



**VNiVERSIDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

**Departamento de Estadística
Máster en Análisis Avanzado de Datos Multivariante
Trabajo de Fin de Máster**

**META-ANÁLISIS DE TESTS DIAGNÓSTICOS PARA EVALUACIÓN DE
ENFERMEDADES CON BAJA PREVALENCIA: UN ESTUDIO COMPARATIVO DESDE
HSROC, MADA Y STATA**

AUTORES:

**SERGIO ALEX BAUZ OLVERA
JOHNY JAVIER PAMBABAY CALERO**

TUTORAS:

**DRA. M^a PURIFICACIÓN GALINDO VILLARDON
DRA. ANA BELÉN NIETO LIBRERO**

2017



A DIOS,
a nuestras familias
y a nuestros amigos.



**VNiVERSiDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

M^a Purificación Galindo Villardon

Profesora titular del Departamento de Estadística de la Universidad de Salamanca

Ana Belén Nieto Librero

Profesora asociada del Departamento de Estadística de la Universidad de Salamanca

CERTIFICA que **D. Sergio Alex Bauz Olvera** y **D. Johny Javier Pambabay Calero** han realizado en la Universidad de Salamanca, bajo su dirección, el trabajo que para optar título de Máster en Análisis Avanzado de Datos Multivariantes presenta con el título **Meta-análisis de tests diagnósticos para evaluación de enfermedades con baja prevalencia: Un estudio comparativo desde HSROC, MADA y STATA**, autorizando expresamente su lectura y defensa.

Y para que conste, firma el presente certificado en Salamanca a 14 de julio de 2016.

M^a Purificación Galindo Villardon

Ana Belén Nieto Librero

AGRADECIMIENTOS

En primer lugar, a las autoridades de la Escuela Superior Politécnica del Litoral - Espol, al señor Rector, a la señora Vicerrectora, Decano, Subdecano y Profesores de la Facultad de Ciencias Naturales y Matemáticas- FCNM por su confianza y apoyo.

A Puri, Directora del Departamento de Estadística de la Universidad de Salamanca, por su guía personal como también científica, Ana Belén Nieto por su orientación y participación activa en el desarrollo de esta investigación y Ana Belén Sánchez que ha sido un motor fundamental para el presente trabajo.

A los profesores y colaboradores del Departamento de Estadística de la Usal por sus enseñanzas, tiempo y colaboración prestada para la realización de este Trabajo de Fin de Máster.

A nuestras madres Noemí†, Eufemia y padres Bibian, Rafael, por sus sabios consejos, valores, que nos han infundido, los cuales nos han permitido ser personas capaces y útiles a la sociedad y en especial por haber estado a nuestro lado en todo momento. A nuestras esposas, Ivonn y Karin, por todos los esfuerzos y sacrificios que han realizado. No hay palabras para describir, cuan agradecidos estamos por su apoyo incondicional recibido durante el Máster.

Por último, a nuestros hijos, Emily, Alejandro y Matías, todavía son muy pequeños, pero con el pasar del tiempo, se darán cuenta que nos trasmitían la fuerza necesaria aún en los momentos más adversos.;

*"La Educación es el pasaporte hacia el futuro,
el mañana pertenece a aquellos que se preparan
para él en el día de hoy"*

Malcolm X

RESUMEN

El crecimiento exponencial de la literatura médica y el uso cada vez más extendido de las tecnologías de la información y la comunicación (TIC), junto con la dispersión de la literatura científica, dificulta al investigador, y sobre todo al profesional sanitario, el acceso a la información relevante. Meta-análisis es una revisión sistemática que incorpora una estrategia estadística para integrar resultados de varios estudios en una sola estimación.

El diagnóstico médico es un proceso dinámico en el que se intenta tomar decisiones idóneas en presencia de incertidumbre. Desde un punto de vista funcional, se considera prueba diagnóstica a cualquier procedimiento realizado para confirmar o descartar un diagnóstico o incrementar o disminuir su verosimilitud. La utilidad de una prueba diagnóstica depende fundamentalmente de su validez y de su fiabilidad, pero también de su rendimiento clínico y de su coste. A menudo se emplean procedimientos diagnósticos de los que se desconocen estos parámetros. Para conocer esta información se debe recurrir a los estudios de evaluación de pruebas diagnósticas publicados. Sin embargo, en muchos de estos estudios no siempre encontramos información válida, relevante o aplicable al entorno clínico Ransohoff and Feinstein [103], Reid et al. [104], Zhou [153].

La sensibilidad y la especificidad son características intrínsecas de la prueba diagnóstica que informan de su validez antes de conocer el resultado. Una vez conocido el resultado de la prueba, son los valores predictivos los que informan sobre la probabilidad de la enfermedad. Sin embargo, los valores predictivos dependen de la prevalencia de enfermedad en el estudio original y por ello no son directamente aplicables a escenarios con probabilidades pre-prueba distintas. Para conocer la probabilidad de enfermedad (post-prueba) en escenarios con probabilidades pre-prueba distintas debemos usar los cocientes de probabilidades (probabilidad de un determinado resultado de la prueba en presencia de enfermedad dividida por la probabilidad de ese mismo resultado en ausencia de enfermedad) que son índices resumen de la sensibilidad y la especificidad, independientes de la probabilidad pre-prueba en el estudio original.

Para poder operar con las razones de verosimilitudes en el cálculo de probabilidades, éstas deben transformarse en ventajas (odds), estos cálculos se simplifican recurriendo a nomogramas (de Fagan). La sensibilidad y la especificidad varían en función del punto de corte que se utilice para separar los valores normales de los anormales. En estos casos, los resultados pueden representarse gráficamente como una

curva ROC, que permite conocer las características de la prueba según diferentes puntos de corte y que puede utilizarse para elegir el más adecuado. Un estimador de la relevancia global de una prueba puede ser el área bajo la curva (AUC). Es preciso tener en cuenta que la información de que se dispone sobre la validez de las pruebas diagnósticas proceden de diferentes poblaciones. Por tanto, las estimaciones obtenidas en dichos estudios están sujetas a variabilidad aleatoria y si los estudios han sido diseñados incorrectamente, a sesgos. La variabilidad de las mediciones va a estar influida por múltiples factores que interesa conocer y controlar. Entre ellos, tiene especial importancia distinguir las variaciones intra e inter estudio.

En el presente trabajo se analiza el modelo jerárquico HSROC (Hierarchical Summary Receiver Operating Characteristic) propuesto por Rutter and Gatsonis [114], el cual en lugar de modelizar directamente S (Sensibilidad) y E (Especificidad), ajusta los datos de clasificación de los estudios mediante una regresión logística de efectos aleatorios. Se compara con la propuesta de Reitsma et al. [105] en la que se propone modelizar los resultados de los estudios mediante una regresión Normal Bivariante (NB), donde S y E se incluyen como efectos aleatorios y se añade un parámetro que refleja la eventual correlación entre ellos. Se presentan los pasos a seguir para llevar a cabo un meta-análisis de tests diagnósticos en situaciones de baja prevalencia (Heterogeneidad (Q , I^2), Sesgo de publicación (Eger's Index), Hair cross y ROCellipses, Parámetros del modelo HSROC y Curva SROC) y se analiza los programas MIDAS (STATA), Dwamena [34]; MADA (R), Doebler [30] y HSROC (R), Schiller et al. [121].

El enfoque SROC es útil en ausencia de heterogeneidad entre estudios (efectos fijos), donde las medidas más representativas de la validez de la prueba están dadas por: razones de verosimilitud, logit de la sensibilidad y especificidad, Q^* , área bajo la curva; el modelo bivariante (efectos aleatorios) estima la correlación que pueda existir entre la sensibilidad y especificidad intra e inter estudios, generando salidas resúmenes como: gráfico forest-plot sensibilidad y especificidad con diamante, gráficos de evaluación de la capacidad del modelo, sensibilidad, especificidad, razones de verosimilitud y diagnostic odds ratio (DOR) con sus respectivos intervalos de confianza, heterogeneidad estadística (I^2), curva SROC con sus respectivas regiones de confianza y predicción, área bajo la curva, funnel-plot, nomograma de Fagan. El enfoque HSROC se caracteriza por modelizar las covariaciones de los parámetros de precisión y los puntos de corte mediante simulaciones de cadena de markov y monte carlo (MCMC), teniendo como salidas principales, las configuraciones generales de la simulación, los tipos de distribuciones utilizadas, resumen de las estimaciones de los parámetros intra estudio, curva SROC con sus respectivas regiones de confianza y predicción, las distribuciones muestrales a posteriori de los parámetros del modelo.

El objetivo de este trabajo es poner de manifiesto que la forma tradicional de valorar la precisión de las pruebas diagnósticas, a partir de tablas tetracóricas, no tiene sentido en este contexto.

Palabras Clave: Meta análisis, Pruebas diagnósticas, SROC, Modelo Bivariante, HSROC, MADA, MIDAS (STATA).

ABSTRACT

The exponential growth of medical literature and the increasing use of Information and Communications technology (ICT), attached to the scientific literature dispersion, make it difficult for researchers especially medical professionals to access relevant information. The Meta-analysis is a systematic revision which adopts an statistical strategy so as to incorporate results of certain studies in a single estimation.

The medical diagnosis is a dynamic process in which you try to take appropriate decisions in the presence of uncertainty. From a functional standpoint, it is considered diagnostic test to any procedure performed to confirm or rule out a diagnosis or increase or decrease its likelihood. The usefulness of a diagnostic test depends primarily on its validity and reliability, but also their clinical performance and cost. Often diagnostic procedures which these parameters are unknown is used. For this information should resort to studies published evaluation of diagnostic tests. However, in many of these studies are not always valid, relevant or applicable information to the clinical setting.

To operate with the reasons likelihoods in the calculation of probabilities, they must become advantages (odds), these calculations are simplified using nomograms (Fagan). The sensitivity and specificity vary depending on the cutoff that is used to separate normal values abnormal. In these cases, the results can be plotted as a ROC curve, which allows to know the characteristics of the test according to different cutoffs and can be used to choose the most appropriate. An estimator of the global importance of a test can be the area under the curve (AUC). It is necessary to note that the information that is available about the validity of diagnostic tests from different populations. Therefore, the estimates obtained in these studies are subject to random variability and whether studies have been improperly designed, to bias. The variability of the measurements will be influenced by many factors that are interested in knowing and controlling. Among them it is particularly important to distinguish variations within and between studies.

The hierarchical HSROC (Hierarchical Summary Receiver Operating Characteristic) model will be analysed as proposed by Rutter and Gatsonis [114], which instead of directly modelling S (sensibilidad) Sensitivity and E (Especificidad) Specificity remodels the classification data of the studies through a logistic regression of random effects. They will be compared with the Reitsma et al. [105] proposal, the studies which propose modelling the study results through a Bivariate Normal (BN) regression in which S and E are included as random effects and a parameter is added, which shows the eventual correlation among

them.

The steps to follow will be submitted in order to carry out a Meta-Analysis of diagnostic tests in low prevalence situations (Heterogeneity (Q , I^2)), publication bias (Eger's Index), Hair Cross and ROC ellipses, HSROC model and SROC Curve parameters and the MIDAS (STATA), Dwamena [34]; MADA (R), Doebler [30] and HSROC (R), Schiller et al. [121] programmes.

The SROC approach is useful in the absence of heterogeneity between studies-effects fixed, where the most representative of the validity of the test measurements are given by: likelihood ratios, logit of sensitivity and specificity, Q^* , area under the curve; the bivariate model (random effects) estimates the correlation of sensitivity and specificity within and between studies, generating summary outputs as graphic forest-plot sensitivity and specificity diamond graphics capability assessment model, sensitivity, specificity, reasons likelihood ratio and diagnostic odds (DOR) with confidence intervals, statistical heterogeneity (I^2), SROC curve with their respective regions of confidence and prediction, area under the curve, funnel-plot, Fagan nomogram. The HSROC approach is characterized by modeling the covariances of the parameters of accuracy and cutoffs by simulating Markov chain and monte carlo (MCMC), with the main outputs, the general settings of the simulation, the types of distributions used, summary estimates of intra study parameters, SROC curve with their respective regions of confidence and prediction, sample posterior distributions of model parameters.

The objective of this work is to explain that the traditional form of rating the diagnostic tests from tetrachoric tables does not have a purpose in this context.

Key Words: Meta-Analysis, Diagnostic Tests, SROC, Bivariate analysis, HSROC, MADA, MIDAS (STATA).

ÍNDICE GENERAL

	Página
ÍNDICE GENERAL	ix
ÍNDICE DE TABLAS	xiii
ÍNDICE DE FIGURAS	xiv
1 INTRODUCCIÓN Y OBJETIVOS	1
1.1 INTRODUCCIÓN	1
1.2 OBJETIVOS	4
1.2.1 OBJETIVO GENERAL	4
1.2.2 OBJETIVOS ESPECÍFICOS	4
2 META-ANÁLISIS	6
2.1 DEFINICIÓN	6
2.2 ORÍGENES Y ANTECEDENTES	6
2.3 META-ANÁLISIS EN LA MEDICINA	7
2.4 ETAPAS O FASES DE UN META-ANÁLISIS	8
2.4.1 FORMULACIÓN DEL PROBLEMA	9
2.4.2 BÚSQUEDA DE LA LITERATURA	10
2.4.2.1 ESTRATEGIA DE BÚSQUEDA	12
2.4.2.2 IDENTIFICACIÓN Y SELECCIÓN DE ESTUDIOS	12
2.4.3 CODIFICACIÓN DE ESTUDIOS	13
2.4.4 MEDIDAS DE LOS RESULTADOS	14
2.4.4.1 VOTE COUTING	15
2.4.4.2 VALOR P	15
2.4.4.3 TAMAÑO DEL EFECTO	15
2.4.4.4 COEFICIENTE DE CORRELACIÓN	15
2.4.4.5 DIFERENCIA DE MEDIAS	16
2.4.4.6 ESTIMACIÓN DE RIESGO	16

2.4.5	ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS	18
2.4.5.1	MODELO DE EFECTOS FIJOS (fixed effects)	18
2.4.5.2	MODELO DE EFECTOS ALEATORIOS(random effects)	18
2.4.5.3	MODELO DE EFECTOS MIXTOS (mixed effects)	19
2.4.5.4	META-REGRESIÓN	19
2.4.5.5	META-PARTICIÓN	19
3	META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS	21
3.1	PRUEBAS DIAGNÓSTICAS	21
3.1.1	CLASIFICACIÓN DE LAS ENFERMEDADES POR DIAGNOSIS	21
3.1.2	CRITERIOS DE CALIDAD DE LAS PRUEBAS	22
3.1.3	DISEÑO DE UNA INVESTIGACIÓN PARA EVALUAR UNA PRUEBA DIAGNÓSTICA	22
3.1.4	PARÁMETROS DE VALORACIÓN DE PRUEBAS DIAGNÓSTICAS	23
3.2	ANÁLISIS DE SENSIBILIDAD Y ESPECIFICIDAD EN LAS MEDIDAS RESUMEN DEL ANÁLISIS DIAGNÓSTICO	35
3.3	MODELOS ESTADÍSTICOS PARA INTEGRACIÓN DE PRUEBAS	38
3.4	ENFOQUE SROC	39
3.4.1	COMPORTAMIENTO EMPÍRICO DE SROC	42
3.4.2	MEDIDAS RESUMEN. ÁREA BAJO LA CURVA Y Q^*	46
3.4.3	ERROR ESTÁNDAR DE AUC Y Q^*	48
3.5	MODELO BIVARIANTE	50
3.6	MODELO DE REGRESIÓN BINOMIAL	53
3.7	REGRESIÓN JERÁRQUICA PARA META-ANÁLISIS DE LAS EVALUACIONES EN LA PRECISIÓN DE LAS PRUEBAS DIAGNÓSTICAS (HSROC)	54
3.7.1	NIVEL I (VARIACIÓN DENTRO DEL ESTUDIO)	54
3.7.2	NIVEL II (VARIACIÓN ENTRE ESTUDIOS)	54
3.7.3	NIVEL III	54
3.7.4	AJUSTE DEL MODELO HSROC	55
3.7.5	ELECCIÓN A PRIORI	56
3.7.6	PARÁMETROS DE ESTIMACIÓN	56
3.7.7	DIAGNÓSTICO	56
3.7.8	RESUMEN ESTADÍSTICO	57
3.8	META-ANÁLISIS BAYESIANO PARA LA PRECISIÓN DE UN TEST EN AUSENCIA DE LA PRUEBA ORO ESTÁNDAR	59
3.8.1	ESTIMACIÓN	60
3.8.2	LA DEPENDENCIA CONDICIONAL ENTRE EL ÍNDICE Y ANÁLISIS DE REFERENCIA	61
3.9	UNIFICACIÓN DE MODELOS PARA META-ANÁLISIS DE LA PRECISIÓN DIAGNÓSTICA	62
3.10	RELACIÓN ENTRE EL MODELO BIVARIANTE Y HSROC	62
3.11	RELACIÓN DE LOS PARÁMETROS DE LOS MODELOS SIN COVARIANZA	65
3.12	RELACIÓN DE LOS PARÁMETROS DE LOS MODELOS CON COVARIANZA	65
3.13	ENFOQUE DE INFERENCIA Y SALIDAS DE LOS MODELOS	66
3.13.1	MODELO HSROC	66
3.13.2	MODELO BIVARIANTE	66

3.14	MODELO ESTADÍSTICO PARA I^2	67
4	RESULTADOS	77
4.1	PRECISIÓN DIAGNÓSTICA CON MADA PARA META-ANÁLISIS	77
4.1.1	RESUMEN	77
4.1.2	INTRODUCCIÓN	77
4.1.3	OBTENIENDO EL PAQUETE	78
4.1.4	INGRESANDO LOS DATOS	78
4.1.5	CELDA VACÍAS	80
4.1.6	ESTADÍSTICA DESCRIPTIVA	80
4.1.7	GRÁFICOS DESCRIPTIVOS	83
4.1.8	ENFOQUE UNIVARIANTE	85
4.1.9	ODDS RATIO DE DIAGNÓSTICO (DOR)	85
4.1.10	MODELO PROPORCIONAL HAZARDS	87
4.1.11	ENFOQUE BIVARIANTE	90
4.1.12	COMPARANDO CURVAS SROC	93
4.1.13	META-REGRESIÓN BIVARIANTE	94
4.2	PRECISIÓN DIAGNÓSTICA CON STATA PARA META-ANÁLISIS	96
4.2.1	DESCRIPCIÓN	96
4.2.2	SINTAXIS DE LAS ÓRDENES DE STATA	98
4.2.3	GESTIONES DE BASE DE DATOS	99
4.2.4	TIPOS DE DATOS	99
4.2.5	CARGAR UN ARCHIVO DE DATOS EN FORMATO STATA	99
4.2.6	OPCIONES EN STATA	100
4.2.6.1	MODELAMIENTO Y OPCIONES DE POST-ESTIMACIÓN	100
4.2.6.2	OPCIONES DE EVALUACIÓN DE LA CALIDAD	101
4.2.6.3	GRÁFICOS EXPLORATORIOS	101
4.2.6.4	OPCIONES DEL REPORTE PRINCIPAL	101
4.2.6.5	OPCIONES DE FOREST PLOT	101
4.2.6.6	OPCIONES DE LA CURVA ROC	102
4.2.6.7	OPCIONES DE HETEROGENEIDAD	102
4.2.6.8	OPCIONES DEL SESGO DE PUBLICACIÓN	102
4.2.6.9	OPCIONES DE UTILIDAD CLÍNICA	102
4.2.6.10	OPCIONES DE MISCELÁNEOS	103
4.2.7	GUARDANDO LOS RESULTADOS	103
4.2.8	EJEMPLO CON UN CONJUNTO DE DATOS (DATASET)	104
4.2.9	MODELOS DE PREDICCIÓN Y DIAGNÓSTICO	108
4.2.10	EVALUACIÓN DE LA CALIDAD	110
4.2.11	ASOCIACIÓN BIVARIANTE	113
4.2.12	RESÚMENES DE LAS ESTIMACIONES DE RENDIMIENTO	113
4.2.13	HETEROGENEIDAD ESTADÍSTICA	114
4.2.14	FOREST-PLOT PARA ANALIZAR LA SENSIBILIDAD Y ESPECIFICIDAD	115
4.2.15	CURVA SROC	116
4.2.16	META-REGRESIÓN	118
4.2.17	TEST DE REGRESIÓN LINEAL DE ASIMETRÍA FUNNEL PLOT	120
4.2.18	FAGAN PLOT (NOMOGRAMA DE BAYES)	121

4.2.19	DIAGRAMA DE DISPERSIÓN DE LOS ODDS RATIO DE MÁXIMA VEROSIMILITUD	122
4.2.20	VALORES PREDICTIVOS Y GRÁFICO DE PROBABILIDAD MODIFICATORIA	123
4.3	PRECISIÓN DIAGNÓSTICA CON HSROC PARA META-ANÁLISIS	124
4.3.1	INTRODUCCIÓN	124
4.3.2	META-ANÁLISIS EN PRESENCIA DE UNA PRUEBA ORO DE REFERENCIA ESTÁNDAR ASUMIENDO INDEPENDENCIA CONDICIONAL	125
4.3.3	PREPARACIÓN DE LOS DATOS	125
4.3.4	EJECUTANDO LA SIMULACIÓN CADENAS DE MARKOV MONTE CARLO (MCMC)	127
4.3.5	INTERPRETANDO LOS ARCHIVOS DE SALIDA	128
4.3.6	ESTADÍSTICA DESCRIPTIVA	128
4.3.7	RESUMEN GRÁFICO	130
5	Discusión	134
5.1	ELECCIÓN DE LAS MEDIDAS	134
5.2	COMPARACIÓN DE LOS ÍNDICES DE LAS PRUEBAS Y ANÁLISIS DEL EFECTO DE LA COVARIACIÓN	135
5.3	PROPIEDADES ESTADÍSTICAS	135
5.4	ESTUDIOS DE LA PRECISIÓN DE DIAGNÓSTICO: LA CALIDAD DEL DISEÑO Y ELABORACIÓN DE INFORMES	136
5.5	R Y STATA	136
6	Conclusiones	138
	ANEXO	139
A	Estadística Descriptiva de los Parámetros a estimar	141
B	Gráficos de dispersión, distribución muestral a posteriori, datos MRI	145
C	Gráficos de densidad de los parámetros inter-estudio, datos MRI	152
	BIBLIOGRAFÍA	159

ÍNDICE DE TABLAS

1.1	Revisiones narrativas vs. Revisiones sistemáticas. Adaptada de Cook (1997)	2
2.1	Estrategias de Búsquedas en Pub-Med sobre Meta-Análisis	7
2.2	Definiciones de riesgos y Odds Ratio	17
2.3	Presencia o ausencia del factor de riesgo	17
3.1	Tabla de doble entrada, ejemplo hipotético	24
3.2	Tabla resumen de datos	38
3.3	AUC_{hom} , Q^* , y sus diferencias para varios valores de odds ratio diagnóstico: caso homogéneo	48
3.4	Resumen de modelos estadísticos en pruebas diagnósticas	72
3.5	Resumen de las funciones empleadas en R para enfoque univariante	73
3.6	Resumen de las funciones empleadas en R para enfoque bivariante	73
3.7	Resumen de las funciones empleadas en Stata para enfoque bivariante	74
3.8	Resumen de las funciones empleadas en R para enfoque jerárquico-HSROC	75
3.9	Resumen de las principales medidas y gráficos para el enfoque univariante, bivariante y HSROC	76
4.1	Datos del estudio i-ésimo en una tabla de 2×2	78
4.2	Orden de las variables a,b,c y d	99
4.3	Tipos de datos en Stata	99
4.4	Interpretación valores AUROC	117

ÍNDICE DE FIGURAS

1.1	Tipos de revisiones según Sackett et al. [116]	3
1.2	Tipos de revisiones. Adaptado de Madhukar Pai	3
2.1	Número de publicaciones sobre Meta-análisis en PubMed (1977-2016)	8
2.2	Etapas de meta-análisis	9
2.3	Metodología PICO	9
2.4	Población objetivo y población accesible en un meta-análisis	11
2.5	Selección de Estudio - Diagrama de flujo	12
2.6	Interpretación coeficiente Kappa.	13
2.7	Diagrama que resume la metodología meta-partición.	20
3.1	Criterio de calidad de las pruebas	22
3.2	Diseño para evaluar una prueba diagnóstica	23
3.3	Variación de la razón de verosimilitud en función de la sensibilidad para diferentes pruebas cuya especificidad es igual a 0,90	29
3.4	Variación de las razones de verosimilitud en función de la especificidad para diferentes pruebas cuya sensibilidad es igual a 0.80	29
3.5	Especificidad Vs Sensibilidad	30
3.6	Especificidad Vs Sensibilidad	30
3.7	Cocientes de Probabilidad, según Evidence-bases Medicine Working Group	31
3.8	Gráfica de Fagan, datos midas_example	32
3.9	Curva ROC, datos simulados	34
3.10	Curva SROC con varios valores de α ($\beta = 0$), datos simulados	43
3.11	Curva SROC con varios valores de β ($\alpha = 2$), datos simulados	44
3.12	Diagrama de enfoques HSROC, Bivariante, SROC, elaboración propia	70
3.13	Esquema metodológico para el meta-análisis en pruebas diagnósticas	71
4.1	Gráficos forest-plot, para sensibilidad y especificidad, datos AuditC	84
4.2	Gráfico ponderado crosshair con colores arbitrarios y regiones de confianza en forma de elipse, datos AuditC	85

4.3	Forest-plot, logaritmos DOR's, ajuste DSL, datos AuditC	87
4.4	Resumen gráfico con modelo PHM, datos AuditC	90
4.5	Curva SROC, modelo de Reitsma, datos AuditC	93
4.6	Comparación de curvas SROC para SAQ (entrevistados) y IAQ (entrevistadores), datos IAQ y SAQ	94
4.7	Ambiente de trabajo en STATA	97
4.8	Menú y paneles en Ambiente de trabajo STATA	98
4.9	Cargar archivo desde barra de Menú en Stata	100
4.10	Comando para leer en disco duro la ruta del archivo <i>midas_example</i>	104
4.11	Acceso a <i>midas_example.dta</i> desde las pantallas	105
4.12	Comando para usar los datos de <i>midas_example.dta</i>	106
4.13	Instalación de Paquete midas desde el repositorio oficial	108
4.14	Bayes empírico predicho frente a los resultados de las pruebas observadas, datos midas_example	109
4.15	Representación gráfica de la normalidad bivariante, bondad de ajuste basada en residuales y los análisis de detección de influencia y atípicos, datos midas_example	110
4.16	Instalación del paquete QUADAS desde el repositorio oficial	111
4.17	Gráfico de barras de la evaluación de la calidad para cada artículo, incluidos los criterios modificados QUADAS, datos midas_example	112
4.18	Diagrama de caja bivariante de los estudios agrupados con siete valores atípicos que sugiere indirectamente un menor grado de heterogeneidad, datos midas_example	113
4.19	Salida de la estadística de heterogeneidad, datos midas_example, elaboración propia	115
4.20	Diagrama forest plot para cada estudio específico (eje derecho) y la media de sensibilidad y especificidad con las estadísticas de heterogeneidad correspondientes, datos midas_example	116
4.21	Resumen de la curva ROC con las regiones de confianza y predicción alrededor de la media de sensibilidad y el punto de especificidad, datos midas_example	117
4.22	Forest plot de una meta-regresión univariante múltiple, datos midas_example	119
4.23	Gráfico Funnel Plot con la línea de regresión superpuesta, datos midas_example	121
4.24	Gráfica de Fagan, datos midas_example	122
4.25	Razón de verosimilitud y diagrama de dispersión, datos midas_example	123
4.26	Gráfica de Probabilidad modificada, datos midas_example	124
4.27	Lista de configuraciones generales de la simulación	128
4.28	Resumen de las distribuciones utilizadas	129
4.29	Resumen de las distribuciones utilizadas	129
4.30	Resumen de las estimaciones de los parámetros intra-estudio	130
4.31	Gráfico de dispersión, de la distribución muestral a posteriori de un sólo parámetro, datos MRI	131
4.32	Gráficos de densidad de algunos de los parámetros inter-estudios, datos MRI.	132
4.33	Curva SROC, datos MRI	133

CAPÍTULO 1

INTRODUCCIÓN Y OBJETIVOS

”En esto descubrieron treinta o cuarenta molinos de viento que hay en aquel campo, y así como Don Quijote los vió, dijo a su escudero: la ventura va guiando nuestras cosas mejor de lo que acertáramos a desear; porque ves allí, amigo Sancho Panza, donde se descubren treinta o poco más desaforados gigantes con quien pienso hacer batalla...”

Miguel de Cervantes Saavedra,
El Ingenioso Hidalgo Don Quijote de la
Mancha

1.1 INTRODUCCIÓN

La necesidad de manejar grandes volúmenes de datos, con características importantes como la falta de actualización de la información, o la pobre calidad de las fuentes de información no son situaciones desconocidas en el campo médico-sanitario, por lo que obligan al experto a investigar artículos de revisión en lugar de artículos originales. Es así, que las revisiones proveen las herramientas necesarias para informarse y actualizarse en el menor tiempo posible. En la práctica se ha desarrollado una revisión, la revisión narrativa, la cual hace uso de métodos informales y hace un análisis muy subjetivo, Sáenz and Ausejo [119]. En las revisiones narrativas existen defectos, tales como que los investigadores no especifican el procedimiento que se ha utilizado para focalizar y evaluar los resultados que han ayudado a obtener sus conclusiones y la más importante es que no se es capaz de reproducir, verificar y contrastar las conclusiones de las revisiones Sáenz and Ausejo [119]. Al nombrar el término de revisiones hay que tener claro que existen dos tipos, revisión y revisión sistemática, Sackett et al. [116]. La primera hace referencia a todo intento que sintetice los resultados y las conclusiones de dos o más publicaciones relacionadas con un tema dado, la segunda revisión, se centra en la identificación exhaustiva a través de toda la literatura de un tema dado, valorando su calidad y sintetizando sus resultados.

Característica	Revisión Narrativa	Revisión Sistemática
Pregunta de investigación	Amplia. No estructurada	Estructurada, clara, concreta y centrada en un problema clínico bien definido.
Búsqueda bibliográfica. Selección de fuentes de información.	No detallada. No sistemática. No orientada a localizar todos los estudios disponibles. Probabilidad alta de sesgo	Búsqueda detallada, sistemática y explícita.
Selección de artículos.	No hay criterios de selección. No reproducible. Probabilidad alta de sesgo.	Selección basada en criterios explícitos. Aplicación uniforme de los criterios de selección, exclusión a todos los artículos.
Valoración de la calidad de los estudios.	No hay valoración.	Valoración, evaluación crítica de la calidad metodológica de los estudios.
Síntesis.	A menudo resumen subjetivo, cualitativo, sin un estimador estadístico.	Basada en la calidad metodológica de los estudios. A menudo resumen cuantificado por un estimador estadístico
Interpretación.	A veces basada en la evidencia. Frecuentemente basada en opiniones personales.	Generalmente basada en la evidencia

Tabla 1.1: Revisiones narrativas vs. Revisiones sistemáticas. Adaptada de Cook (1997)

Las diferencias entre las revisiones narrativas y revisiones sistemáticas aparecen recogidas en la Tabla 1.1.

El creciente volumen de información disponible, ha motivado la creación de técnicas que ayuden a los médicos a filtrar y extraer información de calidad y clínicamente válida como la lectura crítica, Oxman et al. [96] y las revisiones sistemáticas.

Las características esenciales de toda revisión sistemática, Martín Rodero [86], son las siguientes:

- Objetivos claramente definidos con criterios, previamente establecidos, de selección (inclusión/exclusión) de los estudios.
- Metodología explícita y reproducible.
- Búsqueda sistemática y exhaustiva que trate de identificar todos los estudios que cumplan los criterios de elegibilidad.
- Evaluación de la validez de los resultados de los estudios incluidos, por ejemplo a través de la evaluación del riesgo de sesgo.
- Síntesis y presentación sistemática de las características y resultados de los estudios incluidos.

Una de las fortalezas del método científico es que una investigación bajo este contexto debe ser replicable. Es por tal motivo que entre lo que se ha estudiado en el pasado y las futuras investigaciones se tome en consideración la revisión de la literatura y la resolución de las contradicciones empíricas para así validar la consistencia o robustez de un determinado hecho empírico, Galindo-Villardón [40]. Sackett et al. [117] presenta definiciones claras de revisiones narrativas, sistemáticas y meta-análisis, ver Figura

1.1. A partir de la Figura 1.1, se podría expandir hacia un cuarto nivel los tipos de revisiones, ver Figura 1.2, en la cual se encontrarían los meta-análisis de datos individuales de pacientes, Martín Rodero [86].

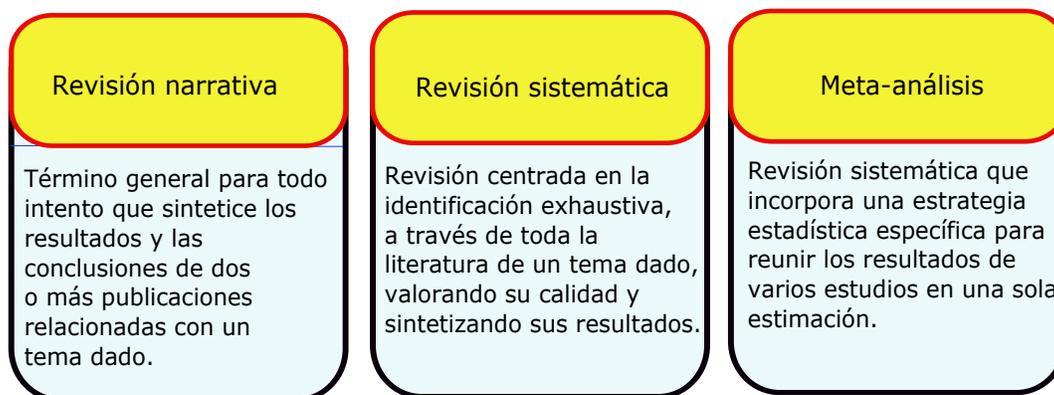


Figura 1.1: Tipos de revisiones según Sackett et al. [116]



Figura 1.2: Tipos de revisiones. Adaptado de Madhukar Pai

En un meta-análisis se diferencian 6 etapas, Rosenthal [110], todas ellas esenciales. En este trabajo centraremos la atención en la fase de integración de resultados y abordaremos la problemática particular

asociada a las pruebas diagnósticas utilizadas en screening masivos de enfermedades de baja prevalencia, pero cuyo diagnóstico precoz es esencial, es decir, cuando los verdaderos positivos son muy bajos en comparación con los verdaderos negativos (autismo, por ejemplo).

En este trabajo se pondrá de manifiesto que la forma tradicional de valorar la precisión de las pruebas diagnósticas, a partir de tablas tetracóricas, no tiene sentido en este contexto, ya que el porcentaje de aciertos en el diagnóstico va a ser altísimo incluso cuando los Verdaderos Positivos (VP) sean muy bajos y/o inferiores a los Falsos Positivos (FP). Tampoco tiene sentido utilizar el OR (conocido como DOR en pruebas diagnósticas) como estimador del tamaño del efecto. Los valores de los DOR pueden llegar a valer cantidades del orden de 500, 1000, 2000, etc., o sea, valores carentes de interpretación clínica. Además, los distintos instrumentos que se utilizan (tests) pueden utilizar puntos de corte diferentes y es frecuente encontrar alta heterogeneidad en la sensibilidad y especificidad de los estudios. Por tanto, todo lo que da excelentes resultados en meta-análisis de Ensayos clínicos y/o en pruebas diagnósticas de enfermedades de alta prevalencia, no es válido en este contexto. El método de Agregación Directa ignora factores de estudio como la prevalencia, y la relación entre S (Sensibilidad) y E (Especificidad), un factor imprescindible para comprender el comportamiento del test en el espacio ROC.

El presente trabajo, tiene la siguiente estructura, se comienza por la introducción, luego se enuncian los objetivos del estudio, para posteriormente detallar los conceptos importantes en meta-análisis y pruebas diagnósticas, luego se estudian los métodos estadísticos que se emplean en las pruebas diagnósticas, y finalmente se presentan los resultados y se termina con las conclusiones más relevantes.

1.2 OBJETIVOS

1.2.1 OBJETIVO GENERAL

Poner de manifiesto que la forma tradicional de valorar la precisión de las pruebas diagnósticas, a partir de tablas tetracóricas, no tiene sentido en este contexto, ya que el porcentaje de aciertos en el diagnóstico va a ser altísimo incluso cuando los Verdaderos Positivos (VP) sean muy bajos y/o inferiores a los Falsos Positivos (FP).

1.2.2 OBJETIVOS ESPECÍFICOS

1. Realizar una exhaustiva revisión bibliográfica que nos permita conocer el estado del arte en relación a la evaluación de las pruebas diagnósticas.
2. Conocer las particularidades que presenta la integración de resultados de pruebas diagnósticas cuando los estudios son heterogéneos, presentan diferentes puntos de corte, y no se dispone de prueba oro.
3. Conocer la problemática asociada a los meta-análisis de pruebas diagnósticas utilizadas en enfermedades con baja prevalencia, donde no es posible utilizar la exactitud para valorar la prueba.
4. Realizar un estudio comparativo entre el Modelo de Moses, el modelo Bivariante y el modelo jerárquico HSROC.
5. Presentar los pasos a seguir para llevar a cabo un meta-análisis de tests diagnósticos en situaciones de baja prevalencia, que pueda servir de referencia a los investigadores de este campo.

6. Realizar un estudio crítico comparativo entre los programas MIDAS (STATA) Dwamena [34], el MADA (R) Doebler [30], y el HSROC (R) Schiller et al. [121], por ser los de más uso en este contexto.

Se puede decir que la prueba diagnóstica es un meta-análisis diferente, por tal razón es necesario conocer algunos conceptos importantes que encierra el meta-análisis antes de abordar el estudio de pruebas diagnósticas, en el Capítulo 2 se enuncian algunos puntos importantes del meta-análisis.

CAPÍTULO 2

META-ANÁLISIS

"Nada ocurre porque sí. Todo en la vida es una sucesión de hechos que, bajo la lupa del análisis, responden perfectamente a causa y efecto."

Richard Feynmann

2.1 DEFINICIÓN

La revisión sistemática nos lleva al meta-análisis, Galindo-Villardón [40], el cual puede definirse como el análisis estadístico de los resultados de estudios individuales con el fin de integrar los resultados, es decir, combina información estadística de estudios similares y realiza análisis estadístico de los resultados generales. En resumen, el meta-análisis puede ser definido como una revisión cuantitativa y síntesis de los resultados de estudios relacionados, los cuales son independientes, Martín Rodero [86].

2.2 ORÍGENES Y ANTECEDENTES

Agrupar investigaciones independientes no es algo nuevo, en 1904 Karl Pearson formuló un procedimiento para sintetizar los resultados de estudios que no tenían potencia con el fin de llegar a conclusiones significativas en términos estadísticos, Casares [14]. El objetivo de un meta-análisis es llevar a cabo un análisis sistemático e integral de la información, sobre un conjunto de resultados de investigaciones empíricas que estudian un mismo problema, con el fin de llegar a conclusiones que puedan generalizarse, para optimizar el proceso cognoscitivo e investigativo en la toma de decisiones, cuando la incertidumbre está presente, sea esta científica, empresarial, educacional, social, etcétera, Avilés Merens et al. [3]. En el año 1976 el Psicólogo americano Gene V. Glass, utiliza por primera vez el término meta-análisis para dar a conocer estudios estadísticos de una serie de resultados mediante ensayos clínicos individuales con el fin de integrarlos, Martín Rodero [86]. Mediante la combinación de información a través de diferentes estudios, un análisis integrado tendrá más poder estadístico para detectar el tamaño del efecto que un análisis basado

en un solo estudio. Cuando varios estudios tienen resultados contradictorios, un meta-análisis puede ser utilizado para estimar un efecto promedio de estudios asociados, Normand [94].

2.3 META-ANÁLISIS EN LA MEDICINA

En Medicina, una de las primeras publicaciones que pueden ser calificadas de meta-análisis es un trabajo de Beechers sobre la evaluación del efecto placebo publicado en la revista JAMA en el año 1955, Egger and Smith [35]. Actualmente existe un crecimiento en las publicaciones de meta-análisis, lo que refleja una importancia muy interesante en estas revisiones logrando aceptación en el campo de la medicina, Martín Rodero [86].

Es importante notar como han evolucionado en el tiempo las revisiones en el tema del meta-análisis, así encontramos la Figura 2.1, generada con la estrategia que se observa en la Tabla 2.1.

1. "meta-analysis"[PT] OR "meta-analysis as Topic"[Mesh]
2. meta-analysis[Title/Abstract] OR metaanalysis[Title/Abstract]
3. #1 OR #2

Tabla 2.1: Estrategias de Búsquedas en Pub-Med sobre Meta-Análisis

Nótese en la Figura 2.1, un crecimiento exponencial en las publicaciones de meta-análisis, a partir del año 1979, algo importante que recalcar, es que este comportamiento creciente coincide con el nacimiento y desarrollo del estudio de la medicina basada en la evidencia. El corte para la generación de este gráfico fue la fecha 15 de abril del año 2016.

El meta-análisis en gran medida se ha empleado para analizar y combinar salidas de ensayos clínicos aleatorizados, pero en la última década es frecuente notar aplicaciones importantes en evaluación de pruebas diagnósticas.

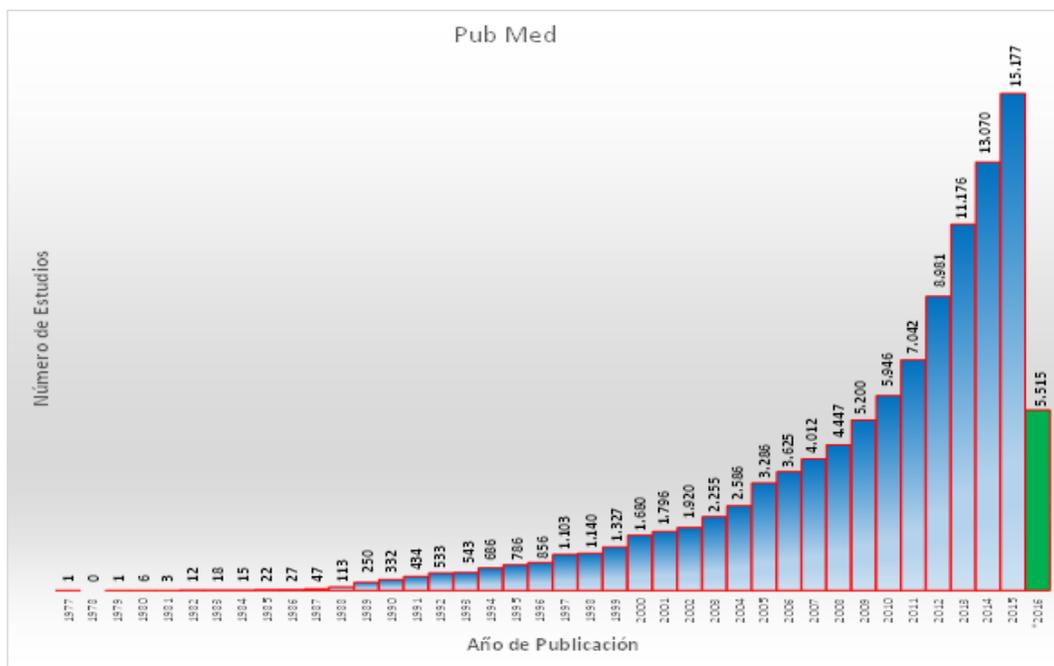


Figura 2.1: Número de publicaciones sobre Meta-análisis en PubMed (1977-2016)

2.4 ETAPAS O FASES DE UN META-ANÁLISIS

El termino meta-análisis es conocido como revisión cuantitativa, integración de investigación o análisis cuantitativo de dominios de investigación. La National Library of Medicine, utiliza el término meta-análisis desde 1989. Para elaborar un meta-análisis robusto y confiable, el mismo debe tener una planificación correcta y ha de contar con un protocolo que determine con claridad los objetivos y las etapas que deben de realizarse, Martín Rodero [86]. El protocolo puede contar de seis fases, Vallejo [140]. Existen cinco etapas por la que debe atravesar una síntesis de investigación, Robert [109], observe Figura 2.2.

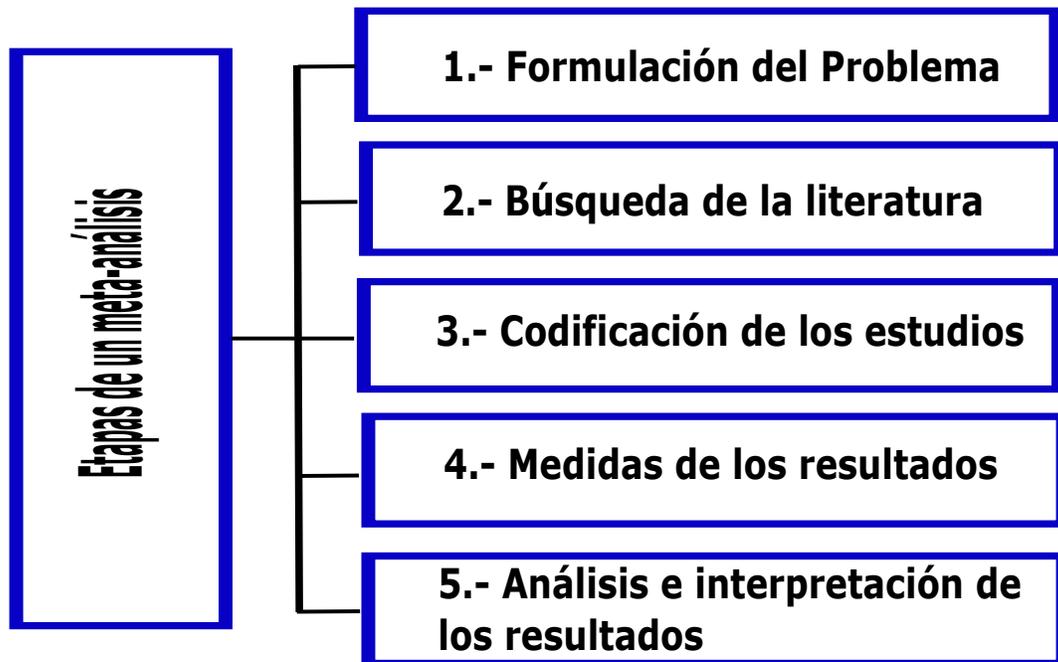


Figura 2.2: Etapas de meta-análisis

2.4.1 FORMULACIÓN DEL PROBLEMA

En este paso se debe establecer con máxima rigurosidad posible el tema sobre que se va a recoger la información, se deben tener claros los objetivos de la investigación y se deben tener definidos los conceptos y las variables de estudio. Para tener un buen tema de investigación se deben contemplar cuatro características de acuerdo a la metodología PICO, Oxman et al. [96], véase Figura 2.3.

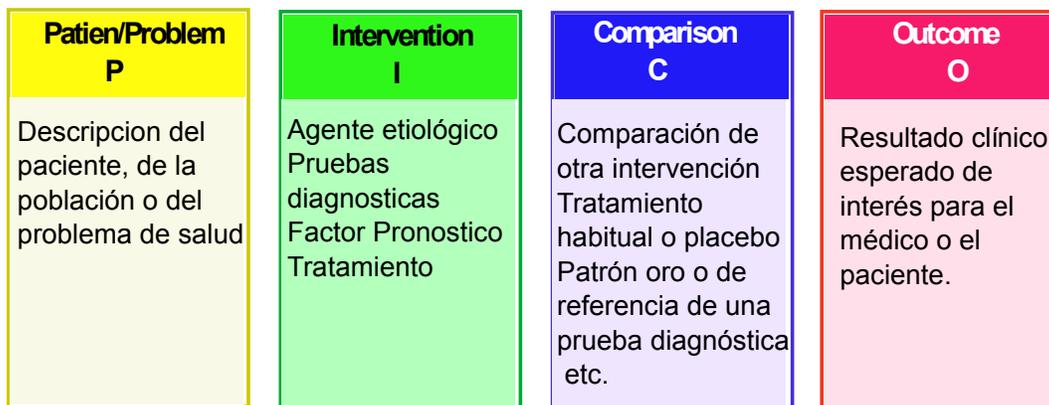


Figura 2.3: Metodología PICO

Es decir, hay que circunscribir el objeto de la revisión, definiendo con precisión las variables y conceptos implicados. Los objetivos con los que debe contar un meta-análisis se establecen de manera clara, Gisbert

and Bonfill [44], es decir:

1. Resumir los resultados de los estudios seleccionados y realizar una valoración cualitativa y cuantitativa.
2. Resolver la incertidumbre
3. Mejorar la precisión del estimador del efecto testado i.e., aumentar el efecto estadístico.
4. Comparar y contrastar los resultados divergentes de los estudios individuales.
5. Responder preguntas nuevas que los estudios originales no habían planteado o no podían responder por no tener suficiente poder estadístico (análisis de subgrupos).
6. Orientar futuros trabajos de investigación, es decir, generar hipótesis. Resumir los resultados de los estudios seleccionados y realizar una valoración cualitativa y cuantitativa.

2.4.2 BÚSQUEDA DE LA LITERATURA

Consiste en identificar todos los estudios que se han realizado sobre el tema de estudio. Es importante en un meta-análisis definir la población objetivo y la población accesible; la primera está constituida por todos los entes o individuos que el revisor espera poder representar en el estudio y la segunda incluye a aquellos individuos que el revisor puede hallar o localizar para el estudio. Se debe tener claro que la población accesible debe ser un subconjunto o una adecuación de la población objetivo, por ello para asegurar esta adecuación se debe contar con dos fuentes de información, fuentes informales y fuentes primarias, sino se cumple lo anterior es decir la población objetivo y accesible no coinciden, observe Figura 2.4. Esto puede suponer una amenaza importante a la validez de los resultados, Galindo-Villardón [40], Martín Rodero [86].

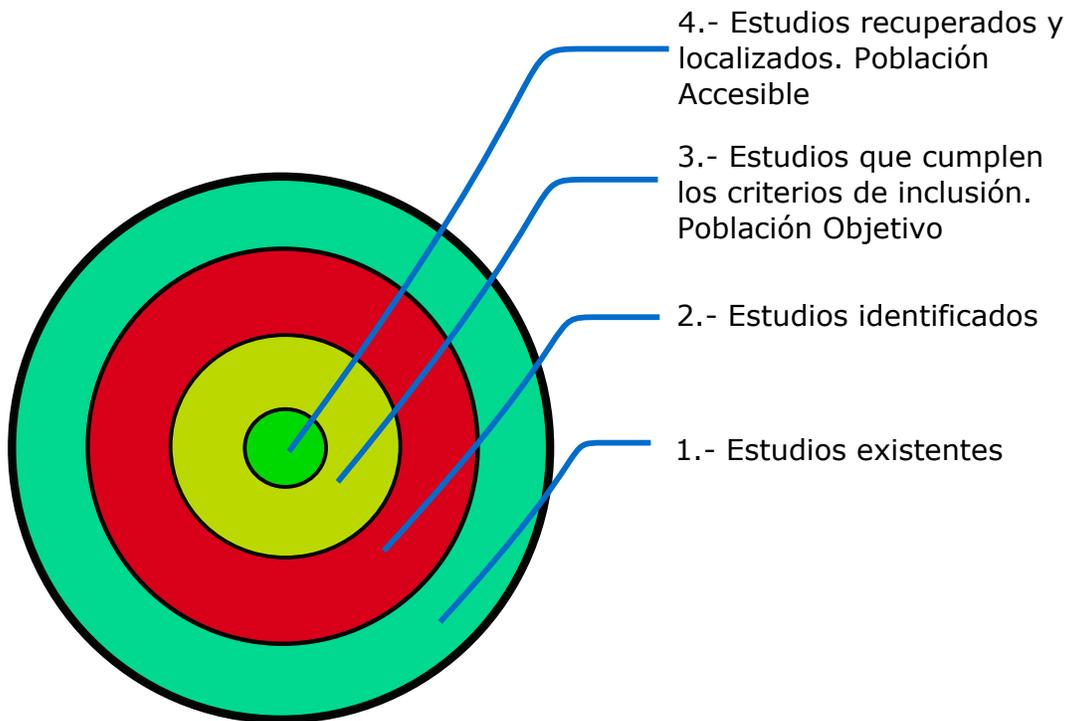


Figura 2.4: Población objetivo y población accesible en un meta-análisis

En las fuentes informales, se pueden encontrar trabajos presentados en congresos profesionales y las conexiones que se hagan con expertos en el tema de interés, con el fin de obtener material nuevo o no publicado por parte de un investigador que trabaje en el campo de estudio; las fuentes primarias son búsquedas o consultas de bibliografías personales y de revistas de investigación, las dos fuentes no son excluyentes entre sí.

En lo relacionado a las localizaciones de estudios existen dos técnicas establecidas, Cooper [20], el procedimiento ascendente y el descendente. En el primero se localizan investigaciones previas en listas de referencias de estudios existentes y disponibles, en cambio en el segundo procedimiento se ubican trabajos consultando índices de citas. La búsqueda por computadoras puede incluir los dos procedimientos antes mencionados, el mismo es más rápido y eficaz.

Existe una crítica a la integración de investigación, y es llamada sesgo de selección editorial a favor de estudios con resultados significativos. Este problema ha sido abordado por Robert [109], pero la solución no es definitiva, la misma se utiliza para estimar la tolerancia (fail-safe number) de un meta-análisis a resultados nulos, es decir, se estima el número de resultados nulos que deben existir en los archivos editoriales para alterar los resultados significativos de un meta-análisis.

En el informe de cualquier meta-análisis se debe especificar claramente la fuente o fuentes de información utilizadas (base de datos electrónicas), periodos de búsquedas y fecha de la última búsqueda, también si fuera el caso se debe incluir las búsquedas manuales realizadas.

2.4.2.1 ESTRATEGIA DE BÚSQUEDA

El procedimiento más utilizado es el efectuado en bases de datos electrónicas. El mismo no es tan simple, ya que existe una gran cantidad de revistas en las diferentes bases de datos. Por tal motivo debe realizarse una búsqueda en las bases de datos bibliográficas como MEDLINE, EMBASE, SCISEARCH, también existe otras bases de datos muy personalizadas como Current Contents, Cochrane Library, National Research Register, etc, y otras transversales como Web of Science, Martín Rodero [86].

2.4.2.2 IDENTIFICACIÓN Y SELECCIÓN DE ESTUDIOS

Los parámetros de inclusión y exclusión de los estudios deben estar establecidos en un protocolo, los mismos deberán estar enmarcados dentro de los objetivos propuestos en el estudio. Se debe considerar, tamaño de la muestra, diseño del estudio, etc. Asimismo se debe decidir el número de evaluadores, la especialización de los mismos y la forma de resolver las disputas u opiniones contrarias que se produzcan. Es importante que en todas la fases de localización y selección de los estudios se debe documentar de manera correcta mediante un diagrama de flujo los artículos seleccionados en cada fase, Martín Rodero [86], note la Figura 2.5. Si existiera diferencia entre los evaluadores, se aplica el índice estadístico kappa, el mismo mide la concordancia de ambos evaluadores, es decir, hasta qué punto los evaluadores coinciden en su evaluación, Martín Rodero [86], el cual se puede interpretar de la forma que se explica en la Figura 2.6.

