u_0
Total	n_1	n_0

En esta Tabla, de los n individuos totales n_1 tienen test positivo y n_0 test negativo. De los n_1 individuos en los que el test ha dado positivo, s_1 están enfermos, r_1 no están enfermos y u_1 no tienen verificado su estado de enfermedad y por tanto se desconoce si están o no enfermos. Análogamente, de los n_0 individuos con test negativo, s_0 están enfermos, r_0 no están enfermos y u_0 no tienen verificado su estado de enfermedad y se desconoce si están o no enfermos. Esta situación se corresponde con estudios de dos fases, (Carroll et al., 1995). En la primera fase el test diagnóstico se aplica a todos los individuos de la muestra y en la segunda fase solamente una parte los individuos de la muestra son verificados con el gold estándar. Estos diseños se utilizan con frecuencia en estudios epidemiológicos (Hand, 1987; Pickles et al., 1995), y aunque este enfoque puede ser práctico y rentable en estudios clínicos, cuando ocurre en estudios para evaluar la exactitud de tests diagnósticos, la estimación de la exactitud puede estar sesgada. A este tipo de sesgo se le denomina sesgo de verificación o workup bias, (Ransohoff y Feinstein, 1978; Begg y Greenes, 1983). El sesgo de verificación surge cuando el estudio de la eficacia de un test diagnóstico se restringe a los individuos con el estado de la enfermedad verificado, dependiendo de la asociación entre la selección para la verificación de la enfermedad y el resultado del test diagnóstico. Esta asociación afecta directamente a las probabilidades de seleccionar un individuo para verificar su estado de enfermedad, ya que la probabilidad de que un individuo sea seleccionado para verificar su estado es alta cuando el resultado del test diagnóstico es positivo y es baja

cuando es negativo, de tal forma que, una fuerte asociación entre la selección para la verificación y el resultado del test produce un gran sesgo, y por el contrario, cuanto mayor es la tasa de pacientes verificados menor es el sesgo de verificación.

Para ilustrar cómo funciona y afecta a los estimadores de la exactitud de un test diagnóstico, se puede considerar el ejemplo de la estimación de la sensibilidad de un test radiográfico en el diagnóstico de una enfermedad coronaria (Tavel, Enas y Woods, 1987), utilizándose como gold estándar una angiografía. Si la sensibilidad de la radiografía es del 80% y se dispone de una muestra de 500 individuos enfermos a los que se les realiza una radiografía, es esperable obtener un resultado positivo en 400 individuos y negativo en 100. Dado que la angiografía es un procedimiento arriesgado y caro, si la probabilidad de verificar a un paciente de radiografía positivo es del 75% y la probabilidad de verificar un individuo con resultado negativo es del 10%, al analizar solamente los individuos verificados se obtiene que la sensibilidad del test radiográfico es del 97%, con lo que se ha sobrestimado la sensibilidad del test. Diferentes estudios también han puesto de manifiesto la existencia de este sesgo de verificación (Philbricks, Horwitz y Feinstein, 1980; Greenes y Begg, 1985; Begg y McNeil, 1988; Bates, Margolis y Evans, 1993; Reid, Lachs y Feinstein, 1995; Cecils et al., 1996).

A continuación se estudia la estimación de la sensibilidad y de la especificidad de un test diagnóstico binario en presencia de verificación parcial de la enfermedad.

1.3.1. Estimación de la sensibilidad y la especificidad

Begg y Greenes (1983) han deducido un método para corregir el sesgo de verificación cuando se estiman la sensibilidad y la especificidad de un test binario. Este método se basa en la suposición de que el proceso de verificación depende solamente del resultado del test diagnóstico y no del estado de la enfermedad, que es equivalente a suponer que el mecanismo de datos faltantes es MAR (missing at random). Zhou (1993) ha ampliado el método de corrección del sesgo de Begg y Greenes, deduciendo las expresiones de los estimadores máximo verosímiles de la sensibilidad y la especificidad de un test diagnóstico binario y sus correspondientes varianzas, tanto sin la presencia de covariables como con ellas, y demostrando que cuando el proceso de verificación no

depende del estado de la enfermedad los estimadores máximo verosímiles coinciden con los estimadores deducidos por Begg y Greenes (1983). Roldán Nofuentes y Luna del Castillo (2007a) han estudiado el tamaño del sesgo de verificación en la estimación de la sensibilidad y de la especificidad cuando no se consideran a los individuos con estado de enfermedad no verificado con el gold estándar. Harel y Zhou (2006) han estudiado el rendimiento de distintos intervalos de confianza para la sensibilidad y la especificidad aplicando imputación múltiple, y han comparado el rendimiento de estos intervalos con los intervalos de confianza obtenidos aplicando el método de corrección del sesgo de Begg y Greenes (1983). En cuanto a los intervalos de confianza, en esta Sección se realiza una aportación, el intervalo de confianza cuadrático para la sensibilidad y para la especificidad, y se compara su rendimiento con los intervalos de Begg y Greenes (1983) y con el método de imputación múltiple de Harel y Zhou (2006). Finalmente, se estudia el comportamiento de los estimadores máximo verosímiles de la sensibilidad y especificidad cuando el proceso de verificación no es MAR (Kosinski y Barnhart, 2003).

1.3.1.1. Estimadores máximo verosímiles de la sensibilidad y de la especificidad sin covariables

Sean las variables aleatorias binarias T , D y V las variables aleatorias definidas como sigue. La variable T modeliza el resultado del test, siendo $T = 1$ cuando el resultado del test es positivo y $T = 0$ cuando es negativo; la variable D modeliza el verdadero estado de la enfermedad o resultado del gold estándar, siendo $D = 1$ cuando el individuo está enfermo y $D = 0$ cuando no está enfermo; y la variable V modeliza el proceso de verificación de la enfermedad, siendo $V = 1$ cuando el individuo ha sido verificado con el gold estándar y $V = 0$ cuando no ha sido verificado. Por consiguiente, el test diagnóstico se aplica a todos los individuos de la muestra, pero sólo a un subconjunto de ellos se les aplica el gold estándar, con lo que no se dispone de información del verdadero estado de la enfermedad para los individuos no verificados, obteniéndose la Tabla 1.12.

Tabla 1.12. Frecuencias observadas bajo verificación parcial de la enfermedad.

	$T = 1$	$T = 0$	Total
$V = 1$			
$D = 1$	s_1	s_0	s
$D = 0$	r_1	r_0	r
$V = 0$	u_1	u_0	u
Total	n_1	n_0	n

Tabla 1.13. Probabilidades de la distribución multinomial.

	$T = 1$	$T = 0$
$V = 1$		
$D = 1$	$pSe\lambda_{11}$	$p(1-Se)\lambda_{10}$
$D = 0$	$(1-p)(1-Sp)\lambda_{01}$	$(1-p)Sp\lambda_{00}$
$V = 0$	$pSe(1-\lambda_{11})+(1-p)(1-Sp)(1-\lambda_{01})$	$p(1-Se)(1-\lambda_{10})+(1-p)Sp(1-\lambda_{00})$

Los datos de la Tabla 1.12 son la realización de una distribución multinomial cuyas probabilidades se muestran en la Tabla 1.13, donde Se es la sensibilidad, Sp la especificidad del test diagnóstico, p la prevalencia de la enfermedad, y $\lambda_{ij} = P(V = 1 | D = i, T = j)$, con $i, j = 0, 1$, las probabilidades de verificación, de tal forma que λ_{11} es la probabilidad de seleccionar para verificar el estado de la enfermedad un individuo enfermo con resultado del test positivo, λ_{01} la probabilidad de seleccionar para verificar el estado de la enfermedad un individuo no enfermo con resultado del test positivo, λ_{10} la probabilidad de seleccionar para verificar el estado de la enfermedad un individuo enfermo con resultado del test negativo y λ_{00} la

probabilidad de seleccionar para verificar el estado de la enfermedad un individuo no enfermo con resultado del test negativo.

La función del logaritmo de la verosimilitud de los datos de la Tabla 1.12 es

$$\begin{aligned}
 l \propto & (s_1 + s_0) \log(p) + (r_1 + r_0) \log(1-p) + \\
 & u_1 \log\{p(1-\lambda_{11})Se + (1-p)(1-\lambda_{01})(1-Sp)\} + \\
 & u_0 \log\{p(1-\lambda_{10})(1-Se) + (1-p)(1-\lambda_{00})Sp\} + s_1 \log(Se) + s_0 \log(1-Se) + \\
 & r_1 \log(1-Sp) + r_0 \log(Sp) + s_1 \log(\lambda_{11}) + s_0 \log(\lambda_{10}) + r_1 \log(\lambda_{01}) + r_0 \log(\lambda_{00}).
 \end{aligned} \tag{1.79}$$

Si se definen los siguientes cocientes

$$k_1 = \frac{P(V=1 | D=1, T=1)}{P(V=1 | D=0, T=1)} = \frac{\lambda_{11}}{\lambda_{01}} \quad \text{y} \quad k_0 = \frac{P(V=1 | D=1, T=0)}{P(V=1 | D=0, T=0)} = \frac{\lambda_{10}}{\lambda_{00}}, \tag{1.80}$$

entonces la función del logaritmo de la verosimilitud se puede reescribir como

$$\begin{aligned}
 l \propto & (s_1 + s_0) \log(p) + (r_1 + r_0) \log(1-p) + s_1 \log(Se) + \\
 & s_0 \log(1-Se) + r_1 \log(1-Sp) + r_0 \log(Sp) + \\
 & (s_1 + r_1) \log(\lambda_{11}) + (s_0 + r_0) \log(\lambda_{10}) + \\
 & u_0 \log\{k_0(1-\lambda_{10})p(1-Se) + (k_0 - \lambda_{10})Sp(1-p)\} - \\
 & (r_1 + u_1) \log(k_1) - (r_0 + u_0) \log(k_0).
 \end{aligned} \tag{1.81}$$

El número de celdas de la distribución multinomial de los datos de la Tabla 1.12 es seis, por lo que solamente se pueden estimar cinco de los siete parámetros de la distribución. Si se supone que k_1 y k_0 son conocidos, los cinco parámetros restantes se pueden estimar por el método de máxima verosimilitud. Los estimadores máximo-verosímiles son las soluciones del siguiente sistema de ecuaciones no lineales:

$$\begin{aligned}
 \frac{\partial l}{\partial Se} = & \frac{s_1}{Se} - \frac{s_0}{1-Se} + \frac{k_1(1-\lambda_{11})p}{k_1(1-\lambda_{11})pSe + (k_1 - \lambda_{11})(1-Sp)(1-p)} u_1 - \\
 & \frac{k_0(1-\lambda_{10})p}{k_0(1-\lambda_{10})(1-Se)p + (k_0 - \lambda_{10})Sp(1-p)} u_0 = 0
 \end{aligned}$$

$$\begin{aligned} \frac{\partial l}{\partial Sp} &= \frac{r_0}{Sp} - \frac{s_1}{1-Sp} + \frac{k_1(1-\lambda_{11})(1-p)}{k_1(1-\lambda_{11})pSe + (k_1-\lambda_{11})(1-Sp)(1-p)} u_1 + \\ &\quad \frac{k_0(1-\lambda_{10})(1-p)}{k_0(1-\lambda_{10})(1-Se)p + (k_0-\lambda_{10})Sp(1-p)} u_0 = 0 \\ \frac{\partial l}{\partial p} &= \frac{s_1+s_0}{p} - \frac{r_1+r_0}{1-p} + \frac{k_1(1-\lambda_{11})p - (k_1-\lambda_{11})(1-Sp)}{k_1(1-\lambda_{11})pSe + (k_1-\lambda_{11})(1-Sp)(1-p)} u_1 + \\ &\quad \frac{k_0(1-\lambda_{10})(1-Se) - (k_0-\lambda_{10})Sp}{k_0(1-\lambda_{10})(1-Se)p + (k_0-\lambda_{10})Sp(1-p)} u_0 = 0 \\ \frac{\partial l}{\partial p_{11}} &= \frac{s_1+r_1}{\lambda_{11}} - \frac{pSek_1 - (1-Sp) - (1-p)}{k_1(1-\lambda_{11})pSe + (k_1-\lambda_{11})(1-Sp)(1-p)} u_1 = 0 \\ \frac{\partial l}{\partial p_{10}} &= \frac{s_0+r_0}{\lambda_{10}} + \frac{(1-Se)pk_0 - Sp(1-p)}{k_0(1-\lambda_{10})(1-Se)p + (k_0-\lambda_{10})Sp(1-p)} u_0 = 0. \end{aligned} \quad (1.82)$$

La obtención de las expresiones explícitas de los estimadores máximo verosímiles a partir de las ecuaciones del sistema no lineal (1.82) no es fácil. A continuación se deducen las expresiones intuitivas de los estimadores de Se , Sp , p , λ_{11} y λ_{10} , y posteriormente se comprueba que dichos estimadores verifican el sistema de ecuaciones no lineales (1.82).

Si se define la función $\#(A)$ como el número total de elementos que satisfacen las condiciones de A , entonces la sensibilidad y la especificidad del test se definen como

$$Se = P(T=1|D=1) = \frac{\#(T=1, D=1)}{\#(D=1)} \quad (1.83)$$

y

$$Sp = P(T=0|D=0) = \frac{\#(T=0, D=0)}{\#(D=0)}, \quad (1.84)$$

y las probabilidades de verificación son

$$\lambda_{11} = \frac{s_1}{\#(T=1, D=1)}, \quad \lambda_{01} = \frac{r_1}{\#(T=1, D=0)}, \quad (1.85)$$

$$\lambda_{10} = \frac{s_0}{\#(T=0, D=1)} \quad \text{y} \quad \lambda_{00} = \frac{r_0}{\#(T=0, D=0)}$$

Sustituyendo las expresiones (1.85) en los coeficientes definidos en (1.80), los valores de k_1 y k_0 vienen dados por las expresiones

$$k_1 = \frac{s_1}{\#(T=1, D=1)} \frac{\#(T=1, D=0)}{r_1} \quad (1.86)$$

y

$$k_0 = \frac{s_0}{\#(T=0, D=1)} \frac{\#(T=0, D=0)}{r_0}. \quad (1.87)$$

Realizando cálculos algebraicos se obtiene que

$$\#(T=1, D=1) = \frac{s_1}{r_1 k_1 + s_1} (s_1 + r_1 + u_1) \quad (1.88)$$

y

$$\#(T=0, D=1) = \frac{s_0}{r_0 k_0 + s_0} (s_0 + r_0 + u_0), \quad (1.89)$$

y por tanto se tiene que

$$\#(D=1) = \frac{s_1}{r_1 k_1 + s_1} (s_1 + r_1 + u_1) + \frac{s_0}{r_0 k_0 + s_0} (s_0 + r_0 + u_0). \quad (1.90)$$

Finalmente, sustituyendo (1.88), (1.89) y (1.90) en las expresiones (1.83), (1.84) y (1.85) se obtienen unos estimadores de los parámetros Se , Sp , p , λ_{11} y λ_{10} , dados por las expresiones

$$\hat{Se} = \frac{s_1 n_1 / (s_1 + k_1 r_1)}{s_1 n_1 / (s_1 + k_1 r_1) + s_0 n_0 / (s_0 + k_0 r_0)}, \quad (1.91)$$

$$\hat{Sp} = \frac{k_0 r_0 n_0 / (s_0 + k_0 r_0)}{(k_1 s_1 n_1) / (s_1 + k_1 r_1) + (k_0 s_0 n_0) / (s_0 + k_0 r_0)}, \quad (1.92)$$

$$\hat{p} = \frac{s_1 n_1 / (s_1 + k_1 r_1) + s_0 n_0 / (s_0 + k_0 r_0)}{n}, \quad (1.93)$$

$$\hat{\lambda}_{11} = \frac{k_1 r_1 + s_1}{n_1}, \quad (1.94)$$

y

$$\hat{\lambda}_{10} = \frac{k_0 r_0 + s_0}{n_0}, \quad (1.95)$$

siendo $n_1 = s_1 + r_1 + u_1$, $n_0 = s_0 + r_0 + u_0$ y $n = n_1 + n_0$. A continuación, se demuestra que estos estimadores son los estimadores máximo verosímiles cuando k_1 y k_0 son conocidos.

Teorema 1.1. (Zhou, 1993)

Cuando los valores k_1 y k_0 son conocidos los estimadores máximo verosímiles de los parámetros Se , Sp , p , λ_{11} y λ_{10} vienen dados por las ecuaciones (1.91) a (1.95), y además

$$\sqrt{n} \begin{pmatrix} \hat{Se} - Se \\ \hat{Sp} - Sp \\ \hat{p} - p \\ \hat{\lambda}_{11} - \lambda_{11} \\ \hat{\lambda}_{10} - \lambda_{10} \end{pmatrix} \xrightarrow{L} N_5 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma^{-1} \right),$$

donde los elementos de la matriz Σ son:

$$\sigma_{11} = \frac{p\lambda_{11}}{Se} + \frac{p\lambda_{10}}{1-Se} + \frac{(1-\lambda_{11})^2 p^2}{(1-\lambda_{11})pSe + (1-\lambda_{11}/k_1)(1-Sp)(1-p)} + \frac{(1-\lambda_{10})^2 p^2}{(1-\lambda_{10})p(1-Se) + (1-\lambda_{10}/k_0)Sp(1-p)},$$

$$\sigma_{22} = \frac{(1-p)\lambda_{11}/k_1}{1-Sp} + \frac{(1-p)\lambda_{10}/k_0}{Sp} + \frac{(1-\lambda_{11}/k_1)^2 (1-p)^2}{(1-\lambda_{11})pSe + (1-\lambda_{11}/k_1)(1-Sp)(1-p)} + \frac{(1-\lambda_{10}/k_0)^2 (1-p)^2}{(1-\lambda_{10})p(1-Se) + (1-\lambda_{10}/k_0)Sp(1-p)},$$

$$\sigma_{33} = \frac{Se\lambda_{11} + (1-Se)\lambda_{11}}{p} + \frac{\lambda_{11}(1-Sp)/k_1 + \lambda_{10}Sp/k_1}{1-p} + \frac{((1-\lambda_{11})Se - (1-\lambda_{11}/k_1)(1-Sp))^2}{(1-\lambda_{11})pSe + (1-\lambda_{11}/k_1)(1-Sp)(1-p)} + \frac{((1-\lambda_{10})(1-Se) - (1-\lambda_{10}/k_0)Sp)^2}{(1-\lambda_{10})p(1-Se) + (1-\lambda_{10}/k_0)Sp(1-p)},$$

$$\sigma_{44} = \frac{Se + (1-Sp)(1-p)/k_1}{\lambda_{11}} + \frac{(pSe + (1-Sp)(1-p)/k_1)^2}{(1-\lambda_{11})pSe + (1-\lambda_{11}/k_1)(1-Sp)(1-p)},$$

$$\sigma_{55} = \frac{(1-Se)p + Sp(1-p)/k_0}{\lambda_{10}} + \frac{((1-Se)p + Sp(1-p)/k_0)^2}{(1-\lambda_{10})(1-Se)p + (1-\lambda_{10}/k_0)Sp(1-p)},$$

$$\sigma_{12} = \frac{(1-p_{11}/k_1)(1-p_{11})p(1-p)}{(1-p_{11})pSe + (1-p_{11}/k_1)(1-Sp)(1-p)} - \frac{(1-p_{10}/k_0)(1-p_{10})p(1-p)}{(1-p_{10})(1-Se)p + (1-p_{10}/k_0)Sp(1-p)},$$

$$\sigma_{13} = \frac{(1-\lambda_{11}/k_1)(1-\lambda_{11})(1-Sp)}{(1-\lambda_{11})pSe + (1-\lambda_{11}/k_1)(1-Sp)(1-p)} + \frac{(1-\lambda_{10}/k_0)(1-\lambda_{10})Sp}{(1-\lambda_{10})(1-Se)p + (1-\lambda_{10}/k_0)Sp(1-p)},$$

$$\sigma_{14} = \frac{(1 - \lambda_{11}/k_1)(1 - \lambda_{11})(1 - Sp)}{(1 - \lambda_{11}) pSe + (1 - \lambda_{11}/k_1)(1 - Sp)(1 - p)},$$

$$\sigma_{15} = \frac{(1 - 1/k_1) p(1 - p) Sp}{(1 - \lambda_{10})(1 - Se) p + (1 - \lambda_{10}/k_0) Sp(1 - p)},$$

$$\sigma_{23} = \frac{(1 - \lambda_{11}/k_1)(1 - \lambda_{11}) Se}{(1 - \lambda_{11}) pSe + (1 - \lambda_{11}/k_1)(1 - Sp)(1 - p)} + \frac{(1 - \lambda_{10}/k_0)(1 - \lambda_{10})(1 - Sp)}{(1 - \lambda_{10})(1 - Se) p + (1 - \lambda_{10}/k_0) Sp(1 - p)},$$

$$\sigma_{24} = \frac{(1 - 1/k_1) p(1 - p) Se}{(1 - \lambda_{11}) pSe + (1 - \lambda_{11}/k_1)(1 - Sp)(1 - p)},$$

$$\sigma_{25} = \frac{(1 - 1/k_0) p(1 - p)(1 - Se)}{(1 - \lambda_{10})(1 - Se) p + (1 - \lambda_{10}/k_0) Sp(1 - p)},$$

$$\sigma_{34} = \frac{(1 - 1/k_1) Se(1 - Sp)}{(1 - \lambda_{11}) pSe + (1 - \lambda_{11}/k_1)(1 - Sp)(1 - p)},$$

$$\sigma_{35} = \frac{(1 - 1/k_0)(1 - Se) Sp}{(1 - \lambda_{10})(1 - Se) p + (1 - \lambda_{10}/k_0) Sp(1 - p)},$$

$$\sigma_{45} = 0.$$

Demostración

La demostración de que los estimadores intuitivos son los estimadores máximo verosímiles consiste en comprobar que los estimadores intuitivos satisfacen el sistema de ecuaciones no lineales (1.82). Realizando operaciones algebraicas se verifican las siguientes igualdades,

$$\hat{p}\hat{S}e = \frac{s_1 n_1}{(k_1 r_1 + s_1) n}, \quad (1 - \hat{S}e)(1 - \hat{p}) = \frac{k_1 r_1 n_1}{(k_1 r_1 + s_1) n}, \quad (1.96)$$

$$k_1(1-\hat{\lambda}_{11})\hat{S}e\hat{p} + (k_1-\hat{\lambda}_{11})(1-\hat{S}p)(1-\hat{p}) = \frac{k_1 u_1}{n} \quad (1.97)$$

y

$$k_0(1-\hat{\lambda}_{10})(1-\hat{S}e)\hat{p} + (k_0-\hat{\lambda}_{10})\hat{S}p(1-\hat{p}) = \frac{k_0 u_0}{n}. \quad (1.98)$$

Por otra parte,

$$\frac{s_1}{\hat{S}e} + \frac{k_1(1-\hat{\lambda}_{11})\hat{p}}{k_1(1-\hat{\lambda}_{11})\hat{S}e\hat{p} + (k_1-\hat{\lambda}_{11})(1-\hat{S}p)(1-\hat{p})} u_1 = \frac{s_1 n_1}{k_1 r_1 + s_1} + \frac{s_0 n_0}{k_0 r_0 + s_0} \quad (1.99)$$

y

$$\frac{s_0}{1-\hat{S}e} + \frac{k_0(1-\hat{\lambda}_{10})\hat{p}}{k_0(1-\hat{\lambda}_{10})(1-\hat{S}e)\hat{p} + (k_0-\hat{\lambda}_{10})\hat{S}p(1-\hat{p})} u_0 = \frac{s_1 n_1}{k_1 r_1 + s_1} + \frac{s_0 n_0}{k_0 r_0 + s_0} \quad (1.100)$$

son dos expresiones iguales entre sí, y por tanto, restándolas se obtiene la primera ecuación del sistema (1.82). De forma similar se obtienen el resto de ecuaciones del sistema (1.82). Por tanto, los estimadores dados por las expresiones (1.91) a (1.95) son los estimadores máximo verosímiles de Se , Sp , p , λ_{11} y λ_{10} .

Como los estimadores máximo verosímiles dependen de los valores de k_1 y k_0 , éstos se pueden utilizar para estimar k_1 y k_0 . Suponiendo que de los u_1 individuos con estado de enfermedad no verificado y resultado del test positivo, s_{11} están enfermos y r_{01} no lo están, y que de los u_0 individuos con estado de la enfermedad no verificado y resultado del test negativo, s_{10} están enfermos y r_{00} no lo están, entonces los estimadores insesgados de k_1 y k_0 son

$$\hat{k}_1 = \frac{s_1(r_1 + u_1 - s_{11})}{(s_1 + s_{11})r_1} \quad (1.101)$$

y

$$\hat{k}_0 = \frac{s_0(r_0 + u_0 - s_{10})}{(s_0 + s_{10})r_0}. \quad (1.102)$$

En la práctica, solamente se conoce que $0 \leq s_{11} \leq u_1$ y $0 \leq s_{10} \leq u_0$, por lo que \hat{k}_1 y \hat{k}_0 se encuentran entre los valores

$$\frac{s_1}{s_1 + u_1} \leq \hat{k}_1 \leq \frac{r_1 + u_1}{r_1} \quad (1.103)$$

y

$$\frac{s_0}{s_0 + u_0} \leq \hat{k}_0 \leq \frac{r_0 + u_0}{r_0}. \quad (1.104)$$

De las ecuaciones (1.103) y (1.104) se deduce que si u_1 es relativamente más pequeño que s_1 y r_1 , entonces \hat{k}_1 es aproximadamente igual a uno; y de forma similar, si u_0 es relativamente más pequeño que s_0 y r_0 , entonces \hat{k}_0 es también aproximadamente igual a uno. En esta situación, la sensibilidad y la especificidad del test diagnóstico se puede estimar utilizando el método de Begg y Greenes (1983), siendo los estimadores de la sensibilidad y especificidad

$$\hat{Se} = \frac{n_1 s_1 / (s_1 + r_1)}{n_1 s_1 / (s_1 + r_1) + n_0 s_0 / (s_0 + r_0)} \quad (1.105)$$

y

$$\hat{Sp} = \frac{n_0 r_0 / (s_0 + r_0)}{n_1 r_1 / (s_1 + r_1) + n_0 r_0 / (s_0 + r_0)}, \quad (1.106)$$

y

$$\hat{p} = \frac{n_1 s_1 / (s_1 + r_1) + n_0 s_0 / (s_0 + r_0)}{n} \quad (1.107)$$

el estimador de la prevalencia de la enfermedad. Para la sensibilidad, el término $n_1s_1/(s_1+r_1)$ es la proporción de individuos enfermos que han sido verificados y en los que el test es positivo ponderado por el total de individuos con resultado del test positivo, y $n_0s_0/(s_0+r_0)$ es la proporción de individuos no enfermos verificados en los que el test es negativo ponderado por el total de individuos con resultado del test negativo. Para la especificidad, $n_0r_0/(s_0+r_0)$ es la proporción de individuos no enfermos que han sido verificados y en los que el test es negativo ponderado por el total de individuos con resultado del test negativo y $n_1r_1/(s_1+r_1)$ es la proporción de individuos no enfermos verificados en los que el test es positivo ponderada por el total de individuos con resultado del test positivo. Si $k_1=1$ y $k_0=1$ entonces $\lambda_{11}=\lambda_{01}$ y $\lambda_{10}=\lambda_{00}$, y por tanto $\lambda_{ij}=\lambda_j$, y el proceso de verificación depende solamente del resultado del test diagnóstico ($P(V=1|D=i,T=j)=P(V=1|T=j)$) y no del estado de la enfermedad. Esto es equivalente a suponer que el proceso de verificación es MAR (Rubin, 1976). Por consiguiente, bajo la suposición MAR se pueden obtener los estimadores máximo verosímiles de la sensibilidad y de la especificidad del test diagnóstico, dados por las ecuaciones (1.105) y (1.106), y que coinciden con los deducido por Begg y Greenes (1983).

1.3.1.2. Estimadores máximo verosímiles en presencia de covariables

Zhou (1993) ha deducido los estimadores máximo verosímiles de la sensibilidad y especificidad de un test diagnóstico binario en presencia del sesgo de verificación cuando en todos los individuos se observan un conjunto de covariables discretas. Sea X un vector de covariables discretas observadas en todos los pacientes, siendo x_j la j -ésima covariable observada, $j=1,\dots,J$. Se asume que X es una muestra aleatoria de un espacio discreto (x_1,\dots,x_J) con probabilidades $\xi=(\xi_1,\dots,\xi_J)$. En la Tabla 1.14, se muestran los datos obtenidos al evaluar el test diagnóstico cuando $X=x_j$.

Tabla 1.14. Frecuencias observadas cuando $X = x_j$.

	$T = 1$	$T = 0$
$V = 1$		
$D = 1$	s_{1j}	s_{0j}
$D = 0$	r_{1j}	r_{0j}
$V = 0$	u_{1j}	u_{0j}
Total	n_{1j}	n_{0j}

En esta situación, se definen las siguientes probabilidades:

$$\begin{aligned}
 Se_j &= P(T = 1 | D = 1, X = x_j), \quad Sp_j = P(T = 0 | D = 0, X = x_j), \quad \xi_j = P(X = x_j), \\
 p_j &= P(D = 1 | X = x_j), \quad \lambda_{11j} = P(V = 1 | T = 1, D = 1, X = x_j), \\
 \lambda_{01j} &= P(V = 1 | T = 1, D = 0, X = x_j), \quad \lambda_{10j} = P(V = 1 | T = 0, D = 1, X = x_j), \\
 \lambda_{00j} &= P(V = 1 | T = 0, D = 0, X = x_j),
 \end{aligned} \tag{1.108}$$

siendo Se_j y Sp_j la sensibilidad y la especificidad del test diagnóstico cuando $X = x_j$.

Sean $k_{1j} = \frac{\lambda_{11j}}{\lambda_{01j}}$ y $k_{0j} = \frac{\lambda_{10j}}{\lambda_{00j}}$, y $\alpha_1 = \sum_{j=1}^J \frac{1}{p_j \xi_j}$ y $\alpha_2 = \sum_{j=1}^J \frac{1}{(1-p_j) \xi_j}$. Como

$\xi = (\xi_1, \dots, \xi_J)^T$ es el vector de probabilidades de una distribución multinomial, su matriz de varianzas-covarianzas es $\Sigma = \text{diag}\{\xi_j\} - \xi \xi^T$. En el siguiente Teorema se obtienen los estimadores máximo verosímiles de la exactitud del test diagnóstico en presencia de covariables.

Teorema 1.2. (Zhou, 1993)

Si los valores k_{1j} y k_{0j} son conocidos, entonces se verifica que:

1. Los estimadores máximo verosímiles de los parámetros Se_j , Sp_j , p_j , λ_{11j} y λ_{10j} son:

$$\hat{Se}_j = \frac{s_{1j}n_{1j}/(s_{1j} + k_{1j}r_{1j})}{s_{1j}n_{1j}/(s_{1j} + k_{1j}r_{1j}) + s_{0j}n_{0j}/(s_{0j} + k_{0j}r_{0j})}, \quad (1.109)$$

$$\hat{Sp}_j = \frac{k_{0j}r_{0j}n_{0j}/(s_{0j} + k_{0j}r_{0j})}{(k_{1j}s_{1j}n_{1j})/(s_{1j} + k_{1j}r_{1j}) + (k_{0j}s_{0j}n_{0j})/(s_{0j} + k_{0j}r_{0j})}, \quad (1.110)$$

$$\hat{p}_j = \frac{s_{1j}n_{1j}/(s_{1j} + k_{1j}r_{1j}) + s_{0j}n_{0j}/(s_{0j} + k_{0j}r_{0j})}{n_j}, \quad (1.111)$$

$$\hat{\lambda}_{11j} = \frac{k_{1j}r_{1j} + s_{1j}}{n_{1j}}, \quad (1.112)$$

y

$$\hat{\lambda}_{10j} = \frac{k_{0j}r_{0j} + s_{0j}}{n_{0j}}, \quad (1.113)$$

siendo $n_j = n_{1j} + n_{0j}$.

2. Los estimadores máximo verosímiles de la sensibilidad y de la especificidad global del test diagnóstico son

$$\hat{Se} = \frac{\sum_{j=1}^J \hat{Se}_j \hat{p}_j n_j}{\sum_{j=1}^J \hat{p}_j n_j}, \quad (1.114)$$

y

$$\hat{Sp} = \frac{\sum_{j=1}^J \hat{Sp}_j (1 - \hat{p}_j) n_j}{\sum_{j=1}^J (1 - \hat{p}_j) n_j}. \quad (1.115)$$

3. Las varianzas asintóticas de los estimadores de la sensibilidad y de la especificidad global vienen dadas por las expresiones:

$$\begin{aligned} \lim_{n \rightarrow \infty} n \text{Var}(\hat{Se}) &= \sum_{j=1}^J \left((\alpha_1 p_j \xi_j)^2 \tilde{\sigma}_{11} + 2\alpha_1^3 p_j \xi_j^2 \sum_{i=1}^J (Se_j - Se_i) \xi_i p_i \tilde{\sigma}_{13} + \right. \\ &\left. \alpha_1^2 \xi_j^2 \left(\sum_{i=1}^J (Se_j - Se_i) \xi_i p_i \right)^2 \tilde{\sigma}_{33} \right) + \left(\frac{\delta Se_1}{\delta \xi_1}, \dots, \frac{\delta Se_J}{\delta \xi_J} \right) \Sigma \left(\frac{\delta Se_1}{\delta \xi_1}, \dots, \frac{\delta Se_J}{\delta \xi_J} \right)^T, \end{aligned} \quad (1.116)$$

y

$$\begin{aligned} \lim_{n \rightarrow \infty} n \text{Var}(\hat{Sp}) &= \sum_{j=1}^J \left((\alpha_2 p_j \xi_j)^2 \tilde{\sigma}_{22} + 2\alpha_2^3 p_j \xi_j^2 \sum_{i=1}^J (Sp_j - Sp_i) \xi_i p_i \tilde{\sigma}_{23} + \right. \\ &\left. \alpha_2^2 \xi_j^2 \left(\sum_{i=1}^J (Sp_j - Sp_i) \xi_i p_i \right)^2 \tilde{\sigma}_{33} \right) + \left(\frac{\delta Sp_1}{\delta \xi_1}, \dots, \frac{\delta Sp_J}{\delta \xi_J} \right) \Sigma \left(\frac{\delta Sp_1}{\delta \xi_1}, \dots, \frac{\delta Sp_J}{\delta \xi_J} \right)^T. \end{aligned} \quad (1.117)$$

Los estimadores de la exactitud del test diagnóstico obtenidos en presencia de covariables son una generalización de los obtenidos en la Sección 1.3.1.1; si no existen covariables, las expresiones (1.109) a (1.113) coinciden con las deducidas en dicha Sección. Finalmente, los estimadores de la sensibilidad y especificidad global del test diagnóstico dados por las expresiones (1.114) y (1.115) son una media ponderada de las sensibilidades y especificidades del test diagnóstico en cada patrón de covariable.

1.3.1.3. Estimadores simples de la sensibilidad y de la especificidad

En presencia de verificación parcial de la enfermedad, si la sensibilidad y la especificidad se estiman utilizando solamente los individuos con estado de enfermedad verificado, los estimadores obtenidos, denominados estimadores simples, están afectados por el sesgo de verificación (Begg y Greenes (1983) y Zhou (1993)). Roldán Nofuentes y Luna del Castillo (2007a) han estudiado el efecto del sesgo de verificación en los estimadores simples de la sensibilidad y de la especificidad de un test diagnóstico.

Si la exactitud de un test diagnóstico se estima sin considerar a los pacientes con estado de enfermedad desconocido, los estimadores simples de la sensibilidad y

especificidad, denotados como $\hat{S}e_{nv}$ y $\hat{S}p_{nv}$, son los estimadores de proporciones binomiales, dados por las expresiones

$$\hat{S}e_{nv} = \frac{s_1}{s_1 + s_0} \quad \text{y} \quad \hat{S}p_{nv} = \frac{r_0}{r_1 + r_0}, \quad (1.118)$$

y sus respectivas esperanzas aproximadas bajo la hipótesis MAR (Roldán Nofuentes y Luna del Castillo, 2007a) son

$$E(\hat{S}e_{nv}) \approx \frac{\lambda_1 Se}{\lambda_1 Se + \lambda_0 (1 - Se)} \quad \text{y} \quad E(\hat{S}p_{nv}) \approx \frac{\lambda_0 Sp}{\lambda_1 (1 - Sp) + \lambda_0 Sp}, \quad (1.119)$$

siendo $\lambda_1 = P(V = 1|T = 1)$ y $\lambda_0 = P(V = 1|T = 0)$. Una vez obtenidas las esperanzas de los estimadores simples, los sesgos de los estimadores simples bajo la hipótesis MAR son

$$\text{Sesgo}(\hat{S}e_{nv}) \approx \frac{Se(1 - Se)(\lambda_0 - \lambda_1)}{\lambda_1 Se + \lambda_0 (1 - Se)} \quad \text{y} \quad \text{Sesgo}(\hat{S}p_{nv}) \approx \frac{Sp(1 - Sp)(\lambda_1 - \lambda_0)}{\lambda_1 (1 - Sp) + \lambda_0 Sp}. \quad (1.120)$$

De las ecuaciones (1.120) se obtiene que si bajo la hipótesis MAR se verifica que $\lambda_1 = \lambda_0$, entonces los estimadores simples son aproximadamente insesgados. Por tanto, si bajo la hipótesis MAR se verifica que $\lambda_1 = \lambda_0$, entonces la probabilidad de seleccionar a un individuo para verificar su estado de enfermedad no depende ni del resultado del test diagnóstico ni del estado de enfermedad. De esta forma, si la selección de un individuo para verificar su estado de enfermedad se realiza completamente al azar (el proceso de verificación es MCAR (missing completely at random)), los estimadores simples de la sensibilidad y de la especificidad son aproximadamente insesgados, y por tanto la evaluación del test diagnóstico se puede realizar utilizando solamente los individuos con estado de enfermedad verificado y aplicando los métodos estadísticos clásicos en tablas 2×2 .

1.3.1.4. Intervalos de confianza para la sensibilidad y la especificidad

En presencia de verificación parcial de la enfermedad, la sensibilidad y la especificidad no son proporciones binomiales y por consiguiente estos parámetros no se pueden estimar utilizando los intervalos de confianza explicados en la Sección 1.2.1. A continuación se estudian varios intervalos de confianza para la sensibilidad y la especificidad, cuando el test diagnóstico se aplica a todos los individuos de una muestra aleatoria de tamaño n y el gold estándar solamente se aplica a un subconjunto de ellos.

a. Intervalo de confianza de Begg y Greenes

Bajo la suposición MAR, los estimadores de la sensibilidad y de la especificidad viene dados por las expresiones (1.105) y (1.106), y aplicando el método delta sus varianzas estimadas son (Begg y Greenes, 1983)

$$\widehat{Var}(\hat{Se}) = \{\hat{Se}(1 - \hat{Se})\}^2 \left(\frac{n}{n_1 n_2} + \frac{r_1}{s_1(s_1 + r_1)} + \frac{r_0}{s_0(s_0 + r_0)} \right) \quad (1.121)$$

y

$$\widehat{Var}(\hat{Sp}) = \{\hat{Sp}(1 - \hat{Sp})\}^2 \left(\frac{n}{n_1 n_2} + \frac{s_1}{r_1(s_1 + r_1)} + \frac{s_0}{r_0(s_0 + r_0)} \right), \quad (1.122)$$

y los respectivos intervalos de confianza a nivel $100(1 - \alpha)\%$ son

$$\hat{Se} \pm z_{1-\alpha/2} \sqrt{\widehat{Var}(\hat{Se})} \quad (1.123)$$

y

$$\hat{Sp} \pm z_{1-\alpha/2} \sqrt{\widehat{Var}(\hat{Sp})}. \quad (1.124)$$

El intervalo de confianza de Begg y Greenes es un intervalo de confianza tipo Wald.

b. Intervalo de confianza logit de Begg y Greenes

En lugar de asumir la normalidad de \hat{Se} y de \hat{Sp} , la transformación logit de cada uno de estos estimadores, $\ln\{\hat{Se}/(1-\hat{Se})\}$ y $\ln\{\hat{Sp}/(1-\hat{Sp})\}$, sigue una distribución normal de media $\ln\{Se/(1-Se)\}$ y $\ln\{Sp/(1-Sp)\}$ respectivamente. De esta forma, los intervalos de confianza para $\text{logit}(Se)$ y $\text{logit}(Sp)$ son

$$\text{logit}(\hat{Se}) \pm z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\text{logit}(\hat{Se}))} \quad (1.125)$$

y

$$\text{logit}(\hat{Sp}) \pm z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\text{logit}(\hat{Sp}))} \quad (1.126)$$

respectivamente, siendo las respectivas varianzas

$$\widehat{\text{Var}}(\text{logit}(\hat{Se})) = \left(\frac{n}{n_1 n_2} + \frac{r_1}{s_1 (s_1 + r_1)} + \frac{r_0}{s_0 (s_0 + r_0)} \right) \quad (1.127)$$

y

$$\widehat{\text{Var}}(\text{logit}(\hat{Sp})) = \left(\frac{n}{n_1 n_2} + \frac{s_1}{r_1 (s_1 + r_1)} + \frac{s_0}{r_0 (s_0 + r_0)} \right). \quad (1.128)$$

Finalmente, los intervalos de confianza logit para la sensibilidad y para la especificidad son

$$\left(\frac{\exp\left\{\text{logit}(\hat{Se}) - z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\text{logit}(\hat{Se}))}\right\}}{1 + \exp\left\{\text{logit}(\hat{Se}) - z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\text{logit}(\hat{Se}))}\right\}}, \frac{\exp\left\{\text{logit}(\hat{Se}) + z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\text{logit}(\hat{Se}))}\right\}}{1 + \exp\left\{\text{logit}(\hat{Se}) + z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\text{logit}(\hat{Se}))}\right\}} \right) \quad (1.129)$$

y

$$\left(\frac{\exp\{\text{logit}(\hat{S}p) - z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\text{logit}(\hat{S}p))}\}}{1 + \exp\{\text{logit}(\hat{S}p) - z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\text{logit}(\hat{S}p))}\}}, \frac{\exp\{\text{logit}(\hat{S}p) + z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\text{logit}(\hat{S}p))}\}}{1 + \exp\{\text{logit}(\hat{S}p) + z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\text{logit}(\hat{S}p))}\}} \right), \quad (1.130)$$

respectivamente.

c. Intervalos de confianza mediante imputación múltiple

La imputación múltiple de Rubin (1987) es una técnica basada en la simulación que consiste en sustituir los datos faltantes por un conjunto de m posibles datos, dando por resultado un sistema de $m > 1$ conjuntos de datos completos, y que requiere que los datos faltantes se originen de forma aleatoria. En cada conjunto de datos completos se calculan los estimadores de los parámetros y sus errores estándares, que, combinados mediante reglas aritméticas, dan un resultado que tiene en consideración los valores faltantes. Harel y Zhou (2006) han aplicado la imputación múltiple de Rubin para estimar la sensibilidad y la especificidad de un test diagnóstico binario en presencia de verificación parcial de la enfermedad, empleando para ello varios intervalos de confianza para proporciones binominales. El método de Harel y Zhou (2006) tiene las siguientes fases:

1) Fase de imputación

La fase principal de la imputación múltiple es obtener la distribución a posteriori de los datos con verdadero estado de enfermedad dado que el individuo no ha sido verificado con el gold estándar (ya sea con test positivo o negativo). Asumiendo que el mecanismo de datos faltantes es MAR, los datos de la Tabla 1.3 son la realización de una distribución multinomial cuyas probabilidades se muestran en la Tabla 1.13 pero considerando que $\lambda_{ij} = \lambda_j$ (hipótesis MAR). Asimismo, de los u_j individuos no verificados, se asumen que u_{1j} están enfermos y u_{0j} no lo están, de tal forma que $u_j = u_{1j} + u_{0j}$ y solamente se observa la frecuencia marginal u_j , con $j = 0, 1$. Aplicando

propiedades de esta distribución, la distribución de los datos faltantes dado los datos observados $(Y_{obs} = \{(s_i, r_i, u_i), i = 0, 1\})$ es una distribución multinomial, esto es,

$$(u_{1j}, u_{0j}) | Y_{obs}, \theta \sim M(u_j, (\theta_{1j} / \theta_{+j}, \theta_{0j} / \theta_{+j})), \quad j = 0, 1, \quad (1.131)$$

donde θ_{ij} es la probabilidad de que una unidad se esté en la celda (i, j) y $\theta_{+j} = \sum_i \theta_{ij}$.

Para los parámetros de la distribución multinomial se elige una distribución a priori de Dirichlet, de tal forma que

$$(s, r) | \theta \sim M(n, \theta) \quad (1.132)$$

$$\theta \sim D(\alpha) \quad (1.133)$$

$$\theta | Y \sim D(\alpha') \quad (1.134)$$

donde $\alpha' = \alpha + (s, r)$ y $D(\alpha)$ es una distribución Dirichlet con parámetro α . Finalmente, la imputación de los datos se realiza mediante modelos log-lineales.

2) Fase de análisis

Tras imputar los datos, se obtienen m conjuntos de datos completos, obteniendo las estimaciones $(\hat{Q}^{(1)}, \hat{Q}^{(2)}, \dots, \hat{Q}^{(m)})$ y las varianzas asociadas $(U^{(1)}, U^{(2)}, \dots, U^{(m)})$ para la sensibilidad y especificidad. Harel y Zhou (2006) han utilizado distintos intervalos de confianza para la sensibilidad y la especificidad con los datos completos, entre ellos los intervalos de Agresti y Coull (1998) y logit de Rubin y Schenker (1987).

3) Combinación de resultados

Tras la obtención de m conjuntos de estimadores y sus varianzas, se utiliza la combinación de reglas de Rubin de la siguiente forma. La estimación global es $\bar{Q} = (1/m) \sum Q^{(i)}, i = 1, \dots, m$, y su varianza es $T = \bar{U} + (1/(m+1))B$, donde $\bar{U} = (1/m) \sum U^i$ es la varianza estimada de los datos completos, y $(1/(m+1))B$ es la

varianza debida a la imputación de los valores faltantes, siendo $B = (1/(m-1)) \sum_{i=1}^m (Q^{(i)} - \bar{Q})^2$. Las inferencias se basan en la aproximación de la distribución t de Student $T^{-1/2}(Q - \hat{Q}) \sim t_\nu$, donde los grados de libertad son $\nu = (m-1) \left[1 + (\bar{U} / (1+m^{-1})B) \right]^2$. Finalmente, el intervalo de confianza a nivel $100(1-\alpha)\%$ es

$$\bar{Q} \pm t_{\nu, 1-\alpha/2} \sqrt{T}. \quad (1.135)$$

Harel y Zhou (2006) han propuesto este método de imputación múltiple para corregir el sesgo de verificación en la estimación de la sensibilidad y de la especificidad, y han realizado unos experimentos de simulación para comparar el rendimiento de distintos intervalos de confianza mediante imputación múltiple con los intervalos de confianza de Beggs y Greenes y logit de Beggs y Greenes, obteniendo que el intervalo que presenta un mejor rendimiento es el intervalo logit de Rubin y Schenker con la imputación múltiple.

d. Intervalo de confianza cuadrático (Aportación)

Asumiendo la normalidad asintótica de $\hat{S}e$ y de $\hat{S}p$, resolviendo las ecuaciones

$$\frac{(\hat{S}e - Se)^2}{Var(Se)} = z_{1-\alpha/2}^2 \quad \text{y} \quad \frac{(\hat{S}p - Sp)^2}{Var(Sp)} = z_{1-\alpha/2}^2, \quad (1.136)$$

sustituyendo $Var(Se)$ por $\widehat{Var}(\hat{S}e)$ y $Var(Sp)$ por $\widehat{Var}(\hat{S}p)$, y realizando las operaciones algebraicas, los intervalos de confianza cuadráticos a nivel $100(1-\alpha)\%$ para la sensibilidad y la especificidad son

$$\frac{z_{1-\alpha/2} \sqrt{\widehat{Var}(\logit(\hat{S}e))} \mp 1 \pm \sqrt{\pm 4 \hat{S}e z_{1-\alpha/2} \sqrt{\widehat{Var}(\logit(\hat{S}e))} + \left(1 \mp z_{1-\alpha/2} \sqrt{\widehat{Var}(\logit(\hat{S}e))} \right)^2}}{2 z_{1-\alpha/2} \sqrt{\widehat{Var}(\logit(\hat{S}e))}} \quad (1.137)$$

y

$$\frac{z_{1-\alpha/2}\sqrt{\widehat{Var}(\logit(\hat{Sp}))} \mp 1 \pm \sqrt{\pm 4\hat{Sp}z_{1-\alpha/2}\sqrt{\widehat{Var}(\logit(\hat{Sp}))} + \left(1 \mp z_{1-\alpha/2}\sqrt{\widehat{Var}(\logit(\hat{Sp}))}\right)^2}}{2z_{1-\alpha/2}\sqrt{\widehat{Var}(\logit(\hat{Sp}))}}, \quad (1.138)$$

respectivamente. Estos intervalos también se pueden obtener añadiendo una corrección por continuidad, de tal forma que las ecuaciones son

$$\frac{(|\hat{\theta} - \theta| - 0.5)^2}{Var(\theta)} = z_{1-\alpha/2}^2,$$

$$\frac{(|\hat{Se} - Se| - 0.5)^2}{Var(Se)} = z_{1-\alpha/2}^2 \quad \text{y} \quad \frac{(|\hat{Sp} - Sp| - 0.5)^2}{Var(Sp)} = z_{1-\alpha/2}^2, \quad (1.139)$$

y repitiendo el proceso anterior se obtiene el intervalo de confianza cuadrático con corrección por continuidad a nivel $100(1-\alpha)\%$ para la sensibilidad y para la especificidad, siendo sus expresiones

$$\frac{n\left(z_{1-\alpha/2}\sqrt{\widehat{Var}(\logit(\hat{Se}))} \mp 1\right) \pm \sqrt{2n(-1 \pm 2n\hat{Se})z_{1-\alpha/2}\sqrt{\widehat{Var}(\logit(\hat{Se}))} + n^2\left(1 \mp z_{1-\alpha/2}\sqrt{\widehat{Var}(\logit(\hat{Se}))}\right)^2}}{2nz_{1-\alpha/2}\sqrt{\widehat{Var}(\logit(\hat{Se}))}} \quad (1.140)$$

y

$$\frac{n\left(z_{1-\alpha/2}\sqrt{\widehat{Var}(\logit(\hat{Sp}))} \mp 1\right) \pm \sqrt{2n(-1 \pm 2n\hat{Sp})z_{1-\alpha/2}\sqrt{\widehat{Var}(\logit(\hat{Sp}))} + n^2\left(1 \mp z_{1-\alpha/2}\sqrt{\widehat{Var}(\logit(\hat{Sp}))}\right)^2}}{2nz_{1-\alpha/2}\sqrt{\widehat{Var}(\logit(\hat{Sp}))}} \quad (1.141)$$

respectivamente.

1.3.1.5. Comparación entre intervalos (Aportación)

El análisis de la cobertura asintótica de los intervalos de confianza de la sensibilidad y de la especificidad explicados en la Sección anterior se ha realizado mediante unos

experimentos de simulación Monte Carlo, que han consistido en generar 5000 muestras aleatorias de distribuciones multinomiales cuyas probabilidades se han calculado a partir de la Tabla 1.13. Como tamaños muestrales se han considerado de 100 a 500 de 100 en 100, de 1000 a 5000 de 1000 en 1000 y por último 10000. Como valores para la sensibilidad y especificidad se han tomado los valores $(Se = 0.85, Sp = 0.75)$ y $(Se = 0.90, Sp = 0.85)$, que son valores que aparecen con cierta frecuencia en la práctica clínica. Como probabilidades de verificación se han tomado los valores $(\lambda_1 = 0.70, \lambda_0 = 0.10)$, $(\lambda_1 = 0.80, \lambda_0 = 0.25)$ y $(\lambda_1 = 0.95, \lambda_0 = 0.40)$, que se pueden considerar como valores de probabilidades de verificación bajos, intermedios y altos, respectivamente; y como prevalencia de la enfermedad se han tomado los valores 10%, 30%, 50%, 70% y 90%. Para cada 5000 muestras aleatorias de una misma distribución multinomial y para cada tamaño muestral, se han calculado la cobertura y la longitud media de cada uno de los intervalos de confianza al 95% de confianza, tanto para la sensibilidad como para la especificidad. Para todo el estudio se ha considerado que un intervalo de confianza tiene un buen rendimiento cuando su cobertura fluctúa en torno al 95%, considerando que dicha fluctuación varía en el intervalo $95\% \pm 2\%$. Todas las muestras aleatorias se han generado de tal forma que ninguna frecuencia s_i y/o r_i es igual a cero, ya que en este caso no se pueden estimar las varianzas de los estimadores de la sensibilidad y de la especificidad.

En las Tablas 1.15 a 1.44 se muestran los resultados obtenidos en los experimentos de simulación.

Tabla 1.15. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad
($\lambda_1 = 0.70$, $\lambda_0 = 0.10$) .

$Se = 0.85$ $Sp = 0.75$ $p = 10\%$										
Intervalos de confianza para la sensibilidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.1138	0.6051	0.1044	0.6005	0.5050	0.7075	0.6826	0.9082	0.5418	0.7242
200	0.5512	0.5928	0.5440	0.5904	0.7674	0.6982	0.9026	0.8699	0.8476	0.7060
300	0.7618	0.5758	0.7586	0.5741	0.8586	0.6612	0.9456	0.7671	0.8682	0.6609
400	0.7892	0.5594	0.7870	0.5581	0.8924	0.6211	0.9578	0.6761	0.9064	0.6156
500	0.8226	0.5452	0.8214	0.5441	0.9100	0.5946	0.9676	0.6062	0.9238	0.5765
1000	0.9034	0.4806	0.9030	0.4799	0.9446	0.4611	0.9842	0.4067	0.9518	0.4317
2000	0.9286	0.3814	0.9280	0.3810	0.9588	0.3342	0.9476	0.2768	0.9644	0.2992
3000	0.9402	0.3061	0.9398	0.3058	0.9640	0.2704	0.9286	0.2264	0.9702	0.2391
4000	0.9392	0.2564	0.9392	0.2562	0.9562	0.2353	0.9132	0.1973	0.9674	0.2054
5000	0.9482	0.2194	0.9478	0.2192	0.9550	0.2087	0.9190	0.1778	0.9648	0.1830
10000	0.9508	0.1407	0.9508	0.1406	0.9510	0.1480	0.9362	0.1276	0.9514	0.1291
Intervalos de confianza para la especificidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9434	0.2379	0.9314	0.2287	0.9512	0.2617	0.9664	0.2416	0.9618	0.2379
200	0.9442	0.1502	0.9336	0.1453	0.9466	0.1598	0.9626	0.1497	0.9546	0.1491
300	0.9492	0.1177	0.9418	0.1144	0.9470	0.1230	0.9532	0.1172	0.9518	0.1170
400	0.9552	0.1000	0.9514	0.0974	0.9548	0.1033	0.9596	0.0996	0.9578	0.0994
500	0.9526	0.0882	0.9488	0.0863	0.9530	0.0911	0.9562	0.0880	0.9534	0.0879
1000	0.9478	0.0610	0.9440	0.0600	0.9492	0.0622	0.9510	0.0609	0.9504	0.0609
2000	0.9498	0.0427	0.9466	0.0422	0.9530	0.0433	0.9520	0.0427	0.9516	0.0427
3000	0.9518	0.0348	0.9496	0.0345	0.9520	0.0353	0.9502	0.0348	0.9506	0.0348
4000	0.9472	0.0302	0.9446	0.0299	0.9496	0.0305	0.9474	0.0302	0.9468	0.0302
5000	0.9462	0.0270	0.9452	0.0268	0.9458	0.0273	0.9462	0.0270	0.9458	0.0270
10000	0.9524	0.0191	0.9508	0.0190	0.9548	0.0193	0.9528	0.0191	0.9532	0.0191

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.16. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad
 $(\lambda_1 = 0.70, \lambda_0 = 0.10)$.

$Se = 0.85, Sp = 0.75, p = 30\%$										
Intervalos de confianza para la sensibilidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.7020	0.5492	0.6898	0.5438	0.8288	0.6107	0.9116	0.6924	0.8404	0.6153
200	0.8376	0.5129	0.8336	0.5099	0.9104	0.5286	0.9600	0.5193	0.9148	0.5138
300	0.8852	0.4792	0.8816	0.4770	0.9372	0.4610	0.9778	0.4186	0.9394	0.4384
400	0.9064	0.4445	0.9044	0.4427	0.9486	0.4086	0.9794	0.3556	0.9502	0.3812
500	0.9260	0.4173	0.9254	0.4159	0.9590	0.3713	0.9794	0.3154	0.9610	0.3415
1000	0.9334	0.2960	0.9326	0.2951	0.9492	0.2618	0.9206	0.2217	0.9618	0.2335
2000	0.9442	0.1880	0.9436	0.1874	0.9490	0.1844	0.9232	0.1594	0.9564	0.1629
3000	0.9434	0.1455	0.9432	0.1451	0.9452	0.1510	0.9326	0.1308	0.9482	0.1325
4000	0.9514	0.1229	0.9510	0.1226	0.9504	0.1302	0.9438	0.1138	0.9540	0.1148
5000	0.9518	0.1083	0.9514	0.1081	0.9554	0.1171	0.9422	0.1020	0.9524	0.1027
10000	0.9496	0.0745	0.9496	0.0744	0.9474	0.0825	0.9414	0.0723	0.9462	0.0725
Intervalos de confianza para la especificidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9506	0.2919	0.9438	0.2832	0.9602	0.3209	0.9702	0.3033	0.9696	0.2946
200	0.9602	0.1846	0.9554	0.1798	0.9594	0.1982	0.9606	0.1841	0.9646	0.1829
300	0.9528	0.1444	0.9494	0.1412	0.9560	0.1531	0.9584	0.1438	0.9584	0.1433
400	0.9550	0.1225	0.9506	0.1200	0.9560	0.1283	0.9564	0.1219	0.9570	0.1216
500	0.9554	0.1082	0.9516	0.1063	0.9586	0.1128	0.9552	0.1078	0.9580	0.1076
1000	0.9546	0.0755	0.9520	0.0745	0.9516	0.0780	0.9536	0.0754	0.9562	0.0753
2000	0.9454	0.0532	0.9432	0.0527	0.9474	0.0547	0.9466	0.0532	0.9464	0.0532
3000	0.9532	0.0434	0.9526	0.0431	0.9556	0.0446	0.9524	0.0434	0.9510	0.0434
4000	0.9472	0.0376	0.9454	0.0373	0.9456	0.0385	0.9470	0.0375	0.9452	0.0375
5000	0.9502	0.0336	0.9490	0.0334	0.9488	0.0344	0.9474	0.0336	0.9464	0.0336
10000	0.9512	0.0237	0.9510	0.0236	0.9494	0.0243	0.9520	0.0237	0.9512	0.0237

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.17. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad
($\lambda_1 = 0.70$, $\lambda_0 = 0.10$).

$Se = 0.85, Sp = 0.75, p = 50\%$										
Intervalos de confianza para la sensibilidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.8022	0.4947	0.7920	0.4886	0.8926	0.4995	0.9402	0.5104	0.8938	0.4988
200	0.8788	0.4452	0.8754	0.4418	0.9336	0.4116	0.9674	0.3735	0.9314	0.3939
300	0.9102	0.3977	0.9074	0.3951	0.9472	0.3496	0.9642	0.3023	0.9462	0.3257
400	0.9198	0.3523	0.9186	0.3501	0.9496	0.3074	0.9396	0.2608	0.9554	0.2797
500	0.9254	0.3157	0.9238	0.3139	0.9522	0.2753	0.9196	0.2328	0.9546	0.2476
1000	0.9424	0.2028	0.9410	0.2017	0.9530	0.1923	0.9248	0.1663	0.9592	0.1709
2000	0.9496	0.1302	0.9494	0.1296	0.9498	0.1369	0.9384	0.1193	0.9526	0.1206
3000	0.9504	0.1034	0.9498	0.1031	0.9492	0.1113	0.9440	0.0977	0.9514	0.0984
4000	0.9514	0.0883	0.9500	0.0880	0.9502	0.0963	0.9470	0.0846	0.9548	0.0851
5000	0.9518	0.0784	0.9514	0.0782	0.9450	0.0860	0.9460	0.0758	0.9518	0.0761
10000	0.9522	0.0546	0.9516	0.0544	0.9444	0.0606	0.9496	0.0537	0.9528	0.0538
Intervalos de confianza para la especificidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9654	0.3635	0.9610	0.3556	0.9726	0.4024	0.9682	0.3915	0.9808	0.3721
200	0.9600	0.2378	0.9550	0.2331	0.9632	0.2583	0.9624	0.2400	0.9700	0.2365
300	0.9580	0.1856	0.9550	0.1824	0.9614	0.1987	0.9506	0.1851	0.9624	0.1838
400	0.9612	0.1575	0.9566	0.1551	0.9614	0.1679	0.9524	0.1567	0.9628	0.1560
500	0.9602	0.1398	0.9570	0.1379	0.9588	0.1483	0.9550	0.1391	0.9596	0.1387
1000	0.9468	0.0976	0.9444	0.0967	0.9456	0.1019	0.9414	0.0973	0.9446	0.0972
2000	0.9476	0.0687	0.9460	0.0682	0.9490	0.0716	0.9460	0.0686	0.9470	0.0685
3000	0.9550	0.0561	0.9534	0.0557	0.9528	0.0583	0.9548	0.0560	0.9558	0.0560
4000	0.9496	0.0484	0.9492	0.0482	0.9514	0.0503	0.9500	0.0484	0.9504	0.0484
5000	0.9500	0.0434	0.9490	0.0432	0.9526	0.0451	0.9522	0.0433	0.9520	0.0433
10000	0.9482	0.0306	0.9474	0.0305	0.9476	0.0318	0.9458	0.0306	0.9470	0.0306

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.18. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad
 $(\lambda_1 = 0.70, \lambda_0 = 0.10)$.

$Se = 0.85, Sp = 0.75, p = 70\%$										
Intervalos de confianza para la sensibilidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.8814	0.4341	0.8718	0.4270	0.9398	0.3941	0.9680	0.3797	0.9302	0.3931
200	0.8996	0.3723	0.8964	0.3683	0.9436	0.3219	0.9564	0.2851	0.9370	0.3055
300	0.9066	0.3144	0.9016	0.3114	0.9446	0.2719	0.9272	0.2646	0.9396	0.2500
400	0.9190	0.2690	0.9166	0.2666	0.9434	0.2341	0.9172	0.2033	0.9440	0.2142
500	0.9326	0.2339	0.9312	0.2319	0.9498	0.2111	0.9244	0.1835	0.9534	0.1909
1000	0.9426	0.1471	0.9404	0.1461	0.9504	0.1482	0.9360	0.1306	0.9536	0.1326
2000	0.9508	0.0980	0.9498	0.0975	0.9492	0.1055	0.9482	0.0930	0.9540	0.0936
3000	0.9486	0.0785	0.9478	0.0781	0.9546	0.0857	0.9454	0.0759	0.9470	0.0762
4000	0.9472	0.0675	0.9468	0.0673	0.9474	0.0741	0.9424	0.0658	0.9462	0.0660
5000	0.9496	0.0602	0.9490	0.0600	0.9526	0.0669	0.9486	0.0590	0.9496	0.0591
10000	0.9532	0.0421	0.9528	0.0420	0.9502	0.0474	0.9488	0.0417	0.9512	0.0418
Intervalos de confianza para la especificidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9776	0.4622	0.9750	0.4555	0.9836	0.5031	0.9504	0.5171	0.9932	0.4795
200	0.9770	0.3341	0.9750	0.3299	0.9744	0.3659	0.9508	0.3524	0.9782	0.3377
300	0.9712	0.2696	0.9690	0.2666	0.9656	0.2934	0.9462	0.2757	0.9676	0.2693
400	0.9648	0.2300	0.9624	0.2276	0.9590	0.2475	0.9440	0.2317	0.9630	0.2285
500	0.9638	0.2039	0.9622	0.2020	0.9546	0.2174	0.9462	0.2041	0.9570	0.2022
1000	0.9570	0.1414	0.9560	0.1404	0.9580	0.1499	0.9486	0.1408	0.9556	0.1403
2000	0.9526	0.0991	0.9506	0.0987	0.9560	0.1052	0.9518	0.0989	0.9516	0.0987
3000	0.9580	0.0807	0.9570	0.0803	0.9574	0.0852	0.9560	0.0805	0.9564	0.0804
4000	0.9454	0.0698	0.9452	0.0695	0.9488	0.0738	0.9444	0.0697	0.9464	0.0696
5000	0.9470	0.0622	0.9456	0.0620	0.9488	0.0660	0.9484	0.0621	0.9476	0.0621
10000	0.9532	0.0440	0.9522	0.0439	0.9530	0.0466	0.9520	0.0439	0.9522	0.0439

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.19. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad
($\lambda_1 = 0.70$, $\lambda_0 = 0.10$).

<i>Se = 0.85, Sp = 0.75, p = 90%</i>										
Intervalos de confianza para la sensibilidad										
<i>n</i>	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9794	0.3587	0.9760	0.3502	0.9918	0.2845	0.9834	0.2626	0.9906	0.2833
200	0.9678	0.2690	0.9638	0.2643	0.9814	0.2252	0.9622	0.2001	0.9800	0.2128
300	0.9536	0.2120	0.9498	0.2087	0.9722	0.1869	0.9520	0.1659	0.9666	0.1734
400	0.9458	0.1721	0.9430	0.1696	0.9622	0.1596	0.9430	0.1427	0.9600	0.1470
500	0.9356	0.1479	0.9302	0.1458	0.9596	0.1421	0.9368	0.1273	0.9498	0.1300
1000	0.9298	0.0947	0.9264	0.0936	0.9462	0.0987	0.9334	0.0896	0.9400	0.0902
2000	0.9474	0.0652	0.9458	0.0647	0.9478	0.0704	0.9440	0.0636	0.9488	0.0638
3000	0.9440	0.0529	0.9428	0.0525	0.9438	0.0575	0.9408	0.0520	0.9454	0.0521
4000	0.9496	0.0456	0.9482	0.0453	0.9500	0.0497	0.9464	0.0450	0.9474	0.0451
5000	0.9546	0.0407	0.9534	0.0405	0.9574	0.0443	0.9544	0.0403	0.9546	0.0403
10000	0.9488	0.0286	0.9478	0.0285	0.9498	0.0313	0.9494	0.0285	0.9482	0.0285

Intervalos de confianza para la especificidad										
<i>n</i>	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9952	0.5917	0.9948	0.5868	0.9988	0.6370	0.9684	0.6868	0.9998	0.6315
200	0.9952	0.5284	0.9952	0.5256	0.9928	0.5545	0.9054	0.5833	0.9942	0.5396
300	0.9932	0.4755	0.9926	0.4733	0.9804	0.5027	0.9004	0.5255	0.9790	0.4843
400	0.9886	0.4342	0.9882	0.4324	0.9764	0.4630	0.9138	0.4801	0.9774	0.4438
500	0.9816	0.4015	0.9812	0.3940	0.9690	0.4349	0.9194	0.4421	0.9696	0.4116
1000	0.9604	0.2954	0.9600	0.2946	0.9588	0.3191	0.9350	0.3058	0.9584	0.2970
2000	0.9620	0.2055	0.9616	0.2051	0.9586	0.2220	0.9444	0.2059	0.9582	0.2040
3000	0.9502	0.1669	0.9494	0.1666	0.9538	0.1803	0.9430	0.1665	0.9508	0.1655
4000	0.9514	0.1440	0.9514	0.1437	0.9502	0.1551	0.9490	0.1435	0.9514	0.1429
5000	0.9528	0.1285	0.9524	0.1283	0.9542	0.1382	0.9522	0.1280	0.9572	0.1277
10000	0.9562	0.0900	0.9560	0.0899	0.9536	0.0968	0.9574	0.0897	0.9578	0.0896

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.20. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad ($\lambda_1 = 0.80, \lambda_0 = 0.25$).