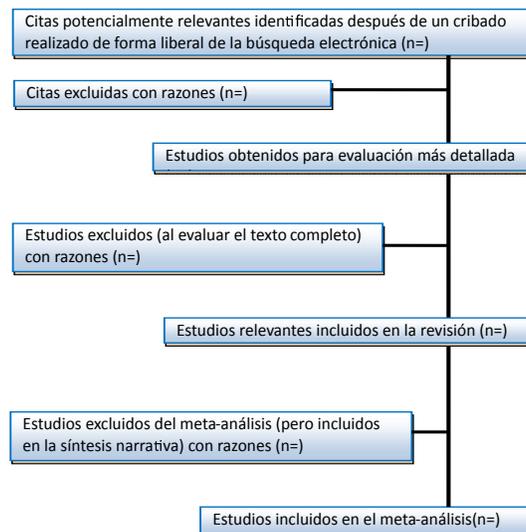


Figura 2.5: Selección de Estudio - Diagrama de flujo

Para evaluar la calidad de un meta-análisis, son elementos importantes: a) la evaluación de la validez de los estudios; b) la detección de la presencia de sesgos y c) la utilización de escalas de calidad de los estudios, Landis and Koch [72].

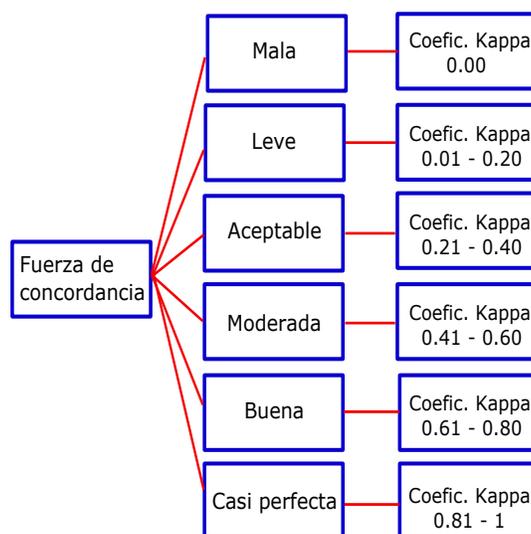


Figura 2.6: Interpretación coeficiente Kappa.

La validez de un meta-análisis debe garantizar la detección y eliminación de errores sistemáticos y sesgo. Existen dos tipos de validez, la interna y la externa. La validez interna se refiere aquello para lo cual ha sido diseñado, al referirse a los conceptos de exactitud, fiabilidad y reproducibilidad. La validez externa es facilidad de generalizar los resultados de un estudio a personas o poblaciones distintas.

En lo referente al sesgo, el mismo se relaciona con las debilidades de la validez interna de una investigación, se cuestiona si los resultados obtenidos son verdaderos en la población a la cual se pretende generalizar los resultados. En la práctica existen cuatro tipos de sesgo, Martín Rodero [86], sesgo de selección, de realización, de desgaste y de detección. El primer sesgo ocurre cuando existen diferencias en la asignación de los participantes al grupo de control y al de tratamiento; el sesgo de realización es causado por las diferencias de los cuidados que reciben los pacientes; el tercero ocurre cuando se dan bajas entre los participantes de estudios afectando los grupos de comparación y el último sesgo es ocasionado por los errores en la medición de los resultados entre los grupos que se comparan.

2.4.3 CODIFICACIÓN DE ESTUDIOS

El revisor debe codificar las características de los estudios que están relacionando con los resultados del tópico de estudio. Glass distingue dos características, los metodológicos y los sustantivos. La primera característica se refiere al número de pacientes en cada estudio, años de publicación del estudio, criterios de inclusión/exclusión, tipo de cegado, número de hombre/mujeres, edades de los pacientes (media, desviación típica, etc.) y la segunda característica considera aspectos como valoración de los análisis estadísticos, valoración del diseño, valoración de la presentación de los resultados, adecuación del tipo de cegado, etc.

Para formalizar el proceso de codificación se realiza un cuestionario que incluya las características relevantes, pero este proceso de codificación presenta un problema de medida, por lo que está sujeto a fiabilidad y validez del estudio. Para que el proceso de codificación sea coherente, se debe elaborar un

manual de codificación que contenga lo más detallado posible las definiciones de las características, las normas y reglas de codificación. Este proceso debe ser conducido por expertos codificadores capacitados con anterioridad. El mismo evaluará una muestra aleatoria de todos los estudios incluidos en el meta-análisis. No existe un procedimiento único para codificar los estudios a combinar. Existe un formulario de manera general que puede contener variables, de Dios and Santamaría [21], como:

1. Identificación del estudio: autores, país, fuente de financiación.
2. Fecha de publicación del estudio.
3. Fecha de inicio y final del trabajo.
4. Problema estudiado o hipótesis formulada.
5. Características de los individuos: sexo, edad.
6. Diseño del estudio: experimental (con aleatorización o no) u observacional (prospectivo o retrospectivo).
7. Exposición o tratamiento estudiado: dosis, pauta.
8. Co-exposición o co-tratamiento.
9. Resultados que permiten medir los efectos principales y secundarios: resultados brutos y medidas del efecto ajustadas con sus intervalos de confianza o valores p asociados e interacciones representativas.
10. Valoración de la calidad o validez de los estudios.
11. Susceptibilidad del estudio a posibles sesgos, aspectos destacables y limitaciones del estudio, y explicaciones que los autores presentan de los resultados.

La parte fundamental de la codificación es materializar el conjunto de peculiaridades que tiene cada uno de los estudios recuperados, Martín Rodero [86]. En el proceso de la codificación se pueden presentar problemas de fiabilidad y puede estar latente el sesgo; para evitar la presencia de sesgo se recomienda que existan al menos dos codificadores. Mientras mayor sea la cantidad de codificadores, la fiabilidad efectiva se incrementa y por ende los errores aleatorios de cada codificador se anulan con los errores de los otros. Galindo-Villardón [40], Vallejo [140].

2.4.4 MEDIDAS DE LOS RESULTADOS

Para la integración de los resultados de un conjunto de estudios es preciso que lo mismos sean cuantificados en una escala de medida común, ya que si las variables dependientes empleadas en los diferentes estudios difieren entre sí, no son comparables entre sí. Es por eso que se aplican tres estrategias, la primera hace referencia a "vote counting", el uso de niveles de significancia (p-valor) y la tercera emplea el uso de medidas de la magnitud del efecto (o tamaño del efecto - TE).

2.4.4.1 VOTE COUNTING

Los resultados encontrados se categorizan de tres maneras, Vallejo [140].

- Resultados significativos en los cuales la diferencia de medias es positiva.
- Resultados significativos en los cuales la diferencia de medias es negativa.
- Resultados no significativos.

En esta estrategia se examinan todos los estudios que aporten relevancia sobre una variable independiente de interés. Se tabula el número de estudios que incluyen dentro de cada una de las tres categorías citadas anteriormente. Si un gran número de estudios se ubica dentro de una de las categorías anteriores y muy pocas en las dos restantes, se infiere que esa categoría modal es asumida como la mejor estimación de la verdadera relación entre la variable dependiente y la independiente, Light and Smith [76]. El inconveniente de este método es que puede llevar a conclusiones erróneas ya que no se toma en cuenta el tamaño de la muestra, la magnitud del efecto no es analizada, Martín Rodero [86].

2.4.4.2 VALOR P

El valor p hace referencia a la probabilidad de que la diferencia observada en una muestra haya ocurrido al azar siendo los grupos comparados realmente semejantes, es decir, bajo la hipótesis nula, esta estrategia presente una debilidad, no proporciona información en relación a la magnitud del efecto experimental. La significancia estadística no provee inferencia sobre la intensidad de la relación (tamaño del efecto), por lo cual es interesante seleccionar un indicador que conlleve a la significancia clínica a través de la estimación de la magnitud del efecto.

2.4.4.3 TAMAÑO DEL EFECTO

Las medidas más informativas del resultado de una investigación es el tamaño o magnitud del efecto, Green and Hall [48], pero fue Glass el responsable de introducir las medidas del TE en la integración cuantitativa de resultados de investigación.

El TE es un valor concreto, distinto de cero, cuanto mayor es este valor, tanto mayor es el grado en que se manifiesta el fenómeno bajo estudio, esto es cuando la hipótesis nula es falsa. El p valor permite determinar si un efecto experimental es distinto de cero, el TE permite conocer en que medida es distinto de cero. Para estimar el TE existen varios procedimientos, entre los cuales tenemos la diferencia de medias, diferencias de medias estandarizadas, Glass [46], etc. Dependiendo de la naturaleza de la investigación que se pretenda integrar, se puede hacer uso de varios índices o indicadores de la magnitud el efecto, Castillo [15].

2.4.4.4 COEFICIENTE DE CORRELACIÓN

Es una medida estandarizada de la relación entre dos variables, es decir mide el grado de intensidad o de relación. Tenemos el coeficiente de correlación de Pearson y de Spearman. Estos índices pueden tomar valores entre -1 hasta +1. Valores cercanos a +1 indican una fuerte relación positiva o directa, valores cercanos a -1 denota una relación negativa o inversa entre las variables, si el índice es igual a cero se puede inferir ausencia de relación lineal entre las variables.

2.4.4.5 DIFERENCIA DE MEDIAS

Estima la magnitud de la diferencia de medias, y se encuentran los índices de Cohen, Glass y Hedge, Vallejo [140]. Se puede calcular la diferencia de media y la medida asociada de precisión para cada estudio, Normand [94]. Sea i el índice de estudio, T el tratamiento, C el grupo de control, n_{Ti} y n_{Ci} , los respectivos tamaños muestrales de las poblaciones. Una posible medida es la diferencia de media $Y = \bar{X}_{Ti} - \bar{X}_{Ci}$ (diferencia de media del tratamiento i con el grupo de control i) con error estándar denotado por S_i y estimado como $S_i = S_{pi}^2 \left(\frac{1}{n_{Ti}} + \frac{1}{n_{Ci}} \right)$, donde $S_{pi}^2 = \frac{(n_{Ti} - 1)S_{Ti}^2 + (n_{Ci} - 1)S_{Ci}^2}{n_{Ti} + n_{Ci} - 2}$, nótese que S_{Ti}^2 y S_{Ci}^2 son las varianzas muestrales de los tratamientos y grupo de control respectivamente para i -ésimo estudio. Otro estimador común para la medida del efecto es la diferencia de media estandarizada, la misma se estima como la diferencia de media dividida para la variabilidad de las medidas, es decir, si asumimos normalidad en los tratamientos y grupos de control con sus medias μ^T y μ^C respectivamente y varianzas iguales σ^2 , en otras palabras $Y_{ij}^T \sim N(\mu^T, \sigma^2)$ e $Y_{ij}^C \sim N(\mu^C, \sigma^2)$, se tiene el estimador insesgado $\delta = \frac{\mu^T - \mu^C}{\sigma}$, donde δ se puede interpretar como la pérdida o ganancia de la fracción de variabilidad de la medida. Un estimador de δ denotado por Hedges' g es definido como $h_i = \frac{\bar{Y}_i^T - \bar{Y}_i^C}{S_p}$, la razón de dividir por la desviación estándar es tener una medida resumen libre de unidades (adimensional) debido que el éxito en cada estudio se mide de diferente manera; donde el estimador de la varianza de h_i es $\frac{1}{n_{Ti}} + \frac{1}{n_{Ci}} + \frac{\hat{\delta}^2}{2(n_{Ti} + n_{Ci})}$, donde $\hat{\delta}^2$ es el estimador muestral de δ^2 .

2.4.4.6 ESTIMACIÓN DE RIESGO

La estimación se obtiene a partir de una tabla de doble entrada (2X2), entre los cuales tenemos diferencia de riesgo, el riesgo relativo, Odds-Ratio (OR) y el número necesario de pacientes a tratar (NNT). La diferencia de riesgo es fácil de interpretar, son proporciones (individuos en una población que tienen la enfermedad o el evento de interés en un momento dado), que pueden tomar valores de 0 hasta 1 (es decir, es la probabilidad de estar enfermo en el instante en que se mide la prevalencia) y siguen aproximadamente una distribución normal para tamaños muestrales pequeños o modestos, Normand [94]. Los riesgos relativos y odds ratios suelen ser analizados en la escala logarítmica, pero, a diferencia de la diferencia de riesgo, no se definen valores límites o fronteras. El riesgo relativo (RR), indica la probabilidad de que se desarrolle la enfermedad en los expuestos a un factor de riesgo, en relación al grupo de los no expuestos, si $RR = 1$ se infiere que no hay asociación entre la enfermedad y el factor de riesgo, si es mayor que 1 el factor de riesgo favorece la enfermedad y si RR es menor que 1 indica un factor protector, Martínez-González et al. [87]. Es importante notar que el RR no puede ser utilizado en estudios retrospectivos ya que no se conocen las probabilidades condicionadas de presentar la enfermedad ya que se fija de entrada los casos totales con enfermedad y los casos totales sin enfermedad y no es posible conocer la proporción real en los subgrupos estudiados. Odd-Ratio es una razón¹, en la cual el numerador es la probabilidad "p" de que ocurra un suceso y el denominador es la probabilidad de que el suceso no ocurra, toma valores entre cero e infinito, tiene como fin saber si un suceso es más probable que su complemento, Martínez-González et al. [87]. Y NNT es el número de individuos que hay que atender con el tratamiento experimental para producir una mejoría respecto al control, es decir cuántos pacientes hay que tratar para obtener un resultado positivo; lo ideal sería que NNT sea igual a uno, es decir que

¹Cociente en el cual el numerador no forma parte del denominador

todos mejoran con el tratamiento y ninguno con el control, cuanto más grande sea NNT o se aleje de la unidad menos eficaz es el tratamiento, Martínez-González et al. [87]. Según, Normand [94] la diferencia de riesgo, el riesgo relativo y Odds ratio se definen como se observa en la Tabla 2.2.

	Diferencia de Riesgo	Riesgo Relativo	Odds Ratio
Parámetro	$D = P_T - P_C$	$R = \frac{P_T}{P_C}$	$\Omega = \frac{P_T(1 - P_T)}{P_C(1 - P_C)}$
Estimador	$d_i = \hat{p}_{Ti} - \hat{p}_{Ci}$	$r_i = \frac{\hat{p}_{Ti}}{\hat{p}_{Ci}}$	$\omega_i = \frac{\hat{p}_{Ti}\hat{q}_{Ci}}{\hat{q}_{Ti}\hat{p}_{Ci}}$
Error Estándar	$S_{d_i} = \sqrt{\frac{\hat{p}_{Ti}\hat{q}_{Ti}}{n_{Ti}}}$	$S \log(r_i) = \frac{\hat{q}_{Ti}}{n_{Ti}\hat{p}_{Ti}} + \frac{\hat{q}_{Ci}}{n_{Ci}\hat{p}_{Ci}}$	$S \log(\omega_i) = \sqrt{\frac{1}{n_a} + \frac{1}{n_b} + \frac{1}{n_c} + \frac{1}{n_d}}$

Tabla 2.2: Definiciones de riesgos y Odds Ratio

Donde T denota el tratamiento, C el grupo de control, $q_i = 1 - p_i$, n_{Ti} y n_{Ci} representan el número de pacientes en el grupo de tratamiento como de control, y a, b, c y d representan el número de observaciones en la tabla de doble entrada que presentan la presencia y ausencia de la enfermedad y factor de riesgo. Otra manera de observar los riesgos es mediante una tabla de doble entrada (2x2), donde las columnas representan la presencia o ausencia de la enfermedad y las filas la presencia o ausencia del factor de riesgo según Martínez-González et al. [87], mayor detalle note Tabla 2.3.

	Enfermedad			
	Presente	Enfermos	No Enfermos	
Factor de Riesgo	Presente	VP (a)	FP (b)	VP + FP
	Ausente	FN (c)	VN (d)	FN + VN
		VP + FN	FP + VN	VP + FN + FP + VN

Tabla 2.3: Presencia o ausencia del factor de riesgo

Donde VP, FP, FN y VN, denotan los verdaderos positivos, falsos positivos, falsos negativos y verdaderos negativos respectivamente, luego el riesgo relativo se define como la ecuación 2.1.

$$RR = \frac{P(E|+)}{P(E|-)} = \frac{\frac{a}{a+b}}{\frac{c}{a+d}} = \frac{\text{Incidencia en expuesto}}{\text{Incidencia en no expuesto}} \quad (2.1)$$

Donde, la probabilidad condicional $P(E|+)$ se interpreta como la probabilidad de que un paciente esté enfermo cuando la prueba da positiva y $P(E|-)$ representa la probabilidad de que el paciente esté enfermo cuando el test es negativo. Con relación a los Odds Ratio, se define el odd ratio prospectivo y retrospectivo con las ecuaciones 2.2 y 2.3, Martínez-González et al. [87]. Donde $P(No E|+)$ y $P(No E|-)$ denotan la probabilidad de no estar enfermo, dado que el test es positivo y la probabilidad de no estar enfermo dado que la prueba fue negativa respectivamente. Asimismo $P(+|No E)$ y $P(-|No E)$ representan la probabilidad que dio positivo en el test dado que no está enfermo y la probabilidad de que el test fue negativo dado que no está enfermo respectivamente.

$$OR = \frac{\frac{P(E|+)}{P(No E|+)}}{\frac{P(E|-)}{P(No E|-)}} = \frac{\frac{\frac{a}{a+b}}{\frac{b}{a+b}}}{\frac{\frac{c}{c+d}}{\frac{d}{c+d}}} = \frac{ad}{bc} = \frac{\text{Razón enfermos en los expuestos}}{\text{Razón enfermos en los no expuestos}} \quad (2.2)$$

$$OR = \frac{\frac{P(+|E)}{P(-|E)}}{\frac{P(+|No E)}{P(-|No E)}} = \frac{\frac{\frac{a}{a+c}}{\frac{c}{a+c}}}{\frac{\frac{b}{b+d}}{\frac{d}{b+d}}} = \frac{ad}{bc} = \frac{\text{Razón de expuestos al riesgo en enfermos}}{\text{Razón de expuestos al riesgo en sanos}} \quad (2.3)$$

2.4.5 ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS

Una vez cuantificados los estudios mediante niveles de significancia o estimadores de TE, los resultados son transformados en un índice global, el cual debe ser representativo del conjunto de estudio. Por otro lado para combinar estadísticamente los resultados en un meta-análisis existen tres modelos, el modelo de efectos fijos (fixed effects), el modelo de efectos aleatorios (random effects) y el modelo de efectos mixtos (mixed effects).

2.4.5.1 MODELO DE EFECTOS FIJOS (fixed effects)

Asume que en cada estudio analizado se mide el mismo parámetro subyacente y que no hay variación en la medición entre los estudios Normand [94], es decir, las diferencias que se observan entre los estudios se deben únicamente a variabilidad aleatoria. Una definición matemática la encontramos en Normand [94] dice, que cada resumen estadístico Y_i es una realización de una población de estudio, la cual se estima con media θ , donde θ es el parámetro principal de interés y se asume que existen k estudios independientes ($i = 1, 2, \dots, k$). Se asume que $E[Y_i] = \theta$ y $S_i^2 = Var(Y_i)$ la varianza resumen del estudio i . Cuando se tienen tamaños de estudio moderadamente grandes Normand [94] cada Y_i se distribuye asintóticamente normal (por el teorema del límite central) y aproximadamente insesgada, es decir $Y_i \sim N(\theta, S_i^2)$ para ($i = 1, 2, \dots, k$) y asume conocida la varianza S_i^2 . El parámetro central de interés es θ que cuantifica el promedio de los efectos del tratamiento.

2.4.5.2 MODELO DE EFECTOS ALEATORIOS(random effects)

No asume la homogeneidad entre los resultados de los estudios. Se suelen aplicar cuando los estudios tienen un grado de heterogeneidad superior al esperado por simple variabilidad aleatoria, Martín Rodero [86]. Según Normand [94], en el modelo se asume que cada resumen estadístico del estudio Y_i analizado son independientes y puede ser representado como una distribución normal con media θ_i y varianza S_i^2 , es decir $Y_i|\theta_i, S_i^2 \sim N(\theta_i, S_i^2)$. Del mismo modo, cada media de los estudios que se analizan θ_i , se asumen que

proviene de una gran población normal de efectos de media θ y varianza τ^2 es decir $\theta_i|\theta, \tau^2 \stackrel{ind}{\sim} N(\theta, \tau^2)$, donde θ y τ^2 son llamados hiper-parámetros y representan la media del efecto del tratamiento y la variación inter-estudio respectivamente. Nótese, que dado los hiper-parámetros de la distribución de Y_i , después de promediar los efectos de cada estudio específico es normal con media θ y varianza $S_i^2 + \tau^2$. Como en el caso del modelo de efectos fijos, el parámetro central a estimar era θ , asimismo la variación inter-estudio τ^2 , juega un rol importante y debe ser estimada. Adicionalmente la media del tratamiento del efecto es posible estimarla mediante los efectos de los estudios específicos, θ_i , los cuales son útiles para identificar particularmente la eficacia de los estudios. La distribución θ_i , condicionada a los datos observados y a los hiper-parámetros es denotada como en la ecuación 2.4, donde $Y = (Y_1, Y_2, \dots, Y_k)$ y B_i es definido como $\frac{S_i^2}{S_i^2 + \tau^2}$, denominado comúnmente **”shrinkage factor”**. Cuanto más grande sea la variación inter-estudio τ^2 , menor será B_i de los efectos de los estudios observados, porque $0 \leq B_i \leq 1$ y la media de θ_i de la ecuación 2.4, es un compromiso de la media de los efectos del tratamiento θ y el resumen estadístico del estudio observado Y_i . Cuando $\tau^2 = 0$, **”shrinkage”** toma su valor máximo con $B_i = 1$ tal que $\theta_1 = \theta_2 = \dots = \theta_k = \theta$, entonces el modelo de efectos aleatorios corresponde al modelo de efectos fijos.

$$\theta_i|Y, \theta, \tau \sim N(B_i\theta + (1 - B_i)Y_i, S_i^2(1 - B_i)) \quad (2.4)$$

2.4.5.3 MODELO DE EFECTOS MIXTOS (mixed effects)

El modelo de efectos aleatorios se basa en una hipótesis poco realista, asumiendo que los estudios provienen de una gran población de estudios y que la variabilidad entre estudios puede ser representada por una sola varianza. El modelo de efectos mixtos puede considerarse como una regresión de efectos aleatorios que explica la heterogeneidad adicional debido a variables externas por lo que, podría ser la solución al problema de la heterogeneidad, ya que este enfoque incluye una variación aleatoria adicional tomando en cuenta diferentes factores o variables que podrían explicar la variabilidad entre los estudios. De esta manera, meta-regresión y la meta-partición proporcionan una alternativa para explorar las asociaciones entre las características y los resultados de los estudios.

2.4.5.4 META-REGRESIÓN

El término meta-regresión hace referencia a los análisis en los cuales las características de los estudios, o las características de los sujetos que intervienen en ellos se utilizan como variables independientes (explicativas) en una ecuación de regresión simple o múltiple, en la cual la variable dependiente (a explicar) es el resultado de cada estudio (efecto medio), Martín Rodero [86], Molinero [92].

2.4.5.5 META-PARTICIÓN

Meta-partición es un procedimiento recursivo que identifica sub-grupos mutuamente excluyentes y exhaustivos de una población de tamaños de efectos Ortega et al. [95]. Es un método que se utiliza en meta-análisis para explicar la heterogeneidad de los tamaños de efectos con variables moderadoras². Por lo tanto, se consideran dos tipos de variables: (1) las variables de respuesta que proporcionan las estimaciones de los tamaños del efecto, y (2) los factores que influyen en los tamaños del efecto, que pueden

²Aquellas que modifican la relación entre la variable independiente y la dependiente

ser variables cualitativas o cuantitativas. El criterio consiste en ajustar la heterogeneidad de los tamaños muestrales de los efectos y particionar esta variabilidad mediante un algoritmo que encuentre el mejor moderador de cada partición. El mejor moderador es el que explica la mayor cantidad de variabilidad entre los subconjuntos posibles de cada partición Ortega et al. [95], note Figura 2.7.

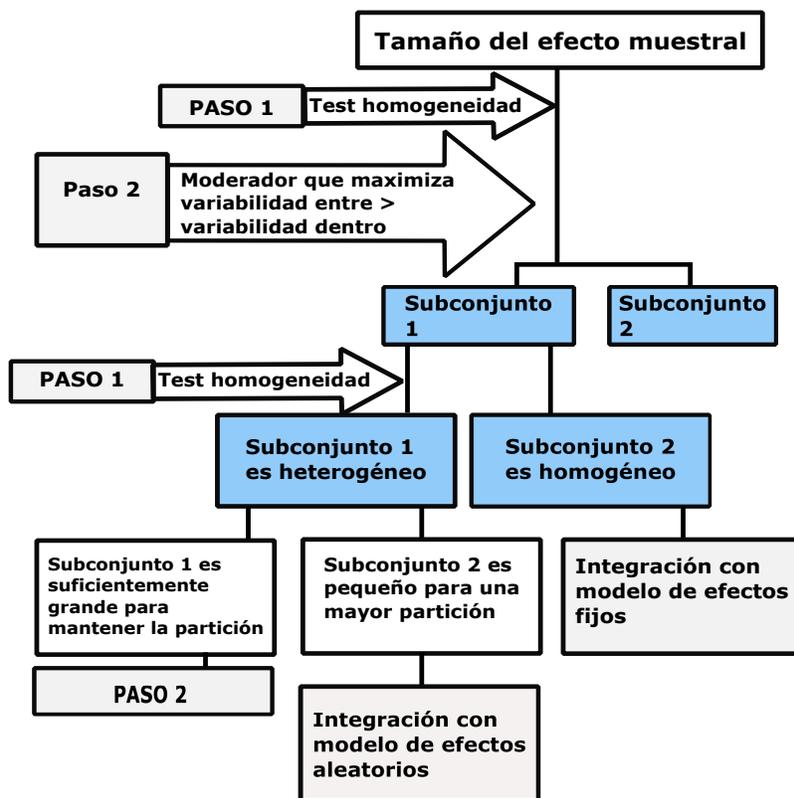


Figura 2.7: Diagrama que resume la metodología meta-partición.

El paso 1 consiste en la evaluación de la heterogeneidad con la prueba de homogeneidad y el valor de I^2 . El paso 2 consiste en seleccionar el mejor moderador de acuerdo con el valor de p , y la etapa 3 implica integrar la magnitud del efecto en los subconjuntos finales.

Como se sabe la medicina es una ciencia que involucra probabilidades y el manejo de la incertidumbre, la misma se extiende no sólo a las actividades preventivas, terapéuticas y pronósticos sino también a las diagnósticas, es por eso que en el Capítulo 3 se analizan los principales conceptos de pruebas diagnósticas desde un punto de vista estadístico.

CAPÍTULO 3

META-ANÁLISIS EN PRUEBAS DIAGNÓSTICAS

”El diagnóstico no es una suma de los datos obtenidos por la clínica y los exámenes complementarios, sino una síntesis de éstos”

M. Soriano Jiménez

3.1 PRUEBAS DIAGNÓSTICAS

Las pruebas diagnósticas son pruebas para determinar si un individuo tiene o no cierta enfermedad (u otra característica). Las estimaciones numéricas que se obtengan de la misma pueden depender de la prevalencia (proporción de individuos que tienen la enfermedad o el evento de interés en el momento dado) y de las características clínicas de los pacientes que se hayan estudiado Sergas [126].

3.1.1 CLASIFICACIÓN DE LAS ENFERMEDADES POR DIAGNOSIS

Dependiendo del grado de avance de la enfermedad en el paciente, las enfermedades pueden clasificarse en tres tipos básicos.

- **Tipo I.-** La peor equivocación que se pueden cometer en el diagnóstico es un Falso Negativo. En esta clasificación están las enfermedades que son curables si se detectan a tiempo.
- **Tipo II.-** La equivocación que se puede cometer en el diagnóstico es un Falso Positivo , es decir la enfermedad es grave, pero difícilmente curable o sin remisión.
- **Tipo III.-** Las enfermedades que no pueden ser clasificadas como de Tipo I o Tipo II. Ejemplo el SIDA.

3.1.2 CRITERIOS DE CALIDAD DE LAS PRUEBAS

La calidad de la prueba, es la capacidad de la prueba para medir efectivamente lo que se quiere medir. Una prueba diagnóstica es considerada de buena calidad si es capaz de clasificar correctamente a los miembros de la población a la que se aplica, dando resultados positivos en las personas enfermas y negativas en los carentes de la enfermedad de estudio. La calidad de una prueba diagnóstica se mide en términos de validez y seguridad.

- **Validez.-** Compara la prueba con otra prueba de probada eficacia en la enfermedad de estudio, a este proceso de comparación se le llama Patrón Áureo (Gold Estándar) o criterio de referencia, Pita Fernández and Pértegas Díaz [100].
- **Seguridad.-** Capacidad para predecir la presencia o ausencia de la enfermedad y se mide en términos de los valores predictivos positivos y negativos, Pita Fernández and Pértegas Díaz [100], observe Figura 3.1.

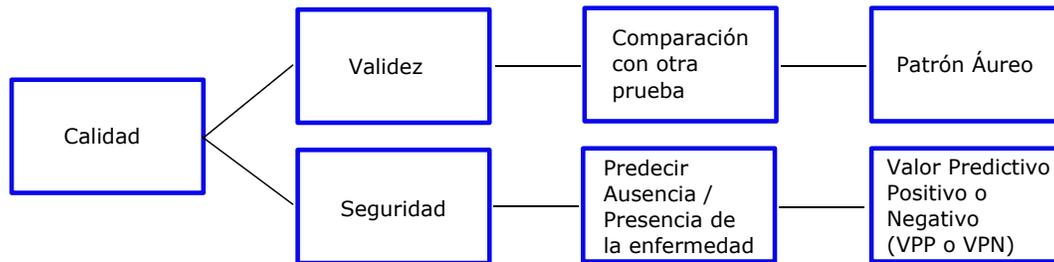


Figura 3.1: Criterio de calidad de las pruebas

Es importante notar que existen tres tipos de pruebas:

- Las que tienen **respuesta dicotómica** (Positiva +, Negativa -), ejemplo, pruebas de embarazo.
- Las pruebas que poseen **respuesta policotómica**, ejemplo, estudios del cáncer.
- Las que poseen **respuesta cuantitativa**, ejemplo, prueba Glucosa/diabetes.

3.1.3 DISEÑO DE UNA INVESTIGACIÓN PARA EVALUAR UNA PRUEBA DIAGNÓSTICA

En este punto, se trata de realizar una investigación descriptiva, en la cual un grupo de participantes con características determinadas acorde a la prueba a probar es sometido a la prueba y concomitantemente al patrón áureo, note Figura 3.2. En toda investigación, existe sesgo, para minimizar el mismo se deben considerar, tres aspectos importantes: la selección de la muestra, comparación e interpretación de los resultados.

El modelo apropiado es un diseño descriptivo, en el cual un grupo de participantes con características determinadas acorde a la prueba a probar es sometido a la prueba y concomitantemente al patrón áureo.

- **Selección de la muestra.-** La muestra debe contemplar participantes que caractericen todas las formas clínicas del daño buscado, guardando la proporción en que la misma se presenta en la práctica cotidiana.
- **Comparación.-** Debe establecerse con un patrón de referencia confiable y apropiado. El patrón áureo debe reunir esas características y a todos los participantes se le deben realizar las dos pruebas (la que se está investigando y el patrón áureo). Al utilizar una prueba costosa como patrón áureo se imposibilita llenar las casillas c y d de la tabla de doble entrada, por lo que no es posible estimar la sensibilidad y la especificidad, a este error de procedimiento (error sistemático o sesgo) se lo denomina error de verificación (work-up bias).
- **Interpretación de los resultados.-** La lectura de los resultados de la prueba analizada debe ser efectuada, sin tener conocimiento de los resultados del patrón áureo, para minimizar o evitar el sesgo hacia una coincidencia espuria.

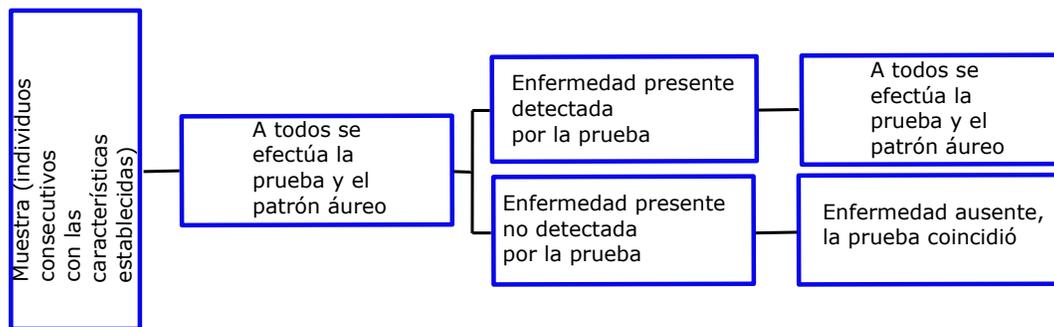


Figura 3.2: Diseño para evaluar una prueba diagnóstica

3.1.4 PARÁMETROS DE VALORACIÓN DE PRUEBAS DIAGNÓSTICAS

Supóngase que se quiere evaluar una prueba analítica para el diagnóstico de una enfermedad, obteniéndose los resultados, que se observan en la Tabla 3.1. Los 27 pacientes que padecen la enfermedad dieron positivo en la prueba, se consideran verdaderos positivos (VP). Los 6 pacientes que están enfermos y que dieron negativo en la prueba de laboratorio son considerados falsos negativos (FN). Los tres pacientes que no tienen la enfermedad y que sus pruebas clínicas denotan un resultado positivo se denominan falso positivo (FP). Y los 164 pacientes que no tienen la enfermedad y cuyos resultados de laboratorio fueron negativos se catalogan como verdaderos negativos (VN). A continuación se procederá a definir los parámetros de valoración de una prueba de diagnóstico.

1. **Sensibilidad.-** Mide el porcentaje de individuos enfermos correctamente diagnosticados, es decir, es la probabilidad de que el test sea positivo en presencia de la enfermedad, Burgueño et al. [13]. Definamos los siguientes eventos:
 - (a) **E:** Individuo con la condición de interés
 - (b) **NE:** Individuo sin la condición de interés
 - (c) **T:** Prueba diagnóstica

- (d) **T+**: Resultado positivo del test
- (e) **T-**: Resultado negativo del test
- (f) **P(E)**: Probabilidad a priori de que un paciente elegido al azar tenga la enfermedad, también denominada prevalencia.

Prueba Diagnóstica	Enfermedad		
		Enfermos	No Enfermos
Presente	27 (VP) a	3 (FP) b	30 (TP)
Ausente	6 (FN) c	164 (VN) d	170 (TN)
	33 (TE)	167 (TNE)	200 (T)

Tabla 3.1: Tabla de doble entrada, ejemplo hipotético

La sensibilidad se puede medir como; $Se = P(T + |E)$, es decir la probabilidad de que el test sea positivo dado que el paciente está enfermo. En otras palabras, mide el porcentaje de individuos enfermos correctamente diagnosticados, e indica la capacidad de la prueba para detectar a un sujeto enfermo. Siguiendo con la Tabla 3.1 la sensibilidad será:

$$Se = \frac{VP}{TE} \tag{3.1}$$

En el ejemplo, 81,8% de los pacientes enfermos dieron positivo en la prueba, por lo que habrá un total de 18,2% de falsos negativos.

2. **Especificidad.-** Es el porcentaje de resultados negativos en pacientes que no padecen la enfermedad, es decir, mide la probabilidad de que el test sea negativo en individuos sanos, $Es = P(T - |NE)$. En otros palabras, valora la capacidad de una prueba para detectar correctamente individuos sanos, una alta especificidad indica una baja frecuencia de falsos positivos, Burgueño et al. [13]; siguiendo el ejemplo de la Tabla 3.1 la especificidad se pueden medir como:

$$Es = \frac{VN}{TNE} \tag{3.2}$$

El 98,2% de individuos sanos dieron negativo en la prueba, y habrá un 1,8% de falsos positivos. Queda claro que la calidad de la prueba puede ser representada por la sensibilidad y especificidad del test, pero en la práctica clínica lo que interesa es responder a las preguntas, ¿cuál es la probabilidad de que esté verdaderamente enfermo si un sujeto ha resultado positivo , $P(E|T+)$ y por el contrario, ¿cuál es la probabilidad de que realmente esté sano si el sujeto resultó negativo en la prueba $P(NE|T-)$, estas dos probabilidades se pueden calcular utilizando el Teorema de Bayes¹ siempre y cuando se conozca la prevalencia de que un sujeto este enfermo antes de ser sometido al

¹Se considera a los parámetros poblacionales como variables aleatorias; los estadísticos que promueven estos conceptos se denominan Bayesianos ya que se basan en el concepto de la probabilidad subjetiva.

test. Para encontrar estas probabilidades denominemos P la probabilidad pre-prueba de padecer la enfermedad, Se la sensibilidad de la prueba y Es la especificidad de la misma, lo cual permitirá estimar la probabilidad post-prueba de que el sujeto esté enfermo cuando resultó positivo o de que este sano cuando resultó negativo.

3. **Valor Predictivo Positivo (VPP).**- Es el porcentaje de pacientes enfermos entre todos los pacientes con resultados positivos, Burgueño et al. [13]. Valora la probabilidad de que una prueba positiva diagnostique correctamente a un individuo enfermo, luego el VPP queda definido como se nota en la ecuación 3.3.

$$P(E|T+) = \frac{P(T+|E) \times P(E)}{P(T+|E) \times P(E) + P(T+|NE) \times P(NE)} = \frac{Se \times P(E)}{Se \times P(E) + (1 - Se) \times (1 - P(E))} = \frac{VP}{VP + FP} \quad (3.3)$$

$$P(E|T+) = \frac{VP}{TP}$$

4. **Valor Predictivo Negativo (VPN).**- Indica la frecuencia de pacientes no enfermos entre todos los pacientes con resultado negativo Burgueño et al. [13]. Pondera la probabilidad de que una prueba negativa diagnostique correctamente a un individuo sano, note ecuación 3.4.

$$P(NE|T-) = \frac{P(T-|NE) \times P(NE)}{P(T-|NE) \times P(NE) + P(T-|E) \times P(E)} = \frac{Es \times P(NE)}{Es \times P(NE) + (1 - Es) \times (1 - P(E))} = \frac{VN}{VN + FN} \quad (3.4)$$

$$P(NE|T-) = \frac{VN}{TN}$$

Luego si seguimos con el ejemplo de la Tabla 3.1, tenemos que, $VPP = \frac{27 \times 100}{30} = 90\%$ y $VPN = \frac{164 \times 100}{170} = 96,5\%$. Un VPP de 90% indica que de cada 100 pacientes que dan la prueba positiva 90 padecen la enfermedad. Un VPN de 96,5% indica que de cada 100 pruebas negativas 96,5 pertenecen a individuos sanos. En resumen un test altamente sensible proporciona pocos falsos negativos y es bueno para descartar, de manera análoga un test altamente específico proporciona pocos falsos positivos y es bueno para confirmar. Es importante conocer las estimaciones para los intervalos de confianza de los valores predictivos positivo y negativo, los cuales se estiman con las expresiones 3.5 y 3.6, donde $\varepsilon = 1,96$. Los valores predictivos dependen mucho de la prevalencia de la enfermedad en la población (cociente entre enfermos y población), no así la sensibilidad y especificidad, que son independientes de la misma. VPP y VPN interesan en la práctica clínica, pero presentan un inconveniente, es que dependen de la proporción de enfermos en la muestra estudiada, por lo que es necesario utilizar otros tipos de índices como las razones de verosimilitud que no dependan de esa proporción de enfermos.

$$VPP \pm \varepsilon \times \sqrt{\frac{VPP \times (1 - VPP)}{VP + FP}} \quad (3.5)$$

$$VPN \pm \varepsilon \times \sqrt{\frac{VPN \times (1 - VPN)}{FN + VN}} \quad (3.6)$$

5. **Punto de Corte.-** Para entender la elección de los puntos de corte en las pruebas diagnóstica es necesario introducir el concepto de coste (o utilidad) en la realización de las pruebas. El costo puede ser interpretado en un sentido estricto de dinero o días perdidos por pacientes saludables, pero también se puede pensar en una combinación de varios componentes Metz [89].

El costo medio (promedio) de la realización de una prueba diagnóstica debe incluir, en primer lugar, el precio promedio que se debe pagar para la realización de la prueba; a este costo lo denotaremos por C_o . Al coste C_o hay que añadirle los costes médicos de cada tipo de decisión diagnóstica. Como se trata de estimar un coste medio, hay que ponderar la decisión de cada uno de estos costos mediante las probabilidades involucradas en cada tipo de decisión en cuestión. Por lo que el coste medio \bar{C} resultante de la prueba diagnóstica queda expresado en la ecuación 3.7 Metz [89]. Donde P, representa la probabilidad de clasificar un caso como verdadero positivo - P(VP), verdadero negativo - P(VN), falso positivo - P(FP) y falso negativo - P(FN). C representa el coste promedio de una decisión medica de clasificar un paciente como, verdadero positivo - C_{VP} , verdadero negativo - C_{VN} , falso positivo - C_{FP} y falso negativo - C_{FN} .

$$\bar{C} = C_o + C_{VP}P(VP) + C_{VN}P(VN) + C_{FP}P(FP) + C_{FN}P(FN) \quad (3.7)$$

La probabilidad P(VP) puede ser reemplazada por $P(E)P(T+|E) = Se \cdot P(E)$. Similarmente, $P(VN) = P(NE)P(T-|NE) = Es \cdot P(NE)$; $P(FN) = P(NE)P(T-|E) = (1 - Se) \cdot P(E)$; y $P(FP) = P(NE)P(T+|NE) = (1 - Es) \cdot P(NE)$.

Además, $P(T-|E) = 1 - P(T+|E)$ y $P(T-|NE) = 1 - P(T+|NE)$. Por lo tanto la expresión 3.7 puede ser expresada como la ecuación 3.8.

$$\begin{aligned} \bar{C} = & - \{[C_{FN} - C_{VP}]P(E)\}P(T+|E) + \{[C_{FP} - C_{VN}]P(NE)\}P(T+|NE) + \\ & + \{C_o + C_{VN}P(NE) + C_{FN}P(E)\} \end{aligned} \quad (3.8)$$

De la ecuación 3.8 se puede inferir que \bar{C} depende proporcionalmente de C_o y de las probabilidades $P(T+|E)$ y $P(T+|NE)$ que representan a la Se y $(1 - Es)$ las cuales son las coordenadas de una curva ROC.

Dado que el tomador de decisiones puede cambiar estas cantidades en conjunto, es decir, puede mover la Se y $(1 - Es)$ a lo largo de la curva ROC mediante el uso de umbrales diferentes, el costo promedio dependerá del umbral utilizado. Por lo tanto, en términos de análisis coste-beneficio el mejor punto de funcionamiento en una curva ROC se lo obtiene minimizando el coste medio (o maximizando los beneficios) en una situación particular de diagnóstico.

Si se deriva implícitamente \bar{C} en términos de $P(T + |NE)$, es decir $P(T + |E)$ es una función de $P(T + |NE)$ e igualando a cero se obtiene la ecuación 3.9.

$$\frac{d\bar{C}}{dP(T + |NE)} = - \{[C_{FN} - C_{VP}] P(E)\} \frac{dP(T + |E)}{dP(T + |NE)} + \{[C_{FP} - C_{VN}] P(NE)\} = 0$$

despejando $\frac{dP(T + |E)}{dP(T + |NE)}$

$$(3.9)$$

\therefore

$$\frac{dP(T + |E)}{dP(T + |NE)} = \frac{C_{FP} - C_{VN}}{C_{FN} - C_{VP}} \cdot \frac{P(NE)}{P(E)}$$

Para considerar el efecto de la prevalencia en el espacio ROC, supóngase que se estudia una enfermedad rara en una población cualquiera, es decir, $P(E)$ es cercano a 0, por lo que $P(NE)$ será cercana a uno. Luego el tomador de decisión deberá trabajar en la parte inferior izquierda de la curva ROC donde TVP es pequeña pero TFP es aún más pequeña. Pero si se considera una enfermedad común $P(NE)$ será cercana a 0, y el mejor punto de operación de la curva ROC se ubicará en la parte superior derecha del espacio ROC - donde TVP es alta pero TFP es aún más.

Si se analizan los costos de la ecuación 3.9 y se nota que $C_{FP} - C_{VN}$ es mayor que $C_{FN} - C_{VP}$ (cuando se aplica un tratamiento o varias pruebas diagnósticas a pacientes saludables, siendo estas medidas perjudiciales a los mismos pero de poco beneficios a pacientes enfermos), entonces el valor de la pendiente es grande. A la inversa, si $C_{FP} - C_{VN}$ es menor que $C_{FN} - C_{VP}$, es decir, se aplicó un tratamiento o algunas pruebas a pacientes sanos, pero las mismas fueron inofensivas para estos pacientes y beneficiosas a pacientes enfermos, entonces el tomador de decisión debería operar en la parte superior derecha del espacio ROC, donde la TVP debería mantenerse con valores altos e incluso a expensas de la TFP.

6. **Razones de Verosimilitud.-** Al cociente entre la tasa de verdaderos positivos y la tasa de falsos positivos se le denomina cociente de probabilidad positiva (**LR+**); también recibe el nombre de cociente de probabilidad positivo o cociente de verosimilitud de una prueba positiva o poder predictivo de una prueba positiva; indica mejor capacidad para diagnosticar la presencia de la enfermedad en pacientes que presentan la alteración en comparación con los que no tienen la enfermedad (deseable que sea lo mayor posible) Sergas [126]. Al cociente entre la tasa de verdaderos negativos y la tasa de falsos negativos se le denomina cociente de probabilidad negativo (**LR-**) también conocido como cociente de verosimilitud negativo o índice de eficiencia pronóstico de una prueba negativa. Indica cuanto menos probable es un resultado negativo en los enfermos en comparación con los pacientes que no presentan la alteración; el inverso de este cociente se llama poder (o potencia) predictiva de una prueba negativa, note ecuaciones 3.10 y 3.11 respectivamente, Sergas [126].

$$LR+ = \frac{P(T + |E)}{P(T + |NE)} = \frac{Se}{1 - Es} \quad (3.10)$$

$$LR- = \frac{P(T - |E)}{P(T - |NE)} = \frac{1 - Se}{Es} \quad (3.11)$$

La importancia o utilidad de los cocientes $LR+$ y $LR-$ frente a los VPP y VPN de la prueba se fundamenta en que estos LR no dependen de la proporción de enfermos en la muestra sino de la sensibilidad y especificidad. Por otro lado si se estima la probabilidad pre-prueba de que un sujeto padezca la enfermedad, utilizando los cocientes de probabilidad, al realizar la prueba se podrá corregir ese valor, de tal manera que la probabilidad aumente o disminuya según que el resultado sea positivo o negativo, aplicando la fórmula 3.12, donde P es la probabilidad a priori o pre-prueba de la enfermedad, LR es el cociente de probabilidad (positivo en el caso de que deseáramos calcular la probabilidad de que padezca la enfermedad o negativo en caso contrario) y P_{post} es la probabilidad post prueba.

$$P_{post} = \frac{P \times LR}{1 + P \times (LR - 1)} \quad (3.12)$$

7. **Variabilidad de las razones de verosimilitud.**-Si la especificidad es constante, las razones de verosimilitud positivas $LR+ = \frac{Se}{1 - Es}$ son función lineal de la sensibilidad. Cuando la sensibilidad aumenta, $LR+$ aumenta y viceversa, Pita Fernández and Pértegas Díaz [100]. La razón de verosimilitud negativa $LR- = \frac{1 - Se}{Es}$ es igualmente una función lineal de sensibilidad pero negativa, cuando aumenta la sensibilidad disminuye $LR-$, Pita Fernández and Pértegas Díaz [100]. Nótese que, cuando $Se = 1 - Es = 0,1$, Figura 3.3 las dos razones de verosimilitud son iguales a 1. Luego la prueba no tiene ningún valor de diagnóstico. El valor diagnóstico del resultado positivo o negativo de la prueba crece linealmente cuando la sensibilidad aumenta de 0,1 a 1. En la Figura 3.4, se pone de manifiesto que cuando $Es = 1 - Se = 0,8$, las dos razones de verosimilitud son iguales a 1 y por lo tanto la prueba carece de utilidad diagnóstica. Se nota que la progresión de $LR+$ es muy rápida cuando la especificidad es superior a 0,7. Se dice que la eficacia diagnóstica de una prueba en la investigación de una enfermedad E está ligada al crecimiento de su especificidad. Comparando los cocientes de probabilidades de dos pruebas diagnósticas existen 4 posibles resultados que se pueden visualizar en un gráfico. En el eje Y ubicamos la sensibilidad y el eje X se representa 1-especificidad. Supóngase ahora que se quiere comparar la prueba A con la prueba B, cuya sensibilidad es de 0,9 y especificidad 0,5. Las zonas de la Figura 3.5, corresponden a las siguientes situaciones:

1. En esta región la prueba B es siempre superior a la A tanto para confirmar presencia como ausencia de enfermedad.
2. Prueba B solo es superior a la A para confirmar ausencia de enfermedad.
3. Prueba B siempre es inferior a A
4. B será superior a A para confirmar la presencia de la enfermedad.

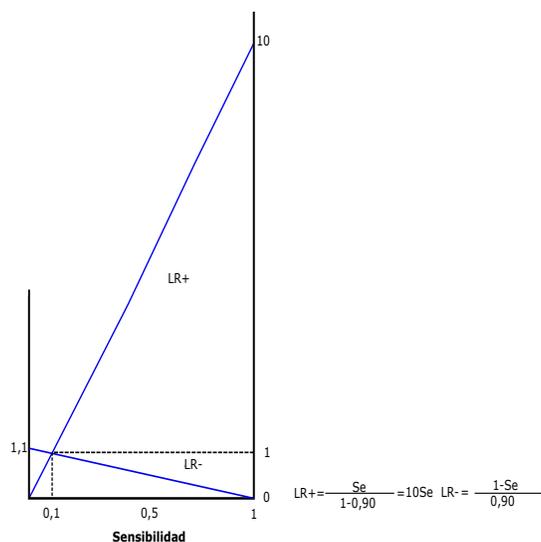


Figura 3.3: Variación de la razón de verosimilitud en función de la sensibilidad para diferentes pruebas cuya especificidad es igual a 0,90

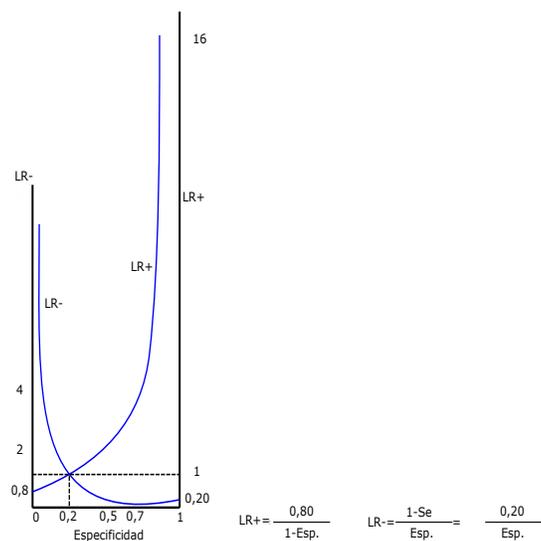


Figura 3.4: Variación de las razones de verosimilitud en función de la especificidad para diferentes pruebas cuya sensibilidad es igual a 0.80

Si se desea comparar la prueba A con otra prueba B cuya sensibilidad es 0,9 y especificidad 0,5, dichas pruebas pueden ser representadas mediante la Figura 3.6, obteniendo que B cae en la zona 2, lo cual indica que es una prueba mejor que A para validar o confirmar ausencia de la enfermedad. Para comparar los parámetros de dos pruebas diagnósticas hay que considerar algunos criterios que las estimaciones realizadas podrán estar sometidos a posibles errores de muestreo, con lo cual habrá que realizar un contraste de hipótesis estadística para inferir si las diferencias encontradas

son suficientemente importantes y no se deben al azar. Si la razón de verosimilitud se acerca a 1, la probabilidad diagnóstica es la misma antes y después de la prueba. Bajo este escenario el test es inútil y no tiene capacidad discriminatoria. Cuanto más se aleje de la unidad la razón de verosimilitud, más interés presenta la prueba. Según Evidence-bases Medicine Working Group, los cocientes de probabilidad tienen la interpretación que se observa en la Figura 3.7.

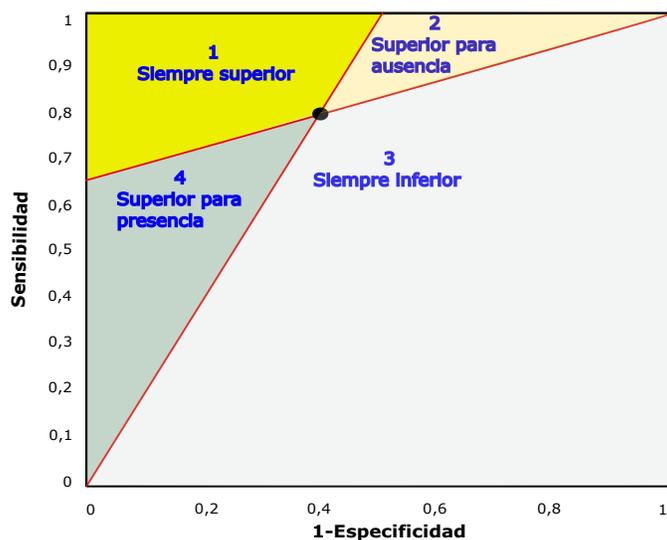


Figura 3.5: Especificidad Vs Sensibilidad

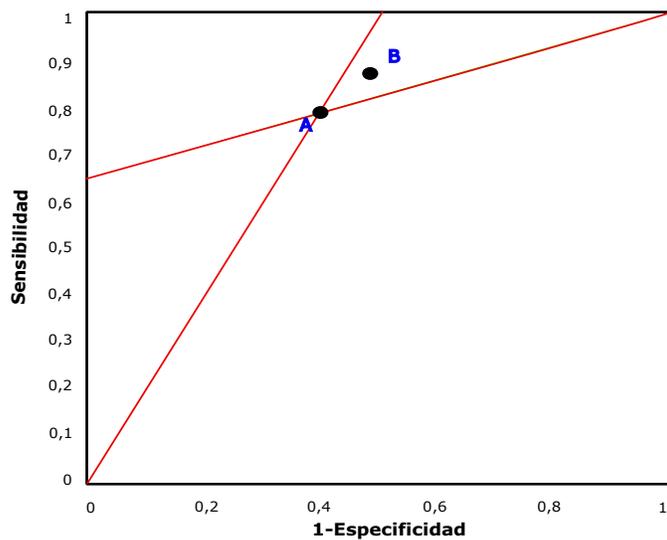


Figura 3.6: Especificidad Vs Sensibilidad

<p>LR+>10 o LR-<0.1</p> <p>Generan cambios amplios y a menudo concluyentes desde una probabilidad pre-prueba hasta una probabilidad post-prueba</p>	<p>5<LR+<10 y 0.1<LR-<0.2</p> <p>Generan cambios moderados desde la probabilidad pre-prueba hasta la probabilidad post-prueba</p>	<p>2<LR+<5 y 0.05<LR-<0.2</p> <p>Generan cambios pequeños (pero en ocasiones importantes) de la probabilidad</p>	<p>1<LR+<2 y 0.5<LR-<0.1</p> <p>Alteran la probabilidad en un grado insignificante (y rara vez importante)</p>
--	--	---	---

Figura 3.7: Cocientes de Probabilidad, según Evidence-bases Medicine Working Group

Las ventajas de utilizar o emplear la razón de verosimilitud, es que es independiente de la prevalencia, reflejando de manera correcta el valor de la prueba independientemente del grupo al que se le aplique, resumiendo la información en un solo parámetro, Seqc [125]. Los cocientes de probabilidad permiten transformar la probabilidad pre-prueba a post-prueba mediante la relación $odds_{post} = LR_T \times odds_{pre}$, donde odds es el cociente entre la probabilidad de que ocurra un evento y la probabilidad de que no ocurra, por lo tanto se tiene la ecuación 3.13.

$$odds = \frac{P}{1 - P} \tag{3.13}$$

$$P = \frac{odds}{1 + odds}$$

Los sub-índices post y pre señalan la post-prueba y pre-prueba y el sub-índice T representa el resultado obtenido positivo o negativo. Fagan [37], describió un nomograma para el teorema de Bayes, basado en la capacidad de convertir al teorema en una función sumatoria lineal simple. El mismo es una herramienta que permite calcular la probabilidad post prueba una vez que se conoce la probabilidad pre prueba y la razón de verosimilitud. El nomograma tiene tres columnas, en la primera columna se tiene la probabilidad pre-prueba, la segunda o la del centro es la razón de verosimilitud y en la columna de la derecha se localiza la probabilidad post-prueba. Al observar la Figura 4.24 podemos notar que el médico ha establecido un diagnóstico de sospecha de 25%, es decir la probabilidad a priori, luego si un paciente obtuvo un test positivo con el nomograma podemos inferir que tendría la enfermedad con una probabilidad del 84%. Por otro lado, si el test fue negativo, se puede inferir que el paciente no tendrá la enfermedad con una probabilidad del 9%. En otras palabras con el nomograma se puede notar la probabilidad post-prueba.