$Se = 0.85, Sp = 0.75, p = 10\%$										
Intervalos de confianza para la sensibilidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.6644	0.6007	0.6498	0.5960	0.8138	0.6792	0.9230	0.8640	0.8654	0.7073
200	0.8178	0.5549	0.8140	0.5522	0.9114	0.5815	0.9690	0.6302	0.9280	0.5920
300	0.8806	0.5194	0.8784	0.5175	0.9378	0.5071	0.9776	0.4978	0.9478	0.5056
400	0.8948	0.4891	0.8926	0.4876	0.9372	0.4508	0.9824	0.4200	0.9464	0.4435
500	0.9094	0.4598	0.9084	0.4585	0.9462	0.4072	0.9818	0.3681	0.9516	0.3954
1000	0.9356	0.3448	0.9342	0.3440	0.9582	0.2889	0.9066	0.2533	0.9644	0.2698
2000	0.9416	0.2233	0.9414	0.2228	0.9488	0.2022	0.9198	0.1810	0.9590	0.1863
3000	0.9404	0.1710	0.9400	0.1707	0.9484	0.1664	0.9298	0.1495	0.9498	0.1520
4000	0.9530	0.1431	0.9524	0.1429	0.9480	0.1438	0.9370	0.1295	0.9542	0.1311
5000	0.9566	0.1259	0.9564	0.1257	0.9488	0.1291	0.9390	0.1164	0.9552	0.1175
10000	0.9486	0.0860	0.9482	0.0859	0.9404	0.0913	0.9410	0.0827	0.9432	0.0831
Intervalos de confianza para la especificidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9472	0.1940	0.936	0.1844	0.9508	0.1963	0.9506	0.1924	0.9536	0.1913
200	0.9506	0.1337	0.9428	0.1288	0.9522	0.1346	0.9568	0.1329	0.9538	0.1325
300	0.9506	0.1080	0.9424	0.1047	0.9498	0.1086	0.9512	0.1075	0.9524	0.1073
400	0.9500	0.0932	0.9440	0.0907	0.9506	0.0936	0.9504	0.0928	0.9482	0.0927
500	0.9514	0.0831	0.9444	0.0811	0.9484	0.0835	0.9448	0.0829	0.9484	0.0828
1000	0.9526	0.0586	0.9466	0.0576	0.953	0.0588	0.9536	0.0585	0.9532	0.0585
2000	0.9536	0.0414	0.9504	0.0409	0.9542	0.0415	0.9542	0.0413	0.9538	0.0413
3000	0.9478	0.0338	0.9456	0.0334	0.9480	0.0339	0.9488	0.0338	0.9468	0.0338
4000	0.9452	0.0292	0.9426	0.0290	0.9460	0.0293	0.9442	0.0292	0.9456	0.0292
5000	0.9574	0.0262	0.9560	0.0260	0.9576	0.0263	0.9566	0.0261	0.9568	0.0261
10000	0.9544	0.0185	0.9534	0.0184	0.9548	0.0185	0.9540	0.0185	0.9544	0.0185

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.21. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad ($\lambda_1 = 0.80, \lambda_0 = 0.25$).

$Se = 0.85, Sp = 0.75, p = 30\%$										
Intervalos de confianza para la sensibilidad										
<i>n</i>	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.8630	0.5073	0.8564	0.5013	0.9216	0.4635	0.9730	0.4818	0.9352	0.4894
200	0.9228	0.4233	0.9188	0.4197	0.9534	0.3594	0.9802	0.3232	0.9614	0.3488
300	0.9294	0.3562	0.9262	0.3534	0.9544	0.2952	0.9192	0.2611	0.9624	0.2796
400	0.9316	0.3038	0.9298	0.3015	0.9604	0.2573	0.9166	0.2266	0.9660	0.2389
500	0.9422	0.2651	0.9398	0.2631	0.9558	0.2295	0.9160	0.2037	0.9702	0.2124
1000	0.9456	0.1667	0.9452	0.1656	0.9498	0.1619	0.9314	0.1461	0.9506	0.1485
2000	0.9482	0.1109	0.9478	0.1104	0.9444	0.1145	0.9374	0.1042	0.9468	0.1049
3000	0.9530	0.0887	0.9518	0.0884	0.9512	0.0937	0.9434	0.0851	0.9532	0.0855
4000	0.9586	0.0762	0.9580	0.0759	0.9548	0.0816	0.9484	0.0739	0.9564	0.0741
5000	0.9548	0.0678	0.9540	0.0676	0.9488	0.0726	0.9508	0.0662	0.9550	0.0664
10000	0.9526	0.0474	0.9522	0.0472	0.9476	0.0512	0.9462	0.0468	0.9502	0.0469
Intervalos de confianza para la especificidad										
<i>n</i>	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9514	0.2296	0.9424	0.2203	0.9512	0.2326	0.95	0.2273	0.9550	0.2255
200	0.9540	0.1575	0.9480	0.1526	0.9510	0.1591	0.9456	0.1561	0.9514	0.1556
300	0.9524	0.1277	0.9470	0.1245	0.9488	0.1289	0.9436	0.1270	0.9490	0.1267
400	0.9474	0.1104	0.9422	0.1079	0.9458	0.1115	0.9440	0.1099	0.9474	0.1097
500	0.9528	0.0985	0.9486	0.0965	0.9506	0.0995	0.9486	0.0981	0.9524	0.0980
1000	0.9518	0.0696	0.9488	0.0686	0.9494	0.0702	0.9490	0.0694	0.9508	0.0694
2000	0.9504	0.0491	0.9480	0.0486	0.9472	0.0496	0.9474	0.0491	0.9482	0.0490
3000	0.9478	0.0401	0.9464	0.0398	0.9502	0.0405	0.9472	0.0401	0.9482	0.0400
4000	0.9510	0.0347	0.9502	0.0345	0.9494	0.0351	0.9512	0.0347	0.9512	0.0347
5000	0.9422	0.0310	0.9398	0.0308	0.9414	0.0314	0.9422	0.0310	0.9424	0.0310
10000	0.9490	0.0219	0.9472	0.0219	0.9506	0.0222	0.9484	0.0219	0.9474	0.0219

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.22. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad ($\lambda_1 = 0.80, \lambda_0 = 0.25$).

<i>Se = 0.85, Sp = 0.75, p = 50%</i>										
Intervalos de confianza para la sensibilidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9006	0.4356	0.8926	0.4281	0.9432	0.3731	0.9748	0.3462	0.9456	0.3705
200	0.9292	0.3265	0.9264	0.3221	0.9552	0.2693	0.9112	0.2396	0.9616	0.2553
300	0.9396	0.2533	0.9370	0.2500	0.9574	0.2201	0.9252	0.1975	0.9662	0.2053
400	0.9402	0.2092	0.9372	0.2066	0.9496	0.1910	0.9230	0.1717	0.9576	0.1764
500	0.9400	0.1799	0.9364	0.1778	0.9506	0.1702	0.9350	0.1544	0.9564	0.1575
1000	0.9582	0.1182	0.9566	0.1172	0.9542	0.1208	0.9446	0.1100	0.9594	0.1110
2000	0.9508	0.0808	0.9498	0.0803	0.9490	0.0855	0.9442	0.0780	0.9480	0.0784
3000	0.9470	0.0654	0.9460	0.0650	0.9424	0.0694	0.9398	0.0639	0.9462	0.0640
4000	0.9436	0.0562	0.9428	0.0560	0.9458	0.0605	0.9382	0.0553	0.9392	0.0554
5000	0.9512	0.0501	0.9496	0.0499	0.9444	0.0538	0.9468	0.0495	0.9510	0.0495
10000	0.9488	0.0352	0.9484	0.0351	0.9476	0.0382	0.9486	0.0350	0.9482	0.0350
Intervalos de confianza para la especificidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9544	0.2821	0.9466	0.2732	0.9536	0.2858	0.9362	0.2790	0.9582	0.2756
200	0.9548	0.1942	0.9490	0.1894	0.9524	0.1966	0.9392	0.1921	0.9542	0.1910
300	0.9570	0.1580	0.9512	0.1548	0.9524	0.1604	0.9448	0.1566	0.9534	0.1561
400	0.9520	0.1362	0.9470	0.1338	0.9532	0.1382	0.9470	0.1353	0.9520	0.1350
500	0.9524	0.1219	0.9504	0.1199	0.9536	0.1235	0.9456	0.1212	0.9534	0.1209
1000	0.9502	0.0858	0.9474	0.0849	0.9494	0.0871	0.9498	0.0856	0.9506	0.0855
2000	0.9494	0.0606	0.9468	0.0601	0.9494	0.0615	0.9480	0.0605	0.9496	0.0604
3000	0.9478	0.0494	0.9466	0.0491	0.9478	0.0502	0.9468	0.0494	0.9482	0.0494
4000	0.9506	0.0428	0.9502	0.0427	0.9504	0.0435	0.9494	0.0428	0.9502	0.0428
5000	0.9416	0.0383	0.9402	0.0381	0.9414	0.0389	0.9422	0.0383	0.9416	0.0383
10000	0.9502	0.0271	0.9496	0.0270	0.9488	0.0275	0.9502	0.0271	0.9496	0.0270

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.23. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad ($\lambda_1 = 0.80, \lambda_0 = 0.25$).

<i>Se = 0.85, Sp = 0.75, p = 70%</i>										
Intervalos de confianza para la sensibilidad										
<i>n</i>	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9118	0.3623	0.9024	0.3540	0.9470	0.2948	0.9428	0.2702	0.9486	0.2901
200	0.9300	0.2455	0.9254	0.2406	0.9494	0.2100	0.9256	0.1920	0.9578	0.1998
300	0.9406	0.1869	0.9374	0.1835	0.9510	0.1721	0.9266	0.1575	0.9566	0.1612
400	0.9404	0.1551	0.9366	0.1525	0.9502	0.1493	0.9354	0.1373	0.9502	0.1395
500	0.9476	0.1354	0.9428	0.1333	0.9526	0.1338	0.9356	0.1231	0.9548	0.1245
1000	0.9550	0.0914	0.9520	0.0903	0.9534	0.0947	0.9442	0.0873	0.9518	0.0878
2000	0.9512	0.0632	0.9496	0.0627	0.9450	0.0667	0.9468	0.0618	0.9494	0.0620
3000	0.9462	0.0513	0.9452	0.0509	0.9452	0.0545	0.9432	0.0505	0.9456	0.0506
4000	0.9522	0.0443	0.9506	0.0440	0.9538	0.0473	0.9524	0.0438	0.9512	0.0438
5000	0.9476	0.0395	0.9468	0.0393	0.9454	0.0425	0.9482	0.0392	0.9474	0.0392
10000	0.9494	0.0278	0.9484	0.0277	0.9462	0.0298	0.9466	0.0277	0.9480	0.0277
Intervalos de confianza para la especificidad										
<i>n</i>	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9610	0.3806	0.9556	0.3728	0.9710	0.3874	0.9302	0.3864	0.9778	0.3747
200	0.9624	0.2659	0.9590	0.2613	0.9608	0.2694	0.9358	0.2631	0.9610	0.2602
300	0.9582	0.2161	0.9546	0.2129	0.9584	0.2198	0.9430	0.2140	0.9576	0.2124
400	0.9582	0.1865	0.9550	0.1841	0.9540	0.1896	0.9414	0.1847	0.9532	0.1837
500	0.9624	0.1665	0.9592	0.1645	0.9560	0.1697	0.9456	0.1651	0.9554	0.1644
1000	0.9540	0.1169	0.9520	0.1159	0.9528	0.1194	0.9490	0.1163	0.9534	0.1161
2000	0.9532	0.0824	0.9520	0.0819	0.9550	0.0843	0.9472	0.0822	0.9516	0.0821
3000	0.9528	0.0673	0.9514	0.0669	0.9534	0.0688	0.9526	0.0671	0.9528	0.0671
4000	0.9482	0.0582	0.9466	0.0580	0.9472	0.0596	0.9460	0.0581	0.9482	0.0581
5000	0.9488	0.0520	0.9480	0.0518	0.9526	0.0535	0.9512	0.0520	0.9498	0.0519
10000	0.9502	0.0368	0.9496	0.0367	0.9540	0.0376	0.9512	0.0367	0.9512	0.0367

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.24. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad ($\lambda_1 = 0.80, \lambda_0 = 0.25$).

<i>Se = 0.85, Sp = 0.75, p = 90%</i>										
Intervalos de confianza para la sensibilidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9544	0.2688	0.9446	0.2591	0.9770	0.2194	0.9520	0.2037	0.9740	0.2148
200	0.9426	0.1674	0.9354	0.1622	0.9554	0.1529	0.9356	0.1432	0.9566	0.1463
300	0.9392	0.1204	0.9346	0.1259	0.9498	0.1247	0.9328	0.1176	0.9514	0.1190
400	0.9518	0.1089	0.9466	0.1063	0.9528	0.1080	0.9460	0.1019	0.9538	0.1027
500	0.9454	0.0957	0.9412	0.0937	0.9498	0.0961	0.9424	0.0909	0.9478	0.0915
1000	0.9502	0.0660	0.9476	0.0650	0.9550	0.0678	0.9488	0.0645	0.9538	0.0646
2000	0.9552	0.0461	0.9520	0.0456	0.9512	0.0480	0.9528	0.0457	0.9518	0.0456
3000	0.9532	0.0375	0.9512	0.0372	0.9504	0.0392	0.9518	0.0372	0.9532	0.0373
4000	0.9514	0.0324	0.9498	0.0322	0.9502	0.0341	0.9516	0.0322	0.9516	0.0323
5000	0.9524	0.0290	0.9512	0.0288	0.9518	0.0303	0.9508	0.0289	0.9538	0.0289
10000	0.9518	0.0204	0.9502	0.0203	0.9510	0.0215	0.9490	0.0204	0.9506	0.0204
Intervalos de confianza para la especificidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9822	0.5710	0.9804	0.5659	0.9920	0.6025	0.9558	0.6829	0.9966	0.6114
200	0.9704	0.4787	0.9696	0.4755	0.9810	0.4899	0.9156	0.5189	0.9872	0.4851
300	0.9698	0.4065	0.9650	0.4040	0.9750	0.4136	0.9204	0.4211	0.9806	0.4038
400	0.9616	0.3540	0.9606	0.3519	0.9648	0.3608	0.9246	0.3590	0.9660	0.3491
500	0.9612	0.3180	0.9602	0.3163	0.9606	0.3239	0.9276	0.3200	0.9624	0.3132
1000	0.9576	0.2242	0.9570	0.2232	0.9608	0.2297	0.9468	0.2232	0.9578	0.2212
2000	0.9542	0.1569	0.9538	0.1564	0.9546	0.1618	0.9456	0.1559	0.9536	0.1553
3000	0.9576	0.1277	0.9570	0.1274	0.9556	0.1320	0.9466	0.1271	0.9558	0.1268
4000	0.9534	0.1104	0.9530	0.1101	0.9514	0.1142	0.9470	0.1099	0.9486	0.1097
5000	0.9510	0.0985	0.9504	0.0983	0.9522	0.1020	0.9486	0.0982	0.9504	0.0980
10000	0.9532	0.0695	0.9528	0.0694	0.9474	0.0722	0.9508	0.0694	0.9530	0.0694

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.25. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad
 $(\lambda_1 = 0.95, \lambda_0 = 0.40)$

<i>Se = 0.85, Sp = 0.75, p = 10%</i>										
Intervalos de confianza para la sensibilidad										
<i>n</i>	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
	<i>Cob.</i>	<i>Long.</i>	<i>Cob.</i>	<i>Long.</i>	<i>Cob.</i>	<i>Long.</i>	<i>Cob.</i>	<i>Long.</i>	<i>Cob.</i>	<i>Long.</i>
100	0.7800	0.5805	0.7734	0.5755	0.8836	0.6183	0.9614	0.7321	0.9132	0.6481
200	0.8830	0.5201	0.8782	0.5171	0.9356	0.4940	0.9826	0.4953	0.9496	0.5037
300	0.9110	0.4705	0.9088	0.4683	0.9512	0.4478	0.9874	0.3827	0.9624	0.4094
400	0.9216	0.4264	0.9200	0.4246	0.9576	0.3554	0.9784	0.3262	0.9618	0.3513
500	0.9240	0.3907	0.9232	0.3891	0.9538	0.3186	0.9334	0.2884	0.9630	0.3104
1000	0.9438	0.2638	0.9430	0.2628	0.9574	0.2233	0.9136	0.2036	0.9662	0.2118
2000	0.9484	0.1666	0.9474	0.1660	0.9550	0.1583	0.9356	0.1464	0.9534	0.1488
3000	0.9576	0.1306	0.9576	0.1303	0.9534	0.1292	0.9448	0.1203	0.9538	0.1215
4000	0.9514	0.1110	0.9510	0.1107	0.9494	0.1115	0.9408	0.1043	0.9528	0.1050
5000	0.9518	0.0982	0.9514	0.0980	0.9510	0.1001	0.9382	0.0935	0.9492	0.0940
10000	0.9526	0.0680	0.9526	0.0679	0.9524	0.0709	0.9498	0.0663	0.9534	0.0665
Intervalos de confianza para la especificidad										
<i>n</i>	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
	<i>Cob.</i>	<i>Long.</i>	<i>Cob.</i>	<i>Long.</i>	<i>Cob.</i>	<i>Long.</i>	<i>Cob.</i>	<i>Long.</i>	<i>Cob.</i>	<i>Long.</i>
100	0.9524	0.1847	0.9378	0.1751	0.9516	0.1836	0.9466	0.1827	0.9528	0.1819
200	0.9564	0.1288	0.9462	0.1239	0.9524	0.1285	0.9436	0.1280	0.9518	0.1277
300	0.9540	0.1049	0.9464	0.1016	0.9534	0.1047	0.9542	0.1044	0.9540	0.1043
400	0.9472	0.0906	0.9410	0.0882	0.9486	0.0905	0.9474	0.0903	0.9482	0.0902
500	0.9528	0.0809	0.9458	0.0789	0.9526	0.0808	0.9518	0.0807	0.9538	0.0806
1000	0.9558	0.0571	0.9498	0.0561	0.9536	0.0571	0.9520	0.0570	0.9538	0.0570
2000	0.9556	0.0404	0.9532	0.0399	0.9560	0.0404	0.9556	0.0404	0.9554	0.0404
3000	0.9510	0.0330	0.9490	0.0326	0.9514	0.0330	0.9492	0.0330	0.9512	0.0330
4000	0.9532	0.0286	0.9510	0.0283	0.9526	0.0286	0.9524	0.0286	0.9524	0.0285
5000	0.9474	0.0255	0.9458	0.0253	0.9466	0.0256	0.9480	0.0255	0.9478	0.0255
10000	0.9548	0.0181	0.9534	0.0180	0.9550	0.0181	0.9548	0.0181	0.9552	0.0181

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.26. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad
($\lambda_1 = 0.95, \lambda_0 = 0.40$)

<i>Se = 0.85, Sp = 0.75, p = 30%</i>										
Intervalos de confianza para la sensibilidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9060	0.4643	0.8974	0.4576	0.9464	0.3999	0.9830	0.3773	0.9574	0.4030
200	0.9320	0.3502	0.9274	0.3459	0.9578	0.2828	0.9160	0.2582	0.9636	0.2752
300	0.9404	0.2762	0.9370	0.2730	0.9586	0.2310	0.9184	0.2113	0.9680	0.2207
400	0.9404	0.2280	0.9374	0.2254	0.9502	0.2001	0.9182	0.1841	0.9624	0.1897
500	0.9446	0.1963	0.9428	0.1943	0.9528	0.1788	0.9322	0.1660	0.9596	0.1697
1000	0.9560	0.1282	0.9534	0.1272	0.9536	0.1262	0.9382	0.1183	0.9564	0.1194
2000	0.9548	0.0874	0.9536	0.0869	0.9542	0.0896	0.9476	0.0840	0.9532	0.0844
3000	0.9492	0.0705	0.9488	0.0701	0.9466	0.0731	0.9440	0.0686	0.9486	0.0688
4000	0.9552	0.0607	0.9544	0.0604	0.9552	0.0636	0.9478	0.0595	0.9534	0.0596
5000	0.9494	0.0540	0.9478	0.0538	0.9548	0.0568	0.9468	0.0532	0.9498	0.0533
10000	0.9494	0.0380	0.9490	0.0379	0.9520	0.0402	0.9488	0.0377	0.9498	0.0377
Intervalos de confianza para la especificidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9534	0.2120	0.9444	0.2025	0.9496	0.2107	0.9418	0.2094	0.9492	0.2081
200	0.9532	0.1484	0.9452	0.1435	0.9538	0.1479	0.9480	0.1473	0.9544	0.1468
300	0.9562	0.1205	0.9496	0.1173	0.9528	0.1203	0.9452	0.1198	0.9520	0.1196
400	0.9494	0.1044	0.9440	0.1020	0.9476	0.1044	0.9432	0.1040	0.9470	0.1038
500	0.9468	0.0933	0.9420	0.0913	0.9464	0.0933	0.9428	0.0929	0.9458	0.0928
1000	0.9508	0.0659	0.9476	0.0650	0.9496	0.0660	0.9482	0.0658	0.9494	0.0658
2000	0.9536	0.0466	0.9512	0.0461	0.9522	0.0467	0.9504	0.0465	0.9514	0.0465
3000	0.9508	0.0380	0.9496	0.0377	0.9498	0.0381	0.9480	0.0380	0.9498	0.0380
4000	0.9512	0.0329	0.9482	0.0327	0.9506	0.0330	0.9500	0.0329	0.9510	0.0329
5000	0.9468	0.0294	0.9454	0.0292	0.9478	0.0295	0.9468	0.0294	0.9456	0.0294
10000	0.9470	0.0208	0.9464	0.0207	0.9452	0.0208	0.9462	0.0208	0.9472	0.0208

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.27. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad
 $(\lambda_1 = 0.95, \lambda_0 = 0.40)$

<i>Se = 0.85, Sp = 0.75, p = 50%</i>										
Intervalos de confianza para la sensibilidad										
<i>n</i>	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9264	0.3737	0.9176	0.3655	0.9544	0.2994	0.9376	0.2760	0.9630	0.2963
200	0.9346	0.2514	0.9298	0.2465	0.9528	0.2123	0.9130	0.1962	0.9608	0.2037
300	0.9450	0.1902	0.9416	0.1867	0.9548	0.1731	0.9326	0.1616	0.9582	0.1651
400	0.9518	0.1587	0.9496	0.1561	0.9534	0.1504	0.9330	0.1404	0.9566	0.1426
500	0.9514	0.1385	0.9496	0.1364	0.9500	0.1345	0.9368	0.1259	0.9530	0.1274
1000	0.9560	0.0937	0.9540	0.0927	0.9562	0.0950	0.9470	0.0895	0.9528	0.0900
2000	0.9510	0.0649	0.9486	0.0644	0.9524	0.0674	0.9500	0.0635	0.9520	0.0637
3000	0.9472	0.0526	0.9466	0.0523	0.9462	0.0549	0.9462	0.0519	0.9492	0.0519
4000	0.9492	0.0455	0.9482	0.0452	0.9488	0.0477	0.9474	0.0449	0.9492	0.0450
5000	0.9464	0.0406	0.9450	0.0403	0.9484	0.0426	0.9478	0.0402	0.9478	0.0402
10000	0.9438	0.0286	0.9432	0.0285	0.9418	0.0301	0.9440	0.0284	0.9444	0.0284
Intervalos de confianza para la especificidad										
<i>n</i>	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9484	0.2564	0.9430	0.2472	0.9498	0.2538	0.9356	0.2524	0.9530	0.2501
200	0.9586	0.1788	0.9518	0.1740	0.9550	0.1781	0.9458	0.1769	0.9564	0.1761
300	0.9538	0.1458	0.9492	0.1425	0.9530	0.1454	0.9466	0.1446	0.9518	0.1442
400	0.9564	0.1260	0.9528	0.1235	0.9540	0.1260	0.9460	0.1252	0.9530	0.1249
500	0.9510	0.1127	0.9462	0.1107	0.9488	0.1129	0.9434	0.1122	0.9492	0.1120
1000	0.9552	0.0795	0.9524	0.0785	0.9564	0.0797	0.9536	0.0793	0.9558	0.0792
2000	0.9516	0.0562	0.9500	0.0557	0.9512	0.0565	0.9492	0.0561	0.9514	0.0561
3000	0.9530	0.0459	0.9508	0.0455	0.9518	0.0460	0.9512	0.0458	0.9534	0.0458
4000	0.9544	0.0397	0.9524	0.0395	0.9528	0.0399	0.9536	0.0397	0.9544	0.0397
5000	0.9534	0.0355	0.9524	0.0353	0.9536	0.0357	0.9530	0.0355	0.9534	0.0355
10000	0.9490	0.0251	0.9480	0.0250	0.9484	0.0252	0.9470	0.0251	0.9482	0.0251

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.28. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad
($\lambda_1 = 0.95, \lambda_0 = 0.40$)

<i>Se = 0.85, Sp = 0.75, p = 70%</i>										
Intervalos de confianza para la sensibilidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9372	0.2972	0.9276	0.2879	0.9584	0.2392	0.9218	0.2236	0.9642	0.2357
200	0.9448	0.1878	0.9372	0.1825	0.9556	0.1696	0.9308	0.1593	0.9596	0.1628
300	0.9440	0.1458	0.9390	0.1423	0.9470	0.1383	0.9302	0.1309	0.9500	0.1327
400	0.9470	0.1228	0.9416	0.1202	0.9416	0.1196	0.9344	0.1137	0.9454	0.1148
500	0.9530	0.1081	0.9490	0.1060	0.9508	0.1071	0.9450	0.1017	0.9592	0.1025
1000	0.9468	0.0744	0.9448	0.0734	0.9492	0.0760	0.9474	0.0722	0.9482	0.0724
2000	0.9490	0.0519	0.9470	0.0513	0.9542	0.0538	0.9476	0.0511	0.9504	0.0512
3000	0.9518	0.0421	0.9502	0.0419	0.9498	0.0439	0.9484	0.0417	0.9510	0.0418
4000	0.9518	0.0364	0.9506	0.0362	0.9522	0.0380	0.9492	0.0361	0.9522	0.0362
5000	0.9480	0.0325	0.9464	0.0323	0.9492	0.0340	0.9468	0.0323	0.9464	0.0324
10000	0.9604	0.0229	0.9588	0.0228	0.9538	0.0241	0.9600	0.0229	0.9606	0.0229
Intervalos de confianza para la especificidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9580	0.3398	0.9516	0.3315	0.9688	0.3344	0.9238	0.3346	0.9686	0.3293
200	0.9612	0.2394	0.9592	0.2348	0.9574	0.2379	0.9344	0.2362	0.9524	0.2343
300	0.9584	0.1944	0.9550	0.1912	0.9538	0.1935	0.9398	0.1922	0.9556	0.1911
400	0.9584	0.1684	0.9552	0.1660	0.9568	0.1682	0.9456	0.1670	0.9572	0.1663
500	0.9564	0.1501	0.9552	0.1481	0.9544	0.1502	0.9474	0.1490	0.9542	0.1485
1000	0.9504	0.1057	0.9494	0.1047	0.9496	0.1061	0.9448	0.1052	0.9458	0.1050
2000	0.9540	0.0747	0.9532	0.0742	0.9546	0.0753	0.9490	0.0746	0.9524	0.0745
3000	0.9544	0.0609	0.9536	0.0605	0.9534	0.0613	0.9540	0.0608	0.9538	0.0607
4000	0.9470	0.0527	0.9460	0.0525	0.9480	0.0532	0.9470	0.0527	0.9458	0.0527
5000	0.9522	0.0471	0.9508	0.0469	0.9546	0.0475	0.9536	0.0471	0.9532	0.0471
10000	0.9468	0.0333	0.9462	0.0332	0.9462	0.0336	0.9444	0.0333	0.9462	0.0333

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.29. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad
($\lambda_1 = 0.95, \lambda_0 = 0.40$)

<i>Se = 0.85, Sp = 0.75, p = 90%</i>										
Intervalos de confianza para la sensibilidad										
<i>n</i>	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9496	0.2189	0.9366	0.2087	0.9666	0.1858	0.9384	0.1766	0.9680	0.1825
200	0.9460	0.1381	0.9404	0.1329	0.9546	0.1302	0.9420	0.1252	0.9554	0.1267
300	0.9502	0.1093	0.9424	0.1058	0.9534	0.1063	0.9480	0.1026	0.9554	0.1034
400	0.9510	0.0932	0.9444	0.0906	0.9512	0.0918	0.9432	0.0889	0.9526	0.0894
500	0.9510	0.0824	0.9466	0.0804	0.9490	0.0820	0.9434	0.0795	0.9480	0.0798
1000	0.9512	0.0572	0.9458	0.0562	0.9478	0.0579	0.9488	0.0562	0.9496	0.0563
2000	0.9526	0.0401	0.9490	0.0396	0.9514	0.0410	0.9506	0.0398	0.9534	0.0398
3000	0.9526	0.0327	0.9512	0.0323	0.9530	0.0334	0.9514	0.0325	0.9534	0.0325
4000	0.9496	0.0283	0.9464	0.0280	0.9472	0.0290	0.9468	0.0281	0.9498	0.0282
5000	0.9460	0.0253	0.9434	0.0251	0.9426	0.0258	0.9430	0.0252	0.9448	0.0252
10000	0.9470	0.0178	0.9450	0.0177	0.9466	0.0183	0.9476	0.0178	0.9472	0.0178
Intervalos de confianza para la especificidad										
<i>n</i>	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9688	0.5448	0.9664	0.5393	0.9872	0.5607	0.9434	0.6309	0.9922	0.5737
200	0.9558	0.4321	0.9536	0.4286	0.9748	0.4309	0.9176	0.4476	0.9802	0.4297
300	0.9546	0.3576	0.9538	0.3549	0.9638	0.3525	0.9208	0.3565	0.9648	0.3489
400	0.9530	0.3093	0.9522	0.3071	0.9604	0.3063	0.9244	0.3067	0.9574	0.3019
500	0.9558	0.2772	0.9540	0.2754	0.9586	0.2761	0.9298	0.2749	0.9566	0.2715
1000	0.9562	0.1948	0.9554	0.1938	0.9590	0.1955	0.9452	0.1930	0.9560	0.1919
2000	0.9536	0.1370	0.9528	0.1365	0.9504	0.1385	0.9500	0.1362	0.9528	0.1358
3000	0.9458	0.1116	0.9454	0.1113	0.9462	0.1129	0.9430	0.1111	0.9450	0.1109
4000	0.9484	0.0964	0.9480	0.0961	0.9484	0.0976	0.9424	0.0960	0.9470	0.0959
5000	0.9542	0.0863	0.9534	0.0861	0.9552	0.0874	0.9534	0.0861	0.9548	0.0860
10000	0.9522	0.0608	0.9518	0.0607	0.9524	0.0617	0.9502	0.0607	0.9528	0.0607

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.30. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad ($\lambda_1 = 0.70, \lambda_0 = 0.10$).

$Se = 0.90, Sp = 0.85, p = 10\%$										
Intervalos de confianza para la sensibilidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.0064	0.6080	0.0056	0.6034	0.2990	0.7147	0.6562	0.9236	0.2540	0.7313
200	0.1138	0.5959	0.1082	0.5935	0.6292	0.6981	0.9134	0.8685	0.7576	0.7078
300	0.4760	0.5808	0.4664	0.5792	0.7728	0.6624	0.9466	0.7596	0.8580	0.6610
400	0.7124	0.5655	0.7106	0.5642	0.8332	0.6222	0.9694	0.6641	0.8424	0.6137
500	0.7636	0.5523	0.7630	0.5512	0.8682	0.5863	0.9722	0.5901	0.8716	0.5725
1000	0.8608	0.4978	0.8602	0.4972	0.9254	0.4529	0.9856	0.3839	0.9358	0.4239
2000	0.9074	0.4131	0.9072	0.4127	0.9516	0.3170	0.9888	0.2482	0.9558	0.2851
3000	0.9218	0.3448	0.9212	0.3445	0.9534	0.2517	0.9354	0.1970	0.9616	0.2220
4000	0.9278	0.2896	0.9276	0.2894	0.9546	0.2150	0.9130	0.1697	0.9634	0.1867
5000	0.9346	0.2445	0.9344	0.2443	0.9558	0.1894	0.9106	0.1527	0.9656	0.1644
10000	0.9446	0.1399	0.9442	0.1398	0.9554	0.1301	0.9284	0.1097	0.9586	0.1134
Intervalos de confianza para la especificidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9140	0.2155	0.8978	0.2054	0.9372	0.2118	0.9582	0.1903	0.9482	0.1928
200	0.9350	0.1303	0.9256	0.1250	0.9450	0.1292	0.9548	0.1214	0.9500	0.1224
300	0.9390	0.1011	0.9318	0.0977	0.9466	0.1012	0.9546	0.0965	0.9504	0.0970
400	0.9460	0.0852	0.9392	0.0827	0.9480	0.0855	0.9548	0.0823	0.9537	0.0826
500	0.9454	0.0751	0.9388	0.0731	0.9460	0.0755	0.9548	0.0731	0.9518	0.0733
1000	0.9456	0.0516	0.9404	0.0506	0.9464	0.0519	0.9506	0.0508	0.9494	0.0509
2000	0.9514	0.0359	0.9490	0.0354	0.9502	0.0362	0.9514	0.0357	0.9508	0.0357
3000	0.9562	0.0292	0.9530	0.0289	0.9534	0.0295	0.9512	0.0291	0.9526	0.0291
4000	0.9458	0.0252	0.9424	0.0250	0.9488	0.0254	0.9438	0.0252	0.9440	0.0252
5000	0.9524	0.0226	0.9498	0.0224	0.9528	0.0228	0.9510	0.0225	0.9524	0.0225
10000	0.9484	0.0159	0.9478	0.0158	0.9484	0.0161	0.9486	0.0159	0.9486	0.0159

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.31. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad ($\lambda_1 = 0.70, \lambda_0 = 0.10$).

<i>Se = 0.90, Sp = 0.85, p = 30%</i>										
Intervalos de confianza para la sensibilidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.4468	0.5567	0.4262	0.5514	0.7398	0.6130	0.9190	0.6907	0.7976	0.6185
200	0.7626	0.5264	0.7592	0.5236	0.8864	0.5244	0.9708	0.5058	0.8962	0.5124
300	0.8330	0.4966	0.8304	0.4945	0.9198	0.4518	0.9830	0.3985	0.9270	0.4325
400	0.8594	0.4698	0.8568	0.4681	0.9274	0.4023	0.9830	0.3343	0.9346	0.3754
500	0.8870	0.4453	0.8852	0.4438	0.9418	0.3614	0.9838	0.2904	0.9466	0.3318
1000	0.9242	0.3380	0.9230	0.3371	0.9578	0.2492	0.9424	0.1939	0.9634	0.2183
2000	0.9394	0.2079	0.9388	0.2074	0.9534	0.1677	0.9168	0.1373	0.9668	0.1459
3000	0.9408	0.1485	0.9400	0.1481	0.9492	0.1347	0.9276	0.1136	0.9602	0.1179
4000	0.9430	0.1197	0.9422	0.1194	0.9460	0.1165	0.9292	0.0987	0.9490	0.1014
5000	0.9482	0.1029	0.9474	0.1026	0.9500	0.1041	0.9374	0.0887	0.9534	0.0905
10000	0.9534	0.0677	0.9532	0.0676	0.9528	0.0732	0.9488	0.0631	0.9562	0.0637
Intervalos de confianza para la especificidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9232	0.2763	0.9096	0.2668	0.9460	0.2652	0.9596	0.2361	0.9632	0.2401
200	0.9378	0.1626	0.9312	0.1574	0.9458	0.1603	0.9598	0.1473	0.9550	0.1490
300	0.9450	0.1240	0.9406	0.1206	0.9454	0.1232	0.9490	0.1158	0.9540	0.1167
400	0.9522	0.1039	0.9460	0.1014	0.9506	0.1040	0.9538	0.0987	0.9578	0.0993
500	0.9554	0.0914	0.9496	0.0893	0.9520	0.0919	0.9506	0.0877	0.9542	0.0881
1000	0.9514	0.0624	0.9468	0.0614	0.9516	0.0632	0.9478	0.0611	0.9496	0.0612
2000	0.9494	0.0435	0.9474	0.0430	0.9454	0.0442	0.9462	0.0431	0.9492	0.0431
3000	0.9488	0.0354	0.9470	0.0351	0.9456	0.0360	0.9470	0.0352	0.9498	0.0352
4000	0.9504	0.0306	0.9490	0.0304	0.9508	0.0312	0.9514	0.0305	0.9518	0.0305
5000	0.9502	0.0274	0.9490	0.0272	0.9506	0.0279	0.9490	0.0273	0.9522	0.0273
10000	0.9488	0.0193	0.9476	0.0192	0.9500	0.0197	0.9482	0.0193	0.9476	0.0193

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.32. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad ($\lambda_1 = 0.70, \lambda_0 = 0.10$).

$Se = 0.90, Sp = 0.85, p = 50\%$										
Intervalos de confianza para la sensibilidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.6906	0.5094	0.6774	0.5034	0.8402	0.4987	0.9476	0.5030	0.8486	0.5006
200	0.8332	0.4710	0.8276	0.4677	0.9104	0.4072	0.9708	0.3576	0.9174	0.3920
300	0.8834	0.4296	0.8794	0.4272	0.9428	0.3424	0.9828	0.2813	0.9460	0.3191
400	0.8978	0.3937	0.8952	0.3917	0.9458	0.2972	0.9782	0.2370	0.9480	0.2708
500	0.9046	0.3578	0.9024	0.3560	0.9468	0.2652	0.9666	0.2096	0.9490	0.2374
1000	0.9310	0.2319	0.9296	0.2308	0.9520	0.1805	0.9116	0.1468	0.9604	0.1577
2000	0.9452	0.1324	0.9444	0.1319	0.9554	0.1240	0.9276	0.1049	0.9562	0.1083
3000	0.9504	0.0996	0.9500	0.0992	0.9504	0.1003	0.9392	0.0863	0.9534	0.0881
4000	0.9504	0.0831	0.9502	0.0828	0.9520	0.0872	0.9352	0.0750	0.9518	0.0760
5000	0.9480	0.0728	0.9472	0.0726	0.9496	0.0779	0.9394	0.0671	0.9460	0.0679
10000	0.9484	0.0495	0.9478	0.0494	0.9468	0.0545	0.9426	0.0476	0.9474	0.0479
Intervalos de confianza para la especificidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9160	0.3569	0.9060	0.3485	0.9532	0.3411	0.9550	0.3048	0.9694	0.3099
200	0.9396	0.2145	0.9312	0.2094	0.9550	0.2080	0.9512	0.1871	0.9616	0.1900
300	0.9502	0.1595	0.9452	0.1560	0.9528	0.1571	0.9536	0.1444	0.9608	0.1460
400	0.9492	0.1320	0.9454	0.1295	0.9542	0.1310	0.9488	0.1223	0.9600	0.1234
500	0.9526	0.1158	0.9484	0.1135	0.9538	0.1158	0.9462	0.1087	0.9536	0.1095
1000	0.9562	0.0783	0.9546	0.0773	0.9506	0.0795	0.9530	0.0759	0.9570	0.0762
2000	0.9466	0.0542	0.9434	0.0537	0.9456	0.0555	0.9430	0.0534	0.9466	0.0535
3000	0.9542	0.0440	0.9516	0.0437	0.9554	0.0450	0.9492	0.0436	0.9536	0.0436
4000	0.9544	0.0380	0.9536	0.0378	0.9536	0.0390	0.9538	0.0377	0.9530	0.0378
5000	0.9492	0.0340	0.9474	0.0339	0.9514	0.0349	0.9464	0.0338	0.9478	0.0338
10000	0.9540	0.0240	0.9524	0.0238	0.9554	0.0246	0.9496	0.0239	0.9520	0.0239

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.33. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad ($\lambda_1 = 0.70, \lambda_0 = 0.10$).

<i>Se = 0.90, Sp = 0.85, p = 70%</i>										
Intervalos de confianza para la sensibilidad										
<i>n</i>	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.7996	0.4525	0.7850	0.4455	0.8992	0.3845	0.9616	0.3652	0.8978	0.3888
200	0.8608	0.4054	0.8542	0.4015	0.9238	0.3127	0.9638	0.2682	0.9154	0.3009
300	0.8992	0.3613	0.8936	0.3584	0.9432	0.2659	0.9640	0.2173	0.9400	0.2461
400	0.9066	0.3173	0.9038	0.3149	0.9436	0.2296	0.9444	0.1858	0.9466	0.2084
500	0.9166	0.2785	0.9144	0.2765	0.9514	0.2033	0.9238	0.1665	0.9512	0.1835
1000	0.9368	0.1636	0.9350	0.1625	0.9532	0.1397	0.9274	0.1184	0.9612	0.1239
2000	0.9388	0.0973	0.9376	0.0967	0.9434	0.0983	0.9356	0.0846	0.9484	0.0863
3000	0.9522	0.0756	0.9508	0.0753	0.9448	0.0795	0.9368	0.0692	0.9526	0.0701
4000	0.9490	0.0641	0.9482	0.0638	0.9456	0.0687	0.9382	0.0600	0.9476	0.0606
5000	0.9444	0.0566	0.9432	0.0564	0.9444	0.0617	0.9424	0.0538	0.9452	0.0542
10000	0.9548	0.0390	0.9536	0.0389	0.9552	0.0434	0.9520	0.0381	0.9566	0.0382
Intervalos de confianza para la especificidad										
<i>n</i>	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9228	0.4731	0.9160	0.4664	0.9668	0.4599	0.9666	0.4282	0.9834	0.4309
200	0.9412	0.3195	0.9370	0.3150	0.9668	0.3002	0.9486	0.2694	0.9784	0.2737
300	0.9438	0.2375	0.9394	0.2342	0.9594	0.2259	0.9396	0.2037	0.9658	0.2070
400	0.9528	0.1933	0.9506	0.1907	0.9608	0.1877	0.9464	0.1706	0.9658	0.1730
500	0.9554	0.1675	0.9534	0.1654	0.9626	0.1637	0.9432	0.1507	0.9620	0.1526
1000	0.9568	0.1102	0.9554	0.1092	0.9586	0.1107	0.9478	0.1041	0.9592	0.1048
2000	0.9514	0.0755	0.9496	0.0750	0.9502	0.0773	0.9464	0.0733	0.9520	0.0736
3000	0.9502	0.0609	0.9494	0.0606	0.9482	0.0626	0.9440	0.0597	0.9482	0.0599
4000	0.9540	0.0524	0.9524	0.0522	0.9524	0.0540	0.9504	0.0517	0.9530	0.0517
5000	0.9488	0.0468	0.9476	0.0466	0.9496	0.0483	0.9442	0.0463	0.9478	0.0463
10000	0.9480	0.0329	0.9472	0.0328	0.9486	0.0342	0.9500	0.0327	0.9502	0.0327

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.34. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad ($\lambda_1 = 0.70, \lambda_0 = 0.10$).

<i>Se = 0.90, Sp = 0.85, p = 90%</i>										
Intervalos de confianza para la sensibilidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9654	0.3787	0.9586	0.3703	0.9866	0.2591	0.9942	0.2323	0.9870	0.2625
200	0.9448	0.3057	0.9382	0.3009	0.9784	0.2118	0.9740	0.1823	0.9714	0.2042
300	0.9292	0.2566	0.9244	0.2532	0.9642	0.1815	0.9448	0.1539	0.9576	0.1703
400	0.9226	0.2124	0.9174	0.2098	0.9514	0.1584	0.9306	0.1349	0.9436	0.1461
500	0.9140	0.1791	0.9100	0.1770	0.9460	0.1413	0.9228	0.1210	0.9204	0.1289
1000	0.9256	0.1030	0.9230	0.1019	0.9454	0.0978	0.9272	0.0863	0.9396	0.0884
2000	0.9454	0.0660	0.9428	0.0655	0.9462	0.0692	0.9446	0.0614	0.9510	0.0620
3000	0.9474	0.0525	0.9458	0.0522	0.9446	0.0563	0.9484	0.0501	0.9514	0.0505
4000	0.9560	0.0451	0.9552	0.0448	0.9538	0.0492	0.9488	0.0436	0.9542	0.0438
5000	0.9540	0.0400	0.9528	0.0398	0.9488	0.0435	0.9460	0.0389	0.9516	0.0391
10000	0.9436	0.0279	0.9434	0.0278	0.9472	0.0308	0.9450	0.0275	0.9442	0.0276
Intervalos de confianza para la especificidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9480	0.6008	0.9438	0.5960	0.9832	0.6336	0.9994	0.6613	0.9946	0.6283
200	0.9516	0.5351	0.9502	0.5323	0.9850	0.5169	0.9726	0.4906	0.9948	0.4956
300	0.9528	0.4759	0.9520	0.4737	0.9826	0.4413	0.9326	0.4108	0.9938	0.4165
400	0.9498	0.4262	0.9478	0.4243	0.9838	0.3902	0.9086	0.3622	0.9924	0.3650
500	0.9452	0.3810	0.9446	0.3793	0.9806	0.3530	0.9184	0.3247	0.9842	0.3262
1000	0.9494	0.2487	0.9488	0.2477	0.9610	0.2365	0.9300	0.2151	0.9652	0.2179
2000	0.9538	0.1604	0.9534	0.1599	0.9572	0.1588	0.9452	0.1460	0.9552	0.1475
3000	0.9556	0.1261	0.9554	0.1258	0.9592	0.1273	0.9506	0.1179	0.9600	0.1188
4000	0.9512	0.1067	0.9500	0.1064	0.9518	0.1089	0.9378	0.1012	0.9486	0.1018
5000	0.9520	0.0948	0.9520	0.0946	0.9496	0.0974	0.9446	0.0909	0.9522	0.0913
10000	0.9490	0.0652	0.9488	0.0651	0.9462	0.0680	0.9432	0.0638	0.9472	0.0640

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.35. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad
($\lambda_1 = 0.80$, $\lambda_0 = 0.25$)

$Se = 0.90, Sp = 0.85, p = 10\%$										
Intervalos de confianza para la sensibilidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.3296	0.6057	0.3086	0.6010	0.6812	0.6835	0.9364	0.8654	0.8302	0.7104
200	0.7542	0.5612	0.7482	0.5586	0.8544	0.5738	0.9792	0.6098	0.8862	0.5856
300	0.7960	0.5302	0.7918	0.5283	0.8938	0.4967	0.9850	0.4747	0.9220	0.4974
400	0.8538	0.5027	0.8508	0.5011	0.9192	0.4352	0.9862	0.3912	0.9344	0.4305
500	0.8684	0.4814	0.8670	0.4801	0.9360	0.3913	0.9896	0.3379	0.9488	0.3826
1000	0.9236	0.3804	0.9228	0.3796	0.9606	0.2694	0.9870	0.2211	0.9644	0.2517
2000	0.9300	0.2482	0.9300	0.2477	0.9564	0.1827	0.9170	0.1551	0.9660	0.1670
3000	0.9368	0.1791	0.9362	0.1787	0.9555	0.1472	0.9212	0.1273	0.9588	0.1334
4000	0.9446	0.1419	0.9446	0.1416	0.9510	0.1264	0.9268	0.1108	0.9570	0.1146
5000	0.9480	0.1202	0.9476	0.1200	0.9530	0.1120	0.9346	0.0995	0.9562	0.1021
10000	0.9576	0.0776	0.9574	0.0775	0.9530	0.0788	0.9416	0.0709	0.9564	0.0718
Intervalos de confianza para la especificidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9284	0.1875	0.9154	0.1771	0.9438	0.1713	0.9374	0.1635	0.9488	0.1663
200	0.9500	0.1178	0.9432	0.1126	0.9498	0.1131	0.9424	0.1103	0.9512	0.1112
300	0.9494	0.0935	0.9426	0.0901	0.9484	0.0912	0.9436	0.0895	0.9510	0.0899
400	0.9454	0.0800	0.9400	0.0774	0.9438	0.0785	0.9448	0.0774	0.9468	0.0777
500	0.9444	0.0710	0.9386	0.0689	0.9474	0.0701	0.9482	0.0691	0.9478	0.0693
1000	0.9464	0.0494	0.9412	0.0484	0.9474	0.0492	0.9442	0.0487	0.9466	0.0488
2000	0.9450	0.0356	0.9404	0.0351	0.9412	0.0358	0.9418	0.0354	0.9432	0.0354
3000	0.9536	0.0282	0.9506	0.0279	0.9496	0.0283	0.9454	0.0281	0.9484	0.0281
4000	0.9456	0.0244	0.9436	0.0242	0.9446	0.0245	0.9442	0.0244	0.9456	0.0244
5000	0.9540	0.0218	0.9514	0.0216	0.9518	0.0219	0.9520	0.0218	0.9534	0.0218
10000	0.9518	0.0158	0.9504	0.0157	0.9500	0.0160	0.9518	0.0158	0.9514	0.0158

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.36. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad
($\lambda_1 = 0.80$, $\lambda_0 = 0.25$)

<i>Se = 0.90, Sp = 0.85, p = 30%</i>										
Intervalos de confianza para la sensibilidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.7960	0.5202	0.7856	0.5143	0.8912	0.4755	0.9790	0.4603	0.9166	0.4827
200	0.8862	0.4519	0.8828	0.4484	0.9360	0.3457	0.9564	0.2956	0.9486	0.3378
300	0.9104	0.3951	0.9070	0.3924	0.9460	0.2796	0.9834	0.2307	0.9550	0.2646
400	0.9134	0.3416	0.9102	0.3394	0.9528	0.2380	0.9414	0.1980	0.9600	0.2224
500	0.9256	0.3005	0.9232	0.2985	0.9542	0.2096	0.9064	0.1752	0.9636	0.1936
1000	0.9408	0.1755	0.9400	0.1743	0.9548	0.1455	0.9216	0.1255	0.9636	0.1315
2000	0.9504	0.1045	0.9488	0.1039	0.9488	0.1005	0.9316	0.0897	0.9538	0.0916
3000	0.9566	0.0811	0.9552	0.0808	0.9490	0.0814	0.9426	0.0736	0.9520	0.0746
4000	0.9466	0.0687	0.9460	0.0684	0.9456	0.0710	0.9406	0.0639	0.9502	0.0646
5000	0.9510	0.0606	0.9504	0.0604	0.9496	0.0633	0.9442	0.0572	0.9510	0.0577
10000	0.9456	0.0417	0.9450	0.0416	0.9436	0.0449	0.9404	0.0405	0.9418	0.0407
Intervalos de confianza para la especificidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9368	0.2222	0.9252	0.2119	0.9502	.01977	0.9404	0.1868	0.9590	0.1910
200	0.9526	0.1411	0.9462	0.1359	0.9524	0.1335	0.9430	0.1290	0.9564	0.1304
300	0.9544	0.1112	0.9464	0.1077	0.9522	0.1075	0.9412	0.1047	0.9538	0.1054
400	0.9558	0.0948	0.9488	0.0922	0.9514	0.0927	0.9422	0.0906	0.9544	0.0911
500	0.9540	0.0841	0.9496	0.0820	0.9522	0.0827	0.9518	0.0811	0.9560	0.0814
1000	0.9546	0.0583	0.9510	0.0573	0.9486	0.0410	0.9432	0.0405	0.9436	0.0574
2000	0.9432	0.0409	0.9408	0.0404	0.9460	0.0410	0.9432	0.0405	0.9436	0.0406
3000	0.9476	0.0333	0.9446	0.0330	0.9512	0.0335	0.9478	0.0331	0.9486	0.0331
4000	0.9524	0.0288	0.9502	0.0285	0.9524	0.0290	0.9508	0.0287	0.9528	0.0287
5000	0.9488	0.0257	0.9474	0.0255	0.9456	0.0259	0.9472	0.0256	0.9502	0.0256
10000	0.9478	0.0182	0.9466	0.0181	0.9482	0.0183	0.9484	0.0181	0.9472	0.0181

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.37. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad
($\lambda_1 = 0.80$, $\lambda_0 = 0.25$)

<i>Se = 0.90, Sp = 0.85, p = 50%</i>										
Intervalos de confianza para la sensibilidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.8636	0.4616	0.8516	0.4548	0.9208	0.3647	0.9820	0.3229	0.9358	0.3628
200	0.9134	0.3668	0.9090	0.3625	0.9542	0.2554	0.9686	0.2132	0.9568	0.2423
300	0.9276	0.2953	0.9230	0.2920	0.9542	0.2053	0.9084	0.1716	0.9628	0.1819
400	0.9280	0.2371	0.9250	0.2344	0.9532	0.1747	0.9138	0.1495	0.9574	0.1607
500	0.9290	0.1975	0.9262	0.1953	0.9526	0.1552	0.9174	0.1342	0.9602	0.1418
1000	0.9450	0.1147	0.9438	0.1136	0.9540	0.1079	0.9362	0.0961	0.9556	0.0985
2000	0.9526	0.0743	0.9512	0.0738	0.9504	0.0756	0.9448	0.0684	0.9566	0.0692
3000	0.9506	0.0591	0.9496	0.0588	0.9462	0.0620	0.9410	0.0560	0.9488	0.0564
4000	0.9520	0.0505	0.9508	0.0502	0.9514	0.0534	0.9434	0.0485	0.9514	0.0487
5000	0.9518	0.0448	0.9504	0.0446	0.9474	0.0478	0.9442	0.0434	0.9490	0.0436
10000	0.9514	0.0312	0.9506	0.0311	0.9492	0.0337	0.9492	0.0307	0.9502	0.0308
Intervalos de confianza para la especificidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9308	0.2839	0.9214	0.2743	0.9510	0.2434	0.9254	0.2256	0.9652	0.2334
200	0.9452	0.1773	0.9388	0.1721	0.9476	0.1629	0.9352	0.1555	0.9514	0.1580
300	0.9610	0.1379	0.9590	0.1344	0.9586	0.1311	0.9458	0.1264	0.9598	0.1277
400	0.9516	0.1170	0.9462	0.1145	0.9472	0.1130	0.9400	0.1096	0.9488	0.1104
500	0.9528	0.1034	0.9488	0.1013	0.9510	0.1008	0.9476	0.0981	0.9516	0.0987
1000	0.9504	0.0713	0.9476	0.0703	0.9490	0.0708	0.9474	0.0694	0.9514	0.0696
2000	0.9454	0.0497	0.9428	0.0492	0.9444	0.0498	0.9450	0.0490	0.9454	0.0491
3000	0.9452	0.0404	0.9436	0.0401	0.9472	0.0467	0.9450	0.0400	0.9456	0.0401
4000	0.9512	0.0349	0.9484	0.0346	0.9516	0.0353	0.9484	0.0347	0.9510	0.0347
5000	0.9504	0.0312	0.9490	0.0310	0.9504	0.0314	0.9530	0.0310	0.9520	0.0311
10000	0.9494	0.0220	0.9490	0.0219	0.9516	0.0223	0.9502	0.0219	0.9504	0.0220

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.38. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad ($\lambda_1 = 0.80, \lambda_0 = 0.25$).

$Se = 0.90, Sp = 0.85, p = 70\%$										
Intervalos de confianza para la sensibilidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.8820	0.3975	0.8718	0.3895	0.9382	0.2834	0.9762	0.2491	0.9402	0.2822
200	0.9246	0.2880	0.9162	0.2831	0.9562	0.1994	0.9270	0.1710	0.9606	0.1887
300	0.9304	0.2157	0.9256	0.2120	0.9492	0.1605	0.9192	0.1400	0.9608	0.1496
400	0.9282	0.1707	0.9244	0.1678	0.9506	0.1380	0.9196	0.1220	0.9548	0.1278
500	0.9350	0.1408	0.9312	0.1385	0.9490	0.1224	0.9298	0.1092	0.9514	0.1130
1000	0.9480	0.0872	0.9466	0.0862	0.9530	0.0864	0.9386	0.0778	0.9536	0.0790
2000	0.9514	0.0584	0.9494	0.0578	0.9466	0.0606	0.9432	0.0552	0.9488	0.0557
3000	0.9508	0.0469	0.9500	0.0465	0.9452	0.0497	0.9464	0.0452	0.9492	0.0455
4000	0.9526	0.0403	0.9514	0.0400	0.9456	0.0430	0.9496	0.0392	0.9508	0.0393
5000	0.9462	0.0358	0.9454	0.0356	0.9460	0.0384	0.9452	0.0351	0.9484	0.0352
10000	0.9490	0.0251	0.9476	0.0250	0.9454	0.0271	0.9490	0.0248	0.9496	0.0248
Intervalos de confianza para la especificidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9260	0.3950	0.9182	0.3871	0.9594	0.3380	0.9324	0.3099	0.9714	0.3248
200	0.9424	0.2554	0.9366	0.2504	0.9594	0.2216	0.9222	0.2069	0.9652	0.2129
300	0.9416	0.1948	0.9376	0.1913	0.9524	0.1773	0.9296	0.1682	0.9542	0.1713
400	0.9518	0.1627	0.9488	0.1601	0.9520	0.1519	0.9334	0.1453	0.9528	0.1473
500	0.9546	0.1425	0.9522	0.1404	0.9528	0.1354	0.9418	0.1302	0.9532	0.1316
1000	0.9558	0.0961	0.9536	0.0951	0.9582	0.0944	0.9494	0.0917	0.9552	0.0922
2000	0.9522	0.0665	0.9506	0.0660	0.9532	0.0666	0.9460	0.0650	0.9516	0.0652
3000	0.9516	0.0540	0.9508	0.0536	0.9522	0.0543	0.9484	0.0531	0.9512	0.0532
4000	0.9500	0.0466	0.9486	0.0463	0.9476	0.0470	0.9480	0.0460	0.9500	0.0461
5000	0.9454	0.0416	0.9444	0.0414	0.9458	0.0420	0.9464	0.0412	0.9456	0.0412
10000	0.9518	0.0292	0.9508	0.0291	0.9496	0.0297	0.9510	0.0291	0.9516	0.0291

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.39. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad ($\lambda_1 = 0.80, \lambda_0 = 0.25$).

<i>Se = 0.90, Sp = 0.85, p = 90%</i>										
Intervalos de confianza para la sensibilidad										
<i>n</i>	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9228	0.3091	0.9170	0.2995	0.9658	0.2048	0.9650	0.1839	0.9638	0.2054
200	0.9234	0.1987	0.9146	0.1932	0.9564	0.1463	0.9280	0.1310	0.9506	0.1400
300	0.9306	0.1431	0.9236	0.1393	0.9486	0.1186	0.9290	0.1078	0.9514	0.1122
400	0.9342	0.1138	0.9286	0.1110	0.9496	0.1018	0.9418	0.0933	0.9494	0.0959
500	0.9446	0.0973	0.9408	0.0951	0.9566	0.0910	0.9424	0.0840	0.9572	0.0857
1000	0.9496	0.0634	0.9464	0.0624	0.9482	0.0639	0.9462	0.0594	0.9508	0.0600
2000	0.9506	0.0434	0.9474	0.0429	0.9480	0.0450	0.9480	0.0421	0.9502	0.0423
3000	0.9546	0.0351	0.9518	0.0348	0.9530	0.0368	0.9544	0.0344	0.9546	0.0345
4000	0.9492	0.0302	0.9472	0.0300	0.9508	0.0319	0.9474	0.0298	0.9490	0.0298
5000	0.9462	0.0270	0.9444	0.0268	0.9428	0.0285	0.9452	0.0266	0.9460	0.0267
10000	0.9420	0.0190	0.9410	0.0189	0.9436	0.0201	0.9440	0.0188	0.9438	0.0189
Intervalos de confianza para la especificidad										
<i>n</i>	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9128	0.5777	0.9086	0.5726	0.9668	0.5841	0.9956	0.6225	0.9848	0.5891
200	0.9284	0.4863	0.9266	0.4830	0.9668	0.4337	0.9534	0.4171	0.9800	0.4275
300	0.9340	0.4101	0.9320	0.4076	0.9720	0.3502	0.9156	0.3249	0.9800	0.3379
400	0.9418	0.3512	0.9396	0.3490	0.9688	0.2949	0.9164	0.2733	0.9766	0.2845
500	0.9482	0.3051	0.9472	0.3033	0.9684	0.2598	0.9204	0.2412	0.9762	0.2496
1000	0.9510	0.1943	0.9506	0.1933	0.9558	0.1788	0.9314	0.1691	0.9548	0.1720
2000	0.9494	0.1284	0.9490	0.1279	0.9494	0.1243	0.9408	0.1194	0.9512	0.1204
3000	0.9508	0.1024	0.9498	0.1020	0.9494	0.1009	0.9456	0.0974	0.9502	0.0979
4000	0.9544	0.0875	0.9542	0.0873	0.9534	0.0874	0.9478	0.0843	0.9564	0.0846
5000	0.9490	0.0775	0.9484	0.0773	0.9484	0.0778	0.9438	0.0752	0.9480	0.0754
10000	0.9502	0.0540	0.9500	0.0539	0.9516	0.0548	0.9498	0.0531	0.9504	0.0532

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.40. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad
($\lambda_1 = 0.95$, $\lambda_0 = 0.40$)

$Se = 0.90, Sp = 0.85, p = 10\%$										
Intervalos de confianza para la sensibilidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.6574	0.5903	0.6422	0.5855	0.8178	0.6239	0.9692	0.7346	0.8786	0.6529
200	0.8152	0.5344	0.8094	0.5316	0.9012	0.4851	0.9874	0.4712	0.9268	0.4961
300	0.8604	0.4913	0.8562	0.4892	0.9246	0.4008	0.9882	0.3602	0.9424	0.4023
400	0.8884	0.4564	0.8854	0.4547	0.9388	0.3441	0.9922	0.2972	0.9512	0.3397
500	0.9050	0.4229	0.9030	0.4214	0.9430	0.3008	0.9888	0.2578	0.9588	0.2854
1000	0.9240	0.2972	0.9234	0.2963	0.9550	0.2024	0.9132	0.1751	0.9634	0.1926
2000	0.9408	0.1730	0.9402	0.1724	0.9508	0.1399	0.9234	0.1252	0.9598	0.1309
3000	0.9480	0.1261	0.9472	0.1257	0.9508	0.1134	0.9336	0.1029	0.9556	0.1058
4000	0.9464	0.1037	0.9462	0.1038	0.9442	0.0974	0.9292	0.0694	0.9548	0.0912
5000	0.9540	0.0901	0.9540	0.0899	0.9516	0.0874	0.9394	0.0803	0.9512	0.0816
10000	0.9578	0.0602	0.9576	0.0601	0.9536	0.0614	0.9454	0.0569	0.9552	0.0573
Intervalos de confianza para la especificidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9378	0.1835	0.9240	0.1731	0.9490	0.1652	0.9370	0.1595	0.9540	0.1623
200	0.9538	0.1202	0.9456	0.1150	0.9554	0.1145	0.9478	0.1121	0.9576	0.1130
300	0.9586	0.0956	0.9484	0.0922	0.9548	0.0929	0.9492	0.0913	0.9544	0.0918
400	0.9478	0.0819	0.9436	0.0793	0.9484	0.0804	0.9488	0.0791	0.9498	0.0794
500	0.9502	0.0727	0.9434	0.0706	0.9468	0.0716	0.9418	0.0707	0.9490	0.0709
1000	0.9510	0.0507	0.9474	0.0496	0.9530	0.0506	0.9474	0.0500	0.9514	0.0500
2000	0.9506	0.0356	0.9470	0.0351	0.9470	0.0357	0.9482	0.0353	0.9492	0.0354
3000	0.9568	0.0290	0.9544	0.0287	0.9542	0.0291	0.9526	0.0289	0.9544	0.0289
4000	0.9504	0.0251	0.9474	0.0248	0.9516	0.0252	0.9508	0.0250	0.9506	0.0250
5000	0.9514	0.0224	0.9496	0.0222	0.9504	0.0226	0.9510	0.0224	0.9520	0.2224
10000	0.9528	0.0158	0.9514	0.0157	0.9536	0.0159	0.9538	0.0158	0.9538	0.0158

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.41. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad
 $(\lambda_1 = 0.95, \lambda_0 = 0.40)$

$Se = 0.90, Sp = 0.85, p = 30\%$										
Intervalos de confianza para la sensibilidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.8656	0.4870	0.8542	0.4805	0.9236	0.3906	0.9876	0.3519	0.9414	0.3944
200	0.9126	0.3892	0.9086	0.3851	0.9548	0.2670	0.9792	0.2275	0.9614	0.2597
300	0.9210	0.3131	0.9182	0.3099	0.9546	0.2105	0.9062	0.1814	0.9616	0.2013
400	0.9290	0.2554	0.9252	0.2527	0.9560	0.1815	0.9198	0.1585	0.9636	0.1712
500	0.9402	0.2153	0.9384	0.2130	0.9584	0.1589	0.9158	0.1407	0.9706	0.1497
1000	0.9460	0.1240	0.9446	0.1229	0.9520	0.1117	0.9388	0.1016	0.9582	0.1044
2000	0.9486	0.0794	0.9478	0.0788	0.9480	0.0783	0.9384	0.0723	0.9482	0.0732
3000	0.9570	0.0629	0.9558	0.0626	0.9546	0.0640	0.9490	0.0592	0.9560	0.0597
4000	0.9470	0.0535	0.9464	0.0533	0.9464	0.0550	0.9400	0.0512	0.9472	0.0515
5000	0.9548	0.0476	0.9538	0.0474	0.9538	0.0490	0.9504	0.0459	0.9536	0.0461
10000	0.9498	0.0330	0.9480	0.0329	0.9480	0.0347	0.9454	0.0324	0.9476	0.0325
Intervalos de confianza para la especificidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9376	0.2301	0.9270	0.2199	0.9516	0.1980	0.9236	0.1884	0.9596	0.1936
200	0.9462	0.1467	0.9378	0.1415	0.9464	0.1374	0.9346	0.1329	0.9498	0.1345
300	0.9466	0.1039	0.9392	0.1004	0.9478	0.0999	0.9380	0.0985	0.9494	0.0991
400	0.9572	0.0987	0.9524	0.0962	0.9532	0.0963	0.9448	0.0939	0.9540	0.0945
500	0.9472	0.0787	0.9430	0.0767	0.9464	0.0770	0.9418	0.0763	0.9456	0.0765
1000	0.9526	0.0608	0.9502	0.0598	0.9536	0.0607	0.9498	0.0596	0.9542	0.0597
2000	0.9528	0.0426	0.9506	0.0421	0.9488	0.0430	0.9482	0.0422	0.9514	0.0422
3000	0.9506	0.0347	0.9480	0.0343	0.9502	0.0351	0.9484	0.0344	0.9494	0.0345
4000	0.9510	0.0271	0.9486	0.0269	0.9524	0.0271	0.9512	0.0270	0.9512	0.0270
5000	0.9512	0.0268	0.9492	0.0266	0.9478	0.0262	0.9488	0.0267	0.9490	0.0267
10000	0.9484	0.0171	0.9478	0.0170	0.9476	0.0171	0.9484	0.0171	0.9478	0.0171

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.42. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad
($\lambda_1 = 0.95$, $\lambda_0 = 0.40$)

<i>Se = 0.90, Sp = 0.85, p = 50%</i>											
Intervalos de confianza para la sensibilidad											
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>		
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	
100	0.9026	0.4114	0.8926	0.4037	0.9450	0.2869	0.9844	0.2481	0.9560	0.2844	
200	0.9206	0.2861	0.9140	0.2811	0.9504	0.1943	0.9114	0.1696	0.9574	0.1861	
300	0.9322	0.2101	0.9278	0.2064	0.9530	0.1557	0.9096	0.1389	0.9648	0.1475	
400	0.9420	0.1658	0.9368	0.1629	0.9532	0.1340	0.9264	0.1213	0.9558	0.1265	
500	0.9464	0.1378	0.9426	0.1355	0.9540	0.1190	0.9346	0.1087	0.9652	0.1123	
1000	0.9498	0.0866	0.9460	0.0855	0.9540	0.0841	0.9430	0.0777	0.9580	0.0789	
2000	0.9490	0.0581	0.9488	0.0576	0.9518	0.0590	0.9426	0.0551	0.9486	0.0555	
3000	0.9564	0.0466	0.9552	0.0462	0.9510	0.0482	0.9490	0.0450	0.9530	0.0452	
4000	0.9508	0.0400	0.9484	0.0398	0.9456	0.0416	0.9456	0.0390	0.9510	0.0391	
5000	0.9534	0.0356	0.9514	0.0354	0.9486	0.0373	0.9490	0.0349	0.9496	0.0350	
10000	0.9484	0.0250	0.9478	0.0249	0.9476	0.0264	0.9472	0.0248	0.9472	0.0248	
Intervalos de confianza para la especificidad											
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>		
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	
100	0.9272	0.2939	0.9194	0.2845	0.9486	0.2421	0.9242	0.2269	0.9592	0.2365	
200	0.9486	0.1605	0.9424	0.1553	0.9514	0.1477	0.9372	0.1436	0.9556	0.1455	
300	0.9570	0.1264	0.9522	0.1229	0.9578	0.1196	0.9412	0.1173	0.9556	0.1183	
400	0.9492	0.1077	0.9458	0.1051	0.9476	0.1033	0.9370	0.1018	0.9464	0.1024	
500	0.9514	0.0952	0.9472	0.0931	0.9494	0.0921	0.9406	0.0909	0.9488	0.0914	
1000	0.9522	0.0659	0.9486	0.0649	0.9520	0.0649	0.9462	0.0644	0.9502	0.0646	
2000	0.9514	0.0461	0.9496	0.0456	0.9506	0.0458	0.9484	0.0456	0.9512	0.0456	
3000	0.9516	0.0375	0.9488	0.0372	0.9514	0.0374	0.9502	0.0372	0.9498	0.0373	
4000	0.9486	0.0324	0.9472	0.0322	0.9496	0.0324	0.9494	0.0323	0.9494	0.0323	
5000	0.9524	0.0290	0.9510	0.0288	0.9530	0.0290	0.9530	0.0288	0.9546	0.0289	
10000	0.9466	0.0230	0.9456	0.0229	0.9462	0.0234	0.9466	0.0229	0.9472	0.0229	

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.43. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad ($\lambda_1 = 0.95, \lambda_0 = 0.40$).