Un ejemplo del uso del nomograma con tres pruebas para el diagnóstico de Sepsis se lo observa en Seqc [125]. Donde se detalla como a partir de una sospecha de la enfermedad (probabilidad a priori) se pasa a una probabilidad post-prueba e indicando además que prueba es más efectiva en la detección de la alteración.

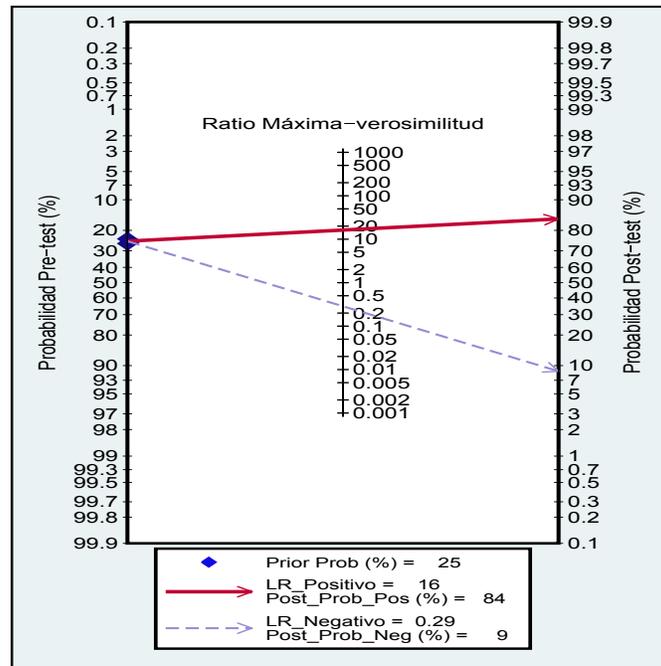


Figura 3.8: Gráfica de Fagan, datos midas_example

Para entender el Nomograma se debe entender la versión de la probabilidad del Teorema de Bayes, donde E, representa enfermedad, \bar{E} , significa ausencia de enfermedad y T es un resultado específico de la prueba; es decir, $Odds\ postprueba = Odds\ preprueba \times LR$, una manera de interpretarla matemáticamente es la que se observa en la ecuación 3.14.

$$\begin{aligned}
 O(E|T) &= O(E) \times \frac{P(T|E)}{P(T|\bar{E})} = \text{si tomamos logaritmo tenemos } \log(\text{Odds postprueba}) \\
 &= \log(O(E|T)) = \log(Oddspreprueba) + \log(LR) \\
 &= \log(O(E)) + \log\left(\frac{P(T|E)}{P(T|\bar{E})}\right) = P(preprueba) \times LR
 \end{aligned} \tag{3.14}$$

∴

$$\log(O(E|T)) = \log(P(preprueba)) + \log(LR)$$

- Odds Ratio Diagnóstica.-** Es un estadístico utilizado en los estudios epidemiológicos que mide la fuerza de asociación entre el factor de riesgo y la enfermedad. En nuestro caso es útil para denotar el grado de intensidad entre el resultado de una prueba y la enfermedad, el índice no se ve influenciado por la prevalencia y se estima como la razón entre odds de estar enfermo si la prueba da positivo y la odds de no estar enfermo si da negativo. Su estimación se obtiene mediante la ecuación 3.15.

$$\text{DOR} = \frac{\frac{VP}{FN}}{\frac{FP}{VN}} = \frac{LR+}{LR-} = \frac{\frac{VPP}{1-VPP}}{\frac{VPN}{1-VPN}} \quad (3.15)$$

Los valores DOR pueden variar desde cero hasta el infinito (cuanto más alto o grande sea DOR, mejor será la prueba). Si DOR=1, es decir, no es discriminante la prueba, la misma es inútil. Si $DOR \geq 1$, es más probable que la prueba de positivo en el caso de enfermos que en sanos, Seqc [125].

9. **Eficiencia diagnóstica de una prueba.-** Representa la proporción de resultados correctos entre el total de resultados. Es la probabilidad de que un paciente sea clasificado correctamente por la prueba; es decir, en el numerador aparecerán los enfermos con resultado positivo y los sanos con test negativo, mientras que en el denominador se encuentran todos los sujetos o pacientes. La eficiencia se puede categorizar en tres tipos: eficiencia, eficiencia si se considera la prevalencia y eficiencia estándar, nótese ecuaciones 3.16, 3.17 y 3.18 respectivamente.
10. **Curvas "receiver-operating characteristics" (ROC).-** Si el resultado de una prueba es una variable continua (cantidad en la sangre de glucemia en mg/dl), hay decidir cuál resultado es positivo para obtener el diagnóstico apropiado; entonces se debe establecer un punto de corte. El mismo determinará la sensibilidad y la especificidad de la prueba, Burgueño et al. [13]. Para establecer dicho punto de corte se recurre a las curvas de rendimiento diagnóstico (ROC). Es un gráfico que relaciona la sensibilidad y la especificidad. En el eje Y se ubican la sensibilidad o los verdaderos positivos, en el eje X se sitúan los falsos positivos ($1 - Es$).

Una prueba con discriminación perfecta, posee una curva ROC que pasa por la esquina superior izquierda, donde Se y Es toman sus máximos valores ($Se = Es = 1$). Una curva ROC que tiene una línea diagonal de 45° , desde la esquina inferior izquierda hasta la esquina superior derecha es característico de una prueba con bajo poder discriminatorio, véase Figura 3.9.

$$\frac{VP + VN}{VP + VN + FP + FN} = \frac{a + d}{a + b + c + d} \quad (3.16)$$

$$\frac{P(E) \times Se}{1 - P(E) \times Es} \quad (3.17)$$

$$\frac{Se + Es}{2} \quad (3.18)$$

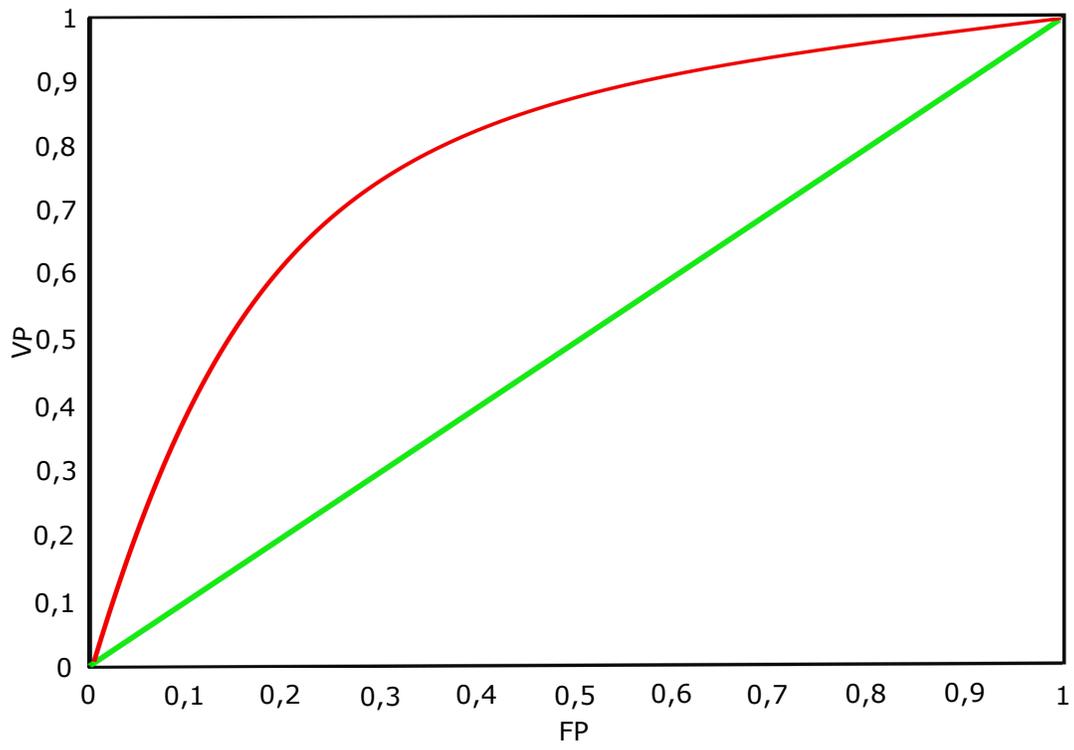


Figura 3.9: Curva ROC, datos simulados

3.2 ANÁLISIS DE SENSIBILIDAD Y ESPECIFICIDAD EN LAS MEDIDAS RESUMEN DEL ANÁLISIS DIAGNÓSTICO

”El camino no es un método; esto debe quedar claro. El método es una técnica, un procedimiento para obtener el control del camino y lograr que sea viable.”

Jacques Derrida

Los estudios sobre la precisión o exactitud diagnóstica son de vital importancia en la evaluación de las tecnologías diagnósticas médicas, es decir ayudan a la validez y reproducibilidad de los procedimientos que son utilizados por el personal de la salud para realizar la mejor aproximación posible a la condición real de los pacientes, Guyatt et al. [50], Knottnerus and Muris [68], Sackett and Haynes [115], y también en la precisión de los estudios, para medir el nivel de concordancia entre los resultados de una prueba, con relación a un patrón de referencia, en un proceso de evaluación. Existen diferentes medidas para la estimación de la precisión diagnóstica Griner et al. [49], Habbema et al. [51], pero la mayoría de los estudios actuales de la precisión diagnóstica se realizan mediante la estimación de la sensibilidad y/o especificidad o en combinación con otras medidas Honest and Khan [62].

Debido a que la gran mayoría de los trabajos de diagnóstica muestran estimaciones de sensibilidad y especificidad, el enfoque que ha seguido el meta-análisis se ha centrado en estas medidas Deeks [22], Honest and Khan [62], Irwig et al. [64], Kardaun and Kardaun [66], Lijmer et al. [78], Moses et al. [93], Rutter and Gatsonis [114], Walter and Jadad [146]. La integración de la sensibilidad y especificidad no es sencilla, debido a que están correlacionadas negativamente en muchos estudios; la necesidad de una revisión sistemática y la sintonización de las evidencias publicadas sobre la precisión de las pruebas diagnósticas, se han incrementado. La información de estas revisiones es un elemento clave en la toma de decisiones en el campo clínico y en las políticas de salud sobre el uso de pruebas diagnóstica; también es esencial para guiar el proceso de desarrollo de la tecnología y la evaluación del diagnóstico en la medicina.

La curva SROC (Summary Receiver Operator Characteristic - SROC), se ha convertido en el método principal en los estudios de meta-análisis, ya que informa tanto la sensibilidad como la especificidad, Hasselblad and Hedges [56], Irwig et al. [65], Kardaun and Kardaun [66], Littenberg and Moses [80], Midgette et al. [91], Moses et al. [93], Walter [144]. La curva SROC convierte cada par de valores de sensibilidad y especificidad en una medida de la precisión Glas et al. [45]. La desventaja de contar con una sola medida de la precisión diagnóstica, es que no tiene la capacidad para detectar a los enfermos (sensibilidad) y la identificación de los sanos (especificidad). La discriminación entre estas medidas (sensibilidad y especificidad) es muy importante para determinar el empleo adecuado (óptimo) de una prueba en la práctica clínica. El modelo bivalente que propone Reitsma et al. [105], tiene la ventaja de preservar la naturaleza bidimensional de los datos subyacentes y logra una estimación de la sensibilidad y especificidad, capturando una posible correlación (negativa) entre estas dos medidas.

Las revisiones diagnósticas empiezan con un conjunto de sensibilidades y especificidades de estudios individuales. Un enfoque intuitivo podría ser separar en dos conjuntos los valores de la sensibilidad y especificidad, usando el método estándar para proporciones, pero sin embargo la relación entre la sensibilidad y la especificidad es inversa (negativa) dentro de los estudios, por lo cual si se ignora este detalle

podría ser inapropiado, Moses et al. [93], Rutter and Gatsonis [114], Walter and Jadad [146]. Una posible causa de esta correlación negativa entre la sensibilidad y la especificidad es que los estudios pueden haber usado diferentes umbrales para definir resultados positivos y negativos de la prueba. En algunos casos, esto puede deberse a una diferencia explícita; por ejemplo, los estudios que utilizaron diferentes puntos de corte para clasificar a una medición (continua) bioquímica como positiva o negativa. En otras situaciones, se presentan variaciones implícitas en los umbrales entre los estudios, debido a diferencias en los observadores, laboratorios o equipos. A diferencia de otras fuentes de variación, una diferencia de umbral conduce a un patrón particular entre sensibilidad y especificidad. Este patrón es bien conocido a partir de estudios que notan el efecto de distintos puntos de cortes en pruebas bioquímicas con resultados continuos, Begg and McNeil [9], Hilden [60], Sorribas et al. [132]. Al reducir el valor del punto de corte, esto ocasionará más pacientes con un resultado positivo, no solo aumentará el número de verdaderos positivos, sino también el número de falsos positivos. Esto significa que la sensibilidad será mayor, a expensas de la especificidad. Esta relación negativa (trade-off) conduce a una curva cóncava, y esto ocurre cuando se grafica la sensibilidad versus uno menos la especificidad, lo cual genera la curva SROC. En muchos estudios que involucran la curva ROC, la sensibilidad se conoce como verdaderos positivo (tasa de verdaderos positivos) (TVP) y uno menos la especificidad como falsos positivos (tasa de falsos positivos) (TFP).

En los siguientes secciones se mostrará como la SROC, enfoque bivariante, modelo de regresión binomial, HSROC, y meta-análisis bayesiano para la precisión de un test, en ausencia de una prueba de referencia para la estimación de la sensibilidad y especificidad, pueden mostrar una posible correlación negativa. Es importante señalar que el enfoque SROC comienza con la representación de los pares ordenados de la sensibilidad y especificidad de cada estudio; la parte central de este enfoque es encontrar una curva suave mediante la unión de los pares ordenadas antes mencionados, asimismo es importante notar que el paso principal consiste en aplicar una transformación a las tasas de verdaderos positivos (TVP) (sensibilidad) y a las tasas de falsos positivos (TFP)(1-especificidad). Por otro lado el modelo bivariante, utiliza como punto de partida los conjuntos de puntos de la sensibilidad y especificidad de los meta-análisis, asume que los valores de la sensibilidad y especificidad de cada estudio (después de aplicarle la transformación logit) de un meta-análisis sigue una distribución normal.

En el modelo binomial, el número de pruebas positivas de cada estudio, siguen distribuciones binomiales independientes, en donde la probabilidad de una prueba positiva queda expresada como una transformación logit. El modelo de regresión binomial implica una relación lineal entre $\text{logit}(\text{TVP})$ y $\text{logit}(\text{TFP})$.

El análisis de regresión jerárquica es una extensión del modelo de regresión binomial para explicar más detalladamente la variabilidad dentro y entre estudios de las tasas de VP y FP. El método bayesiano, es una extensión del HSROC, asume que la variación de la sensibilidad y especificidad en los estudios aparece debido al empleo de diferentes puntos de corte para definir una prueba positiva y/o diferencias en la precisión del diagnóstico, cuando no existe una prueba de referencia.

En general, se puede decir que los métodos bayesianos no han sido aprovechados en gran medida en el campo de la salud. El método más común empleado en las pruebas diagnóstica es el SUMMARY RECEIVER OPERATOR CHARACTERISTIC (SROC). Las pruebas diagnósticas son ampliamente usadas en la medicina para determinar el estatus de una enfermedad, es decir, determinar si la enfermedad está presente o ausente. Las pruebas se basan en un resultado continuo subyacente, para determinar si un valor

está por encima de un umbral y así considerar indicios de la enfermedad, por tal motivo en la Sección 3.3 se analizan diferentes enfoques estadísticos para la valoración de una prueba diagnóstica.

3.3 MODELOS ESTADÍSTICOS PARA INTEGRACIÓN DE PRUEBAS

"La estadística es el único tribunal de apelación para juzgar el nuevo conocimiento."

P. C. Mahalanobis

En meta-análisis, el objetivo está en integrar los resultados de los diferentes artículos en los que se evalúa una prueba diagnóstica. La información de partida es una tabla tetracórica en la cual aparecen los VP, los FP, los FN y los VN. La tabla muestra la concordancia entre el resultado de la prueba en forma binaria y la información de la enfermedad. Cuando se trabaja en ensayos clínicos, el tamaño del efecto se mide generalmente en términos de Odd Ratio, incluso cuando se trabaja con pruebas diagnósticas, pero cuando estamos ante enfermedades de muy baja prevalencia (100 de cada 10 000 por ejemplo), donde además no existe una prueba oro y la decisión de que el sujeto está enfermo se hace a partir de un punto de corte subjetivo, las técnicas estándar no dan resultados óptimos.

Existen distintas alternativas en la literatura. A continuación se mostrará el enfoque univariante denominado SROC, el enfoque bivariante, el modelo de regresión binomial, el modelo jerárquico HSROC, y el meta-análisis bayesiano, todos ellos para evaluar la precisión de un test en ausencia de una prueba de referencia y para llevar a cabo la estimación de la sensibilidad y especificidad cuando ambas presentan una correlación significativa.

Supóngase que se tienen m estudios que contribuyen al vector z_i que denota las covarianzas en los niveles de estudio ($i = 1, 2, \dots, m$) y una tabla resumen de datos de 2×2 , que muestra la concordancia entre el resultado de la prueba en forma binaria y la información de la enfermedad de manera definitiva (o patrón de referencia) y sea la notación de la Tabla 3.2; donde, los resultados de los verdaderos y falsos positivos de las pruebas se definen en la ecuaciones 3.19 y 3.20 respectivamente.

	Enfermedad	
	Sí	No
Prueba diagnóstica	Sí=1	y_{i11} y_{i01}
	No=0	y_{i10} y_{i00}
Total	n_{i1}	n_{i0}

Tabla 3.2: Tabla resumen de datos

$$\widehat{VP} = \frac{y_{i11}}{n_{i1}} \quad (3.19)$$

$$\widehat{FP} = \frac{y_{i01}}{n_{i0}} \quad (3.20)$$

3.4 ENFOQUE SROC

Los test diagnósticos son ampliamente usados en medicina para determinar el status de una enfermedad; es decir, para determinar si la enfermedad esta presente o ausente. En general se supone que el test diagnóstico esta basado en un contenido latente para el cual, valores por encima de un punto de corte (specified threshold) nos llevan a declarar al individuo enfermo. Para investigar el efecto de un punto de corte en la sensibilidad y la especificidad, los investigadores presentan los resultados en forma de curva ROC (Receiver Operator Characteristic) donde la probabilidad de los verdaderos positivos (sensibilidad) es presentada en ordenadas frente a la probabilidad de falsos positivos (1- Especificidad) que se representa en abscisas.

Una forma de resumir el comportamiento de un test diagnóstico desde múltiples estudios es calculando la media de la sensibilidad y la media de la especificidad pero estas medidas resumen no son válidas si hay heterogeneidad entre los estudios y desafortunadamente los test para detectar heterogeneidad no son muy potentes Midgette et al. [91]. Además la sensibilidad y la especificidad dentro de cada estudio están inversamente relacionadas y dependen del punto de corte. Todo esto implica que la sensibilidad media y la especificidad media no son medidas adecuadas Irwig et al. [64]. Para corregir este posible sesgo, Moses et al. [93] desarrollaron la técnica para calcular la SROC (Summary ROC) la cual relaciona el punto de corte con la precisión de la prueba, vía regresión lineal.

El enfoque "Summary Receiver Operator Characteristic (SROC) curve approach" comienza con el trazado de los pares ordenados de la sensibilidad y especificidad de cada estudio. El objetivo principal es encontrar una curva suave a partir estos puntos. El paso principal es transformar la tasa de verdaderos positivos (TVP) (**sensibilidad**) y la tasa de falsos positivos (TFP) (**1-especificidad**) para que la relación se vuelva lineal y así poder realizar un ajuste para los puntos dados, Moses et al. [93].

Se utilizan las transformaciones 3.21 y 3.22 para TVP y TFP. Se define D como la diferencia de los valores de las transformaciones logit de TVP y TFP, mientras que S es la suma de estas transformaciones.

$$D_i = \ln \left(\frac{TVP_i}{1 - TVP_i} \right) - \ln \left(\frac{TFP_i}{1 - TFP_i} \right) = \ln(DOR) \quad (3.21)$$

$$S_i = \ln \left(\frac{TVP_i}{1 - TVP_i} \right) + \ln \left(\frac{TFP_i}{1 - TFP_i} \right) \quad (3.22)$$

$$D_i = \alpha + (\beta S_i) + \varepsilon_i \quad (3.23)$$

D es el logaritmo de la razón del odd ratio diagnóstico (diagnostic odds ratio) (**DOR**) para cada estudio. Es un indicador o índice de la precisión diagnóstica, y muestra con que frecuencia se produce un resultado positivo en los pacientes con la condición de interés, en comparación con los pacientes sin la enfermedad. S se refiere al umbral de cada prueba. Tiene un valor de cero en un estudio cuando la sensibilidad es igual a la especificidad. S es positiva en los estudios donde la sensibilidad es mayor que la especificidad, y es negativa cuando la especificidad toma valores más grandes que la sensibilidad. Una

recta de regresión lineal se ajusta a los puntos de las transformaciones 3.21 y 3.22, que nota como el logaritmo DOR varía en los umbrales implícitos. El modelo 3.23 es ajustado utilizando mínimos cuadrados ponderados o no ponderados, Moses et al. [93] y ε_i es el error aleatorio para el estudio i . La interpretación del intercepto y de la pendiente del modelo 3.23 se lo realiza con mayor detalle en la Sección 3.4.1. Cuando el DOR varía con S , el coeficiente β , no tiene una interpretación directa, pero influye en la forma de la curva ROC [144].

Algunos aspectos de la variabilidad entre los estudios pueden ser manejados dentro del enfoque SROC. Los niveles de covariación de los estudios se pueden incorporar de manera sencilla en la ecuación 3.23 para proporcionar un análisis exploratorio de los efectos de las características del estudio y de la especificidad de la prueba. El modelo puede ser estimado por mínimos cuadrados ordinarios o ponderados o algún método de regresión robusta. Los pesos pueden ser utilizados para cuantificar la diferencia que existen entre estudios, pero no captura de manera simultánea la diferencia en el tamaño de la muestra dentro de los pacientes con la enfermedad (n_{i1}) y los pacientes que no tienen la enfermedad (n_{i0}). Estos dos tamaños muestrales afectan a la precisión de las tasas de verdaderos positivos y falsos positivos. En la práctica, los modelos ponderados y no ponderados pueden producir resultados muy diferentes, y no hay una manera clara de elección entre ellos, Rutter and Gatsonis [114]. El punto Q^* es el punto en la curva SROC donde la sensibilidad y especificidad son iguales. Desafortunadamente, el punto Q^* puede dar valores de sensibilidad y especificidad fuera de los rangos establecidos en los estudios analizados. Estos puntos Q^* , suelen utilizarse para probar la diferencia de precisión (exactitud) global entre las pruebas diagnósticas. La justificación, es que el punto Q^* remueve el efecto de posibles diferencias en los umbrales comparando los odds ratio diagnóstico con valores específicos de S .

El coeficiente β en la ecuación 3.23 representa la dependencia de la precisión de la prueba en el umbral. Si $\beta \approx 0$ entonces los estudios pueden referirse como homogéneos, y pueden ser resumidos por un DOR global, notando que $\alpha = \ln(DOR)$. Si $\beta \neq 0$, los estudios son referidos como heterogéneos con respecto al DOR , en este caso el valor de α puede ser estimado como $\ln(DOR)$ cuando $S = 0$. El modelo está generalmente validado para el supuesto de distribución logística. Si los resultados de las pruebas subyacentes en realidad tienen distribución logística, S para un estudio dado puede ser expresado como una función del punto de corte que da lugar al análisis de la sensibilidad y especificidad para ese estudio, Moses et al. [93].

Si las varianzas de las distribuciones de los resultados de las pruebas de los casos y no casos son iguales, entonces la SROC es simétrica respecto a la línea diagonal, donde $VP = 1 - FP$. Si las varianzas son desiguales, la SROC resultante será asimétrica. No se hacen suposiciones acerca de la distribución de S . Las covariables se pueden añadir al modelo para evaluar si la exactitud de la prueba varía sistemáticamente con otros factores relacionados con el estudio, Deeks [22], Fahey et al. [38].

Si la estimación de la exactitud de la prueba para cada estudio se pondera por la inversa de su varianza, esto supone que el error de muestreo es la única fuente de variabilidad entre los estudios. Sin embargo, este supuesto puede ser poco realista, dado que los estudios tienden a variar en un sin número de aspectos, incluyendo el espectro de la enfermedad, las condiciones bajo las que se administró la prueba, y otros estudios y características de los pacientes que podrían afectar a la precisión de ensayo, Lijmer et al. [77]. A pesar de que los parámetros α y β son tanto fijos porque se supone que son constante en todos los estudios, aplicando pesos iguales a estudios (es decir, ajustando una regresión no ponderada)

se ha demostrado empíricamente que son consistentes con el supuesto de un efecto aleatorio, Moses et al. [93], Rutter and Gatsonis [113]. Esto es porque tanto las variaciones intra y la inter de los estudios se tienen en cuenta, lo que da relativamente más peso a los estudios más pequeños, como ocurriría en un modelo de efectos aleatorios.

Irwig et al. [64] recomienda el análisis no ponderado, porque los pesos basados en varianzas de las muestras pueden dar demasiado énfasis a los estudios inexactos, y por lo tanto dar una curva SROC sesgada. Moses et al. [93] también recomienda el enfoque no ponderado. Una vez que la regresión se ha ajustado, se pueden invertir las transformaciones, representadas por las ecuaciones 3.21 y 3.22 y por lo tanto deducir la relación entre VP y FP quedando expresado como la ecuación 3.24, Walter [144].

$$\begin{aligned}
 \ln\left(\frac{VP}{1-VP}\right) - \ln\left(\frac{FP}{1-FP}\right) &= \alpha + \beta \left(\ln\left(\frac{FP}{1-FP}\right) + \ln\left(\frac{FP}{1-FP}\right) \right) \\
 (1-\beta) \ln\left(\frac{VP}{1-VP}\right) &= \alpha + (1-\beta) \ln\left(\frac{FP}{1-FP}\right) \text{ si elevamos a la ecuación por } e \\
 \frac{VP}{1-VP} &= e^{\frac{\alpha}{1-\beta}} \cdot \left(\frac{FP}{1-FP}\right)^{\frac{1+\beta}{1-\beta}} \text{ si agrupamos por el término VP} \\
 \left(1 + e^{\frac{\alpha}{1-\beta}} \cdot \left(\frac{FP}{1-FP}\right)^{\frac{1+\beta}{1-\beta}}\right) \cdot VP &= e^{\frac{\alpha}{1-\beta}} \cdot \left(\frac{FP}{1-FP}\right)^{\frac{1+\beta}{1-\beta}} \text{ despejando por el término VP} \\
 \therefore \\
 VP &= \frac{e^{\frac{\alpha}{1-\beta}} \cdot \left(\frac{FP}{1-FP}\right)^{\frac{1+\beta}{1-\beta}}}{1 + e^{\frac{\alpha}{1-\beta}} \cdot \left(\frac{FP}{1-FP}\right)^{\frac{1+\beta}{1-\beta}}}
 \end{aligned} \tag{3.24}$$

La expresión 3.24 expresa TVP en términos de TFP y por lo tanto define toda la curva SROC. Si bien puede haber interés en identificar puntos particulares en la curva, también es útil tener una medida global resumen del comportamiento de la curva. Una medida adecuada es el área bajo la curva (AUC), que se puede calcular con la ecuación 3.25, Walter [144].

$$\int_0^1 \frac{e^{\frac{\alpha}{1-\beta}} \cdot \left(\frac{x}{1-x}\right)^{\frac{1+\beta}{1-\beta}}}{1 + e^{\frac{\alpha}{1-\beta}} \cdot \left(\frac{x}{1-x}\right)^{\frac{1+\beta}{1-\beta}}} dx \tag{3.25}$$

En general, la AUC debe ser estimada empleando algún método numérico para aproximar. La adopción de un índice resumen es útil para la información sucinta de un determinado conjunto de datos,

especialmente cuando los datos limitados impiden la identificación fiable de determinados puntos de la curva. La medida AUC es ampliamente utilizado en el análisis ROC, en el que se puede interpretar como la probabilidad de que los valores de prueba para un par aleatoria de individuos enfermos y no enfermos se clasifican correctamente. También representa el promedio (no ponderado) de TVP sobre todos los valores posibles de TFP . Las AUC son también un resumen candidato natural para un análisis SROC. El AUC parcial también se ha propuesto como una medida resumen, siendo este el área bajo una parte restringida de la curva correspondiente a los valores TFP de interés clínico, o en los que se encuentran los datos del estudio.

3.4.1 COMPORTAMIENTO EMPÍRICO DE SROC

La Figura 3.10 muestra un conjunto de tres curvas simétricas ROC con $\beta = 0$, que ocurre cuando los estudios son homogéneos y por lo tanto no presentan ninguna relación entre DOR y el umbral S . Las curvas de la Figura 3.10 se obtuvieron utilizando evaluación numérica de la ecuación 3.24 para $\alpha = 1, 5, \alpha = 2$, y $\alpha = 3$ que corresponden a $DOR = 4,5, 7,4$, y $20,1$, respectivamente. Al incrementarse el valor de α la curva SROC se acerca a su ideal, cerca de la esquina superior izquierda. Si $\alpha \rightarrow \infty$, $AUC \rightarrow 1$, lo que indicaría una prueba perfecta con 100% de sensibilidad y especificidad y no hay errores en la distinción de casos y no casos. En contraste, si $\alpha \simeq 0$ o $DOR = 1$ la curva se encuentra cerca de la diagonal $VP = FP$ en el espacio SROC, luego $AUC = \frac{1}{2}$, por lo que se infiere que la prueba no realiza ninguna mejora que el azar, Walter [144]. Para complementar, si $\alpha < 0$ es decir $DOR < 1$, la prueba discrimina casos y no casos en direcciones equivocadas, dando resultados peores que el azar. Si $DOR \rightarrow 0$, es decir $AUC \rightarrow 0$ la curva se encuentra cerca de la esquina inferior derecha del espacio SROC. Tales situaciones son poco probables que ocurra en la práctica, Walter [144].

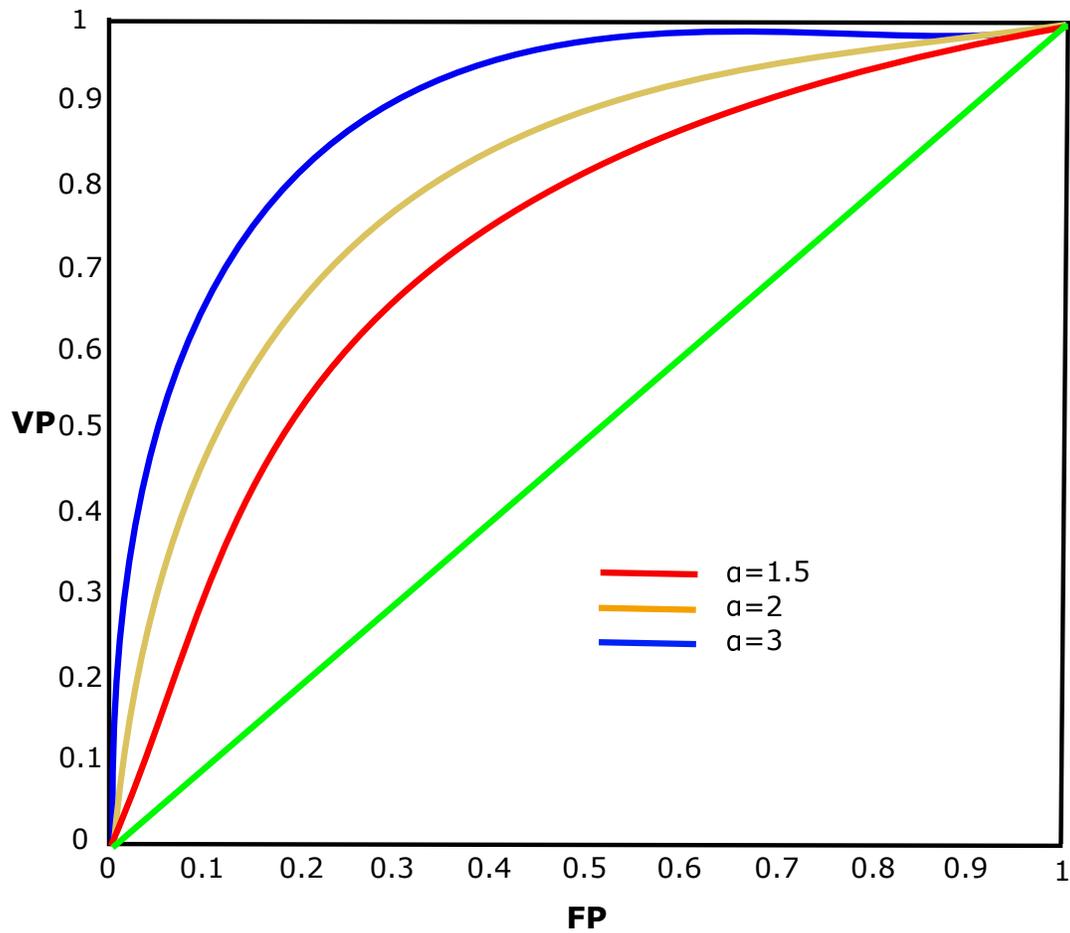


Figura 3.10: Curva SROC con varios valores de α ($\beta = 0$), datos simulados

Se considera el caso de estudios heterogéneos, modelo 3.23 , donde la precisión diagnóstica depende del umbral $\beta \neq 0$. La Figura 3.11 muestra tres curvas SROC derivadas de la ecuación 3.24 , todas con $\alpha = 2$, pero diferentes valores de β .

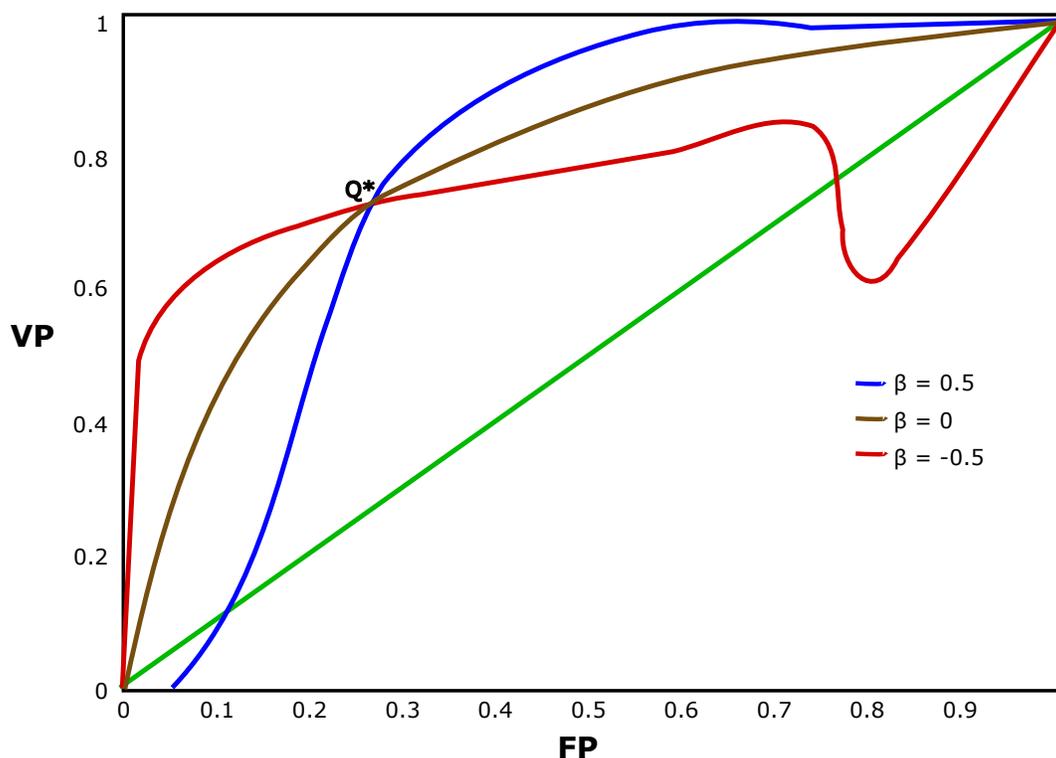


Figura 3.11: Curva SROC con varios valores de β ($\alpha = 2$), datos simulados

Los valores distintos de cero para β producen curvas asimétricas, si $\beta > 0$ tiene una elevación menos pronunciada que la curva simétrica con $\beta = 0$ hasta que la intercepta y a partir del punto de intersección se eleva más abruptamente para alcanzar valores relativamente altos de VP para valores altos de FP. Las curvas ROC con $\beta < 0$ exhiben el comportamiento opuesto, por ejemplo, la curva con $\beta = -0.5$ en la Figura 3.11. Curiosamente, como se sugiere en la Figura 3.11 y como lo muestra Moses et al. [93], la familia de curvas definidas por un valor fijo cruzan por un punto común, que se encuentra en la antidiagonal, donde $TVP = 1 - TFP$, o la sensibilidad es igual a la especificidad, teniendo como coordenadas las ecuaciones, 3.26 y 3.27, Walter [144] que son obtenidas mediante las ecuaciones 3.21, 3.22 y 3.23. Nótese que para obtener las ecuaciones siguientes se asume que la ecuación 3.22 es igual a cero.

$$\begin{aligned}
\ln\left(\frac{VP}{1-VP}\right) + \ln\left(\frac{FP}{1-FP}\right) &= 0 \\
\ln\left(\frac{VP}{1-VP}\right) - \ln\left(\frac{FP}{1-FP}\right) &= \alpha \text{ si sumamos las dos ecuaciones, se tiene que:} \\
2\ln\left(\frac{VP}{1-VP}\right) &= \alpha \text{ aplicando logaritmo a la ecuación, se tiene que:} \quad (3.26) \\
VP + VP\left(e^{\alpha/2}\right) &= e^{\alpha/2} \text{ despejando VP, se tiene que:} \\
&\vdots \\
VP &= \frac{e^{\alpha/2}}{1 + e^{\alpha/2}}
\end{aligned}$$

$$FP = \frac{1}{1 + e^{\alpha/2}} \quad (3.27)$$

Moses et al. [93] denotó el valor de VP en este punto por Q^* . En el análisis ROC, Q^* se ha sugerido como una medida resumen más. Corresponde con el punto más cercano a la esquina superior izquierda ideal del espacio SROC en curvas simétricas. El hecho de que todas las curvas SROC con un valor dado pasen por el mismo punto Q^* significa que Q^* no transmite ninguna información estadística adicional más allá de odds ratio. Sin embargo, el uso de Q^* puede estar motivado por su sencilla interpretabilidad, teniendo en cuenta que Q^* es el punto de la curva donde $VP = 1 - FP$; por lo tanto, representa el umbral de diagnóstico en el que la probabilidad de un diagnóstico correcto es constante para todos los sujetos.

Como consecuencia de asumir el modelo 3.23, $\beta \neq 0$, la curva SROC tiene una región en la que $VP < FP$, que se encuentra por debajo de la diagonal principal. Esta región puede verse, por ejemplo, cerca de la esquina inferior izquierda de la curva de $\beta = 0.5$ en la Figura 3.11. En esta región, la prueba teóricamente estaría clasificando los casos y no casos peor que al azar. Sin embargo, en la práctica, esta región es muy pequeña. Incluso si existe presencia relativamente fuerte de heterogeneidad con $\beta > 0$, la parte "impropia" de la curva SROC (donde $TVP < TFP$) incluye valores muy bajos de TVP, y estos valores corresponden a umbrales que es poco probable que sean aceptables en la práctica clínica. Si la heterogeneidad ocurre con $\beta < 0$, la parte impropia de la curva corresponde a valores altos de FP, lo que de nuevo no tiene relevancia clínica, vea Walter [144] para más detalle. Cuando $\beta \rightarrow 1$ la curva SROC ajustada se vuelve progresivamente más empinada, y en el caso límite en $\beta = 1$ degenera a una línea vertical. Esto podría ocurrir, por ejemplo, si no hay variación en FP entre los estudios. Sin embargo, pasaría por el punto común Q^* definida por las ecuaciones 3.26 y 3.27. Si $\beta > 1$, se invierte la curva y muestra una relación negativa de VP y FP. Esto es plausible, en la práctica, excepto tal vez por casualidad en muestras pequeñas. Un comportamiento similar se observa si β es cerca o por debajo de 1. Si es cercano a -1, la curva es horizontal, indicando que no hay relación de VP y FP, esto podría ocurrir si no hubiera variación en VP entre los estudios. Si $\beta < -1$, la curva se invierte (a una forma diferente) y sugiere una negativa relación entre VP y FP. Todas las curvas $|\beta| > 1$ posee un valor común Q^* , dado por las ecuaciones 3.26 y 3.27. En la práctica, valores $|\beta| > 1$ son poco probables, Walter [144]. La experiencia empírica sugiere que los conjuntos de datos de meta-análisis práctico a menudo tienen β cercano a 0, y son raras veces mayor que 0.5 en valor absoluto, Walter [144].

3.4.2 MEDIDAS RESUMEN. ÁREA BAJO LA CURVA Y Q^*

El AUC es un índice popular del rendimiento global de la prueba, Beck and Shultz [7], DeLong et al. [24], los rangos de las AUC van, desde 1 para una prueba perfecta que diagnostica correctamente siempre; a 0, para una prueba que no diagnostica correctamente Hanley [52], Hanley and McNeil [53], Hilden [60], Metz et al. [90]. En los estudios individuales, AUC se puede interpretar como la probabilidad de que la prueba categoriza correctamente un caso de los no casos, n elegido al azar con respecto a sus valores de prueba, Bamber [6]. Se propone el área bajo la curva (AUC) para que cumpla con la misma función que en los meta-análisis, suponiendo que la curva SROC transmite con precisión el rendimiento de pruebas a nivel de los sujetos que se analizan en la prueba. Aunque se requiere la integración numérica en general para obtener AUC, algunos casos especiales son de interés porque con ellos se obtienen expresiones analíticas exactas y resultados comparativos. Se sabe que AUC se incrementa con el valor de α , manteniendo fijo β . Mediante el examen de la ecuación 3.25, se puede demostrar que para un valor dado de α , AUC se maximiza cuando $\beta = 0$, lo que implica que el AUC es óptima en los estudios "grandes" homogéneos, Walter [144]. Por otra parte, el AUC es simétrica en β , por lo que los valores negativos de β producen el mismo valor de AUC en valor absoluto. Si $\alpha = \beta = 0$ de la ecuación 3.25, se tiene que $\int_0^1 x dx = \frac{1}{2}$. Es evidente que si $AUC = \frac{1}{2}$ cuando $\alpha = 0$, inclusive si $\beta \neq 0$, así que cuando $AUC = \frac{1}{2}$ indica el rendimiento general al azar para cualquier conjunto de estudios. En el caso homogéneo $\beta = 0$, la ecuación 3.25 queda expresada como la ecuación 3.28, Walter [144], obteniendo una solución exacta, mediante la ecuación 3.29, Walter [144].

$$\int_0^1 \frac{e^{\alpha \cdot \left(\frac{x}{1-x}\right)}}{1 + e^{\alpha \cdot \left(\frac{x}{1-x}\right)}} dx \quad (3.28)$$

$$\int_0^1 \frac{e^\alpha \cdot \left(\frac{x}{1-x}\right)}{1 + e^\alpha \cdot \left(\frac{x}{1-x}\right)} dx = e^\alpha \int_0^1 \frac{\frac{x}{\cancel{1-x}}}{\frac{1-x + e^\alpha x}{\cancel{1-x}}} dx = e^\alpha \int_0^1 \frac{x}{1+x(e^\alpha-1)} dx$$

$u = 1 + x(e^\alpha - 1)$; realizando cambio de variable

$$dx = \frac{du}{e^\alpha - 1}$$

si $x = 0$ se tiene que

$$u = 1$$

si $x = 1$ se tiene que

$u = 1 + e^\alpha - 1 = e^\alpha$; luego la integral queda expresada como

$$\begin{aligned} &= \frac{e^\alpha}{(e^\alpha - 1)^2} \int_1^{e^\alpha} \frac{u-1}{u} du = \frac{e^\alpha}{(e^\alpha - 1)^2} \left[\int_1^{e^\alpha} du - \int_1^{e^\alpha} \frac{du}{u} \right] \\ &= \frac{e^\alpha}{(e^\alpha - 1)^2} \left[u \Big|_1^{e^\alpha} - \ln(u) \Big|_1^{e^\alpha} \right] = \frac{e^\alpha}{(e^\alpha - 1)^2} [e^\alpha - 1 - \ln(e^\alpha)] \end{aligned}$$

\therefore

$$AUC_{hom} = \frac{DOR}{(DOR - 1)^2} \cdot [(DOR - 1) - \ln(DOR)]$$

(3.29)

Donde AUC_{hom} indica el AUC para estudios homogéneos, y $DOR = e^\alpha$. Si $\alpha = 0$ u $DOR = 1$, entonces $AUC_{hom} = \frac{1}{2}$. La expresión 3.29 puede ser usada para evaluar AUC para estudios homogéneos, sin necesidad de utilizar integración numérica. La Tabla 3.3 muestra el valor de AUC, Q^* , y su diferencia para un rango de valores de DOR en el caso homogéneo. Note que AUC disminuye con el aumento de β , y que la curva cuando $\beta \rightarrow 1$ pasa a través del punto común Q^* , a partir de las ecuaciones 3.26 y 3.27, se deduce un límite inferior para el AUC en curvas con un determinado valor de α , que se expresa mediante la ecuación 3.30, Walter [144], que es equivalente a la ecuación 3.26.

$$Q^* = \frac{e^{\alpha/2}}{1 + e^{\alpha/2}} = \frac{\sqrt{OR}}{1 + \sqrt{OR}} \quad (3.30)$$

Odds ratio	AUC_{hom}	Q^*	$AUC_{hom} - Q^*$
0.5	0.386	0.414	-0.028
1	0.500	0.500	0.000
1.5	0.567	0.511	0.017
2	0.614	0.586	0.028
3	0.676	0.634	0.042
4	0.717	0.667	0.051
5	0.747	0.691	0.056
10	0.827	0.760	0.067
20	0.887	0.817	0.069
30	0.913	0.846	0.068
40	0.929	0.863	0.065
50	0.939	0.876	0.063

Tabla 3.3: AUC_{hom} , Q^* , y sus diferencias para varios valores de odds ratio diagnóstico: caso homogéneo

El valor de Q^* de la ecuación 3.30 y el máximo valor de AUC_{hom} de la ecuación 3.29 proveen límites inferiores y superiores fácilmente computables para AUC para cualquier valor dado de $\alpha > 0$. Las evaluaciones numéricas de las expresiones 3.29 y 3.30 demuestran que, en el caso homogéneo, la diferencia entre AUC y Q^* aumentan para valores moderados de DOR , pero nunca es más de alrededor del 7%, Walter [144]. La diferencia máxima se produce con un valor de $\alpha = 2.85$ (u $DOR = 17.3$), que es el valor solución de una ecuación trascendental. Para valores mayores de α , la diferencia disminuye muy lentamente, con un valor límite $AUC_{hom} - Q^* = 0$ cuando $\alpha \rightarrow \infty$. La integración numérica de la ecuación 3.25 para el caso heterogéneo permite evaluar el impacto de la heterogeneidad en el valor de AUC. Para un valor fijo de DOR , AUC disminuye lentamente a medida que se incrementa β desde cero (estudios homogéneos) a valores más grandes (el aumento de heterogeneidad). Sin embargo, la dependencia de las AUC en β es débil, y el efecto dominante en el AUC es el valor de DOR (o α). Para $|\beta| < 0.4$, el cambio porcentual en comparación con el AUC caso homogéneo es de menos de 2%. En consecuencia, AUC_{hom} proporciona una buena aproximación a AUC incluso en estudios heterogéneos Walter [144].

3.4.3 ERROR ESTÁNDAR DE AUC Y Q^*

Considérese la variación muestral de \widehat{AUC} , de la ecuación 3.25. Nótese que es una función de regresión de los parámetros α y β y que por lo tanto la variabilidad en \widehat{AUC} es una función de regresión de la variación de $\hat{\alpha}$ y $\hat{\beta}$, usando el método delta² para obtener una aproximación a la varianza de \widehat{AUC} , se tiene la ecuación 3.31, Walter [144] y de la ecuación 3.25 se obtienen las expresiones 3.32 y 3.33, Walter [144].

²Se emplea para obtener la distribución asintótica de una función de variables aleatorias usando la primera derivada para obtener una aproximación por una serie de Taylor de primer orden.

$$var(\widehat{AUC}) = \left(\frac{\partial AUC}{\partial \alpha}\right)^2 var(\hat{\alpha}) + \left(\frac{\partial AUC}{\partial \beta}\right)^2 var(\hat{\beta}) + 2\left(\frac{\partial AUC}{\partial \alpha}\right)\left(\frac{\partial AUC}{\partial \beta}\right)cov(\hat{\alpha}, \hat{\beta}) \quad (3.31)$$

$$\frac{\partial AUC}{\partial \alpha} = \left(\frac{1}{1-\beta}\right) e^{\frac{\alpha}{1-\beta}} \int_0^1 \frac{\left(\frac{x}{1-x}\right)^p}{\left[1 + \left(\frac{x}{1-x}\right)^p e^{\frac{\alpha}{1-\beta}}\right]^2} dx \quad (3.32)$$

$$\frac{\partial AUC}{\partial \beta} = \left(\frac{1}{1-\beta}\right)^2 e^{\frac{\alpha}{1-\beta}} \int_0^1 \frac{\left(\frac{x}{1-x}\right)^p \left[\alpha + 2 \ln\left(\frac{x}{1-x}\right)\right]}{\left[1 + \left(\frac{x}{1-x}\right)^p e^{\frac{\alpha}{1-\beta}}\right]^2} dx \quad (3.33)$$

Donde $p = \frac{1+\beta}{1-\beta}$. En general la evaluación de la $var(\widehat{AUC})$, requiere integración numérica sin embargo, para el caso especial de los estudios homogéneos donde $\beta = 0$, una expresión es posible utilizando el método delta, que se podrá obtener mediante la ecuación 3.34, Walter [144]. De la ecuación 3.34 se tiene que $SE(\widehat{AUC}_{hom})$ es simétrico con $\ln(DOR)$ para valores $DOR > 1$. Cuando $DOR = 1$, la expresión 3.34 se degenera, pero se puede utilizar la regla de L'Hôpital y se puede demostrar que en la vecindad de $DOR = 1$, se tiene la expresión $SE(\widehat{AUC}) \approx \frac{SE(\hat{\alpha})}{6}$, Walter [144]. El método delta también provee una estimación para el error estándar de Q^* , donde se tiene la expresión $SE(\hat{Q}^*) = \frac{\sqrt{DOR}}{2(\sqrt{DOR} + 1)} SE(\hat{\alpha})$, Walter [144].

$$SE(\widehat{AUC}_{hom}) = \frac{DOR}{(DOR - 1)^3} [(DOR + 1) \ln(DOR) - 2(DOR - 1)] SE(\hat{\alpha}) \quad (3.34)$$

3.5 MODELO BIVARIANTE

Hay que señalar que el modelo SROC no cuantifica el error en S, que puede ser una fuente de sesgo en las estimaciones de parámetros y los resúmenes de la curva, Baker [5]. Un enfoque alternativo para la construcción de una curva de SROC fue propuesto por Kardaun and Kardaun [66], que supone que $\left(\ln\left(\frac{TVP_i}{1-TVP_i}\right); \ln\left(\frac{TFP_i}{1-TFP_i}\right)\right)$ sigue una distribución normal bivalente y postuló una relación lineal entre los dos componentes de la media bivalente. El principio de máxima verosimilitud³ es empleado en este enfoque para la estimación de la pendiente y el intercepto de este modelo. La desventaja del uso del DOR como parámetro de resultado para las estimaciones es que al combinar sensibilidad y especificidad en un único indicador se pierde el valor relativo de cada uno de éstos.

El modelo bivalente utiliza un punto de partida diferente para los pares ordenados de sensibilidad y especificidad de los meta-análisis. En vez de transformar estas dos medidas en un indicador diagnóstico de la precisión, como en el enfoque SROC, el modelo bivalente preserva las dos dimensiones de los datos a través del análisis. El modelo bivalente se fundamenta en el siguiente razonamiento Kardaun and Kardaun [66], Van Houwelingen et al. [142] y Van Houwelingen et al. [143].

Se asume que los valores de la sensibilidad de los estudios individuales (después de la transformación logit) de un meta-análisis siguen una distribución normal alrededor del valor de la media y con cierta cantidad de variabilidad alrededor de la misma media. Este es un enfoque de efectos aleatorios, similar al que se usa en los ensayos terapéuticos, para incorporar la variabilidad no explicada en el análisis. Esta variación en la sensibilidad subyacente entre los estudios puede estar relacionada con las diferencias que subsisten en la población de estudio, las diferencias implícitas en el umbral, o variaciones imperceptibles en el protocolo del ensayo clínico. Las mismas consideraciones se aplican a las especificidades de estos estudios. La presencia potencial de una correlación (negativa) entre la sensibilidad y la especificidad dentro de los estudios se aborda mediante la incorporación explícita de esta correlación en el análisis. La combinación de la distribución normal y las transformaciones logit de la sensibilidad y especificidad, reconocen la posible correlación entre ellas, lo que lleva a una distribución normal bivalente Kotz et al. [70], Van Houwelingen et al. [142]. Además de la variabilidad entre los estudios en las sensibilidades y especificidades subyacentes, también existe variación debida al muestreo. Si los estudios difieren en tamaño, la variación es debida probablemente al tamaño de la muestra (muestras pequeñas). El modelo bivalente incorpora la precisión, donde la sensibilidad y especificidad se miden en cada estudio mediante el enfoque de Van Houwelingen et al. [142]. Es decir, en estudios con una estimación más precisa de la sensibilidad, tendrán peso más alto en el análisis de sensibilidad. Lo mismo ocurre para estudios con estudios más precisos de la especificidad. Una descripción más técnica del modelo bivalente es la siguiente.