<i>Se = 0.90, Sp = 0.85, p = 70%</i>										
Intervalos de confianza para la sensibilidad										
<i>n</i>	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9050	0.3378	0.8910	0.3287	0.9460	0.2263	0.9530	0.1988	0.9530	0.2229
200	0.9328	0.2116	0.9234	0.2060	0.9550	0.1550	0.9246	0.1394	0.9650	0.1484
300	0.9376	0.1509	0.9332	0.1470	0.9504	0.1251	0.9308	0.1148	0.9548	0.1192
400	0.9404	0.1215	0.9336	0.1186	0.9476	0.1081	0.9352	0.1000	0.9558	0.1027
500	0.9454	0.1042	0.9406	0.1020	0.9458	0.0963	0.9374	0.0893	0.9512	0.0912
1000	0.9490	0.0683	0.9464	0.0672	0.9468	0.0680	0.9434	0.0636	0.9494	0.0642
2000	0.9480	0.0467	0.9454	0.0461	0.9454	0.0478	0.9442	0.0450	0.9462	0.0453
3000	0.9488	0.0376	0.9458	0.0373	0.9452	0.0390	0.9464	0.0368	0.9482	0.0369
4000	0.9546	0.0324	0.9538	0.0322	0.9542	0.0338	0.9494	0.0319	0.9524	0.0320
5000	0.9494	0.0289	0.9484	0.0287	0.9506	0.0302	0.9486	0.0285	0.9508	0.0285
10000	0.9480	0.0203	0.9468	0.0202	0.9472	0.0214	0.9436	0.0202	0.9460	0.0202
Intervalos de confianza para la especificidad										
<i>n</i>	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9440	0.3522	0.9390	0.3436	0.9720	0.2904	0.9220	0.2695	0.9760	0.2829
200	0.9472	0.2262	0.9410	0.2211	0.9568	0.1962	0.9228	0.1876	0.9594	0.1922
300	0.9544	0.1741	0.9496	0.1706	0.9590	0.1584	0.9410	0.1535	0.9604	0.1559
400	0.9540	0.1463	0.9516	0.1437	0.9556	0.1364	0.9370	0.1332	0.9540	0.1347
500	0.9572	0.1284	0.9542	0.1264	0.9548	0.1215	0.9416	0.1189	0.9550	0.1200
1000	0.9520	0.0877	0.9482	0.0866	0.9502	0.0855	0.9466	0.0843	0.9494	0.0847
2000	0.9516	0.0608	0.9510	0.0603	0.9516	0.0602	0.9504	0.0597	0.9520	0.0598
3000	0.9510	0.0493	0.9496	0.0490	0.9494	0.0491	0.9474	0.0487	0.9496	0.0488
4000	0.9506	0.0426	0.9498	0.0423	0.9526	0.0425	0.9514	0.0422	0.9530	0.0422
5000	0.9438	0.0380	0.9430	0.0378	0.9428	0.0380	0.9460	0.0377	0.9468	0.0377
10000	0.9528	0.0267	0.9520	0.0267	0.9524	0.0268	0.9516	0.0267	0.9524	0.0267

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

Tabla 1.44. Cobertura asintótica de los intervalos de confianza para la sensibilidad y la especificidad ($\lambda_1 = 0.95, \lambda_0 = 0.40$).

<i>Se = 0.90, Sp = 0.85, p = 90%</i>										
Intervalos de confianza para la sensibilidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.9310	0.2576	0.9220	0.2471	0.9640	0.1726	0.9300	0.1564	0.9660	0.1710
200	0.9322	0.1473	0.9228	0.1416	0.9512	0.1195	0.9340	0.1117	0.9512	0.1160
300	0.9380	0.1087	0.9316	0.1049	0.9466	0.0967	0.9338	0.0914	0.9488	0.0936
400	0.9430	0.0894	0.9374	0.0866	0.9494	0.0837	0.9420	0.0793	0.9506	0.0807
500	0.9488	0.0780	0.9448	0.0759	0.9490	0.0746	0.9418	0.0711	0.9494	0.0720
1000	0.9524	0.0526	0.9456	0.0516	0.9486	0.0526	0.9460	0.0503	0.9486	0.0506
2000	0.9546	0.0364	0.9518	0.0359	0.9492	0.0371	0.9478	0.0356	0.9508	0.0357
3000	0.9516	0.0295	0.9484	0.0292	0.9480	0.0303	0.9472	0.0291	0.9486	0.0292
4000	0.9532	0.0255	0.9514	0.0252	0.9504	0.0262	0.9482	0.0252	0.9512	0.0252
5000	0.9440	0.0228	0.9420	0.0226	0.9442	0.0236	0.9452	0.0226	0.9456	0.0226
10000	0.9524	0.0160	0.9504	0.0159	0.9506	0.0166	0.9494	0.0159	0.9508	0.0160
Intervalos de confianza para la especificidad										
	<i>Csc</i>		<i>Ccc</i>		<i>IM</i>		<i>BG</i>		<i>LBG</i>	
<i>n</i>	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
100	0.8960	0.5517	0.8940	0.5462	0.9630	0.5320	0.9940	0.5486	0.9790	0.5372
200	0.9288	0.4429	0.9258	0.4394	0.9656	0.3770	0.9390	0.3565	0.9756	0.3731
300	0.9306	0.3643	0.9280	0.3615	0.9614	0.3011	0.9216	0.2826	0.9750	0.2954
400	0.9398	0.3069	0.9382	0.3046	0.9652	0.2544	0.9194	0.2396	0.9732	0.2489
500	0.9430	0.2650	0.9416	0.2630	0.9626	0.2248	0.9264	0.2137	0.9676	0.2201
1000	0.9532	0.1697	0.9510	0.1686	0.9586	0.1559	0.9376	0.1509	0.9598	0.1530
2000	0.9588	0.1132	0.9578	0.1126	0.9530	0.1087	0.9416	0.1065	0.9530	0.1072
3000	0.9504	0.0906	0.9496	0.0903	0.9498	0.0885	0.9458	0.0870	0.9498	0.0874
4000	0.9550	0.0777	0.9548	0.0775	0.9528	0.0765	0.9464	0.0754	0.9532	0.0756
5000	0.9560	0.0692	0.9550	0.0690	0.9556	0.0686	0.9512	0.0675	0.9546	0.0677
10000	0.9460	0.0483	0.9452	0.0482	0.9474	0.0483	0.9440	0.0477	0.9480	0.0478

Se: sensibilidad. *Sp*: especificidad. *p*: prevalencia. *n*: tamaño muestral.

Cob.: cobertura. *Long.*: longitud media.

Csc: Intervalo de Confianza Cuadrático sin corrección.

Ccc: Intervalo de Confianza Cuadrático con corrección.

IM: Imputación Múltiple.

BG: Intervalo de Confianza de Begg y Greenes.

LBG: Intervalo de Confianza Logit de Begg y Greenes.

En las Figuras 1.3 a 1.14 se muestra como varían las coberturas del intervalo cuadrático sin corrección por continuidad, del intervalo mediante imputación múltiple y del intervalo logit de Begg y Greenes (que son los intervalos que presentan un mejor rendimiento) para los diferentes valores de sensibilidad y especificidad, probabilidades de verificación y para una prevalencia del 10%. En dichos gráficos, tanto el eje de abscisas (tamaño muestral) como el eje de ordenadas (nivel de confianza) se han ajustado a una escala que permite una mejor apreciación de los resultados.

Los resultados de los experimentos de simulación han puesto de manifiesto el comportamiento tan diferente que tienen los intervalos de confianza para la sensibilidad y para la especificidad. Para la sensibilidad, el rendimiento de los intervalos de confianza depende en gran medida de las probabilidades de verificación y de la prevalencia de la enfermedad, mientras que estos parámetros tienen un efecto mucho menor sobre el rendimiento de los intervalos de confianza para la especificidad. Este comportamiento diferencial se ha encontrado en otros estudios sobre la comparación de parámetros de tests diagnósticos binarios en presencia de verificación parcial de la enfermedad (por ejemplo, Roldán Nofuentes y Luna del Castillo, 2007c, 2008; Roldán Nofuentes, Luna del Castillo y Femia Marzo, 2009) y se deben al efecto que tanto las probabilidades de verificación como la prevalencia de la enfermedad tienen sobre los estimadores de la sensibilidad y de la especificidad, ya en la práctica clínica la probabilidad de verificar a un paciente con test positivo (λ_1) es mayor que la probabilidad de verificar a un paciente con test negativo (λ_0), siendo el efecto de λ_1 sobre la sensibilidad más importante que el de λ_0 sobre la especificidad. Por otra parte, en algunos casos, sobre todo cuando las muestras son pequeñas, ninguno de los intervalos de confianza tiene un rendimiento adecuado. Por tanto es necesario investigar nuevos métodos de estimación de la sensibilidad y especificidad en presencia de verificación parcial de la enfermedad, siendo los métodos de inferencia exacta los más apropiados para abordar este problema.

Figura 1.3. Cobertura asintótica de los intervalos de confianza para la sensibilidad
 ($Se = 0.85$, $Sp = 0.75$, $\lambda_1 = 0.70$, $\lambda_0 = 0.10$, $p = 10\%$).

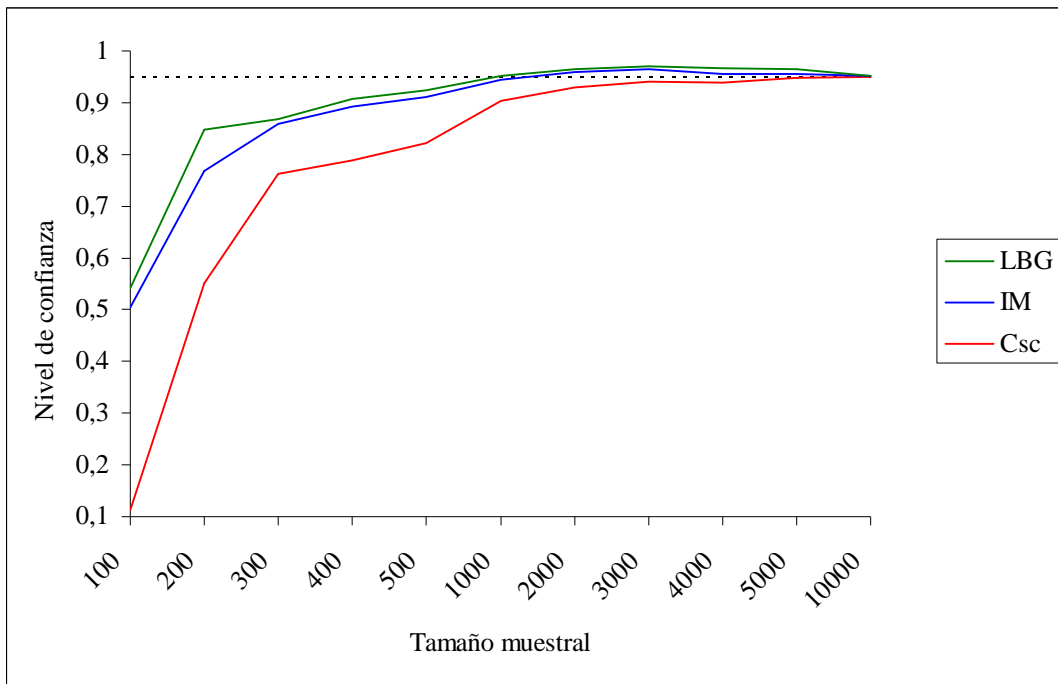


Figura 1.4. Cobertura asintótica de los intervalos de confianza para la especificidad
 ($Se = 0.85$, $Sp = 0.75$, $\lambda_1 = 0.70$, $\lambda_0 = 0.10$, $p = 10\%$).

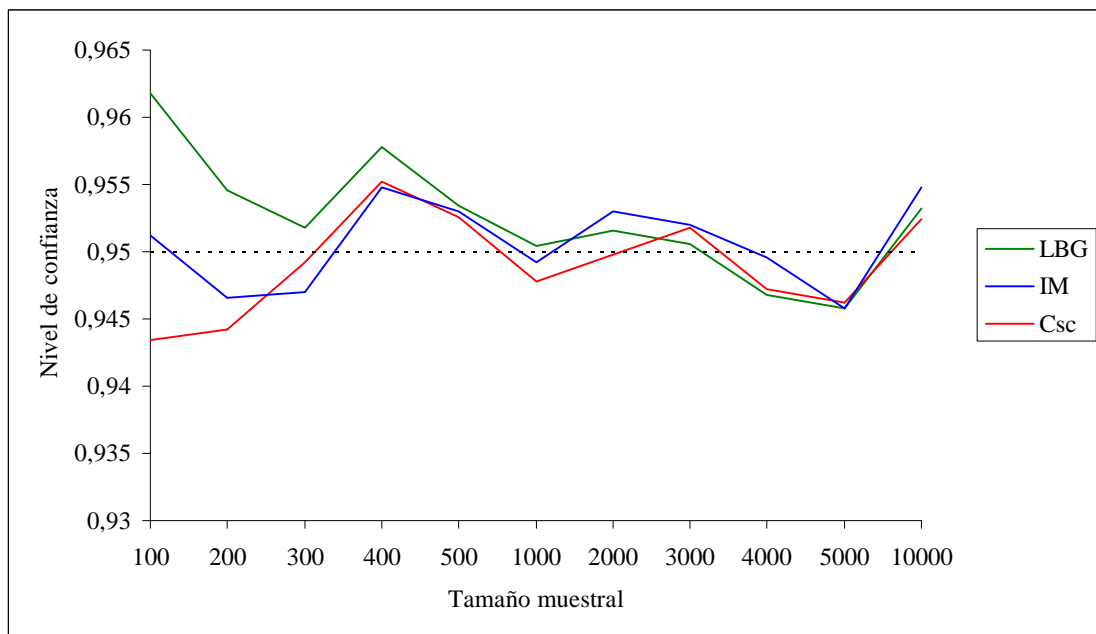


Figura 1.5. Cobertura asintótica de los intervalos de confianza para la sensibilidad
 ($Se = 0.85$, $Sp = 0.75$, $\lambda_1 = 0.80$, $\lambda_0 = 0.25$, $p = 10\%$).

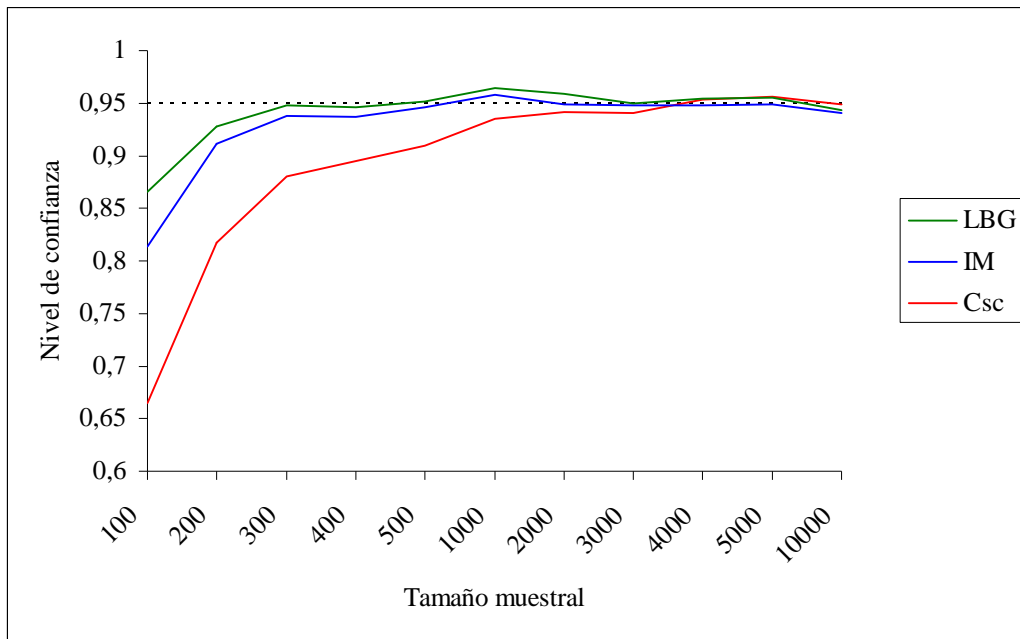


Figura 1.6. Cobertura asintótica de los intervalos de confianza para la especificidad
 ($Se = 0.85$, $Sp = 0.75$, $\lambda_1 = 0.80$, $\lambda_0 = 0.25$, $p = 10\%$).

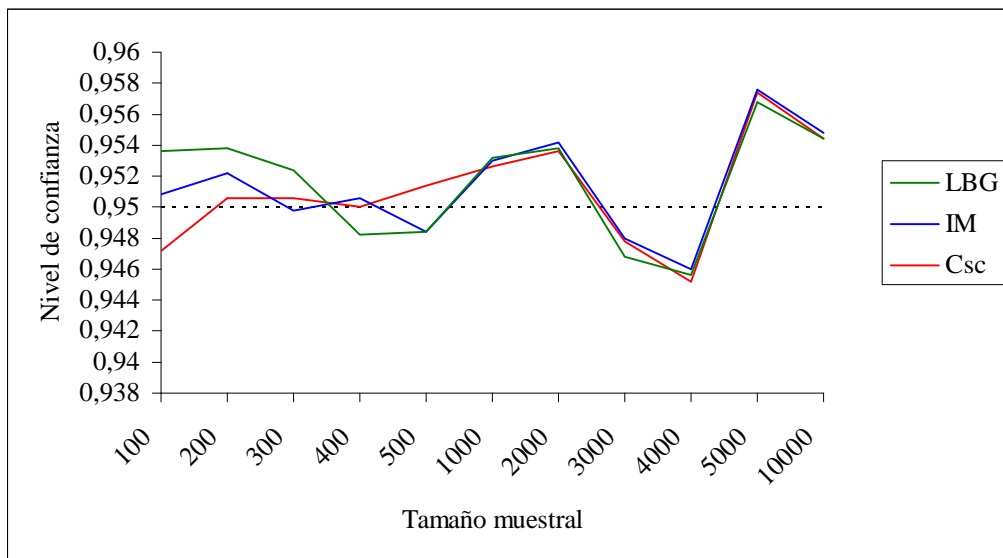


Figura 1.7. Cobertura asintótica de los intervalos de confianza para la sensibilidad
 ($Se = 0.85$, $Sp = 0.75$, $\lambda_1 = 0.95$, $\lambda_0 = 0.40$, $p = 10\%$).

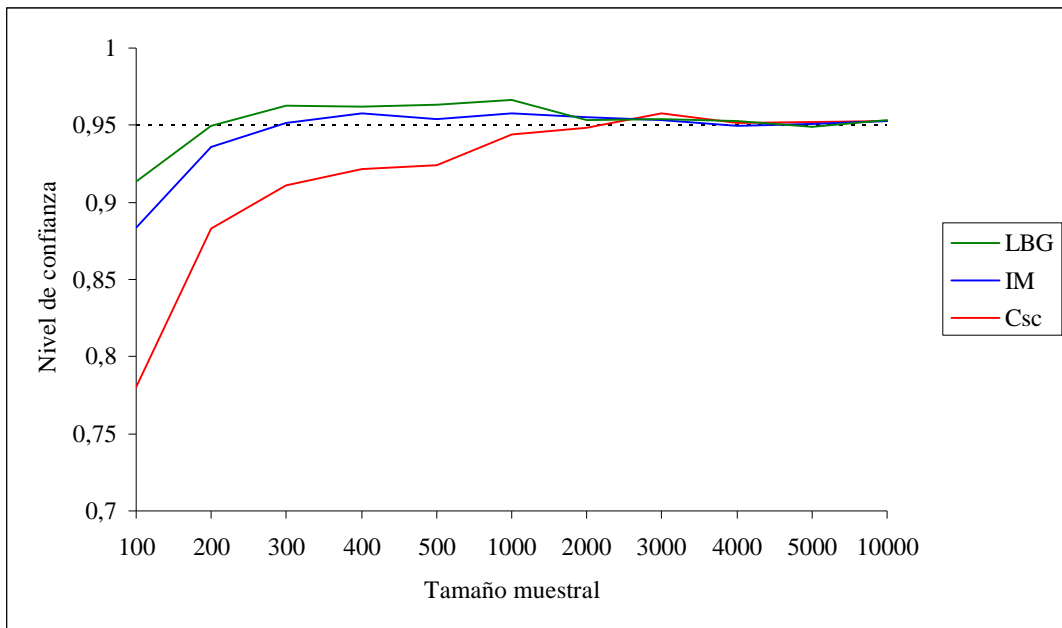


Figura 1.8. Cobertura asintótica de los intervalos de confianza para la especificidad
 ($Se = 0.85$, $Sp = 0.75$, $\lambda_1 = 0.95$, $\lambda_0 = 0.40$, $p = 10\%$).

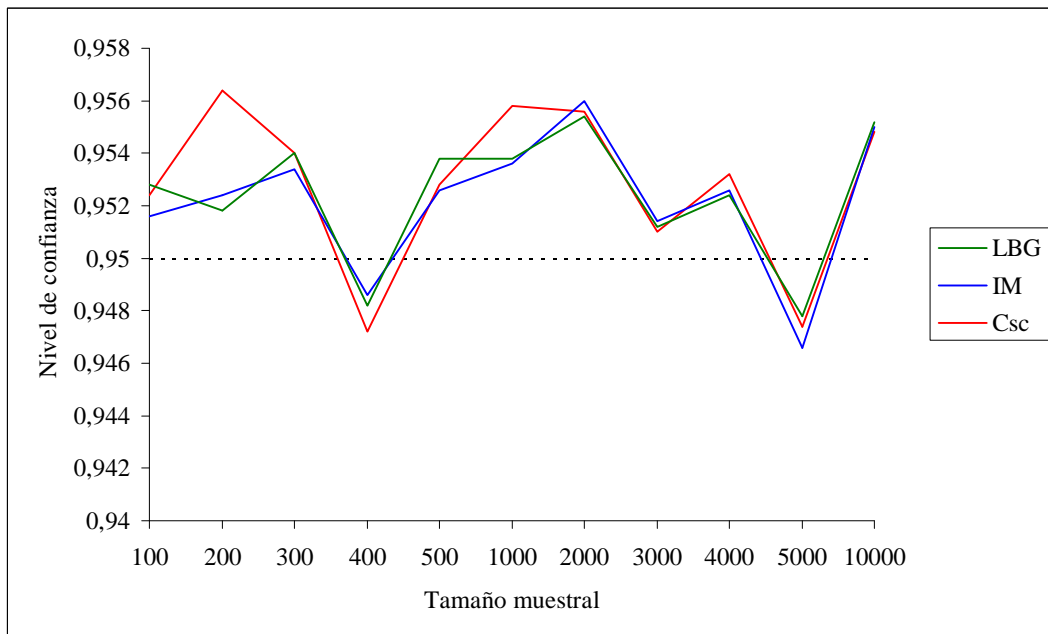


Figura 1.9. Cobertura asintótica de los intervalos de confianza para la sensibilidad
 ($Se = 0.90$, $Sp = 0.85$, $\lambda_1 = 0.70$, $\lambda_0 = 0.10$, $p = 10\%$).

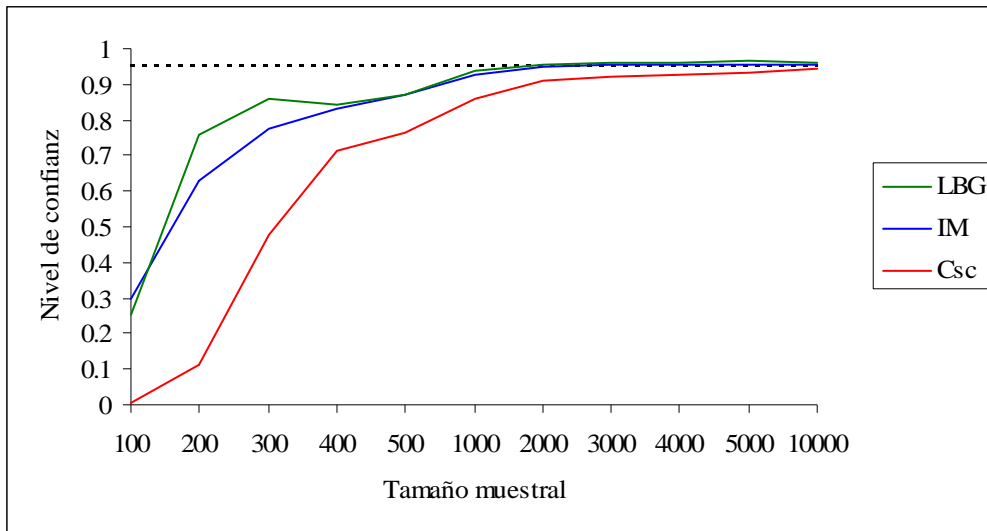


Figura 1.10. Cobertura asintótica de los intervalos de confianza para la especificidad
 ($Se = 0.90$, $Sp = 0.85$, $\lambda_1 = 0.70$, $\lambda_0 = 0.10$, $p = 10\%$).

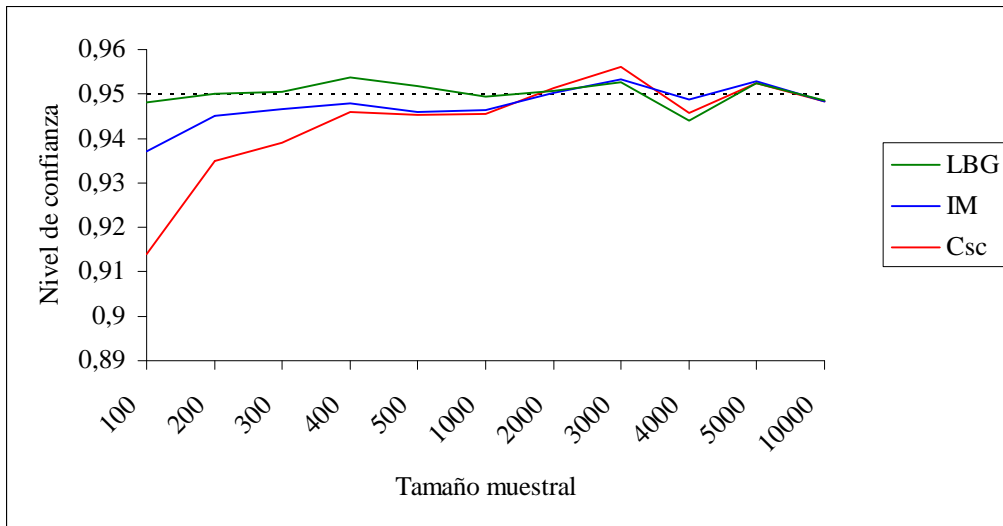


Figura 1.11. Cobertura asintótica de los intervalos de confianza para la sensibilidad
 ($Se = 0.90$, $Sp = 0.85$, $\lambda_1 = 0.80$, $\lambda_0 = 0.25$, $p = 10\%$).

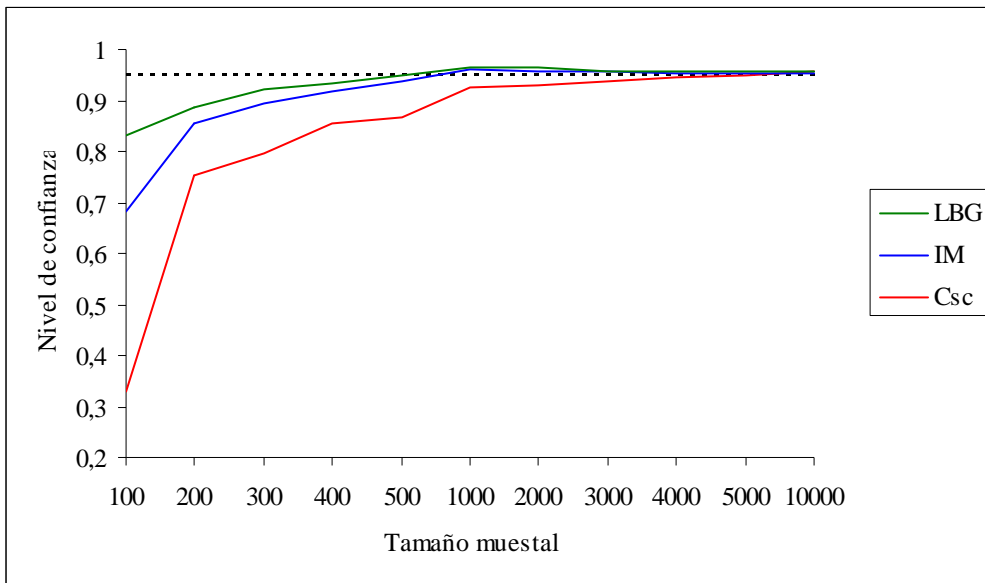


Figura 1.12. Cobertura asintótica de los intervalos de confianza para la especificidad
 ($Se = 0.90$, $Sp = 0.85$, $\lambda_1 = 0.80$, $\lambda_0 = 0.25$, $p = 10\%$).

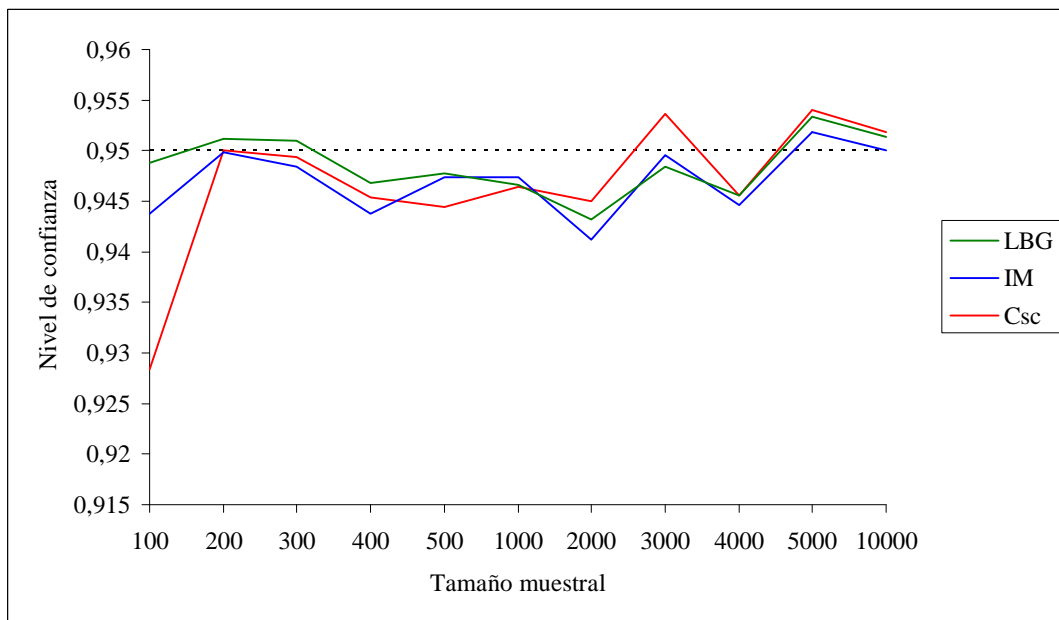


Figura 1.13. Cobertura asintótica de los intervalos de confianza para la sensibilidad
 ($Se = 0.90$, $Sp = 0.85$, $\lambda_1 = 0.95$, $\lambda_0 = 0.40$, $p = 10\%$).

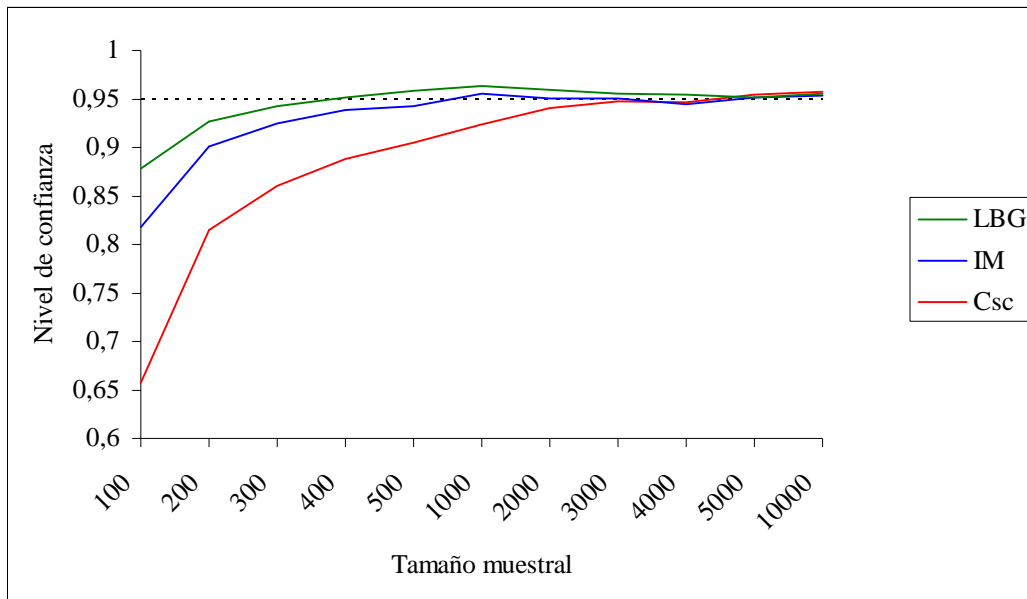
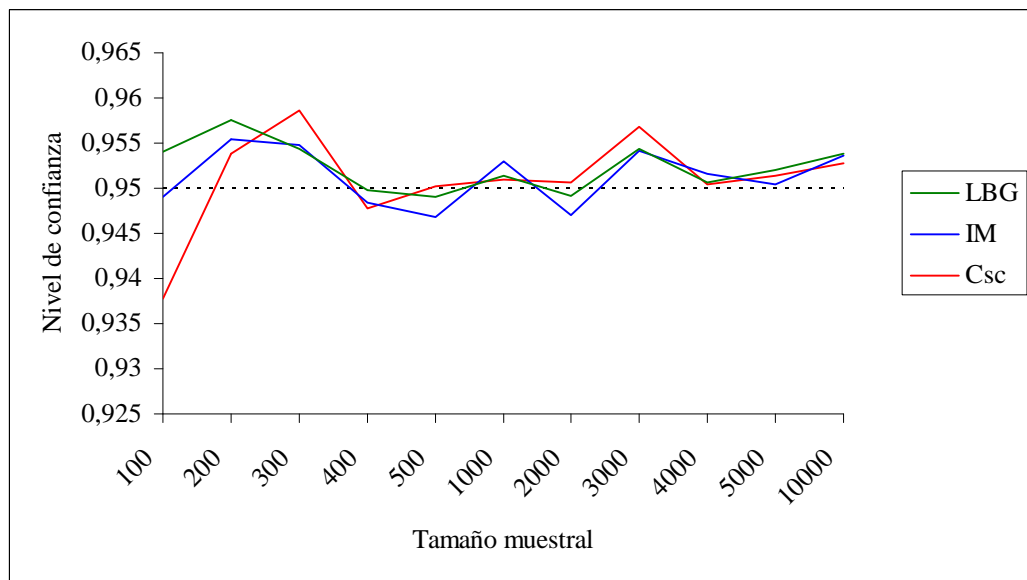


Figura 1.14. Cobertura asintótica de los intervalos de confianza para la especificidad
 ($Se = 0.90$, $Sp = 0.85$, $\lambda_1 = 0.95$, $\lambda_0 = 0.40$, $p = 10\%$).



1.3.1.5.1. Comparación de los intervalos para la sensibilidad

Del análisis de los resultados obtenidos en los experimentos de simulación se obtienen en términos generales las siguientes conclusiones para los intervalos de confianza para la sensibilidad:

- Intervalo cuadrático sin corrección por continuidad (C_{sc}). Cuando las probabilidades de verificación son bajas ($\lambda_1 = 0.70$, $\lambda_0 = 0.10$) y la prevalencia de la enfermedad es pequeña ($p = 10\%$), en términos generales se necesitan muestras muy grandes, de al menos 5000 pacientes, para que la cobertura de este intervalo fluctúe en torno al 95%; cuando la prevalencia es $\geq 30\%$ y $\leq 70\%$, la cobertura del intervalo fluctúa en torno al 95% para tamaños de muestra de al menos 1000-2000 pacientes; y cuando la prevalencia de la enfermedad es muy alta, con muestras de al menos 100-200 pacientes la cobertura del intervalo fluctúa en torno al 95%. Cuando las probabilidades de verificación son intermedias ($\lambda_1 = 0.80$, $\lambda_0 = 0.25$) y la prevalencia de la enfermedad es $\leq 70\%$, se necesita un tamaño muestral de entre 500 y 2000 pacientes para que la cobertura fluctúe en torno al 95%, y cuando la prevalencia es muy alta ($p = 90\%$), con muestras de tamaño 100 la cobertura del intervalo fluctúa en torno al 95%. Cuando las probabilidades de verificación son altas ($\lambda_1 = 0.95$, $\lambda_0 = 0.40$) y la prevalencia de la enfermedad es pequeña ($p = 10\%$), se necesita un tamaño muestral de al menos 2000 pacientes para que la cobertura fluctúe en torno al 95%; cuando la prevalencia de la enfermedad es intermedia ($30\% \leq p \leq 50\%$), se necesitan muestras de al menos 200 pacientes para que el intervalo tenga una buena cobertura, y cuando la prevalencia es muy alta ($p \geq 70\%$), con muestras de tamaño 100 la cobertura del intervalo fluctúa en torno al 95%. En general, una vez que el intervalo cuadrático alcanza la cobertura del 95%, fluctúa siempre en torno a ella.
- Intervalo cuadrático con corrección por continuidad (C_{cc}). La cobertura y amplitud del intervalo cuadrático con corrección por continuidad tienen un comportamiento muy similar a las del intervalo de confianza cuadrático sin corrección por

continuidad. Por consiguiente, la corrección por continuidad no tiene un efecto importante en el comportamiento asintótico del intervalo cuadrático.

- Imputación múltiple (intervalo logit de Rubin y Schenker, 1987). Cuando las probabilidades de verificación son bajas ($\lambda_1 = 0.70$, $\lambda_0 = 0.10$) y la prevalencia de la enfermedad es pequeña ($p = 10\%$), en términos generales se necesitan muestras grandes, de al menos 1000 pacientes, para que la cobertura de este intervalo fluctúe en torno al 95%; cuando la prevalencia es $\geq 30\%$ y $\leq 70\%$, la cobertura del intervalo fluctúa en torno al 95% para tamaños de muestra de entre 200-500 pacientes; y cuando la prevalencia de la enfermedad es muy alta ($p = 90\%$), el intervalo de confianza tiene una cobertura que desborda a la confianza del 95% con muestras de 100-300 pacientes, fluctuando en torno al 95% con muestras de al menos 400 pacientes. Cuando las probabilidades de verificación son intermedias ($\lambda_1 = 0.80$, $\lambda_0 = 0.25$) y la prevalencia de la enfermedad es pequeña ($p = 10\%$), se necesita un tamaño muestral de al menos 500 pacientes para que la cobertura fluctúe en torno al 95%, y para valores de la prevalencia $\geq 30\%$, en general con muestras de entre 100 y 200 pacientes la cobertura del intervalo fluctúa en torno al 95%. Cuando las probabilidades de verificación son altas ($\lambda_1 = 0.95$, $\lambda_0 = 0.40$) y la prevalencia de la enfermedad es pequeña ($p = 10\%$), se necesita un tamaño muestral de al menos 400 pacientes para que la cobertura fluctúe en torno al 95%; y cuando la prevalencia de la enfermedad es $\geq 30\%$, con muestras de al menos 100 pacientes la cobertura del intervalo fluctúa en torno al 95%. En general, el intervalo de confianza logit de Rubin y Schenker (1987) mediante imputación múltiple tiene un buen rendimiento incluso cuando las muestras son pequeñas. Cuando los tamaños muestrales son grandes, el intervalo de confianza cuadrático sin corrección por continuidad tiene una amplitud media levemente menor que el intervalo mediante imputación múltiple, siendo ambas coberturas muy similares.
- Intervalo de confianza de Begg y Greenes. El rendimiento del intervalo de confianza de Begg y Greenes (1983) ha sido objeto de estudios anteriores (Harel y Zhou, 2006), y es un intervalo cuya cobertura presenta un comportamiento muy irregular. En términos muy generales la cobertura de este intervalo fluctúa en torno al 95%

cuando las muestras son grandes, dependiendo de las probabilidades de verificación y de la prevalencia de la enfermedad.

- Intervalo de confianza logit de Begg y Greenes. Este intervalo también ha sido objeto de otros estudios (Harel y Zhou, 2006) y su rendimiento es mejor que el del intervalo de confianza de Begg y Greenes. En términos generales, cuando las probabilidades de verificación y la prevalencia son bajas, se necesitan muestras de al menos 1000 pacientes para que la cobertura del intervalo fluctúe en torno al 95%, disminuyendo el tamaño muestral conforme aumentan las probabilidades de verificación y/o la prevalencia de la enfermedad. Para valores de la prevalencia $\geq 30\%$ y probabilidades de verificación intermedias o altas, la cobertura del intervalo tiene un buen comportamiento en torno al 95% con muestras de al menos 100 pacientes; y conforme aumenta la prevalencia, la cobertura también tiene un buen comportamiento cuando las probabilidades de verificación son bajas. En relación a los intervalos cuadráticos y al intervalo mediante imputación múltiple, el intervalo logit de Begg y Greenes presenta unas fluctuaciones en torno al 95% que son menos estables, aunque su amplitud es levemente menor que la del resto de intervalos.

1.3.1.5.2. Comparación de los intervalos para la especificidad

Del análisis de los resultados obtenidos en los experimentos de simulación para los intervalos de confianza para la especificidad, se obtiene que ni la prevalencia de la enfermedad ni las probabilidades de verificación tienen un efecto importante sobre el rendimiento de estos intervalos de confianza, al contrario de lo que ocurre en los intervalos de confianza para la sensibilidad. En términos generales se obtienen las siguientes conclusiones para los intervalos de confianza para la especificidad.

- Intervalo cuadrático sin corrección por continuidad (C_{sc}). Cuando la prevalencia de la enfermedad es $\leq 70\%$ con muestras de tamaño 200 la cobertura de este intervalo fluctúa en torno al 95%. Cuando la prevalencia es igual al 90%, dependiendo de las probabilidades de verificación, se necesitan muestras entre 100 y 1000 pacientes para que la cobertura fluctúe en torno al 95%. Al igual que ocurre con la

sensibilidad, cuando este intervalo alcanza la cobertura, fluctúa siempre levemente en torno a ella.

- Intervalo cuadrático con corrección por continuidad (C_{cc}). Su rendimiento es similar al intervalo cuadrático sin corrección por continuidad, por lo que la corrección por continuidad no tiene un importante efecto sobre el comportamiento del intervalo cuadrático.
- Imputación múltiple (intervalo logit de Rubin y Schenker, 1987). El rendimiento del intervalo de confianza mediante imputación múltiple es algo similar al del intervalo cuadrático sin corrección por continuidad. En general, cuando la prevalencia de la enfermedad es $\leq 50\%$ con muestras de tamaño 100-200 la cobertura de este intervalo fluctúa en torno al 95%. Cuando la prevalencia es $\geq 70\%$, dependiendo de las probabilidades de verificación, se necesitan muestras entre 100 y 500 pacientes para que la cobertura fluctúe en torno al 95%. En relación al intervalo cuadrático, cuando la cobertura de ambos intervalos fluctúa en torno al 95%, ambos intervalos tienen un rendimiento muy similar tanto en términos de cobertura como de amplitud media.
- Intervalo de confianza de Begg y Greenes. Este intervalo de confianza presenta un mejor rendimiento para la especificidad que para la sensibilidad. En general, su comportamiento es similar al del resto de intervalos para la especificidad, aunque en algunos casos concretos, cuando la prevalencia es muy alta ($p = 90\%$), presenta importantes fluctuaciones en torno al 95%.
- Intervalo de confianza logit de Begg y Greenes. Su comportamiento es muy similar al del intervalo cuadrático sin corrección por continuidad y al del intervalo mediante imputación múltiple.

1.3.1.5.3. Conclusiones

Una vez analizada la cobertura de los intervalos de confianza para la sensibilidad y para la especificidad, comparando el rendimiento de cada intervalo cuando la cobertura fluctúa en torno al 95% se obtienen en términos generales las siguientes conclusiones.

Para la sensibilidad:

- Cuando la prevalencia de la enfermedad es pequeña ($p = 10\%$), utilizar siempre el intervalo de confianza mediante imputación múltiple.
- Para valores de la prevalencia comprendidos entre el 30% y el 70%, para muestras de tamaño menor que 1000 utilizar el intervalo de confianza mediante imputación múltiple, y para muestras de al menos 1000 pacientes utilizar el intervalo cuadrático sin corrección por continuidad. En las situaciones en las que el intervalo cuadrático no tiene un buen rendimiento, utilizar el método de imputación múltiple.
- Cuando la prevalencia de la enfermedad es alta ($p = 90\%$), utilizar siempre el intervalo cuadrático sin corrección por continuidad.

Para la especificidad:

- Cuando la prevalencia de la enfermedad es $\leq 50\%$, para muestras de tamaño 100-200 utilizar el intervalo de confianza mediante imputación múltiple o el intervalo de confianza logit de Begg y Greenes; para muestras de al menos 200 pacientes utilizar cualquiera de los tres intervalos (cuadrático sin corrección por continuidad, imputación múltiple o logit).
- Cuando la prevalencia es $\geq 70\%$, utilizar cualquiera de los tres intervalos (cuadrático sin corrección por continuidad, imputación múltiple o logit).
- En el caso de que se puedan utilizar los tres intervalos (cuadrático, imputación múltiple y logit), los intervalos cuadrático y logit requieren un esfuerzo computacional mucho menor que el intervalo mediante imputación múltiple.

1.3.1.6. Análisis de sensibilidad ante la no verificación de la hipótesis MAR

Kosinski y Barnhart (2003) han estudiado el comportamiento de los estimadores de Begg y Greenes (1983) cuando no se verifica la hipótesis MAR, para lo cual han realizado un análisis de la sensibilidad global de la sensibilidad y de la especificidad de un test cuando no se verifica dicha hipótesis MAR. Este análisis de sensibilidad global proporciona una región con todas las combinaciones de sensibilidad y especificidad

consonante con los datos observados, denominada región TIR (región de ignorancia del test). Si de los u_i individuos no verificados u_{iD} están enfermos y los restantes $(u_i - u_{iD})$ no lo están, se obtiene la Tabla 1.45. En presencia de verificación cuando el proceso de verificación no depende ni del test diagnóstico ni del estado de enfermedad, se dice que el mecanismo de datos faltantes es MCAR (missing completely at random)

Tabla 1.45. Datos cuando u_{1D} y u_{0D} son conocidos.

	$T = 1$	$T = 0$
$V=1$	$s_1 + u_{1D}$	$s_0 + u_{0D}$
$D = 0$	$r_1 + (u_1 - u_{1D})$	$r_0 + (u_0 - u_{0D})$
Total	n_1	n_0

Si u_{1D} y u_{0D} son conocidos, los estimadores de la sensibilidad y la especificidad son

$$\hat{S}e(u_{1D}, u_{0D}) \equiv f_1(u_{1D}, u_{0D}) = \frac{s_1 + u_{1D}}{s_1 + u_{1D} + s_0 + u_{0D}} \quad (1.142)$$

y

$$\hat{S}p(u_{1D}, u_{0D}) \equiv f_0(u_{1D}, u_{0D}) = \frac{r_0 + (u_0 - u_{0D})}{r_1 + (u_1 - u_{1D}) + r_0 + (u_0 - u_{0D})}, \quad (1.143)$$

mientras que si u_{1D} y u_{0D} no son conocidos (como es habitual en presencia de verificación parcial de la enfermedad), sus valores están en algún lugar de la región TIR formada por $[0, u_1] \times [0, u_0]$. El análisis de sensibilidad global estudia el comportamiento de $\hat{S}e(u_{1D}, u_{0D})$ y de $\hat{S}p(u_{1D}, u_{0D})$ cuando u_{1D} y u_{0D} varían según distintos valores. Para ello, primero se consideran los límites generales para los estimadores de la sensibilidad y de la especificidad, que vienen dados por

$$f_1(0, u_0) = \frac{s_1}{s_1 + s_0 + u_0} \leq Se \leq \frac{s_1 + u_1}{s_1 + s_0 + u_1} = f_1(u_1, 0) \quad (1.144)$$

y

$$f_0(0, u_0) = \frac{r_0}{r_1 + r_0 + u_1} \leq Sp \leq \frac{r_0 + u_0}{r_1 + r_0 + u_0} = f_0(u_1, 0). \quad (1.145)$$

Si todos los pacientes con resultado del test positivo tienen su estado de enfermedad verificado, es decir $u_1 = 0$, entonces la sensibilidad, si el mecanismo de datos faltantes es MCAR, es el límite superior para la sensibilidad y la especificidad bajo la condición MCAR es el límite inferior para la especificidad. Por el contrario, si todos los pacientes con resultado del test negativo tienen su estado de enfermedad verificado, es decir $u_0 = 0$, entonces la sensibilidad bajo la condición MCAR es el límite inferior para la sensibilidad y la especificidad bajo la hipótesis MCAR es el límite superior para la especificidad. Cuando u_1 y u_0 son cero los límites generales se reducen a los valores bajo la hipótesis MCAR.

Kosinski y Barnhart (2003) han derivado la región de ignorancia del test (TIR) obteniendo sus límites mediante una transformación de los límites $[0, u_1] \times [0, u_0]$, dados por las siguientes funciones,

$$g_1(Se) \begin{cases} 1 - \frac{n_1 - s_1}{n - \frac{s_1}{Se}} & si \quad f_1(0, u_0) = \frac{s_1}{s_1 + s_0 + u_0} \leq Se \leq \frac{s_1 + u_1}{s_1 + s_0 + u_1} = f_1(0, 0) \\ \frac{n_0 - s_0}{n - \frac{s_0}{(1 - Se)}} & si \quad f_1(0, 0) = \frac{s_1}{s_1 + s_0 + u_0} \leq Se \leq \frac{s_1 + u_1}{s_1 + s_0 + u_1} = f_1(u_1, 0) \end{cases} \quad (1.146)$$

y

$$g_0(Se) \begin{cases} \frac{r_0}{n - \frac{(s_0 + u_0)}{(1 - Se)}} & si \quad f_1(0, u_0) = \frac{s_1}{s_1 + s_0 + u_0} \leq Se \leq \frac{s_1 + u_1}{s_1 + s_0 + u_1} = f_1(u_1, u_0) \\ 1 - \frac{r_1}{n - \frac{(s_1 + u_1)}{Se}} & si \quad f_1(u_1, u_0) = \frac{s_1}{s_1 + s_0 + u_0} \leq Se \leq \frac{s_1 + u_1}{s_1 + s_0 + u_1} = f_1(u_1, 0) \end{cases} \quad (1.147)$$

En la práctica la región TIR puede ser grande, en cuyo caso puede ser interesante visualizar subregiones TIR formadas por más restricciones de u_{1D} y u_{0D} . Para ello puede ser útil la parametrización de estas dos cantidades como $u_{1D} = p_1 u_1$ y $u_{0D} = p_0 u_0$, con $p_1 = P(D=1|T=1, V=0)$ y $p_0 = P(D=1|T=0, V=0)$. De esta forma, para un valor de p_1 fijo en el intervalo $[0,1]$, la especificidad se puede reescribir como una función de la sensibilidad y de p_1 , esto es,

$$Sp = h_1(Se, p_1) = 1 - \frac{n_1 - (s_1 + p_1 u_1)}{n - \frac{(s_1 + p_1 u_1)}{Se}}, \quad (1.148)$$

para la sensibilidad perteneciente al intervalo $[f_1(p_1 u_1, u_0), f_0(p_1 u_1, 0)]$. Para un p_0 fijo en el intervalo $[0,1]$ se tiene que

$$Sp = h_0(Se, p_0) = \frac{n_0 - (s_0 + p_0 u_0)}{n - \frac{(s_0 + p_0 u_0)}{(1-Se)}}, \quad (1.149)$$

para la sensibilidad perteneciente al intervalo $[f_1(0, p_0 u_0), f_0(u_1, p_0 u_0)]$. Añadiendo las curvas h_1 y h_0 al gráfico TIR se pueden visualizar las subregiones de la sensibilidad y de la especificidad para los distintos valores de p_1 y p_0 . En dicha representación se distinguen tres escenarios, dando lugar a su vez a tres subregiones para la sensibilidad y la especificidad:

- I. Donde estarían los pacientes que no se pueden verificar por estar demasiado enfermos, de forma que la probabilidad de tener la enfermedad es alta en los grupos no verificados positivos y negativos comparados con los correspondientes grupos verificados, es decir, $p_1 \geq s_1 / (s_1 + r_1)$ y $p_0 \geq s_0 / (s_0 + r_0)$.
- II. Donde los pacientes no son verificados porque en sentido médico se pueden considerar no enfermos, es decir, $p_1 \leq s_1 / (s_1 + r_1)$ y $p_0 \leq s_0 / (s_0 + r_0)$.

III. Donde el resultado del test tiene impacto en la aplicación del gold estándar debido a:

1. Si el resultado del test es positivo, el paciente es verificado salvo si está muy enfermo.
2. Si el resultado del test es negativo entonces el paciente es expuesto al gold estándar invasivo solamente si hay alguna otra razón para creer que el paciente está enfermo.

En este escenario la probabilidad de enfermedad es alta en el grupo positivo no verificado y baja en el grupo negativo no verificado, es decir $p_1 \geq s_1/(s_1 + r_1)$ y $p_0 \leq s_0/(s_0 + r_0)$.

Por tanto, Kosinski y Barnhart (2003) han realizado un análisis de sensibilidad global para evaluar el rendimiento de un test diagnóstico en presencia de verificación parcial, derivando una región, denominada región TIR, de todos los valores de la sensibilidad y de la especificidad con los datos observados, y proponen que cuando se evalúe un test diagnóstico en presencia de verificación parcial de la enfermedad, en primer lugar se debe proporcionar el gráfico con la región TIR y a partir de este gráfico determinar los posibles pares de valores de la sensibilidad y especificidad.

1.3.2. Estimación de los valores predictivos

Zhou (1994) ha estudiado la estimación de los valores predictivos positivo y negativo de un test diagnóstico en presencia de verificación parcial, demostrando que cuando el proceso de verificación es MAR los estimadores máximo verosímiles de los valores predictivos coinciden con los respectivos estimadores simples.

1.3.2.1. Estimadores simples de los valores predictivos

El valor predictivo positivo se define como la probabilidad de que un individuo esté enfermo dado un resultado positivo del test y el valor predictivo negativo es la probabilidad de que un individuo no esté enfermo dado un resultado negativo del test, esto es,

$$vpp = P(D = 1 | T = 1) \quad (1.150)$$

y

$$vpn = P(D = 0 | T = 0), \quad (1.151)$$

y sus estimadores simples, denotados como \widehat{vpp}_{nv} y \widehat{vpn}_{nv} , obtenidos a partir de los datos de la Tabla 1.12 sin considerar a los individuos no verificados, son

$$\widehat{vpp}_{nv} = \frac{s_1}{s_1 + r_1} \quad (1.152)$$

y

$$\widehat{vpn}_{nv} = \frac{s_0}{s_0 + r_0}. \quad (1.153)$$

Las siguientes proposiciones dan las esperanzas y varianzas de los estimadores simples de los valores predictivos positivo y negativo, tanto en términos generales como bajo la hipótesis MAR.

Proposición 1.1. (Zhou, 1994)

Las esperanzas de \widehat{vpp}_{nv} y de \widehat{vpn}_{nv} son

$$E(\widehat{vpp}_{nv}) = \frac{k_1 vpp}{1 - (1 - k_1) vpp} \quad (1.154)$$

y

$$E(\widehat{vpn}_{nv}) = \frac{k_0 vpn}{1 - (1 - k_0) vpn}, \quad (1.155)$$

respectivamente, y sus varianzas aproximadas vienen dadas por las expresiones

$$\text{Var}\left(\widehat{vpp}_{nv}\right) \approx \frac{\lambda_{11}\lambda_{10}vpp(1-vpp)}{n_1\{\lambda_{11}vpp + \lambda_{10}(1-vpp)\}^3} \quad (1.156)$$

y

$$\text{Var}\left(\widehat{vpn}_{nv}\right) \approx \frac{\lambda_{10}\lambda_{00}vpn(1-vpn)}{n_0\{\lambda_{10}vpn + \lambda_{00}(1-vpn)\}^3} \quad (1.157)$$

siendo $k_1 = \lambda_{11}/\lambda_{01}$ y $k_0 = \lambda_{10}/\lambda_{00}$.

Demostración

Sean $m_0 = s_0 + r_0$ y $m_1 = s_1 + r_1$, entonces los datos (s_1, r_1, u_1) son una realización de una distribución multinomial con probabilidades (Blumenthal, 1968) $\lambda_{11}vpp$, $\lambda_{01}(1-vpp)$ y $1 - \lambda_{11}vpp - \lambda_{01}(1-vpp)$. Por las propiedades de la distribución multinomial (Bishop et al. 1975), la distribución de s_1 condicionado a m_1 es una distribución binomial, esto es

$$s_1 | m_1 \rightarrow B\left(m_1, \frac{\lambda_{11}vpp}{\lambda_{11}vpp + \lambda_{01}(1-vpp)}\right), \quad (1.158)$$

por lo que su esperanza es

$$E\left(\widehat{vpp}_{nv}\right) = E\left(\frac{s_1}{m_1}\right) = E\left(\frac{1}{m_1} E(s_1 | m_1)\right) = \frac{\lambda_{11}vpp}{\lambda_{11}vpp + \lambda_{01}(1-vpp)} \quad (1.159)$$

y su varianza

$$\begin{aligned} \text{Var}\left(\widehat{vpp}_{nv}\right) &= \text{Var}\left(E\left(\widehat{vpp}_{nv} | m_1\right)\right) + E\left(\text{Var}\left(\widehat{vpp}_{nv} | m_1\right)\right) = \\ &= E\left(\frac{\lambda_{11}\lambda_{01}vpp(1-vpp)}{m_1\{\lambda_{11}vpp + \lambda_{01}(1-vpp)\}^2}\right) = \\ &= \frac{\lambda_{11}\lambda_{01}vpp(1-vpp)}{m_1\{\lambda_{11}vpp + \lambda_{01}(1-vpp)\}^2} E\left(\frac{1}{m_1} | m_1 > 0\right), \end{aligned} \quad (1.160)$$

donde m_1 es una variable aleatoria binomial con probabilidad $\lambda_{11}vpp + \lambda_{01}(1-vpp)$.

Mendenhall y Lehmann (1960) han demostrado que

$$E\left(\frac{1}{m_1} E(s_1 | m_1)\right) \approx \frac{1}{n_1 \{\lambda_{11}vpp + \lambda_{01}(1-vpp)\}}, \quad (1.161)$$

por lo que sustituyendo la expresión anterior (1.161) en (1.160) se obtiene que

$$Var(\widehat{vpp}_{nv}) \approx \frac{\lambda_{11}\lambda_{01}vpp(1-vpp)}{n_1 \{\lambda_{11}vpp + \lambda_{01}(1-vpp)\}^3}. \quad (1.162)$$

De forma similar se deduce la ecuación (1.157) de la varianza del estimador simple del valor predictivo negativo.

Proposición 1.2. (Zhou, 1994)

Bajo la hipótesis MAR, los estimadores simples de los valores predictivos positivo y negativo son insesgados y sus varianzas son

$$Var(\widehat{vpp}_{nv}) \approx \frac{vpp(1-vpp)}{n_1\lambda_1} \quad (1.163)$$

y

$$Var(\widehat{vpn}_{nv}) \approx \frac{vpn(1-vpn)}{n_0\lambda_0}, \quad (1.164)$$

siendo $\lambda_1 = \lambda_{01} = \lambda_{11}$ y $\lambda_0 = \lambda_{10} = \lambda_{00}$.

Demostración

Bajo la hipótesis MAR se verifica que $k_1 = k_0 = 1$ y por tanto $\lambda_1 = \lambda_{01} = \lambda_{11}$ y $\lambda_0 = \lambda_{10} = \lambda_{00}$, con lo que operando en las ecuaciones (1.154), (1.155),

(1.156) y (1.157) se obtiene que los estimadores simples son insesgados y sus varianzas vienen dadas por las ecuaciones (1.163) y (1.164).

Bajo la suposición MAR se verifica que $\lambda_1 = \lambda_{01} = \lambda_{11} = P(V=1|T=1)$ y $\lambda_0 = \lambda_{10} = \lambda_{00} = P(V=1|T=0)$, por lo que los estimadores consistentes para λ_1 y λ_0 son

$$\hat{\lambda}_1 = \frac{s_1 + r_1}{n_1} \quad (1.165)$$

y

$$\hat{\lambda}_0 = \frac{s_0 + r_0}{n_0}. \quad (1.166)$$

Sustituyendo los valores predictivos por sus estimadores simples dados por las ecuaciones (1.152) y (1.153), y sustituyendo λ_{11} y λ_{10} por sus estimadores dados (1.165) y (1.166) en las ecuaciones (1.163) y (1.164) se obtienen los estimadores consistentes de las varianzas de los estimadores simples,

$$\widehat{Var}(\widehat{vpp}_{nv}) \approx \frac{s_1 r_1}{(s_1 + r_1)^3} \quad (1.167)$$

y

$$\widehat{Var}(\widehat{vpn}_{nv}) \approx \frac{s_0 r_0}{(s_0 + r_0)^3}. \quad (1.168)$$

1.3.2.2. Estimadores máximo verosímiles de los valores predictivos

Los datos (s_1, r_1, u_1) y (s_0, r_0, u_0) se pueden considerar como las realizaciones de dos muestras independientes de dos distribuciones multinomiales con probabilidades

$$(P(D=1, V=1|T=1), P(D=0, V=1|T=1), P(V=0|T=1))$$

y

$$(P(D=1, V=1|T=0), P(D=0, V=1|T=0), P(V=0|T=0)),$$

respectivamente. La función del logaritmo de la verosimilitud de la muestra multinomial (s_1, r_1, u_1) es

$$l_1 = \log \left\{ (\lambda_{11} vpp)^{s_1} (\lambda_{01} (1-vpp))^{r_1} (1 - \lambda_{11} vpp - \lambda_{01} (1-vpp))^{u_1} \right\} =$$

$$s_1 \log \{ \lambda_{11} vpp \} + r_1 \log \{ \lambda_{01} (1-vpp) \} + u_1 \log \{ 1 - \lambda_{11} vpp - \lambda_{01} (1-vpp) \}, \quad (1.169)$$

y la de la muestra multinomial (s_0, r_0, u_0) es

$$l_0 = \log \left\{ (\lambda_{10} vpn)^{s_0} (\lambda_{00} (1-vpn))^{r_0} (1 - \lambda_{10} vpn - \lambda_{00} (1-vpn))^{u_0} \right\} =$$

$$s_0 \log \{ \lambda_{10} vpn \} + r_0 \log \{ \lambda_{00} (1-vpn) \} + u_0 \log \{ 1 - \lambda_{10} vpn - \lambda_{00} (1-vpn) \}. \quad (1.170)$$

Como el número de grados de libertad de l_1 y de l_0 es dos para ambos, entonces solamente se pueden estimar dos de los cuatro parámetros de l_1 y l_0 . Si se supone que k_1 y k_0 son conocidos, maximizando l_1 respecto de vpp y l_0 respecto de vpn , los estimadores máximo verosímiles de los valores predictivos son

$$\widehat{vpp} = \frac{s_1}{s_1 + r_1 k_1} \quad (1.171)$$

y

$$\widehat{vpn} = \frac{r_0 k_0}{s_0 + r_0 k_0}. \quad (1.172)$$

El suponer que $k_1 = k_0 = 1$ es equivalente a suponer que la probabilidad de seleccionar un individuo para la verificación de la enfermedad depende solamente del resultado del test diagnóstico y por tanto que el proceso de verificación es MAR, es decir

$$k_1 = k_0 = 1 \Rightarrow \frac{\lambda_{11}}{\lambda_{01}} = \frac{\lambda_{10}}{\lambda_{00}} = 1 \Rightarrow \frac{P(V=1|D=1,T=1)}{P(V=1|D=0,T=1)} = \frac{P(V=1|D=1,T=0)}{P(V=1|D=0,T=0)} = 1, \quad (1.173)$$

por lo que

$$\lambda_{11} = \lambda_{01} \Rightarrow P(V=1|D=1,T=1) = P(V=1|D=0,T=1) = P(V=1|T=1) \quad (1.174)$$

y

$$\lambda_{10} = \lambda_{00} \Rightarrow P(V=1|D=1,T=0) = P(V=1|D=0,T=0) = P(V=1|T=0). \quad (1.175)$$

Por lo tanto, bajo la hipótesis MAR, los estimadores máximo-verosímiles coinciden con los estimadores simples, verificándose:

- Si $k_1 = 1 \Rightarrow \lambda_{11} = \lambda_{01} = \lambda_1$, siendo λ_1 la probabilidad de seleccionar para su verificación un individuo con resultado del test positivo, y su estimador máximo verosímil es

$$\hat{\lambda}_1 = \frac{r_1 + s_1}{n_1}, \quad (1.176)$$

con $n_1 = s_1 + r_1 + u_1$.

- Si $k_0 = 1 \Rightarrow \lambda_{10} = \lambda_{00} = \lambda_0$, siendo λ_0 la probabilidad de seleccionar para su verificación un individuo con resultado del test negativo, y su estimador máximo verosímil es

$$\hat{\lambda}_0 = \frac{r_0 + s_0}{n_0}, \quad (1.177)$$

con $n_0 = s_0 + r_0 + u_0$.

Como los estimadores máximo verosímiles de los valores predictivos dependen de k_1 y k_0 , estos valores se pueden estimar a partir de dichos estimadores. Suponiendo que de los u_1 individuos no verificados con resultado del test positivo s_{11} están enfermos y r_{01} no lo están, y que de los u_0 individuos no verificados con resultado del test negativo

s_{10} están enfermos y r_{00} no lo están, entonces los estimadores insesgados de k_1 y k_0 vienen dados por las ecuaciones (1.101) y (1.102). En la práctica, solamente se conoce que $0 \leq s_{11} \leq u_1$ y $0 \leq s_{10} \leq u_0$, por lo que unos intervalos para \hat{k}_1 y \hat{k}_0 son los dados por (1.103) y (1.104).