Sean los estudios individuales ($I = 1, 2, 3, \dots, k$) que muestran sensibilidad $p(A_i)$, determinada en N_A individuos con la condición de interés y especificidad $p(B_i)$ medida en N_B individuos sin esta condición de interés. Sea $\theta_{A,i}$ la transformación logit de la sensibilidad en el estudio i , y $\theta_{B,i}$ la transformación logit de la especificidad. Supóngase que los verdaderos valores logit de sensibilidad de estos estudios se distribuyen

³Sea θ , las cantidades vectoriales no observables o parámetros poblacionales (probabilidad de supervivencia a cada tratamiento), \mathbf{y} datos observados (número de supervivientes y muertos en los grupos de estudio). La inferencia bayesiana obedece al principio de verosimilitud, es decir para una muestra de datos, dos modelos con la misma función de verosimilitud $P(\mathbf{y}|\theta)$ producen la misma inferencia sobre θ , es decir la función de verosimilitud contiene toda la información relevante que aportan los datos. Este principio acepta que la inferencia está condicionada a los datos observados.

normalmente alrededor de algún valor medio θ_A , con una variabilidad entre estudios de σ_A^2 . Se asume el supuesto de efectos aleatorios para la especificidad de los estudios; donde θ_B , denota el valor medio *logit* de la especificidad y σ_B^2 la varianza entre estudios de los *logit* de las especificidades.

Se incorpora de manera explícita la posible correlación entre la sensibilidad (*logit*) y especificidad dentro de los estudios. La combinación de las dos distribuciones normales que se pueden correlacionar lleva al modelo normal bivariante, 3.35. Nótese la representación de Σ en la ecuación 3.36. Es importante notar que Σ y σ_{AB} representan las covarianzas de la transformación *logit* de la sensibilidad y especificidad. El modelo bivariante se extiende mediante la incorporación de la precisión de la sensibilidad y especificidad que se ha estimado en cada estudio. Cuando N_A y N_B son "grandes" y $0 < p_{A,i}; p_{B,i} < 1$, se tiene que la estimación correspondiente para la varianza de los *logit* de la sensibilidad y especificidad en cada estudio viene dada por las ecuaciones 3.37 y 3.38 respectivamente. Si se trata a las varianzas observadas de los *logit* de la sensibilidad y especificidad como cantidades fijas, un enfoque estándar en meta-análisis se puede escribir como la ecuación 3.39, donde C_i es una matriz diagonal de S_i^2 , 3.40. Este modelo puede ser ajustado por el método de máxima verosimilitud, utilizando SAS, proc mixed, R mada, stata entre otros. La salida estándar del modelo bivariante incluye, la media *logit* de la sensibilidad θ_A y especificidad θ_B con sus errores estándar y sus intervalos de confianza al 95%; y las estimaciones entre estudios de la variabilidad de los *logit* de la sensibilidad σ_A^2 y especificidad σ_B^2 y la covarianza entre ellas σ_{AB} .

$$\begin{pmatrix} \theta_{A,i} \\ \theta_{B,i} \end{pmatrix} \sim N \left(\begin{pmatrix} \theta_A \\ \theta_B \end{pmatrix}, \Sigma \right) \quad (3.35)$$

$$\Sigma = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix} \quad (3.36)$$

$$S_{A,i}^2 = \frac{1}{n_{A,i} p_{A,i} (1 - p_{A,i})} \quad (3.37)$$

$$S_{B,i}^2 = \frac{1}{n_{B,i} p_{B,i} (1 - p_{B,i})} \quad (3.38)$$

$$\begin{pmatrix} \hat{\theta}_{A,i} \\ \hat{\theta}_{B,i} \end{pmatrix} \sim N \left(\begin{pmatrix} \theta_A \\ \theta_B \end{pmatrix}, C_i \right) \quad (3.39)$$

$$C_i = \begin{pmatrix} S_{A,i}^2 & 0 \\ 0 & S_{B,i}^2 \end{pmatrix} \quad (3.40)$$

$$LR+ = \frac{\frac{\exp(\theta_A)}{1 + \exp(\theta_A)}}{1 - \frac{\exp(\theta_B)}{1 + \exp(\theta_B)}} \quad (3.41)$$

$$LR- = \frac{1 - \frac{\exp(\theta_A)}{1 + \exp(\theta_A)}}{\frac{\exp(\theta_B)}{1 + \exp(\theta_B)}} \quad (3.42)$$

Basado en los parámetros antes descritos se pueden calcular otras medidas de interés, como, los ratios de máxima verosimilitud para test positivos y negativos, 3.41 y 3.42, la correlación entre el logit de la sensibilidad y especificidad, estimada como $\frac{\sigma_{AB}}{\sqrt{\sigma_A^2}\sqrt{\sigma_B^2}}$. Los parámetros del modelo bivalente se estiman

en un modelo único que incorpora una posible correlación entre la sensibilidad y especificidad. Hasta cierto punto, los modelos de diagnóstico bivalente pueden ser expresados como un análisis longitudinal con dos mediciones (sensibilidad y especificidad) dentro de cada persona (correspondiente a estudios individuales), Reitsma et al. [105]. El modelo bivalente puede ser visto como una mejora y ampliación del enfoque SROC, ya que puede producir los siguientes resultados, Reitsma et al. [105]:

- Estima la cantidad de variación entre los estudios mediante la sensibilidad y especificidad; además provee el grado de correlación entre la sensibilidad y la especificidad. Esto proporciona información importante acerca de los resultados de heterogeneidad entre los estudios donde se pueda notar una diferencia de un umbral implícito.
- Produce estimaciones de sensibilidad y especificidad con sus respectivos intervalos de confianza al 95%. Estos intervalos toman en cuenta la heterogeneidad entre los estudios más allá del azar (modelo de efectos aleatorios).
- Los parámetros de la distribución bivalente permiten calcular ya sea una elipse de confianza alrededor de las medias de los logit de los valores de sensibilidad y especificidad o una elipse de predicción para los valores sensibilidad y especificidad, teniendo en cuenta la posible correlación (negativa) entre ambas.
- Se pueden calcular otras medidas derivadas de la sensibilidad y especificidad, como los odds ratio de diagnóstico y las razones de máxima verosimilitud.
- Las covariaciones pueden ser agregadas al modelo bivalente, lo cual conduce a separar los efectos sobre la sensibilidad y especificidad, pero los efectos sobre los odds ratio de diagnóstico estarán presentes. Esto quiere decir que se puede probar de manera explícita si la sensibilidad o especificidad o ambas son diferentes entre dos tecnologías de diagnóstico.

3.6 MODELO DE REGRESIÓN BINOMIAL

Un modelo de regresión para el estudio en el campo del meta-análisis para relacionar el par ordenado de $(VP_i; FP_i)$ se discutió por primera vez en la referencia, Rutter and Gatsonis [114] y fue motivado por la formulación de un análisis de regresión ordinal para la curva ROC, Baker [5], McCullagh [88], Tosteson and Begg [139]. Si W denota el grado de sospecha de la presencia de una anomalía, denotada en una escala categórica ordinal con J categorías, el modelo paramétrico ROC es equivalente al modelo de regresión ordinal $g = (\theta_j + \alpha x) \exp(-\beta x)$. Donde x es una covariable que denota el verdadero estado de la enfermedad (binario). La base conceptual del modelo es la suposición de las respuestas observadas W que representan una categorización de una variable latente, con una función enlace $g(\cdot)$ que relaciona la parte aleatoria y sistemática del modelo. El probit link (que transforma probabilidades en valores estándar de la función de distribución normal) es una variable latente gaussiana que se utiliza comúnmente para estudios individuales de la curva ROC, Harbord et al. [54]. Se usa un logit link bajo el supuesto de un modelo logit, para estimar los parámetros de la regresión mediante los logaritmos de los odds ratios. En el ajuste binomial con $J = 2$, el uso de la relación logit no requiere la asunción de una variable latente subyacente-log normal. Un modelo de regresión ordinal con una logistic link y $J = 2$, se puede utilizar en los estudios de meta-análisis que reportan el par ordenado $(VP_i; FP_i)$, Rutter and Gatsonis [114].

Bajo el modelo binomial, el número de pruebas positivas de cada estudio, $y_{ij1}, i = 1, 2, \dots, m, j = 0, 1$, siguen distribuciones binomiales independientes, es decir $y_{ij1} \sim Binomial(n_{ij}, \pi_{ij})$, en el que la probabilidad de una prueba positiva queda expresada o modelizada mediante la ecuación 3.43. Donde θ_i representan los puntos de corte, y los valores de VP y FP se incrementan al incrementar los puntos de corte. Por lo tanto, los puntos de corte explican la relación negativa (trade-off) entre los VP y FP. Llamamos a α , parámetro de precisión, ya que mide la diferencia entre VP y FP. Cuando $\beta = 1$, α modeliza al logaritmo de los odds ratio como constante para una prueba positiva de un paciente que tiene la enfermedad versus un paciente que no tiene la enfermedad. Llamamos β el parámetro de escala, ya que permite notar la diferencia que existe en las varianzas de los pacientes con y sin la enfermedad de una población de estudio. Cuando $\beta \neq 1$, los valores de VP y FP aumentan a un ritmo diferente al variar el valor de θ , lo que produce curvas ROC asimétricas. El modelo de regresión binomial se estima por máxima verosimilitud y puede cuantificar una tasa de error para los valores de VP y FP.

El modelo de regresión binomial 3.43 implica una relación lineal entre $\text{logit}(VP)$ y $\text{logit}(FP)$. El modelo 3.43 asume que en todos los estudios, las pruebas (test) tienen la misma precisión y parámetros de escala, con diferencias entre los estudios debido a los diferentes umbrales de positividad θ_i . Estos modelos de regresión binomial se pueden formular para permitir efectos de covarianza de los niveles de estudio, Rutter and Gatsonis [114]. Sin embargo, debido a que cada estudio contribuye sólo con dos grados de libertad, existen limitaciones prácticas sobre el número de covariables que pueden ser incluidas en estos modelos.

$$\text{logit}(\pi_{ij}) = (\theta_i + \alpha_i X_{ij}) \exp(-\beta X_{ij}) \quad (3.43)$$

3.7 REGRESIÓN JERÁRQUICA PARA META-ANÁLISIS DE LAS EVALUACIONES EN LA PRECISIÓN DE LAS PRUEBAS DIAGNÓSTICAS (HSROC)

El objeto del modelo jerárquico es obtener estimaciones resúmenes significativas de la sensibilidad y especificidad y propone un mejor manejo de la variabilidad dentro y entre estudio, aplicando un enfoque bayesiano para la estimación de los parámetros. Este enfoque bayesiano tiene un precio, se basa en las cadena de Markov⁴, simulaciones de Monte Carlo⁵, lo cual requiere programación, simulaciones y una evaluación de la convergencia del modelo y una adecuación y síntesis de los resultados de la simulación. El modelo se divide en tres niveles, Rutter and Gatsonis [114].

3.7.1 NIVEL I (VARIACIÓN DENTRO DEL ESTUDIO)

El número de falsos y verdaderos positivos de la prueba del estudio i -ésimo, y_{i01} e y_{i11} , se suponen que son independientes y siguen una distribución binomial, donde la probabilidad de que la prueba sea positiva viene dada por la ecuación 3.43. Donde X_{ij} , denota el verdadero estado de la enfermedad para los casos que se encuentran en la ij -ésima celda. Bajo este enfoque jerárquico (HSROC), se permiten criterios de positividad, θ_i y un parámetro de exactitud α_i , que pueden variar en cada estudio, ecuación 3.43. Debido a que la estimación del parámetro de escala β requiere información de más de un estudio, β se supone que es constante en todos los estudios. Si se permite que β varíe en todos los estudios, los parámetros en cada estudio serían identificables o estimados sólo a través de sus distribuciones a priori.

3.7.2 NIVEL II (VARIACIÓN ENTRE ESTUDIOS)

Los parámetros de los niveles de estudio en 3.43, α_i y θ_i , se asumen que están distribuidos normalmente, con una media determinada por una función lineal de las covariaciones de los niveles de estudio. En el caso de una sola covarianza Z que afecta al punto de corte y al parámetro de precisión, el modelo puede ser escrito como se nota en la ecuación 3.44, Rutter and Gatsonis [114]. Los coeficientes de γ y λ del modelo sistemático se diferencian en el criterio de positividad y precisión en todos los estudios, debido a la covarianza Z . Sin embargo, se puede hacer una formulación más general del modelo, es decir, se incluye más de una covariable y diferentes entre sí, que se utilizan para los puntos de corte y precisión en las ecuaciones de la regresión.

$$\left. \begin{aligned} \theta_i | \Theta, \gamma, Z_i, \sigma_\theta^2 &\sim N(\Theta + \gamma Z_i, \sigma_\theta^2) \\ \alpha_i | \Lambda, \lambda, Z_i, \sigma_\alpha^2 &\sim N(\Lambda + \lambda Z_i, \sigma_\alpha^2) \end{aligned} \right\} \text{Condicionales independientes} \quad (3.44)$$

3.7.3 NIVEL III

La especificación del modelo jerárquico se completa con la elección de las distribuciones a priori de los parámetros restantes, 3.45, Rutter and Gatsonis [114]. Los parámetros Θ , Λ , β , γ , λ , σ_θ^2 y σ_α^2 se asume

⁴Una cadena de Markov es una sucesión de ensayos similares u observaciones en la cual cada ensayo tiene el mismo número finito de resultados posibles y en donde la probabilidad de cada resultado para un ensayo dado depende sólo del resultado del ensayo inmediatamente precedente y no de cualquier resultado previo.

⁵El método Montecarlo es un método numérico que permite resolver problemas físicos y matemáticos mediante la simulación de variables aleatorias. El método MonteCarlo fue bautizado así por su clara analogía con los juegos de ruleta de los casinos, el más célebre de los cuales es el de Montecarlo, casino cuya construcción fue propuesta en 1856 por el príncipe Carlos III de Mónaco, siendo inaugurado en 1861.

que son mutuamente independientes y los restantes parámetros μ_θ , μ_γ , ξ_θ , μ_α , μ_λ , ξ_α y μ_β se fijan de tal manera que reflejen rangos plausibles. La curva SROC, puede ser obtenida a partir de los valores esperados de $\Lambda + \lambda Z$ y β . Si se le asigna el valor de $\frac{1}{2}$ para los casos que tienen la enfermedad y $-\frac{1}{2}$ para los casos que no tienen la enfermedad, para un valor dado de covarianza Z_i , la tasa de verdaderos positivos queda expresada en 3.46, Rutter and Gatsonis [114]. Luego la curva SROC es dibujada, colocando en un plano los pares $(FP, VP(FP))$, para $FP \in [0, 1]$.

$$\left. \begin{array}{l} \Theta \sim Uniforme [\mu_{\theta_1}, \mu_{\theta_2}] \\ \gamma \sim Uniforme [\mu_{\gamma_1}, \mu_{\gamma_2}] \\ \sigma_\theta^2 \sim \Gamma^{-1} [\xi_{\theta_1}, \xi_{\theta_2}] \\ \Lambda \sim Uniforme [\mu_{\alpha_1}, \mu_{\alpha_2}] \\ \lambda \sim Uniforme [\mu_{\lambda_1}, \mu_{\lambda_2}] \\ \sigma_\alpha^2 \sim \Gamma^{-1} [\xi_{\alpha_1}, \xi_{\alpha_2}] \\ \beta \sim Uniforme [\mu_{\beta_1}, \mu_{\beta_2}] \end{array} \right\} \text{Mutuamente independientes} \quad (3.45)$$

$$VP(FP) = \text{logit}^{-1} \left[\left(\text{logit}(FP) \exp \left\{ \frac{E(\beta)}{2} \right\} + E(\Lambda + \lambda Z_i) \right) \exp \left\{ \frac{E(\beta)}{2} \right\} \right] \quad (3.46)$$

3.7.4 AJUSTE DEL MODELO HSROC

La inferencia del modelo HSROC se fundamenta en una distribución a posteriori de los parámetros del modelo. Debido a que los modelos que se consideran no son conjugados⁶, es decir, son expresiones de forma cerradas, las distribuciones a posteriori no existen. Luego estos valores a posteriori son estimados mediante una simulación de la distribución a posteriori, utilizando cadenas de Markov y Monte Carlo (MCMC), Gelfand and Smith [42]. Estos valores simulados de las distribuciones a posteriori se utilizan para estimar la distribución marginal de interés, tal como distribuciones a posteriori de los parámetros particulares.

Las distribuciones condicionales de los parámetros del nivel II Θ , Λ , σ , θ^2 , σ , σ^2 , γ , y λ tienen una distribución normal estándar. Por ejemplo, la distribución condicional de Λ dado λ , z , σ , σ^2 , α , μ_α es proporcional a una distribución normal con media $\sum_{i=1}^m \frac{\alpha_i - \lambda z_i}{m}$ y varianza $\frac{\sigma_\alpha^2}{m}$ en el rango $[\mu_{\alpha_1}, \mu_{\alpha_2}]$. Las varianzas de los parámetros son previas conjugadas, es decir, una previa que al combinarla (conjugarla) con la verosimilitud, la posteriori resultante pertenece a la misma familia de distribuciones que la previa, ecuación 3.47. De manera similar se obtienen las distribuciones condicionales de Θ y σ_θ^2 . La distribución condicional de $\lambda | \Lambda, \sigma_\alpha^2, \alpha, \mu_\lambda$ es proporcional a una distribución normal con media $\frac{\sum_{i=1}^m z_i (\alpha_i - \Lambda)}{\sum_{i=1}^m z_i^2}$ y varianza $\frac{\sigma_\alpha^2}{\sum_{i=1}^m z_i^2}$ en el rango $[\mu_{\lambda_1}, \mu_{\lambda_2}]$. La distribución condicional del nivel I con parámetros $\theta_i, \alpha_i, \beta$ no está distribuida de manera estándar. La distribución condicional de la precisión (exactitud) de los estudios específicos, $\alpha_i | y_i, n_i, z_i, \Lambda, \sigma_\alpha^2, \theta_i, \beta$, se obtiene multiplicando una binomial con una normal, Rutter and Gatsonis [114], note expresión 3.48. La distribución condicional de θ_i , tiene la misma forma que la anterior

⁶Una distribución previa (a priori) tal que al combinarla (conjugarla) con la verosimilitud, la posterior resultante pertenece a la misma familia de distribuciones que la previa, entonces se dice que es una previa conjugada.

expresión. La distribución condicional de los parámetros de escala $\beta|y, n, Z, \theta, \alpha, \mu_\beta$, es proporcional al producto de $2 \times m$ binomiales, Rutter and Gatsonis [114], note expresión 3.49, con probabilidad en el intervalo $[\mu_{\beta_1}, \mu_{\beta_2}]$.

3.7.5 ELECCIÓN A PRIORI

Los rangos para Θ, Λ, β deben ser seleccionados de tal manera que reflejen conocimiento de las modalidades de diagnósticos de las revisiones que se están analizando. En general, el intervalo $[-10, 10]$ cubre todos los valores posibles de Θ , Rutter and Gatsonis [114]. De manera similar, el intervalo $[-5, 5]$ cubre todos los valores posibles para β . Pero, debido a que se espera que los resultados de las pruebas sean positivas en pacientes que padezcan la enfermedad, el intervalo $[-2, 20]$ cubrirá todos los valores posibles de Λ , Rutter and Gatsonis [114]. La selección a priori de la distribución gamma inversa Γ^{-1} para los parámetros de la varianza entre estudios, σ_θ^2 y σ_α^2 es más complicada. La distribución muestral es potencialmente sensible a la distribución a priori, la cual es usada para modelizar los parámetros de la varianza. La gamma inversa a priori que explica los parámetros de la varianza debería ser seleccionada de tal forma que refleje un rango apropiado de los parámetros de precisión y los puntos de corte específicos del estudio. El objetivo de una previa (priori) adecuada para los parámetros de la varianza, es seleccionar una distribución que no asigne valores de probabilidades no muy grandes (pocos realistas). Afortunadamente la elección a priori de γ , es realista ya que se fundamenta en los rangos de los puntos de corte y de los parámetros de precisión, que se han identificado anteriormente.

3.7.6 PARÁMETROS DE ESTIMACIÓN

El objetivo de la estimación es la descripción de la distribución a posteriori de los parámetros del modelo y los resúmenes estadísticos que son funciones de los parámetros del modelo. Los intervalos de confianza al 95% son estimados empíricamente por los percentiles a posteriori del 2, 5% y 97, 5% de la simulación. La simetría de la distribución a posteriori es aproximada por las medias de los valores simulados de (Θ, Λ, β) . La asimetría de la distribución a posteriori se aproxima por la mediana de los valores simulados de $(\sigma_\theta^2, \sigma_\alpha^2)$, Rutter and Gatsonis [114].

$$\sigma_\alpha^2 | \lambda, z, \alpha, \Lambda, \xi_\alpha \sim \Gamma_{-1} \left(\left(\xi_{\alpha_1} + \frac{m}{2} \right); \left(\frac{1}{2} \sum_{i=1}^m (\alpha_1 - \Lambda - \lambda z_i)^2 + \frac{1}{\xi_{\alpha_2}} \right)^{-1} \right) \quad (3.47)$$

$$\exp \left[-\frac{(\alpha_i - (\Lambda + \lambda z_i))^2}{2\sigma_\alpha^2} \right] \prod_{j=1}^2 \binom{n_{ij}}{y_{ij}} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{(n_{ij} - y_{ij})} \quad (3.48)$$

$$\prod_{i=1}^2 \prod_{j=1}^2 \binom{n_{ij}}{y_{ij}} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{(n_{ij} - y_{ij})} \quad (3.49)$$

3.7.7 DIAGNÓSTICO

En este punto se analizan las estadísticas de diagnóstico para evaluar una mala especificación del modelo, un sobre ajuste del modelo y en general para identificar datos atípicos. Se pueden seguir para este

punto las sugerencias de Weiss [147]. Debido a que el número de verdaderos positivos y falsos positivos dentro de los estudios siguen una distribución binomial, se puede comprobar si la mala especificación del modelo se debe a la evaluación de la distribución a priori de cada estudio específico de los parámetros θ y α .

Nótese que ambos parámetros se distribuyen de una manera normal. Si se asume que, $\gamma = \lambda = 0$, las sumas de cuadrados $S_\theta = \frac{\sum_i (\theta_i - \Theta)^2}{\sigma_\theta^2}$ y $S_\alpha = \frac{\sum_i (\alpha_i - \Lambda)^2}{\sigma_\alpha^2}$, siguen una distribución χ^2 con m grados de libertad, donde m es el número de estudios en el meta-análisis. Si se asume que $\gamma \neq 0$ y $\lambda \neq 0$ en el modelo, la suma de cuadrados tendrá la forma $S_\alpha = \frac{\sum_i (\alpha_i - \Lambda - \lambda Z_i)^2}{\sigma_\alpha^2}$; si las probabilidades en las colas son muy grandes o muy pequeñas, Rutter and Gatsonis [114] sugieren que hay una mala especificación de la distribución a priori. Se tiene que evaluar el supuesto de independencia condicional entre θ y α , estimando estos parámetros directamente de los datos y realizando una inspección mediante un gráfico de dispersión y correlación.

Estas estimaciones se basan en la fórmula 3.43 con $\beta = 0$ y son representadas en las ecuaciones, 3.50, 3.51, donde "cc" denota corrección de continuidad o de Yates⁷. Para la evaluación global de la bondad de ajuste del modelo se suelen usar dos estadísticos de discrepancia chi-cuadrado. El primero, se basa en las estimaciones de los recuentos (counts), note ecuación 3.52, donde y_{ij1} , denota el número de sujeto i que dió positivo en la prueba pero que no tiene la enfermedad ($j = 0$) o que padece la enfermedad ($j = 1$) en el grupo. D_{counts}^2 , se compara con una distribución χ_{2m}^2 . El segundo estadístico de la bondad de ajuste global es el que se fundamenta en la corrección de continuidad de los logaritmos de los odds ratios, y este se estima como en la ecuación 3.53. Donde $\log(OR_{cc})_i$, es la corrección de continuidad de los logaritmos de los odds ratio del i -ésimo estudio.

3.7.8 RESUMEN ESTADÍSTICO

La estimación MCMC, permite estimar los parámetros de las funciones antes descritas, por lo que los resultados pueden ser presentados utilizando una variedad de medidas de resumen. En los resúmenes aparecen medidas como las tasas de verdaderos positivos, falsos positivos y los ratios de máxima verosimilitud, Boyko [12]. En el contexto del meta-análisis, estas medidas son estimaciones a posteriori de las combinaciones de las informaciones de los estudios. Las tasas de VP y FP se estiman como aparecen en las ecuaciones 3.54 y 3.55; donde los ratios de máxima verosimilitud positivos y negativos respectivamente, se estiman a partir de las ecuaciones 3.56 y 3.57. El enfoque jerárquico del resumen ROC (HSROC), es un modelo que combina el par de estimación de la sensibilidad y especificidad de múltiples estudios y que es un enfoque que incorpora la variabilidad dentro y entre los estudios en comparación con el modelo de efectos fijos del resumen ROC (SROC). Esto permite flexibilidad en las estimaciones de los resúmenes estadísticos. El HSROC, describe la variabilidad dentro de los estudios usando una distribución binomial del número de pruebas positivas con pacientes que tienen la enfermedad y no. La variación en los criterios de positividad y la precisión es modelizada usando una distribución normal, con una regresión lineal en la media que permite la dependencia de la covarianza de los niveles en los estudios. Distribuciones con colas más pesadas (como la t o de cauchy) pueden ser utilizadas en el segundo nivel del modelo jerárquico en vez de una normal. El modelo puede ser extendido para permitir la dependencia entre el criterio de

⁷Si la tabla de contingencia (tabla de 2x2), cuenta con un grado de libertad, el valor estadístico esta sobre-estimado. En tal caso, se aplica una corrección en la fórmula del cálculo de χ^2 , que se denomina corrección de Yates. Esta corrección consiste en restar 0.5 cuando la frecuencia observada es mayor que la esperada, y sumar 0.5 en caso contrario.

positividad y la precisión. El modelo HSROC, representado por las ecuaciones 3.50 y 3.51, provee una cuantificación más realista de la variabilidad dentro de los estudios que el enfoque de efectos fijos.

$$\theta_i^{(0)} = \frac{\text{logit}(VP_i^{cc}) - \text{logit}(FP_i^{cc})}{2} \quad (3.50)$$

$$\alpha_i^{(0)} = \text{logit}(VP_i^{cc}) - \text{logit}(FP_i^{cc}) \quad (3.51)$$

$$D_{counts}^2 = \sum_i \sum_j \frac{y_{ij1} - E(y_{ij1} | model, data)^2}{E(y_{ij1} | model, data)} \quad (3.52)$$

$$D_{\lg(OR)}^2 = \sum_i \frac{\lg(OR_{cc})_i - E(\lg(OR_{cc})_i | model, data)^2}{\sqrt{\text{Var}(\lg(OR_{cc})_i | model, data)}} \quad (3.53)$$

$$VP = \text{logit}^{-1} \left(\left(\Theta + \frac{\Lambda}{2} \right) \exp^{-\frac{\beta}{2}} \right) \quad (3.54)$$

$$FP = \text{logit}^{-1} \left(\left(\Theta - \frac{\Lambda}{2} \right) \exp^{-\frac{\beta}{2}} \right) \quad (3.55)$$

$$LR+ = \frac{VP}{FP} \quad (3.56)$$

$$LR- = \frac{1 - VP}{1 - FP} \quad (3.57)$$

3.8 META-ANÁLISIS BAYESIANO PARA LA PRECISIÓN DE UN TEST EN AUSENCIA DE LA PRUEBA ORO ESTÁNDAR

Se puede decir que el modelo de Moses et al. [93] describe un meta-análisis como odds ratio de diagnóstico (DOR). El enfoque HSROC, Rutter and Gatsonis [114] expresa la sensibilidad y la probabilidad de falsos positivos de cada estudio como una función de un modelo normal bivalente subyacente, mientras que el modelo bivalente de Reitsma et al. [105] supone que el vector (logit(sensibilidad), logit(especificidad)) sigue una distribución normal bivalente. En ausencia de covarianza, las funciones de verosimilitud del HSROC y del modelo bivalente son algebraicamente equivalentes, proporcionando "pooled" idénticos de sensibilidad y especificidad y estimaciones de varianza entre estudios Harbord et al. [54]. El modelo HSROC conduce de forma más natural a una curva SROC, mientras que el modelo bivalente genera un "pooled" de sensibilidad y especificidad.

Walter et al. [145], propone una clase de modelo latente, donde asume que el test bajo evaluación y el test de referencia son considerados variables latentes (no observables), del estado verdadero de la enfermedad. Estos autores demostraron que el modelo es identificable al asumir que la sensibilidad y la especificidad siguen siendo la misma, pero la prevalencia varía entre los estudios. En la práctica, la asunción de que la sensibilidad y la especificidad sean idénticas en todos los estudios es difícil de justificar, dada la variabilidad de la población y los aspectos individuales de los diseños de estudios. Chu et al. [17], describen un modelo más general donde la sensibilidad y especificidad de la prueba, tanto en la fase de evaluación y la prueba de referencia, así como la prevalencia, se tratan como efectos aleatorios. Este modelo puede ser conceptualizado como una extensión del modelo de Reitsma et al. [105] para el caso donde no exista el patrón o prueba oro. Sadatsafavi et al. [118], describen un modelo donde un parámetro (i.e., sensibilidad o especificidad de uno de los test) varía en los estudios. Ninguno de los modelos analizados anteriormente se ajusta cuando los estudios utilizan patrones de referencias diferentes en los estudios seleccionados.

Este método es una extensión del HSROC. Asume que la variación de la sensibilidad y especificidad en los estudios aparece debido al empleo de diferentes puntos de corte para definir una prueba positiva y/o diferencias en la precisión del diagnóstico, cuando no existe una prueba de referencia.

Sean J estudios de diagnóstico que son incluidos en un meta-análisis, y que cada estudio provee una tabla de doble entrada, que relaciona el test bajo evaluación T_1 y el test de referencia T_2 . Las dos medidas se asume que son variables dicotómicas, tomando el valor de 1 cuando el test es positivo y 0 cuando es negativo. Ambas variables se supone que son medidas imperfectas de una variable latente subyacente dicotómica D , para el verdadero estado de la enfermedad. Sea t_{1j} y t_{2j} , vectores de resultados del estudio j para T_1 y T_2 respectivamente. Donde la sensibilidad de la prueba de referencia es definida como $S_2 = P(T_2 = 1 | D = 1)$ y su especificidad se define por $C_2 = P(T_2 = 0 | D = 0)$.

Rutter and Gatsonis [114], notan que los resultados de T_1 se basan en una variable latente subyacente. Se asume que la variable latente Z_1 sigue una distribución normal y que un resultado positivo en T_1 corresponde a valores altos de Z_1 en comparación con un resultado negativo. El modelo asume que entre pacientes con $D = 0$, $Z_1 \sim N\left(-\frac{\alpha_j}{2}, \exp\left(-\frac{\beta}{2}\right)\right)$. Este modelo puede ser conceptualizado como un modelo de regresión binomial con un enlace probit (probit link). Dentro del j -ésimo estudio,

la diferencia en medias de estas dos distribuciones es α_j , y el ratio de su desviación estándar es $\exp(\beta)$. Cada estudio asume un punto de corte diferente, θ_j , que es un valor positivo. Se define una distribución a priori jerárquica sobre la diferencia de medias (o precisión diagnóstica), $\alpha_j \sim N(\Lambda, \sigma_\alpha^2)$, permitiendo una variación en la distribución Z_1 en cada estudio. Del mismo modo, una jerárquica a priori $\theta_j \sim N(\Theta, \sigma_\theta^2)$, que permite una variación de los puntos de corte en los estudios. Esta estructura es equivalente a un modelo jerárquico de dos niveles, un nivel dentro del estudio para los parámetros θ_j y α_j y un nivel para estudiar la variabilidad entre los estudios con parámetros Λ , Θ , y β , los cuales son comunes en todos los estudios. En base a las asunciones anteriores, la sensibilidad T_1 en el j -ésimo estudio es dada por, $S_{1j} = \Phi\left(\frac{\theta_j^{-\alpha_j/2}}{\exp(\beta/2)}\right)$, mientras que la especificidad está dada por, $C_{1j} = \Phi\left(\frac{\theta_j^{\alpha_j/2}}{\exp(-\beta/2)}\right)$. Por lo tanto, valores crecientes de θ_j , inducen a una correlación negativa entre la sensibilidad y la especificidad de T_1 a través de los estudios, mientras que aumento de los valores α_j , provoca correlación positiva entre ellos.

Un índice global para la prueba de la sensibilidad y especificidad puede resumirse como, $\Phi\left(-\frac{\Theta^{-\Lambda/2}}{\exp(\beta/2)}\right)$ y $\Phi\left(\frac{\Theta^{-\Lambda/2}}{\exp(\beta/2)}\right)$ respectivamente. La utilidad de un solo "pooled" de estimación dependerá del grado de heterogeneidad entre los estudios, Macaskill et al. [84].

3.8.1 ESTIMACIÓN

Dendukuri et al. [27], proponen la siguiente estimación: La función de verosimilitud de los datos observados de los J estudios puede ser expresada en términos de la sensibilidad y especificidad de cada test, y la prevalencia en el j -ésimo estudio, $P((D = 1 | \text{Estudio} = j) = \pi_j)$, puede ser expresada como en la ecuación 3.58. Para llevar a cabo estimación bayesiana, se tienen que especificar distribuciones a priori (previas) sobre el conjunto de parámetros desconocidos. La estrategia es emplear distribuciones previas para la mayoría de los parámetros, utilizando valores a priori de los parámetros. Las previas (priori) para Λ , Θ , y β son seleccionadas de tal manera que la distribución marginal resultante del "pooled" de la sensibilidad y especificidad sea aproximadamente normal en $(0, 1)$. El "pooled" de diferencias de medias para los parámetros es asumido con una densidad previa (priori) $\Lambda \sim U(-3, 3)$. El logaritmo del ratio entre las dos desviaciones estándar, β , se asume que sigue una distribución uniforme, es decir $U(-0.75, 0.75)$. El "pooled" del punto de corte de los parámetros, Θ , se asume que sigue una distribución uniforme, $U(-1.5, 1.5)$. Los parámetros σ_β y σ_θ se distribuyen como uniforme, $U(0, 2)$. Para los parámetros π_j , S_2 y C_2 se emplea una distribución a priori (previa) beta, si se trata de una distribución objetivo a priori se emplea una Beta $(1, 1)$, Dendukuri et al. [27].

El número total de grados de libertad disponibles es $3J$, con cada estudio contribuyendo 3 grados de libertad. El número total de parámetros a estimar es al menos $J + 7$. Por lo tanto, un mínimo de 4 estudios se requeriría para estimar razonablemente este modelo sin ningún tipo de distribuciones previas informativas. Dado que dos parámetros, S_{1j} y C_{1j} , son definidos como una función que depende de tres parámetros, α_j, θ_j y β , se asume que β es la misma en todos los estudios, para evitar problemas de identificabilidad.

Se puede emplear un paquete de R, HSROC, "An R package for Bayesian meta-analysis of diagnostic

test accuracy” para la estimación de los parámetros antes mencionados, algoritmo Schiller et al. [121]. Para evaluar la convergencia de los modelos se utilizan cinco cadenas diferentes de 50.000 iteraciones a partir de los valores iniciales y se calcula el estadístico de Gelman-Rubin para comparar la variabilidad dentro y entre las cadenas Gelman and Rubin [43].

3.8.2 LA DEPENDENCIA CONDICIONAL ENTRE EL ÍNDICE Y ANÁLISIS DE REFERENCIA

En ausencia de una referencia estándar de oro es importante ajustar la dependencia condicional entre múltiples pruebas realizadas sobre los mismos temas, Dendukuri et al. [27], con el fin de ajustar la correlación no explicada entre los ensayos dentro de cada clase latente, Dendukuri and Joseph [26]. El modelo en 3.58 se puede extender al ajustar la dependencia condicional de distintas maneras, Dendukuri et al. [25]. Una de las maneras de ajustar la dependencia condicional es sumando la covarianza entre la sensibilidad del índice y el test de referencia $cov(S_j)$ en el j -ésimo estudio y entre su especificidad $cov(C_j)$ en el j -ésimo estudio como lo proponen Chu et al. [17], Dendukuri et al. [25], Dendukuri and Joseph [26], teniendo que la probabilidad conjunta entre estos dos test en el j -ésimo estudio se puede representar como se expresa en la ecuación 3.59. Luego se establece una distribución a priori uniforme para la covarianza de los parámetros; $cov(s_j) \sim U(0, \min(S_{1j}, S_2) - S_{1j}S_2)$, $cov(c_j) \sim U(0, \min(C_{1j}, C_2) - C_{1j}C_2)$.

3.9 UNIFICACIÓN DE MODELOS PARA META-ANÁLISIS DE LA PRECISIÓN DIAGNÓSTICA

Existe un creciente interés en las revisiones sistemáticas y meta-análisis de los datos de los estudios de la precisión de diagnóstico, Bossuyt et al. [11], Deeks [22], Devillé et al. [29], Gluud and Gluud [47], Khan et al. [67], Tatsioni et al. [136], Whiting et al. [151]. Los datos de cada estudio primario se resumen en una tabla de 2x2, sobre la base de resultados de una prueba dicotómica contra el verdadero estado de la enfermedad, de la cual se extraen medidas conocidas como, la sensibilidad y la especificidad.

Los métodos estadísticos antes estudiados y propuestos por Dukic and Gatsonis [33], Moses et al. [93], Reitsma et al. [105], Rutter and Gatsonis [114], Siadaty and Shu [129], reflejan dos principales características, primero una correlación negativa entre la sensibilidad y la especificidad es explicada por los diferentes tipos de umbrales en las pruebas, Deeks [22], Moses et al. [93]. Segundo se espera una sustancial heterogeneidad entre los estudios, la cual debe ser incorporada en los modelos, Lijmer et al. [77].

El enfoque inferencial de estos métodos son sujeto de ciertos debates. Algunos autores, proponen estimaciones de las medidas de sensibilidad y especificidad o regiones de predicción, dentro de las cuales pueden contener estudios con poca credibilidad, Reitsma et al. [105], mientras que otros sugieren que en presencia de heterogeneidad, los resultados de los meta-análisis deben presentarse como una curva (SROC), Rutter and Gatsonis [114]. Algunos autores han propuesto un método de generación de una curva SROC, utilizando una regresión lineal simple, que es el enfoque más utilizado con frecuencia, Littenberg and Moses [80]. Asimismo, los supuestos de la regresión lineal no se cumplen, por lo que el método es una aproximación. También existe incertidumbre al momento de elegir los pesos de la regresión, Rutter and Gatsonis [114], Walter [144]. Dos métodos estadísticos se han propuesto, para superar estos problemas descritos anteriormente, Reitsma et al. [105], Rutter and Gatsonis [114], el enfoque bivalente y el HSROC respectivamente.

3.10 RELACIÓN ENTRE EL MODELO BIVARIANTE Y HSROC

El modelo bivalente permite la inclusión de covarianzas que afectan a la sensibilidad y/o especificidad, mientras que el enfoque HSROC permite introducir covariaciones a los parámetros de precisión y/o los umbrales de las pruebas. Un modelo HSROC que introduce covariaciones a la precisión y a los umbrales de las pruebas es equivalente a un enfoque bivalente si en el modelo se introduce las mismas covariaciones del modelo HSROC pero a nivel de sensibilidad y especificidad Harbord et al. [54]. Una interpretación más formal se detalla a continuación:

Sea $b = \exp\left(\frac{\beta}{2}\right)$, entonces el modelo HSROC, ecuación 3.43, nivel I queda expresado en las ecuaciones 3.60 y 3.61. Si en el modelo bivalente, ecuación 3.35, reemplazamos θ_{Ai} y θ_{Bi} por μ_{Ai} y μ_{Bi} , las cuales representan las transformaciones logit de la sensibilidad y especificidad, se obtienen las ecuaciones 3.62 y 3.63 respectivamente. Luego, se pueden relacionar las variables aleatorias que forman la base de los dos modelos, de la manera que se aprecia en las expresiones 3.64 y 3.65. Este par de ecuaciones, describen como μ_{Ai} y μ_{Bi} son combinaciones lineales de dos variables aleatorias, θ_i y α_i , que en el modelo HSROC son variables normales independientes. Luego se puede inferir que cualquier par de combinaciones lineales de variables aleatorias con distribuciones normales, tiene una distribución normal bivalente, Dudewicz and

Mishra [32]. Se sabe que la distribución conjunta de μ_{Ai} y μ_{Bi} es normal bivalente, en el modelo HSROC.

Luego el modelo HSROC es equivalente a un modelo bivalente. Las ecuaciones 3.64 y 3.65 se pueden expresar en forma matricial, 3.66. Invirtiendo el resultado anterior se tiene la ecuación 3.67. Donde S, es la matriz de transformación asociada, con los cambios desde las coordenadas del modelo bivalente (transformaciones logit de la sensibilidad y especificidad) hasta las coordenadas del modelo HSROC (puntos de corte y parámetros de precisión). Nótese que la matriz S, no es ortogonal $S^{-1} \neq S^T$.

$$\begin{aligned}
L(\Theta, \Lambda, S_2, C_2, \sigma_\alpha^2, \sigma_\theta^2, \beta, \pi_j, \alpha_j, j = 1, 2, \dots, J | t_{1j}, t_{2j}, j = 1, 2, \dots, J) = \\
\prod_{j=1}^J \left[\pi_j \Phi \left\{ -\frac{\theta_j^{-\alpha_j/2}}{\exp(\beta/2)} \right\} S_2 + (1 - \pi_j) \Phi \left\{ -\frac{\theta_j^{\alpha_j/2}}{\exp(-\beta/2)} \right\} (1 - C_2) \right]^{t_{ij} \cdot t_{2j}} \times \\
\times \left[\pi_j \Phi \left\{ -\frac{\theta_j^{-\alpha_j/2}}{\exp(\beta/2)} \right\} (1 - S_2) + (1 - \pi_j) \Phi \left\{ -\frac{\theta_j^{\alpha_j/2}}{\exp(-\beta/2)} \right\} C_2 \right]^{t_{ij} \cdot (1-t_{2j})} \times \\
\times \left[\pi_j \Phi \left\{ \frac{\theta_j^{-\alpha_j/2}}{\exp(\beta/2)} \right\} S_2 + (1 - \pi_j) \Phi \left\{ \frac{\theta_j^{\alpha_j/2}}{\exp(-\beta/2)} (1 - C_2) \right\} \right]^{(1-t_{ij}) \cdot t_{2j}} \times \\
\times \left[\pi_j \Phi \left\{ \frac{\theta_j^{-\alpha_j/2}}{\exp(\beta/2)} \right\} (1 - S_2) + (1 - \pi_j) \Phi \left\{ \frac{\theta_j^{\alpha_j/2}}{\exp(-\beta/2)} C_2 \right\} \right]^{(1-t_{ij}) \cdot (1-t_{2j})} \quad (3.58)
\end{aligned}$$

$$\begin{aligned}
P(T_1 = u, T_2 = v | Estudio = j) = \pi_j \left[S_{ij}^u (1 - S_{ij})^{(1-u)} S_2^v (1 - S_2)^{(1-v)} (-1)^{|u-v|} cov(S_j) \right] + \\
= +(1 - \pi_j) \left[C_{ij}^{(1-u)} (1 - C_{ij})^u C_2^{(1-v)} (1 - C_2)^v + (-1)^{|u-v|} cov(C_j) \right] \quad (3.59)
\end{aligned}$$

$$logit(\pi_{i1}) = b^{-1} \left(\theta_i + \frac{1}{2} \alpha_i \right) \quad (3.60)$$

$$logit(\pi_{i0}) = b \left(\theta_i - \frac{1}{2} \alpha_i \right) \quad (3.61)$$

$$\mu_{Ai} = logit(\pi_{i1}) \quad (3.62)$$

$$\mu_{Bi} = logit(1 - \pi_{i0}) = \ln \left(\frac{1 - \pi_{i0}}{\pi_{i0}} \right) = -\ln \left(\frac{\pi_{i0}}{1 - \pi_{i0}} \right) = -logit(\pi_{i0}) \quad (3.63)$$

$$\mu_{Ai} = b^{-1} \left(\theta_i + \frac{1}{2} \alpha_i \right) \quad (3.64)$$

$$\mu_{Bi} = -b \left(\theta_i - \frac{1}{2} \alpha_i \right) \quad (3.65)$$

$$\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} = S^{-1} \begin{pmatrix} \theta_i \\ \alpha_i \end{pmatrix} \text{ donde } S^{-1} = \begin{pmatrix} b^{-1} & \frac{1}{2}b^{-1} \\ -b & \frac{1}{2}b \end{pmatrix} \quad (3.66)$$

$$\begin{pmatrix} \theta_{Ai} \\ \alpha_{Bi} \end{pmatrix} = S \begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} \text{ donde } S = \begin{pmatrix} \frac{1}{2}b & -\frac{1}{2}b \\ b & b^{-1} \end{pmatrix} \quad (3.67)$$

$$E \left(\begin{pmatrix} \theta_i \\ \alpha_i \end{pmatrix} \right) = E \left(S \begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} \right) = \begin{pmatrix} \Theta \\ \Lambda \end{pmatrix} = S \begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix} \quad (3.68)$$

$$\text{var} \left(\begin{pmatrix} \theta_i \\ \alpha_i \end{pmatrix} \right) = \text{var} \left(\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} \right) = \begin{pmatrix} \sigma_\theta^2 & 0 \\ 0 & \sigma_\alpha^2 \end{pmatrix} = S \cdot \text{cov} \left(\begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix} \right) S^T = S \Sigma_{AB} S^T \quad (3.69)$$

$$\begin{pmatrix} \sigma_\theta^2 & 0 \\ 0 & \sigma_\alpha^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{4}b^2\sigma_A^2 - \frac{1}{2}\sigma_{AB} + \frac{1}{4}b^{-2}\sigma_B^2 & \frac{1}{2}b^2\sigma_A^2 - \frac{1}{2}b^{-2}\sigma_B^2 \\ \frac{1}{2}b^2\sigma_A^2 - \frac{1}{2}b^{-2}\sigma_B^2 & b^2\sigma_A^2 + 2\sigma_{AB} + b^{-2}\sigma_B^2 \end{pmatrix} \quad (3.70)$$

$$\begin{pmatrix} \Theta \\ \Lambda \end{pmatrix} = \begin{pmatrix} \frac{1}{2}b\mu_A - \frac{1}{2}b^{-1}\mu_B \\ b\mu_A + b^{-1}\mu_B \end{pmatrix} \quad (3.71)$$

$$\Theta = \frac{1}{2} \left[\sqrt{\frac{\sigma_B}{\sigma_A}} \mu_A - \sqrt{\frac{\sigma_A}{\sigma_B}} \mu_B \right] \quad (3.72)$$

$$\Lambda = \sqrt{\frac{\sigma_B}{\sigma_A}} \mu_A + \left[\sqrt{\frac{\sigma_B}{\sigma_A}} \right]^{-1} \mu_B \quad (3.73)$$

3.11 RELACIÓN DE LOS PARÁMETROS DE LOS MODELOS SIN COVARIANZA

Para expresar la relación entre los parámetros de los dos modelos, sin covarianzas, en términos de la matriz de transformación S , tomando la esperanza y la varianza de la ecuación 3.67, se tienen las expresiones, 3.68 y 3.69. La asunción del modelo HSROC de que θ_i y α_i no están correlacionadas, es decir los elementos fuera de la diagonal principal son ceros, fija el valor de b y por lo tanto la matriz de transformación S , donde S es una matriz no ortogonal que diagonaliza la matriz de varianza y covarianza del modelo bivalente. Si se desarrollan las operaciones del lado derecho de la ecuación 3.69, se tiene la ecuación 3.70.

Donde, los elementos fuera de la diagonal principal de 3.70, son iguales a cero si $b = \sqrt{\frac{\sigma_B}{\sigma_A}}$ o su equivalente $\beta = \log\left(\frac{\sigma_B}{\sigma_A}\right)$. Luego el parámetro de forma del modelo HSROC β , está determinado únicamente por la variación de los logit de la sensibilidad y especificidad del modelo bivalente. Se puede inferir de las ecuaciones 3.68 y 3.69, los otros parámetros del HSROC con el modelo bivalente, 3.71. Reemplazando $b = \sqrt{\frac{\sigma_B}{\sigma_A}}$ en la ecuación 3.71, se obtienen las relaciones 3.72 y 3.73. Reemplazando $b = \sqrt{\frac{\sigma_B}{\sigma_A}}$ en 3.70, se obtiene las ecuaciones 3.74 y 3.75.

$$\sigma_\theta^2 = \frac{1}{2}(\sigma_A\sigma_B - \sigma_{AB}) \quad (3.74)$$

$$\sigma_\alpha^2 = 2(\sigma_A\sigma_B + \sigma_{AB}) \quad (3.75)$$

$$E\begin{pmatrix} \theta_i \\ \alpha_i \end{pmatrix} = S\begin{pmatrix} \mu_A + v_A Z_i \\ \mu_B + v_B Z_i \end{pmatrix} \quad (3.76)$$

$$E\begin{pmatrix} \theta_i \\ \alpha_i \end{pmatrix} = S\begin{pmatrix} \Theta + \gamma Z_i \\ \Lambda + \lambda Z_i \end{pmatrix} \quad (3.77)$$

$$\text{logit}(\text{sensibilidad}) = \Lambda \exp^{\beta/2} - \exp^\beta \text{logit}(\text{especificidad}) \quad (3.78)$$

3.12 RELACIÓN DE LOS PARÁMETROS DE LOS MODELOS CON COVARIANZA

Si el modelo bivalente se extiende, incluyendo una covarianza que afecta a la sensibilidad y especificidad, las ecuaciones 3.67 y 3.68 pueden ser escritas como 3.76 y 3.77. Por lo tanto, un modelo bivalente, que se le agregue una o más covarianzas que afecten la sensibilidad y la especificidad es equivalente a un modelo HSROC, donde las mismas covarianzas afectan tanto a los parámetros de precisión como a los puntos de corte. Sin embargo, si en el modelo bivalente se incluyen covarianzas que afecten a la sensibilidad o especificidad, no será equivalente a un modelo HSROC, a menos que las restricciones se impongan sobre la relación entre los coeficientes del modelo HSROC.

3.13 ENFOQUE DE INFERENCIA Y SALIDAS DE LOS MODELOS

3.13.1 MODELO HSROC

El modelo HSROC, genera una curva SROC, al permitir que el parámetro θ varíe, mientras que el parámetro de precisión α se mantiene fijo en su media. Para el caso donde no exista covarianza, la sensibilidad esperada para una especificidad dada viene dada por 3.78, Macaskill [83], Rutter and Gatsonis [114]. Rutter and Gatsonis [114], sugieren que la curva se restrinja dentro del rango de las especificidades estimadas de los estudios. Si $\beta = 0$, la curva es simétrica (sensibilidad=especificidad). Esta curva ROC, no representa la incertidumbre de los parámetros estimados, pero representa la variabilidad del umbral.

3.13.2 MODELO BIVARIANTE

Reitsma et al. [105], sugieren regiones de confianza y de predicción para la curva ROC, que pueden ser construidas utilizando las estimaciones de un modelo bivalente. A medida que la correlación entre la sensibilidad y la especificidad aumentan, los intervalos de confianza para la media de los logit de la sensibilidad y especificidad (μ_A, μ_B) se separan, lo cual puede ser engañoso; por lo cual es preferible utilizar una región de confianza elíptica para ambos parámetros. Tal región se genera utilizando una representación paramétrica, 3.79, 3.80, Douglas [31].

$$\mu_A = \hat{\mu}_A + S_A C \cos(t) \quad (3.79)$$

$$\mu_B = \hat{\mu}_B + S_B C \cos(t + \arccos(r)) \quad (3.80)$$

Donde S_A, S_B , son las estimaciones de los errores estándar de μ_A y μ_B y r es la estimación de su correlación y variando t , $0 \leq t \leq 2\pi$ se generan los bordes de las elipses. La constante "C" se denomina, constante de borde de la elipse, Alexandersson et al. [1]. Asintóticamente provee una región de confianza de $100(1 - \alpha\%)$, con $C = \sqrt{\chi_{2,\alpha}^2}$. Cuando el número de estudios es pequeño, es recomendable utilizar una medida más conservadora para la región de confianza dada por $C = \sqrt{2F_{2,(n-2);\alpha}}$, donde n es el número de estudios y $F_{2,(n-2);\alpha}$ sigue una distribución de Fisher, con 2 y $(n-2)$ grados de libertad, Chew [16], Dudewicz and Mishra [32]. También es posible construir una región de predicción que incluya los verdaderos valores de la sensibilidad y especificidad de estudios futuros. La matriz de covarianzas de los verdaderos valores de los logit de la sensibilidad y especificidad de estudios futuros quedan expresados como $\Sigma_{AB} = \text{var}\begin{pmatrix} \hat{\mu}_A \\ \hat{\mu}_B \end{pmatrix}$.

3.14 MODELO ESTADÍSTICO PARA I^2

Cuando se estima el tamaño del efecto de cada estudio individual, el meta-análisis los integra calculando un tamaño del efecto promedio, ajustando la heterogeneidad alrededor de la media estimada. En general, el modelo estadístico más realista es el que integra la estimación del efecto del meta-análisis en un modelo de efectos aleatorios, porque incorpora las fuentes de heterogeneidad entre los estudios. La cuantificación del grado de heterogeneidad entre una colección de estudios es uno de los aspectos más problemáticos de un meta-análisis. Es importante porque puede afectar a la decisión sobre la elección del modelo estadístico fijo o de efectos aleatorios. Por otro lado, si se encuentra una variabilidad significativa, se pueden detectar las variables potenciales responsables de esta variabilidad. La varianza entre los estudios, τ^2 , es el parámetro en el modelo estadístico que representa principalmente a la verdadera heterogeneidad entre los verdaderos efectos de los estudios. Por lo tanto, un buen procedimiento para determinar si existe una verdadera heterogeneidad entre una colección de estudios debe ser relacionado positivamente con τ^2 . Al mismo tiempo, no debe ser afectado por el número de estudios, y debe ser libre de escala con el fin de ser comparable entre los meta-análisis que han aplicado diferentes índices.