De la ecuación (1.171) se deduce que \widehat{vpp} que es una función decreciente en k_1 y de (1.172) que \widehat{vpn} que es una función creciente en k_0 , por lo que utilizando las ecuaciones anteriores, (1.103) y (1.104), se obtienen unos intervalos para los estimadores máximo verosímiles, es decir,

$$\frac{s_1}{n_1} \leq \widehat{vpp} \leq \frac{s_1 + u_1}{n_1} \quad (1.178)$$

y

$$\frac{s_0}{n_0} \leq \widehat{vpn} \leq \frac{s_0 + u_0}{n_0}. \quad (1.179)$$

De la ecuación anterior (1.178) se deduce que cuanto más grande es la razón u_1/n_1 , más susceptible es el estimador simple del valor predictivo positivo a alejarse de la hipótesis MAR. De la misma forma, de (1.179) se deduce que cuanto más grande es la razón u_0/n_0 más susceptible es el estimador simple del valor predictivo negativo a alejarse de dicha hipótesis.

1.3.3. Estimación del coeficiente kappa ponderado

Roldán Nofuentes y Luna del Castillo (2007b) han estudiado la estimación del coeficiente kappa ponderado de un test diagnóstico binario en presencia de verificación parcial de la enfermedad, tanto si en todos los individuos se observan o no covariables discretas.

1.3.3.1. Estimación del coeficiente kappa ponderado sin covariables

Cuando el proceso de verificación es MAR, sustituyendo en la expresión del coeficiente kappa ponderado, ecuación (1.16), los parámetros por sus estimadores máximo verosímiles en presencia de verificación parcial dados por las expresiones (1.105), (1.106) y (1.107), el estimador máximo verosímil del coeficiente kappa ponderado del test diagnóstico en presencia de verificación parcial es

$$\hat{\kappa}(c) = \frac{n_1 n_0 (s_1 r_0 - s_0 r_1)}{nc \{s_0 n_0 (s_1 + r_1) - n_1 r_1 (s_0 + r_0)\} + n_1 \{n_1 r_1 (s_0 + r_0) + n_0 r_0 (s_1 + r_1)\}} \quad (1.180)$$

donde $n = n_1 + n_0$. Aplicando el método delta el estimador de la varianza de $\hat{\kappa}(c)$ es

$$\begin{aligned} \widehat{Var}(\hat{\kappa}(c)) = & \\ & \left(\frac{\hat{p}\hat{q} + \hat{p}(\hat{p} + c - 1)\hat{\kappa}(c)}{\hat{p}(1 - \hat{Q})c + \hat{q}\hat{Q}(1 - c)} \right)^2 \widehat{Var}(\hat{S}_e) + \left(\frac{\hat{p}\hat{q} + \hat{q}(\hat{p} + c - 1)\hat{\kappa}(c)}{\hat{p}(1 - \hat{Q})c + \hat{q}\hat{Q}(1 - c)} \right)^2 \widehat{Var}(\hat{S}_p) + \\ & \left(\frac{(\hat{S}_e + \hat{S}_p - 1) - (\hat{S}_p - 2\hat{p}(\hat{S}_e + \hat{S}_p - 1) + (1 - c)(\hat{S}_e + \hat{S}_p - 2))\hat{\kappa}(c)}{\hat{p}(1 - \hat{Q})c + \hat{q}\hat{Q}(1 - c)} \right)^2 \widehat{Var}(\hat{p}), \end{aligned} \quad (1.181)$$

donde $\hat{Q} = \hat{p}\hat{S}_e + \hat{q}(1 - \hat{S}_p)$, $\hat{q} = 1 - \hat{p}$, y las varianzas $\widehat{Var}(\hat{S}_e)$ y $\widehat{Var}(\hat{S}_p)$ vienen dadas por las expresiones (1.121) y (1.122) respectivamente, y

$$\widehat{Var}(\hat{p}) = \frac{n_1 n_0 (s_1 r_0 - s_0 r_1)^2}{n^3 (s_1 + r_1)^2 (s_0 + r_0)^2} + \frac{n_1^2 s_1 r_1}{n^2 (s_1 + r_1)^3} + \frac{n_0^2 s_0 r_0}{n^2 (s_0 + r_0)^3}. \quad (1.182)$$

Si no se puede asumir la hipótesis MAR, no se puede obtener el estimador puntual del coeficiente kappa ponderado, y sólo se puede obtener los rangos de posibles valores entre los que se encuentra, de forma similar a como ocurre con la sensibilidad y especificidad y con los valores predictivos. Así, el estimador máximo verosímil del coeficiente kappa ponderado es

$$\hat{\kappa}(c) = \frac{n_1 n_0 (s_1 k_0 r_0 - s_0 k_1 r_1)}{n_0 \gamma + (1 - c)n\beta}, \quad (1.183)$$

siendo

$$\gamma = nc(s_0 + k_0r_0)(s_1 + k_1r_1) - n_1k_1r_1(s_0 + k_0r_0) - n_0k_0r_0(s_1 + k_1r_1)$$

y

$$\beta = n_1k_1r_1(s_0 + k_0r_0) + n_0k_0r_0(s_1 + k_1r_1).$$

1.3.3.2. Estimación del coeficiente kappa ponderado con covariables

En presencia de verificación parcial de la enfermedad, la estimación del coeficiente kappa también se puede realizar cuando en todos los individuos se observa un vector de covariables discretas, de forma similar a lo estudiado en la Sección 1.3.1.2. Si X es un vector de covariables discretas que tiene J patrones diferentes de covariables (x_1, \dots, x_J) , los datos de la muestra provienen de una mixtura de J distribuciones multinomiales independientes (para cada patrón de covariable se obtiene la Tabla 1.14), siendo $\alpha_j = P(X = x_j)$ la probabilidad de la j -ésima distribución multinomial. Si el proceso de verificación depende solamente del resultado del test diagnóstico y del patrón de covariable, es decir, si

$$P(V = 1|D = i, T = j, X = x_k) = P(V = 1|T = j, X = x_k),$$

el estimador máximo verosímil del coeficiente kappa ponderado es

$$\hat{\kappa}(c) = \frac{(1 - \hat{p}) \sum_{j=1}^J \hat{p}_j \hat{\alpha}_j \hat{S}e_j + \hat{p} \sum_{j=1}^J (1 - \hat{p}_j) \hat{\alpha}_j \hat{S}p_j - \hat{p}(1 - \hat{p})}{\{\hat{p} + c - 1\} \left[\sum_{j=1}^J \hat{\alpha}_j \left\{ \hat{p}_j (1 - \hat{S}e_j) + (1 - \hat{p}_j) \hat{S}p_j \right\} \right] + (1 - c)(1 - \hat{p})}, \quad (1.184)$$

donde $\hat{S}e_j$ y $\hat{S}p_j$ son los estimadores máximo verosímiles de la sensibilidad y especificidad del test diagnóstico en el j -ésimo patrón de covariable, \hat{p}_j es el estimador de la prevalencia de la enfermedad en el mismo patrón de covariable; $\hat{S}e$, $\hat{S}p$ y \hat{p} los estimadores globales de la sensibilidad, especificidad y prevalencia; y $\hat{\alpha}_j = n_j/n$,

siendo n_j el número de individuos con $X = x_j$ y n el número de individuos totales.

Aplicando el método delta, la varianza estimada de $\hat{\kappa}(c)$ es

$$\begin{aligned}
 \widehat{Var}(\hat{\kappa}(c)) = & \frac{\{(1-\hat{p}) + \hat{\kappa}(c)(c + \hat{p} - 1)\}^2}{\hat{\beta}^2} \sum_{j=1}^J \hat{p}_j^2 \hat{\alpha}_j^2 \widehat{Var}(\hat{S}e_j) + \\
 & \frac{\{\hat{p} - \hat{\kappa}(c)(c + \hat{p} - 1)\}^2}{\hat{\beta}^2} \sum_{j=1}^J (1-\hat{p}_j)^2 \hat{\alpha}_j^2 \widehat{Var}(\hat{S}p_j) + \\
 & \sum_{j=1}^J \frac{(\hat{\alpha}_j \hat{\gamma}_{1j} + \hat{\kappa}(c) \hat{\gamma}_{2j})^2}{\hat{\beta}^2} \widehat{Var}(\hat{p}_j) + \sum_{j=1}^J \frac{(\hat{\gamma}_{3j} + \hat{\kappa}(c) \hat{\gamma}_{4j})^2}{\hat{\beta}^2} \widehat{Var}(\hat{\alpha}_j) + \\
 & 2 \sum_{j < k} \frac{(\hat{\gamma}_{3j} + \hat{\kappa}(c) \hat{\gamma}_{4j})(\hat{\gamma}_{3k} + \hat{\kappa}(c) \hat{\gamma}_{4k})}{\hat{\beta}^2} \widehat{Cov}(\hat{\alpha}_j, \hat{\alpha}_k),
 \end{aligned} \tag{1.185}$$

siendo

$$\begin{aligned}
 \widehat{Var}(\hat{S}e_j) = & \left\{ \hat{S}e_j (1 - \hat{S}e_j) \right\}^2 \left\{ \frac{N_j}{n_{1j} n_{0j}} + \frac{r_{1j}}{s_{1j} (s_{1j} + r_{1j})} + \frac{r_{0j}}{s_{0j} (s_{0j} + r_{0j})} \right\}, \\
 \widehat{Var}(\hat{S}p_j) = & \left\{ \hat{S}p_j (1 - \hat{S}p_j) \right\}^2 \left\{ \frac{N_j}{n_{1j} n_{0j}} + \frac{s_{1j}}{r_{1j} (s_{1j} + r_{1j})} + \frac{s_{0j}}{r_{0j} (s_{0j} + r_{0j})} \right\}, \\
 \widehat{Var}(\hat{p}_j) = & \frac{n_{1j} n_{0j} (s_{1j} r_{0j} - s_{0j} r_{1j})^2}{N_j^3 (s_{1j} + r_{1j})^2 (s_{0j} + r_{0j})^2} + \frac{n_{1j}^2 s_{1j} r_{1j}}{N_j^2 (s_{1j} + r_{1j})^3} + \frac{n_{0j}^2 s_{0j} r_{0j}}{N_j^2 (s_{0j} + r_{0j})^3}, \\
 \hat{\beta} = & \left\{ \hat{p} + c - 1 \right\} \left[\sum_{j=1}^J \hat{\alpha}_j \left\{ \hat{p}_j (1 - \hat{S}e_j) + (1 - \hat{p}_j) \hat{S}p_j \right\} \right] + (1 - c)(1 - \hat{p}), \\
 \hat{\gamma}_{1j} = & (1 - \hat{p}) \hat{S}e_j - \hat{p} \hat{S}p_j - \hat{p} \hat{S}e + (1 - \hat{p}) \hat{S}p + 2\hat{p} - 1, \\
 \hat{\gamma}_{2j} = & \hat{\alpha}_j \left\{ \sum_{i=1}^J \hat{\alpha}_i (1 - \hat{p}_i) \hat{S}p_i + \sum_{i=1}^J \hat{\alpha}_i \hat{p}_i (1 - \hat{S}e_i) + c - 1 \right\} + (c + \hat{p} - 1) \left\{ \hat{p}_j (1 - \hat{S}e_j) - \hat{\alpha}_j \hat{S}p_j \right\}, \\
 \hat{\gamma}_{3j} = & \hat{p}_j \left\{ (1 - \hat{p}) \hat{S}p - \hat{p} \hat{S}p_j + (1 - \hat{p}) \hat{S}e_j - \hat{p} \hat{S}e + 2\hat{p} - 1 \right\} - (1 - \hat{S}p_j) \hat{p},
 \end{aligned}$$

$$\hat{\gamma}_{4j} = (1 - \hat{p}_j) \left[1 - c - \sum_{i=1}^J \hat{\alpha}_i (1 - \hat{p}_i) \hat{S}p_i - \sum_{i=1}^J \hat{\alpha}_i \hat{p}_i (1 - \hat{S}e_i) \right] + (\hat{p} + c - 1) \{ (1 - \hat{p}_j) \hat{S}p_j + \hat{p}_j (1 - \hat{S}e_j) \},$$

$$\widehat{Var}(\hat{\alpha}_j) = \hat{\alpha}_j (1 - \hat{\alpha}_j)$$

y

$$\widehat{Cov}(\hat{\alpha}_j, \hat{\alpha}_k) = -\hat{\alpha}_j \hat{\alpha}_k.$$

Capítulo 2

Comparación de tests diagnósticos binarios

La comparación de la exactitud de dos o más tests diagnósticos binarios es uno de los problemas más importantes en el estudio de los métodos estadísticos para el diagnóstico, tanto en presencia de verificación completa como de verificación parcial de la enfermedad. En este Capítulo se estudian distintos tests de hipótesis para comparar las sensibilidades, las especificidades, las razones de verosimilitud (positivas y negativas), los valores predictivos (positivos y negativos) y los coeficientes kappa ponderados de tests diagnósticos binarios, tanto en presencia de verificación completa como de verificación parcial de la enfermedad. En presencia de verificación completa se estudian la comparación de las sensibilidades y especificidades de dos tests binarios con muestras independientes y con muestras apareadas, se estudian la comparación de las razones de verosimilitud, valores predictivos y coeficientes kappa ponderados de dos tests binarios en diseños apareados y por último se estudia la comparación de los

coeficientes kappa ponderados de múltiples tests binarios (aportación). En presencia de verificación parcial de la enfermedad se estudian los tests de hipótesis para comparar los mismos parámetros anteriores cuando se comparan dos y más de dos tests binarios.

2.1.- Verificación completa

Cuando dos tests binarios y el gold estándar se aplican a todos los individuos de una muestra aleatoria, la comparación de la exactitud de ambos tests binarios es un problema de especial relevancia en el ámbito de los métodos estadísticos para el diagnóstico, ya que su resolución implica la comparación de proporciones binomiales o de funciones de ellas. En esta situación, la comparación de la exactitud de dos tests binarios da lugar al análisis de dos tablas 2×2 en el caso de dos muestras independientes, y de una tabla 2×4 en el caso de muestras apareadas.

2.1.1. Comparación de las sensibilidades y de las especificidades

Cuando todos los pacientes de la muestra aleatoria son verificados con el gold estándar, la sensibilidad y especificidad de cada test diagnóstico es una proporción binomial. Por consiguiente, la comparación de las sensibilidades y de las especificidades de dos tests binarios se resuelve aplicando tests de hipótesis de comparación de dos proporciones binomiales. A continuación se estudia la comparación de las sensibilidades y de las especificidades de dos tests diagnósticos binarios, tanto con muestras independientes como con muestras apareadas.

2.1.1.1. Comparación con muestras independientes

Sean dos muestras aleatorias de tamaños n_1 y n_2 extraídas de dos poblaciones independientes, de tal forma que los individuos de la primera muestra han sido clasificados en función de la presencia o ausencia de una enfermedad y del resultado del test binario 1, y los individuos de la segunda muestra han sido clasificados en función de la presencia o ausencia de la misma enfermedad y del resultado del test binario 2, tal y como se muestra en la Tabla 2.1. En esta tabla las variables aleatorias T_1 y T_2

modelizan los resultados de los tests diagnósticos 1 y 2 respectivamente, de tal forma que $T_h = 1$ cuando el resultado del h -ésimo test es positivo y $T_h = 0$ cuando su resultado es negativo ($h=1,2$), y la variable aleatoria D modeliza el verdadero estado de la enfermedad, siendo $D=1$ cuando el individuo está enfermo y $D=0$ cuando está sano. Sean $Se_h = P(T_h = 1|D=1)$ y $Sp_h = P(T_h = 0|D=0)$ la sensibilidad y la especificidad del h -ésimo test diagnóstico, respectivamente.

Tabla 2.1. Frecuencias observadas cuando se comparan dos tests binarios con dos muestras independientes.

	$T_1 = 1$	$T_1 = 0$	Total		$T_2 = 1$	$T_2 = 0$	Total
$D = 1$	s_{11}	s_{10}	$s_{1\cdot}$	$D = 1$	r_{11}	r_{10}	$r_{1\cdot}$
$D = 0$	s_{01}	s_{00}	$s_{0\cdot}$	$D = 0$	r_{01}	r_{00}	$r_{0\cdot}$
Total	$s_{\cdot 1}$	$s_{\cdot 0}$	n_1	Total	$r_{\cdot 1}$	$r_{\cdot 0}$	n_2

Utilizando un planteamiento incondicionado, el estadístico de contraste para el test de hipótesis de comparación de las dos sensibilidades

$$H_0 : Se_1 = Se_2$$

$$H_1 : Se_1 \neq Se_2$$

es

$$z_{\text{exp}} = \frac{|\hat{Se}_1 - \hat{Se}_2| - c}{\sqrt{\hat{Se}(1 - \hat{Se})\left(\frac{1}{s_{1\cdot}} + \frac{1}{r_{1\cdot}}\right)}} \rightarrow N(0,1), \quad (2.1)$$

donde $\hat{Se}_1 = \frac{s_{11}}{s_{1\cdot}}$, $\hat{Se}_2 = \frac{r_{11}}{r_{1\cdot}}$, $\hat{Se} = \frac{s_{11} + r_{11}}{s_{1\cdot} + r_{1\cdot}}$, y $c = \frac{1}{s_{1\cdot} r_{1\cdot}}$ si $s_{1\cdot} \neq r_{1\cdot}$ o $c = \frac{2}{s_{1\cdot} r_{1\cdot}}$ si $s_{1\cdot} = r_{1\cdot}$.

Esta aproximación es válida si $\frac{\text{Max}(s_{11} + r_{11}, s_{10} + r_{10}) \times \text{Max}(s_{1\cdot}, r_{1\cdot})}{s_{1\cdot} + r_{1\cdot}} \geq 14.9$ (Martín

Andrés et al., 2005). De forma análoga, el estadístico de contraste para el test de hipótesis

$$H_0 : Sp_1 = Sp_2$$

$$H_1 : Sp_1 \neq Sp_2$$

es

$$z_{\text{exp}} = \frac{|\hat{Sp}_1 - \hat{Sp}_2| - c}{\sqrt{\hat{Sp}(1 - \hat{Sp})\left(\frac{1}{s_{0\cdot}} + \frac{1}{r_{0\cdot}}\right)}} \rightarrow N(0,1), \quad (2.2)$$

en cuyo caso $\hat{Sp}_1 = \frac{s_{00}}{s_{0\cdot}}$, $\hat{Sp}_2 = \frac{r_{00}}{r_{0\cdot}}$, $\hat{Sp} = \frac{s_{00} + r_{00}}{s_{0\cdot} + r_{0\cdot}}$, y $c = \frac{1}{s_{0\cdot}r_{0\cdot}}$ si $s_{0\cdot} \neq r_{0\cdot}$ o $c = \frac{2}{s_{0\cdot}r_{0\cdot}}$ si

$s_{0\cdot} = r_{0\cdot}$. Esta aproximación es válida si $\frac{\text{Max}(s_{01} + r_{01}, s_{00} + r_{00}) \times \text{Max}(s_{0\cdot}, r_{0\cdot})}{s_{0\cdot} + r_{0\cdot}} \geq 14.9$

(Martín Andrés et al., 2005).

En la práctica clínica, la comparación de la exactitud de dos tests binarios con muestras independientes se utiliza con poca frecuencia, ya que en el caso de que un test de hipótesis sea significativo, la significación se puede deber a otras causas no relacionadas con la habilidad intrínseca de cada test binario para distinguir entre pacientes enfermos y pacientes sanos.

2.1.1.2. Comparación con muestras apareadas

La comparación de las sensibilidades y especificidades de dos tests binarios también se puede realizar bajo un diseño apareado. Si los dos tests binarios y el gold estándar se aplican a todos los individuos de una muestra de tamaño n se obtiene la Tabla 2.2 y el modelo teórico que subyace se muestra en la Tabla 2.3.

Los datos de la Tabla 2.2 son la realización de una distribución multinomial, de manera que la probabilidad de esa realización viene dada por la expresión

$$n! \prod_{i=0}^1 \prod_{j=0}^1 \frac{p_{ij}^{s_{ij}} q_{ij}^{r_{ij}}}{s_{ij}! r_{ij}!}. \quad (2.3)$$

Tabla 2.2. Frecuencias observadas al aplicar dos tests binarios a una muestra.

	$T_1 = 1$		$T_1 = 0$		Total
	$T_2 = 1$	$T_2 = 0$	$T_2 = 1$	$T_2 = 0$	
$D = 1$	s_{11}	s_{10}	s_{01}	s_{00}	s
$D = 0$	r_{11}	r_{10}	r_{01}	r_{00}	r
Total	n_{11}	n_{10}	n_{01}	n_{00}	n

Tabla 2.3. Probabilidades teóricas al aplicar dos tests binarios a una muestra.

	$T_1 = 1$		$T_1 = 0$		
	$T_2 = 1$	$T_2 = 0$	$T_2 = 1$	$T_2 = 0$	
$D = 1$	p_{11}	p_{10}	p_{01}	p_{00}	p
$D = 0$	q_{11}	q_{10}	q_{01}	q_{00}	q
	$p_{11} + q_{11}$	$p_{10} + q_{10}$	$p_{01} + q_{01}$	$p_{00} + q_{00}$	1

A partir de la Tabla 2.3, la sensibilidad y la especificidad del test 1 son

$$Se_1 = \frac{p_{11} + p_{10}}{p} \quad \text{y} \quad Sp_1 = \frac{q_{01} + q_{00}}{q}, \quad (2.4)$$

respectivamente, y las del test 2 son

$$Se_2 = \frac{p_{11} + p_{01}}{p} \quad \text{y} \quad Sp_2 = \frac{q_{01} + q_{00}}{q}. \quad (2.5)$$

El test de hipótesis para contrastar la igualdad de las sensibilidades

$$H_0 : Se_1 = Se_2$$

$$H_1 : Se_1 \neq Se_2$$

es equivalente al test de hipótesis

$$\begin{aligned} H_0 : p_{11} + p_{10} = p_{11} + p_{01} &\Leftrightarrow H_0 : p_{10} = p_{01} \\ H_1 : p_{11} + p_{10} \neq p_{11} + p_{01} &H_1 : p_{10} \neq p_{01}, \end{aligned}$$

por lo que condicionando en los pares discordantes, si $p_{10} = p_{01} \Rightarrow p_{10} = p_{01} = 0.5$, el contraste de hipótesis de igualdad de las sensibilidades es equivalente a contrastar

$$H_0 : p_{10} = 0.5$$

$$H_1 : p_{10} \neq 0.5.$$

Por tanto, el test de hipótesis de comparación de las sensibilidades de dos tests binarios con muestras apareadas es equivalente a un test de hipótesis para una proporción binomial. Bajo la hipótesis nula se verifica que $s_{10} \rightarrow B(s_{10} + s_{01}, 0.5)$ y si

$(s_{10} + s_{01})/2$ es mayor que 5 entonces $s_{10} \rightarrow N\left((s_{10} + s_{01})/2, \sqrt{\frac{(s_{10} + s_{01})}{4}}\right)$, y

finalmente el estadístico para contrastar la igualdad de las dos sensibilidades es

$$z_{\text{exp}} = \frac{|s_{10} - s_{01}| - 1}{\sqrt{s_{10} + s_{01}}} \rightarrow N(0, 1). \quad (2.6)$$

Esta aproximación es válida si $s_{10} + s_{01} > 10$ y el test de hipótesis se denomina test de McNemar. De forma análoga, el test de hipótesis de la igualdad de las dos especificidades

$$H_0 : Sp_1 = Sp_2$$

$$H_1 : Sp_1 \neq Sp_2$$

es equivalente al test de hipótesis

$$H_0 : q_{10} = 0.5$$

$$H_1 : q_{10} \neq 0.5,$$

y aplicando el test de McNemar, el estadístico de contraste es

$$z_{\text{exp}} = \frac{|r_{10} - r_{01}| - 1}{\sqrt{r_{10} + r_{01}}} \rightarrow N(0, 1), \quad (2.7)$$

válido para $r_{01} + r_{10} > 10$.

En la práctica clínica, el diseño muestral que más se utiliza para comparar el rendimiento de dos tests binarios es el diseño con muestras apareadas (los tests binarios se aplican a todos los individuos de una misma muestra aleatoria). Por consiguiente, en lo que sigue se va a emplear este tipo de diseño muestral.

2.1.1.3. Intervalos de confianza para la diferencia de las sensibilidades y de las especificidades con muestras apareadas

A continuación se describen varios intervalos de confianza para la diferencia de las sensibilidades (especificidades) de dos tests diagnósticos binarios, cuando ambos tests se aplican a todos los sujetos de una muestra aleatoria de tamaño n .

a. Intervalo de confianza de Wald:

$$Se_1 - Se_2 \in \hat{Se}_1 - \hat{Se}_2 \pm z_{1-\alpha/2} \sqrt{\frac{\hat{Se}_1(1-\hat{Se}_1) + \hat{Se}_2(1-\hat{Se}_2) + 2(\hat{Se}_1\hat{Se}_2 - \hat{p}_{11})}{s}}, \quad (2.8)$$

donde $z_{1-\alpha/2}$ es el $(1-\alpha/2)$ -ésimo percentil de la distribución normal estándar,

$\hat{p}_{11} = \hat{P}(T_1 = 1, T_2 = 1)$ y s es el número de sujetos enfermos.

b. Intervalo de confianza de Wald con corrección por continuidad:

$$Se_1 - Se_2 \in \hat{Se}_1 - \hat{Se}_2 \pm \left\{ z_{1-\alpha/2} \sqrt{\frac{\hat{Se}_1(1-\hat{Se}_1) + \hat{Se}_2(1-\hat{Se}_2) + 2(\hat{Se}_1\hat{Se}_2 - \hat{p}_{11})}{s}} + \frac{1}{s} \right\}. \quad (2.9)$$

c. *Intervalo de confianza de Wald +2 (Agresti y Min, 2005):*

$$Se_1 - Se_2 \in \frac{s_{10} - s_{01}}{s + 2} \pm z_{1-\alpha/2} \frac{\sqrt{(s_{10} + s_{01} + 1) - \frac{(s_{10} - s_{01})^2}{s + 2}}}{s + 2}. \quad (2.10)$$

d. *Intervalo de confianza de Zhou y Qin (2005):*

$$Se_1 - Se_2 \in \left[\text{Max} \left(-1, \hat{Se}_1 - \hat{Se}_2 - \frac{\hat{\sigma}}{\sqrt{n}} g^{-1}(z_{1-\alpha/2}) \right), \text{Min} \left(1, \hat{Se}_1 - \hat{Se}_2 - \frac{\hat{\sigma}}{\sqrt{n}} g^{-1}(z_{\alpha/2}) \right) \right], \quad (2.11)$$

siendo

$$\begin{aligned} \hat{\sigma} &= \sqrt{\hat{Se}_1(1 - \hat{Se}_1) + \hat{Se}_2(1 - \hat{Se}_2) + 2(\hat{Se}_1\hat{Se}_2 - \hat{p}_{11})}, \\ \hat{a} &= \frac{\hat{d}}{6\hat{\sigma}^2}, \quad \hat{b} = \frac{1 - 2(\hat{Se}_1 - \hat{Se}_2)}{2} - \frac{\hat{d}}{6\hat{\sigma}^2}, \\ \hat{d} &= \hat{Se}_2(1 - \hat{Se}_2)(1 - 2\hat{Se}_2) - \hat{Se}_1(1 - \hat{Se}_1)(1 - 2\hat{Se}_1) + 6(\hat{Se}_2 - \hat{Se}_1)(\hat{p}_{11} - \hat{Se}_1\hat{Se}_2), \\ g^{-1}(z) &= \begin{cases} \frac{\sqrt{n}}{\hat{b}\hat{\sigma}} \left[\left(1 + 3\hat{b}\hat{\sigma} \left(\frac{z}{\sqrt{n}} - \frac{\hat{a}\hat{\sigma}}{n} \right) \right)^{1/3} \right], & \hat{b}\hat{\sigma} \neq 0 \\ z - \frac{\hat{a}\hat{\sigma}}{n}, & \hat{b}\hat{\sigma} = 0. \end{cases} \end{aligned}$$

Los intervalos de confianza para la diferencia de las dos especificidades se obtienen sustituyendo Se por Sp , \hat{p}_{11} por $\hat{p}_{00} = \hat{P}(T_1 = 0, T_2 = 0)$, s_{ij} por r_{ij} , y s por r .

2.1.2. Comparación de las razones de verosimilitud

A continuación se estudia la comparación de las razones de verosimilitud, tanto positivas como negativas, de dos tests diagnósticos binarios en diseños apareados. Este problema de inferencia ha sido estudiado por Roldán Nofuentes y Luna del Castillo (2007d). Cuando los resultados de los dos tests binarios son positivos, el test de hipótesis para contrastar la igualdad de las razones de verosimilitud es

$$\begin{aligned} H_0 : \omega^+ &= 0 \\ H_1 : \omega^+ &\neq 0, \end{aligned}$$

y cuando los resultados de ambos tests son negativos, el test de hipótesis es

$$\begin{aligned} H_0 : \omega^- &= 0 \\ H_1 : \omega^- &\neq 0, \end{aligned}$$

donde $\omega^+ = \log(LR_1^+/LR_2^+)$ y $\omega^- = \log(LR_1^-/LR_2^-)$. Los cocientes LR_1^+/LR_2^+ y LR_1^-/LR_2^- se denominan razones de verosimilitud relativas y han sido estudiadas por Pepe (2003). A continuación se obtienen los estimadores máximo verosímiles de las razones de verosimilitud y el estadístico de contraste del test $H_0 : \omega = 0$ frente a $H_1 : \omega \neq 0$, donde ω es ω^+ o ω^- .

Sean las probabilidades condicionadas $\theta_{ij} = P(D=1|T_1=i, T_2=j)$ y las probabilidades $\eta_{ij} = (T_1=i, T_2=j)$, con $i, j = 0, 1$ y $\eta_{11} = 1 - \eta_{00} - \eta_{01} - \eta_{10}$. Sean los vectores $\boldsymbol{\theta} = (\theta_{00}, \theta_{01}, \theta_{10}, \theta_{11})^T$ y $\boldsymbol{\eta} = (\eta_{00}, \eta_{01}, \eta_{10})^T$. Las razones de verosimilitud de cada test diagnóstico se pueden escribir en términos de las probabilidades θ_{ij} y η_{ij} como

$$LR_1^+ = \left(\frac{1 - \sum_{i,j=0}^1 \theta_{ij} \eta_{ij}}{\sum_{i,j=0}^1 \theta_{ij} \eta_{ij}} \right) \left(\frac{\sum_{j=0}^1 \theta_{1j} \eta_{1j}}{\sum_{j=0}^1 (1 - \theta_{1j}) \eta_{1j}} \right) \text{ y } LR_1^- = \left(\frac{1 - \sum_{i,j=0}^1 \theta_{ij} \eta_{ij}}{\sum_{i,j=0}^1 \theta_{ij} \eta_{ij}} \right) \left(\frac{\sum_{j=0}^1 \theta_{0j} \eta_{0j}}{\sum_{j=0}^1 (1 - \theta_{0j}) \eta_{0j}} \right) \quad (2.12)$$

para el test 1, y

$$LR_2^+ = \left(\frac{1 - \sum_{i,j=0}^1 \theta_{ij} \eta_{ij}}{\sum_{i,j=0}^1 \theta_{ij} \eta_{ij}} \right) \left(\frac{\sum_{i=0}^1 \theta_{i1} \eta_{i1}}{\sum_{i=0}^1 (1 - \theta_{i1}) \eta_{i1}} \right) \text{ y } LR_2^- = \left(\frac{1 - \sum_{i,j=0}^1 \theta_{ij} \eta_{ij}}{\sum_{i,j=0}^1 \theta_{ij} \eta_{ij}} \right) \left(\frac{\sum_{i=0}^1 \theta_{i0} \eta_{i0}}{\sum_{i=0}^1 (1 - \theta_{i0}) \eta_{i0}} \right) \quad (2.13)$$

para el test 2. El logaritmo de la función de verosimilitud de los datos de la Tabla 2.1 es

$$\begin{aligned}
 l(\boldsymbol{\theta}, \boldsymbol{\eta}) &= \sum_{i,j=0}^1 s_{ij} \log \{P(D=1, T_1=i, T_2=j)\} + \sum_{i,j=0}^1 r_{ij} \log \{P(D=0, T_1=i, T_2=j)\} = \\
 & \sum_{i,j=0}^1 s_{ij} \log \{\theta_{ij} \eta_{ij}\} + \sum_{i,j=0}^1 r_{ij} \log \{(1-\theta_{ij}) \eta_{ij}\} = \\
 & \sum_{i,j=0}^1 \{s_{ij} \log \theta_{ij} + r_{ij} \log (1-\theta_{ij})\} + \sum_{i,j=0}^1 n_{ij} \log (\eta_{ij}) = l(\boldsymbol{\theta}) + l(\boldsymbol{\eta}),
 \end{aligned} \tag{2.14}$$

donde $n_{ij} = s_{ij} + r_{ij}$. Maximizando la función (2.14), los estimadores máximo verosímiles de θ_{ij} y η_{ij} vienen dados por las expresiones

$$\hat{\theta}_{ij} = \frac{s_{ij}}{n_{ij}} \quad \text{y} \quad \hat{\eta}_{ij} = \frac{n_{ij}}{n}, \tag{2.15}$$

con $i, j = 0, 1$ y la matriz de información de Fisher de $l(\boldsymbol{\theta}, \boldsymbol{\eta})$ viene dada por

$$I_{(\boldsymbol{\theta}, \boldsymbol{\eta})} = \text{diag} \{I_{\boldsymbol{\theta}}, I_{\boldsymbol{\eta}}\}, \tag{2.16}$$

donde $I_{\boldsymbol{\theta}}$ y $I_{\boldsymbol{\eta}}$ son matrices de información de Fisher de las funciones $l(\boldsymbol{\theta})$, y cuyas expresiones son (Zhou, 1998b)

$$I_{\alpha}^{-1} = \text{diag} \left\{ \frac{\theta_{ij}^2 (1-\theta_{ij})^2}{s_{ij} (1-\theta_{ij})^2 + r_{ij} \theta_{ij}^2} \right\}, \quad i, j = 0, 1 \tag{2.17}$$

y

$$I_{\boldsymbol{\eta}}^{-1} = \text{diag} \left(\frac{n_{00}^2}{n_{00}}, \frac{n_{01}^2}{n_{01}}, \frac{n_{10}^2}{n_{10}} \right) - \frac{1}{\sum_{i,j=0}^1 \frac{\eta_{ij}^2}{n_{ij}}} \left(\frac{n_{00}^2}{n_{00}}, \frac{n_{01}^2}{n_{01}}, \frac{n_{10}^2}{n_{10}} \right)^T \left(\frac{n_{00}^2}{n_{00}}, \frac{n_{01}^2}{n_{01}}, \frac{n_{10}^2}{n_{10}} \right). \tag{2.18}$$

Como ω^+ y ω^- son funciones de las probabilidades θ_{ij} y η_{ij} , aplicando el método delta (Agresti, 2002) la varianza de ω es

$$\text{Var}(\omega) = \left(\frac{\partial \omega}{\partial \boldsymbol{\theta}} \right) I_{\boldsymbol{\theta}}^{-1} \left(\frac{\partial \omega}{\partial \boldsymbol{\theta}} \right)^T + \left(\frac{\partial \omega}{\partial \boldsymbol{\eta}} \right) I_{\boldsymbol{\eta}}^{-1} \left(\frac{\partial \omega}{\partial \boldsymbol{\eta}} \right)^T, \tag{2.19}$$

donde ω es ω^+ o ω^- , y realizando las operaciones matriciales en la ecuación (2.19) se obtiene que

$$\text{Var}(\omega) = \sum_{i,j=0}^1 \frac{\theta_{ij}^2 (1-\theta_{ij})^2}{s_{ij} (1-\theta_{ij})^2 + r_{ij} \theta_{ij}^2} \left(\frac{\partial \omega}{\partial \theta_{ij}} \right)^2 + \sum_{(i,j) \neq (1,1)} \frac{\eta_{ij}^2}{n_{ij}} \left(\frac{\partial \omega}{\partial \eta_{ij}} \right)^2 - \frac{1}{\sum_{i,j=0}^1 \frac{\eta_{ij}^2}{n_{ij}}} \left(\sum_{(i,j) \neq (1,1)} \frac{\eta_{ij}^2}{n_{ij}} \frac{\partial \omega}{\partial \eta_{ij}} \right)^2. \quad (2.20)$$

Sustituyendo en las ecuaciones (2.12) y (2.13) los parámetros θ_{ij} y η_{ij} por sus estimadores máximo verosímiles dados en las ecuaciones (2.15) y aplicando el teorema de Zehna (1966), los estimadores máximo verosímiles de las razones de verosimilitud del test 1 vienen dados por las expresiones

$$\widehat{LR}_1^+ = \frac{(s_{10} + s_{11})/s}{(r_{10} + r_{11})/r} \quad \text{y} \quad \widehat{LR}_1^- = \frac{(s_{00} + s_{01})/s}{(r_{00} + r_{01})/r}, \quad (2.21)$$

y

$$\widehat{LR}_2^+ = \frac{(s_{01} + s_{11})/s}{(r_{01} + r_{11})/r} \quad \text{y} \quad \widehat{LR}_2^- = \frac{(s_{00} + s_{10})/s}{(r_{00} + r_{10})/r} \quad (2.22)$$

para el test 2, donde $s = \sum_{i,j=0}^1 s_{ij}$ y $r = \sum_{i,j=0}^1 r_{ij}$. Finalmente, los estimadores máximo verosímiles de ω^+ y ω^- son

$$\omega^+ = \log \left\{ \frac{(s_{10} + s_{11})(r_{01} + r_{11})}{(s_{01} + s_{11})(r_{10} + r_{11})} \right\} \quad \text{y} \quad \omega^- = \log \left\{ \frac{(s_{01} + s_{00})(r_{10} + r_{00})}{(s_{10} + s_{00})(r_{01} + r_{00})} \right\}, \quad (2.23)$$

y el estadístico para contrastar $H_0 : \omega = 0$ frente $H_1 : \omega \neq 0$ es

$$\frac{\hat{\omega}}{\sqrt{\widehat{\text{Var}}(\hat{\omega})}} \rightarrow N(0,1). \quad (2.24)$$

2.1.2.1. Test de hipótesis global

Los tests de hipótesis deducidos anteriormente comparan los logaritmos de las razones de verosimilitud positivas y negativas de una forma separada. Sin embargo, las razones de verosimilitud positiva y negativa no son independientes, y por consiguiente, es necesario obtener un test de hipótesis conjunto para comparar ambas medidas. Como los datos de la Tabla 2.2 son la realización de una distribución multinomial, aplicando el teorema central del límite multivariante se obtiene que

$$\frac{1}{\sqrt{n}} \begin{pmatrix} \hat{\omega}^+ - \omega^+ \\ \hat{\omega}^- - \omega^- \end{pmatrix} \xrightarrow{n \rightarrow \infty} N \left[\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_{\omega} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \right], \quad (2.25)$$

donde los elementos de la matriz de varianzas-covarianzas se calculan aplicando el método delta. De esta forma, si $\zeta = (\omega^+, \omega^-)^T$ entonces

$$\Sigma_{\zeta} = \left(\frac{\partial \zeta}{\partial \theta} \right) I_{\theta}^{-1} \left(\frac{\partial \zeta}{\partial \theta} \right)^T + \left(\frac{\partial \zeta}{\partial \eta} \right) I_{\eta}^{-1} \left(\frac{\partial \zeta}{\partial \eta} \right)^T, \quad (2.26)$$

y el estadístico de contraste para el test de hipótesis global

$$\begin{aligned} H_0 : \omega^- = 0 \text{ y } \omega^+ = 0 \\ H_1 : \omega^- \neq 0 \text{ y/o } \omega^+ \neq 0 \end{aligned}$$

es

$$Q_{\text{exp}}^2 = \hat{\zeta}^T \hat{\Sigma}_{\zeta}^{-1} \hat{\zeta} \rightarrow \chi_2^2, \quad (2.27)$$

donde $\hat{\Sigma}_{\zeta}$ se obtiene sustituyendo en Σ_{ζ} los parámetros por sus estimadores máximos verosímiles.

2.1.3. Comparación de los valores predictivos

Los valores predictivos representan la exactitud del test diagnóstico cuando éste se aplica a una cohorte de individuos, y por tanto son unos parámetros válidos para

comparar el rendimiento de dos tests binarios. Bennett (1972) ha estudiado la comparación de los valores predictivos de dos tests binarios con muestras apareadas, proponiendo un test de hipótesis basado en la distribución chi-cuadrado. Jamart (1993) ha demostrado que el método de Bennett para comparar los valores predictivos negativos de dos tests binarios depende del orden en que los tests diagnósticos son numerados, y por tanto es un método que presenta algunas incoherencias. Leisenring et al. (2000) han estudiado la estimación y comparación de los valores predictivos mediante modelos GEE. Wang et al. (2006) han estudiado la comparación de los valores predictivos de dos tests binarios en diseños apareados mediante mínimos cuadrados ponderados, y han comparado este método con el de Leisenring et al. obteniendo que ambos métodos son similares en términos de error tipo I y potencia. A continuación se describe el método de comparación de los valores predictivos de Wang et al. (2006), adaptado a la notación de las Tablas 2.2 y 2.3.

En términos de las probabilidades de la Tabla 2.3, los valores predictivos positivo y negativo del test 1 son

$$vpp_1 = \frac{p_{11} + p_{10}}{p_{11} + p_{10} + q_{11} + q_{10}} \quad \text{y} \quad vpn_1 = \frac{q_{01} + q_{00}}{p_{01} + p_{00} + q_{01} + q_{00}}, \quad (2.28)$$

respectivamente, y los del test 2 son

$$vpp_2 = \frac{p_{11} + p_{01}}{p_{11} + p_{01} + q_{11} + q_{01}} \quad \text{y} \quad vpn_2 = \frac{q_{10} + q_{00}}{p_{10} + p_{00} + q_{10} + q_{00}}. \quad (2.29)$$

Sean γ y π dos vectores de dimensión 8 definidos como $\gamma = (s_{11}, s_{10}, s_{01}, s_{00}, r_{11}, r_{10}, r_{01}, r_{00})^T$ y $\pi = (p_{11}, p_{10}, p_{01}, p_{00}, q_{11}, q_{10}, q_{01}, q_{00})^T$. Como π es el vector de probabilidades de una distribución multinomial, la matriz de varianzas-covarianzas de $\hat{\pi}$ es

$$\Sigma_{\hat{\pi}} = \{\text{Diag}(\pi) - \pi\pi^T\} / n, \quad (2.30)$$

y el estimador de π es $\hat{\pi} = \gamma/n$. Sea vpp la diferencia entre los valores predictivos positivos de los dos tests binarios, esto es

$$vpp = vpp_1 - vpp_2 = \frac{p_{11} + p_{10}}{p_{11} + p_{10} + q_{11} + q_{10}} - \frac{p_{11} + p_{01}}{p_{11} + p_{01} + q_{11} + q_{01}}, \quad (2.31)$$

su estimador es

$$\widehat{vpp} = \widehat{vpp}_1 - \widehat{vpp}_2 = \frac{s_{11} + s_{10}}{s_{11} + s_{10} + r_{11} + r_{10}} - \frac{s_{11} + s_{01}}{s_{11} + s_{01} + r_{11} + r_{01}}. \quad (2.32)$$

Como la diferencia vpp es una función de las probabilidades del vector $\boldsymbol{\pi}$, aplicando el método delta (Agresti, 2002) la varianza asintótica de \widehat{vpp} es

$$Var(\widehat{vpp}) = \left(\frac{\partial vpp}{\partial \boldsymbol{\pi}} \right) \Sigma_{\hat{\boldsymbol{\pi}}} \left(\frac{\partial vpp}{\partial \boldsymbol{\pi}} \right)^T, \quad (2.33)$$

y el estadístico para el test de hipótesis

$$\begin{aligned} H_0 : vpp_1 &= vpp_2 \\ H_1 : vpp_1 &\neq vpp_2 \end{aligned}$$

es

$$z_{\text{exp}} = \frac{\widehat{vpp}}{\sqrt{\widehat{Var}(\widehat{vpp})}} \rightarrow N(0,1), \quad (2.34)$$

donde $\widehat{Var}(\widehat{vpp})$ se obtiene sustituyendo la ecuación (2.33) los parámetros por sus estimadores máximos verosímiles. De forma similar se resuelve el test de hipótesis de igualdad de los valores predictivos negativos

$$\begin{aligned} H_0 : vpn_1 &= vpn_2 \\ H_1 : vpn_1 &\neq vpn_2, \end{aligned}$$

siendo en este caso el estadístico de contraste

$$z_{\text{exp}} = \frac{\widehat{vpn}}{\sqrt{\widehat{Var}(\widehat{vpn})}} \rightarrow N(0,1), \quad (2.35)$$

con $\widehat{vpn} = \widehat{vpn}_1 - \widehat{vpn}_2$. Estos dos tests de hipótesis también se pueden resolver aplicando la transformación logarítmica sobre los valores predictivos.

2.1.4. Comparación de los coeficientes kappa ponderados

Bloch (1997) ha estudiado la comparación de los coeficientes kappa ponderados de dos tests binarios cuando los dos tests y el gold estándar se aplican a todos los individuos de una muestra aleatoria. Adaptando el método de Bloch a la terminología de las Tablas 2.2 y 2.3, en términos de las probabilidades del vector $\boldsymbol{\pi}$ definido en la Sección anterior, el coeficiente kappa ponderado del test 1 es

$$\kappa_1(c) = \frac{q \sum_{j=0}^1 p_{1j} + p \sum_{j=0}^1 q_{0j} - pq}{cp \left(1 - \sum_{j=0}^1 p_{1j} - \sum_{j=0}^1 q_{1j} \right) + (1-c)q \left(\sum_{j=0}^1 p_{1j} + \sum_{j=0}^1 q_{1j} \right)}, \quad (2.36)$$

y el del test 2 es

$$\kappa_2(c) = \frac{q \sum_{i=0}^1 p_{i1} + p \sum_{i=0}^1 q_{i0} - pq}{cp \left(1 - \sum_{i=0}^1 p_{i1} - \sum_{i=0}^1 q_{i1} \right) + (1-c)q \left(\sum_{i=0}^1 p_{i1} + \sum_{i=0}^1 q_{i1} \right)}, \quad (2.37)$$

y sus correspondientes estimadores máximo verosímiles son

$$\hat{\kappa}_1(c) = \frac{r \sum_{j=0}^1 s_{1j} + s \sum_{j=0}^1 r_{0j} - sr}{cs \left(n - \sum_{j=0}^1 s_{1j} - \sum_{j=0}^1 r_{1j} \right) + (1-c)r \left(\sum_{j=0}^1 s_{1j} + \sum_{j=0}^1 r_{1j} \right)}, \quad (2.38)$$

y

$$\hat{\kappa}_2(c) = \frac{r \sum_{i=0}^1 s_{i1} + s \sum_{i=0}^1 r_{i0} - sr}{cs \left(n - \sum_{i=0}^1 s_{i1} - \sum_{i=0}^1 r_{i1} \right) + (1-c)r \left(\sum_{i=0}^1 s_{i1} + \sum_{i=0}^1 r_{i1} \right)}, \quad (2.39)$$

respectivamente. Sea el vector $\boldsymbol{\kappa} = (\kappa_1(c), \kappa_2(c))^T$, aplicando de nuevo el método delta la matriz de varianzas–covarianzas asintóticas de $\hat{\boldsymbol{\kappa}}$ es

$$\Sigma_{\hat{\boldsymbol{\kappa}}} = \left(\frac{\partial \boldsymbol{\kappa}}{\partial \boldsymbol{\pi}} \right) \Sigma_{\hat{\boldsymbol{\pi}}} \left(\frac{\partial \boldsymbol{\kappa}}{\partial \boldsymbol{\pi}} \right)^T, \quad (2.40)$$

donde $\Sigma_{\hat{\boldsymbol{\pi}}}$ es la matriz definida en la ecuación (2.30). Finalmente, el estadístico para contrastar

$$\begin{aligned} H_0 &: \kappa_1(c) = \kappa_2(c) \\ H_1 &: \kappa_1(c) \neq \kappa_2(c) \end{aligned}$$

es

$$z_{\text{exp}} = \frac{\hat{\kappa}_1(c) - \hat{\kappa}_2(c)}{\sqrt{\widehat{\text{Var}}(\hat{\kappa}_1(c)) + \widehat{\text{Var}}(\hat{\kappa}_2(c)) - 2\widehat{\text{Cov}}(\hat{\kappa}_1(c), \hat{\kappa}_2(c))}} \rightarrow N(0,1), \quad (2.41)$$

donde los estimadores de las varianzas y de la covarianza se obtienen sustituyendo en la ecuación (2.40) los parámetros por sus estimadores máximo verosímiles.

2.1.5. Comparación de múltiples coeficientes kappa ponderados (Aportación)

Como se ha comentado en la Sección 1.1.5, el coeficiente kappa ponderado de un test diagnóstico binario es una medida del acuerdo clasificatorio entre el test diagnóstico y el gold estándar, y es un parámetro válido para evaluar y comparar el rendimiento de dos o más tests diagnósticos binarios. Bloch (1997) ha deducido los tests de hipótesis para comparar los coeficientes kappa de dos tests diagnósticos binarios en diseños apareados. Roldán Nofuentes y Luna del Castillo (2005) han estudiado la comparación de los coeficientes kappa ponderados de dos tests binarios en presencia del sesgo de verificación. Roldán Nofuentes y Luna del Castillo (2006) han estudiado la estimación del coeficiente kappa ponderado de un test binario en presencia de verificación parcial de la enfermedad.

En la práctica clínica es frecuente que la estimación y comparación del rendimiento de más de dos tests binarios cuando todos los tests y el gold estándar se aplican a todos los pacientes de una misma muestra. El método de Bloch (1997) no permite comparar los coeficientes kappa ponderados de más de dos tests binarios, por tanto, es necesario desarrollar métodos que permitan resolver este problema de inferencia. A continuación se deduce un test de hipótesis conjunto para comparar simultáneamente los coeficientes kappa ponderados de múltiples tests binarios cuando todos los tests diagnósticos y el gold estándar se aplican a una misma muestra de pacientes. Se han realizado unos experimentos de simulación para estudiar el error tipo I y la potencia del test de hipótesis conjunto, y los resultados obtenidos se han aplicado al estudio de Weiner et al. (1979) sobre el diagnóstico de la enfermedad coronaria.

2.1.5.1. Test de hipótesis conjunto

Sean J tests diagnósticos binarios ($J \geq 3$) que se aplican de forma independiente a una misma muestra aleatoria de n pacientes. Sea T_j la variable aleatoria que modeliza el resultado del j -ésimo test binario ($j=1, \dots, J$), de tal forma que $T_j=1$ cuando el resultado del test es positivo y $T_j=0$ cuando el resultado es negativo. Sea D la variable aleatoria que modeliza el resultado del gold estándar, de tal forma que $D=1$ cuando el paciente está enfermo y $D=0$ cuando el paciente está sano. Sea s_{i_1, \dots, i_J} el número de pacientes enfermos en los que $T_1=i_1, T_2=i_2, \dots, T_J=i_J$, con $i_j=0, 1$ y $j=1, \dots, J$; y sea r_{i_1, \dots, i_J} el número de pacientes en los que $T_1=i_1, T_2=i_2, \dots, T_J=i_J$, con

$i_j=0, 1$ y $j=1, \dots, J$. Sean $s = \sum_{i_1, \dots, i_J=0}^1 s_{i_1, \dots, i_J}$, $r = \sum_{i_1, \dots, i_J=0}^1 r_{i_1, \dots, i_J}$, $n_{i_1, \dots, i_J} = s_{i_1, \dots, i_J} + r_{i_1, \dots, i_J}$ y

$n = \sum_{i_1, \dots, i_J=0}^1 n_{i_1, \dots, i_J}$. Sean las probabilidades $p_{i_1, \dots, i_J} = P(D=1, T_1=i_1, T_2=i_2, \dots, T_J=i_J)$ y

$q_{i_1, \dots, i_J} = P(D=0, T_1=i_1, T_2=i_2, \dots, T_J=i_J)$, tal que $\sum_{i_1, \dots, i_J=0}^1 p_{i_1, \dots, i_J} + \sum_{i_1, \dots, i_J=0}^1 q_{i_1, \dots, i_J} = 1$. La

prevalencia de la enfermedad es

$$p = P(D=1) = \sum_{i_1, \dots, i_j=0}^1 p_{i_1, \dots, i_j} \quad (2.42)$$

y la probabilidad de que un paciente no esté enfermo es $q = P(D=0) = \sum_{i_1, \dots, i_j=0}^1 q_{i_1, \dots, i_j}$. A

continuación se deducen los estimadores máximo verosímiles de los coeficientes kappa ponderados de cada test diagnóstico binario y el test de hipótesis conjunto para comparar la igualdad de todos los J coeficientes kappa ponderados. La sensibilidad del j -ésimo test diagnóstico es

$$Se_j = P(T_j = 1 | D = 1) = \frac{\sum_{\substack{i_1, \dots, i_j=0 \\ i_j=1}}^1 p_{i_1, \dots, i_j}}{\sum_{i_1, \dots, i_j=0}^1 p_{i_1, \dots, i_j}} \quad (2.43)$$

y la especificidad

$$Sp_j = P(T_j = 0 | D = 0) = \frac{\sum_{\substack{i_1, \dots, i_j=0 \\ i_j=0}}^1 q_{i_1, \dots, i_j}}{\sum_{i_1, \dots, i_j=0}^1 q_{i_1, \dots, i_j}}, \quad (2.44)$$

y por tanto el coeficiente kappa ponderado del j -ésimo test binario es

$$\kappa_j(c) = \frac{pq(Se_j + Sp_j - 1)}{p(1 - Q_j)c + qQ_j(1 - c)}, \quad (2.45)$$

donde $Q_j = pSe_j + (1 - p)(1 - Sp_j)$. Sustituyendo las expresiones (2.42), (2.43) y (2.44) en la ecuación (2.45) se obtiene que

$$\kappa_j(c) = \frac{q \left(\sum_{\substack{i_1, \dots, i_j=0 \\ i_j=1}}^1 p_{i_1, \dots, i_j} \right) + p \left(\sum_{\substack{i_1, \dots, i_j=0 \\ i_j=0}}^1 q_{i_1, \dots, i_j} \right) - pq}{p \left(p - \sum_{\substack{i_1, \dots, i_j=0 \\ i_j=1}}^1 p_{i_1, \dots, i_j} + \sum_{\substack{i_1, \dots, i_j=0 \\ i_j=0}}^1 q_{i_1, \dots, i_j} \right) c + q \left(q + \sum_{\substack{i_1, \dots, i_j=0 \\ i_j=1}}^1 p_{i_1, \dots, i_j} - \sum_{\substack{i_1, \dots, i_j=0 \\ i_j=0}}^1 q_{i_1, \dots, i_j} \right) (1 - c)}. \quad (2.46)$$

La función del logaritmo de la verosimilitud de los datos observados es

$$l = \sum_{i_1, \dots, i_J=0}^1 s_{i_1, \dots, i_J} \log(p_{i_1, \dots, i_J}) + \sum_{i_1, \dots, i_J=0}^1 r_{i_1, \dots, i_J} \log(q_{i_1, \dots, i_J}), \quad (2.47)$$

por lo que maximizando esta función, los estimadores máximo verosímiles de las probabilidades p_{i_1, \dots, i_J} y q_{i_1, \dots, i_J} son

$$\hat{p}_{i_1, \dots, i_J} = \frac{s_{i_1, \dots, i_J}}{n} \quad \text{y} \quad \hat{q}_{i_1, \dots, i_J} = \frac{r_{i_1, \dots, i_J}}{n}, \quad (2.48)$$

con $i_1, \dots, i_J = 0, 1$. Sustituyendo en la ecuación (2.46) los parámetros por sus estimadores máximo verosímiles dados en las ecuaciones (2.48) se obtienen los estimadores máximo verosímiles de los coeficientes kappa ponderados.

Sean $\boldsymbol{\omega} = (p_{1,1, \dots, 1}, \dots, p_{0,0, \dots, 0}, q_{1,1, \dots, 1}, \dots, q_{0,0, \dots, 0})^T$ y $\boldsymbol{\kappa} = (\kappa_1(r), \kappa_2(r), \dots, \kappa_J(r))^T$ vectores de dimensiones 2^{J+1} y J respectivamente. Como las probabilidades p_{i_1, \dots, i_J} y q_{i_1, \dots, i_J} , $i_1, \dots, i_J = 0, 1$, son probabilidades de una distribución multinomial, la matriz de varianzas–covarianzas de $\hat{\boldsymbol{\omega}}$ es

$$\Sigma_{\hat{\boldsymbol{\omega}}} = \frac{1}{n} (\text{diag}(\boldsymbol{\omega}) - \boldsymbol{\omega} \boldsymbol{\omega}^T), \quad (2.49)$$

y aplicando el método delta, la matriz de varianzas–covarianzas de $\hat{\boldsymbol{\kappa}}$ es

$$\Sigma_{\hat{\boldsymbol{\kappa}}} = \left(\frac{\partial \boldsymbol{\kappa}}{\partial \boldsymbol{\omega}} \right) \Sigma_{\hat{\boldsymbol{\omega}}} \left(\frac{\partial \boldsymbol{\kappa}}{\partial \boldsymbol{\omega}} \right)^T. \quad (2.50)$$

Sustituyendo en (2.50) los parámetros por sus correspondientes estimadores máximo verosímiles, se obtienen las varianzas–covarianzas estimadas de los estimadores máximo verosímiles de los coeficientes kappa ponderados.

Para un mismo valor del índice de ponderación c , el test de hipótesis conjunto para contrastar la igualdad de todos los coeficientes kappa ponderados es

$$H_0 : \kappa_1(c) = \kappa_2(c) = \dots = \kappa_J(c)$$

$$H_1 : \text{al menos una igualdad no es cierta.}$$

Este test de hipótesis es equivalente a contrastar

$$H_0 : \boldsymbol{\varphi}\boldsymbol{\kappa} = \mathbf{0}$$

$$H_1 : \boldsymbol{\varphi}\boldsymbol{\kappa} \neq \mathbf{0},$$

donde $\boldsymbol{\varphi}$ es una matriz de rango completo, de dimensión $(J-1) \times J$ y cuyos valores son constantes conocidas. Así por ejemplo si $J = 2$,

$$\boldsymbol{\varphi} = (1, -1),$$

si $J = 3$ entonces

$$\boldsymbol{\varphi} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix},$$

y si $J = 4$

$$\boldsymbol{\varphi} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

Finalmente, el estadístico para contrastar $H_0 : \boldsymbol{\varphi}\boldsymbol{\kappa} = \mathbf{0}$ frente a $H_1 : \boldsymbol{\varphi}\boldsymbol{\kappa} \neq \mathbf{0}$ es

$$Q_{\text{exp}}^2 = \hat{\boldsymbol{\kappa}}^T \boldsymbol{\varphi}^T \left(\boldsymbol{\varphi} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\kappa}}} \boldsymbol{\varphi}^T \right)^{-1} \boldsymbol{\varphi} \hat{\boldsymbol{\kappa}} \rightarrow \chi_H^2. \quad (2.51)$$

Cuando se comparan los coeficientes kappa ponderados de dos tests binarios ($J = 2$), el estadístico de contraste (2.51) es equivalente al deducido por Bloch (1997).

Para resolver el test de hipótesis conjunto, la matriz de varianzas–covarianzas de los coeficientes kappa ponderados se ha estimado aplicando el método delta. Una alternativa a este método consiste en estimar la matriz de varianzas–covarianzas mediante bootstrap. Por tanto, a partir de la muestra se obtienen K réplicas con reemplazamiento y a partir de estas K muestras se estima la matriz de varianzas–covarianzas entre los coeficientes kappa ponderados. Una vez estimada esta matriz, el estadístico de contraste viene dado por la ecuación (2.51).

El test de hipótesis conjunto se puede resolver mediante otros dos métodos alternativos. El primero consiste en realizar las comparaciones por parejas aplicando el método de Bloch (1997) al error α ; y el segundo consiste en aplicar el método de Bonferroni, es decir, realizar las comparaciones por parejas aplicando el método de Bloch (1997) al error $2\alpha/(J(J-1))$.

2.1.5.2. Experimentos de simulación

Se han realizado unos experimentos de simulación para estudiar los errores tipo I y las potencias de los tres métodos estudiados en la Sección anterior para resolver el test de hipótesis $H_0 : \boldsymbol{\phi\kappa} = \mathbf{0}$ cuando se comparan los coeficientes kappa ponderados de tres tests diagnósticos binarios. Estos experimentos han consistido en generar 10000 muestras aleatorias de distribuciones multinomiales con tamaños 100, 500, 1000 y 1500, y cuyas probabilidades se han calculado aplicando el método de Torrance – Rynard y Walter (1997), esto es,

$$P(T_1 = i_1, T_2 = i_2, T_3 = i_3 | D = 1) = \prod_{j=1}^3 Se_j^{i_j} (1 - Se_j)^{1-i_j} + \sum_{j,k,j < k}^3 (-1)^{|i_j - i_k|} \delta_{jk}$$

y

$$P(T_1 = i_1, T_2 = i_2, T_3 = i_3 | D = 0) = \prod_{j=1}^3 Sp_j^{1-i_j} (1 - Sp_j)^{i_j} + \sum_{j,k,j < k}^3 (-1)^{|i_j - i_k|} \varepsilon_{jk},$$

con $i_j = 0, 1$, $i_k = 0, 1$ y $j, k = 1, 2, 3$, y donde δ_{jk} es el parámetro de dependencia entre el j -ésimo y el k -ésimo test binario cuando el paciente está enfermo ($D = 1$), y ε_{jk} es el parámetro de dependencia entre el j -ésimo test binario y k -ésimo test binario cuando el paciente no está enfermo ($D = 0$). Los parámetros de dependencia δ_{jk} y ε_{jk} están sometidos a unas restricciones (Torrance – Rynard y Walter, 1997) que dependen de los valores de las sensibilidades y especificidades de los tres tests diagnósticos. Para simplificar, en los experimentos de simulación se ha considerado que $\delta_{ij} = \delta$ y $\varepsilon_{ij} = \varepsilon$, por lo que los factores de dependencia verifican las restricciones

$$\delta \leq (1 - Se_1)(1 - Se_2)Se_3, \delta \leq (1 - Se_1)Se_2(1 - Se_3), \delta \leq Se_1(1 - Se_2)(1 - Se_3)$$

$$\varepsilon \leq (1 - Sp_1)(1 - Sp_2)Sp_3, \varepsilon \leq (1 - Sp_1)Sp_2(1 - Sp_3), \varepsilon \leq Sp_1(1 - Sp_2)(1 - Sp_3).$$

Cuando $\delta_{jk} = \varepsilon_{jk} = 0$, $j, k = 1, 2, 3$, los tres tests diagnósticos son condicionalmente independientes al estado de enfermedad. En general, y tal y como ocurre en la mayoría de las situaciones prácticas, los parámetros δ_{jk} y/o ε_{jk} son mayores que cero, por lo que los tests diagnósticos son condicionalmente dependientes al estado de enfermedad.

Los experimentos de simulación se han diseñado a partir de las ecuaciones

$$\kappa(0) = \frac{Sp - (1 - Q)}{Q}, \quad (2.52)$$

$$\kappa(1) = \frac{Se - Q}{1 - Q} \quad (2.53)$$

y

$$\kappa(c) = \frac{p(1 - Q)c\kappa_1 + (1 - p)Q(1 - c)\kappa_0}{p(1 - Q)c + (1 - p)Q(1 - c)}, \quad (2.54)$$

siendo $Q = pSe + q(1 - Sp)$, de la siguiente forma. Como prevalencia de la enfermedad se han tomado los valores 10%, 30% y 50%; como índice de ponderación c se han tomado los valores 0.1, 0.5 y 0.9; y para cada uno de los tres tests diagnósticos se han tomado como $\kappa(0)$ y $\kappa(1)$ los valores $\{0.01, 0.02, \dots, 0.98, 0.99\}$. Una vez fijados los valores de p y c , y los valores de $\kappa(0)$ y $\kappa(1)$ para cada test diagnóstico, la sensibilidad y especificidad de cada test diagnóstico se han calculado resolviendo el sistema de ecuaciones formado por las ecuaciones (2.52) y (2.53); finalmente el valor del coeficiente kappa ponderado de cada test diagnóstico se ha calculado a partir de la ecuación (2.54). Por tanto, las probabilidades de las distribuciones multinomiales se han calculado a partir de valores de coeficientes kappa ponderados, y no se han fijado a priori los valores de las sensibilidades y de las especificidades. Asimismo, los experimentos de simulación se han diseñado de tal forma que en todas las muestras

aleatorias generadas se pueden estimar los coeficientes kappa ponderados y sus varianzas–covarianzas. Como error nominal se ha tomado $\alpha = 5\%$.

En los experimentos de simulación se han estudiado y comparado los errores tipo I y las potencias de tres métodos distintos para resolver el test de hipótesis $H_0 : \kappa_1(c) = \kappa_2(c) = \kappa_3(c)$ frente a H_1 : al menos una igualdad no es cierta. Los tres métodos son: (a) test de hipótesis conjunto basado en la distribución chi-cuadrado (ecuación (2.51)) tomando como error nominal $\alpha = 5\%$; (b) comparación de los tres coeficientes kappa ponderados dos a dos aplicando el método de Bloch (1997) tomando como error nominal $\alpha = 5\%$; (c) comparación de los tres coeficientes kappa ponderados aplicando el método de Bonferroni, es decir, comparación de los tres coeficientes kappa ponderados dos a dos aplicando el método de Bloch (1997) tomando como error nominal $2\alpha/(J(J-1)) = 0.05/3 = 1.67\%$. Asimismo, cada uno de estos tres métodos se ha resuelto de dos formas: (a) estimando la matriz de varianzas–covarianzas de los estimadores de los coeficientes kappa ponderados aplicando el método delta (ecuación (2.50)), y (b) estimando la matriz de varianzas–covarianzas mediante el método bootstrap (generando 1000 muestras bootstrap para cada una de las 10000 muestras aleatorias).

2.1.5.2.1. Error tipo I

En las Tablas 2.4, 2.5, 2.6 y 2.7 se muestran los resultados obtenidos cuando los tres coeficientes kappa ponderados son iguales a 0.2, 0.4, 0.6 y 0.8 respectivamente, para diferentes valores de la prevalencia de la enfermedad (p), del índice de ponderación (c), y para valores intermedios y altos de los factores de dependencia (δ_{ij} y ε_{ij}). En cada caso, para cada uno de los tres tests diagnósticos se indican los valores de $\kappa(0)$ y $\kappa(1)$ (y por tanto los valores de las correspondientes sensibilidades y especificidades) con los que se han generado las muestras aleatorias, mostrándose en las tablas los resultados obtenidos cuando $\kappa(0)$ y $\kappa(1)$ toman valores bajos, intermedios y altos. En estas tablas, el método delta (bootstrap) hace referencia a que la matriz de varianzas–covarianzas de los estimadores de los coeficientes kappa ponderados se ha estimado

mediante el método delta (bootstrap). En términos generales, a partir de los resultados de los experimentos de simulación se obtienen las siguientes conclusiones. El error tipo I del test de hipótesis conjunto estimando la matriz de varianzas–covarianzas mediante el método delta fluctúa en torno al error nominal del 5%, sobretodo a partir de muestras de 500 pacientes. En algunas ocasiones, sobretodo cuando la prevalencia es pequeña ($p = 10\%$) y las muestras son pequeñas ($n = 100$), el error tipo I puede desbordar al error nominal, debido a que el tamaño muestral no es lo suficientemente grande para poder comparar los tres coeficientes kappa ponderados. En cuanto al método basado en las comparaciones por parejas al error $\alpha = 5\%$, su error tipo I siempre desborda al error nominal, por lo que este método da lugar a resultados erróneos. Con respecto al método de Bonferroni, su error tipo I es casi siempre menor que el error nominal, sobretodo para muestras de 500 o más pacientes, por lo que es un método conservador que tiene el clásico comportamiento de un test exacto. Conclusiones similares se obtienen cuando el test de hipótesis conjunto se resuelve estimando la matriz de varianzas–covarianzas mediante bootstrap.