Sea d el mejor estimador del parámetro del tamaño del efecto δ , ecuación 3.81, siendo \bar{Y}_E y \bar{Y}_C las medias muestrales del grupo experimental y control respectivamente y S un estimador de la desviación estándar intra grupo, Hedges and Olkin [57], estimada por la ecuación 3.82, Hedges and Olkin [57], donde S_E^2 , S_C^2 , n_E y n_C representan las varianzas muestrales y los tamaños de la muestra de los grupos experimental y control respectivamente. El término $c(m)$ es un factor de corrección para el sesgo positivo sugerido para estandarizar la diferencia de medias cuando el tamaño de la muestra es pequeño y es estimado por la ecuación 3.83, Hedges and Olkin [57], donde $m = n_E + n_C - 2$. La varianza muestral del índice d es estimada por la ecuación 3.84, Hedges and Olkin [57].

$$d = c(m) \frac{\bar{Y}_E - \bar{Y}_C}{S} \quad (3.81)$$

$$S = \sqrt{\frac{(n_E - 1) S_E^2 + (n_C - 1) S_C^2}{n_E + n_C - 2}} \quad (3.82)$$

$$c(m) = 1 - \frac{3}{4 \cdot m - 1} \quad (3.83)$$

$$S_d^2 = \frac{n_E + n_C}{n_E \cdot n_C} + \frac{d^2}{2(n_E + n_C)} \quad (3.84)$$

Sea T_i , la estimación del i -ésimo efecto de una colección de k estudios ($i = 1, 2, 3, \dots, k$). Donde T_i corresponde a la ecuación 3.81. En un modelo de efectos aleatorios cada efecto T_i , es estimador del tamaño del efecto paramétrico θ_i con varianza condicional σ_i^2 y estimada por $\hat{\sigma}_i^2$. El estimador condicional de la varianza $\hat{\sigma}_i^2$ viene dado por la ecuación 3.84. Luego el modelo puede ser propuesto $T_i = \theta_i + \varepsilon_i$, donde ε_i , son independientes y distribuidos normalmente con media cero y varianza σ_i^2 , $\varepsilon_i \sim N(0, \sigma_i^2)$. La varianza condicional representa la variabilidad intra estudio, que es la variabilidad producida por el error

del muestreo.

A su vez, el parámetro del tamaño del efecto θ_i , se refiere a un parámetro de efecto distribuido con media μ_θ y varianza incondicional τ^2 . Donde cada parámetro puede ser definido como $\theta_i = \mu_\theta + u_i$, se asume que los errores u_i son normales e independientes con media cero y varianza τ^2 [$u_i \sim N(0, \tau^2)$]. El valor τ^2 , representa el verdadero valor de heterogeneidad entre los efectos del estudio producidos por innumerables factores sustantivos (tipo de tratamiento, características del paciente), metodológicos (tipo de diseño, tamaño de la muestra, desgaste, etc.) y características de los estudios, Lipsey [79]. Por otro lado T_i , se distribuye normalmente con media μ_θ y varianza $\tau^2 + \sigma_i^2$ [$T_i \sim N(\mu_\theta, \tau^2 + \sigma_i^2)$]. Cuando no existe variabilidad entre los efectos estimados, la varianza de la variabilidad inter estudio es cero ($\tau^2 = 0$), existiendo variabilidad por el error de muestreo, que es representada en el modelo por la varianza incondicional intra estudio, σ_i^2 . En este caso todos los estudios estiman un único parámetro para el tamaño del efecto $\theta_i = \theta$ y el modelo estadístico simplificado queda expresado como $T_i = \theta + \varepsilon_i$, llegándose a convertir en un modelo de efectos fijos. El modelo de efectos fijos puede ser considerado como un caso particular de un modelo de efectos aleatorios, cuando no existe variabilidad inter estudio, y como consecuencia la estimación del efecto T_i , se ve afectada por el error de muestreo σ_i^2 siguiendo una distribución normal con media $\theta = \mu_\theta$ y varianza σ_i^2 [$T_i \sim N(\mu_\theta, \sigma_i^2)$], cuando el tamaño de la muestra es **grande**.

La estimación usual del tamaño promedio del efecto consiste en estimar cada efecto estimado, T_i , por el inverso de su varianza w_i , note ecuación 3.85. En el modelo de efectos fijos, el peso del factor es estimado como $w_i = \frac{1}{\hat{\sigma}_i^2}$, y cuando el modelo es de efectos aleatorios w_i , es estimado como $w_i = \frac{1}{\tau^2 + \hat{\sigma}_i^2}$. Para el índice **d** la estimación de la varianza intra estudio es denotada por σ_i^2 y definida por la ecuación 3.84. Un estimador para la varianza inter estudio τ^2 , se basa en el método de los momentos propuesto por DerSimonian and Laird [28], ecuación 3.86. El valor de c se estima mediante la ecuación 3.87, donde w_i son los pesos del factor para el i -ésimo estudio asumiendo un modelo de efectos fijos ($w_i = \frac{1}{\hat{\sigma}_i^2}$), k es el número de estudios y Q es el test estadístico para la heterogeneidad propuesto por Cochran [19] y definido por Hedges and Olkin [57], mediante la ecuación 3.88, donde \bar{T} se la estima mediante la ecuación 3.85.

$$\bar{T} = \frac{\sum_i w_i T_i}{\sum_i w_i} \quad (3.85)$$

$$\tau^2 = \begin{cases} \frac{Q - (k - 1)}{c} & \text{para } Q > k - 1 \\ 0 & \text{para } Q \leq k - 1 \end{cases} \quad (3.86)$$

$$c = \sum w_i - \frac{\sum w_i^2}{\sum w_i} \quad (3.87)$$

$$Q = \sum w_i (T_i - \bar{T})^2 \quad (3.88)$$

Un procedimiento para detectar si hay una significativa heterogeneidad entre un conjunto de estudios debe estar relacionado directamente con el valor de τ^2 y a la vez libre de escalas, con el objetivo de ser comparable. Si se asume que las varianzas condicional o intra estudio, σ_i^2 son conocidas, entonces bajo la

hipótesis de homogeneidad ($H_0 : \delta_1 = \delta_2 = \dots = \delta_k$; o $H_0 : \tau^2 = 0$), el estadístico Q tiene una distribución χ^2 con $k - 1$ grados de libertad. Por lo tanto, para valores superiores de Q con relación al punto crítico para un nivel de significación dada (α), permite rechazar la hipótesis nula y concluir que no es estadísticamente significativa la variación inter estudio.

Un problema con el estadístico Q es que su potencia estadística depende del número de estudios, siendo muy bajo o muy alto para un pequeño o un gran número de estudios. Para resolver los problemas del estadístico Q y la no compatibilidad de la varianza inter estudio, τ^2 entre los meta-análisis con diferentes tamaño de efectos (métricas distintas), Higgins and Thompson [58], han propuesto el índice I^2 . El índice I^2 , expresión 3.89 cuantifica el grado de heterogeneidad de un conjunto de tamaños de efectos, comparando el valor de Q con su valor esperado asumiendo homogeneidad, con $k-1$ grados de libertad.

$$I^2 = \begin{cases} \frac{Q - (k - 1)}{Q} \times 100\% & \text{para } Q > k - 1 \\ 0 & \text{para } Q \leq k - 1 \end{cases} \quad (3.89)$$

Cuando el estadístico Q es menor que sus grados de libertad, entonces I^2 es truncado a cero. El índice I^2 puede ser interpretado como un porcentaje de heterogeneidad, esto es, la parte total de variación que es debida a la varianza inter estudios $\hat{\tau}^2$. Existe una relación directa entre τ^2 e I^2 y puede ser formalizada a partir de las ecuaciones 3.86 y 3.89, mediante la expresión 3.90.

$$I^2 = \frac{c \cdot \hat{\tau}^2}{Q} \quad (3.90)$$

Finalmente, los métodos antes descritos, se pueden resumir en la Tabla 3.4, donde se detallan las bondades y supuestos que encierra cada modelo. Es importante notar que el modelo bivariante representa una mejora al enfoque univariante (SROC). Por otro lado el enfoque bivariante si se le añade una o más covariaciones que afecten la sensibilidad y la especificidad será equivalente a un modelo HSROC, vea Figura 3.12.

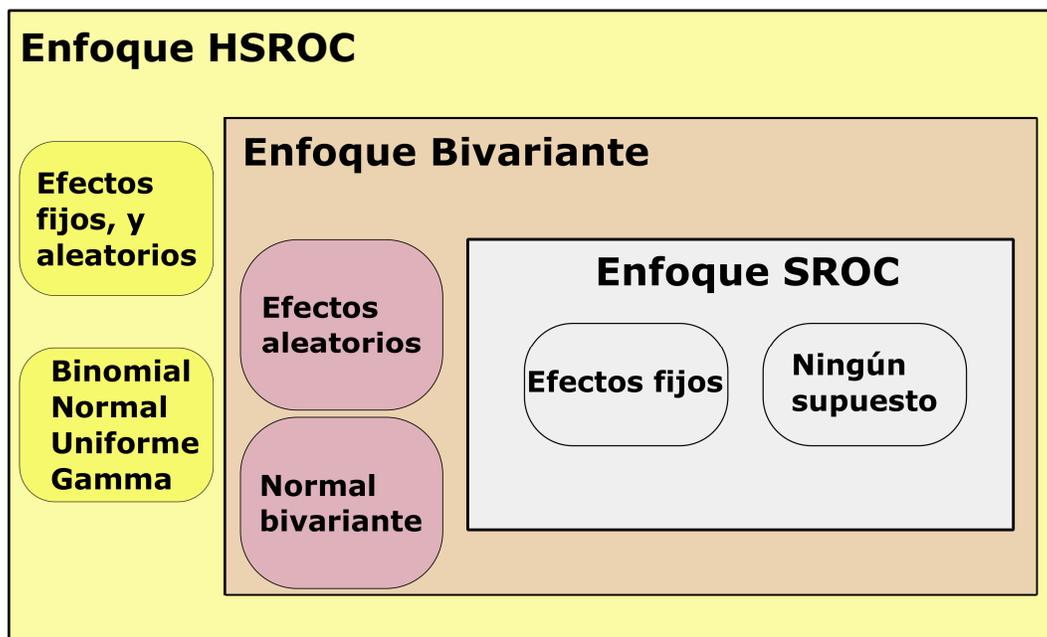


Figura 3.12: Diagrama de enfoques HSROC, Bivariante, SROC, elaboración propia

Para entender el funcionamiento de los modelos analizados en este capítulo, es necesario emplear algún programa estadístico, en nuestro caso se utilizará R y Stata, para obtener las salidas correspondientes que cada enfoque genera, es así que se tiene la Tabla 3.5 y el gráfico 3.13. La misma muestra las principales medidas que se estiman para el enfoque univariante y la función que se debe emplear en R, mayor detalle vea la Sección 4.1.8 del Capítulo 4. Asimismo para el enfoque bivariante se tiene la Tabla 3.6, las salidas para este enfoque se analizan con mayor detalle en la Sección 4.1.11 del Capítulo 4. Con relación al paquete Stata, se tiene la Tabla 3.7, la misma muestra la función a utilizar para realizar un análisis de meta-análisis en pruebas diagnósticas, mayor detalle se aprecia en la Sección 4.2 del Capítulo 4. La Tabla 3.8, donde se muestran las principales salidas del modelo HSROC en R, mayor detalle se analiza en la Sección 4.3 del Capítulo 4, por último, de lo expresado anteriormente se puede construir la Tabla 3.9, donde se muestran las principales medidas y gráficos para los enfoques antes descritos.

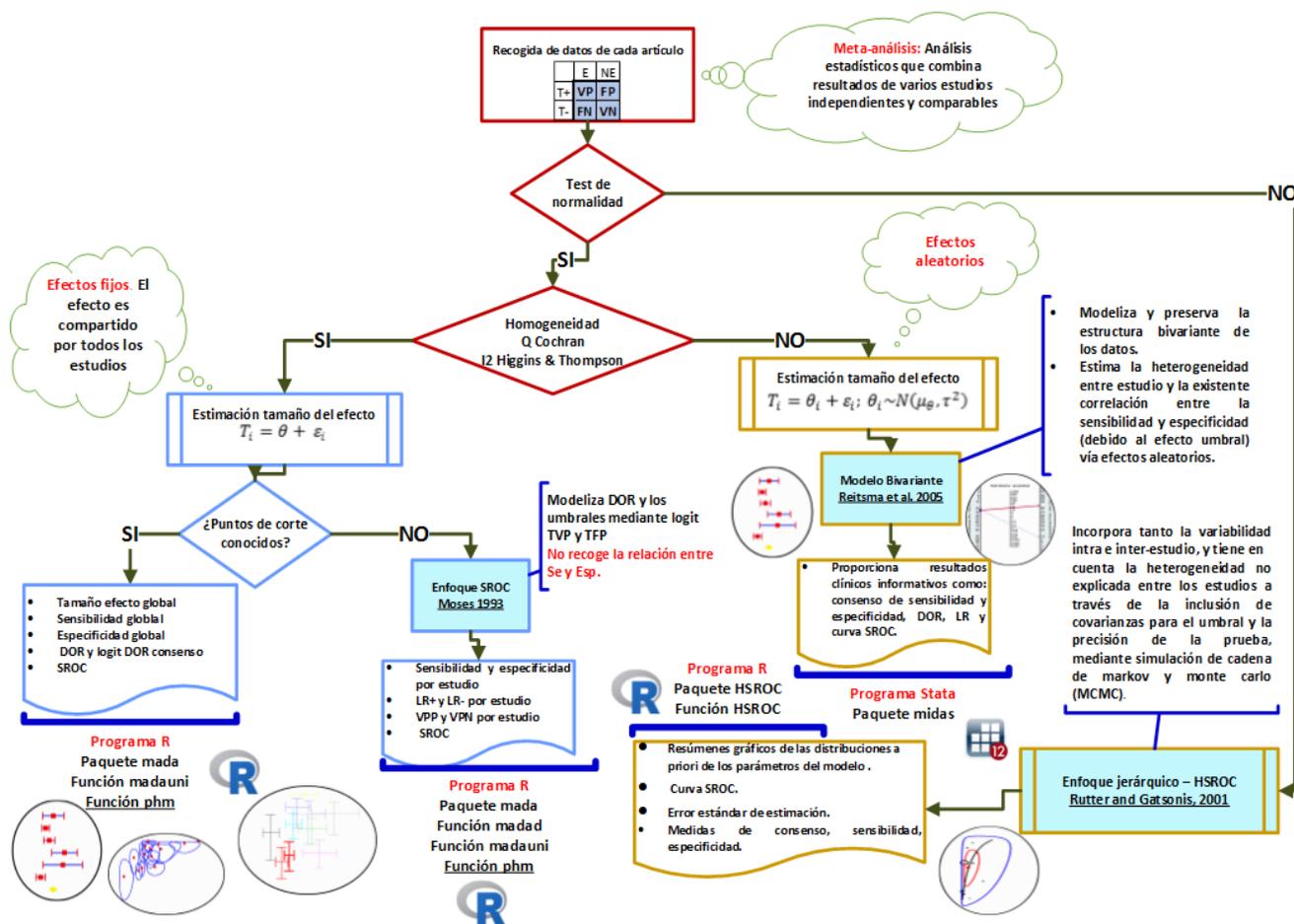


Figura 3.13: Esquema metodológico para el meta-análisis en pruebas diagnósticas

Enfoque	Descripción	Supuestos	Estimación de Parámetros
SROC	Describe un meta-análisis como odds ratio diagnóstico (DOR). Transforma la sensibilidad y especificidad en un indicador de la precisión. No cuantifica el error en S, que puede ser una fuente de sesgo en las estimaciones de parámetros y los resúmenes de la curva. Reconoce únicamente variación debido al tamaño de la muestra.	Ninguno	Mínimos cuadrados
Bivariante	Supone que el vector (<i>logit (sensibilidad)</i> ; <i>logit (especificidad)</i>) sigue una distribución normal bivalente. Preserva las dos dimensiones de los datos a través del análisis. Utiliza un enfoque de efectos aleatorios.	Binomial condicionada y Normal bivalente	Principio de máxima verosimilitud
HSROC	Expresa la sensibilidad y la probabilidad de falsos positivos de cada estudio como una distribución normal bivalente. Propone un mejor manejo de la variabilidad dentro y entre estudios, aplicando un enfoque bayesiano para la estimación de los parámetros	Nivel I (intra) - Binomial. Nivel II (inter) - Normal Nivel III (parámetros) - Uniforme y Gamma	Simulación. Método Montecarlo de Cadena de Markov

Tabla 3.4: Resumen de modelos estadísticos en pruebas diagnósticas

Función-R	Método	Efectos	Estimación	Gráficos
madauni	Mantel-Haenszel - MH	Fijos	Media Ponderada DOR	
madauni	DerSimonian and Laird - DSL	Aleatorio	Media Ponderada logaritmos DOR. \mathbf{Q}, I^2, τ^2	Forest Plot con diamante
phm	Modelo proporcional de Hazards PHM	Fijos y aleatorios	θ, τ, AUC	ROCellipse con bandas de confianza y crosshair

Tabla 3.5: Resumen de las funciones empleadas en R para enfoque univariante

Función-R	Método	Efectos	Estimación	Gráficos
reitsma	The restricted maximum likelihood (REML)	Fijos y Aleatorios	Efectos Fijos: tsen, tfp. Efectos aleatorios: tsen, tfp, sen, fp, $\sigma_A^2, \sigma_B^2, \sigma_{AB}, \sigma_{BA}$. HSROC: $\Theta, \Lambda, \sigma_\beta^2, \sigma_\alpha^2, \alpha$. AUC	Curva SROC con región de confianza
reitsma	REML bivalente	Meta-regresión Aleatorios	Efectos fijos: tsen, tfp, sen, fp, $\sigma_A^2, \sigma_B^2, \sigma_{AB}, \sigma_{BA}$.	

Tabla 3.6: Resumen de las funciones empleadas en R para enfoque bivalente

Función Stata	Método	Efectos	Estimación	Gráficos
midas	Principio de máxima verosimilitud	Fijos y Aleatorios	<ul style="list-style-type: none"> • Sensibilidad, • Especificidad • LR+ • LR- • DOR • Q • I^2 • ICC Sensibilidad • ICC Especificidad • Efectos de heterogeneidad entre estudios • AUC. 	<ul style="list-style-type: none"> • Forest Plot de Sen. y Esp. • Bondad de Ajuste. • Normal bivalente • Anomalías de influencia • Detección de atípicos • Gráfico de barras (calidad) • Gráfica de caja bivalente • Curva SROC con regiones de confianza y predicción • Funnel Plot • Fagan Plot • Diagrama de dispersión odd ratio • Gráfico de probabilidad modificatorio

Tabla 3.7: Resumen de las funciones empleadas en Stata para enfoque bivalente

Función en R	Método	Efectos	Estimación	Gráficos
HSROC, HSROC-Summary	Método de Monte Carlo de Cadena de Markov - MCMC	Fijos y Aleatorios	<ul style="list-style-type: none"> • Tamaño de la muestra, • Información a priori de $\Theta, \Lambda, \sigma_\alpha, \sigma_\theta$ • Información de parámetros inter-estudio, estimación, deviance estándar, MC_error con sus respectivos IC, $\Theta, \Lambda, \beta, \sigma_\alpha, \sigma_\beta, S$ y C • Información de parámetros intra-estudio, θ, α, prevalencia, estimador, deviance, MC_error, sensibilidad, especificidad con sus respectivos IC. • Valores predictivos a posteriori de sensibilidad y especificidad con sus respectivas deviance, MC_error e IC. 	<ul style="list-style-type: none"> • Gráfica de sensibilidad (seguimiento de cada parámetro) • Gráfica de densidad (función a posteriori de cada parámetro) • Curva SROC

Tabla 3.8: Resumen de las funciones empleadas en R para enfoque jerárquico-HSROC

Medida - Gráfico	Stata	Mada-R	HSROC-R
Sensibilidad IC	X	X	X
Especificidad IC	X	X	X
LR+ IC	X		
LR- IC	X		
DOR IC	X	X	
Q, I^2	X	X	
Eger's Index	X		
Curva SROC, regiones de confianza	X	X	X
Forest Plot con diamante	X	X	
Forest Plot sin diamante			
Hair cross - ROCellipses		X	
Parámetros del modelo HSROC			X

Tabla 3.9: Resumen de las principales medidas y gráficos para el enfoque univariante, bivalente y HSROC

CAPÍTULO 4

RESULTADOS

”Capacidad experimental, honestidad en la publicación de los resultados e inteligencia para interpretarlos.”

Richard Phillips Feynman

4.1 PRECISIÓN DIAGNÓSTICA CON MADA PARA META-ANÁLISIS

4.1.1 RESUMEN

El paquete R-mada es una herramienta para la precisión diagnóstica de un meta-análisis. En contraposición con un meta-análisis univariante, el meta-análisis de diagnóstico requiere un enfoque bivariante. El paquete implementa el enfoque de Reitsma, Glas, Rutjes, Scholten, Bossuyt, y Zwinderman (2005), en ausencia de covarianza que es equivalente al enfoque de Rutter and Gatsonis [114].

4.1.2 INTRODUCCIÓN

Si bien existen muchos métodos para meta-análisis de diagnóstico, aún no es un procedimiento rutinario. Una de las razones puede ser la complejidad del enfoque bivariante, pero otra de las razones es que algunos paquetes para el análisis de meta-análisis, por ejemplo Comprehensive Meta-Analysis and RevMan (Biostat, Inc. 2006; The Nordic Cochrane Centre 2011), no incluyen un ajuste apropiado del modelo para el meta-análisis de diagnóstico. Rutter and Gatsonis [114], le dan un enfoque bivariante, pero también existen modelos para meta-análisis de diagnóstico con enfoque bayesiano como por ejemplo WinBUGS, Lunn et al. [82] u OpenBUGS, Lunn et al. [81].

El paquete estadístico **”mada”** escrito para ejecutarse en R (R Core Team 2012), ofrece algunos enfoques actuales para el meta-análisis de diagnóstico, como también funciones que permiten obtener estadísticas descriptivas y gráficos.

4.1.3 OBTENIENDO EL PAQUETE

Si R está instalado y se cuenta con una conexión a internet, el paquete se puede instalar desde el repositorio (CRAN - The Comprehensive R Archive Network). En la dirección <http://r-forge.r-project.org/projects/mada/> se encuentra la versión más actualizada, pero las versiones estables están disponibles desde el CRAN. El paquete se puede cargar escribiendo los siguientes comandos.

```
>install.packages("mada")
> library("mada")
```

4.1.4 INGRESANDO LOS DATOS

En los estudios primarios de diagnóstico, los resultados de un procedimiento estándar de oro, definen la presencia o ausencia de una condición y el resultado de una prueba diagnóstica (normalmente algún tipo de procedimiento de bajo coste). Los datos de un estudio primario pueden ser reportados en una tabla de 2×2 , Tabla 4.1.

	Enfermedad-Condición		
	Sí	No	
Prueba diagnóstica	Sí	y_i	z_i (TP)
	No	$m_i - y_i$	$n_i - z_i$ (TN)
Total		m_i	n_i (T)

Tabla 4.1: Datos del estudio i -ésimo en una tabla de 2×2

Los valores de y_i , z_i , denotan los valores de verdaderos positivos (VP) y falsos positivos (FP) respectivamente, y los valores $m_i - y_i$ y $n_i - z_i$ muestran los valores de falsos negativos (FN) y verdaderos negativos (VN). Las medidas de precisión diagnóstica, se estiman a partir de una tabla de 2×2 . Usando la nomenclatura de la Tabla 4.1, se tienen las ecuaciones 4.1, 4.2 y 4.3.

$$p_i = \frac{y_i}{m_i} = \text{sensibilidad del } i\text{-ésimo estudio} \quad (4.1)$$

$$u_i = \frac{z_i}{n_i} = \text{tasa de falsos positivos del } i\text{-ésimo estudio} \quad (4.2)$$

$$1 - u_i = \frac{n_i - z_i}{n_i} = \text{especificidad del } i\text{-ésimo estudio} \quad (4.3)$$

Básicamente, todas las funciones de mada necesitan una tabla de 2×2 . De la fórmula 4.1 se tiene que $y_i = p_i m_i$. Si en un estudio primario se reporta una sensibilidad de 0.944 y que 142 personas tuvieron la condición, se puede calcular y , pero como el resultado anterior no es un entero, se necesita redondearlo. Note las siguientes líneas de comando, para lograr lo antes mencionado:

```
> y <- 142 * .944
> y
```

```
[1] 134.048
> round(y)
[1] 134
>
```

Se debe asumir que los valores VP, FP, FN y VN son conocidos para cada estudio. Una buena manera de organizar la información en R, es mediante data frame, el cual puede contener varias variables. Donde cada fila del data frame corresponde a un estudio primario. Como ejemplo se ingresarán en un data frame, información de seis estudios de un meta-análisis del AUDIT-C (una prueba corta para detectar problemas de alcohol), Kriston et al. [71], es decir, se tiene la siguiente salida:

```
> AuditC6 <- data.frame(TP = c(47, 126, 19, 36, 130, 84),
+ FN = c(9, 51, 10, 3, 19, 2),
+ FP = c(101, 272, 12, 78, 211, 68),
+ TN = c(738, 1543, 192, 276, 959, 89))
> AuditC6
```

	TP	FN	FP	TN
1	47	9	101	738
2	126	51	272	1543
3	19	10	12	192
4	36	3	78	276
5	130	19	211	959
6	84	2	68	89

Nótese que algunas de las funciones centrales de mada aceptan cuatro vectores de frecuencias (VP, FN, FP, VN) como entrada. Sin embargo, es conveniente para almacenar no sólo las frecuencias observadas, sino también los nombres de estudio en la misma trama de datos. El comando que se muestra a continuación, permite realizar lo antes mencionado.

```
> AuditC6$names <- c("Study 1", "Study 2", "Study 4",
+ "Study 4", "Study 5", "Study 6")
> AuditC6$names
```

```
[1] "Study 1" "Study 2" "Study 4" "Study 4" "Study 5" "Study 6"
```

El conjunto completo de datos del ejemplo se encuentra en mada con 14 estudios, como se muestra a continuación.

```
> data("AuditC")
> AuditC
```

	TP	FN	FP	TN
1	47	9	101	738
2	126	51	272	1543

3	19	10	12	192
4	36	3	78	276
5	130	19	211	959
6	84	2	68	89
7	68	0	112	423
8	752	0	3226	2977
9	59	5	55	136
10	142	50	571	2788
11	137	24	107	358
12	57	3	103	437
13	34	1	21	56
14	152	51	88	264

4.1.5 CELDAS VACÍAS

En el análisis de las tablas de 2×2 , se puede dar el caso que ciertas celdas tengan ceros, lo cual puede conducir a problemas o artificios estadísticos, ya que no se pueden estimar ciertos ratios. En lugar de que las celdas contengan ceros, se hace uso de la corrección por continuidad. Una sugerencia como corrección por continuidad es utilizar el valor de 0.5. Esta estrategia la utiliza mada por defecto.

4.1.6 ESTADÍSTICA DESCRIPTIVA

Las estadísticas descriptivas para un conjunto de datos incluyen la sensibilidad, especificidad y tasa de falso positivos de los estudios primarios y sus ratios de verosimilitud positivo y negativo ($LR+$, $LR-$) y sus odds ratios de diagnóstico, Glas et al. [45]. Están definidas como, $LR+ = \frac{p}{u} = \frac{\text{sensibilidad}}{\text{tasa de falsos positivos}}$, $LR- = \frac{1-p}{1-u}$ y $DOR = \frac{LR+}{LR-} = \frac{VP \cdot VN}{FN \cdot FP}$. Todas estas medidas se calculan fácilmente usando la función "madad", junto con sus intervalos de confianza. La función "madad" usa las fórmulas proporcionadas por Deeks [22]. madad utiliza un test de χ^2 para evaluar la heterogeneidad de las sensibilidades y especificidades. La hipótesis nula en ambos casos es que todos son iguales. Por último, la correlación de las sensibilidades y las tasas de falsos positivos se calcula para dar una pista de si existe un problema con los puntos de corte. Se pueden observar estas medidas con la siguiente sentencia:

```
> madad(AuditC)

Descriptive summary of AuditC with 14 primary studies.
Confidence level for all calculations set to 95 %
Using a continuity correction of 0.5 if applicable

Diagnostic accuracies
sens 2.5% 97.5% spec 2.5% 97.5%
[1,] 0.833 0.716 0.908 0.879 0.855 0.899
[2,] 0.711 0.640 0.772 0.850 0.833 0.866
[3,] 0.650 0.471 0.795 0.939 0.898 0.964
[4,] 0.912 0.785 0.967 0.779 0.733 0.819
[5,] 0.870 0.807 0.915 0.819 0.796 0.840
```

```
[6,] 0.971 0.912 0.991 0.566 0.489 0.641
[7,] 0.993 0.934 0.999 0.790 0.754 0.822
[8,] 0.999 0.994 1.000 0.480 0.468 0.492
[9,] 0.915 0.822 0.962 0.711 0.643 0.770
[10,] 0.738 0.672 0.795 0.830 0.817 0.842
[11,] 0.849 0.786 0.896 0.769 0.729 0.805
[12,] 0.943 0.854 0.979 0.809 0.773 0.840
[13,] 0.958 0.838 0.990 0.724 0.616 0.811
[14,] 0.748 0.684 0.802 0.749 0.702 0.792
```

```
Test for equality of sensitivities:
X-squared = 272.3603, df = 13, p-value = <2e-16
Test for equality of specificities:
X-squared = 2204.8, df = 13, p-value = <2e-16
```

Diagnostic OR and likelihood ratios

DOR	2.5%	97.5%	posLR	2.5%	97.5%	negLR	2.5%	97.5%	
[1,]	36.379	17.587	75.251	6.897	5.556	8.561	0.190	0.106	0.339
[2,]	13.913	9.818	19.717	4.736	4.100	5.470	0.340	0.270	0.429
[3,]	28.600	11.133	73.469	10.660	5.862	19.384	0.373	0.229	0.608
[4,]	36.732	11.925	113.144	4.127	3.320	5.129	0.112	0.041	0.306
[5,]	30.361	18.440	49.986	4.817	4.201	5.523	0.159	0.105	0.240
[6,]	44.162	12.077	161.486	2.240	1.868	2.687	0.051	0.015	0.173
[7,]	515.729	31.687	8393.755	4.730	4.009	5.581	0.009	0.001	0.145
[8,]	1388.854	86.752	22234.711	1.922	1.876	1.968	0.001	0.000	0.022
[9,]	26.607	10.523	67.271	3.167	2.506	4.001	0.119	0.053	0.266
[10,]	13.768	9.865	19.216	4.341	3.879	4.857	0.315	0.249	0.400
[11,]	18.716	11.572	30.271	3.679	3.079	4.397	0.197	0.136	0.284
[12,]	69.444	23.113	208.648	4.927	4.099	5.922	0.071	0.026	0.196
[13,]	60.442	10.948	333.681	3.477	2.411	5.014	0.058	0.012	0.277
[14,]	8.850	5.949	13.165	2.982	2.448	3.632	0.337	0.264	0.430

```
Correlation of sensitivities and false positive rates:
rho 2.5 % 97.5 %
0.677 0.228 0.888
```

La función **madad** tiene una gama de opciones con respecto a los detalles de los cálculos. Por ejemplo, se pueden calcular intervalos de confianza al 80%, mediante la sentencia `madad(AuditC, level = 0.80)`.

```
> madad(AuditC, level = 0.80)
```

```
Descriptive summary of AuditC with 14 primary studies.
Confidence level for all calculations set to 80 %
Using a continuity correction of 0.5 if applicable
```

```

Diagnostic accuracies
sens  10%  90%  spec  10%  90%
[1,] 0.833 0.761 0.887 0.879 0.864 0.893
[2,] 0.711 0.665 0.752 0.850 0.839 0.860
[3,] 0.650 0.533 0.751 0.939 0.914 0.957
[4,] 0.912 0.838 0.955 0.779 0.749 0.806
[5,] 0.870 0.831 0.901 0.819 0.805 0.833
[6,] 0.971 0.938 0.987 0.566 0.516 0.616
[7,] 0.993 0.964 0.999 0.790 0.767 0.812
[8,] 0.999 0.997 1.000 0.480 0.472 0.488
[9,] 0.915 0.860 0.950 0.711 0.667 0.751
[10,] 0.738 0.696 0.777 0.830 0.821 0.838
[11,] 0.849 0.809 0.881 0.769 0.743 0.793
[12,] 0.943 0.892 0.970 0.809 0.786 0.829
[13,] 0.958 0.892 0.985 0.724 0.655 0.784
[14,] 0.748 0.707 0.784 0.749 0.719 0.778

```

```

Test for equality of sensitivities:
X-squared = 272.3603, df = 13, p-value = <2e-16
Test for equality of specificities:
X-squared = 2204.8, df = 13, p-value = <2e-16

```

```

Diagnostic OR and likelihood ratios
DOR      10%      90%  posLR  10%      90%  negLR  10%      90%
[1,]  36.379  22.618  58.513  6.897  5.987  7.944  0.190  0.130  0.277
[2,]  13.913  11.077  17.475  4.736  4.310  5.204  0.340  0.293  0.396
[3,]  28.600  15.433  53.001 10.660  7.210 15.760  0.373  0.271  0.513
[4,]  36.732  17.603  76.651  4.127  3.579  4.757  0.112  0.058  0.216
[5,]  30.361  21.914  42.063  4.817  4.405  5.268  0.159  0.121  0.208
[6,]  44.162  18.918 103.093  2.240  1.989  2.523  0.051  0.023  0.113
[7,] 515.729  83.223 3195.961  4.730  4.245  5.270  0.009  0.002  0.056
[8,]1388.854 226.547 8514.419  1.922  1.892  1.952  0.001  0.000  0.008
[9,]  26.607  14.508  48.797  3.167  2.718  3.690  0.119  0.070  0.201
[10,]  13.768  11.072  17.121  4.341  4.033  4.672  0.315  0.270  0.368
[11,]  18.716  13.667  25.630  3.679  3.275  4.134  0.197  0.155  0.250
[12,]  69.444  33.825 142.573  4.927  4.369  5.557  0.071  0.036  0.138
[13,]  60.442  19.778 184.715  3.477  2.737  4.417  0.058  0.021  0.161
[14,]   8.850   6.826  11.474  2.982  2.621  3.392  0.337  0.287  0.395

```

```

Correlation of sensitivities and false positive rates:
rho  2.5 %  97.5 %
0.677  0.228  0.888

```

Las salidas de madad pueden ser asignadas a un objeto. Por ejemplo, las tasas de falsos positivos, con sus respectivos intervalos de confianza, pueden ser extraídos utilizando el constructo \$, mediante la

sentencia `AuditC.d <- madad(AuditC)`, vea la siguiente salida.

```
> AuditC.d <- madad(AuditC)
> AuditC.d\$fpr

$fpr
[1] 0.12083333 0.15005507 0.06097561 0.22112676 0.18061486 0.43354430
[7] 0.20988806 0.52006770 0.28906250 0.17008929 0.23068670 0.19131238
[13] 0.27564103 0.25070822

$fpr.ci
2.5%      97.5%
[1,] 0.10050071 0.1446182
[2,] 0.13436927 0.1672182
[3,] 0.03560834 0.1024938
[4,] 0.18106960 0.2671547
[5,] 0.15963660 0.2036817
[6,] 0.35875648 0.5114869
[7,] 0.17753840 0.2463665
[8,] 0.50762738 0.5324832
[9,] 0.22957510 0.3568250
[10,] 0.15776394 0.1831681
[11,] 0.19473290 0.2710443
[12,] 0.16038974 0.2265879
[13,] 0.18879306 0.3835508
[14,] 0.20834216 0.2984416
```

4.1.7 GRÁFICOS DESCRIPTIVOS

Para los datos de AUDIT-C, la prueba χ^2 , sugiere la heterogeneidad entre la sensibilidad y la especificidad. El correspondiente forest plot, note Figuras 4.1a y 4.1b, confirma esta afirmación, mediante las sentencias:

```
> forest(madad(AuditC), type = "sens")
> forest(madad(AuditC), type = "spec")
```

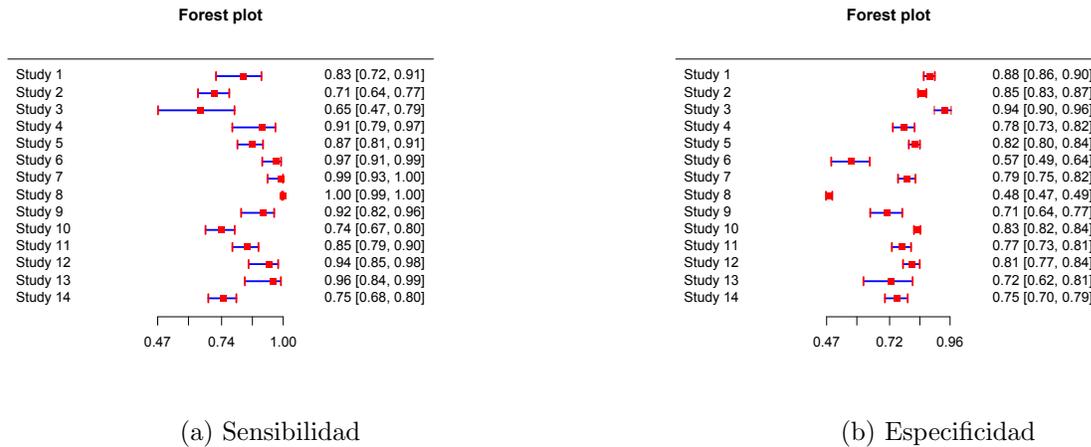


Figura 4.1: Gráficos forest-plot, para sensibilidad y especificidad, datos AuditC

Además de los gráficos anteriores, mada provee una variedad de gráficos para estudiar los datos en el espacio ROC. Hay que tener en cuenta que para fines de exploración es a menudo útil emplear el color y otros detalles con los que R cuenta para dar a los gráficos. Existen dos grandes niveles de gráficos en mada: "crosshair" que producen gráficos "crosshair" Phillips et al. [99], y ROCellipse. El gráfico crosshair puede producir colores a los crosshair y hace que los crosshairs sean más amplios mientras se incrementa el tamaño de la muestra, el gráfico se genera con las sentencias `rs <- rowSums(AuditC)`, `weights <- 4 * rs / max(rs)`, `crosshair(AuditC, xlim = c(0,0.6), ylim = c(0.4,1), col = 1:14, lwd = weights)`, vea Figura 4.2a.

```
> rs <- rowSums(AuditC)
> weights <- 4 * rs / max(rs)
> crosshair(AuditC, xlim = c(0,0.6), ylim = c(0.4,1),
+ col = 1:14, lwd = weights)
```

Asimismo, muestra un gráfico ROCellipse, el cual muestra una región de confianza que describe la incertidumbre de la pareja sensibilidad y especificidad. Estas regiones son elipses sobre el espacio logit ROC, y son generadas con las sentencias, `ROCellipse(AuditC, pch = "")`, `points(fpr(AuditC), sens(AuditC))`, vea Figura 4.2b.

```
> ROCellipse(AuditC, pch = "")
> points(fpr(AuditC), sens(AuditC))
```

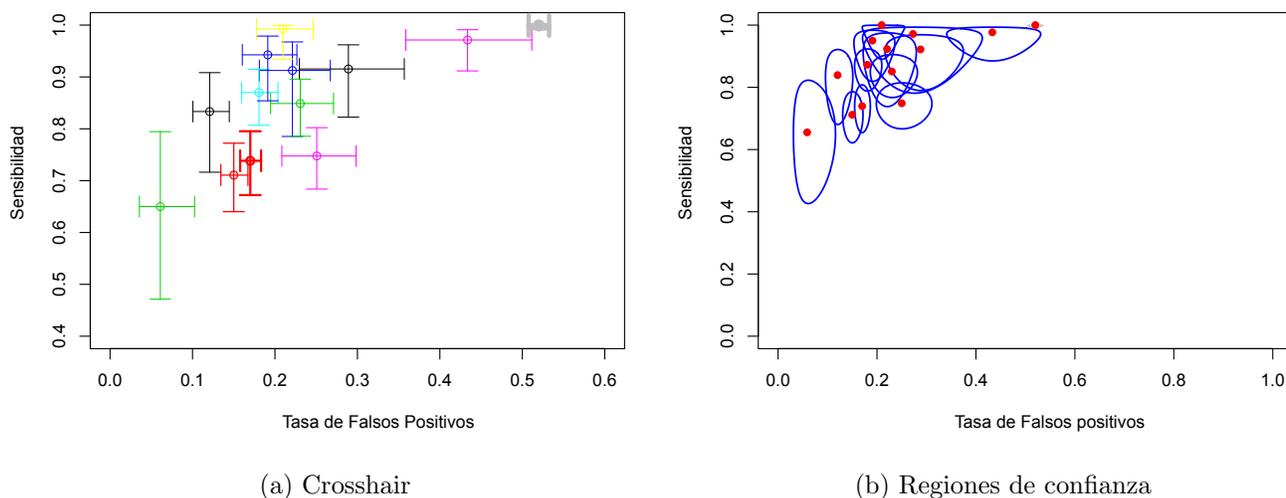


Figura 4.2: Gráfico ponderado crosshair con colores arbitrarios y regiones de confianza en forma de elipse, datos AuditC

El gráfico crosshair es una combinación de un gráfico ROC y un forest-plot que demuestra la relación bivalente y el grado de heterogeneidad entre la sensibilidad y la tasa de falsos positivos.

4.1.8 ENFOQUE UNIVARIANTE

Antes de la llegada del enfoque bivalente, por [114] y [105], algunos enfoques univariantes eran muy populares. El enfoque bivalente no es recomendable cuando el tamaño de la muestra es muy pequeño. El modelo bivalente de Reitsma et al. [105], tiene 5 parámetros, que serían claramente demasiados para un número pequeño de estudios.

4.1.9 ODDS RATIO DE DIAGNÓSTICO (DOR)

El enfoque DOR, desarrollado por Glas et al. [45] es estimado por el método de Mantel-Haenszel (MH) véase por ejemplo Deeks [22]. Mada proporciona los resultados correspondientes. El modelo de efectos fijos subyacente tiene la forma, $DOR_i = \mu + \varepsilon_i$, donde μ es la media de los DOR's y ε_i , son errores independientes con varianza específica. El estimador MH es una media ponderada de los DOR's observados de los estudios primarios y es robusto para celdas con valores nulos. Se estima como $\hat{\mu} = \sum_i \frac{w_i^{MH} DOR_i}{\sum_i w_i^{MH}}$,

donde $w_i^{MH} = \frac{z_i(m_i - y_i)}{m_i + n_i}$ son los pesos de Mantel-Haenszel. Un estimador para un modelo de efectos aleatorios sigue el enfoque de DerSimonian and Laird [28]. El modelo se formula en términos de los logaritmos de los DOR's, luego se tiene que $\log(DOR_i) = \mu + \varepsilon_i + \delta_i$, donde μ es la media de los $\log(DOR_i)$ y $\varepsilon_i + \delta_i$, son independientes con media cero, y la varianza ε_i , denotada por σ_i^2 , se estima como $\hat{\sigma}_i^2 = \frac{1}{y_i} + \frac{1}{m_i - y_i} + \frac{1}{z_i} + \frac{1}{n_i - z_i}$, y la varianza de δ_i , denotada por τ^2 , cuyo valor tiene que ser estimado.

El DSL es un estimador ponderado, por lo que se tiene que $\hat{\mu} = \sum_i \frac{w_i^{DSL} DOR_i}{\sum_i w_i^{DSL}}$, donde $w_i^{DSL} = \frac{1}{\hat{\sigma}_i^2 + \tau_2}$ y τ^2 es estimado mediante el estadístico Q de Cochran. La función "madauni" puede ayudar a estimar τ^2 , mediante la sentencia `fit.DOR.DSL <- madauni(AuditC)`.

```
> (fit.DOR.DSL <- madauni(AuditC))
```

```
Call:
madauni(x = AuditC)
```

```
DOR tau^2
26.337  0.311
```

Asimismo, se puede obtener una estimación del DOR mediante el método MH con la sentencia, `fit.DOR.MH <- madauni(AuditC, method = "MH")`.

```
> (fit.DOR.MH <- madauni(AuditC, method = "MH"))
```

```
Call:
madauni(x = AuditC, method = "MH")
```

```
DOR
17.93335
```

La función madauni no es muy informativa, así que empleando la sentencia `summary(fit.DOR.DSL)`, se puede obtener más detalle para el enfoque DSL, como se muestra a continuación.

```
> (summary(fit.DOR.DSL))
```

```
Call:
madauni(x = AuditC)
```

```
Estimates:
DSL estimate  2.5 % 97.5 %
DOR           26.337 17.971 38.596
lnDOR         3.271  2.889  3.653
tau^2         0.311  0.000  3.787
tau           0.557  0.000  1.946
```

```
Cochran's Q: 19.683 (13 df, p = 0.103)
Higgins' I^2: 33.955%
```

Además de notar los intervalos de confianza, el estadístico Q, Cochran [19], también se observa el I^2 , Higgins et al. [59]. Elaborar un gráfico de forest-plot de los logaritmos de los DOR's es muy sencillo, se utiliza el método forest de la clase madauni mediante la sentencia `forest(fit.DOR.DSL)`, como se muestra en la Figura 4.3.

```
> forest(fit.DOR.DSL)
```

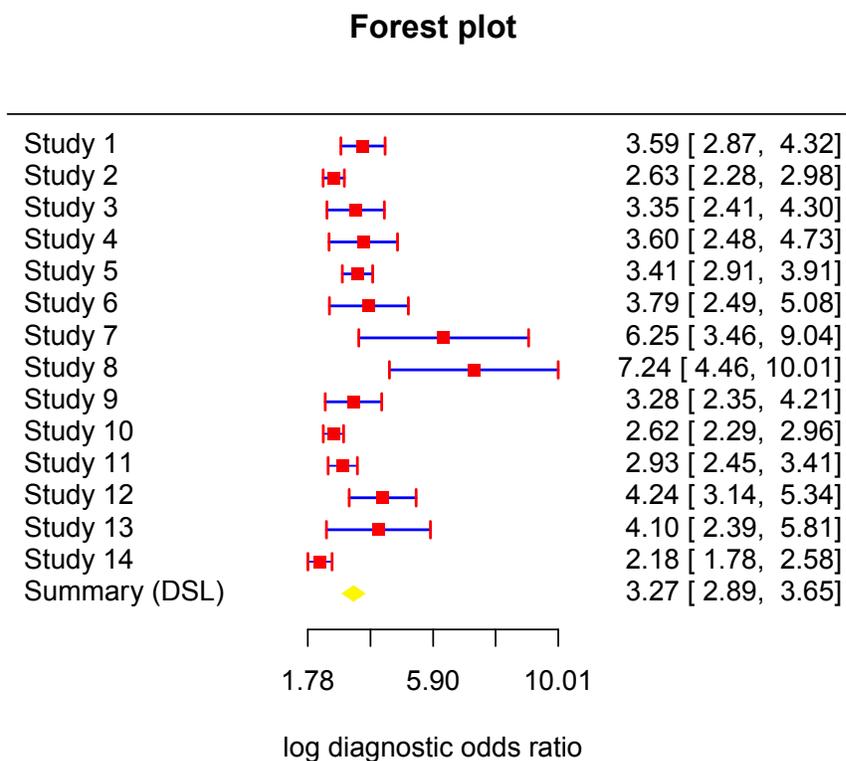


Figura 4.3: Forest-plot, logaritmos DOR's, ajuste DSL, datos AuditC

4.1.10 MODELO PROPORCIONAL HAZARDS

El modelo proporcional de hazards (PHM; vea Holling et al. [61]), se construye sobre la asunción de una curva ROC simple y se asumen las condiciones del modelo de Lehmann, Le [73]. Sea p_i y u_i , los valores de la sensibilidad y la tasa de falsos positivos del i -ésimo estudio respectivamente. Entonces la relación entre p_i y u_i es: $p_i = u_i^\theta$ y θ es el parámetro de la precisión del diagnóstico. Cuanto más pequeño sea el valor de θ , mayor será el área bajo la curva ROC, lo cual indica que el test es más preciso en el diagnóstico. Para estimar θ , el estimador APMLE es implementado en mada, para el caso de homogeneidad (efectos fijos), se realiza con el comando `fit.phm.homo <- phm(AuditC, hetero = FALSE)`.

```
> (fit.phm.homo <- phm(AuditC, hetero = FALSE))
```

Call:

```
phm.default(data = AuditC, hetero = FALSE)
```

Coefficients:

```
theta
0.004586893
```

Para el caso de heterogeneidad, se realiza con el comando `fit.phm.het <- phm(AuditC)`.

```
> (fit.phm.het <- phm(AuditC))
```

```
Call:
phm.default(data = AuditC)
```

```
Coefficients:
theta      taus_sq
0.084631351 0.003706143
```

La salida de la función `phm`, donde el método `summary` es más informativo, se realiza mediante la sentencia `summary(fit.phm.homo)`.

```
> summary(fit.phm.homo)
```

```
Call:
phm.default(data = AuditC, hetero = FALSE)
```

```
Estimate      2.5 %      97.5 %
theta 0.004586893 0.003508507 0.00566528
```

```
Log-likelihood: -61.499 on 1 degrees of freedom
AIC: 125
BIC: 125.6
```

```
Chi-square goodness of fit test (Adjusted Profile Maximum Likelihood
under homogeneity)
```

```
data: x
Chi-square = 222.47, df = 1, p-value < 2.2e-16
```

```
AUC 2.5 % 97.5 % pAUC 2.5 % 97.5 %
0.995 0.997 0.994 0.994 0.995 0.992
```

```
>
```

La prueba de bondad de ajuste χ^2 , rechaza la homogeneidad. Es decir, la hipótesis de heterogeneidad se acepta, y se comprueba con la sentencia `summary(fit.phm.het)`.

```
> summary(fit.phm.het)
```

```

Call:
phm.default(data = AuditC, hetero = TRUE)

Estimate      2.5 %      97.5 %
theta  0.084631351  0.047449859  0.121812844
taus_sq 0.003706143 -0.001277798  0.008690085

Log-likelihood: 31.121 on 2 degrees of freedom
AIC: -58.2
BIC: -57

Chi-square goodness of fit test (Adjusted Profile Maximum Likelihood
under heterogeneity)

data: x
Chi-square = 13.726, df = 2, p-value = 0.3185

AUC  2.5 % 97.5 %  pAUC  2.5 % 97.5 %
0.922 0.955 0.891 0.891 0.937 0.848

```

La estimación de θ , se obtiene a partir de la curva SROC. Graficando la curva en conjunto con las bandas de confianza obtenidas de los intervalos de confianza de θ a partir de `summary`, que es ejecutada con el método gráfico, mediante las sentencias `plot(fit.phm.het, xlim = c(0,0.6), ylim = c(0.4,1))`, `ROCellipse(AuditC, add = TRUE)`.

```

> plot(fit.phm.het, xlim = c(0,0.6), ylim = c(0.4,1))
> ROCellipse(AuditC, add = TRUE)

```

Nótese, que la curva SROC, no se extrapola más allá del rango de los datos originales. El área bajo la curva (AUC), forma parte de la salida del `summary`. Para PHM, es calculada por $AUC = \frac{1}{\theta + 1}$. note Figura 4.4.

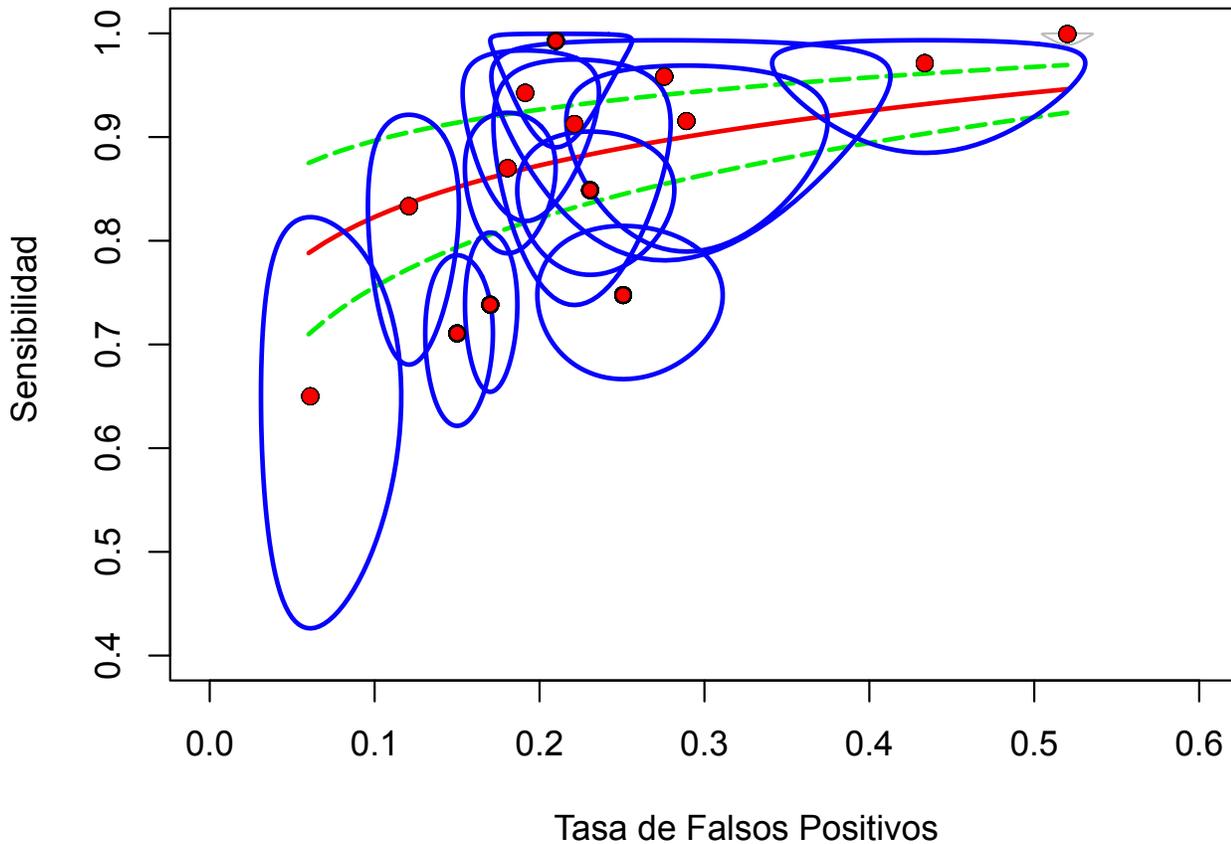


Figura 4.4: Resumen gráfico con modelo PHM, datos AuditC

4.1.11 ENFOQUE BIVARIANTE

Típicamente, la sensibilidad y la especificidad de una prueba diagnóstica dependen de los puntos de corte que se elijen en cada estudio, es decir, al incrementar la sensibilidad, la especificidad disminuye. En un entorno meta-analítico, la relación entre estas medidas es negativa. Es decir, existe una correlación positiva entre la sensibilidad y la tasa de falsos positivos. Dado que estas cantidades están relacionadas entre sí, el enfoque bivalente para meta-análisis en la precisión del diagnóstico ha tenido acogida, Arends et al. [2], Harbord et al. [54], Reitsma et al. [105], Rutter and Gatsonis [114], Van Houwelingen et al. [142].

Se asume que la sensibilidad y la tasa de falsos positivos siguen un modelo binomial condicionado y las transformaciones logit de la sensibilidad y la tasa de los falsos positivos un modelo normal bivalente. Existen dos maneras de llegar al modelo final: un modelo no lineal mixto o un modelo lineal mixto,

Arends et al. [2]. El último enfoque se encuentra en mada como la función **reitsma**. Se puede entender lo anteriormente expresado como una meta-regresión multivariante y el paquete **mvmeta** sirve para este enfoque, Gasparrini et al. [41].