Tabla 2.4. Error tipo I de los tests de hipótesis cuando $\kappa_1 = \kappa_2 = \kappa_3 = 0.2$.

$p = 10\% \quad c = 0.9$						
Test 1: $\kappa(0) = 0.03, \kappa(1) = 0.54, Se_1 = 0.8467, Sp_1 = 0.3533, Q_1 = 0.6667$						
Test 2: $\kappa(0) = 0.20, \kappa(1) = 0.20, Se_2 = 0.28, Sp_2 = 0.92, Q_2 = 0.10$						
Test 3: $\kappa(0) = 0.38, \kappa(1) = 0.19, Se_3 = 0.2326, Sp_3 = 0.9674, Q_3 = 0.0526$						
$\delta_{ij} = 0.004 \quad \varepsilon_{ij} = 0.0004$						
n	Método Delta			Método Bootstrap		
	Test global	(1)	(2)	Test global	(1)	(2)
100	0.076	0.136	0.063	0.047	0.084	0.038
500	0.059	0.135	0.049	0.063	0.134	0.049
1000	0.060	0.126	0.048	0.058	0.128	0.047
1500	0.054	0.136	0.046	0.051	0.130	0.046
$\delta_{ij} = 0.008 \quad \varepsilon_{ij} = 0.0008$						
n	Método Delta			Método Bootstrap		
	Test global	(1)	(2)	Test global	(1)	(2)
100	0.067	0.133	0.047	0.032	0.073	0.032
500	0.066	0.149	0.048	0.065	0.141	0.048
1000	0.055	0.124	0.047	0.057	0.124	0.047
1500	0.052	0.126	0.043	0.051	0.131	0.045
$p = 30\% \quad c = 0.5$						
Test 1: $\kappa(0) = 0.12, \kappa(1) = 0.60, Se_1 = 0.8727, Sp_1 = 0.40, Q_1 = 0.6818$						
Test 2: $\kappa(0) = 0.30, \kappa(1) = 0.15, Se_2 = 0.44, Sp_2 = 0.76, Q_2 = 0.30$						
Test 3: $\kappa(0) = 0.60, \kappa(1) = 0.12, Se_3 = 0.1895, Sp_3 = 0.9684, Q_3 = 0.0789$						
$\delta_{ij} = 0.005 \quad \varepsilon_{ij} = 0.0015$						
n	Método Delta			Método Bootstrap		
	Test global	(1)	(2)	Test global	(1)	(2)
100	0.049	0.109	0.043	0.045	0.093	0.034
500	0.048	0.145	0.039	0.053	0.145	0.041
1000	0.052	0.145	0.047	0.052	0.143	0.048
1500	0.054	0.119	0.043	0.052	0.123	0.045
$\delta_{ij} = 0.010 \quad \varepsilon_{ij} = 0.0030$						
n	Método Delta			Método Bootstrap		
	Test global	(1)	(2)	Test global	(1)	(2)
100	0.053	0.126	0.050	0.050	0.112	0.040
500	0.063	0.123	0.049	0.064	0.126	0.048
1000	0.059	0.130	0.046	0.060	0.134	0.047
1500	0.063	0.143	0.048	0.063	0.145	0.046
$p = 50\% \quad c = 0.1$						
Test 1: $\kappa(0) = 0.03, \kappa(1) = 0.54, Se_1 = 0.7933, Sp_1 = 0.46, Q_1 = 0.6667$						
Test 2: $\kappa(0) = 0.08, \kappa(1) = 0.24, Se_2 = 0.2309, Sp_2 = 0.8673, Q_2 = 0.1818$						
Test 3: $\kappa(0) = 0.38, \kappa(1) = 0.19, Se_3 = 0.0811, Sp_3 = 0.9758, Q_3 = 0.0526$						
$\delta_{ij} = 0.0015 \quad \varepsilon_{ij} = 0.0005$						
n	Método Delta			Método Bootstrap		
	Test global	(1)	(2)	Test global	(1)	(2)
100	0.055	0.105	0.047	0.044	0.093	0.041
500	0.057	0.119	0.047	0.057	0.124	0.048
1000	0.053	0.127	0.047	0.055	0.128	0.047
1500	0.061	0.133	0.048	0.062	0.132	0.046
$\delta_{ij} = 0.0030 \quad \varepsilon_{ij} = 0.0010$						
n	Método Delta			Método Bootstrap		
	Test global	(1)	(2)	Test global	(1)	(2)
100	0.055	0.107	0.050	0.050	0.098	0.043
500	0.044	0.104	0.035	0.052	0.107	0.042
1000	0.058	0.125	0.047	0.063	0.127	0.046
1500	0.057	0.120	0.045	0.056	0.119	0.045

(1) Comparaciones por parejas con $\alpha = 5\%$. (2) Método de Bonferroni ($\alpha = 1.67\%$).

Tabla 2.5. Error tipo I de los tests de hipótesis cuando $\kappa_1 = \kappa_2 = \kappa_3 = 0.4$.

$p = 10\% \quad c = 0.1$						
Test 1: $\kappa(0) = 0.38, \kappa(1) = 0.76, Se_1 = 0.8036, Sp_1 = 0.8873, Q_1 = 0.1818$						
Test 2: $\kappa(0) = 0.52, \kappa(1) = 0.13, Se_2 = 0.1535, Sp_2 = 0.9870, Q_2 = 0.0270$						
Test 3: $\kappa(0) = 0.84, \kappa(1) = 0.07, Se_3 = 0.0785, Sp_3 = 0.9985, Q_3 = 0.0092$						
$\delta_{ij} = 0.001 \quad \varepsilon_{ij} = 0.000005$						
n	Método Delta			Método Bootstrap		
	Test global	(1)	(2)	Test global	(1)	(2)
100	0.112	0.168	0.079	0.075	0.132	0.061
500	0.063	0.162	0.051	0.063	0.156	0.049
1000	0.057	0.148	0.047	0.058	0.150	0.043
1500	0.055	0.144	0.045	0.052	0.138	0.042
$\delta_{ij} = 0.002 \quad \varepsilon_{ij} = 0.000010$						
n	Método Delta			Método Bootstrap		
	Test global	(1)	(2)	Test global	(1)	(2)
100	0.116	0.198	0.095	0.095	0.156	0.080
500	0.063	0.164	0.052	0.062	0.161	0.051
1000	0.056	0.161	0.047	0.058	0.163	0.047
1500	0.054	0.150	0.048	0.055	0.149	0.044
$p = 30\% \quad c = 0.9$						
Test 1: $\kappa(0) = 0.07, \kappa(1) = 0.84, Se_1 = 0.9740, Sp_1 = 0.2214, Q_1 = 0.8372$						
Test 2: $\kappa(0) = 0.40, \kappa(1) = 0.40, Se_2 = 0.58, Sp_2 = 0.82, Q_2 = 0.30$						
Test 3: $\kappa(0) = 0.76, \kappa(1) = 0.38, Se_3 = 0.4894, Sp_3 = 0.9576, Q_3 = 0.1765$						
$\delta_{ij} = 0.0025 \quad \varepsilon_{ij} = 0.0006$						
n	Método Delta			Método Bootstrap		
	Test global	(1)	(2)	Test global	(1)	(2)
100	0.067	0.142	0.059	0.073	0.147	0.061
500	0.046	0.116	0.046	0.052	0.113	0.044
1000	0.051	0.128	0.046	0.053	0.128	0.046
1500	0.064	0.117	0.044	0.061	0.117	0.044
$\delta_{ij} = 0.0050 \quad \varepsilon_{ij} = 0.0012$						
n	Método Delta			Método Bootstrap		
	Test global	(1)	(2)	Test global	(1)	(2)
100	0.064	0.167	0.061	0.072	0.169	0.060
500	0.047	0.117	0.043	0.048	0.118	0.043
1000	0.050	0.122	0.048	0.051	0.125	0.045
1500	0.047	0.110	0.043	0.046	0.108	0.042
$p = 50\% \quad c = 0.5$						
Test 1: $\kappa(0) = 0.28, \kappa(1) = 0.70, Se_1 = 0.9143, Sp_1 = 0.4857, Q_1 = 0.7143$						
Test 2: $\kappa(0) = 0.45, \kappa(1) = 0.36, Se_2 = 0.6444, Sp_2 = 0.7556, Q_2 = 0.4444$						
Test 3: $\kappa(0) = 0.60, \kappa(1) = 0.30, Se_3 = 0.5333, Sp_3 = 0.8667, Q_3 = 0.3333$						
$\delta_{ij} = 0.007 \quad \varepsilon_{ij} = 0.006$						
n	Método Delta			Método Bootstrap		
	Test global	(1)	(2)	Test global	(1)	(2)
100	0.061	0.138	0.052	0.060	0.134	0.053
500	0.057	0.128	0.047	0.061	0.136	0.048
1000	0.056	0.134	0.045	0.055	0.141	0.047
1500	0.047	0.116	0.041	0.050	0.116	0.042
$\delta_{ij} = 0.014 \quad \varepsilon_{ij} = 0.012$						
n	Método Delta			Método Bootstrap		
	Test global	(1)	(2)	Test global	(1)	(2)
100	0.066	0.139	0.055	0.069	0.140	0.056
500	0.042	0.119	0.040	0.042	0.122	0.037
1000	0.044	0.110	0.033	0.042	0.108	0.036
1500	0.044	0.119	0.039	0.042	0.124	0.043

(1) Comparaciones por parejas con $\alpha = 5\%$. (2) Método de Bonferroni ($\alpha = 1.67\%$).

Tabla 2.6. Error tipo I de los tests de hipótesis cuando $\kappa_1 = \kappa_2 = \kappa_3 = 0.6$.

$p = 10\% \quad c = 0.5$						
Test 1: $\kappa(0) = 0.45, \kappa(1) = 0.90, Se_1 = 0.9182, Sp_1 = 0.90, Q_1 = 0.1818$						
Test 2: $\kappa(0) = 0.60, \kappa(1) = 0.60, Se_2 = 0.64, Sp_2 = 0.96, Q_2 = 0.10$						
Test 3: $\kappa(0) = 0.90, \kappa(1) = 0.45, Se_3 = 0.4789, Sp_3 = 0.9947, Q_3 = 0.0526$						
$\delta_{ij} = 0.006 \quad \varepsilon_{ij} = 0.0008$	Método Delta			Método Bootstrap		
n	Test global	(1)	(2)	Test global	(1)	(2)
100	0.104	0.174	0.071	0.076	0.140	0.065
500	0.059	0.141	0.049	0.064	0.141	0.048
1000	0.058	0.133	0.047	0.060	0.131	0.047
1500	0.057	0.117	0.046	0.057	0.117	0.046
$\delta_{ij} = 0.012 \quad \varepsilon_{ij} = 0.0016$	Método Delta			Método Bootstrap		
n	Test global	(1)	(2)	Test global	(1)	(2)
100	0.114	0.175	0.092	0.068	0.133	0.055
500	0.051	0.135	0.047	0.050	0.132	0.048
1000	0.059	0.132	0.048	0.059	0.129	0.047
1500	0.052	0.119	0.047	0.053	0.117	0.046
$p = 30\% \quad c = 0.9$						
Test 1: $\kappa(0) = 0.15, \kappa(1) = 0.90, Se_1 = 0.972, Sp_1 = 0.388, Q_1 = 0.72$						
Test 2: $\kappa(0) = 0.33, \kappa(1) = 0.66, Se_2 = 0.8169, Sp_2 = 0.6908, Q_2 = 0.4615$						
Test 3: $\kappa(0) = 0.87, \kappa(1) = 0.58, Se_3 = 0.6733, Sp_3 = 0.9711, Q_3 = 0.2222$						
$\delta_{ij} = 0.0015 \quad \varepsilon_{ij} = 0.0015$	Método Delta			Método Bootstrap		
n	Test global	(1)	(2)	Test global	(1)	(2)
100	0.068	0.146	0.058	0.072	0.146	0.064
500	0.061	0.145	0.048	0.062	0.139	0.048
1000	0.042	0.099	0.034	0.041	0.103	0.035
1500	0.055	0.135	0.041	0.050	0.139	0.039
$\delta_{ij} = 0.0030 \quad \varepsilon_{ij} = 0.0030$	Método Delta			Método Bootstrap		
n	Test global	(1)	(2)	Test global	(1)	(2)
100	0.068	0.137	0.055	0.068	0.137	0.054
500	0.056	0.125	0.046	0.057	0.122	0.047
1000	0.044	0.135	0.039	0.044	0.135	0.039
1500	0.047	0.118	0.042	0.046	0.120	0.041
$p = 50\% \quad c = 0.1$						
Test 1: $\kappa(0) = 0.58, \kappa(1) = 0.87, Se_1 = 0.948, Sp_1 = 0.748, Q_1 = 0.60$						
Test 2: $\kappa(0) = 0.63, \kappa(1) = 0.42, Se_2 = 0.652, Sp_2 = 0.852, Q_2 = 0.40$						
Test 3: $\kappa(0) = 0.90, \kappa(1) = 0.15, Se_3 = 0.2714, Sp_3 = 0.9857, Q_3 = 0.1429$						
$\delta_{ij} = 0.002 \quad \varepsilon_{ij} = 0.0006$	Método Delta			Método Bootstrap		
n	Test global	(1)	(2)	Test global	(1)	(2)
100	0.062	0.137	0.050	0.064	0.136	0.053
500	0.053	0.129	0.046	0.050	0.123	0.039
1000	0.059	0.139	0.047	0.061	0.139	0.041
1500	0.058	0.129	0.045	0.052	0.123	0.046
$\delta_{ij} = 0.004 \quad \varepsilon_{ij} = 0.0012$	Método Delta			Método Bootstrap		
n	Test global	(1)	(2)	Test global	(1)	(2)
100	0.066	0.134	0.053	0.062	0.131	0.049
500	0.053	0.128	0.046	0.055	0.126	0.048
1000	0.053	0.131	0.048	0.054	0.131	0.047
1500	0.045	0.111	0.039	0.044	0.116	0.042

(1) Comparaciones por parejas con $\alpha = 5\%$. (2) Método de Bonferroni ($\alpha = 1.67\%$).

Tabla 2.7. Error tipo I de los tests de hipótesis cuando $\kappa_1 = \kappa_2 = \kappa_3 = 0.8$.

$p = 10\% \quad c = 0.5$						
Test 1: $\kappa(0) = 0.72, \kappa(1) = 0.90, Se_1 = 0.9122, Sp_1 = 0.9659, Q_1 = 0.1220$						
Test 2: $\kappa(0) = 0.80, \kappa(1) = 0.80, Se_2 = 0.82, Sp_2 = 0.98, Q_2 = 0.10$						
Test 3: $\kappa(0) = 0.90, \kappa(1) = 0.72, Se_3 = 0.7429, Sp_3 = 0.9918, Q_3 = 0.0816$						
$\delta_{ij} = 0.005 \quad \varepsilon_{ij} = 0.00006$	Método Delta			Método Bootstrap		
n	Test global	(1)	(2)	Test global	(1)	(2)
100	0.089	0.144	0.047	0.055	0.100	0.028
500	0.056	0.119	0.049	0.058	0.117	0.049
1000	0.055	0.124	0.043	0.060	0.130	0.048
1500	0.054	0.127	0.047	0.050	0.127	0.045
$\delta_{ij} = 0.010 \quad \varepsilon_{ij} = 0.00012$	Método Delta			Método Bootstrap		
n	Test global	(1)	(2)	Test global	(1)	(2)
100	0.102	0.159	0.051	0.077	0.129	0.038
500	0.052	0.118	0.042	0.049	0.111	0.040
1000	0.058	0.124	0.047	0.060	0.124	0.046
1500	0.057	0.132	0.048	0.060	0.131	0.047
$p = 30\% \quad c = 0.1$						
Test 1: $\kappa(0) = 0.80, \kappa(1) = 0.80, Se_1 = 0.86, Sp_1 = 0.94, Q_1 = 0.30$						
Test 2: $\kappa(0) = 0.88, \kappa(1) = 0.44, Se_2 = 0.5388, Sp_2 = 0.9788, Q_2 = 0.1765$						
Test 3: $\kappa(0) = 0.96, \kappa(1) = 0.32, Se_3 = 0.405, Sp_3 = 0.995, Q_3 = 0.1250$						
$\delta_{ij} = 0.01 \quad \varepsilon_{ij} = 0.00004$	Método Delta			Método Bootstrap		
n	Test global	(1)	(2)	Test global	(1)	(2)
100	0.078	0.123	0.051	0.073	0.119	0.046
500	0.063	0.123	0.047	0.063	0.125	0.046
1000	0.046	0.114	0.039	0.047	0.114	0.041
1500	0.049	0.122	0.037	0.047	0.124	0.039
$\delta_{ij} = 0.02 \quad \varepsilon_{ij} = 0.00008$	Método Delta			Método Bootstrap		
n	Test global	(1)	(2)	Test global	(1)	(2)
100	0.073	0.120	0.042	0.066	0.118	0.032
500	0.055	0.133	0.048	0.054	0.124	0.046
1000	0.062	0.140	0.047	0.061	0.142	0.048
1500	0.047	0.111	0.041	0.047	0.106	0.045
$p = 50\% \quad c = 0.9$						
Test 1: $\kappa(0) = 0.32, \kappa(1) = 0.96, Se_1 = 0.99, Sp_1 = 0.49, Q_1 = 0.75$						
Test 2: $\kappa(0) = 0.56, \kappa(1) = 0.84, Se_2 = 0.936, Sp_2 = 0.736, Q_2 = 0.60$						
Test 3: $\kappa(0) = 0.80, \kappa(1) = 0.80, Se_3 = 0.90, Sp_3 = 0.90, Q_3 = 0.50$						
$\delta_{ij} = 0.00025 \quad \varepsilon_{ij} = 0.005$	Método Delta			Método Bootstrap		
n	Test global	(1)	(2)	Test global	(1)	(2)
100	0.080	0.138	0.058	0.079	0.144	0.062
500	0.066	0.145	0.050	0.076	0.146	0.049
1000	0.059	0.129	0.047	0.063	0.132	0.046
1500	0.055	0.130	0.048	0.048	0.132	0.042
$\delta_{ij} = 0.00050 \quad \varepsilon_{ij} = 0.010$	Método Delta			Método Bootstrap		
n	Test global	(1)	(2)	Test global	(1)	(2)
100	0.071	0.134	0.052	0.070	0.143	0.050
500	0.045	0.118	0.040	0.048	0.122	0.037
1000	0.061	0.147	0.048	0.062	0.147	0.048
1500	0.044	0.138	0.042	0.046	0.138	0.042

(1) Comparaciones por parejas con $\alpha = 5\%$. (2) Método de Bonferroni ($\alpha = 1.67\%$).

2.1.5.2.2. Potencia

En las Tablas 2.8 y 2.9 se muestran los resultados obtenidos para diferentes valores de los tres coeficientes kappa ponderados, para diferentes valores de la prevalencia de la enfermedad (p), del índice de ponderación (c), y para valores intermedios y altos de los factores de dependencia (δ_{ij} y ε_{ij}). Al igual que para el error tipo I, en cada caso se indican los valores de $\kappa(0)$ y $\kappa(1)$ (y por tanto los valores de las correspondientes sensibilidades y especificidades) para cada test diagnóstico con los que se han generado las muestras aleatorias, mostrándose en las tablas los resultados obtenidos cuando $\kappa(0)$ y $\kappa(1)$ toman valores intermedios. En estas tablas, el método delta (bootstrap) hace referencia a la misma cuestión que en el error tipo I. A partir de los resultados se obtienen las siguientes conclusiones. La potencia del test de hipótesis conjunto estimando la matriz de varianzas–covarianzas mediante el método delta y del test de hipótesis conjunto estimando la matriz de varianzas–covarianzas mediante bootstrap son muy similares, y en términos generales se necesitan muestras de entre 500 y 1000 pacientes para que la potencia sea elevada (superior al 90%). El método basado en las comparaciones por parejas al error $\alpha = 5\%$ (en sus dos versiones, método delta y bootstrap) tiene una potencia más elevada que el test conjunto (en sus dos versiones, método delta y bootstrap) debido a que su error tipo I es mayor. En cuanto al método de Bonferroni, su potencia es menor debido a que su error tipo I es también menor.

Tabla 2.8. Potencia de los tests de hipótesis cuando $\kappa_1 = 0.4$, $\kappa_2 = 0.4$ y $\kappa_3 = 0.6$.

$p = 10\% \quad c = 0.9$						
Test 1: $\kappa(0) = 0.76$, $\kappa(1) = 0.38$, $Se_1 = 0.4126$, $Sp_1 = 0.9874$, $Q_1 = 0.0526$						
Test 2: $\kappa(0) = 0.07$, $\kappa(1) = 0.84$, $Se_1 = 0.9314$, $Sp_1 = 0.4686$, $Q_2 = 0.5714$						
Test 3: $\kappa(0) = 0.42$, $\kappa(1) = 0.63$, $Se_1 = 0.6829$, $Sp_1 = 0.9171$, $Q_3 = 0.1429$						
$\delta_{ij} = 0.004 \quad \varepsilon_{ij} = 0.0002$						
n	Método Delta			Método Bootstrap		
	Test global	(1)	(2)	Test global	(1)	(2)
100	0.295	0.381	0.262	0.208	0.292	0.194
500	0.666	0.798	0.640	0.665	0.800	0.640
1000	0.929	0.970	0.918	0.929	0.968	0.920
1500	0.977	0.990	0.975	0.977	0.993	0.976
$\delta_{ij} = 0.008 \quad \varepsilon_{ij} = 0.0004$						
n	Método Delta			Método Bootstrap		
	Test global	(1)	(2)	Test global	(1)	(2)
100	0.277	0.376	0.242	0.198	0.279	0.182
500	0.669	0.813	0.652	0.670	0.812	0.648
1000	0.930	0.967	0.916	0.926	0.965	0.917
1500	0.986	0.994	0.984	0.985	0.994	0.984
$p = 30\% \quad c = 0.5$						
Test 1: $\kappa(0) = 0.28$, $\kappa(1) = 0.70$, $Se_1 = 0.8552$, $Sp_1 = 0.6276$, $Q_1 = 0.5172$						
Test 2: $\kappa(0) = 0.70$, $\kappa(1) = 0.28$, $Se_1 = 0.3854$, $Sp_1 = 0.9561$, $Q_2 = 0.1463$						
Test 3: $\kappa(0) = 0.60$, $\kappa(1) = 0.60$, $Se_1 = 0.72$, $Sp_1 = 0.88$, $Q_3 = 0.30$						
$\delta_{ij} = 0.007 \quad \varepsilon_{ij} = 0.0015$						
n	Método Delta			Método Bootstrap		
	Test global	(1)	(2)	Test global	(1)	(2)
100	0.433	0.586	0.403	0.428	0.575	0.400
500	0.977	0.990	0.975	0.976	0.990	0.973
1000	0.999	1	0.999	0.999	1	0.999
1500	1	1	1	1	1	1
$\delta_{ij} = 0.014 \quad \varepsilon_{ij} = 0.0030$						
n	Método Delta			Método Bootstrap		
	Test global	(1)	(2)	Test global	(1)	(2)
100	0.459	0.578	0.421	0.457	0.577	0.421
500	0.976	0.987	0.968	0.976	0.987	0.968
1000	1	1	1	1	1	1
1500	1	1	1	1	1	1
$p = 50\% \quad c = 0.1$						
Test 1: $\kappa(0) = 0.39$, $\kappa(1) = 0.52$, $Se_1 = 0.7943$, $Sp_1 = 0.6514$, $Q_1 = 0.5714$						
Test 2: $\kappa(0) = 0.84$, $\kappa(1) = 0.07$, $Se_1 = 0.1415$, $Sp_1 = 0.9877$, $Q_2 = 0.0769$						
Test 3: $\kappa(0) = 0.58$, $\kappa(1) = 0.87$, $Se_1 = 0.948$, $Sp_1 = 0.748$, $Q_3 = 0.60$						
$\delta_{ij} = 0.0006 \quad \varepsilon_{ij} = 0.0009$						
n	Método Delta			Método Bootstrap		
	Test global	(1)	(2)	Test global	(1)	(2)
100	0.379	0.545	0.332	0.415	0.569	0.360
500	0.973	0.991	0.962	0.972	0.992	0.965
1000	1	1	0.999	0.999	1	0.999
1500	1	1	1	1	1	1
$\delta_{ij} = 0.0012 \quad \varepsilon_{ij} = 0.0018$						
n	Método Delta			Método Bootstrap		
	Test global	(1)	(2)	Test global	(1)	(2)
100	0.391	0.536	0.362	0.417	0.553	0.374
500	0.966	0.987	0.953	0.970	0.988	0.954
1000	1	1	1	1	1	1
1500	1	1	1	1	1	1

(1) Comparaciones por parejas con $\alpha = 5\%$. (2) Método de Bonferroni ($\alpha = 1.67\%$).

Tabla 2.9. Potencia de los tests de hipótesis cuando $\kappa_1 = 0.6$, $\kappa_2 = 0.6$ y $\kappa_3 = 0.8$.

$p = 10\%$ $c = 0.5$						
Test 1: $\kappa(0) = 0.45$, $\kappa(1) = 0.90$, $Se_1 = 0.9182$, $Sp_1 = 0.90$, $Q_1 = 0.1818$						
Test 2: $\kappa(0) = 0.90$, $\kappa(1) = 0.45$, $Se_2 = 0.4789$, $Sp_2 = 0.9947$, $Q_2 = 0.0526$						
Test 3: $\kappa(0) = 0.80$, $\kappa(1) = 0.80$, $Se_3 = 0.82$, $Sp_3 = 0.98$, $Q_3 = 0.10$						
$\delta_{ij} = 0.003$ $\varepsilon_{ij} = 0.00004$	Método Delta			Método Bootstrap		
n	Test global	(1)	(2)	Test global	(1)	(2)
100	0.340	0.462	0.268	0.260	0.362	0.212
500	0.870	0.930	0.844	0.865	0.931	0.837
1000	0.992	0.996	0.986	0.992	0.996	0.988
1500	1	1	1	1	1	1
$\delta_{ij} = 0.006$ $\varepsilon_{ij} = 0.00008$	Método Delta			Método Bootstrap		
n	Test global	(1)	(2)	Test global	(1)	(2)
100	0.359	0.474	0.300	0.283	0.373	0.233
500	0.872	0.929	0.855	0.871	0.925	0.849
1000	0.992	0.998	0.986	0.993	0.998	0.986
1500	1	1	1	1	1	0.999
$p = 30\%$ $c = 0.1$						
Test 1: $\kappa(0) = 0.58$, $\kappa(1) = 0.87$, $Se_1 = 0.9209$, $Sp_1 = 0.8357$, $Q_1 = 0.3913$						
Test 2: $\kappa(0) = 0.90$, $\kappa(1) = 0.15$, $Se_2 = 0.2067$, $Sp_2 = 0.9933$, $Q_2 = 0.0667$						
Test 3: $\kappa(0) = 0.88$, $\kappa(1) = 0.44$, $Se_3 = 0.5388$, $Sp_3 = 0.9788$, $Q_3 = 0.1765$						
$\delta_{ij} = 0.003$ $\varepsilon_{ij} = 0.00005$	Método Delta			Método Bootstrap		
n	Test global	(1)	(2)	Test global	(1)	(2)
100	0.425	0.541	0.370	0.419	0.530	0.354
500	0.944	0.973	0.932	0.946	0.972	0.935
1000	0.999	1	0.998	1	1	0.998
1500	1	1	1	1	1	1
$\delta_{ij} = 0.006$ $\varepsilon_{ij} = 0.0001$	Método Delta			Método Bootstrap		
n	Test global	(1)	(2)	Test global	(1)	(2)
100	0.427	0.552	0.369	0.424	0.544	0.376
500	0.952	0.979	0.935	0.956	0.979	0.932
1000	0.999	1	0.998	0.999	1	0.998
1500	1	1	1	1	1	1
$p = 50\%$ $c = 0.9$						
Test 1: $\kappa(0) = 0.15$, $\kappa(1) = 0.90$, $Se_1 = 0.9857$, $Sp_1 = 0.2714$, $Q_1 = 0.8571$						
Test 2: $\kappa(0) = 0.87$, $\kappa(1) = 0.58$, $Se_2 = 0.748$, $Sp_2 = 0.948$, $Q_2 = 0.40$						
Test 3: $\kappa(0) = 0.56$, $\kappa(1) = 0.84$, $Se_3 = 0.936$, $Sp_3 = 0.736$, $Q_3 = 0.60$						
$\delta_{ij} = 0.0003$ $\varepsilon_{ij} = 0.0015$	Método Delta			Método Bootstrap		
n	Test global	(1)	(2)	Test global	(1)	(2)
100	0.476	0.615	0.422	0.487	0.620	0.427
500	0.986	0.994	0.982	0.985	0.994	0.983
1000	1	1	1	1	1	1
1500	1	1	1	1	1	1
$\delta_{ij} = 0.0006$ $\varepsilon_{ij} = 0.003$	Método Delta			Método Bootstrap		
n	Test global	(1)	(2)	Test global	(1)	(2)
100	0.481	0.620	0.430	0.481	0.625	0.437
500	0.988	0.992	0.984	0.988	0.993	0.986
1000	1	1	1	1	1	1
1500	1	1	1	1	1	1

(1) Comparaciones por parejas con $\alpha = 5\%$. (2) Método de Bonferroni ($\alpha = 1.67\%$).

Finalmente, a partir de los resultados de los experimentos de simulación, se puede concluir que el test de hipótesis conjunto basado en la distribución chi-cuadrado tiene el rendimiento clásico de un test de hipótesis asintótico, siendo necesarias, en términos generales, muestras de entre 500 y 1000 pacientes para que la potencia sea elevada.

2.1.5.3. Aplicación

Los resultados obtenidos se han aplicado al estudio de Weiner et al. (1979) sobre el diagnóstico de la enfermedad de la arteria coronaria. Weiner et al. investigaron el diagnóstico de la enfermedad de la arteria coronaria utilizando como tests diagnósticos el test de ejercicio, la historia clínica del paciente y el electrocardiograma, y como gold estándar utilizaron la arteriografía coronaria. En la Tabla 2.10 se muestran los resultados obtenidos al aplicar los tests diagnósticos y el gold estándar a una muestra de 1465 hombres, donde T_1 modeliza el resultado del test de ejercicio, T_2 el resultado de la historia clínica, T_3 el resultado del electrocardiograma y D el resultado del gold estándar.

Tabla 2.10. Datos del estudio de Weiner et al. (1979).

	$T_1 = 1$				$T_1 = 0$				Total
	$T_2 = 1$		$T_2 = 0$		$T_2 = 1$		$T_2 = 0$		
	$T_3 = 1$	$T_3 = 0$	$T_3 = 1$	$T_3 = 0$	$T_3 = 1$	$T_3 = 0$	$T_3 = 1$	$T_3 = 0$	
$D = 1$	215	571	9	20	31	152	1	24	1023
$D = 0$	22	47	13	33	16	160	25	126	442
Total	237	618	22	53	47	312	26	150	1465

En la Tabla 2.11 se muestran los resultados obtenidos al comparar los coeficientes kappa ponderados de los tres tests diagnósticos para diferentes valores del índice c , siendo el test de hipótesis conjunto significativo en todos los casos. Así por ejemplo, para $c = 0.5$ los valores de los estimadores máximo verosímiles de los coeficientes

kappa de Cohen son $\hat{\kappa}_1 = 0.51$, $\hat{\kappa}_2 = 0.45$ y $\hat{\kappa}_3 = 0.05$, y la matriz de varianzas-covarianzas estimadas de $\hat{\boldsymbol{\kappa}} = (\hat{\kappa}_1, \hat{\kappa}_2, \hat{\kappa}_3)^T$ es

$$\hat{\Sigma}_{\hat{\boldsymbol{\kappa}}} = \begin{pmatrix} 0.000558 & 0.000038 & 0.000069 \\ 0.000038 & 0.000663 & -0.000014 \\ 0.000069 & -0.000014 & 0.000247 \end{pmatrix}.$$

El valor del estadístico de contraste para el test de hipótesis $H_0 : \kappa_1 = \kappa_2 = \kappa_3$ es $Q_{\text{exp}}^2 = 371.76$ ($p\text{-valor} = 0$), y se rechaza la hipótesis de igualdad de todos los coeficientes kappa de Cohen. Para investigar las causas de la significación se realizan las comparaciones por parejas al error $\alpha = 1.67\%$: no se rechaza la hipótesis de igualdad de los coeficientes kappa de Cohen del test de ejercicio y de la historia clínica del paciente, y los coeficientes kappa de Cohen del test de ejercicio y de la historia clínica son significativamente mayores que el coeficiente kappa de Cohen del electrocardiograma.

Tabla 2.11. Comparación de los tres tests diagnósticos para diferentes valores de c .

c	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\kappa}_3$	$H_0 : \kappa_1 = \kappa_2 = \kappa_3$		$H_0 : \kappa_1 = \kappa_2$		$H_0 : \kappa_1 = \kappa_3$		$H_0 : \kappa_2 = \kappa_3$	
				Q_{exp}^2	$p\text{-valor}$	z_{exp}	$p\text{-valor}$	z_{exp}	$p\text{-valor}$	z_{exp}	$p\text{-valor}$
0.1	0.57	0.35	0.14	105.58	0	6.35	$< 10^{-8}$	9.58	0	4.29	$< 10^{-4}$
0.2	0.55	0.37	0.10	166.36	0	5.38	$< 10^{-6}$	12.88	0	6.92	$< 10^{-10}$
0.3	0.54	0.39	0.08	237.23	0	4.26	$< 10^{-8}$	15.20	0	9.15	0
0.4	0.52	0.42	0.06	307.88	0	3.04	0.0023	16.68	0	11.09	0
0.5	0.51	0.45	0.05	371.76	0	1.73	0.0836	17.49	0	12.89	0
0.6	0.49	0.48	0.05	426.17	0	0.31	0.7566	17.78	0	14.53	0
0.7	0.48	0.52	0.04	470.27	0	1.24	0.2150	17.69	0	16.12	0
0.8	0.47	0.57	0.04	501.38	0	2.92	0.0035	17.37	0	17.51	0
0.9	0.45	0.62	0.03	510.73	0	4.71	$< 10^{-4}$	16.91	0	18.39	0

En la Tabla 2.11 también se puede observar el efecto que el índice c tiene sobre la comparación de los tres coeficientes kappa ponderados. Cuando $c < 0.5$, y por tanto los tests diagnósticos se van a utilizar previamente a un tratamiento invasivo, el coeficiente kappa ponderado del test de ejercicio es significativamente mayor que el coeficiente

kappa ponderado de la historia clínica, que a su vez es significativamente mayor que el del electrocardiograma. Cuando $c > 0.5$, y por tanto los tests diagnósticos se van a utilizar como tests de screening, los coeficientes kappa ponderados del test de ejercicio y de la historia clínica son distintos dependiendo del valor del índice c . Para c igual a 0.6 y 0.7 no existen diferencias significativas entre los dos coeficientes kappa ponderados, y para c igual a 0.8 y 0.9 el coeficiente kappa ponderado de la historia clínica es significativamente mayor que el del test de ejercicio, siendo siempre y el coeficiente kappa ponderado del electrocardiograma significativamente menor que el del test de ejercicio y que el de la historia clínica del paciente.

Por otra parte, como la estimación de las matrices de varianzas-covarianzas de los estimadores de los coeficientes kappa ponderados es complicada y requiere el uso de un software matemático, se ha comparado este método de estimación con la estimación de estas matrices mediante bootstrap. Para ello se han realizado unos experimentos de simulación que han consistido en generar 10000 muestras aleatoria de distribuciones multinomiales de tamaños 100, 500, 1000 y 1465, y cuyas probabilidades son las frecuencias relativas del ejemplo de Weiner et al. (1979) (Tabla 2.11). Para cada una de las 10000 muestras aleatorias se han calculado los intervalos de confianza al 95% $\left(\hat{\kappa}_i - \hat{\kappa}_j \pm 1.96 \sqrt{\widehat{Var}(\hat{\kappa}_i) + \widehat{Var}(\hat{\kappa}_j) - 2\widehat{Cov}(\hat{\kappa}_i, \hat{\kappa}_j)} \right)$ para la diferencia de los coeficientes kappa ponderados $(\kappa_1 - \kappa_2, \kappa_1 - \kappa_3$ y $\kappa_2 - \kappa_3)$ para distintos valores del índice c . Para cada una de las 10000 muestras aleatorias se han generado 1000 muestras bootstrap y se han calculado los intervalos de confianza al 95% para la diferencia de los coeficientes kappa ponderados de la forma $\hat{\kappa}_{Bi} - \hat{\kappa}_{Bj} \pm 1.96 \sqrt{\widehat{Var}_B(\hat{\kappa}_{Bi} - \hat{\kappa}_{Bj})}$, donde $\hat{\kappa}_{Bi}$ es el estimador bootstrap del i -ésimo coeficiente kappa ponderado y $\widehat{Var}_B(\hat{\kappa}_{Bi} - \hat{\kappa}_{Bj})$ es el estimador bootstrap de la varianza de $\hat{\kappa}_{Bi} - \hat{\kappa}_{Bj}$. Finalmente se han calculado la cobertura y amplitud media de los intervalos de confianza tanto para los estimadores por máxima verosimilitud como para los estimadores por bootstrap, para lo cual se ha asumido que los verdaderos valores de los coeficientes kappa ponderados son los dados en la Tabla 2.11. En la Tabla 2.12 se muestran los resultados obtenidos, obteniéndose que en términos de cobertura y amplitud media, no existen prácticamente diferencias entre ambos tipos de intervalos de confianza.

Tabla 2.12. Cobertura y amplitud media de los intervalos de confianza para los diferentes coeficientes kappa ponderados.

$n = 100$												
c	IC máximo verosímil						IC Bootstrap					
	$\kappa_1 - \kappa_2$		$\kappa_1 - \kappa_3$		$\kappa_2 - \kappa_3$		$\kappa_1 - \kappa_2$		$\kappa_1 - \kappa_3$		$\kappa_2 - \kappa_3$	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
0.1	0.941	0.516	0.936	0.649	0.937	0.701	0.942	0.522	0.939	0.646	0.941	0.700
0.2	0.941	0.504	0.948	0.515	0.949	0.570	0.943	0.508	0.949	0.514	0.951	0.570
0.3	0.947	0.500	0.932	0.444	0.947	0.509	0.948	0.504	0.933	0.445	0.949	0.510
0.4	0.945	0.501	0.951	0.408	0.938	0.474	0.944	0.503	0.954	0.410	0.939	0.476
0.5	0.941	0.502	0.947	0.385	0.939	0.455	0.942	0.504	0.949	0.387	0.940	0.458
0.6	0.941	0.507	0.944	0.375	0.935	0.446	0.942	0.509	0.946	0.377	0.931	0.450
0.7	0.948	0.508	0.938	0.370	0.942	0.443	0.949	0.512	0.939	0.372	0.944	0.448
0.8	0.941	0.517	0.944	0.369	0.937	0.451	0.944	0.521	0.945	0.371	0.939	0.457
0.9	0.947	0.532	0.944	0.372	0.930	0.468	0.948	0.538	0.945	0.375	0.932	0.475
$n = 500$												
c	IC máximo verosímil						IC Bootstrap					
	$\kappa_1 - \kappa_2$		$\kappa_1 - \kappa_3$		$\kappa_2 - \kappa_3$		$\kappa_1 - \kappa_2$		$\kappa_1 - \kappa_3$		$\kappa_2 - \kappa_3$	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
0.1	0.947	0.233	0.946	0.299	0.940	0.321	0.943	0.234	0.944	0.298	0.939	0.320
0.2	0.942	0.229	0.943	0.235	0.954	0.259	0.944	0.229	0.941	0.234	0.955	0.259
0.3	0.952	0.227	0.956	0.202	0.954	0.230	0.952	0.227	0.956	0.202	0.954	0.230
0.4	0.950	0.226	0.944	0.184	0.951	0.214	0.952	0.226	0.942	0.184	0.951	0.214
0.5	0.955	0.227	0.957	0.173	0.948	0.205	0.955	0.227	0.959	0.174	0.948	0.206
0.6	0.947	0.228	0.951	0.168	0.942	0.200	0.946	0.228	0.950	0.168	0.942	0.201
0.7	0.945	0.229	0.948	0.166	0.948	0.200	0.945	0.230	0.948	0.166	0.948	0.200
0.8	0.955	0.233	0.946	0.166	0.947	0.203	0.956	0.233	0.945	0.166	0.947	0.204
0.9	0.951	0.242	0.947	0.167	0.943	0.215	0.950	0.242	0.946	0.167	0.944	0.216
$n = 1000$												
c	IC máximo verosímil						IC Bootstrap					
	$\kappa_1 - \kappa_2$		$\kappa_1 - \kappa_3$		$\kappa_2 - \kappa_3$		$\kappa_1 - \kappa_2$		$\kappa_1 - \kappa_3$		$\kappa_2 - \kappa_3$	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
0.1	0.953	0.166	0.942	0.211	0.943	0.228	0.956	0.166	0.942	0.211	0.944	0.228
0.2	0.945	0.162	0.950	0.166	0.951	0.184	0.945	0.162	0.952	0.166	0.953	0.184
0.3	0.954	0.160	0.954	0.143	0.953	0.163	0.954	0.160	0.956	0.143	0.952	0.163
0.4	0.949	0.160	0.938	0.130	0.943	0.151	0.950	0.160	0.937	0.130	0.945	0.152
0.5	0.946	0.160	0.942	0.123	0.951	0.145	0.944	0.161	0.941	0.123	0.950	0.145
0.6	0.952	0.161	0.949	0.119	0.941	0.142	0.952	0.161	0.949	0.119	0.940	0.142
0.7	0.954	0.162	0.958	0.117	0.953	0.141	0.954	0.162	0.958	0.117	0.953	0.141
0.8	0.949	0.165	0.951	0.117	0.947	0.144	0.950	0.165	0.949	0.117	0.946	0.144
0.9	0.948	0.170	0.937	0.118	0.942	0.152	0.948	0.170	0.936	0.118	0.942	0.152
$n = 1465$												
c	IC máximo verosímil						IC Bootstrap					
	$\kappa_1 - \kappa_2$		$\kappa_1 - \kappa_3$		$\kappa_2 - \kappa_3$		$\kappa_1 - \kappa_2$		$\kappa_1 - \kappa_3$		$\kappa_2 - \kappa_3$	
	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.	Cob.	Long.
0.1	0.952	0.137	0.949	0.175	0.945	0.188	0.954	0.137	0.949	0.175	0.946	0.188
0.2	0.949	0.134	0.952	0.137	0.957	0.152	0.950	0.134	0.952	0.137	0.957	0.152
0.3	0.948	0.133	0.956	0.118	0.950	0.134	0.950	0.133	0.956	0.118	0.949	0.134
0.4	0.950	0.132	0.942	0.107	0.953	0.125	0.948	0.132	0.940	0.107	0.951	0.125
0.5	0.944	0.133	0.945	0.101	0.949	0.120	0.944	0.133	0.944	0.101	0.948	0.120
0.6	0.951	0.133	0.956	0.098	0.950	0.117	0.948	0.133	0.956	0.098	0.949	0.117
0.7	0.949	0.134	0.946	0.097	0.951	0.117	0.948	0.134	0.946	0.097	0.951	0.117
0.8	0.956	0.136	0.949	0.097	0.952	0.119	0.953	0.136	0.950	0.097	0.955	0.119
0.9	0.954	0.141	0.958	0.098	0.950	0.126	0.955	0.141	0.959	0.098	0.950	0.126

IC: intervalo de confianza. Cob.: cobertura. Long.: longitud.

2.1.5.4. Conclusiones

El coeficiente kappa ponderado es una medida del acuerdo clasificatorio entre un test diagnóstico binario y el gold estándar, y es una medida válida para evaluar y comparar el rendimiento de tests diagnósticos binarios. Se ha estudiado un test de hipótesis conjunto para comparar la igualdad de los coeficientes kappa ponderados (para un mismo índice de ponderación c) de J tests binarios respecto al mismo gold estándar, cuando todos los tests binarios y el gold estándar se aplican a la misma muestra de individuos. El test de hipótesis conjunto se basa en la distribución chi-cuadrado, donde la matriz de varianzas-covarianzas de los J coeficientes kappa se puede estimar aplicando el método delta o bien mediante el método bootstrap. Se han realizado unos experimentos de simulación para estudiar el error tipo I y la potencia del test de hipótesis conjunto cuando se comparan los coeficientes kappa ponderados de tres tests binarios, tanto si la matriz de varianzas-covarianzas se estima mediante el método delta o por bootstrap, obteniéndose que no hay diferencias importantes entre los resultados obtenidos por ambos métodos. Asimismo, se han comparado el error tipo I y la potencia del test de hipótesis conjunto con los errores tipo I y potencias de otros dos métodos para comparar los tres coeficientes kappa ponderados: un primer método basado en las comparaciones de los tres coeficientes kappa ponderados por parejas al error $\alpha = 5\%$ (comparaciones por parejas sin la corrección de Bonferroni), y un segundo método basado en las comparaciones por parejas con la corrección de Bonferroni ($\alpha = 0.05/3 = 1.67\%$). Los experimentos de simulación han demostrado que el primer método tiene un error tipo I que desborda claramente al error nominal, por lo que es un método que puede dar lugar a resultados erróneos; y el método de Bonferroni es un método conservador, cuyo error tipo I es siempre más pequeño que el error nominal, sobretodo para muestras de 500 o más pacientes. En cuanto al test de hipótesis conjunto, su error tipo I fluctúa alrededor del error nominal, aunque en algunas ocasiones, cuando la muestra y la prevalencia son pequeñas ($n = 100$ y $p = 10\%$), el error tipo I puede desbordar al error nominal. Este problema es debido a que las celdas de la distribución multinomial no están lo suficientemente bien representadas, por lo que las varianzas de los estimadores pueden ser muy grandes dando lugar a resultados imprecisos. Este problema aparece tanto si la matriz de varianzas-covarianzas se estima

mediante el método delta o mediante bootstrap. Por tanto, el test de hipótesis conjunto basado en la distribución chi-cuadrado es un test de hipótesis que tiene el clásico comportamiento de un test asintótico, y que ha demostrado tener un buen rendimiento, tanto en términos de error tipo I como de potencia, cuando las muestras son grandes ($n \geq 500$).

Por tanto, una vez analizados los resultados de los experimentos de simulación, se propone el siguiente método para comparar los coeficientes kappa ponderados (para un mismo índice de ponderación c) de J tests diagnósticos binarios: 1) aplicar el test de hipótesis conjunto basado en la distribución chi-cuadrado al error α , estimando la matriz de varianzas-covarianzas mediante el método delta o mediante bootstrap; 2) si el test de hipótesis conjunto no es significativo al error α , la homogeneidad de los J coeficientes kappa ponderados no se rechaza; pero si el test conjunto es significativo, para buscar las causas de la significación se aplica el método de Bonferroni, es decir se realizan las comparaciones por parejas al error $2\alpha/(J(J-1))$. La razón por la que se realizan las comparaciones por parejas al error $2\alpha/(J(J-1))$ y no al error α , es que el segundo método ha demostrado tener un error tipo I que desborda al error nominal, tal y como se ha comprobado en los experimentos de simulación.

2.2.- Verificación parcial

La comparación de la exactitud de dos tests diagnósticos binarios en presencia de verificación parcial de la enfermedad ha sido objeto de numerosos estudios (Zhou, 1998a; Roldán Nofuentes y Luna del Castillo, 2005, 2006, 2008; Harel y Zhou, 2007a). En presencia de verificación parcial de la enfermedad, la estimación y comparación de la exactitud de dos tests binarios (o de múltiples tests binarios) no se puede realizar aplicando los métodos estudiados en la Sección 2.1 prescindiendo de los pacientes no verificados con el gold estándar, ya que los resultados que se obtendrían estarían afectados por el sesgo de verificación y por tanto las conclusiones podrían ser erróneas. A continuación se estudian los tests de hipótesis para comparar diferentes parámetros de dos tests binarios y de múltiples tests binarios en presencia de verificación parcial de la enfermedad.

2.2.1. Comparación de dos tests binarios

Sean dos tests diagnósticos binarios que se aplican de forma independiente a una misma muestra aleatoria de n pacientes. Sean T_1 , T_2 y D las variables aleatorias binarias definidas en la Sección 2.1, y sea V la variable aleatoria que modeliza el estado de verificación de la enfermedad, siendo $V = 1$ cuando el paciente ha sido verificado con el gold estándar y $V = 0$ cuando no ha sido verificado. Sean Se_h y Sp_h la sensibilidad y la especificidad de cada test diagnóstico ($h = 1, 2$), p la prevalencia de la enfermedad, y $\lambda_{ijk} = P(V = 1 | T_1 = i, T_2 = j, D = k)$ la probabilidad de verificar un paciente con resultados $T_1 = i$, $T_2 = j$ y $D = k$, con $i, j, k = 0, 1$.

La aplicación de los dos tests diagnósticos a todos los pacientes de la muestra aleatoria de tamaño n y la aplicación del gold estándar a una parte de la muestra da lugar a la Tabla de frecuencias 2.13.

Tabla 2.13. Frecuencias observadas cuando se comparan dos tests binarios en presencia de verificación parcial de la enfermedad.

	$T_1 = 1$		$T_1 = 0$		Total
	$T_2 = 1$	$T_2 = 0$	$T_2 = 1$	$T_2 = 0$	
$V = 1$					
$D = 1$	s_{11}	s_{10}	s_{01}	s_{00}	s
$D = 0$	r_{11}	r_{10}	r_{01}	r_{00}	r
$V = 0$	u_{11}	u_{10}	u_{01}	u_{00}	u
Total	n_{11}	n_{10}	n_{01}	n_{00}	n

2.2.1.1. Comparación de las sensibilidades y de las especificidades

A continuación se estudia la comparación de las sensibilidades y de las especificidades de dos tests diagnósticos binarios en presencia de verificación parcial de la enfermedad (Zhou, 1998b).

2.2.1.1.1. Comparación de sensibilidades y especificidades sin covariables

Cuando la probabilidad de seleccionar a un individuo para verificar su estado de enfermedad depende solamente de los resultados de los dos tests y no del resultado del gold estándar, es decir cuando

$$P(V|T_1, T_2, D) = P(V|T_1, T_2), \quad (2.55)$$

el proceso de verificación es MAR (Rubin, 1976) y se verifica que $\lambda_{ijk} = \lambda_{ij}$, $i, j, k = 0, 1$.

Bajo esta suposición, sean θ_{ij} y η_{ij} las probabilidades definidas como

$$\theta_{ij} = P(D=1|T_1=i, T_2=j) \quad \text{y} \quad \eta_{ij} = P(T_1=i, T_2=j) \quad (2.56)$$

para $i, j = 0, 1$, donde $\eta_{11} = 1 - \eta_{00} - \eta_{10} - \eta_{01}$. Sean $\boldsymbol{\theta} = (\theta_{00}, \theta_{10}, \theta_{01}, \theta_{11})$ y $\boldsymbol{\eta} = (\eta_{00}, \eta_{10}, \eta_{01})$. En términos de las probabilidades θ_{ij} y η_{ij} , la sensibilidad de cada test binario se escribe como

$$Se_1 = \frac{\sum_{j=0}^1 \theta_{1j} \eta_{1j}}{p} \quad (2.57)$$

y

$$Se_2 = \frac{\sum_{i=0}^1 \theta_{i1} \eta_{i1}}{p}, \quad (2.58)$$

y la especificidad como

$$Sp_1 = \frac{\sum_{j=0}^1 (1 - \theta_{0j}) \eta_{0j}}{(1-p)} \quad (2.59)$$

y

$$Sp_2 = \frac{\sum_{i=0}^1 (1 - \theta_{i0}) \eta_{i0}}{(1 - p)}, \quad (2.60)$$

donde $p = \sum_{i,j=0}^1 \theta_{ij} \eta_{ij}$. Zhou (1998b) dedujo los estimadores máximo verosímiles de las probabilidades θ_{ij} y η_{ij} , dados por las expresiones

$$\hat{\theta}_{ij} = \frac{s_{ij}}{s_{ij} + r_{ij}} \quad \text{y} \quad \hat{\eta}_{ij} = \frac{n_{ij}}{n}, \quad (2.61)$$

donde n es el tamaño de la muestra. Sustituyendo en las ecuaciones (2.57) a (2.60) los parámetros θ_{ij} y η_{ij} por sus estimadores dados en las ecuaciones (2.61), los estimadores máximo verosímiles de las sensibilidades son

$$\hat{Se}_1 = \frac{\sum_{j=0}^1 \frac{s_{1j}}{s_{1j} + r_{1j}} n_{1j}}{\sum_{i=0}^1 \sum_{j=0}^1 \frac{s_{ij}}{s_{ij} + r_{ij}} n_{ij}} \quad \text{y} \quad \hat{Se}_2 = \frac{\sum_{i=0}^1 \frac{s_{i1}}{s_{i1} + r_{i1}} n_{i1}}{\sum_{i=0}^1 \sum_{j=0}^1 \frac{s_{ij}}{s_{ij} + r_{ij}} n_{ij}}, \quad (2.62)$$

y los de las especificidades

$$\hat{Sp}_1 = \frac{\sum_{j=0}^1 \frac{r_{0j}}{s_{0j} + r_{0j}} n_{0j}}{\sum_{i=0}^1 \sum_{j=0}^1 \frac{r_{ij}}{s_{ij} + r_{ij}} n_{ij}} \quad \text{y} \quad \hat{Sp}_2 = \frac{\sum_{i=0}^1 \frac{r_{i0}}{s_{i0} + r_{i0}} n_{i0}}{\sum_{i=0}^1 \sum_{j=0}^1 \frac{r_{ij}}{s_{ij} + r_{ij}} n_{ij}}. \quad (2.63)$$

Como las sensibilidades y las especificidades son funciones de θ_{ij} y η_{ij} , las matrices de varianzas–covarianzas asintóticas de los estimadores de las sensibilidades y de las especificidades se obtienen aplicando el método delta (Agresti, 2002). De esta forma, la matriz de varianzas–covarianzas asintóticas de las sensibilidades viene dada por la expresión

$$\frac{\partial (Se_1, Se_2)^T}{\partial \boldsymbol{\theta}} I_{\boldsymbol{\theta}}^{-1} \frac{\partial (Se_1, Se_2)}{\partial \boldsymbol{\theta}} + \frac{\partial (Se_1, Se_2)^T}{\partial \boldsymbol{\eta}} I_{\boldsymbol{\eta}}^{-1} \frac{\partial (Se_1, Se_2)}{\partial \boldsymbol{\eta}}, \quad (2.64)$$

donde I_{θ}^{-1} y I_{η}^{-1} son las matrices inversas de las matrices de información de Fisher de θ y η , dadas por las ecuaciones

$$I_{\theta}^{-1} = \text{diag} \left\{ \frac{\theta_{ij}^2 (1 - \theta_{ij})^2}{s_{ij} (1 - \theta_{ij})^2 + r_{ij} \theta_{ij}^2} \right\}, \quad i, j = 0, 1 \quad (2.65)$$

y

$$I_{\eta}^{-1} = \text{diag} \left(\frac{\eta_{00}^2}{n_{00}}, \frac{\eta_{01}^2}{n_{01}}, \frac{\eta_{10}^2}{n_{10}} \right) - \frac{1}{\sum_{i,j=0}^1 \frac{\eta_{ij}^2}{n_{ij}}} \left(\frac{\eta_{00}^2}{n_{00}}, \frac{\eta_{01}^2}{n_{01}}, \frac{\eta_{10}^2}{n_{10}} \right)^T \left(\frac{\eta_{00}^2}{n_{00}}, \frac{\eta_{01}^2}{n_{01}}, \frac{\eta_{10}^2}{n_{10}} \right). \quad (2.66)$$

Realizando las operaciones algebraicas sobre la matriz (2.64), el elemento (r, s) -ésimo de la matriz de varianzas–covarianzas asintóticas de las sensibilidades es

$$\begin{aligned} \sigma_{rs} = & \sum_{i,j=0}^1 \frac{\theta_{ij}^2 (1 - \theta_{ij})^2}{s_{ij} (1 - \theta_{ij})^2 + r_{ij} \theta_{ij}^2} \frac{\partial Se_r}{\partial \theta_{ij}} \frac{\partial Se_s}{\partial \theta_{ij}} + \sum_{(i,j) \neq (1,1)} \frac{\eta_{ij}^2}{n_{ij}} \frac{\partial Se_r}{\partial \eta_{ij}} \frac{\partial Se_s}{\partial \eta_{ij}} \\ & - \frac{1}{\sum_{i,j=0}^1 \frac{\eta_{ij}^2}{n_{ij}}} \left(\sum_{(i,j) \neq (1,1)} \frac{\eta_{ij}^2}{n_{ij}} \frac{\partial Se_r}{\partial \eta_{ij}} \right) \left(\sum_{(i,j) \neq (1,1)} \frac{\eta_{ij}^2}{n_{ij}} \frac{\partial Se_s}{\partial \eta_{ij}} \right), \end{aligned} \quad (2.67)$$

con $r, s = 1, 2$. Sustituyendo en esta ecuación cada parámetro por su estimador se obtiene la expresión del correspondiente estimador del (r, s) -ésimo elemento de esta matriz. Una vez obtenidos los estimadores máximo verosímiles de las sensibilidades y sus varianzas–covarianzas asintóticas, aplicando el teorema de Slutski, el estadístico para el test de hipótesis de igualdad de las dos sensibilidades

$$\begin{aligned} H_0 : Se_1 &= Se_2 \\ H_1 : Se_1 &\neq Se_2 \end{aligned}$$

es

$$z_{\text{exp}} = \frac{\hat{S}e_1 - \hat{S}e_2}{\sqrt{\widehat{\text{Var}}(\hat{S}e_1) + \widehat{\text{Var}}(\hat{S}e_2) - 2\widehat{\text{Cov}}(\hat{S}e_1, \hat{S}e_2)}} \xrightarrow{n \rightarrow \infty} N(0,1). \quad (2.68)$$

De forma análoga se deduce el test de hipótesis para comparar las dos especificidades. La matriz de varianzas–covarianzas asintóticas de las especificidades es

$$\frac{\partial(Sp_1, Sp_2)^T}{\partial\theta} I_{\theta}^{-1} \frac{\partial(Sp_1, Sp_2)}{\partial\theta} + \frac{\partial(Sp_1, Sp_2)^T}{\partial\eta} I_{\eta}^{-1} \frac{\partial(Sp_1, Sp_2)}{\partial\eta}, \quad (2.69)$$

y su (r, s) -ésimo elemento

$$\begin{aligned} \sigma_{rs} = & \sum_{i,j=0}^1 \frac{\theta_{ij}^2 (1-\theta_{ij})^2}{s_{ij} (1-\theta_{ij})^2 + r_{ij} \theta_{ij}^2} \frac{\partial Sp_r}{\partial \theta_{ij}} \frac{\partial Sp_s}{\partial \theta_{ij}} + \sum_{(i,j) \neq (1,1)} \frac{\eta_{ij}^2}{n_{ij}} \frac{\partial Sp_r}{\partial \eta_{ij}} \frac{\partial Sp_s}{\partial \eta_{ij}} \\ & - \frac{1}{\sum_{i,j=0}^1 \frac{\eta_{ij}^2}{n_{ij}}} \left(\sum_{(i,j) \neq (1,1)} \frac{\eta_{ij}^2}{n_{ij}} \frac{\partial Sp_r}{\partial \eta_{ij}} \right) \left(\sum_{(i,j) \neq (1,1)} \frac{\eta_{ij}^2}{n_{ij}} \frac{\partial Sp_s}{\partial \eta_{ij}} \right), \end{aligned} \quad (2.70)$$

con $r, s = 1, 2$. Finalmente, el estadístico para el test de hipótesis

$$\begin{aligned} H_0 : Sp_1 &= Sp_2 \\ H_1 : Sp_1 &\neq Sp_2 \end{aligned}$$

es

$$z_{\text{exp}} = \frac{\hat{S}p_1 - \hat{S}p_2}{\sqrt{\widehat{\text{Var}}(\hat{S}p_1) + \widehat{\text{Var}}(\hat{S}p_2) - 2\widehat{\text{Cov}}(\hat{S}p_1, \hat{S}p_2)}} \xrightarrow{n \rightarrow \infty} N(0,1). \quad (2.71)$$

2.2.1.1.2. Comparación de sensibilidades y especificidades con covariables

Zhou (1998b) también ha deducido los estimadores máximo verosímiles de las sensibilidades y la especificidades y los correspondientes tests de hipótesis cuando, en presencia de verificación parcial de la enfermedad, se observa en todos los pacientes un vector \mathbf{X} de covariables discretas. En esta situación, la suposición (2.55) no es válida. Si la probabilidad de seleccionar a un paciente para verificar su estado de enfermedad

depende sólo de los resultados de los dos tests y del vector de covariables \mathbf{X} y no del verdadero estado de la enfermedad, es decir si

$$P(V|D, T_1, T_2, \mathbf{X}) = P(V|T_1, T_2, \mathbf{X}), \quad (2.72)$$

entonces este problema de inferencia se puede resolver. Suponiendo que el vector \mathbf{X} de covariables discretas tiene un número I de patrones diferentes de covariables, donde $x_i = (x_{i0}, x_{i1}, \dots, x_{ik})$ es el i -ésimo patrón de covariables de k covariables, y $x_{i0} = 1$, $i = 1, \dots, I$, y sea $\xi_i = P(\mathbf{X} = x_i)$. Zhou (1998b) ha deducido los estimadores máximo verosímiles de la sensibilidad y especificidad de cada test binario,

$$\hat{S}e_1 = \frac{\sum_{i=1}^I \sum_{l=0}^1 \hat{\theta}_{i1l} \hat{\eta}_{i1l} \hat{\xi}_i}{\hat{p}} \quad \text{y} \quad \hat{S}p_1 = \frac{\sum_{i=1}^I \sum_{l=0}^1 (1 - \hat{\theta}_{i0l}) \hat{\eta}_{i0l} \hat{\xi}_i}{1 - \hat{p}} \quad (2.73)$$

para el test 1, y

$$\hat{S}e_2 = \frac{\sum_{i=1}^I \sum_{j=0}^1 \hat{\theta}_{ij1} \hat{\eta}_{ij1} \hat{\xi}_i}{\hat{p}} \quad \text{y} \quad \hat{S}p_2 = \frac{\sum_{i=1}^I \sum_{j=0}^1 (1 - \hat{\theta}_{ij0}) \hat{\eta}_{ij0} \hat{\xi}_i}{1 - \hat{p}} \quad (2.74)$$

para el test 2, y donde

$$\hat{p} = \sum_{i=1}^I \sum_{j,l=0}^1 \hat{\theta}_{ijl} \hat{\eta}_{ijl} \hat{\xi}_i \quad (2.75)$$

es el estimador máximo verosímil de la prevalencia de la enfermedad. Sean $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^T$, $\boldsymbol{\alpha} = (\alpha_{000}, \alpha_{001}, \alpha_{010}, \alpha'_{100}, \alpha'_{101}, \alpha'_{110})^T$ y $\boldsymbol{\xi} = (\xi_1, \dots, \xi_{I-1})^T$. Los estimadores de $\hat{\theta}_{ijl}$, $\hat{\eta}_{ijl}$ y $\hat{\xi}_i$ vienen dados por las ecuaciones

$$\hat{\theta}_{ijl} = \hat{P}(D=1|T_1=j, T_2=l, \mathbf{X}=x_i) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 j + \hat{\beta}_2 l + \hat{\beta}_3 x_i)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 j + \hat{\beta}_2 l + \hat{\beta}_3 x_i)},$$

$$\hat{\eta}_{ijl} = \hat{P}(T_1 = j, T_2 = l | \mathbf{X} = x_i) = \frac{\exp(\hat{\alpha}_{0jl} + \hat{\alpha}_{1jl}^T x_i)}{\sum_{h_1, h_2=0}^1 \exp(\hat{\alpha}_{0jl} + \hat{\alpha}_{1h_1 h_2}^T x_i)}, \quad (2.76)$$

$$\hat{\xi}_i = \hat{P}(\mathbf{X} = x_i) = \frac{n_i}{n},$$

para $j, l = 0, 1, i = 1, \dots, I$ y donde $\hat{\alpha}_{011} = \hat{\alpha}_{011}^T = 0$. Cada estimador $\hat{\theta}_{ijl}$ sigue un modelo de regresión logística y cada $\hat{\eta}_{ijl}$ un modelo logit multinomial.

Aplicando el método delta, la matriz de varianzas–covarianzas asintóticas de las sensibilidades es

$$\begin{aligned} & \frac{\partial (Se_1, Se_2)^T}{\partial \alpha} I_{\alpha}^{-1} \frac{\partial (Se_1, Se_2)}{\partial \alpha} + \frac{\partial (Se_1, Se_2)^T}{\partial \beta} I_{\beta}^{-1} \frac{\partial (Se_1, Se_2)}{\partial \beta} \\ & + \\ & \frac{\partial (Se_1, Se_2)^T}{\partial \xi} I_{\xi}^{-1} \frac{\partial (Se_1, Se_2)}{\partial \xi} \end{aligned} \quad (2.77)$$

evaluado en $\alpha = \hat{\alpha}$, $\beta = \hat{\beta}$ y $\xi = \hat{\xi}$. De esta forma, el estadístico para el test de hipótesis

$$\begin{aligned} H_0 : Se_1 &= Se_2 \\ H_1 : Se_1 &\neq Se_2 \end{aligned}$$

es

$$z_{\text{exp}} = \frac{\hat{Se}_1 - \hat{Se}_2}{\sqrt{\widehat{Var}(\hat{Se}_1) + \widehat{Var}(\hat{Se}_2) - 2\widehat{Cov}(\hat{Se}_1, \hat{Se}_2)}} \xrightarrow{n \rightarrow \infty} N(0, 1). \quad (2.78)$$

El test de hipótesis para contrastar la igualdad de las dos especificidades en presencia de covariables se resuelve de forma similar sustituyendo Se por Sp .

2.2.1.2. Comparación de las razones de verosimilitud

Roldán Nofuentes y Luna del Castillo (2005) han deducido los estimadores máximo verosímiles de las razones de verosimilitud positiva y negativa de dos tests binarios y han deducido los tests de hipótesis para comparar estos parámetros en presencia de verificación parcial de la enfermedad.

2.2.1.2.1. Comparación de las razones de verosimilitud sin covariables

En términos de las probabilidades θ_{ij} y η_{ij} las razones de verosimilitud de cada test diagnóstico son

$$LR_1^+ = \left(\frac{1 - \sum_{ij=0}^1 \theta_{ij} \eta_{ij}}{\sum_{ij=0}^1 \theta_{ij} \eta_{ij}} \right) \left(\frac{\sum_{j=0}^1 \theta_{1j} \eta_{1j}}{\sum_{j=0}^1 (1 - \theta_{1j}) \eta_{1j}} \right) \quad (2.79)$$

y

$$LR_1^- = \left(\frac{1 - \sum_{ij=0}^1 \theta_{ij} \eta_{ij}}{\sum_{ij=0}^1 \theta_{ij} \eta_{ij}} \right) \left(\frac{\sum_{j=0}^1 \theta_{0j} \eta_{0j}}{\sum_{j=0}^1 (1 - \theta_{0j}) \eta_{0j}} \right) \quad (2.80)$$

para el test 1, y

$$LR_2^+ = \left(\frac{1 - \sum_{ij=0}^1 \theta_{ij} \eta_{ij}}{\sum_{ij=0}^1 \theta_{ij} \eta_{ij}} \right) \left(\frac{\sum_{i=0}^1 \theta_{i1} \eta_{i1}}{\sum_{i=0}^1 (1 - \theta_{i1}) \eta_{i1}} \right) \quad (2.81)$$

y

$$LR_2^- = \left(\frac{1 - \sum_{ij=0}^1 \theta_{ij} \eta_{ij}}{\sum_{ij=0}^1 \theta_{ij} \eta_{ij}} \right) \left(\frac{\sum_{i=0}^1 \theta_{i0} \eta_{i0}}{\sum_{i=0}^1 (1 - \theta_{i0}) \eta_{i0}} \right) \quad (2.82)$$

para el test 2, y sus correspondientes estimadores máximo verosímiles son

$$\widehat{LR}_1^+ = \begin{pmatrix} \frac{\sum_{ij=0}^1 \frac{r_{ij}}{s_{ij} + r_{ij}} n_{ij}}{\sum_{ij=0}^1 \frac{s_{ij}}{s_{ij} + r_{ij}} n_{ij}} & \frac{\sum_{j=0}^1 \frac{s_{1j}}{s_{1j} + r_{1j}} n_{1j}}{\sum_{j=0}^1 \frac{r_{1j}}{s_{1j} + r_{1j}} n_{1j}} \end{pmatrix} \quad (2.83)$$

y

$$\widehat{LR}_1^- = \begin{pmatrix} \frac{\sum_{ij=0}^1 \frac{r_{ij}}{s_{ij} + r_{ij}} n_{ij}}{\sum_{ij=0}^1 \frac{s_{ij}}{s_{ij} + r_{ij}} n_{ij}} & \frac{\sum_{j=0}^1 \frac{s_{0j}}{s_{0j} + r_{0j}} n_{0j}}{\sum_{j=0}^1 \frac{r_{0j}}{s_{0j} + r_{0j}} n_{0j}} \end{pmatrix}, \quad (2.84)$$

y para el test 2

$$\widehat{LR}_2^+ = \begin{pmatrix} \frac{\sum_{ij=0}^1 \frac{r_{ij}}{s_{ij} + r_{ij}} n_{ij}}{\sum_{ij=0}^1 \frac{s_{ij}}{s_{ij} + r_{ij}} n_{ij}} & \frac{\sum_{i=0}^1 \frac{s_{i1}}{s_{i1} + r_{i1}} n_{i1}}{\sum_{i=0}^1 \frac{r_{i1}}{s_{i1} + r_{i1}} n_{i1}} \end{pmatrix} \quad (2.85)$$

y

$$\widehat{LR}_2^- = \begin{pmatrix} \frac{\sum_{ij=0}^1 \frac{r_{ij}}{s_{ij} + r_{ij}} n_{ij}}{\sum_{ij=0}^1 \frac{s_{ij}}{s_{ij} + r_{ij}} n_{ij}} & \frac{\sum_{i=0}^1 \frac{s_{i0}}{s_{i0} + r_{i0}} n_{i0}}{\sum_{i=0}^1 \frac{r_{i0}}{s_{i0} + r_{i0}} n_{i0}} \end{pmatrix}. \quad (2.86)$$

Aplicando el método delta (Agresti, 2002), la matriz de varianzas–covarianzas asintóticas de las razones de verosimilitud (positivas o negativas) es

$$\frac{\partial(LR_1, LR_2)^T}{\partial\theta} I_{\theta}^{-1} \frac{\partial(LR_1, LR_2)}{\partial\theta} + \frac{\partial(LR_1, LR_2)^T}{\partial\eta} I_{\eta}^{-1} \frac{\partial(LR_1, LR_2)}{\partial\eta}, \quad (2.87)$$

donde LR es LR^+ o LR^- . Realizando operaciones algebraicas, (r, s) -ésimo elemento de esta matriz es

$$\begin{aligned} \sigma_{rs} = & \sum_{i,j=0}^1 \frac{\theta_{ij}^2 (1-\theta_{ij})^2}{s_{ij} (1-\theta_{ij})^2 + r_{ij} \theta_{ij}^2} \frac{\partial LR_r}{\partial \theta_{ij}} \frac{\partial LR_s}{\partial \theta_{ij}} + \sum_{(i,j) \neq (1,1)} \frac{\eta_{ij}^2}{n_{ij}} \frac{\partial LR_r}{\partial \eta_{ij}} \frac{\partial LR_s}{\partial \eta_{ij}} \\ & - \frac{1}{\sum_{i,j=0}^1 \frac{\eta_{ij}^2}{n_{ij}}} \left(\sum_{(i,j) \neq (1,1)} \frac{\eta_{ij}^2}{n_{ij}} \frac{\partial LR_r}{\partial \eta_{ij}} \right) \left(\sum_{(i,j) \neq (1,1)} \frac{\eta_{ij}^2}{n_{ij}} \frac{\partial LR_s}{\partial \eta_{ij}} \right), \end{aligned} \quad (2.88)$$

con $r, s = 1, 2$. Finalmente, el estadístico para el test de hipótesis

$$H_0 : LR_1 = LR_2$$

$$H_1 : LR_1 \neq LR_2$$

es

$$z_{\text{exp}} = \frac{\widehat{LR}_1 - \widehat{LR}_2}{\sqrt{\widehat{Var}(\widehat{LR}_1) + \widehat{Var}(\widehat{LR}_2) - 2\widehat{Cov}(\widehat{LR}_1, \widehat{LR}_2)}} \xrightarrow{n \rightarrow \infty} N(0,1). \quad (2.89)$$

2.2.1.2.2. Comparación de las razones de verosimilitud con covariables

Si en los n pacientes de la muestra se observa un vector \mathbf{X} de covariables discretas y la probabilidad de selección de un paciente para verificar su estado de enfermedad depende sólo de los resultados de los dos tests y del vector de covariables \mathbf{X} , los estimadores máximo verosímiles de las razones de verosimilitud positiva y negativa del test 1 son

$$\widehat{LR}_1^+ = \left(\frac{1 - \sum_{i=1}^l \sum_{j,l=0}^1 \hat{\theta}_{ijl} \hat{\eta}_{ijl} \hat{\xi}_i}{\sum_{i=1}^l \sum_{j,l=0}^1 \hat{\theta}_{ijl} \hat{\eta}_{ijl} \hat{\xi}_i} \right) \left(\frac{\sum_{i=1}^l \sum_{l=0}^1 \hat{\theta}_{iil} \hat{\eta}_{iil} \hat{\xi}_i}{1 - \sum_{i=1}^l \sum_{l=0}^1 \hat{\eta}_{i0l} \hat{\xi}_i - \sum_{i=1}^l \sum_{l=0}^1 \hat{\theta}_{iil} \hat{\eta}_{iil} \hat{\xi}_i} \right) \quad (2.90)$$

y

$$\widehat{LR}_1^- = \left(\frac{1 - \sum_{i=1}^I \sum_{j,l=0}^1 \hat{\theta}_{ijl} \hat{\eta}_{ijl} \hat{\xi}_i}{\sum_{i=1}^I \sum_{j,l=0}^1 \hat{\theta}_{ijl} \hat{\eta}_{ijl} \hat{\xi}_i} \right) \left(\frac{\sum_{i=1}^I \sum_{l=0}^1 \hat{\theta}_{i0l} \hat{\eta}_{i0l} \hat{\xi}_i}{\sum_{i=1}^I \sum_{l=0}^1 (1 - \hat{\theta}_{i0l}) \hat{\eta}_{i0l} \hat{\xi}_i} \right), \quad (2.91)$$

y

$$\widehat{LR}_2^+ = \left(\frac{1 - \sum_{i=1}^I \sum_{j,l=0}^1 \hat{\theta}_{ijl} \hat{\eta}_{ijl} \hat{\xi}_i}{\sum_{i=1}^I \sum_{j,l=0}^1 \hat{\theta}_{ijl} \hat{\eta}_{ijl} \hat{\xi}_i} \right) \left(\frac{\sum_{i=1}^I \sum_{j=0}^1 \hat{\theta}_{ij1} \hat{\eta}_{ij1} \hat{\xi}_i}{1 - \sum_{i=1}^I \sum_{j=0}^1 \hat{\eta}_{ij0} \hat{\xi}_i - \sum_{i=1}^I \sum_{j=0}^1 \hat{\theta}_{ij1} \hat{\eta}_{ij1} \hat{\xi}_i} \right) \quad (2.92)$$

y

$$\widehat{LR}_2^- = \left(\frac{1 - \sum_{i=1}^I \sum_{j,l=0}^1 \hat{\theta}_{ijl} \hat{\eta}_{ijl} \hat{\xi}_i}{\sum_{i=1}^I \sum_{j,l=0}^1 \hat{\theta}_{ijl} \hat{\eta}_{ijl} \hat{\xi}_i} \right) \left(\frac{\sum_{i=1}^I \sum_{j=0}^1 \hat{\theta}_{ij0} \hat{\eta}_{ij0} \hat{\xi}_i}{\sum_{i=1}^I \sum_{j=0}^1 (1 - \hat{\theta}_{ij0}) \hat{\eta}_{ij0} \hat{\xi}_i} \right). \quad (2.93)$$

para el test 2. De forma similar al caso de las sensibilidades y especificidades, la matriz de varianzas–covarianzas de las razones de verosimilitud (positivas o negativas) es

$$\begin{aligned} & \frac{\partial(LR_1, LR_2)^T}{\partial \alpha} I_{\alpha}^{-1} \frac{\partial(LR_1, LR_2)}{\partial \alpha} + \frac{\partial(LR_1, LR_2)^T}{\partial \beta} I_{\beta}^{-1} \frac{\partial(LR_1, LR_2)}{\partial \beta} \\ & + \\ & \frac{\partial(LR_1, LR_2)^T}{\partial \xi} I_{\xi}^{-1} \frac{\partial(LR_1, LR_2)}{\partial \xi}, \end{aligned} \quad (2.94)$$

donde LR es LR^+ o LR^- , y el estadístico para el test de hipótesis

$$\begin{aligned} H_0 &: LR_1 = LR_2 \\ H_1 &: LR_1 \neq LR_2 \end{aligned}$$

es

$$z_{\text{exp}} = \frac{\widehat{LR}_1 - \widehat{LR}_2}{\sqrt{\widehat{Var}(\widehat{LR}_1) + \widehat{Var}(\widehat{LR}_2) - 2\widehat{Cov}(\widehat{LR}_1, \widehat{LR}_2)}} \xrightarrow{n \rightarrow \infty} N(0,1), \quad (2.95)$$

2.2.1.3. Comparación de los valores predictivos

Roldán Nofuentes y Luna del Castillo (2008) han deducido los tests de hipótesis para comparar los valores predictivos de dos tests diagnósticos cuando ambos tests se aplican a la misma muestra de pacientes en presencia de verificación parcial de la enfermedad, tanto si no se observan covariables como si en todos los pacientes se observan covariables discretas.

2.2.1.3.1. Comparación de valores predictivos sin covariables

En términos de las probabilidades θ_{ij} y η_{ij} dadas por las ecuaciones (2.56), los valores predictivos de cada test binario se escriben como:

$$vpp_1 = \frac{\sum_{j=0}^1 \theta_{1j} \eta_{1j}}{\sum_{j=0}^1 \eta_{1j}} \quad \text{y} \quad vpn_1 = \frac{\sum_{j=0}^1 (1 - \theta_{0j}) \eta_{0j}}{\sum_{j=0}^1 \eta_{0j}}, \quad (2.96)$$

y

$$vpp_2 = \frac{\sum_{i=0}^1 \theta_{i1} \eta_{i1}}{\sum_{i=0}^1 \eta_{i1}} \quad \text{y} \quad vpn_2 = \frac{\sum_{i=0}^1 (1 - \theta_{i0}) \eta_{i0}}{\sum_{i=0}^1 \eta_{i0}}. \quad (2.97)$$

Sus estimadores máximo verosímiles son

$$\widehat{vpp}_1 = \frac{n_{11}s_{11}(s_{10} + r_{10}) + n_{10}s_{10}(s_{11} + r_{11})}{(n_{11} + n_{10})(s_{11} + r_{11})(s_{10} + r_{10})} \quad (2.98)$$

y

$$\widehat{vpn}_1 = \frac{n_{01}r_{01}(s_{00} + r_{00}) + n_{00}r_{00}(s_{01} + r_{01})}{(n_{01} + n_{00})(s_{01} + r_{01})(s_{00} + r_{00})} \quad (2.99)$$

para el test 1, y

$$\widehat{vpp}_2 = \frac{n_{11}s_{11}(s_{01} + r_{01}) + n_{01}s_{01}(s_{11} + r_{11})}{(n_{11} + n_{01})(s_{11} + r_{11})(s_{01} + r_{01})} \quad (2.100)$$

y

$$\widehat{vpn}_2 = \frac{n_{10}r_{10}(s_{00} + r_{00}) + n_{00}r_{00}(s_{10} + r_{10})}{(n_{10} + n_{00})(s_{10} + r_{10})(s_{00} + r_{00})} \quad (2.101)$$

para el test 2. Como los valores predictivos son funciones de θ_{ij} y η_{ij} , aplicando de nuevo el método delta (Agresti, 2002), la matriz de varianzas–covarianzas asintóticas de los valores predictivos (positivos o negativos) es

$$\frac{\partial(vp_1, vp_2)^T}{\partial\theta} I_{\theta}^{-1} \frac{\partial(vp_1, vp_2)}{\partial\theta} + \frac{\partial(vp_1, vp_2)^T}{\partial\eta} I_{\eta}^{-1} \frac{\partial(vp_1, vp_2)}{\partial\eta}, \quad (2.102)$$

donde vp es el valor predictivo positivo o el negativo. Como en las Secciones anteriores, el estadístico para el test de hipótesis

$$H_0 : vp_1 = vp_2$$

$$H_1 : vp_1 \neq vp_2$$

es

$$\frac{\widehat{vp}_1 - \widehat{vp}_2}{\sqrt{\widehat{Var}(\widehat{vp}_1) + \widehat{Var}(\widehat{vp}_2) - 2\widehat{Cov}(\widehat{vp}_1, \widehat{vp}_2)}} \xrightarrow{n \rightarrow \infty} N(0,1). \quad (2.103)$$

2.2.1.3.2. Comparación de los valores predictivos con covariables

Si en todos los pacientes se observa el vector \mathbf{X} de covariables discretas, los valores predictivos positivo y negativo para el test 1 vienen dados por las expresiones

$$vpp_1 = \frac{\sum_{i=1}^I \sum_{l=0}^1 \theta_{i0l} \eta_{i0l} \xi_i}{1 - \sum_{i=1}^I \sum_{l=0}^1 \eta_{i0l} \xi_i} \quad (2.104)$$

y

$$vpn_1 = \frac{\sum_{i=1}^I \sum_{l=0}^1 (1 - \theta_{i0l}) \eta_{i0l} \xi_i}{\sum_{i=1}^I \sum_{l=0}^1 \eta_{i0l} \xi_i}, \quad (2.105)$$

y para el test 2

$$vpp_2 = \frac{\sum_{i=1}^I \sum_{j=0}^1 \theta_{ij1} \eta_{ij1} \xi_i}{1 - \sum_{i=1}^I \sum_{j=0}^1 \eta_{ij0} \xi_i} \quad (2.106)$$

y

$$vpn_2 = \frac{\sum_{i=1}^I \sum_{j=0}^1 (1 - \theta_{ij0}) \eta_{ij0} \xi_i}{\sum_{i=1}^I \sum_{j=0}^1 \eta_{ij0} \xi_i}. \quad (2.107)$$

De forma análoga a como se ha resuelto este problema de inferencia con las sensibilidades, especificidades y razones de verosimilitud, la matriz de varianzas-covarianzas asintóticas de los valores predictivos es

$$\begin{aligned} & \frac{\partial (vp_1, vp_2)^T}{\partial \alpha} I_{\alpha}^{-1} \frac{\partial (vp_1, vp_2)}{\partial \alpha} + \frac{\partial (vp_1, vp_2)^T}{\partial \beta} I_{\beta}^{-1} \frac{\partial (vp_1, vp_2)}{\partial \beta} \\ & + \\ & \frac{\partial (vp_1, vp_2)^T}{\partial \xi} I_{\xi}^{-1} \frac{\partial (vp_1, vp_2)}{\partial \xi}, \end{aligned} \quad (2.108)$$

y el estadístico para el test de hipótesis de igualdad de los valores predictivos positivos o negativos en presencia de covariables es similar al de la Sección anterior.

2.2.1.3.3. Comparación de valores predictivos con los estimadores simples

Zhou (1994) estudió el efecto del sesgo de verificación en los valores predictivos de un test diagnóstico binario, demostrando que cuando el proceso de verificación de la enfermedad sólo depende de los resultados del test diagnóstico, los estimadores máximo verosímiles de los valores predictivos coinciden con sus respectivos estimadores simples.

Cuando se comparan dos tests diagnósticos binarios utilizando sólo los pacientes con el estado de la enfermedad verificado, los estimadores simples de los valores predictivos para el test 1 son

$$\widehat{vpp}_{1nv} = \frac{s_{11} + s_{10}}{(s_{11} + r_{11}) + (s_{10} + r_{10})} \quad (2.109)$$

y

$$\widehat{vpn}_{1nv} = \frac{r_{01} + r_{00}}{(s_{01} + r_{01}) + (s_{00} + r_{00})}, \quad (2.110)$$

y los del test 2 son

$$\widehat{vpp}_{2nv} = \frac{s_{11} + s_{01}}{(s_{11} + r_{11}) + (s_{01} + r_{01})} \quad (2.111)$$

y

$$\widehat{vpn}_{2nv} = \frac{r_{10} + r_{00}}{(s_{10} + r_{10}) + (s_{00} + r_{00})}. \quad (2.112)$$

Bajo la hipótesis MAR, las probabilidades de selección de un paciente para la verificación de la enfermedad son $\lambda_{ij} = P(V = 1 | T_1 = i, T_2 = j)$, con $i, j = 0, 1$, siendo sus estimadores máximo verosímiles

$$\hat{\lambda}_{ij} = \frac{s_{ij} + r_{ij}}{n_{ij}}. \quad (2.113)$$

Cuando se comparan los valores predictivos de dos tests binarios en presencia de verificación parcial de la enfermedad, los resultados obtenidos por Zhou (1994) no son validos debido a que los estimadores simples de los valores predictivos de cada test binario están afectados por el sesgo de verificación. En la Tabla 2.14 se presentan las distribuciones marginales de cada test diagnóstico, de manera que si en presencia de verificación parcial de la enfermedad y bajo la hipótesis MAR, dos tests diagnósticos binarios son aplicados a todos los pacientes de una misma muestra aleatoria, la evaluación de cada test diagnóstico no puede ser realizada desde su distribución marginal, es decir, cada test no puede ser evaluado desde la correspondiente tabla 3×2

(Roldán Nofuentes y Luna del Castillo, 2008), ya que en general $P(V = 1|T_1 = i) \neq \lambda_{i1} + \lambda_{i0}$ y $P(V = 1|T_2 = j) \neq \lambda_{1j} + \lambda_{0j}$.

Tabla 2.14. Distribución marginal de los tests 1 y 2.

Distribución marginal del test 1	$T_1 = 1$	$T_1 = 0$
$V = 1$		
$D = 1$	$s_{11} + s_{10}$	$s_{01} + s_{00}$
$D = 0$	$r_{11} + r_{10}$	$r_{01} + r_{00}$
$V = 0$	$u_{11} + u_{10}$	$u_{01} + u_{00}$
Total	$n_{11} + n_{10}$	$n_{01} + n_{00}$
Distribución marginal del test 2	$T_2 = 1$	$T_2 = 0$
$V = 1$		
$D = 1$	$s_{11} + s_{01}$	$s_{10} + s_{00}$
$D = 0$	$r_{11} + r_{01}$	$r_{10} + r_{00}$
$V = 0$	$u_{11} + u_{01}$	$u_{10} + u_{00}$
Total	$n_{11} + n_{01}$	$n_{10} + n_{00}$

2.2.1.4. Comparación de coeficientes kappa ponderados

De forma similar a los parámetros anteriores, los coeficientes kappa ponderados de los tests 1 y 2 (Roldán Nofuentes y Luna del Castillo, 2006) vienen dados por las expresiones

$$\kappa_1(c) = \frac{\sum_{j=0}^1 \theta_{1j} n_{1j} - \left(1 - \sum_{j=0}^1 \eta_{0j}\right) \left(\sum_{i,j=0}^1 \theta_{ij} n_{ij}\right)}{\left(1 - c - \sum_{j=0}^1 \eta_{0j}\right) \left(\sum_{i,j=0}^1 (1 - \theta_{ij}) n_{ij}\right) + c \sum_{j=0}^1 \eta_{0j}} \quad (2.114)$$

y

$$\kappa_2(c) = \frac{\sum_{i=0}^1 \theta_{i1} \eta_{i1} - \left(1 - \sum_{i=0}^1 \eta_{i0}\right) \left(\sum_{i,j=0}^1 \theta_{ij} \eta_{ij}\right)}{\left(1 - c - \sum_{i=0}^1 \eta_{i0}\right) \left(\sum_{i,j=0}^1 (1 - \theta_{ij}) \eta_{ij}\right) + c \sum_{i=0}^1 \eta_{i0}}, \quad (2.115)$$

y sus estimadores máximo verosímiles son

$$\hat{\kappa}_1(c) = \frac{\sum_{j=0}^1 \frac{s_{1j}}{s_{1j} + r_{1j}} n_{1j} - \frac{1}{n} \left(\sum_{j=0}^1 n_{1j}\right) \left(\sum_{i,j=0}^1 \frac{s_{ij}}{s_{ij} + r_{ij}} \eta_{ij}\right)}{\left(c - \frac{1}{n} \sum_{i,j=0}^1 \frac{r_{ij}}{s_{ij} + r_{ij}} n_{ij}\right) \left(\sum_{j=0}^1 n_{0j}\right) + (1-c) \left(\sum_{i,j=0}^1 \frac{r_{ij}}{s_{ij} + r_{ij}} n_{ij}\right)} \quad (2.116)$$

y

$$\hat{\kappa}_2(c) = \frac{\sum_{i=0}^1 \frac{s_{i1}}{s_{i1} + r_{i1}} n_{i1} - \frac{1}{n} \left(\sum_{j=0}^1 n_{i1}\right) \left(\sum_{i,j=0}^1 \frac{s_{ij}}{s_{ij} + r_{ij}} \eta_{ij}\right)}{\left(c - \frac{1}{n} \sum_{i,j=0}^1 \frac{r_{ij}}{s_{ij} + r_{ij}} n_{ij}\right) \left(\sum_{i=0}^1 n_{i0}\right) + (1-c) \left(\sum_{i,j=0}^1 \frac{r_{ij}}{s_{ij} + r_{ij}} n_{ij}\right)}. \quad (2.117)$$

Aplicando de nuevo el método delta (Agresti, 2002), el elemento (r, s) -ésimo de la matriz de varianzas–covarianzas asintóticas de los coeficiente kappa ponderado es

$$\sigma_{rs} = \sum_{i,j=0}^1 \frac{\theta_{ij}^2 (1 - \theta_{ij})^2}{s_{ij} (1 - \theta_{ij})^2 + r_{ij} \theta_{ij}^2} \frac{\partial \kappa_r(c)}{\partial \theta_{ij}} \frac{\partial \kappa_s(c)}{\partial \theta_{ij}} + \sum_{(i,j) \neq (1,1)} \frac{\eta_{ij}^2}{n_{ij}} \frac{\partial \kappa_r(c)}{\partial \eta_{ij}} \frac{\partial \kappa_s(c)}{\partial \eta_{ij}} \quad (2.118)$$

$$- \frac{1}{\sum_{i,j=0}^1 \frac{\eta_{ij}^2}{n_{ij}}} \left(\sum_{(i,j) \neq (1,1)} \frac{\eta_{ij}^2}{n_{ij}} \frac{\partial \kappa_r(c)}{\partial \eta_{ij}} \right) \left(\sum_{(i,j) \neq (1,1)} \frac{\eta_{ij}^2}{n_{ij}} \frac{\partial \kappa_s(c)}{\partial \eta_{ij}} \right),$$

con $r, s = 1, 2$. Una vez obtenidos los estimadores de los coeficientes kappa ponderados y sus varianzas–covarianzas asintóticas, el estadístico para contrastar la igualdad de los dos coeficientes kappa ponderados (para un mismo índice de ponderación)

$$H_0 : \kappa_1(c) = \kappa_2(c)$$

$$H_1 : \kappa_1(c) \neq \kappa_2(c)$$

es

$$z_{\text{exp}} = \frac{\hat{\kappa}_1(c) - \hat{\kappa}_2(c)}{\sqrt{\widehat{\text{Var}}(\hat{\kappa}_1(c)) + \widehat{\text{Var}}(\hat{\kappa}_2(c)) - 2\widehat{\text{Cov}}(\hat{\kappa}_1(c), \hat{\kappa}_2(c))}} \xrightarrow{n \rightarrow \infty} N(0,1). \quad (2.119)$$

La comparación de los dos coeficientes kappa ponderados en presencia de covariables discretas se realiza empleando la misma metodología que para el resto de parámetros anteriores.

2.2.1.5. Imputación múltiple para comparar dos tests

Harel y Zhou (2007a) han estudiado la comparación de las sensibilidades y de las especificidades de dos tests binarios en presencia de verificación parcial mediante imputación múltiple. La imputación múltiple es una técnica basada en la simulación para resolver problemas de inferencia con datos faltantes. En la situación de verificación parcial, cada valor faltante se reemplaza por m conjuntos de posibles valores, dando lugar a m sistemas de conjuntos completos que sólo difieren en los valores imputados, y para cada conjunto completo se calculan las estimaciones de la sensibilidades y de las especificidades y sus errores estándares, que combinados mediante reglas aritméticas simples, dan un resultado que toma en consideración los valores faltantes. Las fases del método de imputación múltiple para comparar dos tests binarios en presencia de verificación parcial de la enfermedad son:

1. Fase de imputación

El paso principal de la imputación múltiple es obtener la distribución a posteriori de los datos con el verdadero estado de enfermedad. A lo largo del procedimiento de imputación, se utiliza el aumento de los datos para imputar los valores perdidos. Bajo la hipótesis de que no se conoce la estructura de los datos, dada la Tabla 2.5, se puede considerar que los datos provienen de una distribución del multinomial. Utilizando propiedades de la distribución multinomial, se tiene que

$$\left(s_{jk}, r_{jk} \right) \Big| Y_{\text{obs}}, \theta \sim M \left(u_{jk}, \left(\theta_{1jk} / \theta_{+jk}, \theta_{0jk} / \theta_{+jk} \right) \right), \quad j, k = 0, 1, \quad (2.120)$$

donde θ_{ijk} es la probabilidad de que una unidad se divida en celdas (i, j, k) y $\theta_{+jk} = \sum_i \theta_{ijk}$. Si se elige una distribución a priori Dirichlet para los parámetros de la distribución multinomial, se obtienen los siguientes resultados

$$(s, r) | \theta \sim M(n, \theta), \quad (2.121)$$

$$\theta \sim D(\alpha), \quad (2.122)$$

y

$$\theta | Y \sim D(\alpha'), \quad (2.123)$$

donde $\alpha' = \alpha + (s, r)$ y $D(\alpha)$ es una distribución Dirichlet con parámetro α .

Para la siguiente etapa de imputación se utiliza la distribución a priori de Jeffrey, práctica común en el análisis bayesiano cuando se quiere utilizar una distribución a priori no informativa.

2. Fase de análisis

Después de imputar los datos perdidos, se obtienen m conjuntos de datos completos, las estimaciones $(\hat{Q}^{(1)}, \hat{Q}^{(2)}, \dots, \hat{Q}^{(m)})$ y las varianzas asociadas $(U^{(1)}, U^{(2)}, \dots, U^{(m)})$ para cada sensibilidad y especificidad.

3. Combinación de resultados

Después de tener m conjuntos de estimaciones y sus varianzas, se utilizan las reglas de Rubin (1978), de manera que la estimación global es $\bar{Q} = (1/m) \sum Q^{(i)}$, $i = 1, \dots, m$, y su varianza es $T = \bar{U} + (1/(m+1))B$, donde $\bar{U} = (1/m) \sum U^i$ es la varianza estimada de los datos completos, y $(1/(m+1))B$ es la varianza adicional debida a la imputación de los valores faltantes. La diferencia entre la imputación, B , se calcula como

$$B = \frac{1}{m-1} \sum_{i=1}^m (Q^{(i)} - \bar{Q})^2. \quad (2.124)$$

Las estimaciones se basan en la aproximación de la distribución *t-Student* $T^{-1/2}(Q - \hat{Q}) \sim t_\nu$, donde los grados de libertad son

$$\nu = (m-1) \left[1 + \frac{\bar{U}}{(1+m^{-1})B} \right]^2. \quad (2.125)$$

Además, el intervalo de confianza para la estimación, a un nivel de confianza $100(1-\alpha)$ viene dado por

$$\bar{Q} \pm t_{\nu, 1-\alpha/2} \sqrt{T}. \quad (2.126)$$

Harel y Zhou (2007a) han propuesto este método de imputación múltiple para comparar dos tests binarios utilizando diferentes intervalos de confianza para la diferencia de las sensibilidades y de las especificidades, concluyendo que el mejor método para comparar estos parámetros es combinar la imputación múltiple junto con el intervalo de confianza de Wald con corrección por continuidad.

2.2.1.6. Algoritmos EM y SEM para comparar dos tests

Roldán Nofuentes, Luna del Castillo y Femia Marzo (2009) han estudiado la comparación de las sensibilidades y especificidades de dos tests binarios en presencia de verificación parcial de la enfermedad mediante los algoritmos EM y SEM. El algoritmo EM permite estimar las sensibilidades y especificidades de los dos tests diagnósticos, y el algoritmo SEM (Meng y Rubin, 1991) permite estimar la matriz de varianzas-covarianzas asintóticas.

En presencia de verificación parcial de la enfermedad la información faltante es el verdadero estado de enfermedad de los pacientes no verificados. Esta información es reconstruida en el paso E del algoritmo, y en el paso M se calculan los estimadores máximo verosímiles a partir de los datos reconstruidos en el paso anterior. Si de los u_{ij}

pacientes no verificados, x_{ij} están enfermos y $u_{ij} - x_{ij}$ no están enfermos, entonces los datos de la Tabla 2.14 se pueden escribir en forma de una tabla 2×4 con frecuencias $s_{ij} + x_{ij}$ para $D=1$ y frecuencias $r_{ij} + u_{ij} - x_{ij}$ para $D=0$. Si el proceso de verificación es MAR, la función del logaritmo de la verosimilitud de los datos de esta tabla 2×4 es

$$\begin{aligned}
 l(Se_1, Sp_1, Se_2, Sp_2, p, \alpha_1, \alpha_0) \propto & \sum_{i,j=0}^1 (s_{ij} + x_{ij}) \log \{P(T_1 = i, T_2 = j, D = 1)\} \\
 & + \\
 & \sum_{i,j=0}^1 (r_{ij} + u_{ij} - x_{ij}) \log \{P(T_1 = i, T_2 = j, D = 0)\}.
 \end{aligned} \tag{2.127}$$

El algoritmo EM es el siguiente. Sea $x_{ij}^{(k)}$ el valor de x_{ij} en la k -ésima iteración del algoritmo, entonces los estimadores máximo verosímiles vienen dados por las expresiones:

$$\begin{aligned}
 \hat{Se}_1^{(k)} &= \frac{\sum_{j=0}^1 (s_{1j} + x_{1j}^{(k)})}{s + x^{(k)}}, & \hat{Sp}_1^{(k)} &= \frac{\sum_{j=0}^1 (r_{0j} + u_{0j} - x_{0j}^{(k)})}{r + u - x^{(k)}}, \\
 \hat{Se}_2^{(k)} &= \frac{\sum_{i=0}^1 (s_{i1} + x_{i1}^{(k)})}{s + x^{(k)}}, & \hat{Sp}_2^{(k)} &= \frac{\sum_{i=0}^1 (r_{i0} + u_{i0} - x_{i0}^{(k)})}{r + u - x^{(k)}}, \\
 \hat{p}^{(k)} &= \frac{s + x^{(k)}}{n}, \\
 \hat{\alpha}_1^{(k)} &= \frac{(s + x^{(k)})(s_{11} + x_{11}^{(k)})}{\left\{ \sum_{i=0}^1 (s_{i1} + x_{i1}^{(k)}) \right\} \left\{ \sum_{j=0}^1 (s_{1j} + x_{1j}^{(k)}) \right\}}, \\
 \hat{\alpha}_0^{(k)} &= \frac{(u + r - x^{(k)})(r_{11} + u_{11} - x_{11}^{(k)})}{\left\{ \sum_{i=0}^1 (r_{i1} + u_{i1} - x_{i1}^{(k)}) \right\} \left\{ \sum_{j=0}^1 (r_{1j} + u_{1j} - x_{1j}^{(k)}) \right\}},
 \end{aligned} \tag{2.128}$$

siendo $s = \sum_{i,j=0}^1 s_{ij}$, $r = \sum_{i,j=0}^1 r_{ij}$, $u = \sum_{i,j=0}^1 u_{ij}$, $x^{(k)} = \sum_{i,j=0}^1 x_{ij}^{(k)}$, $n = s + r + u$, y $\hat{\alpha}_1^{(k)}$ y $\hat{\alpha}_0^{(k)}$ son

los estimadores de las covarianzas entre los dos tests diagnósticos cuando $D=1$ y cuando $D=0$, respectivamente. En la siguiente iteración los estimadores son:

$$\begin{aligned}\hat{S}e_1^{(k+1)} &= \frac{\sum_{j=0}^1 (s_{1j} + x_{1j}^{(k+1)})}{s + x^{(k+1)}}, & \hat{S}p_1^{(k+1)} &= \frac{\sum_{j=0}^1 (r_{0j} + u_{0j} - x_{0j}^{(k+1)})}{r + u - x^{(k+1)}}, \\ \hat{S}e_2^{(k+1)} &= \frac{\sum_{i=0}^1 (s_{i1} + x_{i1}^{(k+1)})}{s + x^{(k+1)}}, & \hat{S}p_2^{(k+1)} &= \frac{\sum_{i=0}^1 (r_{i0} + u_{i0} - x_{i0}^{(k+1)})}{r + u - x^{(k+1)}}, \\ \hat{p}^{(k+1)} &= \frac{s + x^{(k+1)}}{n}, \\ \hat{\alpha}_1^{(k+1)} &= \frac{(s + x^{(k+1)})(s_{11} + x_{11}^{(k+1)})}{\left\{ \sum_{i=0}^1 (s_{i1} + x_{i1}^{(k+1)}) \right\} \left\{ \sum_{j=0}^1 (s_{1j} + x_{1j}^{(k+1)}) \right\}}, \\ \hat{\alpha}_0^{(k+1)} &= \frac{(u + r - x^{(k+1)})(r_{11} + u_{11} - x_{11}^{(k+1)})}{\left\{ \sum_{i=0}^1 (r_{i1} + u_{i1} - x_{i1}^{(k+1)}) \right\} \left\{ \sum_{j=0}^1 (r_{1j} + u_{1j} - x_{1j}^{(k+1)}) \right\}},\end{aligned}\tag{2.129}$$

siendo $x_{ij}^{(k+1)} = u_{ij} \frac{P^{(k)}(T_1 = i, T_2 = j, D = 1)}{P^{(k)}(T_1 = i, T_2 = j, D = 1) + P^{(k)}(T_1 = i, T_2 = j, D = 0)}$, con $i, j = 0, 1$, y

$P^{(k)}(T_1 = i, T_2 = j, D)$ es $P(T_1 = i, T_2 = j, D)$ en la k -ésima iteración del algoritmo.

Este algoritmo se repite hasta que los estimadores convergen. Como valores iniciales $x_{ij}^{(0)}$ se puede tomar cualquier valor comprendido entre 0 y u_{ij} . El algoritmo EM se detiene cuando la diferencia entre los valores de las funciones del logaritmo de la verosimilitud de dos iteraciones consecutivas es menor que un valor suficientemente pequeño, por ejemplo 10^{-10} o 10^{-12} . Este algoritmo se puede aplicar cuando alguna frecuencia s_{ij} o r_{ij} es igual a cero, en cuyo caso el valor $x_{ij}^{(0)}$ debe ser mayor que cero y

menor o igual que u_{ij} . Esto representa una ventaja sobre los métodos de Zhou (1998) y sobre el método de imputación múltiple de Harel y Zhou (2007a), ya que estos dos métodos no se pueden aplicar cuando alguna frecuencia s_{ij} o r_{ij} es cero, debido que en el primer método no se pueden estimar las matrices de varianzas-covarianzas asintóticas y en el segundo método no se pueden aplicar los modelos log-lineales para imputar los datos faltantes.

Una vez obtenidos los estimadores de las sensibilidades, especificidades, prevalencia y covarianzas entre los dos tests binarios, la estimación de la matriz de varianzas-covarianzas asintóticas de $\hat{S}e_1, \hat{S}p_1, \hat{S}e_2, \hat{S}p_2, \hat{p}, \hat{\alpha}_1$ y $\hat{\alpha}_0$ se puede obtener mediante la aplicación del algoritmo SEM (Meng y Rubin, 1991).

Sea Σ la matriz de varianzas-covarianzas de un vector de parámetros, Dempster et al. (1977) han demostrado que

$$\Sigma = I_{oc}^{-1} (I - DM)^{-1}, \quad (2.130)$$

donde $DM = I_{mis} I_{oc}^{-1}$, siendo I_{oc} la matriz de información de Fisher de los datos completos y I_{mis} la matriz de información de Fisher de los datos faltantes. El algoritmo SEM es un procedimiento numérico basado en el algoritmo EM para aproximar la matriz de varianzas-covarianzas del estimador de un vector de parámetros y consiste en tres partes: (1) evaluación de la matriz I_{oc}^{-1} , (2) evaluación de la matriz DM, y (3) evaluación de la matriz Σ . Por tanto, la aplicación del algoritmo SEM requiere en primer lugar evaluar la matriz I_{oc}^{-1} . Esta matriz es la inversa de la matriz de información de Fisher de los datos completos y se calcula a partir de la función del logaritmo de la verosimilitud de la última tabla 2×4 obtenida en la aplicación del algoritmo EM descrito anteriormente.

La segunda parte del algoritmo SEM consiste en evaluar la matriz DM. Los elementos d_{ij} , $i, j = 1, \dots, 7$, de la matriz DM se obtienen aplicando el siguiente algoritmo:

INPUT: $\hat{\theta} = (\hat{S}e_1, \hat{S}p_1, \hat{S}e_2, \hat{S}p_2, \hat{p}, \hat{\alpha}_1, \hat{\alpha}_0)$ y $\theta^{(t)} = (Se_1^{(t)}, Sp_1^{(t)}, Se_2^{(t)}, Sp_2^{(t)}, p^{(t)}, \alpha_1^{(t)}, \alpha_0^{(t)})$.

Paso 1: Calcular $\theta^{(t+1)} = (Se_1^{(t+1)}, Sp_1^{(t+1)}, Se_2^{(t+1)}, Sp_2^{(t+1)}, p^{(t+1)}, \alpha_1^{(t+1)}, \alpha_0^{(t+1)})$ aplicando el algoritmo EM descrito anteriormente.

Paso 2: Obtener los vectores $\theta_1^{(t)} = (Se_1^{(t)}, \hat{S}p_1, \hat{S}e_2, \hat{S}p_2, \hat{p}, \hat{\alpha}_1, \hat{\alpha}_0)$, $\theta_2^{(t)} = (\hat{S}e_1, Sp_1^{(t)}, \hat{S}e_2, \hat{S}p_2, \hat{p}, \hat{\alpha}_1, \hat{\alpha}_0)$, ..., $\theta_7^{(t)} = (Se_1, \hat{S}p_1, \hat{S}e_2, \hat{S}p_2, \hat{p}, \hat{\alpha}_1, \alpha_0^{(t)})$. Para cada uno de ellos ejecutar la primera iteración del algoritmo EM considerando a $\theta_i^{(t)}$ como el valor inicial de θ y obtener los valores $\tilde{\theta}_1^{(t+1)}$, $\tilde{\theta}_2^{(t+1)}$, ..., $\tilde{\theta}_7^{(t+1)}$.

Paso 3: Calcular los ratios

$$d_{ij}^{(t)} = \frac{\tilde{\theta}_{ij}^{(t+1)} - \hat{\theta}_j}{\theta_i^{(t)} - \hat{\theta}_i}, \quad i, j = 1, \dots, 7,$$

donde $\tilde{\theta}_{ij}^{(t+1)}$ es la j -ésima componente de $\tilde{\theta}_i^{(t+1)}$.

OUTPUT: $\theta^{(t+1)}$ y $d_{ij}^{(t)}$, $i, j = 1, \dots, 7$.

Este proceso se repite hasta que $|d_{ij}^{(t+1)} - d_{ij}^{(t)}| \leq \sqrt{\delta}$, siendo δ el valor del criterio de parada del algoritmo EM. Por tanto, cuanto mayor es δ mayores errores numéricos contiene la matriz DM, afectando a la estimación de la matriz de varianzas-covarianzas. Una vez imputada la matriz DM, la matriz de varianzas-covarianzas asintóticas se obtiene aplicando la ecuación (2.130). La matriz de varianzas-covarianzas estimadas no suele ser simétrica debido a los errores numéricos cometidos en la imputación de la matriz DM.

Una vez obtenidos los estimadores máximo verosímiles de las sensibilidades y especificidades y sus varianzas-covarianzas, el estadístico para contrastar la igualdad de las dos sensibilidades (especificidades) es

$$z_{\text{exp}} = \frac{\hat{S}e_1 - \hat{S}e_2}{\sqrt{\widehat{\text{Var}}(\hat{S}e_1) + \widehat{\text{Var}}(\hat{S}e_2) - 2\widehat{\text{Cov}}(\hat{S}e_1, \hat{S}e_2)}} \xrightarrow{n \rightarrow \infty} N(0,1), \quad (2.131)$$

y el intervalo de confianza al $100(1-\alpha)\%$ para la diferencia de las dos sensibilidades es

$$Se_1 - Se_2 \in \hat{S}e_1 - \hat{S}e_2 \pm z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{S}e_1) + \widehat{\text{Var}}(\hat{S}e_2) - 2\widehat{\text{Cov}}(\hat{S}e_1, \hat{S}e_2)}. \quad (2.132)$$

El estadístico para el test de hipótesis de igualdad de las dos especificidades y el intervalo de confianza para la diferencia de las dos especificidades se obtiene de forma similar sustituyendo Se por Sp .

2.2.2. Comparación de múltiples tests binarios

En la práctica clínica es frecuente que se compare la exactitud de más de dos tests binarios respecto a un mismo gold estándar. En presencia de verificación parcial de la enfermedad, Marín Jiménez (2009) ha estudiado diferentes tests de hipótesis para comparar diferentes parámetros (sensibilidades, especificidades, razones de verosimilitud, valores predictivos y coeficientes kappa ponderados) de más de dos tests binarios. En esta situación, sea T_j la variable aleatoria que modeliza el resultado del j -ésimo test binario ($j=1, \dots, J$), de tal forma que $T_j=1$ cuando el resultado del test es positivo y $T_j=0$ cuando el resultado es negativo. Sea V la variable aleatoria que modeliza el proceso de verificación, de tal forma que $V=1$ cuando el paciente es verificado con el gold estándar y $V=0$ cuando el paciente no es verificado con el gold estándar. Sea D la variable aleatoria que modeliza el resultado del gold estándar, de tal forma que $D=1$ cuando el paciente está enfermo y $D=0$ cuando el paciente está sano. Sea s_{i_1, \dots, i_j} el número de pacientes verificados en los que $T_1=i_1, T_2=i_2, \dots, T_j=i_j$ y $D=1$, r_{i_1, \dots, i_j} el número de pacientes verificados en los que $T_1=i_1, T_2=i_2, \dots, T_j=i_j$ y $D=0$, y u_{i_1, \dots, i_j} el número de pacientes no verificados en los

que $T_1 = i_1, T_2 = i_2, \dots, T_J = i_J$ con $i_j = 0, 1$ y $j = 1, \dots, J$. Sean

$n_{i_1, \dots, i_J} = s_{i_1, \dots, i_J} + r_{i_1, \dots, i_J} + u_{i_1, \dots, i_J}$ y $n = \sum_{i_1, \dots, i_J=0}^1 n_{i_1, \dots, i_J}$. Sean las probabilidades

$$\begin{aligned} \phi_{i_1, \dots, i_J} &= P(V = 1, D = 1, T_1 = i_1, T_2 = i_2, \dots, T_J = i_J) \\ \varphi_{i_1, \dots, i_J} &= P(V = 1, D = 0, T_1 = i_1, T_2 = i_2, \dots, T_J = i_J) \\ \gamma_{i_1, \dots, i_J} &= P(V = 1, T_1 = i_1, T_2 = i_2, \dots, T_J = i_J) \end{aligned} \tag{2.133}$$

con $i_j = 0, 1$ y verificándose que $\sum_{i_1, \dots, i_J=0}^1 \phi_{i_1, \dots, i_J} + \sum_{i_1, \dots, i_J=0}^1 \varphi_{i_1, \dots, i_J} + \sum_{i_1, \dots, i_J=0}^1 \gamma_{i_1, \dots, i_J} = 1$.

En la Tabla 2.15 se muestran los resultados obtenidos cuando se comparan tres tests binarios respecto a un mismo gold estándar en presencia de verificación parcial de la enfermedad.

Tabla 2.15. Clasificación cruzada de los resultados de tres tests diagnósticos con el estado de enfermedad y el estado de verificación.

	$T_1 = 1$				$T_1 = 0$			
	$T_2 = 1$		$T_2 = 0$		$T_2 = 1$		$T_2 = 0$	
	$T_3 = 1$	$T_3 = 0$	$T_3 = 1$	$T_3 = 0$	$T_3 = 1$	$T_3 = 0$	$T_3 = 1$	$T_3 = 0$
$V = 1$								
$D = 1$	s_{111}	s_{110}	s_{101}	s_{100}	s_{011}	s_{010}	s_{001}	s_{000}
$D = 0$	r_{111}	r_{110}	r_{101}	r_{100}	r_{011}	r_{010}	r_{001}	r_{000}
$V = 0$								
u_{111}	u_{111}	u_{110}	u_{101}	u_{100}	u_{011}	u_{010}	u_{001}	u_{000}
Total	n_{111}	n_{110}	n_{101}	n_{100}	n_{011}	n_{010}	n_{001}	n_{000}

2.2.2.1. Comparación de las sensibilidades y de las especificidades

Si el proceso de verificación depende solamente de los resultados de los J tests diagnósticos y no del estado de enfermedad, la sensibilidad y la especificidad del j -

ésimo test diagnóstico se pueden escribir en términos de las probabilidades (2.133) como:

$$Se_j = P(T_j = 1 | D = 1) = \frac{\sum_{i_1, \dots, i_j=0}^1 \frac{\phi_{i_1, \dots, i_j} \eta_{i_1, \dots, i_j}}{\phi_{i_1, \dots, i_j} + \varphi_{i_1, \dots, i_j}}}{p} \quad (2.134)$$

y

$$Sp_j = P(T_j = 0 | D = 0) = \frac{\sum_{i_1, \dots, i_j=0}^1 \frac{\varphi_{i_1, \dots, i_j} \eta_{i_1, \dots, i_j}}{\phi_{i_1, \dots, i_j} + \varphi_{i_1, \dots, i_j}}}{(1-p)}, \quad (2.135)$$

donde

$$p = P(D = 1) = \sum_{i_1, \dots, i_j=0}^1 \frac{\phi_{i_1, \dots, i_j} \eta_{i_1, \dots, i_j}}{\phi_{i_1, \dots, i_j} + \varphi_{i_1, \dots, i_j}} \quad (2.136)$$

es la prevalencia de la enfermedad y $\eta_{i_1, \dots, i_j} = \phi_{i_1, \dots, i_j} + \varphi_{i_1, \dots, i_j} + \gamma_{i_1, \dots, i_j}$. Sea $\omega = (\phi_{1, \dots, 1}, \phi_{0, \dots, 0}, \varphi_{1, \dots, 1}, \varphi_{1, \dots, 1}, \gamma_{1, \dots, 1}, \gamma_{1, \dots, 1})^T$ un vector de dimensión $3(2^J)$ cuyas componentes son las probabilidades $\phi_{i_1, \dots, i_j}, \varphi_{i_1, \dots, i_j}$ y γ_{i_1, \dots, i_j} , con $i_j = 0, 1$ y $j = 1, \dots, J$. Si el proceso de verificación depende solamente de los resultados de los J tests binarios, esto es equivalente a suponer que el proceso de verificación es MAR (Rubin, 1976), y todos los parámetros del vector ω se puede estimar mediante el método de máxima verosimilitud. El logaritmo de la función de verosimilitud de los datos observados es

$$l(\omega) = \sum_{i_1, \dots, i_j=0}^1 s_{i_1, \dots, i_j} \log(\phi_{i_1, \dots, i_j}) + \sum_{i_1, \dots, i_j=0}^1 r_{i_1, \dots, i_j} \log(\varphi_{i_1, \dots, i_j}) + \sum_{i_1, \dots, i_j=0}^1 u_{i_1, \dots, i_j} \log(\gamma_{i_1, \dots, i_j}). \quad (2.137)$$

Bajo la suposición MAR, maximizando esta función, los estimadores máximo verosímiles de las probabilidades $\phi_{i_1, \dots, i_j}, \varphi_{i_1, \dots, i_j}$ y γ_{i_1, \dots, i_j} vienen dados por las expresiones

$$\hat{\phi}_{i_1, \dots, i_j} = \frac{s_{i_1, \dots, i_j}}{n}, \quad \hat{\varphi}_{i_1, \dots, i_j} = \frac{r_{i_1, \dots, i_j}}{n} \quad \text{y} \quad \hat{\gamma}_{i_1, \dots, i_j} = \frac{u_{i_1, \dots, i_j}}{n}. \quad (2.138)$$

Sustituyendo en las ecuaciones (2.134) y (2.135) los parámetros por sus estimadores máximo verosímiles dados en las ecuaciones (2.138), los estimadores máximo verosímiles de la sensibilidad y especificidad del j -ésimo test diagnóstico son

$$\hat{S}e_j = \frac{\sum_{i_1, \dots, i_j=0}^1 \frac{S_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{S_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}}}{\sum_{i_1, \dots, i_j=0}^1 \frac{S_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{S_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}}} \quad (2.139)$$

y

$$\hat{S}p_j = \frac{\sum_{i_1, \dots, i_j=0}^1 \frac{r_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{S_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}}}{\sum_{i_1, \dots, i_j=0}^1 \frac{r_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{S_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}}} \quad (2.140)$$

Como ω es el vector de probabilidades de una distribución multinomial, la matriz de varianzas–covarianzas de $\hat{\omega}$ es

$$\Sigma_{\hat{\omega}} = \{diag(\omega) - \omega^T \omega\} / n. \quad (2.141)$$

Sean los vectores $\mathbf{Se} = (Se_1, \dots, Se_J)^T$ y $\mathbf{Sp} = (Sp_1, \dots, Sp_J)^T$ de dimensiones J . Aplicando el método delta (Agresti, 2002), las matrices de varianzas–covarianzas asintóticas de los vectores $\hat{\mathbf{Se}}$ y $\hat{\mathbf{Sp}}$ son

$$\Sigma_{\hat{\mathbf{Se}}} = \left(\frac{\partial \mathbf{Se}}{\partial \omega} \right) \Sigma_{\hat{\omega}} \left(\frac{\partial \mathbf{Se}}{\partial \omega} \right)^T \quad (2.142)$$

y

$$\Sigma_{\hat{\mathbf{Sp}}} = \left(\frac{\partial \mathbf{Sp}}{\partial \omega} \right) \Sigma_{\hat{\omega}} \left(\frac{\partial \mathbf{Sp}}{\partial \omega} \right)^T, \quad (2.143)$$

respectivamente. Sustituyendo en estas dos matrices de varianzas–covarianzas asintóticas los parámetros por sus respectivos estimadores máximo verosímiles, se

obtienen las varianzas–covarianzas estimadas de los estimadores máximo verosímiles de las sensibilidades y especificidades de los J tests diagnósticos binarios, matrices que deben ser evaluadas con un software matemático o estadístico debido a la complejidad de sus componentes.

Una vez obtenidos los estimadores máximo verosímiles de las sensibilidades y sus varianzas–covarianzas asintóticas, el test de hipótesis para contrastar la igualdad de todas las sensibilidades es

$$H_0 : Se_1 = Se_2 = \dots = Se_J$$

$$H_1 : \text{al menos una igualdad no es cierta,}$$

que es equivalente al test

$$H_0 : \boldsymbol{\psi} \mathbf{Se} = \mathbf{0}$$

$$H_1 : \boldsymbol{\psi} \mathbf{Se} \neq \mathbf{0},$$

donde $\boldsymbol{\psi}$ es una matriz de rango completo de dimensión $(J-1) \times J$ cuyos valores son constantes conocidas. Así, para $J = 3$ la matriz de diseño es

$$\boldsymbol{\psi} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}$$

y para $J = 4$

$$\boldsymbol{\psi} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

Finalmente, el estadístico para contrastar la igualdad de sensibilidades (Rao, 1968), es

$$Q_{\text{exp}}^2 = \hat{\mathbf{S}} \mathbf{e}^T \hat{\boldsymbol{\psi}}^T \left(\hat{\boldsymbol{\psi}} \hat{\boldsymbol{\Sigma}}_{\hat{\mathbf{S}}e} \hat{\boldsymbol{\psi}}^T \right)^{-1} \hat{\boldsymbol{\psi}} \hat{\mathbf{S}} \mathbf{e} \rightarrow \chi_{(J-1)}^2. \quad (2.144)$$

Si el test de hipótesis de igualdad de las sensibilidades es significativo al error α , para investigar las causas de la significación se realizan las comparaciones por parejas

de sensibilidades al error $2\alpha/(J(J-1))$. En este caso el estadístico de contraste se calcula a partir de los resultados obtenidos al aplicar el test de hipótesis global, por lo que el estadístico de contraste para comparar las sensibilidades del j -ésimo y del k -ésimo test binario es

$$z_{\text{exp}} = \frac{\hat{S}e_j - \hat{S}e_k}{\sqrt{\hat{V}ar(\hat{S}e_j) + \hat{V}ar(\hat{S}e_k) - 2\hat{C}ov(\hat{S}e_j, \hat{S}e_k)}} \rightarrow N(0,1), \quad (2.145)$$

donde las varianzas y covarianzas se obtienen a partir de la matriz de varianzas-covarianzas dada por la ecuación (2.142), de manera que el test de hipótesis $H_0: Se_j = Se_k$ será significativo cuando el p-valor a dos colas sea menor que $2\alpha/(J(J-1))$. El test de hipótesis de igualdad de las J especificidades se resuelve de forma similar sustituyendo en las ecuaciones anteriores las sensibilidades por las especificidades.

La comparación del rendimiento de los J tests diagnósticos binarios también se puede realizar mediante transformaciones sobre las sensibilidades (especificidades), como son la transformación logarítmica y la transformación logit, que son transformaciones que se utilizan con frecuencia cuando se evalúa o se compara la exactitud de tests diagnósticos binarios.

Si se realiza la transformación logarítmica sobre $\boldsymbol{\theta}$, donde $\boldsymbol{\theta}$ puede ser la sensibilidad o la especificidad, el test de hipótesis para contrastar la igualdad de los J logaritmos de las sensibilidades o especificidades es

$$\begin{aligned} H_0 &: \boldsymbol{\psi} \log(\boldsymbol{\theta}) = \mathbf{0} \\ H_1 &: \boldsymbol{\psi} \log(\boldsymbol{\theta}) \neq \mathbf{0}, \end{aligned}$$

y el estadístico de contraste es

$$Q_{\text{exp}}^2 = \left(\log(\hat{\boldsymbol{\theta}}) \right)^T \hat{\boldsymbol{\Psi}}^T \left(\hat{\boldsymbol{\Psi}} \hat{\boldsymbol{\Sigma}}_{\log(\hat{\boldsymbol{\theta}})} \hat{\boldsymbol{\Psi}}^T \right)^{-1} \hat{\boldsymbol{\Psi}} \left(\log(\hat{\boldsymbol{\theta}}) \right) \rightarrow \chi_{J-1}^2, \quad (2.146)$$

donde $\log(\boldsymbol{\theta}) = (\log(\theta_1), \log(\theta_2), \dots, \log(\theta_J))$ y

$$\Sigma_{\log(\hat{\theta})} = \left(\frac{\partial \log(\boldsymbol{\theta})}{\partial \boldsymbol{\omega}} \right) \Sigma_{\hat{\omega}} \left(\frac{\partial \log(\boldsymbol{\theta})}{\partial \boldsymbol{\omega}} \right)^T. \quad (2.147)$$

Si el test de hipótesis de igualdad de los logaritmos de $\boldsymbol{\theta}$ es significativo al error α , las causas de la significación se investigan de forma similar a como se investigan cuando no se realiza ninguna transformación sobre los parámetros.

Si en lugar de la transformación logarítmica se realiza la transformación logit, el estadístico para contrastar la igualdad de los J logits de las sensibilidades o especificidades es

$$Q_{\text{exp}}^2 = \left(\text{logit}(\hat{\boldsymbol{\theta}}) \right)^T \hat{\Psi}^T \left(\hat{\Psi} \hat{\Sigma}_{\text{logit}(\hat{\boldsymbol{\theta}})} \hat{\Psi}^T \right)^{-1} \hat{\Psi} \left(\text{logit}(\hat{\boldsymbol{\theta}}) \right) \rightarrow \chi_{J-1}^2, \quad (2.148)$$

y el resto del método es similar al caso con transformación logarítmica.

2.2.2.2. Comparación de las razones de verosimilitud

En esta Sección se van a deducir tests de hipótesis para comparar las razones de verosimilitud de más de dos tests diagnósticos binarios en presencia de verificación parcial de la enfermedad.

2.2.2.2.1. Comparación de las razones de verosimilitud positivas

Cuando el proceso de verificación es MAR, la razón de verosimilitud positiva del j -ésimo test diagnóstico se escribe en términos de las probabilidades (2.133) como

$$LR_j^+ = \frac{(1-p) \sum_{i_j=1}^1 \frac{\phi_{i_1, \dots, i_j} \eta_{i_1, \dots, i_j}}{\phi_{i_1, \dots, i_j} + \varphi_{i_1, \dots, i_j}}}{p \left((1-p) - \sum_{i_j=0}^1 \frac{\phi_{i_1, \dots, i_j} \eta_{i_1, \dots, i_j}}{\phi_{i_1, \dots, i_j} + \varphi_{i_1, \dots, i_j}} \right)}, \quad (2.149)$$

donde p es la prevalencia de la enfermedad definida en la ecuación (2.136). Sustituyendo en la ecuación (2.149) los parámetros por sus estimadores máximo

verosímiles, el estimador de la razón de verosimilitud positiva del j -ésimo test diagnóstico es

$$\widehat{LR}_j^+ = \frac{\left(\sum_{i_1, \dots, i_j=0}^1 \frac{r_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{S_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}} \right) \left(\sum_{i_j=1}^1 \frac{S_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{S_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}} \right)}{\left(\sum_{i_1, \dots, i_j=0}^1 \frac{S_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{S_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}} \right) \left(\sum_{i_1, \dots, i_j=0}^1 \frac{r_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{S_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}} - \sum_{i_j=0}^1 \frac{r_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{S_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}} \right)}. \quad (2.150)$$

Sea $\mathbf{LR}^+ = (LR_1^+, \dots, LR_J^+)^T$ un vector de dimensión J . Aplicando el método delta (Agresti, 2002), la matriz de varianzas–covarianzas de vector $\widehat{\mathbf{LR}}^+$ es:

$$\Sigma_{\widehat{\mathbf{LR}}^+} = \left(\frac{\partial \mathbf{LR}^+}{\partial \boldsymbol{\omega}} \right) \Sigma_{\boldsymbol{\omega}} \left(\frac{\partial \mathbf{LR}^+}{\partial \boldsymbol{\omega}} \right)^T. \quad (2.151)$$

El test de hipótesis para contrastar la igualdad de todas las razones de verosimilitud positivas es

$$\begin{aligned} H_0 : LR_1^+ = LR_2^+ = \dots = LR_J^+ \\ H_1 : \text{al menos una igualdad no es cierta,} \end{aligned}$$

que es equivalente a contrastar

$$\begin{aligned} H_0 : \boldsymbol{\psi} \mathbf{LR}^+ = \mathbf{0} \\ H_1 : \boldsymbol{\psi} \mathbf{LR}^+ \neq \mathbf{0}, \end{aligned}$$

y siguiendo con la misma metodología, el estadístico de contraste es

$$Q_{\text{exp}}^2 = \left(\widehat{\mathbf{LR}}^+ \right)^T \widehat{\boldsymbol{\psi}}^T \left(\widehat{\boldsymbol{\psi}} \widehat{\Sigma}_{\widehat{\mathbf{LR}}^+} \widehat{\boldsymbol{\psi}}^T \right)^{-1} \widehat{\boldsymbol{\psi}} \left(\widehat{\mathbf{LR}}^+ \right) \rightarrow \chi_{(J-1)}^2. \quad (2.152)$$

Si el test de hipótesis de igualdad de las J razones de verosimilitud es significativo al error α , las causas de la significación se investigan de forma análoga al caso de las sensibilidades y especificidades.

La comparación de las J razones de verosimilitud también se puede realizar utilizando la transformación logarítmica (la transformación logit no se puede utilizar ya la razón de verosimilitud positiva puede resultar mayor que la unidad) de forma similar al caso de las sensibilidades y especificidades.

2.2.2.2.2. Comparación de las razones de verosimilitud negativas

Cuando el resultado del j -ésimo test es negativo, su razón de verosimilitud es

$$LR_j^- = \frac{(1-p) \left(p - \sum_{\substack{i_1, \dots, i_j=0 \\ i_j=1}}^1 \frac{\phi_{i_1, \dots, i_j} \eta_{i_1, \dots, i_j}}{\phi_{i_1, \dots, i_j} + \varphi_{i_1, \dots, i_j}} \right)}{p \sum_{\substack{i_1, \dots, i_j=0 \\ i_j=0}}^1 \frac{\varphi_{i_1, \dots, i_j} \eta_{i_1, \dots, i_j}}{\phi_{i_1, \dots, i_j} + \varphi_{i_1, \dots, i_j}}} \quad (2.153)$$

y su estimador

$$\widehat{LR}_j^- = \frac{\left(\sum_{i_1, \dots, i_j=0}^1 \frac{r_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{s_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}} \right) \left(\sum_{i_1, \dots, i_j=0}^1 \frac{s_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{s_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}} - \sum_{\substack{i_1, \dots, i_j=0 \\ i_j=1}}^1 \frac{s_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{s_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}} \right)}{\left(\sum_{i_1, \dots, i_j=0}^1 \frac{s_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{s_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}} \right) \left(\sum_{i_j=0}^1 \frac{r_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{s_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}} \right)}. \quad (2.154)$$

De forma análoga a las razones de verosimilitud positivas, el estadístico de contraste para comparar las J razones de verosimilitud negativas es

$$Q_{\text{exp}}^2 = \left(\widehat{LR}^- \right)^T \widehat{\Psi}^T \left(\widehat{\Psi} \widehat{\Sigma}_{\widehat{LR}^-} \widehat{\Psi}^T \right)^{-1} \widehat{\Psi} \left(\widehat{LR}^- \right) \rightarrow \chi_{(J-1)}^2. \quad (2.155)$$

La investigación de las causas de la significación y el test de hipótesis conjunto con la transformación logarítmica son similares a los casos anteriores.

2.2.2.3. Comparación de los valores predictivos

El valor predictivo positivo del j -ésimo test diagnóstico se expresa en términos de las probabilidades (2.133) como

$$vpp_j = \frac{\sum_{\substack{i_1, \dots, i_j=0 \\ i_j=1}}^1 \frac{\phi_{i_1, \dots, i_j} \eta_{i_1, \dots, i_j}}{\phi_{i_1, \dots, i_j} + \varphi_{i_1, \dots, i_j}}}{\sum_{\substack{i_1, \dots, i_j=0 \\ i_j=1}}^1 \frac{\phi_{i_1, \dots, i_j} \eta_{i_1, \dots, i_j}}{\phi_{i_1, \dots, i_j} + \varphi_{i_1, \dots, i_j}} + (1-p) - \sum_{\substack{i_1, \dots, i_j=0 \\ i_j=0}}^1 \frac{\varphi_{i_1, \dots, i_j} \eta_{i_1, \dots, i_j}}{\phi_{i_1, \dots, i_j} + \varphi_{i_1, \dots, i_j}}} \quad (2.156)$$

y el valor predictivo negativo como

$$vpn_j = \frac{\sum_{\substack{i_1, \dots, i_j=0 \\ i_j=0}}^1 \frac{\varphi_{i_1, \dots, i_j} \eta_{i_1, \dots, i_j}}{\phi_{i_1, \dots, i_j} + \varphi_{i_1, \dots, i_j}}}{\sum_{\substack{i_1, \dots, i_j=0 \\ i_j=0}}^1 \frac{\varphi_{i_1, \dots, i_j} \eta_{i_1, \dots, i_j}}{\phi_{i_1, \dots, i_j} + \varphi_{i_1, \dots, i_j}} + p - \sum_{\substack{i_1, \dots, i_j=0 \\ i_j=1}}^1 \frac{\phi_{i_1, \dots, i_j} \eta_{i_1, \dots, i_j}}{\phi_{i_1, \dots, i_j} + \varphi_{i_1, \dots, i_j}}}, \quad (2.157)$$

y sus estimadores son

$$\widehat{vpp}_j = \frac{\sum_{\substack{i_1, \dots, i_j=0 \\ i_j=1}}^1 \frac{s_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{s_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}}}{\sum_{\substack{i_1, \dots, i_j=0 \\ i_j=1}}^1 \frac{s_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{s_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}} + \sum_{\substack{i_1, \dots, i_j=0 \\ i_j=0}}^1 \frac{r_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{s_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}} - \sum_{\substack{i_1, \dots, i_j=0 \\ i_j=0}}^1 \frac{r_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{s_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}}}. \quad (2.158)$$

y

$$\widehat{vpn}_j = \frac{\sum_{\substack{i_1, \dots, i_j=0 \\ i_j=0}}^1 \frac{r_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{s_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}}}{\sum_{\substack{i_1, \dots, i_j=0 \\ i_j=0}}^1 \frac{r_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{s_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}} + \sum_{\substack{i_1, \dots, i_j=0 \\ i_j=1}}^1 \frac{s_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{s_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}} - \sum_{\substack{i_1, \dots, i_j=0 \\ i_j=1}}^1 \frac{s_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{s_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}}}, \quad (2.159)$$

respectivamente. Aplicando la misma metodología, el estadístico para contrastar la igualdad de los valores predictivos positivos o negativos es

$$Q_{\text{exp}}^2 = \left(\widehat{\mathbf{vp}} \right)^T \widehat{\Psi}^T \left(\widehat{\Psi} \widehat{\Sigma}_{\widehat{\mathbf{vp}}} \widehat{\Psi}^T \right)^{-1} \widehat{\Psi} \left(\widehat{\mathbf{vp}} \right) \rightarrow \chi_{(J-1)}^2, \quad (2.160)$$

donde \mathbf{vp} es el vector cuyas componentes son los J valores predictivos positivos o negativos. La comparación de los J valores predictivos también se puede realizar utilizando las transformaciones logarítmica y logit.

2.2.2.4. Comparación de los coeficientes kappa ponderados

El coeficiente kappa ponderado del j -ésimo test es

$$\kappa_j(c) = \frac{q \sum_{\substack{i_1, \dots, i_j=0 \\ i_j=1}}^1 \frac{\phi_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{\phi_{i_1, \dots, i_j} + \varphi_{i_1, \dots, i_j}} + p \sum_{\substack{i_1, \dots, i_j=0 \\ i_j=0}}^1 \frac{\varphi_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{\phi_{i_1, \dots, i_j} + \varphi_{i_1, \dots, i_j}} + pq}{(p+c-1) \left\{ \frac{\sum_{i_j=0}^1 \frac{\varphi_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{\phi_{i_1, \dots, i_j} + \varphi_{i_1, \dots, i_j}}}{1-p} - \sum_{\substack{i_1, \dots, i_j=0 \\ i_j=1}}^1 \frac{\phi_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{\phi_{i_1, \dots, i_j} + \varphi_{i_1, \dots, i_j}} - p \sum_{\substack{i_1, \dots, i_j=0 \\ i_j=0}}^1 \frac{\varphi_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{\phi_{i_1, \dots, i_j} + \varphi_{i_1, \dots, i_j}} - p \right\} + c'q} \quad (2.161)$$

siendo y $c' = 1 - c$ y $q = 1 - p$, y su estimador es

$$\hat{\kappa}_j(c) = \frac{AE + BD + BA}{\{B + c - 1\} \left\{ \frac{D}{A} - E - BD - B \right\} + c'B}, \quad (2.162)$$

donde

$$A = \sum_{i_1, \dots, i_j=0}^1 \frac{r_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{S_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}}, \quad B = \sum_{i_1, \dots, i_j=0}^1 \frac{S_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{S_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}},$$

$$D = \sum_{\substack{i_1, \dots, i_j=0 \\ i_j=0}}^1 \frac{r_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{S_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}}, \quad E = \sum_{\substack{i_1, \dots, i_j=0 \\ i_j=1}}^1 \frac{S_{i_1, \dots, i_j} n_{i_1, \dots, i_j}}{S_{i_1, \dots, i_j} + r_{i_1, \dots, i_j}}.$$

Aplicando la misma metodología, el estadístico de contraste para comparar los coeficientes kappa ponderados de los J tests binarios es

$$Q_{\text{exp}}^2 = (\hat{\mathbf{k}}(c))^T \hat{\Psi}^T \left(\hat{\Psi} \hat{\Sigma}_{\hat{\mathbf{k}}(c)} \hat{\Psi}^T \right)^{-1} \hat{\Psi} (\hat{\mathbf{k}}(c)) \rightarrow \chi_{(J-1)}^2, \quad (2.163)$$

siendo $\hat{\mathbf{k}}(c)$ el vector cuyas componentes son los coeficientes kappa ponderados de cada test diagnóstico. Este test de hipótesis también se puede resolver aplicando las transformaciones logarítmica y logit, y las causas de la significación se investigan de la misma forma que para los parámetros anteriores.