Sea p_i y u_i , la sensibilidad y la tasa de falsos positivos del i -ésimo estudio respectivamente, donde p_i y u_i denotan sus estimaciones a partir de las frecuencias observadas. Entonces, dado un modelo binomial que se asume condicional sobre p_i , la varianza de $\text{logit}(p_i)$ puede ser aproximada como $\frac{\hat{p}_i(1-\hat{p}_i)}{m_i}$ y la varianza de $\text{logit}(u_i)$ como $\frac{\hat{u}_i(1-\hat{u}_i)}{n_i}$. Para los niveles dentro de los estudios, se asume condicional sobre p_i y u_i . La variación observada puede ser descrita por estas varianzas y un modelo normal. Sea D_i una matriz de 2×2 con dos varianzas en su diagonal principal. Sobre los niveles de estudio se supone una media global $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}^T$ y la matriz de covarianzas Σ , vea ecuación 4.4, la cual describe la heterogeneidad del par ordenado $(\text{logit}(\hat{p}_i), \text{logit}(\hat{u}_i))$. Luego el modelo del i -ésimo estudio es $(\text{logit}(\hat{p}_i), \text{logit}(\hat{u}_i))^T \sim N(\mu, \Sigma + D_i)$, y el modelo en mada es similar a las otras funciones, y se tienen varias salidas utilizando la sentencia `fit.reitsma <- reitsma(AuditC)`.

```
> (fit.reitsma <- reitsma(AuditC))

Call: reitsma.default(data = AuditC)

Fixed-effects coefficients:
tsens      tfpr
(Intercept) 2.0997 -1.2637

14 studies, 2 fixed and 3 random-effects parameters
logLik      AIC      BIC
31.5640    -53.1279  -46.4669
```

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma \\ \sigma & \sigma_2^2 \end{pmatrix} \quad (4.4)$$

Las salidas del objeto `reitsma` son escasas, pero se puede obtener más información con `summary(summary(fit.reitsma))`.

```
> summary(fit.reitsma)

Call: reitsma.default(data = AuditC)

Bivariate diagnostic random-effects meta-analysis
Estimation method: REML

Fixed-effects coefficients
Estimate Std. Error      z Pr(>|z|) 95%ci.lb 95%ci.ub
tsens.(Intercept)    2.100      0.338 6.215    0.000    1.438    2.762 ***
```

```

tfpr.(Intercept)  -1.264      0.174 -7.249    0.000   -1.605   -0.922 ***
sensitivity        0.891          -      -          -     0.808    0.941
false pos. rate   0.220          -      -          -     0.167    0.285
---

```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Variance components: between-studies Std. Dev and correlation matrix

```

Std. Dev tsens  tfpr
tsens    1.175 1.000 .
tfpr     0.638 0.854 1.000

```

```

logLik    AIC    BIC
31.564 -53.128 -46.467

```

AUC: 0.887

Partial AUC (restricted to observed FPRs and normalized): 0.861

HSROC parameters

```

Theta      Lambda      beta sigma2theta sigma2alpha
-0.083      3.262      -0.610      0.695      0.218

```

Nótese que la sensibilidad y la tasa de falsos positivos son las transformaciones de μ_1 y μ_2 . Luego entonces se puede graficar la curva SROC de este modelo. Por defecto, la representación del par ordenado de la sensibilidad y la tasa de falsos positivos se representan con una región de confianza. Es posible graficar la curva SROC un poco más gruesa con el argumento **sroclwd** así como agregar un título, mediante las sentencias:

```

> plot(fit.reitsma, sroclwd = 2,
+ main = "SROC curve (bivariate model) for AUDIT-C data")
> points(fpr(AuditC), sens(AuditC), pch = 2)
> legend("bottomright", c("data", "summary estimate"), pch = c(2,1))
> legend("bottomleft", c("SROC", "conf. region"), lwd = c(2,1))
>

```

Por defecto, la curva SROC no extrapola más allá del rango de los datos originales, observe Figura 4.5.

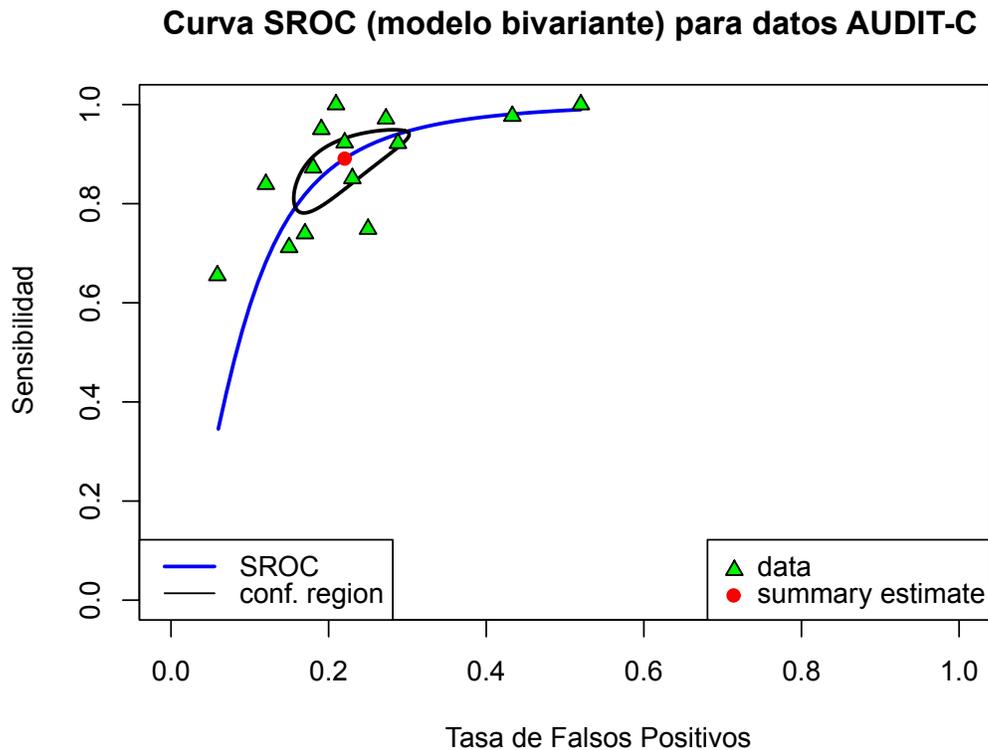


Figura 4.5: Curva SROC, modelo de Reitsma, datos AuditC

4.1.12 COMPARANDO CURVAS SROC

En este apartado se mostrará como comparar curvas SROC. Patrick et al. [97], condujeron un meta-análisis para averiguar la eficacia de los cuestionarios, los cuales fueron cumplimentados por el entrevistado (SAQ) y por el entrevistador (IAQ), para detectar el consumo de nicotina. El conjunto de datos está constituido por los subconjuntos de datos IAQ y SAQ. Primero los subconjuntos de datos se ajustan a un modelo bivalente mediante las sentencias `data("IAQ")`, `data("SAQ")`, `fit.IAQ <- reitsma(IAQ)`, `fit.SAQ <- reitsma(SAQ)`.

```
> data("IAQ")
> data("SAQ")
> fit.IAQ <- reitsma(IAQ)
> fit.SAQ <- reitsma(SAQ)
```

Luego se grafica la curva SROC de los ajustes, comenzando con el ajuste de IAQ y agregando después al gráfico el ajuste de SAQ, mediante las sentencias:

```
> plot(fit.IAQ, xlim = c(0,.5), ylim = c(.5,1),
+ main = "Comparison of IAQ and SAQ")
```

```

> lines(sroc(fit.SAQ), lty = 2)
> ROCellipse(fit.SAQ, lty = 2, pch = 2, add = TRUE)
> points(fpr(IAQ), sens(IAQ), cex = .5)
> points(fpr(SAQ), sens(SAQ), pch = 2, cex = 0.5)
> legend("bottomright", c("IAQ", "SAQ"), pch = 1:2, lty = 1:2)
>

```

Nótese que el argumento `lty` es utilizado para distinguir las curvas de ambos ajustes vea Figura 4.6.

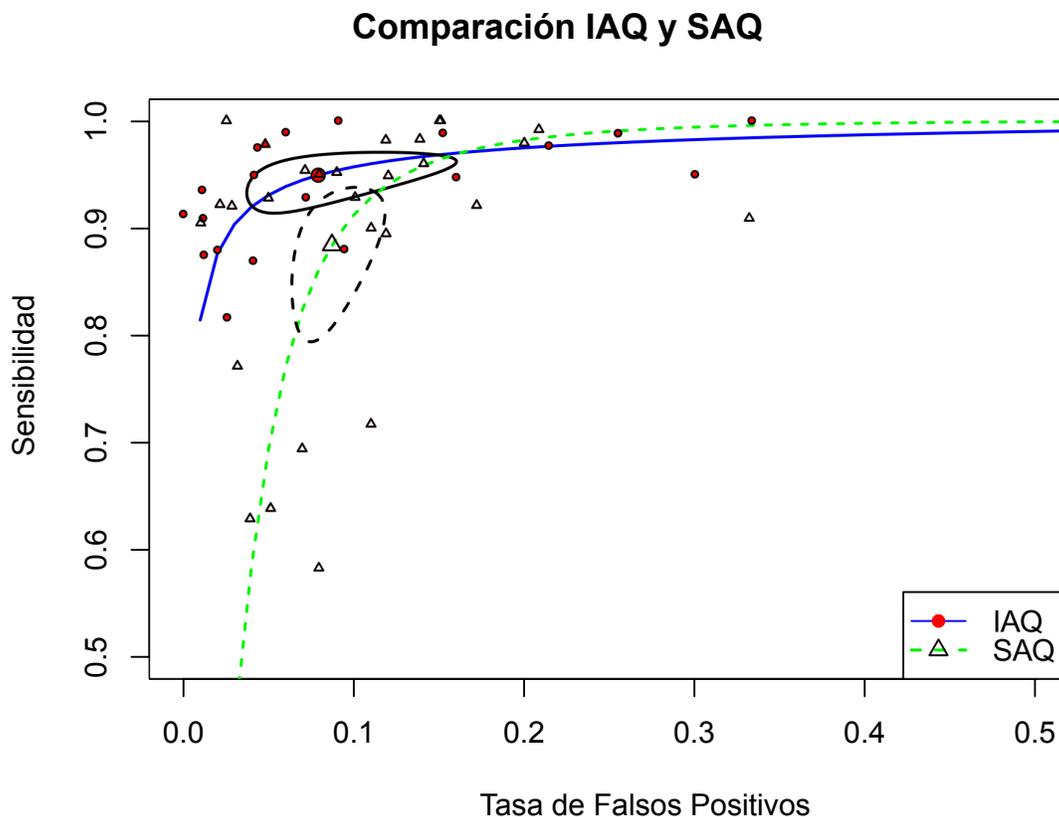


Figura 4.6: Comparación de curvas SROC para SAQ (entrevistados) y IAQ (entrevistadores), datos IAQ y SAQ

4.1.13 META-REGRESIÓN BIVARIANTE

Usaremos el conjunto de datos de Patrick et al. [97], para demostrar el enfoque bivalente de meta-regresión. Los datos se cargan mediante la sentencia `data("smoking")`. El conjunto de datos contiene las mismas variables que el subconjunto de datos IAQ y SAQ, pero con tipos de variables diferentes, y se puede verificar mediante la sentencia `summary(smoking$type)`.

```

> data("smoking")
> summary(smoking$type)

```

```
IAQ SAQ
20 31
```

Se utiliza "type" como factor de covariable en la meta-regresión de diagnóstico. Nótese que el lado izquierdo del objeto tiene que tener la forma `cbind(tsens, tfpr)`, donde `tsens` y `tfpr` representan las transformaciones de la sensibilidad y la tasa de falsos positivos respectivamente, generando la salida, mediante las sentencias:

```
> fit.smoking.type <- reitsma(smoking,
+ formula = cbind(tsens, tfpr) ~ type)
> summary(fit.smoking.type)

Call:  reitsma.default(data = smoking, formula = cbind(tsens, tfpr) ~
type)

Bivariate diagnostic random-effects meta-analysis
Estimation method: REML

Fixed-effects coefficients
Estimate Std. Error      z Pr(>|z|) 95%ci.lb 95%ci.ub
tsens.(Intercept)    3.007      0.307  9.809    0.000    2.406    3.608 ***
tsens.typeSAQ        -0.999      0.395 -2.531    0.011   -1.772   -0.225  *
tfpr.(Intercept)    -2.407      0.243 -9.912    0.000   -2.883   -1.931 ***
tfpr.typeSAQ         0.019      0.307  0.061    0.952   -0.583    0.621
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Variance components: between-studies Std. Dev and correlation matrix
Std. Dev tsens  tfpr
tsens    1.286 1.000 .
tfpr     1.003 0.510 1.000

logLik      AIC      BIC
126.622 -239.244 -220.869
```

La salida anterior puede ser interpretada de la siguiente manera: El valor de z para los coeficientes de regresión, para la sensibilidad es significativa, indicando que la realización de una entrevista ofrece una mejor sensibilidad. Es interesante notar que las tasas de falsos positivos no indican ningún efecto. Una vez que se utiliza meta-regresión, no se pueden trazar curvas SROC razonables, ya que se necesitan valores para la covarianzas. Los valores de AUC no tienen sentido. Se puede comparar el ajuste de dos meta-regresiones bivariantes con la prueba del ratio de máxima verosimilitud. Para esto, se necesita que los modelos se ajusten por el método de máxima verosimilitud, y como la prueba del ratio de máxima verosimilitud emplea la teoría asintótica, la cual solo es válida si se emplea este método de estimación, lo anteriormente mencionado se realiza mediante las sentencias:

```
> fit.smoking.ml.type <- reitsma(smoking,
+ formula = cbind(tsens, tfpr) ~ type, method = "ml")
```

```

> fit.smoking.ml.intercept <- reitsma(smoking,
+ formula = cbind(tsens, tfpr) ~ 1,
+ method = "ml")
> anova(fit.smoking.ml.type, fit.smoking.ml.intercept)

Likelihood-ratio test
Model 1: cbind(tsens, tfpr) ~ type
Model 2: cbind(tsens, tfpr) ~ 1

ChiSquared Df Pr(>ChiSquared)
7.858 2 0.0197 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

>
>

```

La meta-regresión confirma que el "type" explica algo de heterogeneidad entre los estudios primarios.

4.2 PRECISIÓN DIAGNÓSTICA CON STATA PARA META-ANÁLISIS

4.2.1 DESCRIPCIÓN

Midas es un programa exhaustivo de rutinas estadísticas y gráficas para entender el meta-análisis de las pruebas de diagnóstico en Stata, el cual es un paquete de software estadístico creado en 1985 por StataCorp. El índice y la prueba de referencia (prueba oro) son dicotómicas. La síntesis de los datos primarios es desarrollada dentro de una regresión bivalente de efectos mixtos, centrada en hacer inferencia sobre la media de la sensibilidad y especificidad. El enfoque bivalente fue desarrollado para ensayos de tratamiento en meta-análisis, Van Houwelingen et al. [142, 143] y modificado para resumir los datos de una prueba de diagnóstico usando una aproximación normal dentro de los estudios, Arends et al. [2], Reitsma et al. [105], Riley et al. [106], Riley et al. [107], Riley et al. [108]. Una interpretación binomial del modelo bivalente, se obtiene a partir de, Arends et al. [2], Chu and Cole [18], Riley et al. [107], donde se asume que distribuciones binomiales son independientes para los verdaderos positivos y verdaderos negativos condicionadas a la sensibilidad y especificidad en cada estudio.

La estimación basada en máxima verosimilitud para el enfoque binomial es desarrollada adoptando una cuadratura gaussiana, usando el comando `xtmelogit` o `gllamm`, Rabe-Hesketh et al. [101, 102]. Midas permite analizar la heterogeneidad (relacionado con los umbrales y covarianza) y los sesgos de publicación o precisión.

En general, bajo el paquete estadístico Stata se trabaja mediante lenguaje de programación, y conviene recordar que es **case-sensitive**, es decir, diferencia entre letras mayúsculas y minúsculas. En relación a la capacidad del manejo de grandes volúmenes de datos, el paquete Stata requiere hacer una copia a la base de datos con la que se va a trabajar en la memoria RAM del computador donde se trabajará, por lo que la memoria disponible del computador de trabajo debe estar acorde con el tamaño de las bases de datos a utilizar. El ambiente de trabajo se subdivide en cuatro ventanas, además de la barra de menú y

de herramienta, review, variables, command, result. En la ventana básica **review** se detalla el listado de los comandos ejecutados desde que se inició la aplicación; en la ventana **variables** aparecen las variables contenidas en la base de datos cargada; en la ventana **command** se escriben las instrucciones que se le da a la aplicación y por último en la ventana **result** aparecen los resultados de las órdenes ejecutadas, note Figura 4.7. En la esquina inferior izquierda de la pantalla, Figura 4.8, se tiene el directorio de trabajo. Para cambiar de directorio se puede realizar de dos maneras: escribiendo en la sub-ventana command el comando **cd** seguido de la nueva ruta o directorio, por ejemplo **cd c;** o vía ventana ingresando a la opción **Change Working Directory** ubicada en la opción File de la barra menú.

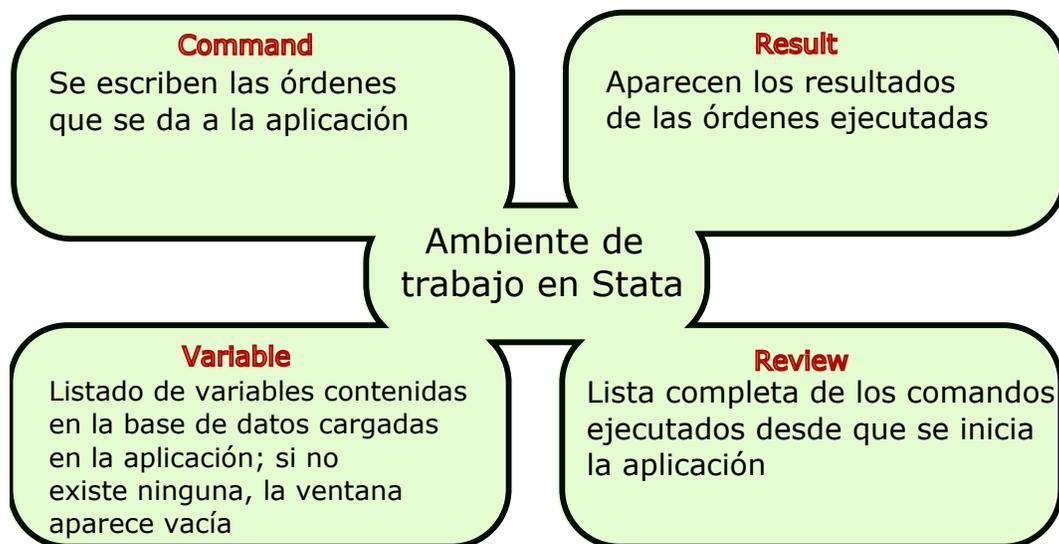


Figura 4.7: Ambiente de trabajo en STATA

En la Sección 4.2.8 se trabajará con el archivo de datos *midas_example.dta*, el cual se obtuvo como parte de una revisión sistemática y un meta-análisis de la literatura publicada sobre el rendimiento de tomografías por emisión de positrones axilar (FDG-PET) en el cáncer de mama para la identificación de la calidad de los estudios primarios, la cuantificación de los resultados globales de clasificación (sensibilidad y especificidad), el poder discriminatorio (odds ratio de diagnóstico) y el valor informativo (ratios de máxima verosimilitud de diagnóstico) y la evaluación de las características técnicas de la prueba y sesgo de publicación en las estimaciones de precisión diagnóstica.

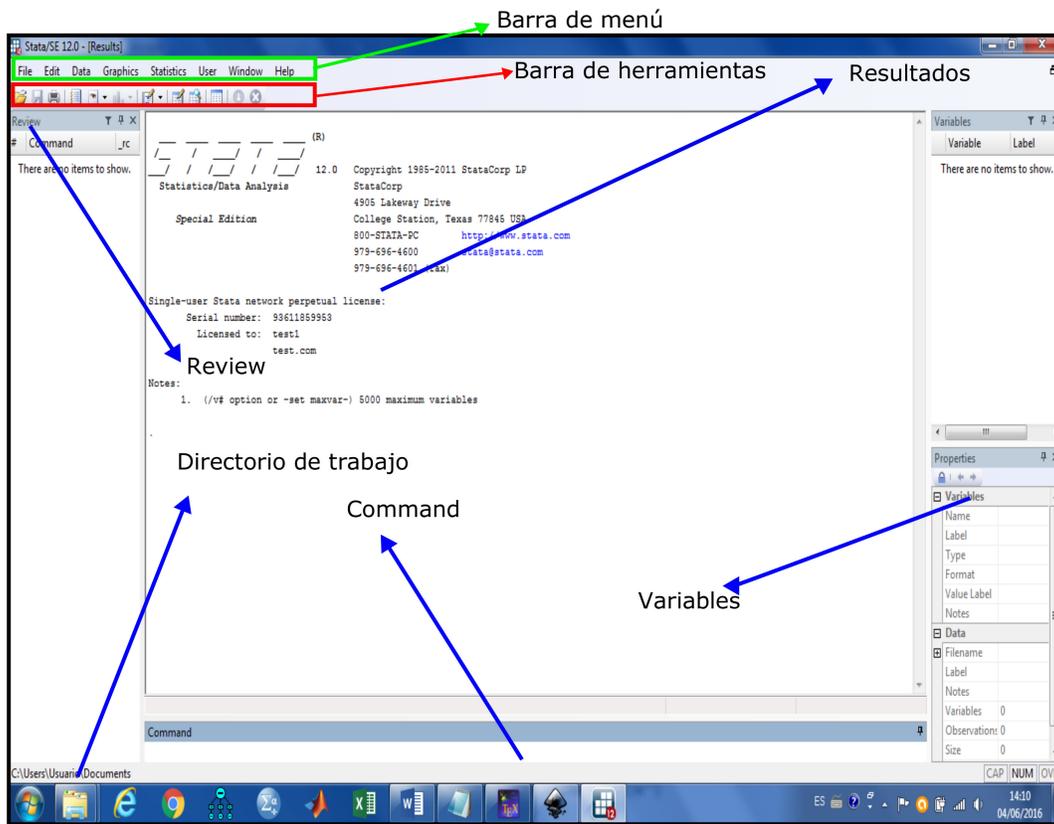


Figura 4.8: Menú y paneles en Ambiente de trabajo STATA

4.2.2 SINTAXIS DE LAS ÓRDENES DE STATA

Stata cuenta con un lenguaje de programación avanzado que respeta normas de sintaxis. Cualquier sentencia debe contener la sintaxis "[**prefix :**] **command** [**varlist**] [=**exp**] [**if**] [**in**] [**weight**] [**using filename**] [, **options**]". Pero la mayoría de las veces se utiliza de la siguiente forma, mucho más simplificada, "**command** [**varlist**] [**if**] [**in**] [**weight**] [, **options**]", donde todo lo que aparece entre corchetes "[...]" es opcional, "**if**" seguida de una expresión lógica que indica que sólo los datos que verifiquen dicha condición serán incluidos en el análisis, **in** sirve para indicar el rango de observaciones que deseamos analizar, **weight** sirve para indicar una variable de ponderación y **options** son las opciones específicas del comando que estemos utilizando. Para el caso de las pruebas de diagnóstico se tendría lo siguiente "**midas varlist** [**min=4 max=4**] [**if**] [**in**] [**weight**] [, **options**]", donde en **varlist** se almacenarán los datos de la tabla de contingencia de 2×2 de los índices y resultados de la prueba de referencia.

El usuario provee los datos de un arreglo rectangular para los elementos a , b , c y d . Cada fila de datos contiene una tabla de datos 2×2 , es decir, los datos de cada estudio. **varlist** debe contener los valores de las variables a , b , c y d en el orden de la Tabla 4.2.

Prueba de referencia			
Test		Positivo	Negativo
		Positivo	a=verdaderos positivos
	Negativo	c=falsos positivos	d=verdaderos negativos

Tabla 4.2: Orden de las variables a,b,c y d

4.2.3 GESTIONES DE BASE DE DATOS

Los archivos de datos en Stata se denominan dataset. Un dataset es una tabla, donde las columnas representan variables y las filas observaciones o casos. Para conocer la naturaleza de los datos contenidos en el dataset se utiliza la instrucción `description`.

4.2.4 TIPOS DE DATOS

Al utilizar el comando `describe`, aparece una variable denominada **storage type** que indica el formato de almacenamiento, es decir, el número de bytes y, por tanto, la precisión de la variable, Tabla 4.3.

Tipos	Menor valor	Mayor valor	Valor cercano a 0	Bytes	Precisión
Byte	-127	100	± 11	1	
Int	-32,767	32,767	± 1	2	
Long	-2,147,483,647	2,147,483,647	± 1	4	
Float	$-1.70141173319 \times 10^{38}$	$1.70141173319 \times 10^{38}$	$\pm 10^{-36}$	4	3.795×10^{-8}
Double	$-8.9884656743 \times 10^{307}$	$8.9884656743 \times 10^{307}$	$\pm 10^{-323}$	8	1.414×10^{-16}

Tabla 4.3: Tipos de datos en Stata

4.2.5 CARGAR UN ARCHIVO DE DATOS EN FORMATO STATA

Para cargar un archivo de datos en formato Stata (extensión *.dta), se tienen dos opciones: realizarlo desde la barra de menú o mediante una línea de código en la ventana command. Desde la barra de menú vea Figura 4.9 y con instrucción en la ventana command mediante la sentencia `use "nombre del archivo"`. Si existe un archivo de datos en memoria, la aplicación no permitirá cargarlo en memoria pues borraría el anterior. Por ello, debería de utilizarse la opción `clear` para liberar de la memoria el anterior archivo de datos `use "nombre del archivo", clear`. La sintaxis de este comando es bastante flexible, permitiendo controlar qué parte del archivo se desea cargar `use [varlist] [if] [in] using filename [, clear nolabel]`.

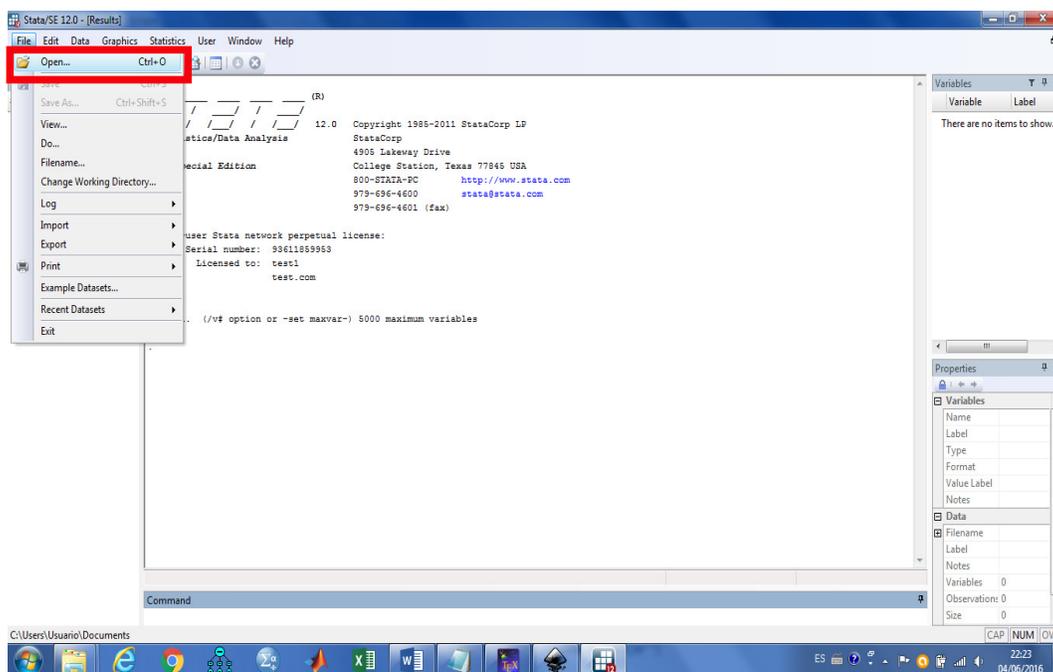


Figura 4.9: Cargar archivo desde barra de Menú en Stata

4.2.6 OPCIONES EN STATA

4.2.6.1 MODELAMIENTO Y OPCIONES DE POST-ESTIMACIÓN

- **nip** (número entero) especifica el número de puntos de integración utilizados para la estimación de máxima verosimilitud basada en una cuadratura gaussiana. Valores altos mejoran la precisión a expensas del tiempo de ejecución. Usando `xtmelogit` con `nip(1)`, el modelo se estima mediante una aproximación Laplaciana. Esto disminuye el tiempo computacional, dando resultados confiables en las estimaciones para modelos de efectos fijos. Pero puede producir estimaciones sesgadas de las componentes de la varianza.
- **ebpred(for|roc)** genera un forest-plot y una curva ROC empírica Bayesiana frente a las estimaciones observadas de sensibilidad y especificidad.
- **modchk(gof|bvn|inf|out|all)** provee una forma gráfica para verificar capacidades; **modchk(gof)** despliega gráficos de cuartiles de la prueba de bondad de ajuste de los residuos; **modchk(bvn)** muestra un gráfico de probabilidades de la chi-cuadrado de las distancias al cuadrado de Mahalanobis para evaluar los supuestos de la normal bivalente; **modchk(inf)** realiza un spikeplot para evaluar observaciones particularmente influyentes mediante la distancia de Cook; **modchk(out)** muestra un gráfico de dispersión para la comprobación de los valores atípicos utilizando efectos aleatorios de los valores pronosticados tipificados y **modchk(all)** provee un gráfico que integra los cuatro gráficos anteriores.

4.2.6.2 OPCIONES DE EVALUACIÓN DE LA CALIDAD

- **qtab(varlis)** crea una tabla que muestra las frecuencias de la calidad metodológica de los ítems.
- **qbar(varlis)** calcula las puntuaciones de la calidad, específica para cada estudio y traza un gráfico de barras de la calidad metodológica. `qtab` se puede combinar con `qbar`, para usar etiquetas en las variables para tablas o gráficos de barras de los ítems metodológicos.

4.2.6.3 GRÁFICOS EXPLORATORIOS

- **bivbox** implementa una generalización bivariante de un gráfico de caja para datos univariantes similar a un gráfico de caja bivariante, Rousseeuw et al. [111]. Es empleado para detectar localización, spread, correlación, asimetría y valores atípicos.
- **chiplot** crea un gráfico `chiplot`, Fisher and Switzer [39] para juzgar si los índices de rendimiento emparejados son independientes aumentando el gráfico de dispersión mediante una pantalla auxiliar.

4.2.6.4 OPCIONES DEL REPORTE PRINCIPAL

- **results(all)** provee un resumen estadístico del desempeño de los índices, de la variación entre estudios, la prueba estadística del ratio de máxima verosimilitud y otros test globales de homogeneidad.
- **results(het)** provee un resumen estadístico, de la variación entre estudios, la prueba estadística del ratio de máxima verosimilitud y otros test globales de homogeneidad.
- **results(sum)** provee un resumen estadístico para el desempeño de los índices, es decir, sensibilidad/especificidad, ratios de máxima verosimilitud positivo/negativo, y resultados de diagnósticos/odds ratios.
- **table(dss|dlor|dlr)** crea una tabla de las estimaciones del desempeño de los estudios específicos con resúmenes estadísticos de las estimaciones de las medidas y resultados de homogeneidad (chi-cuadrado), y el test de inconsistencia I^2 . `dss|dlor|dlr` representan las medidas de desempeño de las parejas de sensibilidad/especificidad, ratios de máxima verosimilitud positivo/negativo, y resultados de diagnósticos/odds ratio.

4.2.6.5 OPCIONES DE FOREST PLOT

- **id(varlis)** provee etiquetas para estudios que permiten hasta 4 variables.
- **bforest(dss|dlr|dlor)** crea un resumen gráfico con estudios específicos (box) y overall (diamante) con sus respectivos intervalos de confianza para cada par de índices usando `graph combine`.
- **uforest(dss|dlr|dlor)** crean resúmenes de gráficos univariantes donde los puntos estimados y los intervalos de confianza pueden extenderse. `fordata` añade estimaciones de los desempeños de los estudios específicos y sus intervalos de confianza al 95%.
- **forstats** añade estadísticos de heterogeneidad siguiendo estimaciones puntuales.

4.2.6.6 OPCIONES DE LA CURVA ROC

- **rocplane** grafica los datos observados en un espacio ROC para la evaluación visual del efecto umbral, una fuente de heterogeneidad única para meta-análisis de diagnóstico. Cuanto mayor sea el valor de corte, mayor será la especificidad y menor la sensibilidad. Esta interdependencia entre estas medidas basada en la variabilidad del umbral puede ser probada a priori utilizando una prueba de correlación de rango como el rho de Spearman. En **midas**, la proporción de variación debido a los efectos de umbral se calcula como el coeficiente de correlación al cuadrado a partir del parámetro de covarianza entre los estudios.
- **sroc(none|pred|conf|both)** grafica los datos observados en forma de puntos, resumen de la sensibilidad y especificidad y curva SROC con o sin las regiones de confianza o predicción.
- **sroc(nnoc|pnoc|cnoc|bnoc)** grafica los datos observados en forma de puntos, resumen de la sensibilidad y especificidad con o sin los contornos de confianza y predicción. No se grafica la curva SROC.

4.2.6.7 OPCIONES DE HETEROGENEIDAD

- **galb(tpr|tnr|dlor|lrp|lrn)** produce el gráfico de Galbraith de las transformaciones estandarizadas de los logit de (TVP, TFP) y el inverso de sus precisiones. Una recta de regresión que pasa por el origen es estimada, junto con una región de 95% de confianza. Los estudios fuera de estas regiones se consideran datos atípicos.
- **regvars(varlist)** permite realizar un análisis univariante de meta-regresión de una o múltiples covariables dicotómicas o continuas, reportando los resultados en una tabla y forest-plot.

4.2.6.8 OPCIONES DEL SESGO DE PUBLICACIÓN

- **pubbias** cuando esta opción se invoca, midas realiza una regresión de los log de los odds ratios con el inverso de la raíz de los tamaños muestrales como una prueba para la asimetría del gráfico en forma de embudo en el meta-análisis de diagnóstico. Un coeficiente diferente de cero es sugerido para la pendiente cuando el sesgo del estudio es no significativo (p-value < 0.10).

4.2.6.9 OPCIONES DE UTILIDAD CLÍNICA

- **fagan(0-0.99)** crea un gráfico mostrando la relación entre la probabilidad a priori especificada por el usuario, el ratio de máxima verosimilitud (combinación de sensibilidad y especificidad) y la probabilidad de prueba a posteriori.
- **pddam(lbp ubp)** produce una línea gráfica de la probabilidad de prueba a posteriori versus la probabilidad a priori, los valores se encuentran entre 0 y 1 usando los resúmenes de los ratios de máxima verosimilitud.
- **lrmatrix** crea un gráfico de dispersión entre los ratios de máxima verosimilitud positivo y negativo. El gráfico se divide en cuadrantes basado en la intensidad o relación de los umbrales para determinar medidas informativas del test.

4.2.6.10 OPCIONES DE MISCELÁNEOS

- **level(#)** especifica los niveles de significancia para las pruebas estadísticas, regiones de confianza, regiones de predicción, e intervalos de confianza.
- **mscale(#)** afecta al tamaño de los puntos estimados en los forest-plots.
- **scheme(string)** permite la elección del esquema gráfico.
- **textscale(#)** permite la elección del tamaño del texto para gráficos, en especial con respecto a las etiquetas de los forest-plots.
- **zcf(#)** define una corrección de continuidad fija para cuando un estudio contiene una celda con cero.

4.2.7 GUARDANDO LOS RESULTADOS

Midas guarda los resultados como escalares en `r()` como:

- **r(mtpr)** Media de la sensibilidad.
- **r(mtprse)** Error estándar de la media de la sensibilidad.
- **r(mtnr)** Media de la especificidad.
- **r(mtnrse)** Error estándar de la media de la especificidad.
- **r(mlrp)** Media del ratio de máxima verosimilitud del resultado de una prueba positiva.
- **r(mlrpse)** Error estándar de la media del ratio de máxima verosimilitud del resultado de una prueba positiva
- **r(mlrn)**: Media del ratio de máxima verosimilitud del resultado de una prueba negativa.
- **r(mlrnse)** Error estándar de la media del ratio de máxima verosimilitud del resultado de una prueba negativa.
- **r(mdor)** Media de los odds ratio de diagnóstico.
- **r(mdorse)** Error estándar de la media de los odds ratio de diagnóstico.
- **r(AUC)** Área bajo la curva SROC.
- **r(AUClo)** Límite inferior del área bajo la curva SROC.
- **r(AUChi)** Límite superior del área bajo la curva SROC.
- **r(rho)** Correlación entre los logits de la sensibilidad y especificidad.
- **r(reffs1)** Varianza de los logit de la sensibilidad.
- **r(reffs1se)** Error estándar de la varianza del logit de la sensibilidad.
- **r(reffs2)** Varianza del logit de la sensibilidad.

- **r(refs2se)** Error estándar de la varianza del logit de la especificidad.
- **r(Islrt)** Índice de inconsistencia global del ratio de la prueba de máxima verosimilitud.
- **r(Islrtlo)** Límite inferior del índice global de inconsistencia.
- **r(Islrthi)** Límite superior del índice global de inconsistencia.

4.2.8 EJEMPLO CON UN CONJUNTO DE DATOS (DATASET)

Para acceder al conjunto de datos *midas_example.dta*, se puede realizar de dos maneras, vía sentencias o por ventanas. Vía sentencia: seleccionar la ruta donde se encuentra el archivo *midas_example.dta* con la sentencia, en nuestro caso `cd "C:\Program Files (x86)\Stata12\ado\base\m"`, Figura 4.10 o vía ventana, Figura 4.11, si se utiliza sentencia se debe cargar el archivo con la sentencia `. use midas_example`, Figura 4.12.

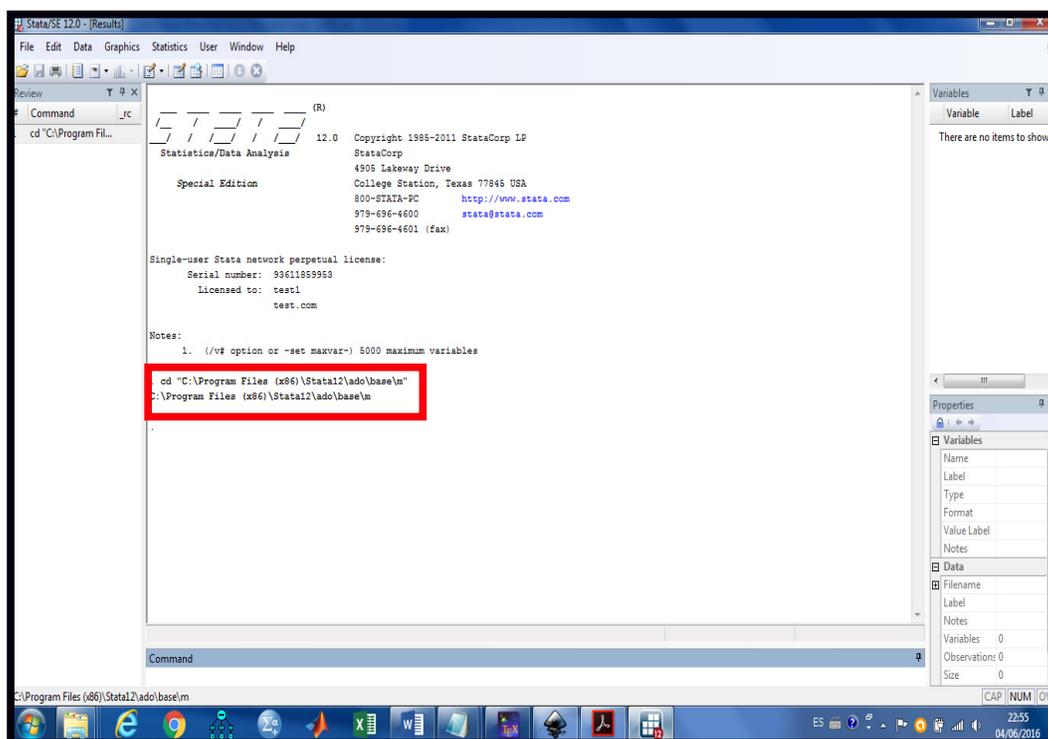
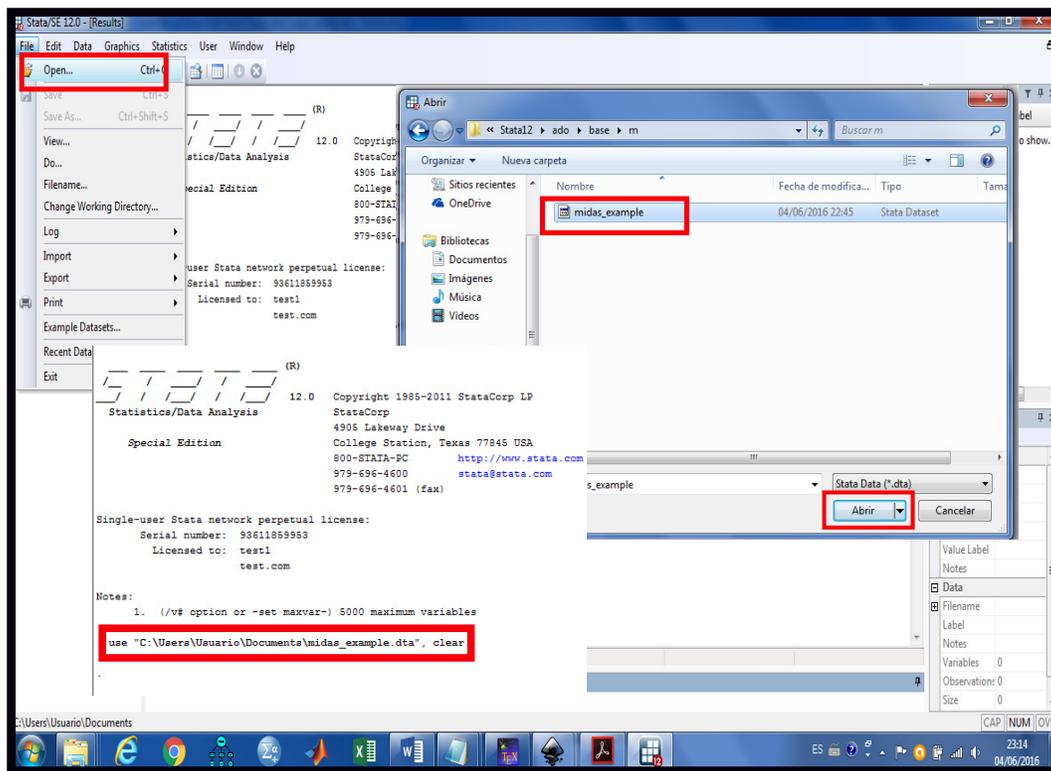


Figura 4.10: Comando para leer en disco duro la ruta del archivo *midas_example*

Figura 4.11: Acceso a *midas_example.dta* desde las pantallas

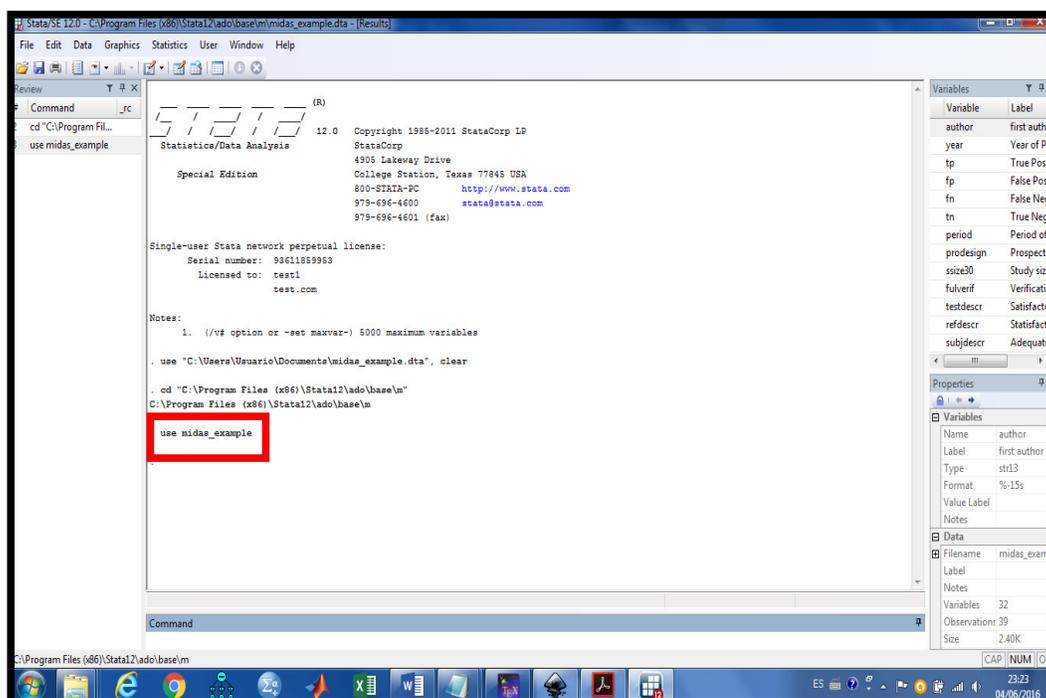


Figura 4.12: Comando para usar los datos de *midas_example.dta*

Para conocer la naturaleza de los datos contenidos en el dataset se utilizan las sentencias, `describe` o la sentencia `list author year pmet sampsize tp fp fn tn in 1\20, sep(1) ab(32) abs noo`.

```

. describe
Contains data from C:\Program Files (x86)\Stata12\ado\base\m\midas_example.dta
obs:          39
vars:         32                2 Feb 2009 14:41
size:        2,457

```

variable name	storage type	display format	value	label	variable label
author	str13	%-15s		first author	
year	int	%8.0g		Year of Publication	
tp	int	%8.0g		True Positive	
fp	byte	%8.0g		False Positive	
fn	float	%9.0g		False Negative	
tn	float	%9.0g		True Negative	
period	byte	%8.0g		Period of publication(before or	
after 1998)					
prodesign	byte	%10.0g		Prospective design	
ssize30	byte	%7.0g		Study size greater than 30	
fulverif	byte	%8.0g		Verification of all by same	
method					
testdescr	byte	%9.0g		Satisfactory description of index	
test					
refdescr	byte	%9.0g		Satisfactory description of ref	
test					
subjdescr	byte	%9.0g		Adequate description of study	
subjects					
report	byte	%7.0g		Satisfactory reporting of results	

consel	byte	%7.0g	Consecutive selection of subjects
brdspect	byte	%10.0g	Broad spectrum of disease
blinded	byte	%7.0g	Blinded interpretation of ref and
index tests			
sampsize	int	%8.0g	Number Data Units Per Study
qscore	byte	%9.0g	Quality score
age	byte	%8.0g	Mean Age
fast	byte	%8.0g	Prior Fasting of at least 4-6
hours			
res	byte	%8.0g	Resolution of PET Camera
dose	int	%8.0g	Appropriate dose used
upttime	byte	%8.0g	Uptake Period specified
acqtime	byte	%8.0g	Duration of image acquisition
testcrit	byte	%8.0g	Test criteria described
multiobs	byte	%8.0g	Multiple Observers
attcor	byte	%9.0g	Attenuation correction of PET
images			
axonly	byte	%8.0g	Dedicated axillary imaging
pmet	float	%9.2g	Study-specific Prevalence of
axillary Metastases			
Analysis	str7	%9s	Unit of Data Analysis(Patient
versus Nodes)			
patient	byte	%10.0g	Mode of analysis

Sorted by:

. list author year pmet sampsize tp fp fn tn in 1/20, sep(1) ab(32) abs noo

author	year	pmet	sampsize	tp	fp	fn	tn
Tse	1992	.7	10	4	0	3	3
Adler1	1993	.47	18	8	0	1	10
Hoh	1993	.64	14	6	0	3	5
Crowe	1994	.5	20	9	0	1	10
Avril	1996	.47	51	19	1	5	26
Bassa	1996	.81	16	10	0	3	3
Scheidhauer	1996	.5	18	9	1	0	8
Utech	1996	.35	124	44	20	0	60
Adler2	1997	.38	50	19	11	0	20
Palmedo	1997	.3	20	5	0	1	14
Noh	1998	.54	24	12	0	1	11
Smith	1998	.42	50	19	1	2	28
Rostom	1999	.65	74	42	0	6	26
Yutani1	1999	.38	26	8	0	2	16
Hubner	2000	.27	22	6	0	0	16
Ohta	2000	.59	32	14	0	5	13
Yutani2	2000	.42	38	8	0	8	22
Greco	2001	.43	167	68	13	4	82

Schirrmeister	2001	.3	113	27	6	7	73
Yang	2001	.33	18	3	0	3	12

4.2.9 MODELOS DE PREDICCIÓN Y DIAGNÓSTICO

Para poder utilizar las sentencias de midas, se debe instalar en STATA con la sentencia `findit midas`, después se procede a dar click a cada uno de los recuadros rojos, Figura 4.13, con lo cual midas queda instalado.

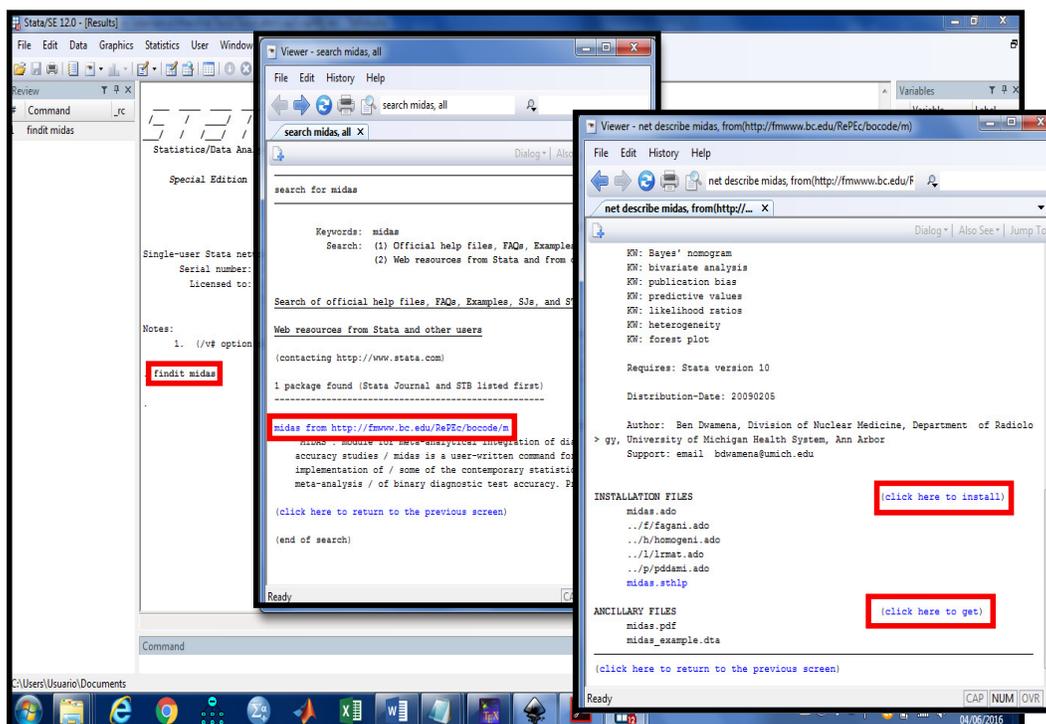


Figura 4.13: Instalación de Paquete midas desde el repositorio oficial

Las predicciones de las estimaciones a posteriori se pueden obtener a partir de modelos estimados, estimaciones de parámetros y estimaciones empíricas de Bayes de efectos aleatorios donde los errores típicos se estiman con el método delta. Los logit de las predicciones pueden ser transformados para obtener predicciones de sensibilidad y especificidad. Midas genera un forest-plot o una curva ROC de una empírica bayesiana versus estimaciones observadas de sensibilidad y especificidad usando `ebpre(for|roc)`, utilizando la sentencia `midas tp fp fn tn, eb(for)`, se obtiene lo anteriormente mencionado, Figura 4.14 .

```
. midas tp fp fn tn, eb(for)
```



Figura 4.14: Bayes empírico predicho frente a los resultados de las pruebas observadas, datos mi-das_example

Los modelos de diagnóstico no suelen ser empleados de manera masiva, debido a que muchos investigadores no tienen la base estadística y otros consideran al meta-análisis como un procedimiento para tratar datos en lugar de un ajuste del modelo Sutton and Higgins [134]. Sin embargo, la aplicación de complejos modelos de máxima verosimilitud para el meta-análisis es importante para evaluar la posible mala especificación del modelo, bondad de ajuste, etc. Midas proporciona un modelo gráfico de evaluación de capacidades, gráfico de cuartiles de residuales basado en la bondad de ajuste, gráficos de probabilidades de la chi-cuadrado de las distancias de Mahalanobis al cuadrado para la evaluación de las hipótesis de la normal bivalente, gráficos spikeplot que permiten detectar observaciones particularmente influyentes utilizando la distancia de Cook y diagramas de dispersión para la comprobación de los valores atípicos utilizando los pronósticos tipificados de efectos aleatorios con la sintaxis `midas tp fp fn tn, modchk(all)`, vea Figura 4.15.

```
. midas tp fp fn tn, modchk(all)
```

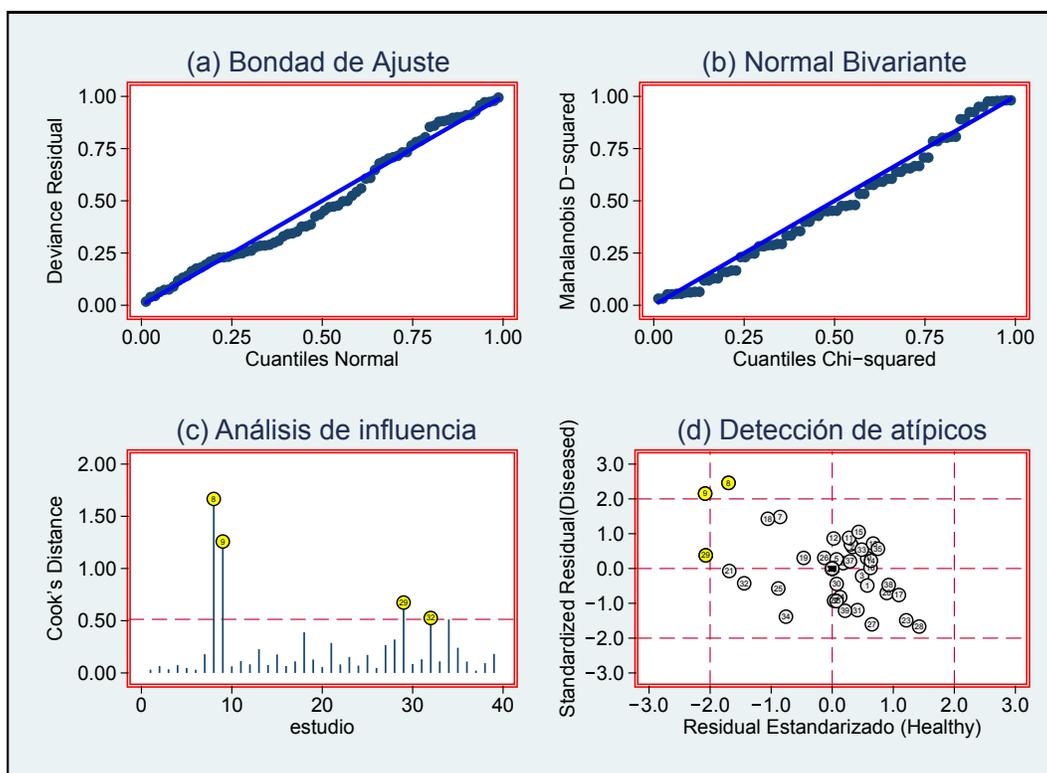


Figura 4.15: Representación gráfica de la normalidad bivalente, bondad de ajuste basada en residuales y los análisis de detección de influencia y atípicos, datos `midas_example`

4.2.10 EVALUACIÓN DE LA CALIDAD

La calidad metodológica de un estudio es el grado en que se pueden mostrar todos los aspectos del diseño y realización de un estudio, para protegerlo contra sesgo sistemático y sesgo no sistemático, que puedan surgir en los estudios que no están bien diseñados y en los errores de inferencia. La herramienta para evaluar la calidad de los estudios de diagnóstico es (QUADAS) Whiting et al. [148, 149, 150, 152]. Es

una herramienta rigurosamente construida y validada que puede ser utilizada por los investigadores para nuevas revisiones sistemáticas. La herramienta QUADAS contiene 14 ítems que cubren el espectro del paciente, patrón de referencia, sesgo de progresión de la enfermedad, verificación y revisión del sesgo, sesgo de revisión clínica, sesgo de incorporación, ejecución de la prueba, retiros de los estudios y los resultados intermedios. Los métodos posibles para hacer frente a las diferencias de calidad son el análisis de sensibilidad, análisis de subgrupos o meta-análisis de regresión. Antes de utilizar QUADAS hay que instalarlo en STATA, con el mismo procedimiento que se empleó para la instalación de midas, es decir, se escribe la sentencia `findit quadas`, Figura 4.16.

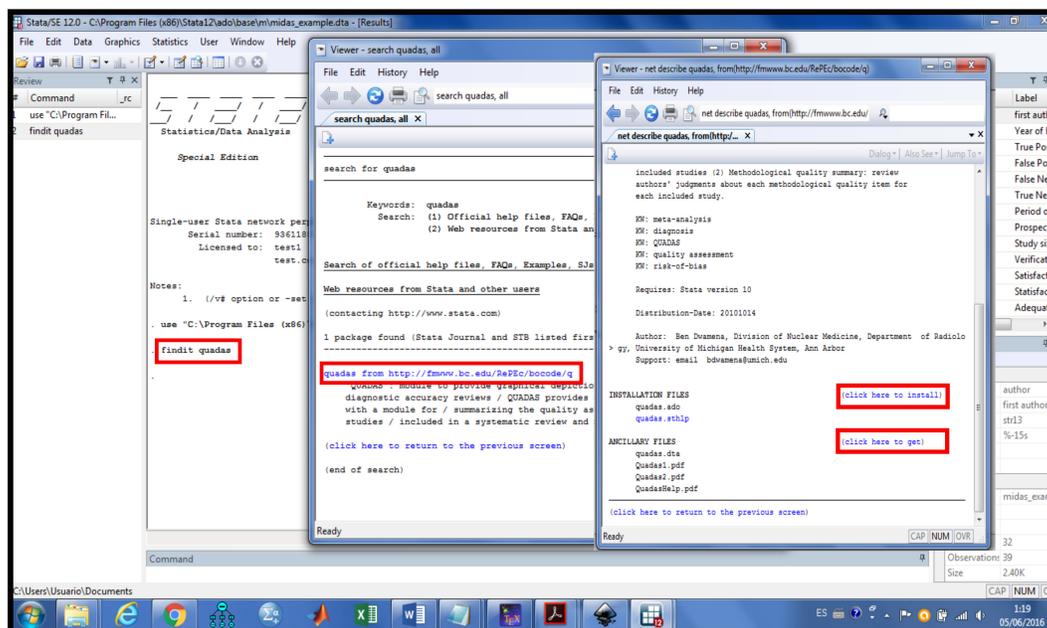


Figura 4.16: Instalación del paquete QUADAS desde el repositorio oficial

La evaluación de la calidad puede resumirse para cada elemento QUADAS como un gráfico de barras, note Figura 4.17, el cual se genera mediante la sentencia:

```

. midas tp fp fn tn, qbar(prodesign ssize30 fulverif testdescr refdescri subjdes
> cr report brdspect blinded) qlab

```

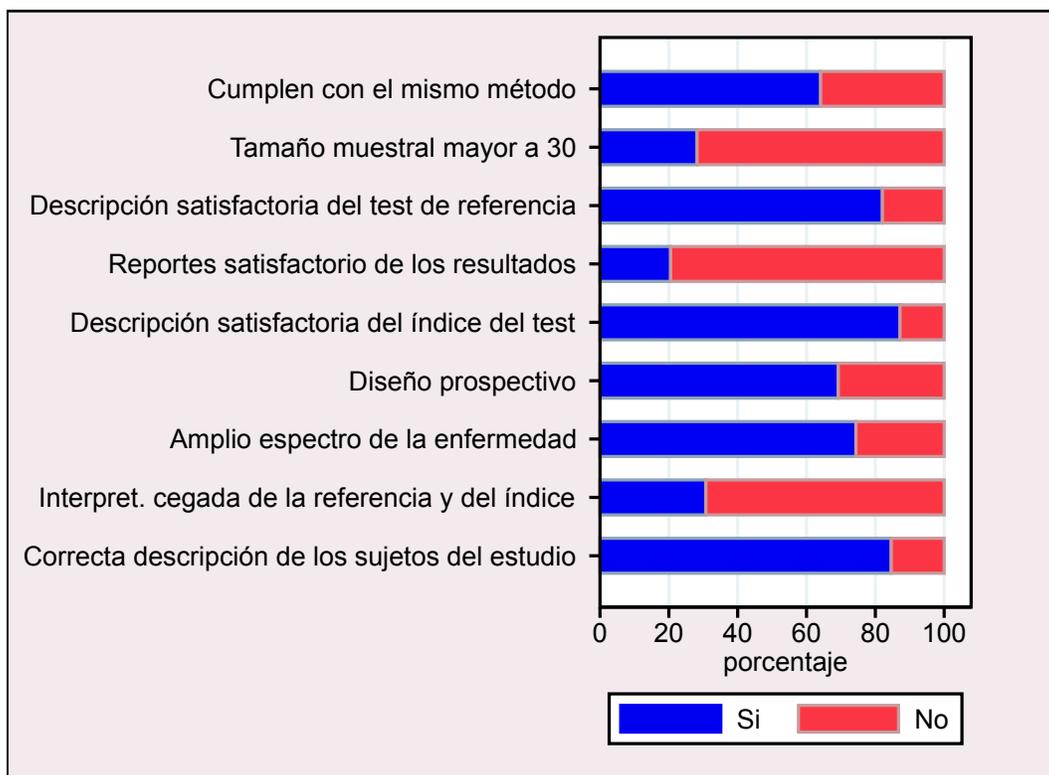


Figura 4.17: Gráfico de barras de la evaluación de la calidad para cada artículo, incluidos los criterios modificados QUADAS, datos midas_example

Como alternativa al uso de `qbar`, existe `qtab` (lista de variables) la cual produce tabla de frecuencias de las puntuaciones de calidad. La tabla se genera mediante la sentencia:

```
. midas tp fp fn tn, qtab(prodesign ssize30 fulverif testdescr refdescr subjdes
> cr report brdspect blinded) qlab
```

METHODOLOGICAL QUALITY ASSESSMENT

Criterion	Yes	Yes(%)	No	No(%)
Prospective design	27	69	12	31
Study size greater than 30	11	28	28	72
Verification of all by same method	25	64	14	36
Satisfactory description of index test	34	87	5	13
Satisfactory description of ref test	32	82	7	18
Adequate description of study subjects	33	85	6	15
Satisfactory reporting of results	8	21	31	79
Broad spectrum of disease	29	74	10	26
Blinded interpretation of ref and index	12	31	27	69

4.2.11 ASOCIACIÓN BIVARIANTE

Midas utiliza un modelo de efectos aleatorios bivalente que modela la sensibilidad y especificidad, por lo que se espera que el desempeño de este par de medidas sean independientes. El diagrama de cajas bivalente describe el grado de interdependencia, incluyendo la ubicación central y cualquier valor atípico. El óvalo interior representa la distribución media de los datos. El óvalo exterior representa la región de confianza al 95%. La figura se obtiene con la sintaxis `midas tp fp fn tn, bivbox`, vea Figura 4.18.

```
. midas tp fp fn tn, bivbox
```

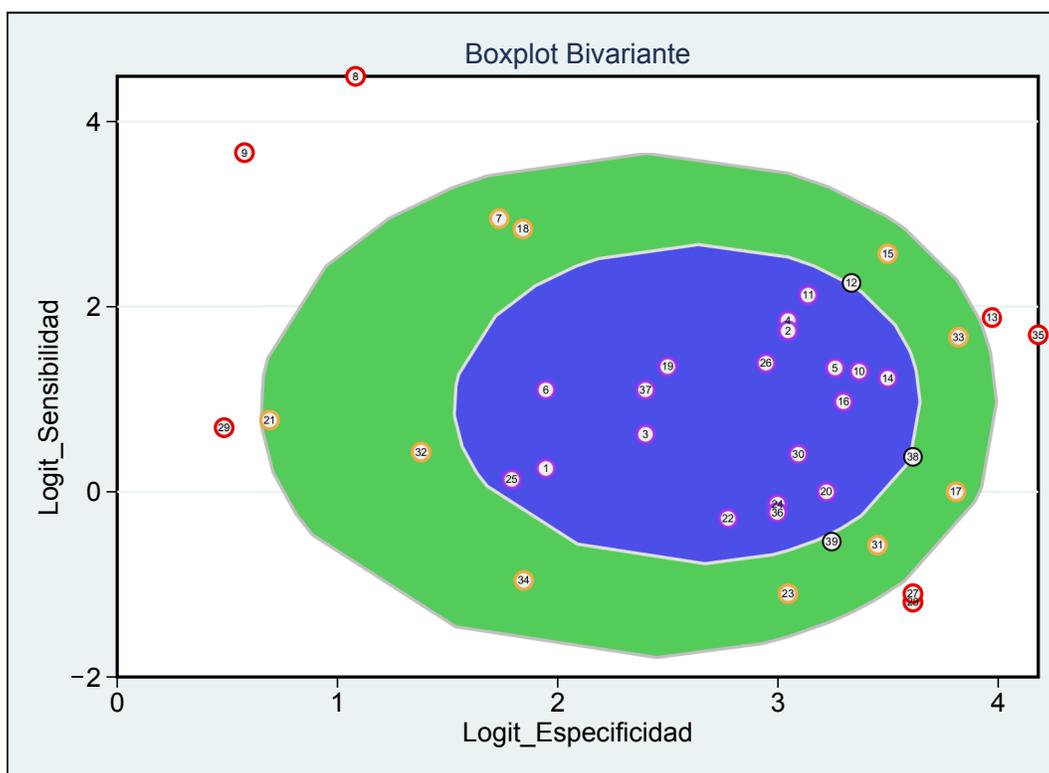


Figura 4.18: Diagrama de caja bivalente de los estudios agrupados con siete valores atípicos que sugiere indirectamente un menor grado de heterogeneidad, datos `midas_example`

El gráfico muestra una asimetría de las medias de rendimiento de la prueba hacia una mayor especificidad con menor sensibilidad, proporcionando evidencia indirecta de cierta variabilidad del umbral.

4.2.12 RESÚMENES DE LAS ESTIMACIONES DE RENDIMIENTO

Los resúmenes de las estimaciones de la sensibilidad y especificidad y sus intervalos de confianza al 95% pueden ser calculados después de la transformación anti-log de la media de los logaritmos de la sensibilidad y especificidad y sus errores estándar. Estos intervalos tienen en cuenta la heterogeneidad entre los estudios más allá del azar (modelo de efectos aleatorios). Esto es posible con la sentencia `midas tp fp fn tn, res(sum) nip(1)`.

```
. midas tp fp fn tn, res(sum) nip(1)

SUMMARY PERFORMANCE ESTIMATES

Parameter          Estimate      95% CI
Sensitivity         0.73 [ 0.63, 0.80]
Specificity         0.96 [ 0.92, 0.97]
Positive Likelihood Ratio 16.2 [ 9.7, 27.3]
Negative Likelihood Ratio 0.29 [ 0.21, 0.39]
Diagnostic Odds Ratio 57 [ 30, 107]
```

4.2.13 HETEROGENEIDAD ESTADÍSTICA

El impacto de la heterogeneidad no observada tradicionalmente se evalúa estadísticamente mediante el I^2 , el cual describe el porcentaje de variación total de los estudios que se debe a la heterogeneidad en lugar del azar Higgins and Thompson [58]. I^2 se puede calcular a partir de la heterogeneidad estadística y los grados de libertad, Higgins and Thompson [58]. Por otra parte, para los modelos mixtos, es el coeficiente de correlación intra-clase expresado como un porcentaje. En midas se implementa por separado para la sensibilidad y especificidad. I^2 puede tomar valores entre 0% y 100%. Un valor de 0% indica ausencia de heterogeneidad y valores por encima de 50% se puede considerar heterogeneidad significativa. La ventaja principal del I^2 radica en que no depende intrínsecamente del número de estudios en el meta-análisis, Higgins and Thompson [58]. La salida de la heterogeneidad en midas, se realiza con la sentencia `midas tp fp fn tn, res(het) nip(1)`, Figura 4.19.

```
. midas tp fp fn tn, res(het) nip(1)

HETEROGENEITY STATISTICS
Heterogeneity (Chi-square): LRT_Q = 146.822, df =2.00, LRT_p =0.000
Inconsistency (I-square): LRT_I2 = 99, 95% CI = [ 98- 99]

Proportion of heterogeneity likely due to threshold effect = 0.11
Interstudy variation in Sensitivity: ICC_SEN = 0.31, 95% CI = [ 0.18- 0.45]
Interstudy variation in Sensitivity: MED_SEN = 0.76, 95% CI = [ 0.70- 0.83]
Interstudy variation in Specificity: ICC_SPE = 0.29, 95% CI = [ 0.12- 0.45]
Interstudy variation in Specificity: MED_SPE = 0.75, 95% CI = [ 0.68- 0.84]
```

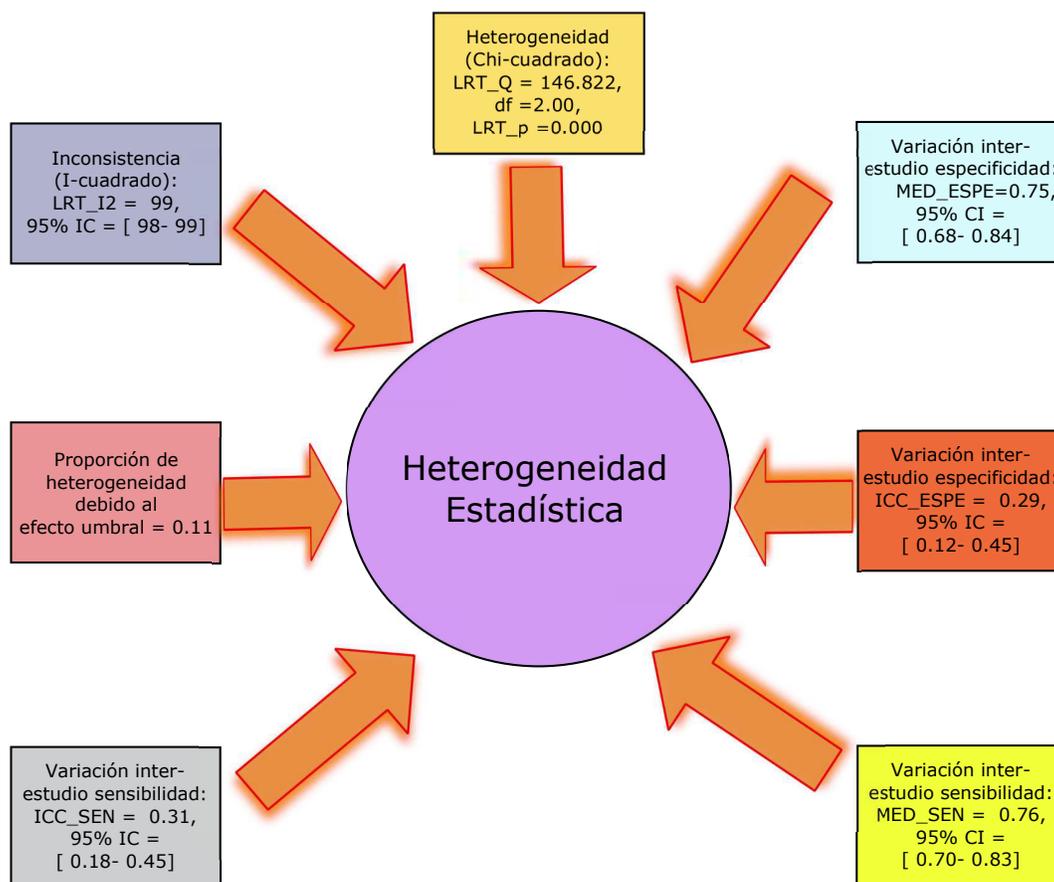


Figura 4.19: Salida de la estadística de heterogeneidad, datos midas_example, elaboración propia

4.2.14 FOREST-PLOT PARA ANALIZAR LA SENSIBILIDAD Y ESPECIFICIDAD

En midas, los forest-plot se pueden crear para cada parámetro de rendimiento de la prueba de manera individual o pueden aparecer como par ordenado de la sensibilidad y especificidad. Lo anterior se puede visualizar con la sentencia `midas tp fp fn tn, texts(0.60) bfor(dss) id(author year) ford fors`, Figura 4.20.

```
. midas tp fp fn tn, texts(0.60) bfor(dss) id(author year) ford fors
```

Es importante notar que los Forest Plot de sensibilidad y especificidad ayudan en la evaluación de la heterogeneidad de los aspectos individuales de la exactitud de la prueba, pero no permiten la evaluación inmediata de si la variación observada es la esperada por la relación de dos variables, es decir, la sensibilidad disminuye a medida que la especificidad se eleva, Phillips et al. [99].

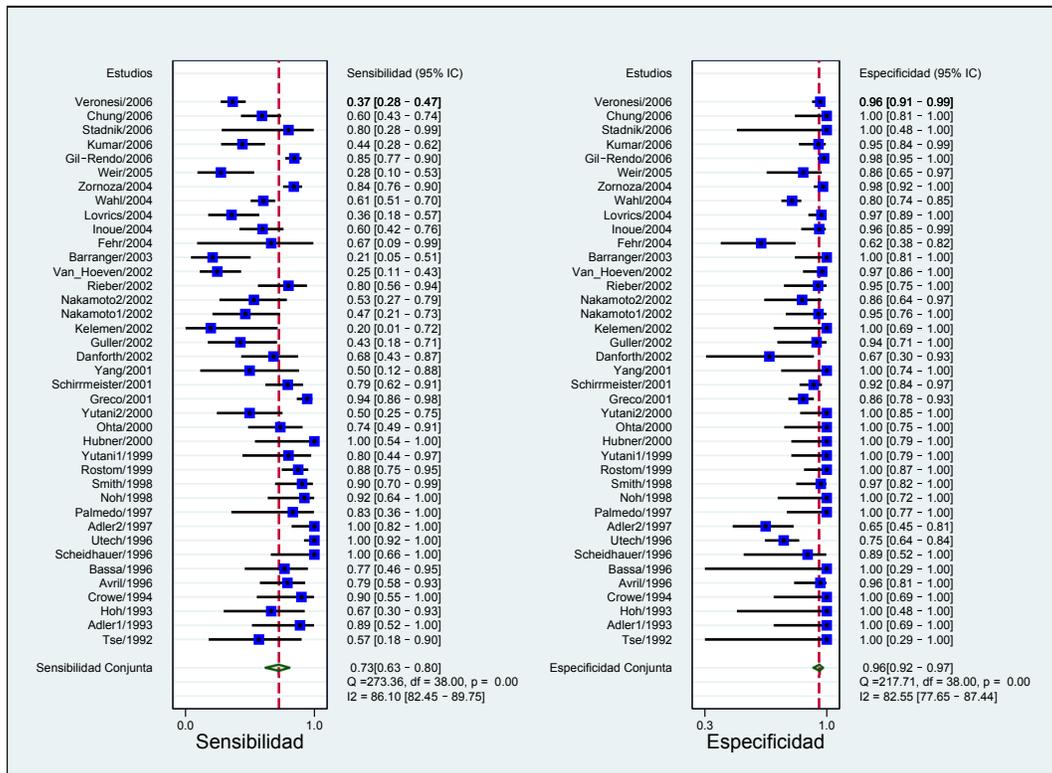


Figura 4.20: Diagrama forest plot para cada estudio específico (eje derecho) y la media de sensibilidad y especificidad con las estadísticas de heterogeneidad correspondientes, datos midas_example

4.2.15 CURVA SROC

Sobre la base de las estimaciones de los parámetros del modelo bivalente, las curva ROC resumen pueden ser expresadas como líneas de regresión que relacionan los logit de la sensibilidad y la especificidad. En midas las estimaciones logit derivadas de la sensibilidad, especificidad y sus varianzas respectivas se utilizan para construir una curva jerárquica SROC. Así se puede construir la curva ROC con la sentencia `midas tp fp fn tn, sroc(both)`, note Figura 4.21.

```
. midas tp fp fn tn, sroc(both)
```

La línea entrecortada alrededor del punto resumen representa la región de confianza al 95%, el área bajo la curva (AUROC), es utilizada como una medida global del desempeño de la prueba. El AUROC es el promedio de TVP sobre la TFP. La Tabla 4.4 se ha sugerido para dar una interpretación a los valores de AUROC, Swets [135].

Intervalos de AUROC	Interpretación
$0.5 \leq \text{AUROC} < 0.7$	Moderado
$0.7 \leq \text{AUROC} < 0.9$	Intermedio
$0.9 \leq \text{AUROC} < 1$	Precisión

Tabla 4.4: Interpretación valores AUROC

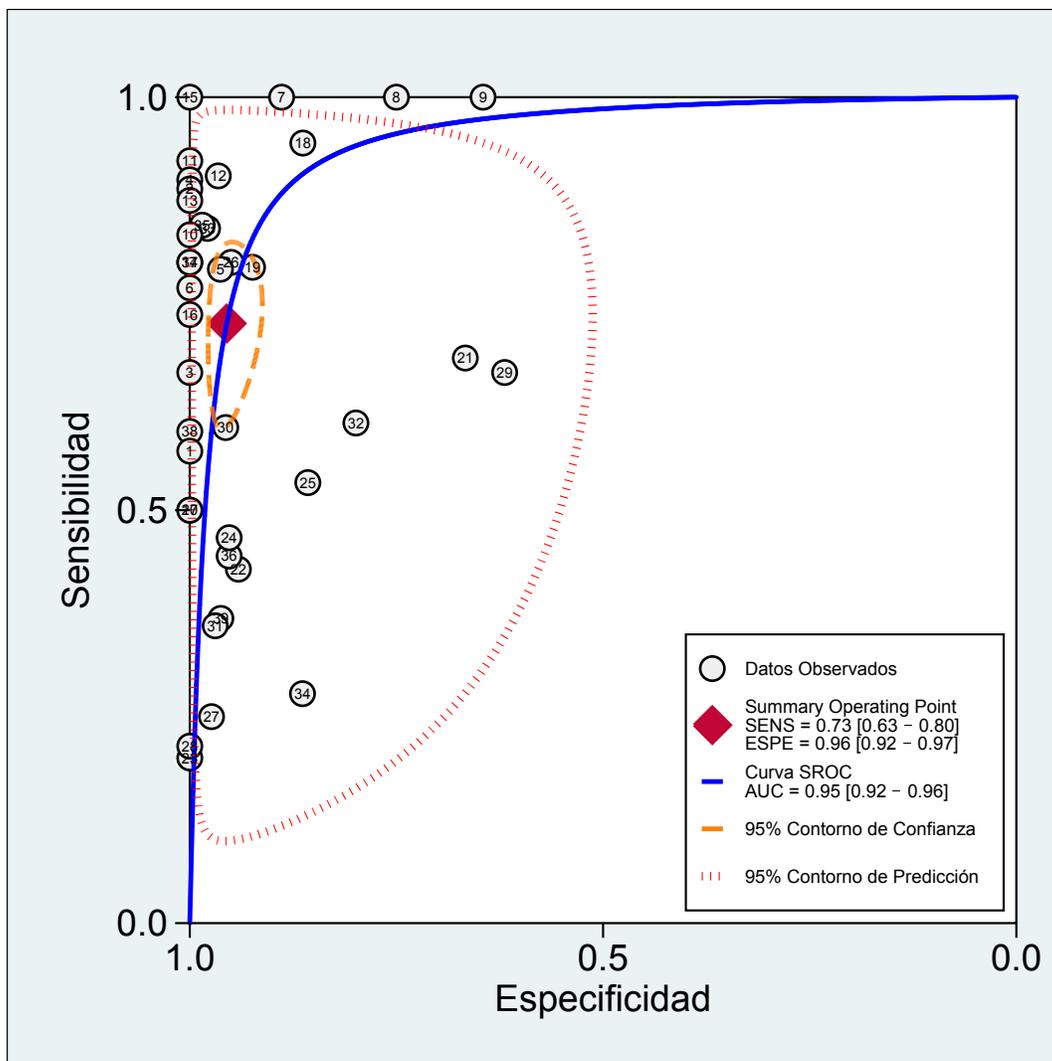


Figura 4.21: Resumen de la curva ROC con las regiones de confianza y predicción alrededor de la media de sensibilidad y el punto de especificidad, datos midas_example

4.2.16 META-REGRESIÓN

Meta-regresión usa los métodos de regresión que incorpora los efectos de la covarianza de los factores sobre las medidas resúmenes de desempeño y ha sido empleada para explorar la heterogeneidad entre estudios terapéuticos. En los estudios de diagnóstico, la heterogeneidad en la sensibilidad y especificidad puede ser resultado de muchas causas relacionadas con las definiciones de las pruebas estándar y referencias, las características operativas del test, métodos de recolección de datos y características del paciente.

Las covarianzas pueden ser introducidas en el modelo como cualquier medida de rendimiento de la prueba. El tamaño de la muestra corresponderá al número de estudio en el análisis, para tamaño de muestra pequeño se limita el poder de la regresión para detectar efectos significativos. Las siguientes sintaxis generan lo antes mencionado, note Figura 4.22:

```
. midas tp fp fn tn, reg(prodesign age qscore sampsize fulverif testdescr refde
> scr clindescr report spectrum blindref blindtest)
variable clindescr not found
in option regvars()
r(111);

. midas tp fp fn tn, reg(prodesign ssize30 fulverif testdescr refdescr subjdesc
> r report brdspect blinded)
```

Estimating Covariate Effect Of: prodesign

Estimating Covariate Effect Of: ssize30

Estimating Covariate Effect Of: fulverif

Estimating Covariate Effect Of: testdescr

Estimating Covariate Effect Of: refdescr

Estimating Covariate Effect Of: subjdescr

Estimating Covariate Effect Of: report

Estimating Covariate Effect Of: brdspect

Estimating Covariate Effect Of: blinded

Sensitivity and Specificity

Parameter	category	nstudies	Sensitivity	p1	Specificity	p2
prodesign	Yes	27	0.73 [0.63 - 0.84]	0.40	0.94 [0.91 -0.97]	0.00
	No	12	0.71 [0.55 - 0.87]	.	0.98 [0.95 -1.00]	.
ssize30	Yes	11	0.74 [0.60 - 0.89]	0.35	0.94 [0.90 -0.99]	0.00
	No	28	0.72 [0.61 - 0.83]	.	0.96 [0.94 -0.99]	.
fulverif	Yes	25	0.73 [0.62 - 0.84]	0.29	0.95 [0.93 -0.98]	0.01
	No	14	0.72 [0.57 - 0.87]	.	0.96 [0.92 -0.99]	.
testdescr	Yes	34	0.72 [0.63 - 0.82]	0.51	0.95 [0.92 -0.98]	0.22
	No	5	0.74 [0.51 - 0.97]	.	0.98 [0.96 -1.00]	.
refdescr	Yes	32	0.72 [0.63 - 0.82]	0.38	0.95 [0.93 -0.98]	0.09
	No	7	0.75 [0.55 - 0.95]	.	0.97 [0.93 -1.00]	.
subjdescr	Yes	33	0.68 [0.59 - 0.78]	0.01	0.95 [0.93 -0.98]	0.35
	No	6	0.90 [0.80 - 1.00]	.	0.98 [0.93 -1.00]	.
report	Yes	8	0.76 [0.58 - 0.93]	0.57	0.92 [0.85 -0.99]	0.00
	No	31	0.72 [0.62 - 0.82]	.	0.96 [0.94 -0.99]	.
brdspect	Yes	29	0.74 [0.65 - 0.84]	0.82	0.95 [0.92 -0.98]	0.02
	No	10	0.66 [0.46 - 0.86]	.	0.97 [0.94 -1.00]	.

blinded	Yes	12	0.54 [0.37 - 0.71]	0.00	0.95 [0.91 -0.99]	0.01
	No	27	0.79 [0.71 - 0.87]	.	0.96 [0.93 -0.99]	.

Joint Model

Parameter	category	LRTChi2	Pvalue	I2	I2lo	I2hi
prodesign	Yes	2.73	0.26	27	0	100
	No
ssize30	Yes	0.67	0.71	0	0	100
	No
fulverif	Yes	0.03	0.99	0	0	100
	No
testdescr	Yes	1.79	0.41	0	0	100
	No
refdescr	Yes	0.61	0.74	0	0	100
	No
subjdescr	Yes	6.75	0.03	70	34	100
	No
report	Yes	2.75	0.25	27	0	100
	No
brdspect	Yes	1.15	0.56	0	0	100
	No
blinded	Yes	8.35	0.02	76	48	100
	No

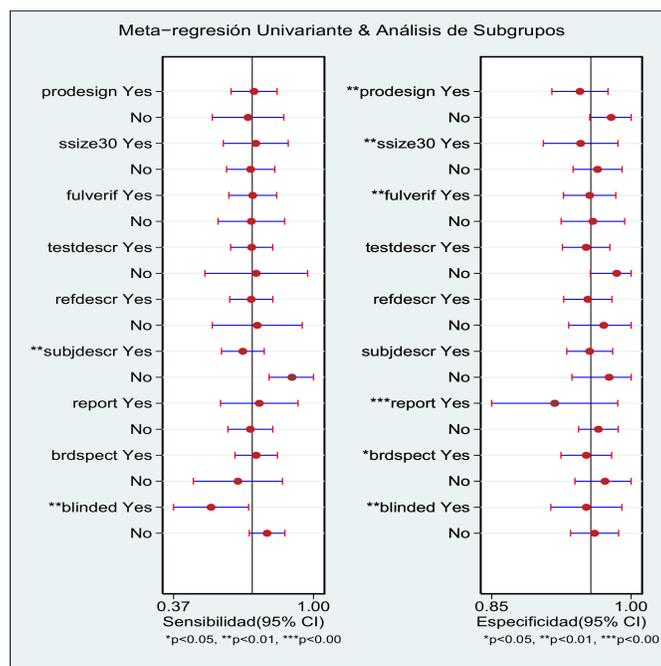


Figura 4.22: Forest plot de una meta-regresión univariante múltiple, datos midas_example

4.2.17 TEST DE REGRESIÓN LINEAL DE ASIMETRÍA FUNNEL PLOT

El gráfico funnel plot es generalmente considerado una buena herramienta de exploración para investigar el sesgo de publicación, Light and Pillemer [75]. Se genera mediante el trazado de la medida del tamaño del efecto versus una medida de la precisión. Si la gráfica es simétrica, se infiere que no existe el sesgo de publicación. Sin embargo, la evaluación por este método puede ser subjetiva. Se han desarrollado métodos de regresión lineal paramétricos y no paramétricos para probar formalmente la simetría del gráfico funnel plot, Begg and Mazumdar [8], Egger and Smith [36], Harbord et al. [55], Macaskill et al. [85], Peters et al. [98], Rücker et al. [112], Schwarzer et al. [123, 124]. El uso de estos métodos para evaluar el sesgo de publicación en los estudios de pruebas diagnósticas puede producir resultados engañosos, Deeks et al. [23], Song et al. [131]. Las pruebas formales para el sesgo de publicación pueden ser llevadas a cabo por una regresión lineal, donde, la variable independiente son los logaritmos de los odds ratio de diagnóstico y la variable dependiente es $\frac{1}{\sqrt{\text{tamaño real de la muestra}}}$, donde un P-valor < 0.10 para la pendiente de la recta de regresión indica asimetría significativa Deeks et al. [23]. Antes de ejecutar la sentencia `midas tp fp fn tn, pubbias`, se debe instalar **mylabels** como en los casos anteriores. La salida también genera un gráfico, note Figura 4.23.

```
. midas tp fp fn tn, pubbias
```

STATISTICAL TESTS FOR SMALL STUDY EFFECTS/PUBLICATION BIAS

yb	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Bias	-3.206976	4.332084	-0.74	0.464	-11.98461	5.57066
Intercept	4.255449	.5420326	7.85	0.000	3.157187	5.353711

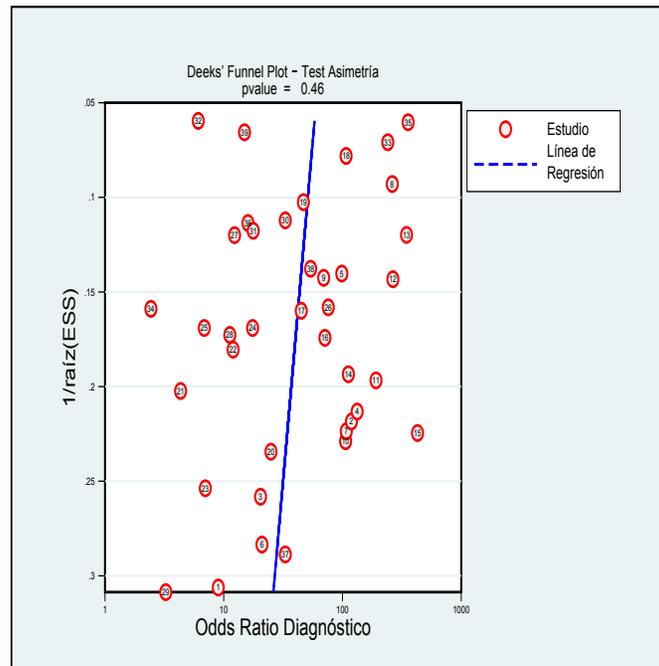


Figura 4.23: Gráfico Funnel Plot con la línea de regresión superpuesta, datos midas_example

El valor p de 0.46 para el coeficiente de la pendiente sugiere la simetría en los datos y una baja probabilidad de sesgo de publicación. Sin embargo se conoce que el test tiene baja potencia, Deeks et al. [23].

4.2.18 FAGAN PLOT (NOMOGRAMA DE BAYES)

La utilidad clínica o relevancia del paciente de una prueba de diagnóstico es evaluada usando los ratios de máxima verosimilitud para calcular la probabilidad post-test (PTP). Basada en el teorema de Bayes, este concepto se muestra en el nomograma de Fagan, Fagan [37]. Cuando el teorema de Bayes se expresa en términos de los logaritmos de los odds ratio, el logaritmo a posteriori es una función lineal de los logaritmos a priori de los odds ratio y de los logaritmos de los ratios de máxima verosimilitud. Un gráfico de Fagan consiste en un eje vertical del lado izquierdo de los log odds ratio a priori, un eje en el centro del gráfico que representa a los log de los odds ratio de máxima verosimilitud y un eje vertical en el lado derecho de la gráfica que denota los log de los odds ratio a posteriori. Las líneas son dibujadas desde el lado izquierdo, pasando por el centro hasta el lado derecho del gráfico, con la sentencia `midas tp fp fn tn, fagan(0.25)`, note Figura 4.24. Al observar la Figura 4.24 podemos notar, que si un paciente obtuvo un test positivo, con el nomograma podemos inferir que tendría la enfermedad con una probabilidad del 84%. Por otro lado, si el test fue negativo, se puede inferir que el paciente no tendrá la enfermedad con una probabilidad del 9%.

```
. midas tp fp fn tn, fagan(0.25)
```

FAGAN'S (BAYESIAN) NOMOGRAM

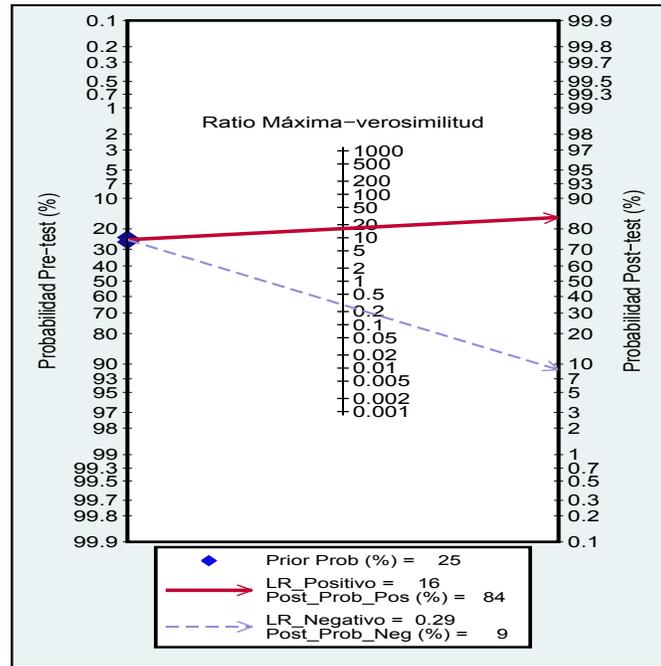


Figura 4.24: Gráfica de Fagan, datos midas_example

4.2.19 DIAGRAMA DE DISPERSIÓN DE LOS ODDS RATIO DE MÁXIMA VEROSIMILITUD

Los resúmenes pueden ser representados de manera gráfica en forma de matriz o por un diagrama de dispersión de los odds ratios de máxima verosimilitud, Stengel et al. [133]. En él se definen cuadrantes sobre la base de los umbrales establecidos:

- Cuadrante superior izquierdo-CSI, ratio de máxima verosimilitud positivo > 10 , ratio de máxima verosimilitud negativo < 0.1 ; Exclusión & Confirmación.
- Cuadrante superior derecho-CSD, ratio de máxima verosimilitud positivo > 10 , ratio de máxima verosimilitud negativo > 0.1 ; Solo confirmación.
- Cuadrante inferior izquierdo-CII, ratio de máxima verosimilitud positivo < 10 , ratio de máxima verosimilitud negativo < 0.1 ; Solo exclusión.
- Cuadrante inferior derecho-CID, ratio de máxima verosimilitud < 10 , ratio de máxima verosimilitud negativo > 0.1 ; No exclusión o confirmación.

Para generar el gráfico se utiliza la sentencia `midas tp fp fn tn, lrm`, note Figura 4.25.

```
. midas tp fp fn tn, lrm
LIKELIHOOD RATIO SCATTERGRAM
```

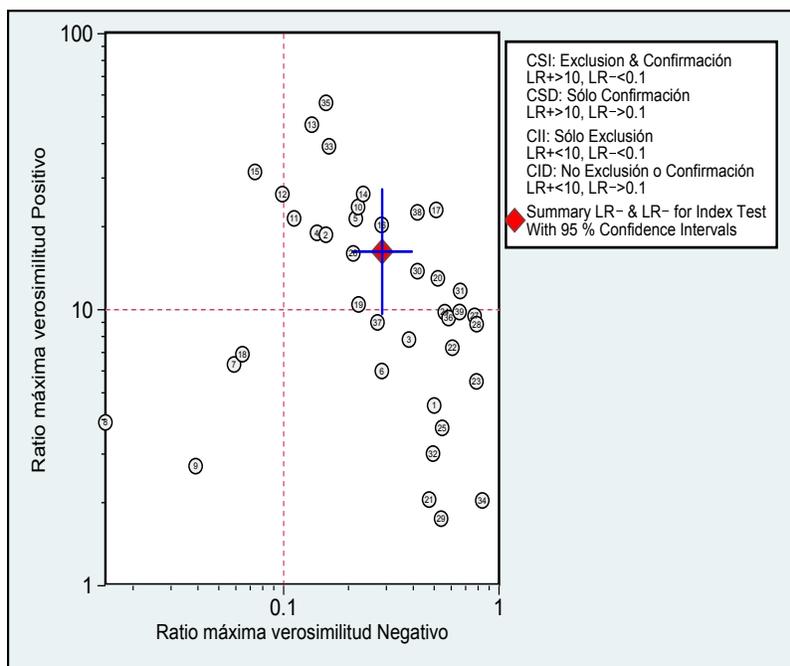


Figura 4.25: Razón de verosimilitud y diagrama de dispersión, datos midas_example

El gráfico de dispersión de los ratio de máxima verosimilitud muestra los puntos de la razón de verosimilitud obtenidos como funciones de la media de la sensibilidad y especificidad, Leeflang et al. [74]. El cuadrante superior derecho sugiere que la prueba es útil para la confirmación de la presencia de la enfermedad (cuando es positiva) y no para su exclusión (cuando es negativa).

4.2.20 VALORES PREDICTIVOS Y GRÁFICO DE PROBABILIDAD MODIFICATORIA

La probabilidad condicional de la enfermedad cuando la prueba es positiva o negativa, los llamados valores predictivos positivos (negativos), son de importancia crítica para la aplicación clínica de un procedimiento de diagnóstico. Estas medidas no dependen solamente de la sensibilidad y especificidad sino también de la prevalencia de la enfermedad. La gráfica de probabilidad modificatoria es un análisis de sensibilidad gráfica del valor predictivo de un proceso continuo que define prevalencia baja para poblaciones de alto riesgo. Se representan curvas separadas para las pruebas positivas y negativas, con la sentencia `midam tp fp fn tn, pddam(0.25 0.75)`, note Figura 4.26.

```
. midam tp fp fn tn, pddam(0.25 0.75)
```

CONDITIONAL PROBABILITY PLOT AND UNCONDITIONAL PREDICTIVE VALUES

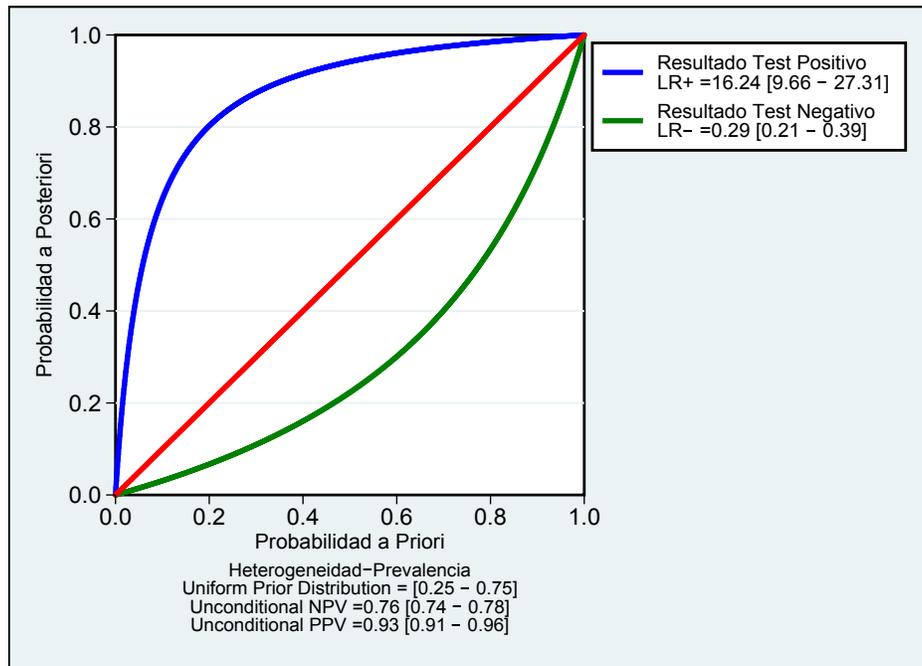


Figura 4.26: Gráfica de Probabilidad modificada, datos midas_example

4.3 PRECISIÓN DIAGNÓSTICA CON HSROC PARA META-ANÁLISIS

4.3.1 INTRODUCCIÓN

Esta sección está destinada a servir como una guía para el uso del paquete HSROC que es empleado en un entorno de software estadístico libre. Por lo tanto, el objetivo principal será describir las funciones en el paquete a través de diferentes ejemplos, sin énfasis en la teoría estadística detrás de ellos. Para profundizar más en la parte matemática, consultar en Dendukuri et al. [27].

El paquete HSROC consta de 4 funciones y 2 conjuntos de datos. Las dos funciones principales son HSROC y HSROCSummary y las 2 restantes funciones son funciones secundarias.

- **HSROC**, el cual debe ejecutarse primero, se utiliza para implementar el algoritmo de Cadenas de Markov Monte Carlo.
- **HSROCSummary** produce el resumen de parámetros para el modelo HSROC.
- **simdata** simula un conjunto de datos basado en el modelo de meta-análisis de diagnóstico sobre HSROC.
- **beta.parameter** devuelve los parámetros de la función de probabilidad Beta (α, β) correspondiente a un rango dado.

En las siguientes secciones se presenta un ejemplo para explicar cómo hacer uso de las funciones dentro del paquete de HSROC. Se basa en una prueba bajo supervisión que es comparada con un ensayo de referencia perfecta cuando ambas pruebas son independientes dado el verdadero estado de la enfermedad.

4.3.2 META-ANÁLISIS EN PRESENCIA DE UNA PRUEBA ORO DE REFERENCIA ESTÁNDAR ASUMIENDO INDEPENDENCIA CONDICIONAL

Primero se empieza instalando el paquete y la librería HSROC en R, mediante las sentencias:

```
> install.packages("HSROC")
> library("HSROC")
```

Empezamos con el modo más simple, donde se asume que la prueba de referencia es la prueba oro (i.e, la sensibilidad y especificidad de la prueba de referencia ambas igual a 100 %). Para este ejemplo, se utilizan los datos de la resonancia magnética (MR) a partir de 10 estudios realizado por Scheidler et al. [120]. Este es un subconjunto del conjunto de datos utilizado en el documento que describe el modelo HSROC de Rutter and Gatsonis [114].

4.3.3 PREPARACIÓN DE LOS DATOS

Los datos de las imágenes de resonancia magnética están incluidas dentro del paquete y pueden ser cargados y leídos con las líneas de comando `data("MRI")` y `MRI`. La salida de las sentencias previas se nota a continuación.

```
> data(MRI)
> MRI

++ +- -+ --
1   9  2  2 44
2   3  6  5 32
3   3  2  1 16
4   3  1 12 44
5   1  1  6 16
6   7  2 22 167
7  12  4  4 29
8  23  5 14 230
9   8  5  5 53
10 16  2  2 22
```

Las columnas ++, +-, -+, --, representan los resultados de las tablas cruzadas entre MRI (prueba de evaluación) y muestras Histológicas/citológicas obtenidas por la cirugía de los ganglios linfáticos, biopsia (prueba de referencia). Con el fin de estimar los parámetros del modelo condicional independiente se utiliza la función HSROC. Para ver los parámetros de la función HSROC se ejecuta la línea de comando `args(HSROC)`, cuya salida es:

```
> args(HSROC)

function (data, iter.num, init = NULL, sub_rs = NULL, first.run = TRUE,
path = getwd(), refresh = 100, prior.Seref = NULL, prior.SPref = NULL,
prior.PI = c(0, 1), prior.LAMBDA = c(-3, 3), prior.THETA = c(-1.5,
1.5), prior.sd.alpha = list(0, 2, "sd"), prior.sd.theta = list(0,
2, "sd"), prior.beta = c(-0.75, 0.75))
NULL
```

La función anterior resulta de extraer una muestra de la distribución a posteriori del modelo de una vía de la distribución de la simulación de Cadena de Markov y Monte Carlo (MCMC). La función contiene una serie de argumentos, los cuales determinan las distribuciones a priori, y tienen por defecto ciertos valores. En particular, los valores por defecto `prior.Seref = NULL` y `prior.SPref = NULL` definen una referencia a la prueba de oro estándar. La simulación de Cadenas Markov y Monte Carlo (MCMC) requiere de valores iniciales que se proporcionarán en este ejemplo. Por lo tanto, lo único que se altera de los valores predeterminados en este ejemplo, es el argumento `init`. Cuando este argumento se deja igual a su valor predeterminado, los valores iniciales requeridos para la simulación de la Cadena de Markov y Monte Carlo (MCMC) son seleccionados al azar por la función `HSROC`, basado en las distribuciones a priori.

El argumento `init` es un objeto tipo lista compuesto por los valores iniciales de los parámetros intra estudio como primer elemento de la lista y los valores iniciales de los parámetros inter estudios como el segundo elemento de la lista. El elemento intra estudio debe ser un objeto tipo matriz donde cada columna corresponde a un parámetro diferente y cada fila corresponde a un diferente estudio, mientras que el elemento inter estudio debe ser un vector con cada elemento correspondiente a un parámetro diferente. Supóngase que se desean los siguientes valores iniciales para cada uno de los parámetros intra estudio α_i (precisión diagnóstica del estudio i), θ_i (el valor de corte para un resultado positivo en el estudio i), S_{1i} (sensibilidad del estudio i para la prueba en evaluación), C_{1i} (especificidad del estudio i para la prueba en evaluación) y π_i (la prevalencia del estudio i). Para lograr la inicialización de los valores antes mencionados se deben realizar las siguientes sentencias:

```
> init.alpha = c(2.51, 2.54, 3.81, 2.41, 2.64, 2.70, 3.31, 3.39, 3.11, 2.99)
> init.theta = c(-0.51, -0.39, 0.33, -2.06, -0.14, -0.08, 1.11, 0.38, -0.86,
+ -0.38)
> init.s1 = rep(0.9,10)
> init.c1 = rep(0.9,10)
> init.pi = c(0.38, 0.17, 0.78, 0.07, 0.74, 0.84, 0.52, 0.95, 0.07, 0.56)
```

Luego se debe crear una matriz de valores iniciales intra-estudio, mediante el comando:

```
> init_within = cbind(init.alpha, init.theta, init.s1, init.c1, init.pi)
```

El orden de los valores iniciales de la función `cbind` es importante y no debe ser alterado. Es decir, la colocación de `init.theta` antes que `init.alpha`, con el comando:

```
> init_within = cbind(init.theta, init.alpha, init.s1, init.c1, init.pi)
```

Esto dará lugar a la inicialización de los parámetros α_i y θ_i . Esto podría dar lugar a una convergencia más lenta del muestreo en el caso que los valores de partida no sean adecuados. Ahora, para los parámetros inter estudios Θ (el valor de la media global para el punto de corte para pruebas positivas), σ_θ (el punto de corte para la desviación estándar inter estudios), Λ (la exactitud de los diagnósticos generales), σ_α (la desviación estándar de la diferencia de medias inter-estudios), y β (el logaritmo del ratio de la desviación estándar de los resultados de las pruebas entre los pacientes con la enfermedad y entre los pacientes sin la enfermedad), se supondrán los siguientes valores de partida:

```

> init.THETA = -0.16
> init.sig.theta = 0.75
> init.LAMBDA = 2.58
> init.sig.alpha = 0.5
> init.beta = 0.25

```

Luego se debe crear el vector de valores iniciales inter estudios, con la sentencia siguiente:

```
> init_between = c(init.THETA, init.sig.theta, init.LAMBDA, init.sig.alpha, init.beta)
```

El orden en el vector anterior debe mantenerse. Por último, simplemente hay que fusionar el intra-estudio y los valores iniciales inter-estudios anteriores en una lista, con la sentencia:

```
> init = list(init_within, init_between)
```

4.3.4 EJECUTANDO LA SIMULACIÓN CADENAS DE MARKOV MONTE CARLO (MCMC)

Para completar "la llamada a la función" necesitamos proveer dos argumentos adicionales (**data** e **iter.num**) los datos y el número de iteraciones, para el conjunto de datos y el número de iteraciones en la simulación de Cadenas de Markov y Monte Carlo (MCMC). En este ejemplo, vamos a ejecutar la Simulación para 50.000 iteraciones, llamando a la función:

```
> HSROC(data=MRI, iter.num=50000, init=init )
```

```
Starting the Gibbs sampler for 50000 iterations ...
```

```
100 iterations completed out of 50000 ...
```

```
200 iterations completed out of 50000 ...
```

```
300 iterations completed out of 50000 ...
```

```
⋮
```

```
49900 iterations completed out of 50000 ...
```

```
The files created during the Gibbs sampler process are in "C:/Trabajo/..."
```

En lugar de manejar grandes matrices de muestras a posteriori para cada parámetro dentro de R, que crecen en tamaño y en número de iteraciones que se incrementan, la función crea archivos de texto en el directorio de trabajo predeterminado, o en cualquier directorio de trabajo especificado por el argumento de ruta. Esto ayudará a aumentar la velocidad de cálculo con el número de iteraciones. Una vez que la función ha alcanzado el número de iteraciones indicadas, aparecerá un mensaje, **The files created during the Gibbs sampler process are in "C:/Trabajo/..."**, donde C: \... es el directorio de trabajo donde se crean y guardan los archivos tipo texto.

4.3.5 INTERPRETANDO LOS ARCHIVOS DE SALIDA

La función `HSROCSummary` se puede utilizar para obtener las estadísticas descriptivas y gráficos del ejemplo anterior. Los argumentos de la función se obtienen mediante la sentencia `args(HSROCSummary)`.

```
> args(HSROCSummary)

function (data, burn_in = 0, iter.keep = NULL, Thin = 1, sub_rs = NULL,
point_estimate = c("median", "mean"), summary.path = getwd(),
chain = getwd(), tv = NULL, digit = 6, print_plot = FALSE,
plot.ind.studies = TRUE, cred_region = TRUE, predict_region = TRUE,
col.pooled.estimate = "red", col.predict.region = "blue",
lty.cred.region = "dotted", lty.predict.region = "dotted",
region_level = 0.95, trunc_low = 0.025, trunc_up = 0.025)
NULL
```

El argumento `data` de la función sirve para referenciar al conjunto de datos. En nuestro caso el argumento `burn_in` indica cuántas iteraciones se deben realizar antes del cálculo de las estimaciones. El argumento `Thin` define la reducción del intervalo. Finalmente `print_plot = TRUE` permite la creación de herramientas gráficas para ayudar al usuario a evaluar si se ha logrado la convergencia de MCMC.

4.3.6 ESTADÍSTICA DESCRIPTIVA

La función `HSROCSummary` devuelve un resumen de los resultados de la interfaz de usuario de R y más detallados en un archivo de texto en el directorio de trabajo. La interfaz de usuario de R enumera las estimaciones puntuales y la densidad más alta a posteriori con su respectivo intervalo de confianza del 95 % para los parámetros inter-estudios e intra-estudios. El archivo de texto se guarda en el directorio de trabajo y se divide en tres secciones.