Capítulo 3

Cálculo del tamaño muestral para evaluar y comparar parámetros de tests diagnósticos binarios

En este último Capítulo se estudia el cálculo del tamaño de muestra necesario para evaluar y comparar la exactitud de tests diagnósticos binarios, tanto si todos los pacientes son verificados con el gold estándar como si no.

Así, una importante cuestión en la evaluación y la comparación de la exactitud de tests diagnósticos binarios es la determinación del tamaño de muestra necesario para llevar a cabo el estudio. Al intentar evaluar la exactitud de un test binario, se calcula el tamaño de la muestra necesario para estimar la sensibilidad o la especificidad con una precisión determinada; y cuando se pretende comparar la exactitud de dos tests diagnósticos binarios, se debe calcular el tamaño de la muestra necesario para estimar la diferencia entre las sensibilidades (especificidades) con una determinada precisión o comparar las sensibilidades (especificidades) de dos tests binarios con un error α y a

una potencia $1-\beta$, en cuyo caso se controla tanto el error α como el error β . Una revisión de los métodos para calcular el tamaño de la muestra para evaluar y comparar los métodos diagnósticos puede verse en Obuchowski (1998) y en Zhou et al. (2002). Li y Fine (2004) han estudiado el cálculo del tamaño de muestra en la estimación de la sensibilidad y especificidad de un test diagnóstico en estudios prospectivos. Fosgate (2005) ha desarrollado un algoritmo para calcular el tamaño de muestra para estimar la exactitud de un test diagnóstico binario.

En este Capítulo se propone un método para calcular el tamaño de muestra cuando se pretende estimar el coeficiente kappa ponderado de un test binario en verificación completa de la enfermedad, y cuando se pretende estimar la sensibilidad y especificidad de un test diagnóstico binario y comparar las sensibilidades y especificidades de dos tests diagnósticos binarios en presencia de verificación parcial de la enfermedad.

3.1.- Verificación completa

Una cuestión importante en el estudio de los métodos estadísticos de diagnóstico es la determinación del tamaño de muestra necesario para poder estimar el rendimiento del test diagnóstico binario con una determinada precisión a una confianza $100(1-\alpha)\%$. En presencia de verificación total de la enfermedad, la determinación del tamaño de la muestra tanto para evaluar la exactitud de un test diagnóstico como para comparar la exactitud de dos tests binarios se realiza aplicando métodos tradicionales en tablas 2×2 o 2×4 respectivamente, donde la sensibilidad y especificidad se estiman como proporciones binomiales.

3.1.1. Tamaño de muestra para la estimación por intervalos de confianza

Sea un test diagnóstico binario cuyo rendimiento es evaluado con respecto a un gold estándar. Sea T la variable aleatoria que modeliza el resultado del test diagnóstico, de tal forma que $T=1$ cuando el resultado del test es positivo (indicando la presencia provisional de la enfermedad) y $T=0$ cuando el resultado del test es negativo

(indicando la ausencia provisional de la enfermedad); y sea D la variable aleatoria que sigue el modelo del resultado del gold estándar, de tal forma que $D=1$ cuando el paciente tiene la enfermedad y $D=0$ cuando el paciente no tiene la enfermedad. Sean $Se = P(T=1|D=1)$ y $Sp = P(T=0|D=0)$ la sensibilidad y la especificidad del test diagnóstico respectivamente, y sea $p = P(D=1)$ la prevalencia de la enfermedad y $q = 1 - p$. Cuando el test diagnóstico y el gold estándar se aplica a todos los pacientes de una muestra de tamaño n se obtiene la Tabla 3.1.

Tabla 3.1. Frecuencias observadas al aplicar un test binario y el gold estándar a una muestra.

	$T = 1$	$T = 0$	Total
$D = 1$	s_1	s_0	s
$D = 0$	r_1	r_0	r
Total	n_1	n_0	n

3.1.1.1. Tamaño de muestra para estimar la sensibilidad y la especificidad

En el cálculo del tamaño de muestra necesario para estimar la sensibilidad y la especificidad hay que tener en cuenta el tipo de diseño y el tipo de muestreo utilizado, ya que éste condiciona la fórmula que se utiliza para calcular el error estándar de la estimación y es diferente según sea el diseño del estudio. Por tanto, si se pretende obtener el tamaño de muestra aplicando un método exacto (Newcombe, 1998a; Martín Andrés y Luna del Castillo, 2004) ocurre, como se indica en el Capítulo 1, que para obtener un intervalo de confianza exacto para la sensibilidad (especificidad), éste depende de los grados de libertad de una variable aleatoria continua F de Snedecor. Un método exacto necesita que el tamaño muestral sea bastante grande, por lo que se prefiere utilizar fórmulas aproximadas para la obtención de dicho tamaño muestral. Por tanto, si se supone que la población es infinita o suficientemente grande, el tamaño de muestra para la sensibilidad (especificidad) se obtiene como:

$$m = \frac{z_{1-\alpha/2}^2 \theta(1-\theta)}{d^2}, \quad (3.1)$$

donde $z_{1-\alpha/2}$ es el $100(1-\alpha/2)$ -ésimo percentil de la distribución normal estándar, θ es la sensibilidad o la especificidad y d es la precisión de la estimación (longitud media del intervalo de confianza) dada de antemano, es decir $|\hat{\theta}-\theta|\leq d$. Se denota como tamaño de muestra la letra m en lugar de n , como ocurre usualmente, para no confundir al lector con el tamaño de muestra de pacientes objeto de estudio utilizado a lo largo de los capítulos.

El problema que surge cuando se utiliza la ecuación (3.1) es que para estimar el tamaño muestral, éste depende de θ , es decir de la sensibilidad o especificidad. Este valor es desconocido, de forma que para un valor de d fijo, esta fórmula toma su valor máximo cuando $\theta=0.5$. Pero d es la precisión en la estimación de θ , es decir, de la sensibilidad o especificidad, y está claro que la magnitud de esta precisión no se puede fijar si no se tiene alguna idea respecto a la sensibilidad o especificidad a estimar. Por ello, cuando no se dispone de información alguna acerca de los posibles valores de la sensibilidad o especificidad, lo que se suele hacer es utilizar una muestra piloto (se toma una muestra pequeña al inicio para estimar el parámetro, se calcula el intervalo de confianza para el parámetro y se sustituye en la fórmula (3.1) θ por el valor del intervalo más próximo a 0.5 que sea compatible con la información y una vez calculado el tamaño muestral se amplía esta muestra) para estimar los parámetros desconocidos, o se sustituye la sensibilidad (especificidad) por 0.5, de tal forma que $\theta(1-\theta)=1/4$ y la fórmula (3.1) queda como

$$m = \frac{z_{1-\alpha/2}^2}{4d^2}. \quad (3.2)$$

El tamaño de muestra también se puede calcular cuando se pretende contrastar

$$\begin{aligned} H_0 : \theta &= \theta_0 \\ H_1 : \theta &\neq \theta_0 \end{aligned}$$

con un error α y una potencia $1-\beta$. Para calcular el tamaño de muestra es necesario especificar la mínima diferencia δ con respecto a θ_0 que se desea que el test detecte. En este caso $\theta_1 = \theta_0 + \delta$ (sin pérdida de generalidad) será el primer valor de la

alternativa que se desea que el test de hipótesis detecte con una potencia $1-\beta$. Entonces la ecuación para calcular el tamaño muestral viene dada por la expresión

$$m = \left(\frac{z_{1-\alpha/2} \sqrt{\theta_0(1-\theta_0)} + z_{1-\beta} \sqrt{\theta_1(1-\theta_1)}}{\delta} \right)^2 \quad (3.3)$$

donde $\sqrt{\theta_0(1-\theta_0)/m}$ es la varianza de $\hat{\theta}$ bajo la hipótesis nula y $\sqrt{\theta_1(1-\theta_1)/m}$ es la varianza de $\hat{\theta}$ bajo la hipótesis alternativa.

Aunque es menos conocido, cuando la variable aleatoria es discreta, como ocurre en este caso ya que procede de una distribución binomial, se debe realizar una corrección por continuidad, siendo el nuevo tamaño muestral

$$m' = \frac{m}{4} \left\{ 1 + \sqrt{1 + \frac{2}{m\delta}} \right\}^2, \quad (3.4)$$

donde θ es la sensibilidad o la especificidad, m viene dado por la ecuación (3.3) y $\delta = \theta_1 - \theta_0$.

3.1.1.2. Tamaño de muestra para estimar el coeficiente kappa ponderado (Aportación)

La determinación del tamaño de muestra necesario para estimar el coeficiente kappa ponderado con una determinada precisión a una confianza $100(1-\alpha)\%$ se basa en el intervalo de confianza tipo Wald, por lo que la ecuación para calcular el tamaño muestral para construir un intervalo de confianza a dos colas para el coeficiente kappa ponderado del test diagnóstico es

$$m = \frac{z_{1-\alpha/2}^2}{d^2} \text{Var}(\hat{\kappa}(c)), \quad (3.5)$$

donde $z_{\alpha/2}$ es el $100(1-\alpha/2)$ -ésimo percentil de la distribución normal estándar, d es la precisión de la estimación y $\text{Var}(\hat{\kappa}(c))$ es la función varianza (McCullagh y Nelder,

1989) de $\hat{\kappa}(c)$. La ecuación (3.5) presenta el problema de estimar la función varianza de $\hat{\kappa}(c)$, para lo cual Roldán Nofuentes, Luna del Castillo y Montero Alonso (2009a) han propuesto el siguiente método.

Sean las variables aleatorias T y D definidas en la Sección 3.1.1, y sean las probabilidades

$$\xi_i = P(D = 1, T = i) = pSe^i(1 - Se)^{1-i} \quad (3.6)$$

y

$$\psi_i = P(D = 0, T = i) = (1 - p)Sp^{1-i}(1 - Sp)^i, \quad (3.7)$$

con $i = 0, 1$. Sea $\omega = (\xi_1, \xi_0, \psi_1, \psi_0)^T$. Como ω es el vector de probabilidades de una distribución multinomial, la matriz de varianzas-covarianzas de $\hat{\omega}$ es

$$\Sigma_{\hat{\omega}} = \{diag(\omega) - \omega^T \omega\} / n.$$

La sensibilidad, especificidad y la prevalencia de la enfermedad se escriben en términos de las probabilidades anteriores como $Se = \frac{\xi_1}{\xi_0 + \xi_1}$, $Sp = \frac{\psi_0}{\psi_0 + \psi_1}$ y $p = \xi_0 + \xi_1$, por lo que el coeficiente kappa ponderado del test binario se escribe en términos de estas mismas probabilidades como

$$\kappa(c) = \frac{\xi_1 \psi_0 - \xi_0 \psi_1}{(\xi_0 + \xi_1)(\xi_0 + \psi_0)c + (\psi_0 + \psi_1)(\xi_1 + \psi_1)(1 - c)}, \quad (3.8)$$

y aplicando de nuevo el método delta (Agresti, 2002) se obtiene que la función varianza de $\hat{\kappa}(c)$ es

$$Var(\hat{\kappa}(c)) = \left(\frac{\partial \kappa(c)}{\partial \omega} \right) \Sigma_{\hat{\omega}} \left(\frac{\partial \kappa(c)}{\partial \omega} \right)^T. \quad (3.9)$$

Esta expresión es larga y complicada, y necesita ser evaluada con un software matemático o estadístico. Una vez calculada la función varianza del coeficiente kappa

ponderado, el tamaño muestral mínimo para estimar el coeficiente kappa ponderado con una precisión d y una confianza del $100(1-\alpha)\%$ se calcula aplicando la ecuación (3.5). Por tanto, el método que se propone para calcular el tamaño muestral exige conocer la sensibilidad, especificidad y prevalencia (y por consiguiente el coeficiente kappa ponderado). Estos valores se pueden determinar a partir de una muestra piloto o bien a partir de otros estudios previos similares.

Este método también se puede utilizar para calcular el tamaño muestral necesario para contrastar

$$H_0 : \kappa(c) = \kappa_0(c)$$

$$H_1 : \kappa(c) \neq \kappa_0(c)$$

a un error α y una potencia $1-\beta$. Para ello es necesario especificar un valor para la diferencia entre κ y κ_0 bajo la hipótesis alternativa. Sea $\delta = \kappa - \kappa_0$ el mínimo valor conjeturado de la diferencia entre κ y κ_0 bajo la hipótesis alternativa, la ecuación para calcular el tamaño muestral viene dada por la expresión

$$m = \left(\frac{z_{1-\alpha/2} \sqrt{\text{Var}_0(\hat{\kappa}(c))} + z_{1-\beta} \sqrt{\text{Var}_1(\hat{\kappa}(c))}}{\delta} \right)^2, \quad (3.10)$$

donde $\text{Var}_0(\hat{\kappa}(c))$ es la varianza de $\hat{\kappa}(c)$ bajo la hipótesis nula y $\text{Var}_1(\hat{\kappa}(c))$ es la varianza bajo la hipótesis alternativa. Si se conocen los valores de $\hat{\kappa}$ bajo la hipótesis nula y bajo la hipótesis alternativa entonces las varianzas se pueden calcular mediante el mismo método utilizado anteriormente. Finalmente, el tamaño muestral se calcula utilizando la ecuación (3.10).

3.1.1.2.1. Experimentos de simulación

Para estudiar el método propuesto en la Sección anterior se han realizado unos experimentos de simulación Monte Carlo con el objetivo de calcular el tamaño muestral mínimo necesario para estimar el coeficiente kappa ponderado con una determinada

precisión y una confianza $100(1-\alpha)\%$, donde se requiere conocer la sensibilidad, especificidad y la prevalencia de la enfermedad. En la práctica estos parámetros no se conocen por lo que hay que estimarlos a partir de una muestra piloto o bien a partir de otros estudios previos, y son unos valores que evidentemente variarán de una muestra a otra. Por tanto es necesario estudiar como de robusto es el método que se propone para calcular el tamaño muestral, ya que este método se basa en valores muestrales y no en valores poblacionales. Los experimentos de simulación han consistido en generar 10000 muestras aleatorias de distribuciones multinomiales cuyas probabilidades viene dadas por las probabilidades (3.6) y (3.7). Como valores de sensibilidades y especificidades se han tomado los valores $(Se = 0.80, Sp = 0.75)$, $(Se = 0.85, Sp = 0.70)$ y $(Se = 0.90, Sp = 0.85)$, que son valores que aparecen con cierta frecuencia en la práctica clínica. Como valores de la prevalencia se han considerado los valores 10%, 30%, 50%, 70% y 90%; y como valores del índice de ponderación se han tomado $c = \{0.10, 0.30, 0.50, 0.70, 0.90\}$. Por tanto, se han generado muestras aleatorias multinomiales a partir de diferentes valores del coeficiente kappa ponderado. Como valores de la precisión d se han tomado los valores 0.025, 0.05 y 0.10, que son unos valores razonables de precisión de la estimación, y como error α se ha tomado el 5%.

Para cada valor de sensibilidad, especificidad, prevalencia, índice de ponderación y precisión se han calculado el tamaño muestral necesario para estimar el coeficiente kappa ponderado y se han generado las 10000 muestras aleatorias multinomiales con ese tamaño muestral. Con cada una de las 10000 muestras aleatorias se ha calculado el tamaño muestral a partir de los valores estimados de los parámetros, y para cada 10000 muestras aleatorias hemos calculado el tamaño muestral medio y la raíz del error cuadrático medio relativo (*RECMR*) dado por la expresión

$$RECMR = \sqrt{\sum_{i=1}^M \frac{(\hat{m}_i - m)^2}{M-1}} / m, \quad (3.11)$$

donde $M = 10000$, m es el tamaño muestral calculado a partir de los valores de los parámetros con los que se han generado las muestras y \hat{m}_i es el tamaño muestral calculado a partir de cada una de las 10000 muestras aleatorias generadas. El *RECMR* es

una medida de la calidad de un estimador y en la práctica se considera que un estimador es válido cuando su *RECMR* es inferior al 30%. En la Tablas 3.2, 3.3 y 3.4 se muestran los resultados obtenidos para los diferentes valores estudiados, de los cuales se obtiene que en todos los casos el *RECMR* es inferior al 25%, incluso en la mayoría de los casos estudiados el *RECMR* es inferior al 10% (sobretudo para $d = 0.025$). Por consiguiente, los resultados obtenidos permiten confirmar la validez del método que se propone para calcular el tamaño muestral.

Tabla 3.2. Robustez del método del cálculo del tamaño muestral cuando $Se = 0.80$ y $Sp = 0.75$.

$c = 0.1$										
p	$\kappa(c)$	$d = 0.025$			$d = 0.05$			$d = 0.10$		
		m	\bar{m}	$RECMR$	m	\bar{m}	$RECMR$	m	\bar{m}	$RECMR$
0.10	0.1949	3230	3,236	0.0610	808	809	0.1190	202	204	0.2427
0.30	0.2324	4185	4,183	0.0427	1047	1,047	0.0856	262	262	0.1706
0.50	0.2878	5592	5,589	0.0262	1398	1,398	0.0517	350	348	0.1065
0.70	0.3779	7760	7,754	0.0204	1940	1,937	0.0406	485	483	0.0804
0.90	0.5500	11984	11,982	0.0362	2996	2,987	0.0724	749	745	0.1425
$c = 0.3$										
p	$\kappa(c)$	$d = 0.025$			$d = 0.05$			$d = 0.10$		
		m	\bar{m}	$RECMR$	m	\bar{m}	$RECMR$	m	\bar{m}	$RECMR$
0.10	0.4140	4648	4,649	0.0214	1162	1,162	0.0426	291	291	0.0854
0.30	0.4512	4644	4,644	0.0144	1161	1,161	0.0292	291	289	0.0589
0.50	0.4957	4692	4,689	0.0179	1173	1,172	0.0356	294	292	0.0718
0.70	0.5500	5042	5,043	0.0254	1261	1,260	0.0511	316	315	0.1027
0.90	0.6176	6446	6,442	0.0273	1612	1,608	0.0528	403	401	0.1085
$c = 0.5$										
p	$\kappa(c)$	$d = 0.025$			$d = 0.05$			$d = 0.10$		
		m	\bar{m}	$RECMR$	m	\bar{m}	$RECMR$	m	\bar{m}	$RECMR$
0.10	0.5288	5189	5,188	0.0129	1298	1,296	0.0254	325	323	0.0533
0.30	0.5392	4529	4,526	0.0154	1133	1,131	0.0315	284	282	0.0633
0.50	0.5500	4277	4,274	0.0199	1070	1,068	0.0399	268	267	0.0800
0.70	0.5612	4549	4,549	0.0199	1138	1,136	0.0396	285	284	0.0796
0.90	0.5729	5484	5,482	0.0152	1371	1,368	0.0311	343	341	0.0640
$c = 0.7$										
p	$\kappa(c)$	$d = 0.025$			$d = 0.05$			$d = 0.10$		
		m	\bar{m}	$RECMR$	m	\bar{m}	$RECMR$	m	\bar{m}	$RECMR$
0.10	0.5863	6572	6,567	0.0206	1643	1,642	0.0418	411	408	0.0838
0.30	0.5500	5284	5,283	0.0236	1321	1,320	0.0469	331	330	0.0949
0.50	0.5179	4953	4,954	0.0202	1239	1,237	0.0398	310	309	0.0800
0.70	0.4894	5098	5,097	0.0149	1275	1,274	0.0298	319	317	0.0607
0.90	0.4639	5456	5,454	0.0170	1364	1,364	0.0342	341	340	0.0693
$c = 0.9$										
p	$\kappa(c)$	$d = 0.025$			$d = 0.05$			$d = 0.10$		
		m	\bar{m}	$RECMR$	m	\bar{m}	$RECMR$	m	\bar{m}	$RECMR$
0.10	0.5500	13114	13,121	0.0310	3279	3,274	0.0606	820	814	0.1241
0.30	0.4091	8982	8,976	0.0208	2246	2,243	0.0427	562	557	0.0836
0.50	0.3257	6981	6,980	0.0209	1746	1,742	0.0436	437	434	0.0857
0.70	0.2705	5585	5,585	0.0330	1397	1,396	0.0668	350	347	0.1311
0.90	0.2313	4548	4,541	0.0471	1137	1,139	0.0958	285	286	0.1949

p : prevalencia; c : índice de ponderación; d : precisión m : tamaño muestral; \bar{m} : tamaño muestral medio
 $RMSE$: raíz del error cuadrático medio relativo.

Tabla 3.3. Robustez del método del cálculo del tamaño muestral cuando $Se = 0.85$ y $Sp = 0.70$.

$c = 0.1$										
p	$\kappa(c)$	$d = 0.025$			$d = 0.05$			$d = 0.10$		
		m	\bar{m}	RECMR	m	\bar{m}	RECMR	m	\bar{m}	RECMR
0.10	0.1684	2342	2343	0.0741	586	590	0.1494	147	151	0.2809
0.30	0.2037	3156	3157	0.0532	789	789	0.1050	198	199	0.2137
0.50	0.2578	4446	4441	0.0329	1112	1110	0.0667	278	277	0.1344
0.70	0.3511	6593	6585	0.0213	1649	1645	0.0430	413	409	0.0861
0.90	0.5500	10615	10615	0.0417	2654	2647	0.0837	664	659	0.1653
$c = 0.3$										
p	$\kappa(c)$	$d = 0.025$			$d = 0.05$			$d = 0.10$		
		m	\bar{m}	RECMR	m	\bar{m}	RECMR	m	\bar{m}	RECMR
0.10	0.3738	3893	3893	0.0276	974	974	0.0540	244	244	0.1097
0.30	0.4185	4125	4124	0.0171	1032	1030	0.0338	258	257	0.0699
0.50	0.4753	4345	4343	0.0160	1087	1085	0.0323	272	271	0.0659
0.70	0.5500	4698	4697	0.0276	1175	1173	0.0551	294	293	0.1079
0.90	0.6525	6008	6001	0.0358	1502	1498	0.0716	376	373	0.1446
$c = 0.5$										
p	$\kappa(c)$	$d = 0.025$			$d = 0.05$			$d = 0.10$		
		m	\bar{m}	RECMR	m	\bar{m}	RECMR	m	\bar{m}	RECMR
0.10	0.4911	4761	4760	0.0141	1191	1189	0.0281	298	297	0.0578
0.30	0.5189	4387	4384	0.0123	1097	1096	0.0255	275	273	0.0513
0.50	0.5500	4191	4190	0.0197	1048	1047	0.0393	262	261	0.0791
0.70	0.5851	4418	4417	0.0249	1105	1104	0.0500	277	275	0.0980
0.90	0.6250	5525	5522	0.0221	1382	1378	0.0448	346	343	0.0926
$c = 0.7$										
p	$\kappa(c)$	$d = 0.025$			$d = 0.05$			$d = 0.10$		
		m	\bar{m}	RECMR	m	\bar{m}	RECMR	m	\bar{m}	RECMR
0.10	0.5580	6469	6468	0.0173	1618	1616	0.0341	405	403	0.0713
0.30	0.5500	5425	5422	0.0213	1357	1356	0.0423	340	339	0.0842
0.50	0.5423	5107	5109	0.0224	1277	1276	0.0458	320	319	0.0893
0.70	0.5347	5404	5401	0.0187	1351	1350	0.0377	338	336	0.0768
0.90	0.5274	6218	6215	0.0159	1555	1552	0.0317	389	387	0.0644
$c = 0.9$										
p	$\kappa(c)$	$d = 0.025$			$d = 0.05$			$d = 0.10$		
		m	\bar{m}	RECMR	m	\bar{m}	RECMR	m	\bar{m}	RECMR
0.10	0.5500	14007	14002	0.0269	3502	3495	0.0540	876	873	0.1094
0.30	0.4459	10237	10240	0.0227	2560	2555	0.0444	640	637	0.0907
0.50	0.3750	8633	8630	0.0192	2159	2155	0.0381	540	536	0.0776
0.70	0.3235	7517	7513	0.0249	1880	1879	0.0511	470	468	0.1017
0.90	0.2845	6599	6603	0.0364	1650	1649	0.0722	413	413	0.1453

p : prevalencia; c : índice de ponderación; d : precisión m : tamaño muestral; \bar{m} : tamaño muestral medio
 RMSE: raíz del error cuadrático medio relativo.

Tabla 3.4. Robustez del método del cálculo del tamaño muestral cuando $Se = 0.90$ y $Sp = 0.85$.

$c = 0.1$										
p	$\kappa(c)$	$d = 0.025$			$d = 0.05$			$d = 0.10$		
		m	\bar{m}	$RECMR$	m	\bar{m}	$RECMR$	m	\bar{m}	$RECMR$
0.10	0.3553	6316	6,320	0.0322	1579	1,579	0.0637	395	396	0.1288
0.30	0.4091	6978	6,972	0.0239	1745	1,744	0.0477	437	435	0.0959
0.50	0.4821	7439	7,434	0.0231	1860	1,860	0.0448	465	464	0.0918
0.70	0.5870	7357	7,357	0.0388	1840	1,845	0.0762	460	462	0.1513
0.90	0.7500	7003	7,013	0.0806	1751	1,746	0.1626	438	444	0.3208
$c = 0.3$										
p	$\kappa(c)$	$d = 0.025$			$d = 0.05$			$d = 0.10$		
		m	\bar{m}	$RECMR$	m	\bar{m}	$RECMR$	m	\bar{m}	$RECMR$
0.10	0.6176	4802	4,800	0.0139	1201	1,198	0.0276	301	298	0.0595
0.30	0.6563	4086	4,083	0.0223	1022	1,020	0.0446	256	254	0.0912
0.50	0.7000	3421	3,419	0.0357	856	857	0.0716	214	214	0.1411
0.70	0.7500	3057	3,057	0.0564	765	764	0.1092	192	192	0.2141
0.90	0.8077	3517	3,514	0.0674	880	878	0.1377	220	216	0.2756
$c = 0.5$										
p	$\kappa(c)$	$d = 0.025$			$d = 0.05$			$d = 0.10$		
		m	\bar{m}	$RECMR$	m	\bar{m}	$RECMR$	m	\bar{m}	$RECMR$
0.10	0.7212	3972	3,968	0.0252	993	990	0.0499	249	245	0.1026
0.30	0.7353	3125	3,121	0.0356	782	779	0.0710	196	195	0.1432
0.50	0.75	2683	2,682	0.0435	671	671	0.0872	168	168	0.1731
0.70	0.7653	2761	2,757	0.0507	691	691	0.0998	173	172	0.1988
0.90	0.7813	3501	3,500	0.0453	876	872	0.0910	219	214	0.1863
$c = 0.7$										
p	$\kappa(c)$	$d = 0.025$			$d = 0.05$			$d = 0.10$		
		m	\bar{m}	$RECMR$	m	\bar{m}	$RECMR$	m	\bar{m}	$RECMR$
0.10	0.7683	4296	4,297	0.0451	1074	1,072	0.0875	269	266	0.1826
0.30	0.7500	3413	3,414	0.0484	854	853	0.0995	214	213	0.1993
0.50	0.7326	3336	3,333	0.0442	834	836	0.0877	209	209	0.1766
0.70	0.7159	3866	3,865	0.0360	967	966	0.0722	242	241	0.1482
0.90	0.7000	4848	4,845	0.0255	1212	1,209	0.0509	303	300	0.1045
$c = 0.9$										
p	$\kappa(c)$	$d = 0.025$			$d = 0.05$			$d = 0.10$		
		m	\bar{m}	$RECMR$	m	\bar{m}	$RECMR$	m	\bar{m}	$RECMR$
0.10	0.7500	8668	8,663	0.0653	2167	2,162	0.1283	542	543	0.2630
0.30	0.6429	7974	7,979	0.0455	1994	2,002	0.0897	499	503	0.1801
0.50	0.5625	8503	8,502	0.0290	2126	2,128	0.0585	532	531	0.1180
0.70	0.5000	8884	8,885	0.0220	2221	2,218	0.0441	556	554	0.0895
0.90	0.4500	8951	8,947	0.0236	2238	2,236	0.0476	560	559	0.0944

p : prevalencia; c : índice de ponderación; d : precisión m : tamaño muestral; \bar{m} : tamaño muestral medio
 $RECMR$: raíz del error cuadrático medio relativo.

3.1.1.2.2. Aplicación

El método del cálculo del tamaño muestral para estimar el coeficiente kappa ponderado se ha aplicado al estudio de Weiner et al. (1979) sobre el diagnóstico de la enfermedad coronaria, utilizando como test diagnóstico un test de ejercicio y como gold estándar la arteriografía coronaria. En la Tabla 3.5 se muestran los resultados obtenidos por Weiner et al., (1979), y donde la variable T modeliza el resultado del test de ejercicio y la variable D modeliza el resultado de la angiografía coronaria.

Tabla 3.5. Datos del estudio de Weiner et al. (1979).

	$T = 1$	$T = 0$	Total
$D = 1$	815	208	1023
$D = 0$	115	327	442
Total	930	535	1465

En la Tabla 3.6 se muestran los valores estimados de los coeficientes kappa ponderados para distintos valores del índice de ponderación. De estos resultados se obtiene que tanto si el test de ejercicio se utiliza antes de un tratamiento intensivo ($0 < c < 0.5$) como si se utiliza como un test de screening ($0.5 < c < 1$), el coeficiente kappa ponderado tiene un valor intermedio para cada valor del índice c , lo que indica que el acuerdo clasificatorio entre el test de ejercicio y la angiografía coronaria es moderado (Landis y Koch, 1977).

Tabla 3.6. Coeficientes kappa ponderados estimados del test de ejercicio.

c	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\kappa}(c)$	0.55	0.54	0.52	0.51	0.49	0.48	0.46	0.45	0.44

Para ilustrar como se aplica el método del cálculo del tamaño muestral se ha planteado qué tamaño muestral es necesario para estimar el coeficiente kappa ponderado (por ejemplo cuando $c = 0.1$) con una precisión $d = 5\%$ y una confianza del

95%. Considerando los datos de la Tabla 3.5 como una muestra piloto, los valores de los estimadores máximo verosímiles de la sensibilidad, especificidad y prevalencia son $\hat{S}_e = 80\%$, $\hat{S}_p = 74\%$ y $\hat{p} = 70\%$. Aplicando la ecuación (3.9) se obtiene que $V(\hat{\kappa}(0.1)) = 1.08487$ y aplicando la ecuación (3.5) se obtiene que el tamaño muestral es 1668. Por tanto es necesario añadir 203 nuevos pacientes a la muestra inicial para poder estimar el coeficiente kappa ponderado ($c = 0.1$) con una precisión del 5% y una confianza del 95%.

3.1.1.2.3. Conclusión

Se ha propuesto un método para calcular el tamaño muestral mínimo cuando se pretende estimar el coeficiente kappa ponderado de un test diagnóstico binario. Dicho método requiere conocer previamente la sensibilidad y la especificidad del test diagnóstico, y la prevalencia de la enfermedad (y por tanto el coeficiente kappa ponderado del test diagnóstico). Estos parámetros se pueden estimar a partir de una muestra piloto o a partir de otros estudios similares. Por tanto, el método se basa en los estimadores de estos parámetros y no en valores poblacionales, por lo que ha sido necesario estudiar la robustez del método del cálculo del tamaño muestral. Para ello se han realizado unos experimentos de simulación cuyos resultados han demostrado la validez del método que se propone para calcular el tamaño muestral. La ecuación del tamaño muestral para estimar el coeficiente kappa ponderado (ecuación (3.5)) se basa en el intervalo de confianza tipo Wald del coeficiente kappa ponderado $\left(\kappa(c) \in \hat{\kappa}(c) \pm z_{\alpha} \sqrt{\widehat{Var}(\hat{\kappa}(c))} \right)$, y por lo tanto tiene el mismo inconveniente que tiene el intervalo de confianza de Wald. Por tanto, cuando el coeficiente kappa ponderado es cercano a 0 o a 1, el método del cálculo del tamaño muestral está afectado por el mal rendimiento de este intervalo de confianza.

3.1.2. Tamaño muestral para la comparación de dos tests binarios

En esta Sección se estudia el problema de la determinación del tamaño muestral para comparar las sensibilidades y/o las especificidades de dos tests diagnósticos binarios, tanto si los dos tests diagnósticos se van a comparar mediante un diseño con muestras

independientes como si se van a comparar con un diseño con muestras apareadas. Este problema ha sido estudiado por diversos autores (Newcombe, 1998a, b, c y 2001a, b; Alonzo et al., 2002; Martín Andrés y Luna del Castillo, 2004).

3.1.2.1. Tamaño muestral para comparar las sensibilidades y/o especificidades con dos muestras independientes

Sean Se_1 y Sp_1 la sensibilidad y la especificidad del test binario 1, y Se_2 y Sp_2 la sensibilidad y la especificidad del test binario 2. El problema que se plantea es determinar el tamaño muestral para que el test de hipótesis

$$H_0 : Se_1 = Se_2$$

$$H_1 : Se_1 \neq Se_2$$

al error α detecte una diferencia $\delta = |Se_1 - Se_2|$ entre las dos sensibilidades con una potencia $100(1 - \beta)\%$, cuando cada test diagnóstico se aplica a una muestra (siendo ambas muestras independientes). Asumiendo la normalidad de $\hat{Se}_1 - \hat{Se}_2$, que el diseño es óptimo ($m = m_1 = m_2$) y que $\delta = Se_1 - Se_2$, el tamaño muestral para comparar las dos sensibilidades viene dado por la ecuación (Martín Andrés y Luna del Castillo, 2004)

$$m = \left(\frac{z_{1-\alpha/2} \sqrt{(Se_1 + Se_2) \frac{(1 - Se_1) + (1 - Se_2)}{2}} + z_{1-\beta} \sqrt{Se_1(1 - Se_1) + Se_2(1 - Se_2)}}{\delta} \right)^2. \quad (3.12)$$

Por tanto, es necesario tomar dos muestras de tamaño m y aplicar el test binario 1 a una de muestra y el test binario 2 a la otra muestra.

La ecuación (3.12) requiere conocer las sensibilidades de cada test binario. Cuando no se tiene ningún conocimiento acerca de las sensibilidades, se sustituye el valor $Se_1(1 - Se_1) + Se_2(1 - Se_2)$ por su valor máximo bajo $\delta = |Se_1 - Se_2|$, obteniéndose dicho máximo cuando las dos sensibilidades valen $0.5 \pm \delta/2$ y siendo $Máx\{Se_1(1 - Se_1) + Se_2(1 - Se_2)\} = (1 - \delta^2)/2$. Finalmente el tamaño muestral es

$$m = \frac{1}{2} \left(\frac{z_{\alpha} + z_{2\beta} \sqrt{1-\delta}}{\delta} \right)^2. \quad (3.13)$$

Cuando se dispone de alguna información acerca de las dos sensibilidades, se sustituyen en la ecuación (3.12) las sensibilidades por los valores que sean lo más cercanos a $0.5 \pm \delta/2$, compatibles con la información disponible y con δ .

Una vez calculado el tamaño muestral (sin información o con ella), se aplica una corrección por continuidad, siendo el tamaño muestral final

$$m' = \frac{m}{4} \left(1 + \sqrt{1 + \frac{4}{m\delta}} \right)^2. \quad (3.14)$$

La expresión (3.12) se ha construido asumiendo que los tamaños de cada muestra son iguales. Cuando el costo de cada observación es diferente, el tamaño de muestra también debe ser diferente. De esta forma, basándose en un diseño óptimo de $n_2 = t \times n_1$, y despejando n_1 , se obtiene que

$$n_1 = \frac{1}{t} \left(\frac{z_{1-\alpha/2} \sqrt{(t+1) Se(1-Se)} + z_{1-\beta} \sqrt{Se_1(1-Se_1) + Se_2(1-Se_2)}}{\delta} \right)^2, \quad (3.15)$$

con $Se = \frac{Se_1 + tSe_2}{1+t}$ y $t = \sqrt{\frac{c_1}{c_2}}$, y siendo c_i el costo de cada unidad de cada muestra. De

forma análoga a lo comentado anteriormente, el valor máximo de esta expresión cuando no se tiene información acerca de las sensibilidades es

$$n_1 = \frac{t+1}{4t\delta^2} \left\{ z_{1-\alpha/2} + z_{1-\beta} \sqrt{1 - \frac{4(t^3+1)}{(t+1)^3} \delta^2} \right\}^2. \quad (3.16)$$

El tamaño muestral para comparar las dos especificidades se calcula de forma similar sustituyendo Se por Sp .

3.1.2.2. Tamaño muestral para comparar las sensibilidades y/o especificidades con dos muestras apareadas

Cuando dos tests binarios y el gold estándar se aplican a todos los individuos de una misma muestra aleatoria se obtiene la Tabla de probabilidades 3.7.

Tabla 3.7. Probabilidades teóricas al aplicar dos tests binarios y el gold estándar a una muestra.

	$T_1 = 1$		$T_1 = 0$		
	$T_2 = 1$	$T_2 = 0$	$T_2 = 1$	$T_2 = 0$	
$D = 1$	p_{11}	p_{10}	p_{01}	p_{00}	p
$D = 0$	q_{11}	q_{10}	q_{01}	q_{00}	q
	$p_{11} + q_{11}$	$p_{10} + q_{10}$	$p_{01} + q_{01}$	$p_{00} + q_{00}$	1

En términos de estas probabilidades, la sensibilidad y la especificidad del test 1 se escriben como

$$Se_1 = \frac{p_{11} + p_{10}}{p} \quad \text{y} \quad Sp_1 = \frac{q_{01} + q_{00}}{q}, \quad (3.17)$$

y las del test 2 como

$$Se_2 = \frac{p_{11} + p_{01}}{p} \quad \text{y} \quad Sp_2 = \frac{q_{10} + q_{00}}{q}. \quad (3.18)$$

El test de hipótesis de igualdad de las dos sensibilidades

$$H_0 : Se_1 = Se_2$$

$$H_1 : Se_1 \neq Se_2$$

es equivalente al test de hipótesis

$$\begin{aligned} H_0 &: p_{10} = p_{01} \\ H_1 &: p_{10} \neq p_{01}, \end{aligned}$$

y el tamaño muestral para que este test de hipótesis al error α detecte una diferencia $\delta = |Se_1 - Se_1| = |p_{01} - p_{10}|$ con una potencia $100(1-\beta)\%$ es (Martín Andrés y Luna del Castillo, 2004)

$$m = \left(\frac{z_{1-\alpha/2} \sqrt{p_{10} + p_{01}} + z_{1-\beta} \sqrt{p_{10} + p_{01} - \delta^2}}{\delta} \right)^2. \quad (3.19)$$

En la práctica, las probabilidades p_{ij} y q_{ij} son desconocidas. Por tanto, si no se dispone de ninguna información acerca de $p_{10} + p_{01}$, este se sustituye por su valor máximo, que se obtiene cuando p_{10} y p_{01} son $0.5 \pm \delta/2$, siendo entonces el tamaño muestral

$$m = \left(\frac{z_{1-\alpha/2} + z_{1-\beta} \sqrt{1 - \delta^2}}{\delta} \right)^2. \quad (3.20)$$

Cuando se dispone de alguna información sobre las probabilidades p_{ij} y q_{ij} (a partir de una muestra piloto o un estudio similar), como valores de p_{10} y p_{01} se toman los valores más cercanos a $0.5 \pm \delta/2$ y que sean compatibles con la información y con δ .

Una vez calculado el tamaño muestral (sin información o con ella), se aplica una corrección por continuidad, siendo el tamaño muestral final similar al dado en la ecuación (3.14).

De forma análoga se calcula el tamaño muestral para comparar las sensibilidades de dos tests binarios con muestras apareadas, sustituyendo Se por Sp , p_{10} por q_{01} y p_{01} por q_{10} .

3.2.- Verificación parcial

En presencia de verificación parcial, la determinación del tamaño muestral tanto para evaluar la exactitud de un test binario como para comparar la exactitud de dos tests binarios no se puede realizar aplicando los métodos analizados en la Sección anterior, ya que en esta situación la sensibilidad y especificidad no son proporciones binomiales. En las siguientes Secciones se proponen unos métodos para calcular el tamaño muestral cuando se evalúa la sensibilidad y la especificidad de un test diagnóstico binario y cuando se comparan las sensibilidades y las especificidades de dos tests diagnósticos binarios en presencia de verificación parcial de la enfermedad.

3.2.1. Tamaño muestral para estimar la sensibilidad y la especificidad de un test binario en presencia de verificación parcial de la enfermedad (Aportación)

La ecuación para calcular el tamaño muestral para estimar la sensibilidad o la especificidad de un test binario mediante un intervalo de confianza es

$$m = \frac{z_{1-\alpha/2}^2}{d^2} \text{Var}(\hat{\theta}), \quad (3.21)$$

donde $z_{1-\alpha/2}$ es el $100(1-\alpha/2)$ -ésimo percentil de la distribución normal estándar, d es la precisión de la estimación y $\text{Var}(\hat{\theta})$ es la función varianza (McCullagh y Nelder, 1989) de $\hat{\theta}$, donde θ es la sensibilidad o la especificidad. La ecuación (3.21), presenta el problema de estimar la función varianza de $\hat{\theta}$. En presencia de verificación parcial de la enfermedad, el estado de la enfermedad es desconocido para algunos pacientes, por lo que la evaluación del test binario no se puede realizar sólo con los pacientes verificados con el gold estándar (Begg y Greenes, 1983; Zhou, 1993), y por consiguiente no es posible calcular el tamaño muestral para estimar la exactitud del test diagnóstico utilizando los métodos tradicionales en tablas 2×2 estudiados en la Sección 3.1.1.1. Para resolver este problema Roldán Nofuentes, Luna del Castillo y Montero Alonso (2009a) han propuesto el siguiente método.

Sea un test diagnóstico binario que se aplica a todos los pacientes de una muestra aleatoria de tamaño n . Sean las variables aleatorias T y D que modelizan el resultado del test el verdadero estado de la enfermedad, respectivamente, tal y como se han definido anteriormente. La variable V modeliza el proceso de verificación de la enfermedad, siendo $V = 1$ cuando el individuo ha sido verificado con el gold estándar y $V = 0$ cuando no ha sido verificado. Cuando el test diagnóstico se aplica a todos los pacientes de la muestra de tamaño n y el gold estándar sólo a una parte de ellos, se obtiene la Tabla 3.8.

Tabla 3.8. Frecuencias observadas en presencia de verificación parcial de la enfermedad.

	$T = 1$	$T = 0$	Total
$V = 1$			
$D = 1$	s_1	s_0	s
$D = 0$	r_1	r_0	r
$V = 0$	u_1	u_0	u
Total	n_1	n_0	n

Sean Se , Sp y p la sensibilidad, la especificidad y la prevalencia de la enfermedad, respectivamente. Sea $\lambda_{ij} = P(V = 1 | D = i, T = j)$ la probabilidad de verificar a un paciente con el gold estándar cuando $D = i$ y $T = j$, con $i, j = 0, 1$. Si se verifica que $\lambda_{ij} = P(V = 1 | D = i, T = j) = P(V = 1 | T = j) = \lambda_j$, entonces el proceso de verificación sólo depende de los resultados del test diagnóstico y el proceso de verificación es MAR (Rubin, 1976). Bajo esta suposición, las probabilidades de cada celda de la Tabla 3.8 vienen dadas por las expresiones

$$\xi_j = P(V = 1, D = 1, T = j) = p\lambda_j Se^j (1 - Se)^{1-j},$$

$$\psi_j = P(V = 1, D = 0, T = j) = (1 - p)\lambda_j Sp^{1-j} (1 - Sp)^j, \quad (3.22)$$

$$\zeta_j = P(V = 0, T = j) = (1 - \lambda_j) \left\{ pSe^j (1 - Se)^{1-j} + (1 - p)Sp^{1-j} (1 - Sp)^j \right\},$$

con $j = 0, 1$. Sea $\boldsymbol{\omega} = (\xi_1, \xi_0, \psi_1, \psi_0, \zeta_1, \zeta_0)^T$ un vector de dimensión 6. Como $\boldsymbol{\omega}$ es el vector de probabilidades de una distribución multinomial, la matriz de varianzas-covarianzas de $\hat{\boldsymbol{\omega}}$ viene dada por la expresión

$$\Sigma_{\hat{\boldsymbol{\omega}}} = \{diag(\boldsymbol{\omega}) - \boldsymbol{\omega}\boldsymbol{\omega}^T\} / n. \quad (3.23)$$

La sensibilidad y la especificidad del test diagnóstico se escriben en términos de las probabilidades ξ_j , ψ_j y ζ_j como

$$Se = \frac{\xi_1 \phi_1}{\xi_1 + \psi_1} / \sum_{i=0}^1 \frac{\xi_i \phi_i}{\xi_i + \psi_i} \quad (3.24)$$

y

$$Sp = \frac{\xi_0 \phi_0}{\xi_0 + \psi_0} / \sum_{i=0}^1 \frac{\xi_i \phi_i}{\xi_i + \psi_i}, \quad (3.25)$$

con $\phi_j = \xi_j + \psi_j + \zeta_j$, y aplicando del método delta (Agresti, 2002) se obtiene que

$$Var(\hat{Se}) = \left(\frac{\partial Se}{\partial \boldsymbol{\omega}} \right) \Sigma_{\hat{\boldsymbol{\omega}}} \left(\frac{\partial Se}{\partial \boldsymbol{\omega}} \right)^T \quad (3.26)$$

y

$$Var(\hat{Sp}) = \left(\frac{\partial Sp}{\partial \boldsymbol{\omega}} \right) \Sigma_{\hat{\boldsymbol{\omega}}} \left(\frac{\partial Sp}{\partial \boldsymbol{\omega}} \right)^T, \quad (3.27)$$

donde las derivadas parciales $\partial Se / \partial \boldsymbol{\omega}$ y $\partial Sp / \partial \boldsymbol{\omega}$ se calculan de la forma rutinaria. Realizando operaciones algebraicas se obtienen las expresiones

$$Var(\hat{Se}) = \frac{Se(1-Se) \left[q(SeSp - I)(pI - Sp)\lambda_0 - Se \{ qSp(pI - Sp + 1) + p(1-Se)\lambda_0 \} \lambda_1 \right]}{p(pI - Sp)(pI - Sp + 1)\lambda_0 \lambda_1} \quad (3.28)$$

y

$$Var(\hat{Sp}) = \frac{Sp(1-Sp) \left[pSeSp(Sp-pI)\lambda_0 - (1-Sp) \{ p(Se-1)(pI-Sp+1) - qSp\lambda_0 \} \lambda_1 \right]}{p(pI-Sp)(pI-Sp+1)\lambda_0\lambda_1}, \quad (3.29)$$

donde $q = 1 - p$ y $I = Se + Sp - 1$ es el índice de Youden del test diagnóstico. Como las probabilidades dadas en (3.22) son las probabilidades de una distribución multinomial, sus estimadores máximo verosímiles vienen dados por las expresiones

$$\hat{\xi}_j = \frac{s_j}{n}, \quad \hat{\psi}_j = \frac{r_j}{n} \quad \text{y} \quad \hat{\zeta}_j = \frac{u_j}{n}. \quad (3.30)$$

Sustituyendo en las ecuaciones (3.24) y (3.25) los parámetros por sus estimadores máximo verosímiles, se obtienen los estimadores de la sensibilidad y especificidad, que son los mismos que los obtenidos por Begg y Greenes (1983) y Zhou (1993), y que vienen dados por las expresiones

$$\hat{Se} = \frac{n_1 s_1 / (s_1 + r_1)}{n_1 s_1 / (s_1 + r_1) + n_0 s_0 / (s_0 + r_0)} \quad (3.31)$$

y

$$\hat{Sp} = \frac{n_0 r_0 / (s_0 + r_0)}{n_1 r_1 / (s_1 + r_1) + n_0 r_0 / (s_0 + r_0)}. \quad (3.32)$$

Una vez calculadas las funciones varianzas de la sensibilidad y especificidad, el tamaño muestral se calcula fácilmente aplicando la ecuación (3.21).

Este método del cálculo del tamaño muestral también se puede utilizar para calcular el tamaño muestral necesario para contrastar

$$\begin{aligned} H_0 : \theta &= \theta_0 \\ H_1 : \theta &\neq \theta_0 \end{aligned}$$

al error α y a la potencia $100(1 - \beta)\%$, donde θ es la sensibilidad o la especificidad. Por tanto, para calcular el tamaño de la muestra, es necesario especificar un valor para la diferencia entre θ y θ_0 bajo la hipótesis alternativa. Si $\delta = \theta - \theta_0$ es el valor

asignado para la diferencia entre θ y θ_0 bajo la hipótesis alternativa, la ecuación para calcular el tamaño muestral viene dado por la expresión

$$m = \left(\frac{z_{1-\alpha/2} \sqrt{\text{Var}_0(\hat{\theta})} + z_{1-\beta} \sqrt{\text{Var}_1(\hat{\theta})}}{\delta} \right)^2, \quad (3.33)$$

donde z_γ es el 100γ -ésimo percentil de la distribución normal estándar, $\text{Var}_0(\hat{\theta})$ es la función varianza de $\hat{\theta}$ bajo la hipótesis nula y $\text{Var}_1(\hat{\theta})$ es la función varianza de $\hat{\theta}$ bajo la hipótesis alternativa. Si se conocen los valores de $\hat{\theta}$ bajo la hipótesis nula y bajo la hipótesis alternativa, la prevalencia de la enfermedad y las probabilidades de verificación λ_1 y λ_0 , las funciones varianza se pueden calcular aplicando el mismo método utilizado anteriormente. Finalmente, el tamaño de muestra se calcula aplicando la ecuación (3.33).

3.2.1.1. Experimentos de simulación

El método anterior para calcular el tamaño muestral requiere conocer la sensibilidad y la especificidad del test diagnóstico, la prevalencia de la enfermedad y las probabilidades de verificación. La sensibilidad, la especificidad y la prevalencia se pueden a partir de una muestra piloto u otro estudio similar, y las probabilidades de verificación se pueden estimar también a partir de la muestra piloto o pueden establecerse previamente en función de unos costos de verificación. Una vez estimados los parámetros para calcular el tamaño de la muestra, se asume que los estimadores son los verdaderos valores de los parámetros. Por tanto, es necesario estudiar la robustez del método para calcular el tamaño de la muestra. Para este propósito se han realizado unos experimentos de simulación que han consistido en generar 5000 muestras aleatorias de distribuciones multinomiales cuyas probabilidades han sido calculadas a partir de las ecuaciones (3.22). Para la sensibilidad y especificidad se han tomado los valores ($Se = 0.80, Sp = 0.70$) y ($Se = 0.85, Sp = 0.75$), que son valores que aparecen con una cierta frecuencia en la

práctica clínica; como valores para la prevalencia se han tomado 10%, 30%, 50%, 70% y 90%. En la práctica clínica, la probabilidad de verificar a un paciente con un test positivo (λ_1) es mayor que la probabilidad de verificar a un paciente con un test negativo (λ_0), por tanto, para las probabilidades de verificación se han tomado los valores siguientes: (a) $\lambda_1 = 0.80$ y $\lambda_0 = 0.10$, (b) $\lambda_1 = 0.80$ y $\lambda_0 = 0.30$, (c) $\lambda_1 = 0.95$ y $\lambda_0 = 0.10$, (d) $\lambda_1 = 0.95$ y $\lambda_0 = 0.30$ y (e) $\lambda_1 = \lambda_0 = 1$. El caso (a) puede considerarse como un escenario con baja verificación, los casos (b) y (c) pueden considerarse como un escenario de verificación intermedia, el caso (d) es un escenario con altas probabilidades de verificación, y en el caso (e) todos los pacientes son verificados con el gold estándar y no hay sesgo de verificación. Como precisión d se ha tomado 0.05 y $\alpha = 5\%$. Para cada una de las 5000 muestras aleatorias se ha calculado el tamaño muestral, considerando a éstas como muestra piloto, aplicando el método propuesto anteriormente, excepto cuando $\lambda_1 = 1$, donde el tamaño de la muestra ha sido calculado aplicando tomando $Var(\hat{\theta}) = \theta(1-\theta)$ (Arkin y Wachtel, 1990). Para cada una de las 5000 muestras (en presencia de verificación parcial de la enfermedad) se ha calculado el tamaño muestral promedio y la raíz del error cuadrático medio relativo (RECMR), que viene dado por la expresión

$$RECMR = \sqrt{\frac{\sum_{i=1}^{5000} (\hat{m}_i - m)^2}{4999}} / m, \quad (3.34)$$

donde m es el tamaño de muestra calculado a partir de los valores de los parámetros con los que se han generado las muestras, \hat{m}_i es el valor estimado del tamaño de muestra calculado a partir de los estimadores máximo verosímiles de la sensibilidad y especificidad dados por las ecuaciones (3.31) y (3.32), y de prevalencia de la enfermedad y probabilidades de verificación dados por las ecuaciones (Begg y Greenes, 1983, Zhou, 1993)

$$\hat{p} = \frac{n_1 s_1 / (s_1 + r_1) + n_0 s_0 / (s_0 + r_0)}{n} \quad \text{y} \quad \hat{\lambda}_i = \frac{s_i + r_i}{n_i}. \quad (3.35)$$

El RECMR es una medida de la calidad de un estimador y aunque no hay ningún valor concreto del RECMR a partir del cual el estimador es válido, se ha considerado, como ocurre en muchas situaciones prácticas, que un valor de RECMR inferior a un 30% indican que el estimador tiene una calidad buena, mientras que un valor superior indica que el estimador no es válido.

En las Tablas 3.9 y 3.10 se muestran los resultados obtenidos cuando la sensibilidad y la especificidad toman como valores ($Se = 0.80$, $Sp = 0.70$) y ($Se = 0.85$, $Sp = 0.75$), y para los que se obtienen las conclusiones siguientes. El tamaño de muestra para estimar la sensibilidad disminuye cuando aumenta la prevalencia de la enfermedad, y el tamaño de muestra para estimar la especificidad disminuye cuando disminuye la prevalencia. Con respecto al efecto de las probabilidades de verificación, el tamaño de muestra para estimar la sensibilidad disminuye cuando aumentan las probabilidades de verificación, siendo el efecto de λ_0 mucho más importante que el de λ_1 , para valores de λ_1 suficientemente grandes e independientemente de la prevalencia de la enfermedad. De forma similar, el tamaño de muestra para estimar la especificidad disminuye cuando aumentan las probabilidades de verificación, aunque en este caso el efecto de λ_0 es de menor importancia que en el caso de sensibilidad. Por consiguiente, cuando las probabilidades de verificar a un paciente con un test positivo son relativamente altas (como normalmente ocurre en la práctica clínica), para que el tamaño de muestra estimado sea más pequeño es necesario aumentar la probabilidad de verificación de los pacientes con un test negativo, y no es necesario aumentar las probabilidades de verificación de los pacientes con test positivo.

Con respecto a la verificación total, la verificación parcial de la enfermedad requiere un aumento importante del tamaño muestral para estimar la sensibilidad, sobre todo cuando la prevalencia no es grande, mientras que para la estimación de la especificidad requiere un aumento del tamaño de muestra cuando la prevalencia es alta. Con respecto a la raíz del error cuadrático medio relativo, este es generalmente menor que el 30%, por lo que los resultados obtenidos confirman la calidad del método que se propone para calcular el tamaño de la muestra cuando se evalúa la exactitud de un test diagnóstico binario en presencia de verificación parcial de la enfermedad.

En algunos casos, el tamaño muestral obtenido aplicando el método desarrollado anteriormente es muy grande, sobre todo cuando la prevalencia de la enfermedad y las probabilidades de verificación son pequeñas. En esta situación, se recomienda tomar unas probabilidades de verificación grandes para que el tamaño de muestra estimado sea más bajo que cuando las probabilidades de verificación son bajas. Así, por ejemplo, en la Tabla 3.9 cuando $p = 10\%$, $\lambda_1 = 0.80$ y $\lambda_0 = 0.10$, el tamaño muestral es muy grande ($m = 19711$), el cual hace difícil llevar a cabo el estudio, por lo que incrementando las probabilidades de verificación ($\lambda_1 = 0.95$, $\lambda_0 = 0.30$) el tamaño muestral disminuye a 6927 pacientes.

Tabla 3.9. Tamaño de muestra para estimar la sensibilidad y la especificidad de un test binario.

$Se = 0.80 \quad Sp = 0.70$									
Tamaño de muestra para la sensibilidad									
$\lambda_1 = 0.80 \quad \lambda_0 = 0.10$					$\lambda_1 = 0.80 \quad \lambda_0 = 0.30$			$\lambda_1 = \lambda_0 = 1$	
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m
0.10	19711	34.5	19527	0.05	7002	47.5	6957	0.07	246
0.30	6096	41.5	6040	0.04	2202	52.5	2185	0.07	246
0.50	3253	48.5	3218	0.04	1213	57.5	1204	0.06	246
0.70	1871	55.5	1850	0.06	748	62.5	742	0.06	246
0.90	825	62.5	809	0.11	417	67.5	414	0.09	246
Tamaño de muestra para la especificidad									
$\lambda_1 = 0.95 \quad \lambda_0 = 0.10$					$\lambda_1 = 0.95 \quad \lambda_0 = 0.30$			$\lambda_1 = \lambda_0 = 1$	
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m
0.10	19637	39.8	19471	0.05	6927	52.7	6878	0.07	246
0.30	6081	48.3	6017	0.04	2187	59.2	2172	0.06	246
0.50	3247	56.8	3214	0.04	1208	65.7	1198	0.06	246
0.70	1870	65.3	1845	0.06	746	72.2	739	0.06	246
0.90	825	73.8	814	0.20	417	78.7	413	0.09	246
Tamaño de muestra para la especificidad									
$\lambda_1 = 0.80 \quad \lambda_0 = 0.10$					$\lambda_1 = 0.80 \quad \lambda_0 = 0.30$			$\lambda_1 = \lambda_0 = 1$	
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m
0.10	403	34.5	443	0.18	381	47.5	382	0.06	323
0.30	640	41.5	658	0.21	540	52.5	541	0.08	323
0.50	1115	48.5	1141	0.23	828	57.5	831	0.10	323
0.70	2400	55.5	2463	0.22	1540	62.5	1552	0.11	323
0.90	10043	62.5	10257	0.20	5396	67.5	5435	0.12	323
Tamaño de muestra para la especificidad									
$\lambda_1 = 0.95 \quad \lambda_0 = 0.10$					$\lambda_1 = 0.95 \quad \lambda_0 = 0.30$			$\lambda_1 = \lambda_0 = 1$	
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m
0.10	392	39.8	435	0.20	370	52.7	372	0.06	323
0.30	606	48.3	626	0.23	506	59.2	506	0.09	323
0.50	1050	56.8	1077	0.24	764	65.7	769	0.11	323
0.70	2272	65.3	2336	0.24	1411	72.2	1422	0.12	323
0.90	9615	73.8	9884	0.21	4968	78.7	5024	0.14	323

p : prevalencia; m : tamaño muestral; %V: porcentaje medio de pacientes verificados; \bar{m} : tamaño muestral medio; RECMR: raíz del error cuadrático medio relativo.

Tabla 3.10. Tamaño de muestra para estimar la sensibilidad y la especificidad de un test binario.

$Se = 0.85 \quad Sp = 0.75$									
Tamaño de muestra para la sensibilidad									
$\lambda_1 = 0.80 \quad \lambda_0 = 0.10$					$\lambda_1 = 0.80 \quad \lambda_0 = 0.30$			$\lambda_1 = \lambda_0 = 1$	
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m
0.10	16675	31.7	16409	0.09	5814	45.5	5755	0.11	196
0.30	5265	40.1	5198	0.08	1857	51.5	1837	0.11	196
0.50	2894	48.5	2851	0.07	1043	57.5	1031	0.10	196
0.70	1741	56.9	1706	0.06	660	63.5	651	0.08	196
0.90	813	65.3	789	0.16	373	69.5	366	0.10	196
$\lambda_1 = 0.95 \quad \lambda_0 = 0.10$					$\lambda_1 = 0.95 \quad \lambda_0 = 0.30$			$\lambda_1 = \lambda_0 = 1$	
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m
0.10	16633	36.4	16406	0.09	5772	50.1	5709	0.11	196
0.30	5257	46.6	5184	0.08	1849	57.9	1830	0.10	196
0.50	2891	56.8	2849	0.07	1041	65.7	1025	0.10	196
0.70	1741	67.0	1705	0.06	659	73.5	650	0.08	196
0.90	813	77.2	786	0.16	372	81.3	366	0.10	196
Tamaño de muestra para la especificidad									
$\lambda_1 = 0.80 \quad \lambda_0 = 0.10$					$\lambda_1 = 0.80 \quad \lambda_0 = 0.30$			$\lambda_1 = \lambda_0 = 1$	
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m
0.10	353	31.7	394	0.19	341	45.5	344	0.07	289
0.30	531	40.1	551	0.19	477	51.5	477	0.09	289
0.50	876	48.5	903	0.24	716	57.5	718	0.11	289
0.70	1808	56.9	1872	0.25	1299	63.5	1306	0.12	289
0.90	7572	65.3	7804	0.24	4485	69.5	4526	0.14	289
$\lambda_1 = 0.95 \quad \lambda_0 = 0.10$					$\lambda_1 = 0.95 \quad \lambda_0 = 0.30$			$\lambda_1 = \lambda_0 = 1$	
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m
0.10	340	36.4	382	0.20	328	50.1	331	0.08	289
0.30	495	46.6	522	0.22	441	57.9	442	0.10	289
0.50	810	56.8	847	0.27	650	65.7	654	0.12	289
0.70	1682	67.0	1750	0.28	1173	73.5	1185	0.14	289
0.90	7159	77.2	7400	0.26	4072	81.3	4117	0.15	289

p : prevalencia; m : tamaño muestral; %V: porcentaje medio de pacientes verificados; \bar{m} : tamaño muestral medio; RECMR: raíz del error cuadrático medio relativo.

En las Figuras 3.1 y 3.2 se muestra al efecto que la probabilidad λ_0 tiene en el tamaño muestral para estimar la sensibilidad y especificidad, respectivamente, para un mismo valor $\lambda_1 = 0.80$, valor que se puede considerar como un valor intermedio de probabilidad de verificación de los pacientes con test positivo. En la Figura 3.1 puede

observarse cómo, para la misma prevalencia, el aumento en λ_0 tiene un efecto importante en el tamaño muestral para estimar la sensibilidad (generalmente cuando la prevalencia de la enfermedad no es muy alta). La probabilidad de verificar a un paciente con un test negativo (λ_0) tiene un efecto mayor en el cálculo del tamaño muestral que la probabilidad de verificar a un paciente con un test positivo (λ_1). Un aumento en λ_0 , cuando λ_1 permanece constante, implica una reducción importante en el tamaño muestral necesario para estimar la sensibilidad de un test diagnóstico, sobre todo cuando la prevalencia de la enfermedad no es muy alta. No obstante, como puede observarse en Figura 3.2, el efecto que λ_0 tiene en el tamaño muestral para estimar la especificidad es menos importante que en el caso de la sensibilidad. Por consiguiente, al intentar calcular el tamaño de la muestra para evaluar la exactitud (sensibilidad o especificidad) de un test diagnóstico binario, los valores altos λ_0 implican un tamaño de muestra menor que con valores bajos (para un mismo valor de λ_1).

Figura 3.1. Efecto de las probabilidades de verificación en el tamaño muestral para estimar la sensibilidad.

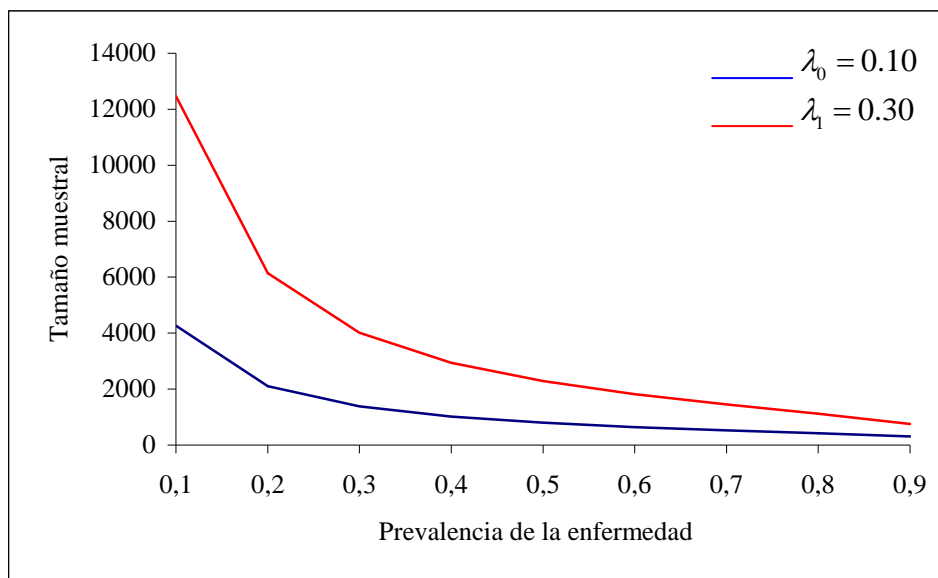
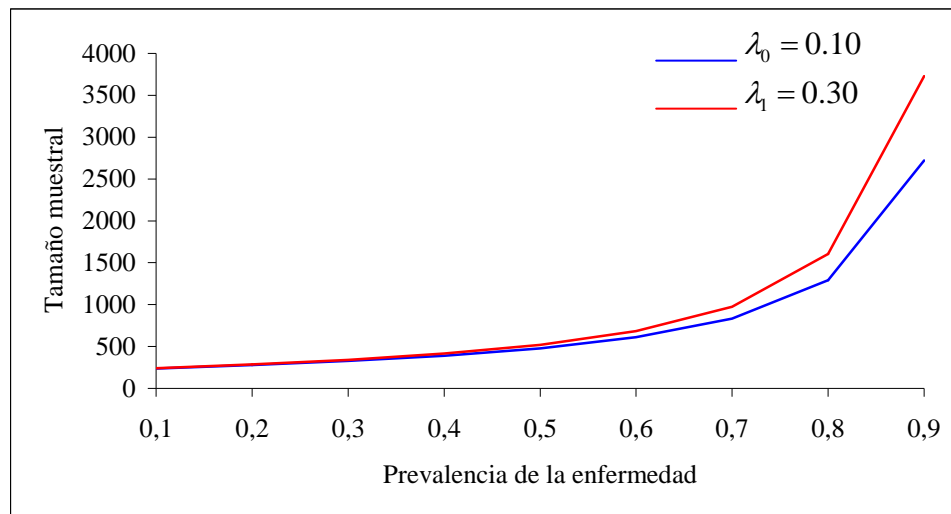


Figura 3.2. Efecto de las probabilidades de verificación en el tamaño muestral para estimar la especificidad.



3.2.1.2. Aplicación

El método del cálculo del tamaño muestral se ha aplicado al diagnóstico de la estenosis coronaria. La estenosis coronaria es una enfermedad que consiste en la obstrucción de la arteria coronaria, y su diagnóstico se puede realizar mediante una ecocardiografía con dobutamina y como gold estándar se utiliza la angiografía coronaria. Como la angiografía puede causar reacciones en los pacientes (infarto, infecciones, trombosis,...), no todos los pacientes son verificados. En la Tabla 3.11 se muestran los resultados obtenidos al aplicar la ecocardiografía con dobutamina a una muestra de 655 personas. La variable aleatoria T_1 modeliza el resultado de la ecocardiografía con dobutamina y la variable aleatoria D modeliza el resultado de la angiografía coronaria. En todo el estudio la selección de los pacientes para verificar el estado de enfermedad con la angiografía se realizó solamente en función de los resultados de la ecocardiografía con dobutamina, por lo que el proceso de verificación es MAR. A partir de los datos de la Tabla 3.11, la sensibilidad y la especificidad estimadas de la ecocardiografía con dobutamina son 0.86 y 0.78, y los correspondientes intervalos de confianza al 95% para la sensibilidad y especificidad son $Se \in (0.77, 0.95)$ y $Sp \in (0.74, 0.82)$, siendo sus respectivas precisiones el 9% y el 4%. La prevalencia de la enfermedad estimada es el 23% y las estimaciones de las probabilidades de verificación son $\hat{\lambda}_1 = 0.83$ y $\hat{\lambda}_0 = 0.32$.

Utilizando los resultados obtenidos, se plantea la cuestión de calcular el tamaño muestral necesario para estimar la sensibilidad de la ecocardiografía con dobutamina con una precisión del 5% ($d = 0.05$) a la confianza del 95%. Aplicando el método propuesto anteriormente se obtiene que $Var(\hat{Se}) = 1.4384$ y aplicando la ecuación (3.21) se obtiene que $m = 2211$. Por tanto, se necesita una muestra de al menos 2211 pacientes para estimar la sensibilidad con una precisión del 5% a la confianza del 95%, y por consiguiente es necesario agregar a la muestra inicial de 665 pacientes un mínimo de 1546 pacientes para estimar la sensibilidad con la precisión y confianza deseadas.

Tabla 3.11. Evaluación de la ecocardiografía con dobutamina.

	$T_1 = 1$	$T_1 = 0$	Total
$V = 1$			
$D = 1$	109	7	116
$D = 0$	90	126	216
$V = 0$	40	283	323
Total	239	416	655

3.2.2. Tamaño de muestra para la comparación de dos tests binarios en presencia de verificación parcial (Aportación)

El problema del cálculo del tamaño muestral en presencia de verificación parcial de la enfermedad también surge cuando se compara la exactitud de dos tests binarios. Al igual que ocurre con un único test diagnóstico, cuando se compara la exactitud de dos tests binarios en presencia de verificación parcial de la enfermedad, el cálculo del tamaño muestral no se puede realizar aplicando los métodos analizados en la Sección 3.1.2.2, ya que la sensibilidad y la especificidad de cada test binario no son proporciones binomiales. En presencia de verificación parcial de la enfermedad, Roldán Nofuentes, Luna del Castillo y Montero Alonso (2009a) han propuesto el siguiente método para calcular el tamaño muestral para comparar la exactitud de dos tests binarios.

Sean dos tests diagnósticos binarios que se aplican de forma independiente a una misma muestra aleatoria de n pacientes. Sean T_1 y T_2 dos variables aleatorias que modelizan los resultados de los tests diagnósticos 1 y 2 respectivamente, de tal forma que $T_h = 1$ cuando el resultado del h -ésimo test es positivo y $T_h = 0$ cuando el resultado del h -ésimo test es negativo. La variable D modeliza el verdadero estado de la enfermedad, siendo $D = 1$ cuando el individuo está enfermo y $D = 0$ cuando está sano. La variable V modeliza el estado de la verificación de la enfermedad, de tal forma que $V = 1$ cuando el individuo ha sido verificado y $V = 0$ cuando no ha sido verificado. Sean Se_h y Sp_h la sensibilidad y la especificidad del h -ésimo test diagnóstico ($h = 1, 2$), $p = P(D = 1)$ la prevalencia de la enfermedad, y $\lambda_{ijk} = P(V = 1 | T_1 = i, T_2 = j, D = k)$ la probabilidad de verificar un paciente con resultados $T_1 = i$, $T_2 = j$ y $D = k$, con $i, j, k = 0, 1$. La aplicación de los dos tests diagnósticos a todos los pacientes de la muestra aleatoria de tamaño n y la aplicación del gold estándar a una parte de la muestra da lugar a la Tabla 3.12.

Tabla 3.12. Frecuencias observadas cuando se comparan dos tests binarios, en presencia de verificación parcial de la enfermedad.

	$T_1 = 1$		$T_1 = 0$		Total
	$T_2 = 1$	$T_2 = 0$	$T_2 = 1$	$T_2 = 0$	
$V = 1$					
$D = 1$	s_{11}	s_{10}	s_{01}	s_{00}	s
$D = 0$	r_{11}	r_{10}	r_{01}	r_{00}	r
$V = 0$	u_{11}	u_{10}	u_{01}	u_{00}	u
Total	n_{11}	n_{10}	n_{01}	n_{00}	n

La ecuación para calcular el tamaño muestral cuando se pretende estimar mediante un intervalo de confianza a dos colas la diferencia entre las dos sensibilidades o especificidades es

$$m = \frac{z_{1-\alpha/2}^2}{d^2} \text{Var}(\hat{\theta}_1 - \hat{\theta}_2), \quad (3.36)$$

donde $z_{1-\alpha/2}$ es el $100\alpha/2$ -ésimo percentil de la distribución normal estándar, d es la precisión de la estimación y $\text{Var}(\hat{\theta}_1 - \hat{\theta}_2)$ es la función varianza de $\hat{\theta}_1 - \hat{\theta}_2$ (McCullagh y Nelder, 1989), siendo θ_i la sensibilidad (Se_i) o la especificidad (Sp_i) de cada test diagnóstico. Por tanto, para calcular el tamaño muestral es necesario determinar previamente $\text{Var}(\hat{\theta}_1 - \hat{\theta}_2)$, para lo cual se va a aplicar un método similar al propuesto en la Sección anterior.

Cuando el proceso de verificación depende solamente de los resultados de los dos tests binarios y no del resultado del gold estándar, se verifica que $\lambda_{ijk} = \lambda_{ij}$, $i, j, k = 0, 1$. Esta suposición es equivalente a suponer que el proceso de verificación es MAR (Rubin, 1976). Bajo la suposición MAR sean las probabilidades

$$\begin{aligned} \xi_{ij} &= P(V = 1, D = 1, T_1 = i, T_2 = j), \\ \psi_{ij} &= P(V = 1, D = 1, T_1 = i, T_2 = j), \\ \zeta_{ij} &= P(V = 0, T_1 = i, T_2 = j), \end{aligned} \quad (3.37)$$

con $i, j = 0, 1$, y $\sum_{i,j=0}^1 \xi_{ij} + \sum_{i,j=0}^1 \psi_{ij} + \sum_{i,j=0}^1 \zeta_{ij} = 1$. En general, tal y como ocurre en la mayoría de las situaciones prácticas, los dos tests diagnósticos son condicionalmente dependientes a la enfermedad (Vacek, 1985; Torrance-Rynard y Walter, 1997), por lo que las probabilidades (3.37) se pueden expresar en términos de las sensibilidades, especificidades, prevalencia, probabilidades de verificación y factores de dependencia como

$$\xi_{ij} = p\lambda_{ij} \left\{ Se_1^i (1 - Se_1)^{1-i} Se_2^j (1 - Se_2)^{1-j} + \delta_{ij} Se_1 Se_2 (\alpha_1 - 1) \right\},$$

$$\psi_{ij} = (1-p)\lambda_{ij} \left\{ Sp_1^{1-i} (1-Sp_1)^i Sp_2^{1-j} (1-Sp_2)^j + \delta_{ij} (1-Sp_1)(1-Sp_2)(\alpha_0 - 1) \right\},$$

$$\zeta_{ij} = \frac{(1-\lambda_{ij})}{\lambda_{ij}} (\xi_{ij} + \psi_{ij}),$$
(3.38)

donde $\delta_{ij} = 1$ si $i = j$ y $\delta_{ij} = -1$ si $i \neq j$, y los parámetros α_1 y α_0 son los factores de dependencia entre los dos tests diagnósticos. El factor de dependencia α_1 (α_0) es la covarianza entre los dos tests diagnósticos cuando $D = 1$ ($D = 0$) (Berry et al., 2002), y se verifica que $1 \leq \alpha_1 \leq 1/\max(Se_h)$ y $1 \leq \alpha_0 \leq 1/\max(1-Sp_h)$. Si $\alpha_1 = \alpha_0 = 1$ los dos tests diagnósticos son condicionalmente independientes a la enfermedad. Cuando $D = 1$ la correlación entre los dos tests diagnósticos viene dada por la expresión

$$\rho_1 = Corr(T = 1, T_2 = 1 | D = 1) = (\alpha_1 - 1) \sqrt{\frac{Se_1 Se_2}{(1-Se_1)(1-Se_2)}} \quad (3.39)$$

y cuando $D = 0$ la correlación es

$$\rho_0 = Corr(T_1 = 1, T_2 = 1 | D = 0) = (\alpha_0 - 1) \sqrt{\frac{(1-Sp_1)(1-Sp_2)}{Sp_1 Sp_2}}. \quad (3.40)$$

Una correlación positiva significa que la proporción de resultados concordantes (positivos o negativos) de ambos tests diagnósticos es mayor cuando ambos tests son condicionalmente independientes. Una correlación negativa indica una proporción mayor de resultados discordantes.

Sea $\boldsymbol{\omega} = (\xi_{11}, \xi_{10}, \xi_{01}, \xi_{00}, \psi_{11}, \psi_{10}, \psi_{01}, \psi_{00}, \zeta_{11}, \zeta_{10}, \zeta_{01}, \zeta_{00})^T$. Como ξ_{ij} , ψ_{ij} y ζ_{ij} son las probabilidades de una distribución multinomial, la matriz de varianzas-covarianzas de $\hat{\boldsymbol{\omega}}$ viene dada por la expresión

$$\Sigma_{\hat{\boldsymbol{\omega}}} = \{diag(\boldsymbol{\omega}) - \boldsymbol{\omega}^T \boldsymbol{\omega}\} / n. \quad (3.41)$$

Sea $\phi_{ij} = \xi_{ij} + \psi_{ij} + \zeta_{ij}$ con $i, j = 0, 1$. La sensibilidad y especificidad de cada test diagnóstico se escriben en términos de las probabilidades anteriores como

$$Se_1 = \frac{\sum_{j=0}^1 \frac{\xi_{1j}\phi_{1j}}{\xi_{1j} + \psi_{1j}}}{\sum_{i,j=0}^1 \frac{\xi_{ij}\phi_{ij}}{\xi_{ij} + \psi_{ij}}} \quad y \quad Sp_1 = \frac{\sum_{j=0}^1 \frac{\psi_{0j}\phi_{0j}}{\xi_{0j} + \psi_{0j}}}{\sum_{i,j=0}^1 \frac{\psi_{ij}\phi_{ij}}{\xi_{ij} + \psi_{ij}}} \quad (3.42)$$

para el test diagnóstico 1, y

$$Se_2 = \frac{\sum_{i=0}^1 \frac{\xi_{i1}\phi_{i1}}{\xi_{i1} + \psi_{i1}}}{\sum_{i,j=0}^1 \frac{\xi_{ij}\phi_{ij}}{\xi_{ij} + \psi_{ij}}} \quad y \quad Sp_2 = \frac{\sum_{i=0}^1 \frac{\psi_{i0}\phi_{i0}}{\xi_{i0} + \psi_{i0}}}{\sum_{i,j=0}^1 \frac{\psi_{ij}\phi_{ij}}{\xi_{ij} + \psi_{ij}}} \quad (3.43)$$

para el test 2. A partir de estas expresiones se obtiene que

$$Se_1 - Se_2 = \left\{ \frac{\xi_{10}\phi_{10}}{\xi_{10} + \psi_{10}} - \frac{\xi_{01}\phi_{01}}{\xi_{01} + \psi_{01}} \right\} / \sum_{i,j=0}^1 \frac{\xi_{ij}\phi_{ij}}{\xi_{ij} + \psi_{ij}} \quad (3.44)$$

y

$$Sp_1 - Sp_2 = \left\{ \frac{\psi_{01}\phi_{01}}{\xi_{01} + \psi_{01}} - \frac{\psi_{10}\phi_{10}}{\xi_{10} + \psi_{10}} \right\} / \sum_{i,j=0}^1 \frac{\psi_{ij}\phi_{ij}}{\xi_{ij} + \psi_{ij}}, \quad (3.45)$$

por lo que aplicando el método delta (Agresti, 2002) se obtiene que las funciones varianza son

$$Var(\hat{Se}_1 - \hat{Se}_2) = \left(\frac{\partial(Se_1 - Se_2)}{\partial \boldsymbol{\omega}} \right) \Sigma_{\hat{\boldsymbol{\omega}}} \left(\frac{\partial(Se_1 - Se_2)}{\partial \boldsymbol{\omega}} \right)^T \quad (3.46)$$

y

$$Var(\hat{Sp}_1 - \hat{Sp}_2) = \left(\frac{\partial(Sp_1 - Sp_2)}{\partial \boldsymbol{\omega}} \right) \Sigma_{\hat{\boldsymbol{\omega}}} \left(\frac{\partial(Sp_1 - Sp_2)}{\partial \boldsymbol{\omega}} \right)^T, \quad (3.47)$$

donde las derivadas parciales $\partial(Se_1 - Se_2)/\partial \boldsymbol{\omega}$ y $\partial(Sp_1 - Sp_2)/\partial \boldsymbol{\omega}$, y por consiguiente $Var(\hat{Se}_1 - \hat{Se}_2)$ y $Var(\hat{Sp}_1 - \hat{Sp}_2)$ son expresiones largas y complicadas que necesitan ser evaluadas con un software matemático o estadístico. Por lo tanto, si se conoce, ya

sea a partir de una muestra piloto o de algún otro estudio previo, la sensibilidad y especificidad de cada test de diagnóstico, las probabilidades de verificación, las covarianzas α_1 y α_0 (o las correlaciones ρ_1 y ρ_0) y la prevalencia de la enfermedad, aplicando la ecuación (3.36) se obtiene el tamaño muestral necesario para comparar la diferencia de las sensibilidades o especificidades con una precisión dada a la confianza $100(1-\alpha)\%$.

El método que se propone en esta sección también se puede aplicar cuando se quiere calcular el tamaño muestral para contrastar

$$\begin{aligned} H_0 : \theta_1 &= \theta_2 \\ H_1 : \theta_1 &\neq \theta_2 \end{aligned}$$

donde θ_i es la sensibilidad o la especificidad. En esta situación, para calcular el tamaño muestral hay que especificar un valor para la diferencia entre θ_1 y θ_2 bajo la hipótesis alternativa. Si se denota por δ a la diferencia entre θ_1 y θ_2 , esto es $\delta = \theta_1 - \theta_2$, la fórmula para determinar el tamaño muestral en diseños apareados viene dado por la expresión

$$m = \left(\frac{z_{1-\alpha/2} + z_{1-\beta}}{\delta} \right)^2 \text{Var}(\hat{\theta}_1 - \hat{\theta}_2), \quad (3.48)$$

siendo z_γ el 100γ -ésimo percentil de la distribución normal estándar y donde la varianza de $\theta_1 - \theta_2$ se calcula aplicando el método que se ha descrito en anteriormente.

El método para calcular el tamaño muestral exige conocer las sensibilidades y especificidades, la prevalencia, las probabilidades de verificación y las covarianzas (correlaciones) entre los dos tests binarios. Sin embargo, en la práctica el cálculo del tamaño muestral es difícil, ya que las covarianzas α_1 y α_0 (o las correlaciones ρ_1 y ρ_0) son difíciles de estimar. Sin embargo, estas covarianzas (correlaciones) se pueden estimar aplicando el algoritmo *EM* deducido por Roldán Nofuentes, Luna del Castillo y Femia Marzo (2009) explicado en la Sección 2.2.1.6.

3.2.2.1. Experimentos de simulación

De forma similar a la Sección 3.2.1.1 se han realizado experimentos de simulación para estudiar la robustez del método propuesto para calcular el tamaño muestral necesario para estimar la diferencia entre dos sensibilidades o dos especificidades. Para ello se han generado 5000 muestras aleatorias de distribuciones multinomiales, cuyas probabilidades vienen dadas por las ecuaciones (3.38). Para las sensibilidades y especificidades se han tomado los valores

$$(Se_1 = 0.90, Sp_1 = 0.85, Se_2 = 0.85, Sp_2 = 0.80),$$

$$(Se_1 = 0.90, Sp_1 = 0.85, Se_2 = 0.80, Sp_2 = 0.70)$$

y

$$(Se_1 = 0.95, Sp_1 = 0.95, Se_2 = 0.90, Sp_2 = 0.90),$$

que son valores que aparecen con cierta frecuencia en la práctica clínica. Para la prevalencia se han tomado los valores 0.1, 0.3, 0.5, 0.7 y 0.9, y como probabilidades de verificación se han tomado

$$(\lambda_{11} = 0.80, \lambda_{10} = \lambda_{01} = 0.40, \lambda_{00} = 0.10),$$

$$(\lambda_{11} = 0.95, \lambda_{10} = \lambda_{01} = 0.60, \lambda_{00} = 0.30)$$

y

$$(\lambda_{11} = \lambda_{10} = \lambda_{01} = \lambda_{00} = 1).$$

Tanto como valor de la precisión d como del error α se ha tomado el 5%. El tamaño muestral ha sido calculado a través del método propuesto anteriormente, y los estimadores máximo verosímiles para las sensibilidades, especificidades, prevalencia y correlaciones se obtuvieron aplicando el algoritmo *EM* deducido por Roldán Nofuentes, Luna del Castillo y Femia Marzo (2009), y los estimadores máximo verosímiles de las probabilidades de verificación se han calculado con las ecuaciones

$$\hat{\lambda}_{ij} = \frac{s_{ij} + r_{ij}}{n_{ij}}, \quad (3.49)$$

con $i, j = 0, 1$. Cuando $\lambda_{ij} = 1$, los estimadores se han obtenido aplicando las ecuaciones clásicas en tablas 2×4 . Para cada 5000 muestras con la misma distribución multinomial se han calculado el tamaño muestral medio y la raíz del error cuadrático medio relativo (*RECMR*) de forma similar que como se ha hecho en las simulaciones anteriores, de tal forma que el método de cálculo del tamaño muestral es válido cuando el valor de *RECMR* del estimador del tamaño de muestra es menor que el 30%.

En las Tablas 3.13, 3.14, 3.15, 3.16, 3.17 y 3.18 se muestran los resultados obtenidos para valores nulos de las correlaciones ($\rho_1 = \rho_0 = 0$), para valores intermedios de ρ_1 y ρ_0 , y para valores muy altos de ρ_1 y ρ_0 . De los resultados de los experimentos se obtienen las siguientes conclusiones. Para las sensibilidades, en términos generales, cuando la correlación ρ_1 entre los dos tests diagnósticos binarios es baja o intermedia, (independiente del valor de la correlación ρ_0), el estimador del tamaño muestral para comparar las dos sensibilidades tiene un *RECMR* menor del 30%; mientras que si la correlación ρ_1 es alta, su *RECMR* es mayor del 30%. Con respecto a las especificidades, en términos generales, cuando la correlación ρ_0 entre los dos tests binarios es baja o intermedia, el estimador del tamaño para comparar las dos especificidades tiene un *RECMR* inferior al 30%, aunque si la correlación ρ_0 es alta su *RECMR* es mayor que el 30%. Por tanto, las correlaciones ρ_1 y ρ_0 entre los dos tests diagnósticos tienen un importante efecto en el método que del cálculo del tamaño muestral, y esto se debe al poco efecto que tienen las correlaciones sobre la proporción de pacientes verificados (para una misma prevalencia y probabilidades de verificación), ya que el aumento de ρ_1 y/o ρ_0 no tiene un importante efecto en el incremento del porcentaje de pacientes verificados. En cuanto a la prevalencia de la enfermedad, su aumento o disminución tiene poco efecto sobre la robustez del método del cálculo del tamaño muestral, sobre todo cuando se calcula el tamaño muestral para estimar la diferencia entre las sensibilidades.

Tabla 3.13. Tamaño muestral para comparar las sensibilidades.

$Se_1 = 0.90 \quad Sp_1 = 0.85 \quad Se_2 = 0.85 \quad Sp_2 = 0.80 \quad d = Se_1 - Se_2 = 0.05$												
$\lambda_{11} = 0.80 \quad \lambda_{10} = \lambda_{01} = 0.40 \quad \lambda_{00} = 0.10$												
$\rho_1 = \rho_0 = 0$					$\rho_1 = 0.36 \quad \rho_0 = 0.37$				$\rho_1 = 0.71 \quad \rho_0 = 0.74$			
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR
0.10	7892	25.7	7772	0.09	5063	26.2	4984	0.16	2090	26.8	2336	0.36
0.30	2341	35.6	2304	0.09	1483	36.1	1462	0.16	562	36.6	673	0.43
0.50	1222	45.5	1200	0.10	773	45.9	763	0.16	292	46.4	356	0.47
0.70	732	55.3	716	0.10	467	55.8	459	0.17	187	56.2	212	0.40
0.90	444	65.2	439	0.10	287	65.6	297	0.18	127	66.0	153	0.46
$\lambda_{11} = 0.95 \quad \lambda_{10} = \lambda_{01} = 0.60 \quad \lambda_{00} = 0.30$												
$\rho_1 = \rho_0 = 0$					$\rho_1 = 0.36 \quad \rho_0 = 0.37$				$\rho_1 = 0.71 \quad \rho_0 = 0.74$			
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR
0.10	5363	45.2	5335	0.10	3450	45.5	3432	0.17	1473	45.7	1668	0.38
0.30	1659	54.4	1653	0.10	1059	54.6	1052	0.16	431	54.8	505	0.40
0.50	915	63.5	910	0.10	584	63.7	581	0.16	239	63.9	286	0.42
0.70	591	72.6	586	0.10	379	72.8	378	0.16	161	73.0	190	0.43
0.90	404	81.8	401	0.11	262	82.0	265	0.16	118	82.2	137	0.41
$\lambda_{11} = \lambda_{10} = \lambda_{01} = \lambda_{00} = 1$												
$\rho_1 = \rho_0 = 0$					$\rho_1 = 0.36 \quad \rho_0 = 0.37$				$\rho_1 = 0.71 \quad \rho_0 = 0.74$			
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR
0.10	3343	100	3333	0.12	2167	100	2169	0.18	992	100	1092	0.40
0.30	1115	100	1117	0.11	723	100	724	0.17	331	100	361	0.37
0.50	669	100	669	0.11	434	100	432	0.17	199	100	219	0.38
0.70	478	100	477	0.11	310	100	309	0.17	142	100	157	0.38
0.90	372	100	372	0.10	241	100	242	0.17	111	100	124	0.38

p : prevalencia; m : tamaño muestral; %V: % promedio de pacientes verificados; \bar{m} : tamaño muestral medio; RECMR: raíz del error cuadrático medio relativo.

Tabla 3.14. Tamaño muestral para comparar las especificidades.

$Se_1 = 0.90 \quad Sp_1 = 0.85 \quad Se_2 = 0.85 \quad Sp_2 = 0.80 \quad d = Sp_1 - Sp_2 = 0.05$												
$\lambda_{11} = 0.80 \quad \lambda_{10} = \lambda_{01} = 0.40 \quad \lambda_{00} = 0.10$												
$\rho_1 = \rho_0 = 0$					$\rho_1 = 0.36 \quad \rho_0 = 0.37$				$\rho_1 = 0.71 \quad \rho_0 = 0.74$			
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR
0.10	550	25.7	573	0.09	350	26.2	385	0.17	148	26.8	199	0.56
0.30	863	35.6	882	0.08	546	36.1	563	0.14	217	36.6	271	0.46
0.50	1444	45.5	1459	0.08	906	45.9	915	0.13	337	46.4	414	0.47
0.70	2848	55.3	2852	0.08	1780	55.8	1791	0.14	637	56.2	743	0.44
0.90	10126	65.2	10125	0.08	6408	65.6	6442	0.16	2433	66.0	2712	0.41
$\lambda_{11} = 0.95 \quad \lambda_{10} = \lambda_{01} = 0.60 \quad \lambda_{00} = 0.30$												
$\rho_1 = \rho_0 = 0$					$\rho_1 = 0.36 \quad \rho_0 = 0.37$				$\rho_1 = 0.71 \quad \rho_0 = 0.74$			
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR
0.10	517	45.2	523	0.08	329	45.5	336	0.13	140	45.7	158	0.34
0.30	734	54.4	737	0.08	465	54.6	468	0.12	191	54.8	214	0.33
0.50	1133	63.5	1132	0.08	714	63.7	715	0.13	281	63.9	311	0.34
0.70	2083	72.6	2082	0.08	1308	72.8	1308	0.13	500	73.0	545	0.34
0.90	6948	81.8	6936	0.08	4388	82.0	4388	0.14	1717	82.2	1840	0.35
$\lambda_{11} = \lambda_{10} = \lambda_{01} = \lambda_{00} = 1$												
$\rho_1 = \rho_0 = 0$					$\rho_1 = 0.36 \quad \rho_0 = 0.37$				$\rho_1 = 0.71 \quad \rho_0 = 0.74$			
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR
0.10	491	100	490	0.08	312	100	311	0.13	133	100	139	0.30
0.30	632	100	629	0.08	401	100	400	0.13	171	100	177	0.30
0.50	884	100	882	0.08	561	100	560	0.13	239	100	249	0.31
0.70	1473	100	1471	0.08	935	100	934	0.14	397	100	410	0.31
0.90	4418	100	4419	0.09	2805	100	2808	0.14	1191	100	1244	0.32

p : prevalencia; m : tamaño muestral; %V: % promedio de pacientes verificados; \bar{m} : tamaño muestral medio; RECMR: raíz del error cuadrático medio relativo.

Tabla 3.15. Tamaño muestral para comparar las sensibilidades.

$Se_1 = 0.90 \quad Sp_1 = 0.85 \quad Se_2 = 0.80 \quad Sp_2 = 0.70 \quad d = Se_1 - Se_2 = 0.05$												
$\lambda_{11} = 0.80 \quad \lambda_{10} = \lambda_{01} = 0.40 \quad \lambda_{00} = 0.10$												
$\rho_1 = \rho_0 = 0$					$\rho_1 = 0.30 \quad \rho_0 = 0.32$				$\rho_1 = 0.60 \quad \rho_0 = 0.64$			
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR
0.10	2215	47.8	2065	0.16	1524	48.1	1450	0.23	555	48.3	795	0.65
0.30	649	56.0	602	0.18	433	56.2	413	0.24	171	56.5	261	0.82
0.50	341	64.2	316	0.18	228	64.4	220	0.25	100	64.6	140	0.68
0.70	209	72.3	194	0.19	143	72.5	140	0.26	70	72.7	92	0.58
0.90	130	80.5	134	0.20	92	80.7	102	0.31	53	80.9	71	0.61
$\lambda_{11} = 0.95 \quad \lambda_{10} = \lambda_{01} = 0.60 \quad \lambda_{00} = 0.30$												
$\rho_1 = \rho_0 = 0$					$\rho_1 = 0.30 \quad \rho_0 = 0.32$				$\rho_1 = 0.60 \quad \rho_0 = 0.64$			
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR
0.10	1516	28.4	1466	0.17	1053	28.9	1033	0.24	466	29.4	652	0.64
0.30	466	37.2	452	0.17	318	37.7	315	0.25	149	38.2	218	0.70
0.50	258	46.1	249	0.17	177	46.5	177	0.25	89	47.0	127	0.68
0.70	169	54.9	163	0.17	118	55.3	119	0.25	63	55.7	84	0.59
0.90	117	63.8	118	0.18	83	64.2	89	0.27	47	64.5	65	0.64
$\lambda_{11} = \lambda_{10} = \lambda_{01} = \lambda_{00} = 1$												
$\rho_1 = \rho_0 = 0$					$\rho_1 = 0.30 \quad \rho_0 = 0.32$				$\rho_1 = 0.60 \quad \rho_0 = 0.64$			
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR
0.10	961	100	951	0.20	684	100	691	0.26	408	100	495	0.50
0.30	321	100	318	0.19	228	100	230	0.26	136	100	165	0.48
0.50	193	100	191	0.18	137	100	138	0.26	82	100	101	0.47
0.70	138	100	137	0.18	98	100	99	0.25	59	100	73	0.47
0.90	107	100	107	0.17	76	100	78	0.25	46	100	58	0.49

p : prevalencia; m : tamaño muestral; %V: % promedio de pacientes verificados; \bar{m} : tamaño muestral medio; RECMR: raíz del error cuadrático medio relativo.

Tabla 3.16. Tamaño muestral para comparar las especificidades.

$Se_1 = 0.90 \quad Sp_1 = 0.85 \quad Se_2 = 0.80 \quad Sp_2 = 0.70 \quad d = Sp_1 - Sp_2 = 0.05$												
$\lambda_{11} = 0.80 \quad \lambda_{10} = \lambda_{01} = 0.40 \quad \lambda_{00} = 0.10$												
$\rho_1 = \rho_0 = 0$					$\rho_1 = 0.30 \quad \rho_0 = 0.32$				$\rho_1 = 0.60 \quad \rho_0 = 0.64$			
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR
0.10	73	47.8	123	0.90	51	48.1	101	0.92	26	48.3	87	0.92
0.30	113	56.0	155	0.57	77	56.2	124	0.89	36	56.5	167	0.98
0.50	185	64.2	231	0.45	124	64.4	182	0.75	55	64.6	123	0.60
0.70	359	72.3	417	0.35	240	72.5	312	0.60	106	72.7	227	0.71
0.90	1307	80.5	1383	0.29	921	80.7	1023	0.48	451	80.9	942	0.53
$\lambda_{11} = 0.95 \quad \lambda_{10} = \lambda_{01} = 0.60 \quad \lambda_{00} = 0.30$												
$\rho_1 = \rho_0 = 0$					$\rho_1 = 0.30 \quad \rho_0 = 0.32$				$\rho_1 = 0.60 \quad \rho_0 = 0.64$			
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR
0.10	68	28.4	80	0.29	47	28.9	63	0.52	25	29.4	47	0.89
0.30	96	37.2	105	0.23	66	37.7	80	0.41	32	38.2	49	0.75
0.50	146	46.1	155	0.23	99	46.5	115	0.39	45	47.0	68	0.90
0.70	266	54.9	274	0.22	179	55.3	197	0.37	75	55.7	94	0.42
0.90	893	63.8	894	0.23	613	64.2	652	0.38	223	64.5	235	0.39
$\lambda_{11} = \lambda_{10} = \lambda_{01} = \lambda_{00} = 1$												
$\rho_1 = \rho_0 = 0$					$\rho_1 = 0.30 \quad \rho_0 = 0.32$				$\rho_1 = 0.60 \quad \rho_0 = 0.64$			
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR
0.10	65	100	65	0.18	45	100	49	0.28	26	100	51	0.52
0.30	83	100	83	0.19	58	100	60	0.27	34	100	69	0.46
0.50	116	100	114	0.20	80	100	82	0.28	48	100	108	0.77
0.70	193	100	191	0.21	134	100	136	0.29	85	100	157	0.49
0.90	577	100	570	0.22	400	100	411	0.31	300	100	354	0.57

p : prevalencia; m : tamaño muestral; %V: % promedio de pacientes verificados; \bar{m} : tamaño muestral medio; RECMR: raíz del error cuadrático medio relativo.

Tabla 3.17. Tamaño muestral para comparar las sensibilidades.

$Se_1 = 0.95$ $Sp_1 = 0.95$ $Se_2 = 0.90$ $Sp_2 = 0.90$ $d = Se_1 - Se_2 = 0.05$												
$\lambda_{11} = 0.80$ $\lambda_{10} = \lambda_{01} = 0.40$ $\lambda_{00} = 0.10$												
$\rho_1 = \rho_0 = 0$					$\rho_1 = 0.26$ $\rho_0 = 0.32$				$\rho_1 = 0.52$ $\rho_0 = 0.64$			
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR
0.10	4803	40.1	4618	0.15	3482	40.1	3373	0.21	1885	40.2	2077	0.33
0.30	1368	51.1	1317	0.16	967	51.2	935	0.22	511	51.3	614	0.44
0.50	706	62.1	677	0.16	502	62.2	489	0.23	281	62.3	333	0.44
0.70	428	73.2	412	0.17	311	73.3	307	0.23	187	73.4	207	0.38
0.90	271	84.3	271	0.19	202	84.3	213	0.26	132	84.4	152	0.41
$\lambda_{11} = 0.95$ $\lambda_{10} = \lambda_{01} = 0.60$ $\lambda_{00} = 0.30$												
$\rho_1 = \rho_0 = 0$					$\rho_1 = 0.26$ $\rho_0 = 0.32$				$\rho_1 = 0.52$ $\rho_0 = 0.64$			
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR
0.10	3308	20.5	3254	0.16	2428	20.7	2396	0.22	1425	20.9	1564	0.34
0.30	999	32.4	976	0.16	724	32.6	719	0.22	423	32.8	494	0.39
0.50	549	44.3	539	0.16	399	44.5	396	0.22	242	44.7	280	0.40
0.70	358	56.2	351	0.16	264	56.4	263	0.22	167	56.6	187	0.37
0.90	251	68.1	251	0.16	188	68.3	192	0.23	124	68.4	139	0.36
$\lambda_{11} = \lambda_{10} = \lambda_{01} = \lambda_{00} = 1$												
$\rho_1 = \rho_0 = 0$					$\rho_1 = 0.26$ $\rho_0 = 0.32$				$\rho_1 = 0.52$ $\rho_0 = 0.64$			
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR
0.10	2113	100	2103	0.18	1588	100	1594	0.24	1062	100	1139	0.36
0.30	705	100	702	0.18	530	100	527	0.24	354	100	381	0.35
0.50	423	100	421	0.17	318	100	319	0.23	213	100	229	0.35
0.70	302	100	302	0.17	227	100	227	0.23	152	100	164	0.35
0.90	235	100	236	0.17	177	100	179	0.23	118	100	129	0.35

p : prevalencia; m : tamaño muestral; %V: % promedio de pacientes verificados; \bar{m} : tamaño muestral medio; RECMR: raíz del error cuadrático medio relativo.

Tabla 3.18. Tamaño muestral para comparar las especificidades.

$Se_1 = 0.95 \quad Sp_1 = 0.95 \quad Se_2 = 0.90 \quad Sp_2 = 0.90 \quad d = Sp_1 - Sp_2 = 0.05$												
$\lambda_{11} = 0.80 \quad \lambda_{10} = \lambda_{01} = 0.40 \quad \lambda_{00} = 0.10$												
$\rho_1 = \rho_0 = 0$					$\rho_1 = 0.26 \quad \rho_0 = 0.32$				$\rho_1 = 0.52 \quad \rho_0 = 0.64$			
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR
0.10	270	40.1	310	0.26	192	40.1	236	0.40	110	40.2	173	0.86
0.30	428	51.1	452	0.19	298	51.2	334	0.31	156	51.3	244	0.82
0.50	706	62.1	733	0.18	482	62.2	524	0.28	241	62.3	387	0.85
0.70	1369	73.2	1402	0.17	933	73.3	973	0.25	465	73.4	706	0.77
0.90	4815	84.3	4799	0.16	3370	84.3	3411	0.26	1829	84.4	2330	0.56
$\lambda_{11} = 0.95 \quad \lambda_{10} = \lambda_{01} = 0.60 \quad \lambda_{00} = 0.30$												
$\rho_1 = \rho_0 = 0$					$\rho_1 = 0.26 \quad \rho_0 = 0.32$				$\rho_1 = 0.52 \quad \rho_0 = 0.64$			
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR
0.10	251	20.5	261	0.18	178	20.7	191	0.26	104	20.9	136	0.55
0.30	358	32.4	364	0.17	251	32.6	262	0.24	139	32.8	183	0.55
0.50	549	44.3	553	0.17	381	44.5	389	0.24	205	44.7	276	0.57
0.70	1000	56.2	999	0.16	692	56.4	701	0.24	369	56.6	485	0.55
0.90	3311	68.1	3290	0.17	2320	68.3	2347	0.25	1288	68.4	1588	0.49
$\lambda_{11} = \lambda_{10} = \lambda_{01} = \lambda_{00} = 1$												
$\rho_1 = \rho_0 = 0$					$\rho_1 = 0.26 \quad \rho_0 = 0.32$				$\rho_1 = 0.52 \quad \rho_0 = 0.64$			
p	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR	m	%V	\bar{m}	RECMR
0.10	235	100	235	0.17	167	100	169	0.24	99	100	117	0.44
0.30	302	100	303	0.17	215	100	215	0.24	127	100	149	0.44
0.50	423	100	421	0.18	300	100	301	0.25	177	100	207	0.44
0.70	705	100	700	0.17	500	100	502	0.25	295	100	340	0.44
0.90	2113	100	2106	0.18	1499	100	1514	0.26	884	100	1031	0.46

p : prevalencia; m : tamaño muestral; %V: % promedio de pacientes verificados; \bar{m} : tamaño muestral medio; RECMR: raíz del error cuadrático medio relativo.

En la Tabla 3.19 se resumen las condiciones generales para las cuales es posible aplicar el método del cálculo del tamaño de muestra necesario para comparar las sensibilidades (especificidades) en presencia de verificación parcial de la enfermedad. Para las sensibilidades el método es válido cuando ρ_1 es baja o intermedia, mientras

que para las especificidades el método es válido cuando ρ_0 es baja o intermedia. En cualquier otro caso el método propuesto para calcular el tamaño muestral no ofrece resultados satisfactorios.

Tabla 3.19. Condiciones de aplicación del método del cálculo de tamaño muestral para comparar la exactitud de dos tests binarios.

Comparación de dos sensibilidades			
		ρ_0	
		Baja o intermedia	Alta
ρ_1	Baja o intermedia	Método válido	Método válido
	Alta	Método no válido	Método no válido
Comparación de dos especificidades			
		ρ_0	
		Baja o intermedia	Alta
ρ_1	Baja o intermedia	Método válido	Método no válido
	Alta	Método válido	Método no válido

3.2.2.2. Aplicación

El método propuesto en la Sección anterior se ha aplicado al diagnóstico de la estenosis coronaria. El diagnóstico de la estenosis coronaria se puede realizar mediante una ecocardiografía con dobutamina o mediante una ecocardiografía con esfuerzo, y como gold estándar se utiliza la angiografía coronaria. Como la angiografía puede causar reacciones en los pacientes (infarto, infecciones, trombosis,...) no todos los pacientes son verificados. En la Tabla 3.20 se muestran los resultados obtenidos al aplicar la ecocardiografía con dobutamina y la ecocardiografía con esfuerzo, a una muestra de 655 personas. La variable aleatoria T_1 modeliza el resultado de la ecocardiografía con dobutamina, la variable aleatoria T_2 modeliza el resultado de la ecocardiografía con esfuerzo y la variable aleatoria D modeliza el resultado de la angiografía coronaria. En todo el estudio la selección de los pacientes para verificar el estado de enfermedad con la angiografía se realizó solamente en función de los resultados de la ecocardiografía con dobutamina y con esfuerzo, por lo que el proceso de verificación con el gold estándar es MAR.

Tabla 3.20. Comparación de las ecocardiografías con dobutamina y con esfuerzo.

	$T_1 = 1$		$T_1 = 0$		Total
	$T_2 = 1$	$T_2 = 0$	$T_2 = 1$	$T_2 = 0$	
$V = 1$					
$D = 1$	95	14	5	2	116
$D = 0$	70	20	47	79	216
$V = 0$	30	10	24	259	323
Total	195	44	76	340	655

Aplicando el método de Zhou (1998b), las sensibilidades y especificidades estimadas de cada test diagnóstico y los intervalos de confianza al 95% para la diferencia de las sensibilidades y de las especificidades son $\hat{S}e_1 = 0.89$, $\hat{S}e_2 = 0.82$, $\hat{S}p_1 = 0.79$, $\hat{S}p_2 = 0.70$, $Se_1 - Se_2 \in (-0.0007, 0.1487)$ y $Sp_1 - Sp_2 \in (0.0461, 0.1221)$. Asimismo la prevalencia y las probabilidades de verificación estimadas son $\hat{p} = 0.22$, $\hat{\lambda}_{11} = 0.85$, $\hat{\lambda}_{10} = 0.77$, $\hat{\lambda}_{01} = 0.68$ y $\hat{\lambda}_{00} = 0.24$. Aplicando el algoritmo *EM* deducido por Roldán Nofuentes, Luna del Castillo y Femia Marzo (2009), las covarianzas (correlaciones) estimadas son $\hat{\alpha}_1 = 1.052$ ($\hat{\rho}_1 = 0.32$) y $\hat{\alpha}_0 = 2.560$ ($\hat{\rho}_0 = 0.53$). Si se pretende calcular el tamaño muestral necesario para estimar la diferencia de las sensibilidades con una precisión $d = 5\%$ a un nivel de confianza del 95%, aplicando el método anterior se obtiene que $Var(\hat{S}e_1 - \hat{S}e_2) = 0.976$ y aplicando la ecuación (3.36) el tamaño muestral mínimo es $m = 1500$. Por tanto, es necesario una muestra de al menos 1500 pacientes para estimar la diferencia de las dos sensibilidades con una precisión del 5% con una confianza del 95% (es necesario añadir a la muestra inicial de 665 pacientes un mínimo de 835 pacientes).

3.2.2.3. Conclusiones

En la Sección 3.2 se ha propuesto un método para calcular el tamaño muestral cuando se evalúa la exactitud de un único test binario y cuando se compara la exactitud de dos

tests binarios, en presencia del sesgo de verificación. En cada caso el método que se propone requiere que el proceso de verificación sea MAR. No obstante, en algunas situaciones prácticas el proceso de verificación depende no sólo del resultado del test diagnóstico (o de los resultados del test diagnóstico) sino que también depende del estado de la enfermedad. Entonces, si el proceso de verificación depende de las variables no observadas, las cuales están relacionadas con el estado de la enfermedad, el proceso de verificación no es MAR y, por consiguiente el método que se propone no se puede aplicar. También es necesario que se conozcan los valores de la sensibilidad y especificidad de cada test, prevalencia de la enfermedad, covarianzas entre los dos tests diagnósticos (cuando se comparan dos tests binarios) y probabilidades de verificación (valores que se pueden conjeturar mediante un estudio inicial o una muestra piloto), así como de las especificaciones para calcular el tamaño muestral (confianza y precisión).

Se han realizado unos experimentos de simulación para estudiar la robustez del método del cálculo del tamaño muestral, tanto si se evalúa un único test diagnóstico como si se comparan dos tests diagnósticos. En el primer caso (un único test diagnóstico) se ha estudiado el efecto que las probabilidades de verificación y la prevalencia tienen sobre el tamaño muestral, concluyendo que el método propuesto ofrece unos resultados muy satisfactorios, ya que los experimentos de simulación han demostrado que en todos los casos estudiados, la raíz del error cuadrático medio relativo es siempre menor que el 30%, lo que indican la validez del método para calcular el tamaño muestral para evaluar la exactitud de un test binario en presencia de verificación parcial de la enfermedad. En el segundo caso (dos tests diagnósticos) se ha estudiado el efecto que tienen las probabilidades de verificación, la prevalencia y las correlaciones (ρ_1 y ρ_0) sobre el tamaño muestral para estimar la diferencia en las sensibilidades (especificidades) a una precisión d con una confianza $100(1-\alpha)\%$. Los resultados de los experimentos de simulación han mostrado que las correlaciones ρ_1 y ρ_0 son los parámetros que tienen el efecto más importante en el tamaño muestral, concluyéndose que el método para calcular el tamaño muestral cuando se comparan la exactitud de dos tests diagnósticos en presencia de verificación parcial de la enfermedad es válido cuando la correlación ρ_1 (ρ_0) es baja o intermedia, y el método no es bueno cuando la correlación ρ_1 (ρ_0)

es alta. Las correlaciones pueden estimarse a partir de una muestra piloto o de estudios anteriores aplicando el algoritmo *EM*.

Por otra parte, el método propuesto para calcular el tamaño muestral se basa en la aproximación normal de los estimadores. Cuando esta aproximación no es válida, el método no se puede aplicar. Otra limitación del método radica en que las ecuaciones del tamaño muestral se basan en intervalos de confianza de Wald, por lo que cuando los valores de la exactitud del test diagnóstico son muy cercanos a 0 ó a 1 el método para calcular el tamaño muestral está afectado por el mal rendimiento del intervalo de confianza (Agresti y Coull, 1998; Brown et al., 2001). Una posible solución a estos problemas podría encontrarse aplicando métodos de inferencia exacta. Por tanto, es necesario investigar métodos de inferencia exacta cuando, en presencia de verificación parcial de la enfermedad, se evalúa o comparan tests diagnósticos binarios. Otra solución a este problema podría ser la aplicación de imputación múltiple y el cálculo de tamaño muestral aplicando los métodos en tablas 2×2 .

Otra cuestión importante cuando se evalúa la exactitud de un test diagnóstico en presencia de verificación parcial es cómo determinar la estrategia para establecer las probabilidades de verificación. De los experimentos de simulación se obtienen en términos generales algunas conclusiones. Cuando se calcula el tamaño muestral para estimar la sensibilidad con una precisión d y una confianza $100(1-\alpha)\%$, es preferible tomar probabilidades de verificación altas, ya que el costo de verificar a los pacientes con el gold estándar es más bajo debido a que el tamaño muestral es más pequeño cuando las probabilidades de verificación son altas que cuando son bajas, ya que el número medio de pacientes verificados es más bajo cuando las probabilidades de verificación son altas que cuando éstas son bajas. Así por ejemplo, para $Se = 0.80$, $Sp = 0.70$ y $p = 50\%$, si $\lambda_1 = 0.80$ y $\lambda_0 = 0.10$, el tamaño muestral necesario para estimar la sensibilidad con una precisión $d = 5\%$ y un nivel de confianza del 95% es $m = 3253$, verificándose un 48.5% de pacientes por término medio (1578 pacientes); si $\lambda_1 = 0.95$ y $\lambda_0 = 0.30$, el tamaño muestral es de $m = 1208$, verificándose un 65.7% de los pacientes (794 pacientes en promedio). Por consiguiente, en el segundo caso los costes médicos de verificación son menores debido a que el tamaño muestral es mucho más pequeño. No obstante, cuando se calcula el tamaño muestral para estimar la

especificidad a una precisión d a una confianza $100(1-\alpha)\%$, la prevalencia de la enfermedad tiene un efecto importante a la hora de fijar las probabilidades de verificación. En general, cuando $p \leq 30\%$ es recomendable fijar probabilidades de verificación bajas para que los costos de verificación sean más bajos, ya que el número medio de pacientes verificados es menor cuando las probabilidades de verificación son menores que cuando estas probabilidades son mayores; y cuando $p > 30\%$ es recomendable fijar probabilidades de verificación altas para que el número medio de pacientes verificados sea menor. Esta misma cuestión ocurre al comparar la exactitud de dos tests binarios en presencia de verificación parcial. Al calcular el tamaño muestral para estimar la diferencia entre dos sensibilidades, para un mismo valor de la prevalencia y las correlaciones ρ_1 y ρ_0 , es recomendable fijar probabilidades de verificación bajas, ya que el número medio de pacientes verificados es menor cuando las probabilidades de verificación λ_{ij} son bajas que cuando éstas son altas. Asimismo, cuando se calcula el tamaño de muestra para estimar la diferencia entre especificidades, la prevalencia de la enfermedad tiene un efecto más importante en la elección de las probabilidades de verificación. En general, cuando $p \leq 50\%$ es recomendable fijar unas probabilidades de verificación bajas (por las mismas razones anteriores), y cuando $p > 50\%$ es recomendable establecer probabilidades de verificación altas.

Conclusiones

La evaluación y la comparación del rendimiento de tests diagnósticos binarios es uno de los tópicos más importantes en el estudio de los métodos estadísticos para el diagnóstico médico. Clásicamente el rendimiento de un test diagnóstico binario respecto a un gold estándar se mide en términos de la sensibilidad y la especificidad. En el Capítulo 1 se estudia la estimación de la sensibilidad y especificidad de un test diagnóstico binario y de otros parámetros del rendimiento del test diagnóstico, como son las razones de verosimilitud, los valores predictivos y el coeficiente kappa ponderado, desde dos perspectivas diferentes: cuando el gold estándar se aplica a todos los individuos objeto de estudio (verificación completa de la enfermedad) y cuando el gold estándar solamente se aplica a un subconjunto de estos individuos (verificación parcial de la enfermedad). En este Capítulo se han realizado dos aportaciones: la estimación por intervalos de confianza del coeficiente kappa ponderado de un test binario en presencia de verificación completa, y los intervalos de confianza cuadráticos para la sensibilidad y especificidad de un test binario en presencia de verificación parcial. Para el coeficiente kappa ponderado, se recomienda utilizar el intervalo de confianza tipo Wald (o Bootstrap) cuando el tamaño de la muestra es relativamente pequeño y el intervalo logia cuando el tamaño es relativamente mayor; para la sensibilidad y especificidad, el intervalo cuadrático tiene un buen rendimiento principalmente cuando las muestras son grandes y las probabilidades de verificación no son bajas. Todos los resultados

obtenidos se basan en la normalidad asintótica de los correspondientes estimadores, por lo que cuando las muestras son pequeñas o relativamente pequeñas los resultados obtenidos pueden ser erróneos. Por tanto es necesario investigar métodos de inferencia exacta que permitan estimar estos parámetros (coeficiente kappa ponderado en presencia de verificación completa, y sensibilidad y especificidad en presencia de verificación parcial) cuando las muestras son pequeñas.

En el Capítulo 2 se ha estudiado la comparación de los parámetros anteriores de dos tests diagnósticos binarios en las mismas situaciones que el Capítulo 1 (verificación completa y parcial). En este Capítulo se ha realizado una aportación: la comparación de los coeficientes kappa ponderados de múltiples tests binarios cuando todos los tests diagnósticos y el gold estándar se aplican a todos los individuos de una muestra aleatoria. Se ha deducido un test de hipótesis global para comparar simultáneamente, para un mismo índice de ponderación, múltiples coeficientes kappa ponderados y se han realizado unos experimentos de simulación para estudiar el error tipo I y la potencia de este test de hipótesis cuando se comparan tres coeficientes kappa ponderados, tanto si la matriz de varianzas–covarianzas de los estimadores de los coeficientes kappa ponderados se estima mediante el método delta o mediante bootstrap, obteniéndose que no hay diferencias importantes entre los resultados obtenidos por ambos métodos. Asimismo, se han comparado el error tipo I y la potencia del test de hipótesis conjunto con los errores tipo I y potencias de otros dos métodos para comparar los tres coeficientes kappa ponderados: un primer método basado en las comparaciones de los tres coeficientes kappa ponderados por parejas al error α (comparaciones por parejas sin la corrección de Bonferroni), y un segundo método basado en las comparaciones por parejas con la corrección de Bonferroni ($\alpha' = 2\alpha/\{J(J-1)\}$). A partir de los resultados obtenidos en los experimentos de simulación se propone el siguiente método para comparar múltiples coeficientes kappa ponderados: 1) aplicar el test de hipótesis global basado en la distribución chi-cuadrado al error α , estimando la matriz de varianzas–covarianzas mediante el método delta o mediante bootstrap; 2) si el test de hipótesis conjunto no es significativo al error α , la homogeneidad de los múltiples coeficientes kappa ponderados no se rechaza; pero si el test conjunto es significativo, para buscar las causas de la significación se aplica el método de Bonferroni, es decir se realizan las comparaciones por parejas al error $2\alpha/(J(J-1))$. El método propuesto para resolver

este problema de inferencia se basa en la aproximación normal de la distribución multinomial, por lo que cuando las muestras son pequeñas o relativamente pequeñas ($n < 500$) los resultados obtenidos pueden ser erróneos ya que el error tipo I suele desbordar al error nominal. Una posible solución a este problema es la aplicación de método de inferencia exacta, siendo por tanto necesario investigar métodos de inferencia exacta para resolver el test de hipótesis global para comparar múltiples coeficientes kappa ponderados.

Finalmente, en el Capítulo 3 se estudia el problema del cálculo del tamaño muestral para estimar algunos parámetros de un test diagnóstico binario en presencia de verificación completa y de verificación parcial de la enfermedad. En este Capítulo se han realizado tres aportaciones. La primera consiste en calcular el tamaño muestral para estimar el coeficiente kappa ponderado de un test binario en presencia de verificación completa, la segunda consiste en calcular el tamaño muestral para estimar la sensibilidad (especificidad) de un test binario en presencia de verificación parcial de la enfermedad, y la tercera aportación consiste en calcular el tamaño muestral para comparar las sensibilidades (especificidades) de dos tests binarios en presencia de verificación parcial de la enfermedad. En cada una de las aportaciones se han realizado unos experimentos de simulación para estudiar la bondad del método propuesto para calcular el tamaño muestral. En todos los casos, el método del tamaño muestral se basa en las estimaciones del correspondiente parámetro a partir de una muestra aleatoria inicial y en el intervalo de confianza de Wald del correspondiente parámetro, por lo que el método puede presentar algunos inconvenientes cuando la muestra inicial es pequeña y el valor del estimador está cercano a cero o a uno. De nuevo la investigación de métodos de inferencia exacta puede ser una solución a este problema de especial relevancia.

Bibliografía

Agresti, A., (2002). *Categorical Data Analysis*. John Wiley and Sons, New York.

Agresti, A., Coull, B.A., (1998). Approximate is better than 'exact' for interval estimation of binomial proportions. *The American Statistician*, **52**:119 – 126.

Agresti, A., Min, Y., (2005). Sample improved confidence intervals for comparing matched proportions. *Statistics in Medicine*, **24**: 729 – 740.

Alonzo, T.A., Pepe, M.S., Moskowitz, C.S., (2002). Sample size calculations for comparative studies of medical tests for detecting presence of disease. *Statistics in Medicine*, **21**: 835 – 852.

Arkin, C.F., Wachtel, M.S., (1990). How many patients are necessary to assess test performance? *Journal of the American Medical Association*, **263**: 275 – 278.

Barratt, A., Irwing, L.M., Glasziou, P.P., Cumming, R.G., Raffle. A., Hicks. N., Gray, J.A., Guyatt, G.H., (1999). User's guides to the medical literature: XVII. How to use guidelines and recommendations about screening. Evidence-based medicine working group. *Journal of the American Medical Association*, **281**: 2039 – 2034.

Bates, A.S., Margolis, P.A., Evans, N.S., (1993). Verification bias in pediatric studies evaluating diagnostic tests. *Journal of pediatrics*, **122**(44): 585 – 590.

Begg, C.B., Greenes, R.A., (1983). Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics*, **39**: 207 – 215.

Begg, C.B., McNeil, B.J., (1988). Assessment of radiologic tests: control of bias and other design considerations. *Radiology*, **167**(2): 565 – 569.

- Bennett, B.M., (1972). On comparison of sensitivity, specificity and predictive value of a number of diagnostic procedures. *Biometrics*, **28**: 793 – 800.
- Bennett, B.M., (1983). Further results on indices of diagnostic screening. *Biometrical Journal*, **24**: 59 – 62.
- Berry, G., Smith, C.L., Macaskill, P., Irwig, L., (2002). Analytic methods for comparing two dichotomous screening or diagnostic tests applied to two populations of differing disease prevalence when individuals negative on both tests are unverified. *Statistics in Medicine*, **21**: 853 – 862.
- Bloch, D.A., (1997). Comparing two diagnostic tests against the same “gold standard” in the same sample. *Biometrics*, **53**: 73 – 85.
- Blumenthal, S., (1968). Multinomial sampling with partially categorized data. *Journal Amer. Statist. Assoc.*, **63**: 542 – 551.
- Brown, L.D., Cai, T.T., Dasgupta, A., (2001). Interval estimation for a binomial proportion. *Statistical Science*, **16** (2):101 – 133.
- Carroll R.J., Ruppert, D., Stefanski, L.A., (1995). *Measurement error in non-linear models*. Chapman and Hall, London.
- Cecil, M.P., Kosinski, A.S., Jones, M.T., Taylor, A., Alazraki, N.P., Pettigrew, R.I., Weintraub, W.S., (1996). The importance of work-up (verification) bias correction in assessing the accuracy of SPECT thallium-201 testing for the diagnosis of coronary artery disease. *Journal of Clinical Epidemiology*, **49**: 735 –742.
- Cohen, J., (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, **20**: 37 – 46.
- Dempster, A., Laird, N., Rubin, D., (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *J. Roy Stat Soc Ser B Methodol*, **39**: 1 – 38.
- Efron B., Tibshirani R., (1993). *An introduction to the Bootstrap*. Chapman & Hall: London.
- Feinstein, A.R., (1975). Clinical biostatistics on the sensitivity, specificity, and discrimination of diagnostic tests. *Clin Pharmacol Theory*, **17**: 926 – 930.

- Fosgate, G.T., (2005). Modified exact sample size for a binomial proportion with special emphasis on diagnostic test parameter estimation. *Statistics in Medicine*, **24**: 2857 – 2866.
- Gart, J.J., Nam J., (1988). Aproximate interval estimation of the ratio of binomial parameters: a review and corrections for skewness. *Biometrics*, **44**: 323 – 338.
- Greenes R., Begg C.B., (1985). Assessment of diagnostic technologies: methodology for unbiased estimation from samples of selective verified patients. *Investigative Radiology*, **20**: 751 – 756.
- Hand, D.J., (1987). Screening vs prevalence estimation. *Journal of Royal Statistical Society Series C Applied Statistics*, **36**: 1 – 7.
- Harel, O., Zhou, X.H., (2006). Multiple imputation for correcting verification bias. *Statistics in Medicine*, **25**: 3769 – 3786.
- Harel, O., Zhou, X.H., (2007a). Multiple imputation for the comparison of two screening tests in two-phase Alzheimer studies. *Statistics in Medicine*, **26**: 2370 – 2388.
- Jamart, J., (1993). Letter to the editor: on tests for equality of predictive values for t diagnostic procedures. *Statistics in Medicine*, **12**: 185 – 186.
- Kosinski, A.S., Barnhart, H.X., (2003). A global sensitivity analysis of performance of a medical diagnostic test when verification bias is present. *Statistics in Medicine*, **22**: 2711 – 2721.
- Kraemer, H.C., (1992). *Evaluating medical tests. Objective and quantitative guidelines*. Sage Publications: Newbury Park.
- Kraemer, H.C., Periyakoil, V.S., Noda, A., (2002). Kappa coefficients in medical research. *Statistics in Medicine*, **21**: 2109 – 2129.
- Landis, J.R., Koch., G.G., (1977). The measurement of observer agreement for categorical data. *Biometrics*, **33**: 159 – 174.
- Leisenring, W., Alonzo, T.A., Pepe, M.S., (2000). Comparison of predictive values of binary medical diagnostic tests for paired designs. *Biometrics*, **56**: 345 – 351.
- Li, J., Fine, J. (2004). On sample size for sensitivity and specificity in prospective diagnostic studies. *Statistics in Medicine*, **23**: 2537 – 2550.

- Marín Jiménez, A. E., (2009). *Inferencia exacta y asintótica para parámetros de tests diagnósticos discretos en presencia de verificación parcial*. Tesis Doctoral, Universidad de Granada.
- Martín Andrés, A., Luna del Castillo, J.D., (2004). *Bioestadística para las Ciencias de la Salud (+)*. 1ª Edición (5ª). Ed. Norma-Capitel: Madrid.
- Martín Andrés, A., Sánchez Quevedo, M.J., Tapia García, J.M., Silva Mato, A., (2005). On the validity condition of the chi-squared test in 2×2 tables. *Test* **14**: 99 – 128.
- McCullagh, P., Nelder, J.A., (1989). *Generalized linear models*. 2 Ed. Chapman and Hall: New York, USA.
- McNeil, B., Adelstein, J., (1976). Determining the Value of Diagnostic and Screening Tests. *Journal of Nuclear Medicine*, **17**(6): 439 – 448.
- Mendenhall, W., Lehmann, E. H. Jr., (1960). An Approximation to the Negative Moments of the Positive Binomial useful in Life-testing. *Technometrics*, **2**: 227 – 241.
- Meng, X.L., Rubin, D.B., (1991). Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *Journal of the American Statistical Association*, **86**: 899 – 909.
- Newcombe, R.G., (1998a). Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistic in Medicine*, **17**: 857 – 872.
- Newcombe, R.G., (1998b). Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine*, **17**: 873 – 890.
- Newcombe, R.G., (1998c). Improved confidence intervals for the difference between binomial proportions based on paired data. *Statistics in Medicine*, **17**: 2635 – 2650.
- Newcombe, R.G., (2001a). Simultaneous comparison of sensitivity and specificity of two tests in the paired design. A straightforward graphical approach. *Statistics in Medicine*, **20**: 907 – 915.
- Newcombe, R.G., (2001b). Estimating the difference between differences: measurement of additive scale interaction for proportions. *Statistics in Medicine*, **20**: 2885 – 2893.
- Obuchowski, N.A., (1998). Sample size calculations in studies of test accuracy. *Statistical Methods in Medical Research*, **7**: 371 – 392.

- Pepe, M.S. (2003). *The Statistical Evaluation of Medical Tests of Classification and Prediction*. Oxford University Press: Oxford.
- Pickles, A., Dunn, G., Vázquez-Barquero, J.L., (1995). Screening for stratification in two-phase epidemiological survey. *Statistical Methods in Medical Research*, **4**:73 – 89.
- Philbrick, J T, Horwitz, R.I., Feinstein, A.R., (1980). Methodologic problems of exercise testing for coronary artery disease: groups, analysis and bias. *American Journal of Cardiology*, **46**(5): 807 – 812,
- Ransohoff D.J., Feinstein A.R., (1978). Problems of spectrum an bias in evaluating the efficacy of diagnostic tests. *The New England Journal of Medicine* **299**: 926 – 930.
- Rao, C.R. (1968). *Linear statistical inference and its applications*. John Wiley and Sons: New York.
- Reid M.C., Lachs M.S., Feinstein A.R., (1995). Use of methodological standards in diagnostic test research: Getting better but still no good. *JAMA*, **274**: 645 – 651.
- Roldán Nofuentes, J.A., (2003): *Estimación y contrastes de hipótesis sobre parámetros de tests diagnósticos discretos en presencia de verificación parcial*. Tesis Doctoral, Universidad de Granada.
- Roldán Nofuentes, J.A., Luna del Castillo, J.D., (2005). Comparing the likelihood ratios of two binary diagnostic tests in the presence of partial verification. *Biometrical Journal*, **47**: 442 – 457.
- Roldán Nofuentes, J.A., Luna del Castillo, J.D., (2006). Comparing two binary diagnostic tests in the presence of verification bias. *Computational Statistics and Data Analysis*, **50**: 1551 – 1564.
- Roldán Nofuentes, J.A., Luna del Castillo, J.D., (2007a). The effect of verification bias in the naive estimators of accuracy of a binary diagnostic test. *Communications in Statistics - Simulation and Computation*, **36**: 959 – 972.
- Roldán Nofuentes, J.A., Luna del Castillo, J.D., (2007b). Risk of error and kappa coefficient of a binary diagnostic test in presence of partial verification. *Journal of Applied Statistics*, **34**: 887 – 898.

Roldán Nofuentes, J.A., Luna del Castillo, J.D., (2007c). EM algorithm for comparing two binary diagnostic tests when not all the patients are verified. *Journal of Statistical Computation and Simulation*, **78**: 19 – 35.

Roldán Nofuentes J.A., Luna del Castillo J.D., (2007d) Comparing of the likelihood ratios of two binary diagnostic tests in paired designs. *Statistics in Medicine*, **26**: 4179 – 4201.

Roldán Nofuentes, J.A., Luna del Castillo, J.D., (2008). The effect of verification bias on the comparison of predictive values of two binary diagnostic tests. *Journal of Statistical planning and Inference*, **138**: 959 – 963.

Roldán Nofuentes, J.A., Luna del Castillo, J.D., Femia Marzo, P., (2009). Computational methods for comparing two binary diagnostic tests in the presence of partial verification. *Computational Statistics*, **24**: 695 – 718.

Roldán Nofuentes, J.A., Luna del Castillo, J.D., Montero Alonso, M.A., (2009a). Determining sample size to evaluate and compare the accuracy of binary diagnostic tests in the presence of partial disease verification. *Computational Statistics and Data Analysis*, **53**(3): 742 – 755.

Roldán Nofuentes J.A., Luna del Castillo J.D., Montero Alonso, M.A., (2009b). Confidence intervals of weighted kappa coefficient of a binary diagnostic test. *Communications in Statistics. Simulation and Computation*, **38**: 1562 – 1578.

Rubin, D.B., (1976). Inference and missing data. *Biometrika*, **4**: 73 – 89.

Rubin D.B., (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley: New York.

Rubin D.B., Schenker, N., (1987). Logit-based interval estimation for binomial data using the Jeffreys prior. *Sociological Methodology*, **17**: 131 – 144.

Serfling, R.J., (1980). *Approximation Theorems of Mathematical Statistics*. Wiley: New York.

Simel D.L., Samsa, G.P., Matchar, D.B., (1991). Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J. Clin Epidemiology*, **44**(8): 763-770.

Sox H.C. Jr, Blatt M.A., Higgins, M.C., Marton, K.I., (1989). *Medical decision making*. Butterworths-Heinemann: Boston.

- Swets, J.A., (1979). ROC analysis applied to the evaluation of medical imaging techniques. *Investigative Radiology*, **14**:109 – 21.
- Swets, J.A., Pickett, R.M., Whitehead, S.F., Getty, D.J., Schnur, J.A., Swets, J.B., Freeman, B.A., (1979). Assesment of diagnostic technologies. *Science*, **4407**: 753-759.
- Tavel M.E., Enas N.H., Woods J.R., (1987). Screening tests for enteropathy in children. *American Journal of Cardiology*, **60**: 1167 – 1169.
- Thibodeau, L.A., (1981). Evaluating diagnostic tests. *Biometrics*, **37**: 801-804.
- Torrance-Rynard, V.L., Walter, S.D., (1997). Effects of dependent errors in the assessment of diagnostic test performance. *Statistics in Medicine*, **16**: 2157 – 2175.
- Vacek, P.M., (1985). The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics*, **41**: 959 – 968.
- Wang, W., Davis, C.S., Soong, S., (2006). Comparison of predictive values of two diagnostic tests from the same sample of subjects using weighted least squares. *Statistics in Medicine*, **25**: 2215 – 2229.
- Weiner, D.A., Ryan, T.J., McCabe, C.H., Kennedy, J.W., Schloss, M., Tristani, F., Chaitman, B.R. y Fisher, L.D., (1979). Exercise stress testing. Correlations among history of angina, ST-segment response and prevalence of coronary-artery disease in the coronary artery surgery study (CASS). *The New England Journal of Medicine*, **301**: 230-235.
- Youden, W.J., (1950). Index for rating diagnostic tests. *Cancer*, **3**: 32 – 35.
- Zehna, P.W., (1966). Invariance of maximum likelihood estimation. *Annals of Mathematical Statistics*, **37**: 744.
- Zhou, X.H., (1993). Maximum likelihood estimators of sensitivity and specificity corrected for verification bias. *Communication in Statistics – Theory and Methods*, **22**: 3177 – 3198.
- Zhou, X.H., (1994). Effect of verification bias on positive and negative predictive values. *Statistics in Medicine*, **13**: 1737 – 1745.
- Zhou, X.H., (1998a). Correcting for verification bias in studies of a diagnostic test's accuracy. *Statistical methods in medical research*, **7**: 337 – 353.

Zhou, X.H., (1998b). Comparing accuracies of two screening tests in a two-phase study for dementia. *Journal of Royal Statistical Society Series C Applied Statistics*, **47**: 135 – 147.

Zhou XH., Obuchowski, N.A., McClish, D.K., (2002). *Statistical methods in diagnostic medicine*. John Wiley and Sons: New York.

Zhou, X.H., Qin, G., (2005). A new confidence interval for the difference between two binomial proportions of paired data. *Journal of Statistical Planning and Inference*, **128**: 527 – 542.