La primera sección tiene configuraciones y ajustes generales. El número de Cadenas de Markov es de 1 en este ejemplo. Un total de 50.000 iteraciones se utilizaron con un punto de quiebre (para del algoritmo) de 10.000 dentro de cada cadena. Cada valor por segundo es guardado para las estimaciones (la disminución del intervalo). Esto nos deja con una muestra de 20.000 valores para extraer la distribución a posteriori de cada parámetro que interesa. Esta sección también muestra la ubicación del archivo en el disco duro, la fecha de creación del resumen a presentar y la naturaleza de la prueba de referencia (perfecta o imperfecta). El conjunto de datos también está en la lista, vea Figura 4.27.

La siguiente sección de salida proporciona un resumen de las distribuciones anteriores utilizadas en este ejemplo. La distribución a priori utilizada para la prevalencia es una distribución Beta (1,1), esta es equivalente a una uniforme a priori de la prevalencia, permitiendo a todos los valores posibles de ser igualmente probables. El logaritmo de la proporción entre las dos desviaciones estándar, β se supone que sigue una distribución U (-0,75, 0,75). La media general de corte Θ sigue una distribución uniforme desde -1,5 hasta 1,5. La precisión del diagnóstico combinado Λ sigue una

```
Summary for N = 20000
Number of chains = 1
Number of iteration within a chain = 50000
Burn in within each chain = 10000
Thinning interval = 2
Total number of iteration kept = 20000
File Location : c:/Users/Usuario/Documents
Date : 2016-06-06 01:23:10

Perfect reference standard

SAMPLE SIZE

Total ++ +- --
Study 1 57 9 2 2 44
Study 2 46 3 6 5 32
Study 3 22 3 2 1 16
Study 4 60 3 1 12 44
Study 5 24 1 1 6 16
Study 6 138 7 2 22 167
Study 7 49 12 4 4 29
Study 8 272 23 5 14 230
Study 9 71 8 5 5 53
Study 10 42 16 2 2 22
```

Figura 4.27: Lista de configuraciones generales de la simulación

distribución uniforme de -3 a 3. Finalmente, los parámetros σ_α y σ_θ se supone que ambas siguen una distribución $U(0, 2)$, vea Figura 4.28.

Finalmente, en esta sección se enumeran los estadísticos descriptivos de los parámetros del modelo. Adicional a la estimación puntual e intervalos más altos de la densidad a posteriori (HPD), también incluye el error y la desviación estándar de la muestra a posteriori en la simulación de Monte Carlo (MC). El error MC se calcula por el método de medias por lote descrito por Ioannis [63]. Por esta razón, el usuario puede observar que no se entregará ninguna estimación si es inferior a 100 iteraciones, teniendo en cuenta la condición de parada y afinamiento del intervalo, generando un mensaje de error, si se emplea menos de 100 iteraciones, `HSROCSummary(data = MRI, burn_in = 1, Thin = 1, i = 50, print_plot = T)`.

```

PRIOR INFORMATION
-----
Prior of prevalence (pi) is beta ( 1 , 1 ), <=> pi in [ 0 , 1 ]
Prior of beta is Uniform(-0.75 , 0.75 )
Prior of THETA is Uniform( -1.5 , 1.5 )
Prior of LAMBDA is Uniform( -3 , 3 )
Prior of sigma_alpha is uniform( 0 , 2 )
Prior of sigma_theta is uniform( 0 , 2 )
    
```

Figura 4.28: Resumen de las distribuciones utilizadas

```

> Error in HSROCSummary(data = MRI, burn_in = 1, Thin = 1, i = 50, print_plot = T):
You don't have enough iterations to estimate the MC error.
After taking into account the "burn in" and "thinning interval",
you need at least 100 iterations to proceed.
    
```

La suma del error MC (Monte Carlo) y el error estándar proporcionan medios para examinar la precisión de la estimación. Uno podría querer ejecutar la simulación MCMC hasta que el error MC sea suficientemente pequeño. Por ejemplo, un criterio deseable podría ser tener un error MC para cada parámetro menor al 10 % de su desviación estándar posterior. Las estimaciones de los parámetros intra-estudios e inter-estudio se observan en las Figuras 4.29, 4.30, mayor detalle note ANEXO-A.

Por ejemplo, el punto de corte para el parámetro Θ fue estimado en 0.928372 y un intervalo de (0.469583, 1.365921) al 95% de confianza para el valor más alto de densidad-posterior. La desviación estándar y el error MC fue estimado en 0.228954 y 0.011726 respectivamente.

BETWEEN-STUDY parameters (Point estimate = median)					
	Estimate	Standard_Dev	MC_error	C.I._lower	C.I._upper
THETA	0.928372	0.228954	0.011726	0.469583	1.365921
LAMBDA	2.106404	0.340238	0.019787	1.420487	2.752459
beta	-0.415436	0.264169	0.021773	-0.749537	0.120237
sigma.alpha	0.515597	0.290335	0.006616	5.6e-05	1.090918
sigma.theta	0.512763	0.188981	0.00337	0.230685	0.921319
S overall	0.55776	0.102648	0.00121	0.349044	0.752855
C overall	0.947616	0.023182	0.000312	0.899218	0.981592

WITHIN-STUDY PARAMETERS	
theta	

Figura 4.29: Resumen de las distribuciones utilizadas

Vemos que el error MC es menor que el 10% de la desviación estándar posteriori para Θ . Otros cálculos se incluyen como los de Λ , β , σ_θ y σ_α . Las estimaciones de la sensibilidad y la especificidad agrupada, denotan S en general y C general, y se obtienen como funciones de Θ , Λ y β . Un análisis similar de los datos de las imágenes de MR fue realizado por Rutter and Gatsonis [114] utilizando un modelo HSROC con una función de enlace logit. Reportaron una estimación de la sensibilidad agrupada de 0.541 con una intervalo de (0.286, 0.771) al 95% de confianza con igualdad de colas y una estimación de la especificidad agrupada de 0.953 con un intervalo de (0.907, 0.981) al 95% de confianza con igualdad de colas. A pesar de que ambos modelos utilizan una función de enlace diferente, es alentador notar que las estimaciones de S General y C general son satisfactoriamente cercanas a las estimaciones de Rutter y Gatsonis. La diferencia se nota en los intervalos de confianza, dada por el hecho de que la función HSROCSummary retorna un intervalo al 95% de confianza para el valor más alto de densidad posterior. La Figura 4.30 muestra una parte de las estimaciones de los parámetros

inter-estudios. Se nota que las estimaciones para estudios individuales se agrupan por parámetros. Por ejemplo, las estimaciones de $\theta_1, \theta_2, \theta_3, \dots, \theta_{10}$ (punto de corte en el estudio 1, 2, 3, $\dots, 10$), se muestran en la primera sub-sección de los parámetros intra-estudios.

4.3.7 RESUMEN GRÁFICO

La función HSROCSummary crea varios gráficos para que el usuario pueda juzgar si la estadística descriptiva es fiable. Entre ellos, una gráfica de seguimiento para cada parámetro, puede ser utilizada para evaluar la convergencia del muestreo. Cada tramo de la gráfica es un gráfico de dispersión de la muestra posteriori de un solo parámetro versus el número de la iteración de la simulación. Las gráficas de sensibilidad de MR en los estudios 1, 2, 3, 6, 7 y 8 semuestran en la Figura 4.31, para mayor detalle observe ANEXO-B. Para el ejemplo, notamos que la convergencia se logra con bastante rapidez. Otro resumen gráfico producido por la función HSROCSummary es el diagrama de densidad. En él se traza una gráfica suavizada de la función de densidad a posteriori para cada parámetro, Figura 4.32, mayor detalle vea ANEXO-C. La función también produce una curva (SROC). La curva SROC resume la relación entre sensibilidad y $(1 - \text{especificidad})$ en los estudios, teniendo en cuenta la heterogeneidad inter-estudios. La curva SROC para la MR de los datos se muestra en la Figura 4.33, los estudios individuales se representan por un círculo claro. El radio del círculo es proporcional al tamaño de la muestra del estudio. El círculo rojo marca la combinación de la sensibilidad y especificidad a través de los 10 estudios en este meta-análisis. La predicción al 95% de la región se define por el punteado de la curva azul. El trazo y puntos de la curva de color rojo marca el límite de la región creíble 95% para las estimaciones agrupadas de sensibilidad y especificidad a través de los 10 estudios.

WITHIN-STUDY PARAMETERS					
theta					
	Estimate	Standard_Dev	MC_error	C.I._lower	C.I._upper
Study 1	0.717434	0.269566	0.013283	0.200333	1.252638
Study 2	0.761313	0.263249	0.013039	0.24062	1.267222
Study 3	0.628859	0.332931	0.011389	-0.032148	1.273068
Study 4	1.479935	0.276003	0.012005	0.916126	2.002552
Study 5	1.318358	0.34507	0.01275	0.678148	2.033561
Study 6	1.580732	0.213855	0.012518	1.158249	1.998937
Study 7	0.523696	0.252035	0.013539	0.02064	1.000077
Study 8	1.084833	0.191897	0.013538	0.704547	1.453734
Study 9	0.749527	0.23466	0.013352	0.274661	1.19431
Study 10	0.436283	0.279488	0.0128	-0.120885	0.979411
alpha					
	Estimate	Standard_Dev	MC_error	C.I._lower	C.I._upper
Study 1	2.46736	0.461626	0.01854	1.605338	3.413924
Study 2	1.546263	0.507919	0.018999	0.527815	2.492814
Study 3	2.091117	0.502131	0.018207	1.06714	3.070678
Study 4	1.924161	0.498969	0.025052	0.869954	2.825598
Study 5	1.767675	0.578755	0.023734	0.506963	2.73959
Study 6	2.124205	0.432142	0.027618	1.245024	2.956996
Study 7	2.07992	0.395889	0.016659	1.265104	2.815339
Study 8	2.55992	0.364965	0.021331	1.806921	3.225544
Study 9	2.028452	0.392633	0.017885	1.257335	2.793918
Study 10	2.430977	0.445813	0.016628	1.60232	3.348222
Prevalence					
	Estimate	Standard_Dev	MC_error	C.I._lower	C.I._upper
Study 1	0.200152	0.051609	0.000367	0.106957	0.306094
Study 2	0.184025	0.056113	0.00042	0.085939	0.300807
Study 3	0.200346	0.081228	0.000601	0.059669	0.363462
Study 4	0.255629	0.053353	0.000375	0.150532	0.364636
Study 5	0.303807	0.089419	0.000619	0.137763	0.48196
Study 6	0.148772	0.025197	0.00016	0.100658	0.198386
Study 7	0.332415	0.065783	0.000413	0.207833	0.462481
Study 8	0.137648	0.02098	0.00014	0.099626	0.18145
Study 9	0.188975	0.045677	0.000336	0.103775	0.279726
Study 10	0.431199	0.07431	0.000587	0.286243	0.574721

Figura 4.30: Resumen de las estimaciones de los parámetros intra-estudio

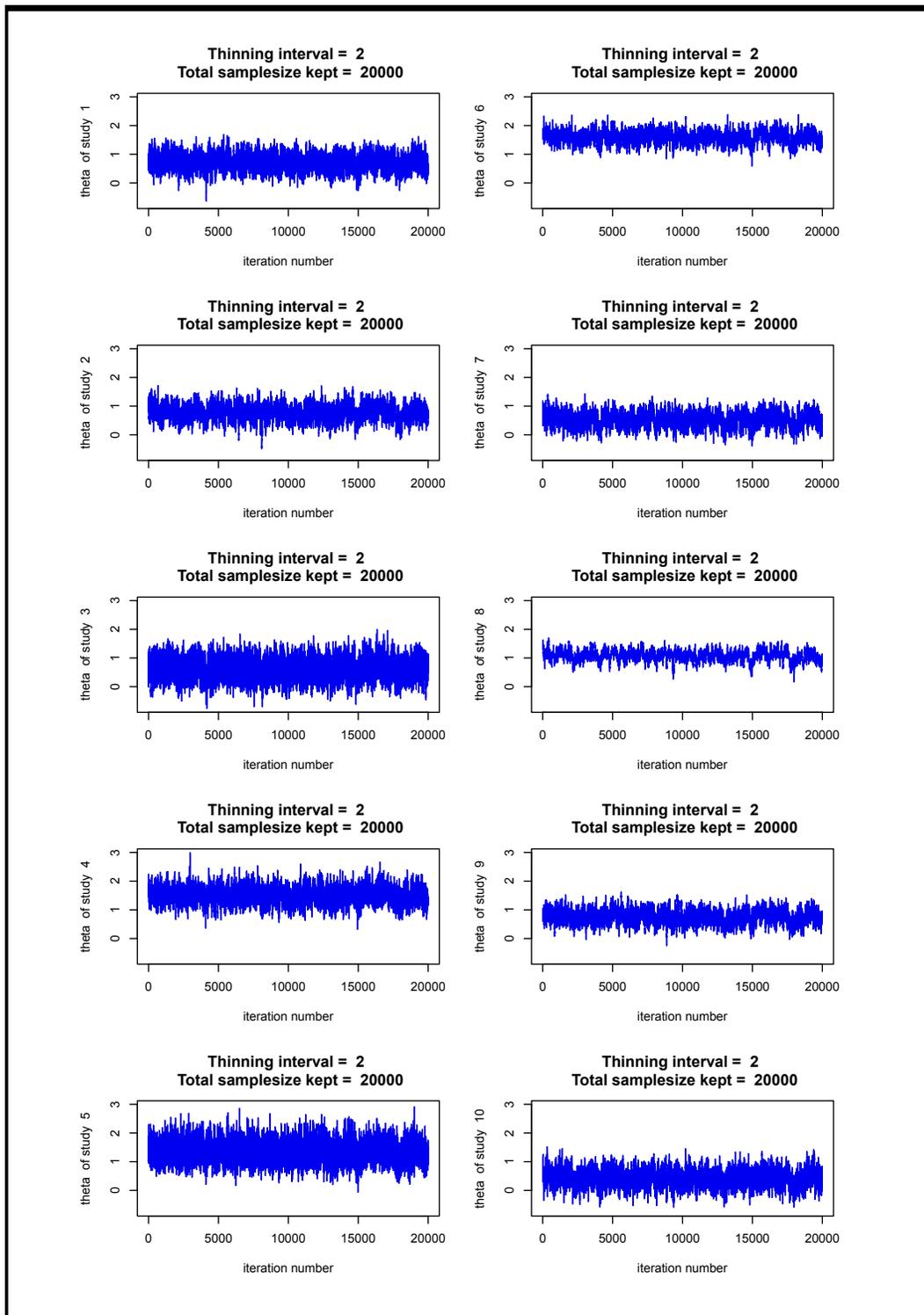


Figura 4.31: Gráfico de dispersión, de la distribución muestral a posteriori de un sólo parámetro, datos MRI

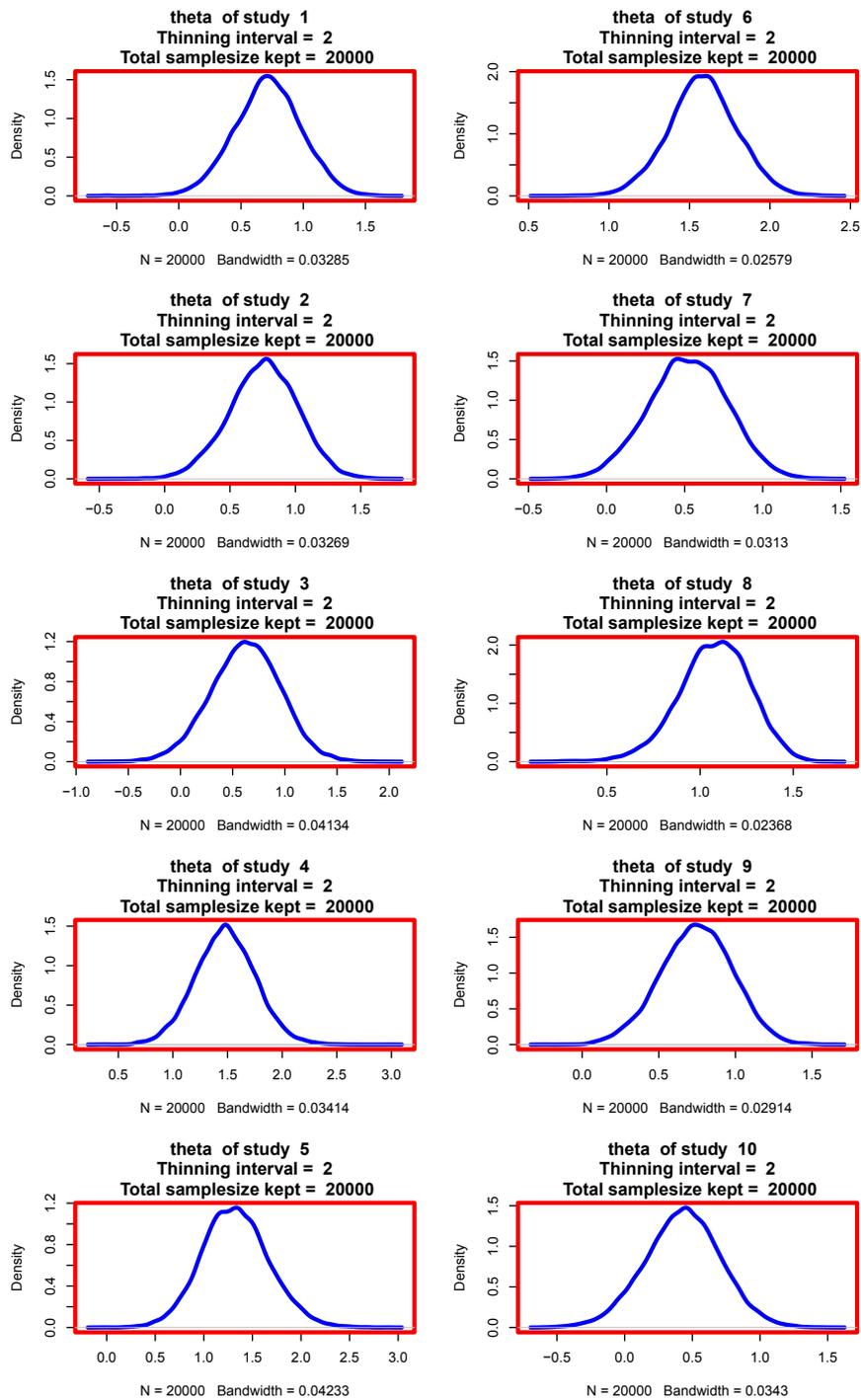


Figura 4.32: Gráficos de densidad de algunos de los parámetros inter-estudios, datos MRI.

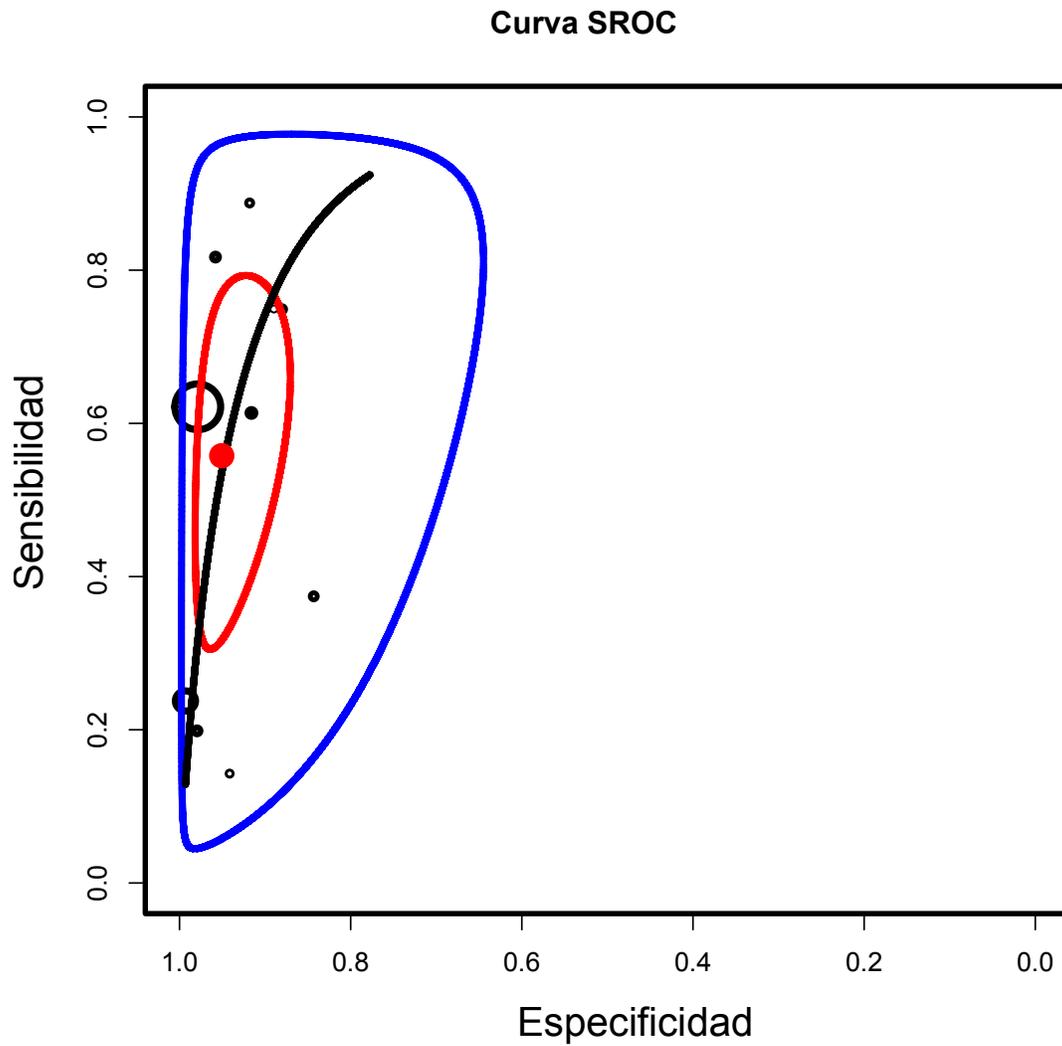


Figura 4.33: Curva SROC, datos MRI

Una interpretación de los resultados obtenidos en los Capítulos 3 y 4 se nota en el Capítulo 5, dando relevancia a la interpretación cualitativa, es decir, se analiza la parte conceptual, metodológica y empírica de los modelos y resultados.

CAPÍTULO 5

DISCUSIÓN

”Creo que en la discusión de los problemas naturales, deberíamos comenzar no con las escrituras, sino con experimentos y demostraciones.”

Galileo Galilei

5.1 ELECCIÓN DE LAS MEDIDAS

Existen varias razones para preferir la sensibilidad y especificidad como las principales medidas de resultados en el meta-análisis de los estudios de diagnóstico de las pruebas que producen como resultados índices dicotómicos. En primer lugar, la gran parte de la precisión de los estudios primarios informan sus resultados como sensibilidad y especificidad Honest and Khan [62]. En segundo lugar, la comprensión de las fuentes de variabilidad y sesgo que afectan a las estimación de la precisión diagnóstico se fundamenta en gran medida en estas dos medidas Bossuyt et al. [11], Whiting et al. [150]. En tercer lugar, la sensibilidad y especificidad son el núcleo de la teoría y enseñanza de las pruebas diagnósticas, y por lo tanto, estas medidas son más familiares a las personas que se desenvuelven en este campo de la medicina. Cuarto, otras medidas alternativas se pueden derivar de estas medidas, tales como los cocientes de máxima verosimilitud y los odds ratio de diagnóstico.

El modelo bivalente analiza directamente estas medidas de la precisión diagnóstica, aunque también reconoce la posibilidad de un efecto umbral implícito. Asimismo, el modelo utiliza un enfoque de efectos aleatorios en las estimaciones de la sensibilidad y especificidad con sus correspondientes intervalos de confianza al 95%. Debido a que el modelo bivalente calcula la intensidad y la forma de la correlación entre la sensibilidad y la especificidad, se puede graficar una elipse de confianza al 95% alrededor de los valores medios o dibujar una elipse de predicción al 95% de confianza para los valores individuales de sensibilidad y especificidad. Esto ayudará a una mejor comprensión de la heterogeneidad de los resultados entre los estudios y la correlación dentro (intra) estudios.

En contraste, el enfoque SROC se centra en los odds ratio de diagnóstico y no produce estimaciones de resúmenes únicas para la sensibilidad y especificidad, lo cual limita enormemente su aplicación clínica. Los valores resumen de la sensibilidad y especificidad en el llamado punto Q^* , son elecciones arbitrarias de valores posibles y pueden no reflejar la verdadera sensibilidad y especificidad que reportan los estudios primarios. La curva SROC describe la interacción dentro de los estudios de la sensibilidad y especificidad, pero en la curva ROC se refleja el efecto de variar el valor o punto de corte de una medición continua. Entender una curva ROC podría sugerir que toda la variación está relacionada con un efecto umbral, pero es solo una posible explicación. Otra dificultad en el enfoque SROC es la elección de los rangos de valores que se ubican por encima de la curva, ya que ésta puede ser engañosa. Los límites de confianza o de predicción de las elipses en el modelo bivalente están bien definidos.

5.2 COMPARACIÓN DE LOS ÍNDICES DE LAS PRUEBAS Y ANÁLISIS DEL EFECTO DE LA COVARIACIÓN

El enfoque SROC utiliza los odds ratio para comparar la exactitud (precisión) diagnóstica entre las pruebas. Esto elimina un posible efecto umbral, pero al mismo tiempo puede enmascarar importantes diferencias clínicas de los rendimientos de las pruebas. En el modelo bivalente se puede mostrar o probar que existen diferencias en la sensibilidad, especificidad o entre ambas. Examinar, cuantificar y explicar las fuentes de variabilidad y sesgo en el meta-análisis es un problema importante, en especial en estudios de diagnóstico, ya que existen muchas diferencias en el diseño, selección de pacientes y en el protocolo de ensayo entre los estudios, Bailey [4], Lijmer et al. [77], Lijmer et al. [78], Normand [94], [130], Thompson and Sharp [138], van Houwelingen and Senn [141], Van Houwelingen et al. [142], Whiting et al. [149]. El efecto de estas diferencias en el diseño se estima con un enfoque SROC cambiando las escalas de los odds ratio de diagnóstico Lijmer et al. [77, 78]. El modelo bivalente permite un análisis de los efectos sobre la sensibilidad y especificidad por separado.

5.3 PROPIEDADES ESTADÍSTICAS

Las propiedades estadísticas del modelo bivalente para la realización de meta-análisis de diagnóstico son sólidas o robustas. Primero, incorporan y estiman la correlación que pueda existir entre las estimaciones de la sensibilidad y especificidad dentro de los estudios. Eso significa que el modelo bivalente produce resultados válidos, si esta correlación es alta, media o está ausente. Esto simplifica el enfoque del meta-análisis como se indica en Devillé et al. [29], Midgette et al. [91], donde se analiza el grado de correlación entre la sensibilidad y especificidad. Si existe una correlación "moderada", el enfoque SROC es defendible. Si la correlación es "pequeña" se separa la sensibilidad y especificidad para sus análisis. El modelo bivalente se ocupa de manera automática de estas dos situaciones. En segundo lugar, el modelo bivalente es una extensión directa de los métodos utilizados para meta-análisis, empleado en los ensayos terapéuticos, Van Houwelingen et al. [142, 143]. El modelo toma en cuenta las diferencias de precisión en la medición de la sensibilidad y especificidad que se realizaron dentro y a través de los estudios e incorpora y estima la cantidad de variabilidad entre los estudios, tanto en la sensibilidad y especificidad (modelo de efectos aleatorios). El enfoque SROC se basa en una regresión lineal de D en S y tiene dos deficiencias estadísticas. En primer lugar, existe un error de medición en la variable dependiente (D) y en la variable independiente (S). A pesar de la existencia de métodos conocidos que tienen en cuenta este error en las variables dependientes e independientes, el enfoque SROC no los emplea. Existe un problema mayor en este enfoque, es que ignora la covarianza que pudiera existir entre D y S, debido a que se definen como la

diferencia y la suma de estas dos variables. La corrección de estos dos errores llevaría hacia un modelo bivalente de regresión de D y S. Por lo tanto el modelo bivalente, puede ser visto como una ampliación mejorada del modelo SROC de Moses et al. [93]. Las deficiencias estadísticas del enfoque SROC se han descrito antes en Deeks [22], Rutter and Gatsonis [113, 114], Shapiro [127]. Han desarrollado un enfoque jerárquico HSROC. Su aplicación en el campo de la medicina ha sido escaso, aunque recientemente una forma menos complicada del modelo HSROC se ha descrito en Macaskill [83]. En contraste, el modelo bivalente por su sencillez y debido al auge de los paquetes estadísticos comerciales, es de fácil aplicación y es por esto que se notan ejemplos que se han publicado Bipat et al. [10], Glas et al. [45], Koelemay et al. [69], Scholten et al. [122].

5.4 ESTUDIOS DE LA PRECISIÓN DE DIAGNÓSTICO: LA CALIDAD DEL DISEÑO Y ELABORACIÓN DE INFORMES

A pesar de la disponibilidad de modelos estadísticos avanzados, la realización de un estudio de meta-análisis parece un poco complicado Irwig et al. [64, 65]. Primero, una revisión sistemática de la precisión de los estudios de diagnóstico y otras revisiones, se ve amenazada por el sesgo de publicación Song et al. [131]. Segundo, muchos informes de estudios de diagnóstico carecen de información sobre elementos claves del diseño y la realización Reid et al. [104]. Sin notaciones completas y exactas no se pueden identificar correctamente las fuentes potenciales de sesgo y variabilidad, Bossuyt et al. [11] dificultando un análisis estadístico de las fuentes de heterogeneidad. En tercer lugar, muchos estudios de la exactitud de diagnóstico tienen importantes deficiencias en el diseño y la realización del mismo, Reid et al. [104], Sheps and Schechter [128].

5.5 R Y STATA

R y Stata son dos paquetes relacionados con la computación estadística. Entre las ventajas que se podría hablar de Stata se tiene que, es fácil de aprender, ofrece amplia gama de análisis estadísticos, tiene un alto grado de flexibilidad, presenta los resultados en formato claro y está bien controlado por StataCorp, que es la compañía que distribuye el software, Thompson [137]. Para el caso de R, es distribuido por el Cran (The Comprehensive R Archive Network), es un paquete estadístico cuyos beneficios son, en R-Studio su interfaz es más amigable y fácil de usar, R es gratis porque se distribuye bajo licencia GNU, por ser de código libre se integra también con otros software libres como L^AT_EX, su capacidad de procesamiento no está limitada por el número de variables, es decir, la capacidad de procesamiento esta dada por los recursos disponibles en cada computador de trabajo, tiene su fortaleza en la automatización y personalización de nuevas funciones por medio de scripts en lenguaje-R, Thompson [137].

R se está convirtiendo en un estándar en la sociedad científica, por hacer figuras de calidad de publicación, además de poder exportarse a diferentes formatos como html, Word, pdf e incluso en diapositivas, Thompson [137].

Por su facilidad de uso, en el futuro Stata podría convertirse en una herramienta para cursos introductorios y de estadística destinados a personas que sólo quieren hacer análisis estadísticos básicos. Una desventaja de Stata es que se vuelve más complicado cuando se trata de programar una tarea no estándar. La tendencia es que más estudiantes, profesores e investigadores continúen con el uso masivo de R en sus

trabajos, Thompson [137].

Finalmente en el capítulo siguiente se resumen los puntos principales del presente trabajo, es decir, se recogen las relaciones encontradas de los elementos estudiados, mayor detalle vease Capítulo 6.

CAPÍTULO 6

CONCLUSIONES

*"La vida es el arte de sacar conclusiones
suficientes a partir de datos insuficientes."*

Samuel Butler

- La revisión bibliográfica sobre meta-análisis de pruebas diagnósticas pone de manifiesto que es frecuente encontrar los métodos de integración utilizados en ensayos clínicos aplicados en este contexto, sin tener en cuenta que en estos análisis no se dispone de una prueba de referencia, los puntos de corte de los estudios pueden no ser los mismos, y la sensibilidad y la especificidad están correlacionadas. Por tanto muchos de los resultados publicados o son incorrectos o al menos no son óptimos.
- En estudios de screening de enfermedades de muy baja prevalencia el DOR puede tomar valores de 500, 1000, 2000, etc., por tanto no es directamente interpretable. Los modelos integran sus valores pero en escala log o en escala logit.
- Los test diagnósticos, para evaluar enfermedades de muy baja prevalencia, no pueden ser valorados desde la perspectiva de la exactitud de la prueba ya que está basada en el número de aciertos y los verdaderos negativos son muy elevados. Por esta razón es necesario trabajar con modelos que tengan en cuenta este hecho.
- En estos estudios, se trabaja con las razones de verosimilitud positiva y negativa ya que pueden expresarse en términos de sensibilidad y especificidad, sin tener en cuenta la probabilidad a priori de la enfermedad.
- Si además se conoce la probabilidad a priori es usual, en este contexto, estimar gráficamente las probabilidades a posteriori (VPP y VPN) con el nomograma de Fagan en lugar de utilizar el teorema de Bayes.
- Los puntos de corte además de ser diferentes, en la mayoría de las publicaciones no aparecen. La literatura es muy rica a la hora de investigar sobre estos puntos, habiéndose encontrado varias decenas de propuestas para estimarlos, relacionadas con la alta/ baja prevalencia.

- La revisión bibliográfica ha puesto de manifiesto que los tres modelos más utilizados en meta-análisis de pruebas diagnósticas son el Modelo de Moses (enfoque SROC) propuesto en 1993, el modelo bivalente de Reitsma de 2005 y el modelo HSROC de Rutter y Gatsonis, propuesto en 2001.
- El modelo de Moses no modeliza directamente los verdaderos positivos y los falsos positivos de cada estudio sino que toma como variable independiente los puntos de corte y estos se calculan a partir de ambos valores y ciertos términos de coste.
- El modelo bivalente asume que los valores de la sensibilidad y especificidad de los estudios individuales (después de la transformación logit) de un meta-análisis siguen una distribución normal bivalente, teniendo en cuenta la posible correlación entre ellas.
- El modelo jerárquico (HSROC) incorpora tanto la variabilidad intra e inter-estudio, y tiene en cuenta la heterogeneidad no explicada entre los estudios a través de la inclusión de covarianzas para el umbral y la precisión de la prueba.
- En el caso de que no se introduzcan correlaciones entre la sensibilidad y especificidad, el modelo bivalente se comporta como un modelo de efectos fijos.
- Cuando las variaciones de la sensibilidad, especificidad, puntos de corte y precisión coinciden, los modelos HSROC y bivalente son equivalentes.
- El paquete *mada* permite modelizar efectos fijos y aleatorios, generando salidas resúmenes de la sensibilidad, especificidad, Q^* , DOR , AUC , razones de verosimilitudes, regiones de confianzas y de predicción y la curva SROC.
- El paquete *midas* modeliza el enfoque bivalente, creando salidas resúmenes de la sensibilidad, especificidad, razones de verosimilitud, correlaciones, medidas consenso, validación del modelo, curva SROC, AUC , Q^* , análisis de heterogeneidad y nomogramas.
- El paquete HSROC mediante simulaciones de cadena de markov y monte carlo (MCMC) genera resúmenes gráficos de las distribuciones a priori de los parámetros del modelo jerárquico, curva SROC, error estándar de estimación, medidas de consenso, sensibilidad, especificidad.

ANEXO

ANEXO A

ESTADÍSTICA DESCRIPTIVA DE LOS PARÁMETROS A ESTIMAR

Summary for N = 20000

Number of chains = 1
 Number of iteration within a chain = 50000 Burn in within each chain = 10000
 Thinning interval = 2
 Total number of iteration kept = 20000

File location : C:/Users/Usuario/Documents

Date : 2016-06-06 01:23:10

Perfect reference standard

SAMPLE SIZE

	Total	++	+-	-+	--
Study 1	57	9	2	2	44
Study 2	46	3	6	5	32
Study 3	22	3	2	1	16
Study 4	60	3	1	12	44
Study 5	24	1	1	6	16
Study 6	198	7	2	22	167
Study 7	49	12	4	4	29
Study 8	272	23	5	14	230
Study 9	71	8	5	5	53
Study 10	42	16	2	2	22

PRIOR INFORMATION

Prior of prevalence (pi) is beta (1 , 1), <=> pi in [0 , 1]

Prior of beta is Uniform(-0.75 , 0.75)
 Prior of THETA is Uniform(-1.5 , 1.5)
 Prior of LAMBDA is Uniform(-3 , 3)
 Prior of sigma_alpha is uniform(0 , 2)
 Prior of sigma_theta is uniform(0 , 2)

BETWEEN_STUDY parameters (Point estimate = median)

	Estimate	Standard_Dev	MC_error	C.I._lower	C.I._upper
THETA	0.928372	0.228954	0.011726	0.469583	1.365921
LAMBDA	2.106404	0.340238	0.019787	1.420487	2.752459
beta	-0.415436	0.264169	0.021773	-0.749537	0.120237
sigma.alpha	0.515597	0.290335	0.006616	5.6e-05	1.090918
sigma.theta	0.512763	0.188981	0.00337	0.230685	0.921319
S overall		0.55776	0.102648	0.00121	0.349044
C overall		0.947616	0.023182	0.000312	0.899218

WITHIN-STUDY PARAMETERS

theta

Summary for N = 20000

	Estimate	Standard_Dev	MC_error	C.I._lower	C.I._upper
Study 1	0.717434	0.269566	0.013283	0.200333	1.252638
Study 2	0.761313	0.263249	0.013039	0.24062	1.267722
Study 3	0.628859	0.332931	0.011389	-0.032148	1.273068
Study 4	1.479935	0.276003	0.012005	0.916126	2.002552
Study 5	1.318358	0.34507	0.01275	0.678148	2.033561
Study 6	1.580732	0.213855	0.012518	1.158249	1.998937
Study 7	0.523696	0.252035	0.013539	0.02064	1.000077
Study 8	1.084833	0.191897	0.013538	0.704547	1.453734
Study 9	0.749527	0.23466	0.013352	0.274661	1.19431
Study 10	0.436283	0.279488	0.0128	-0.120885	0.979411

alpha

	Estimate	Standard_Dev	MC_error	C.I._lower	C.I._upper
Study 1	2.46736	0.461626	0.01854	1.605338	3.413924
Study 2	1.546263	0.507919	0.018999	0.527815	2.492814
Study 3	2.091117	0.502131	0.018207	1.06714	3.070678
Study 4	1.924161	0.498969	0.025052	0.869954	2.825598
Study 5	1.767675	0.578755	0.023734	0.506963	2.73959
Study 6	2.124205	0.432142	0.027618	1.245024	2.956996
Study 7	2.07992	0.395889	0.016659	1.265104	2.815339
Study 8	2.55992	0.364965	0.021331	1.806921	3.225544
Study 9	2.028452	0.392633	0.017885	1.257335	2.793918
Study 10	2.430977	0.445813	0.016628	1.60232	3.348222

Prevalence

	Estimate	Standard_Dev	MC_error	C.I._lower	C.I._upper
Study 1	0.200152	0.051609	0.000367	0.106957	0.306094
Study 2	0.184025	0.056113	0.00042	0.085939	0.300807
Study 3	0.200346	0.081228	0.000601	0.059669	0.363462
Study 4	0.255629	0.055353	0.000375	0.150532	0.364656
Study 5	0.303807	0.089419	0.000619	0.137763	0.48196
Study 6	0.148772	0.025197	0.00016	0.100658	0.198386
Study 7	0.332415	0.065783	0.000413	0.207833	0.462481
Study 8	0.137648	0.02098	0.00014	0.099626	0.18145
Study 9	0.188975	0.045677	0.000336	0.103775	0.279726
Study 10	0.431199	0.07431	0.000587	0.286243	0.574721

sensitivity of test 1 (S1)

	Estimate	Standard_Dev	MC_error	C.I._lower	C.I._upper
Study 1	0.73265	0.114159	0.001984	0.508552	0.939193
Study 2	0.503716	0.15282	0.004078	0.207627	0.783176
Study 3	0.689266	0.154176	0.002652	0.374697	0.952972
Study 4	0.263379	0.102553	0.002416	0.08324	0.472445
Study 5	0.290994	0.141337	0.002812	0.041681	0.566245
Study 6	0.264194	0.075501	0.002041	0.129942	0.421196
Study 7	0.73125	0.097905	0.002059	0.534789	0.905117
Study 8	0.591832	0.076565	0.001938	0.442881	0.741026
Study 9	0.621113	0.113705	0.002667	0.397975	0.835105
Study 10	0.826643	0.083212	0.002049	0.656092	0.96268

specificity of test 1 (C1)

	Estimate	Standard_Dev	MC_error	C.I._lower	C.I._upper
Study 1	0.944918	0.029476	0.000613	0.883059	0.989993
Study 2	0.892726	0.04737	0.001147	0.794837	0.968542

Summary for N = 20000						
Study	3	0.91476	0.052183	0.000844	0.800297	0.986408
Study	4	0.977134	0.019736	0.000422	0.934283	0.998795
Study	5	0.963833	0.036067	0.00067	0.882996	0.999173
Study	6	0.984777	0.00881	0.00023	0.965322	0.997326
Study	7	0.899674	0.044179	0.001189	0.805666	0.969912
Study	8	0.973305	0.010456	0.000292	0.951125	0.99041
Study	9	0.925114	0.030508	0.000709	0.858719	0.972403
Study	10	0.911237	0.047686	0.00113	0.811584	0.984596

Posterior predictive value of Sensitivity of test under evaluation (S1)

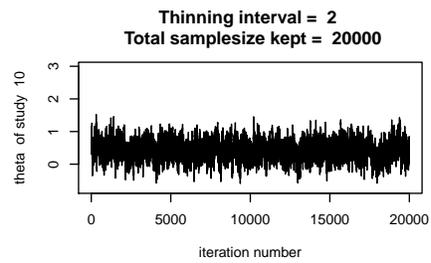
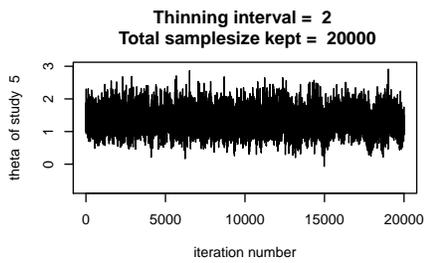
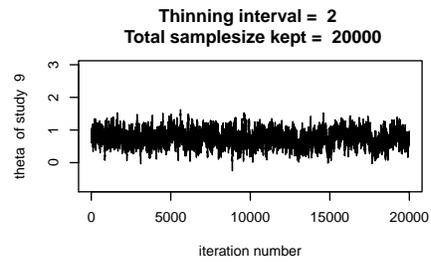
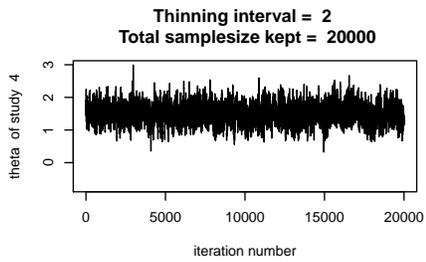
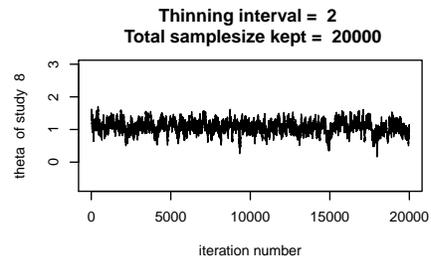
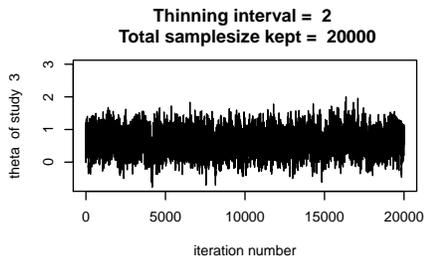
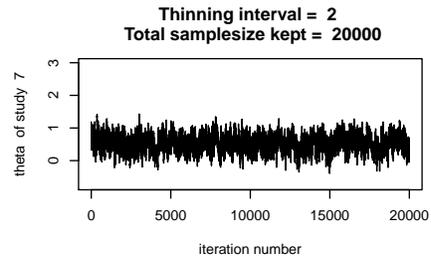
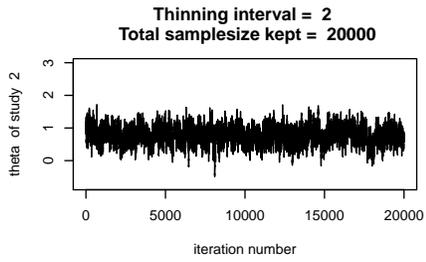
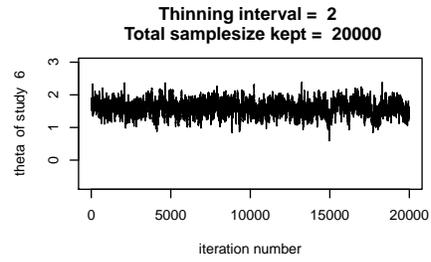
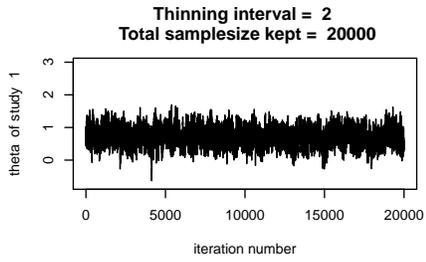
	Estimate	Standard_Dev	MC_error	C.I._lower	C.I._upper
Sensitivity	0.549645	0.223674	0.001684	0.126226	0.971096

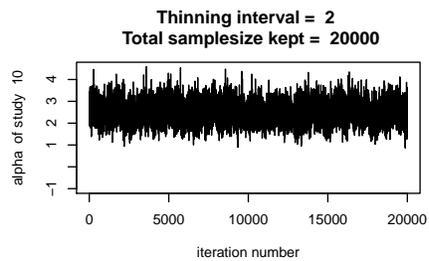
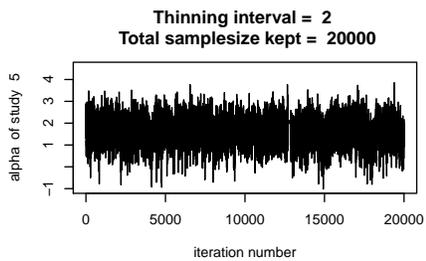
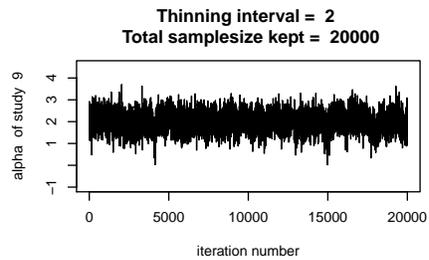
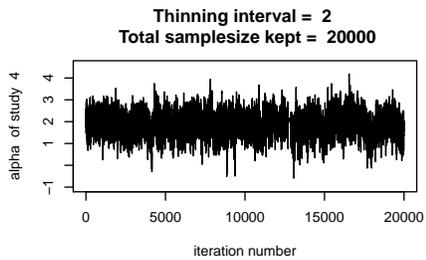
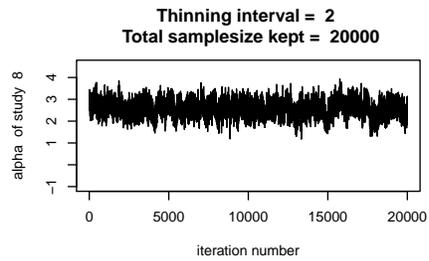
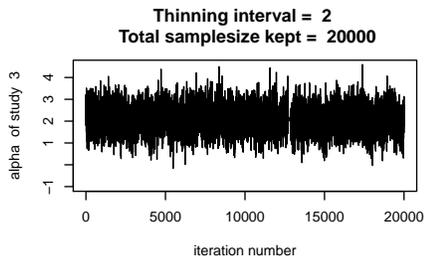
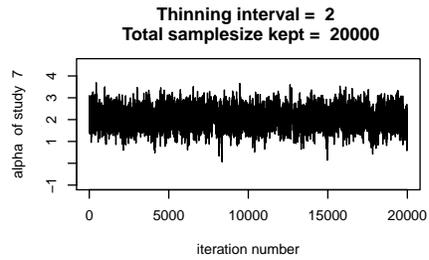
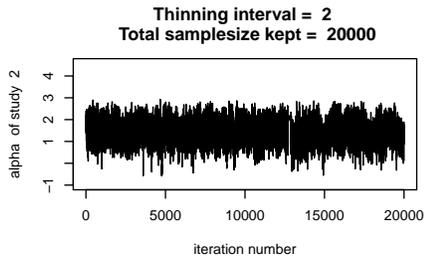
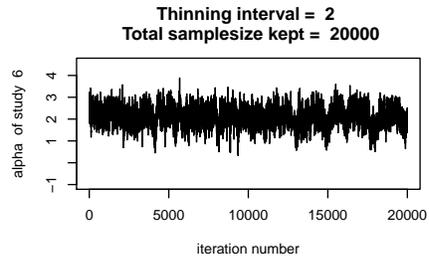
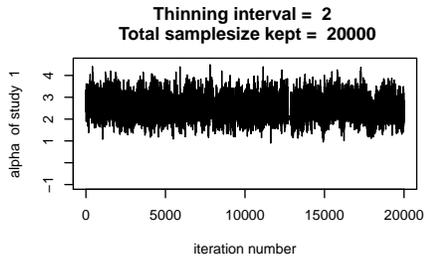
Posterior predictive value of Specificity of test under evaluation (C1)

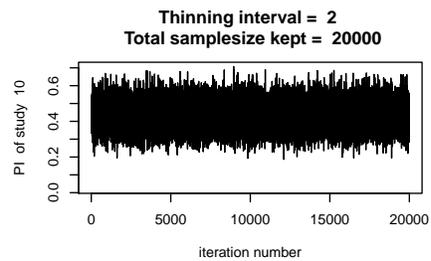
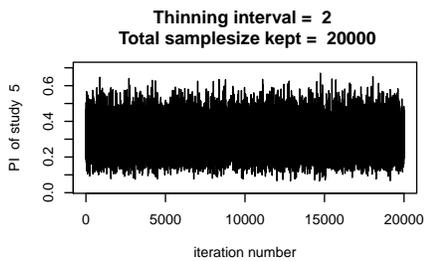
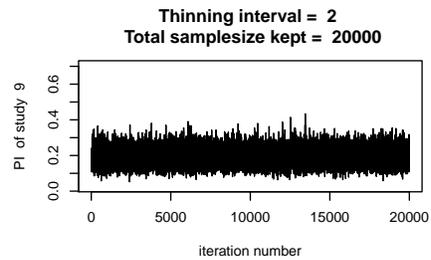
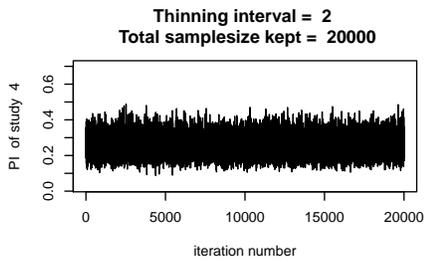
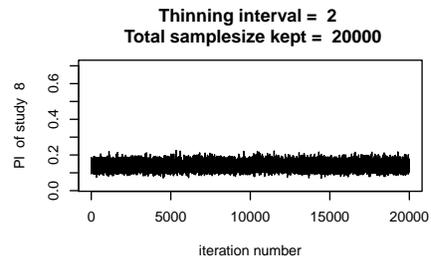
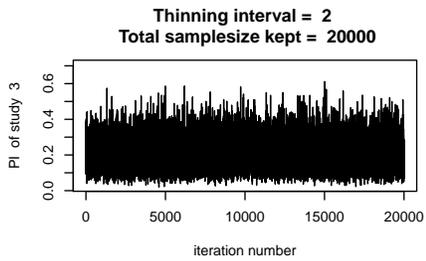
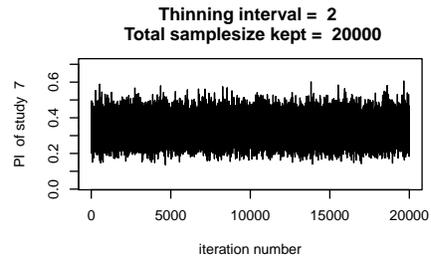
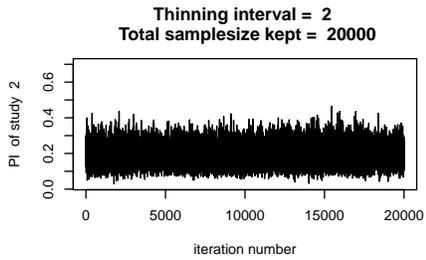
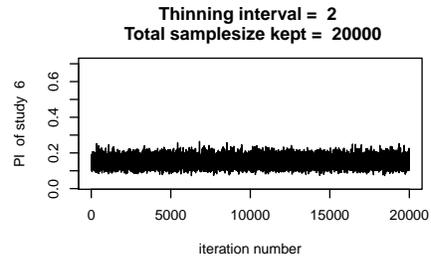
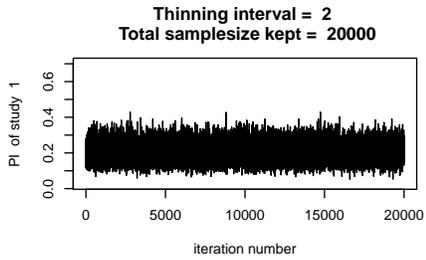
	Estimate	Standard_Dev	MC_error	C.I._lower	C.I._upper
Specificity	0.973323	0.095454	0.002266	0.759643	1

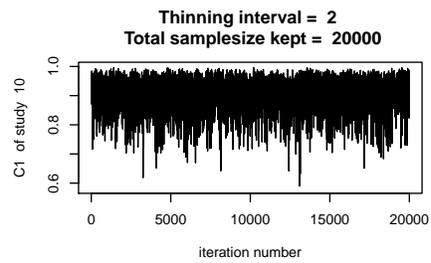
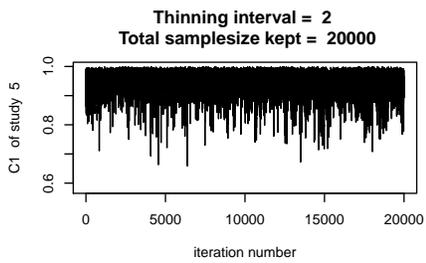
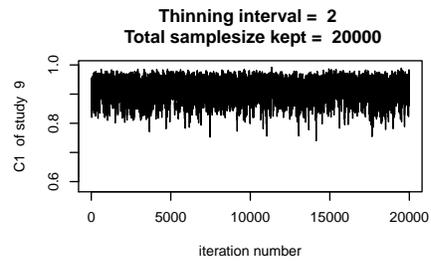
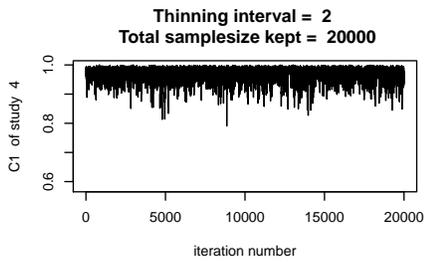
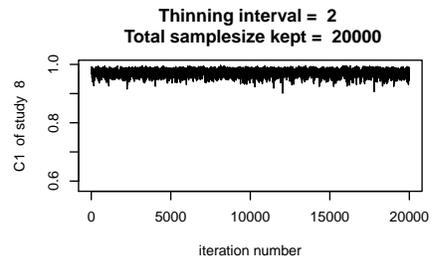
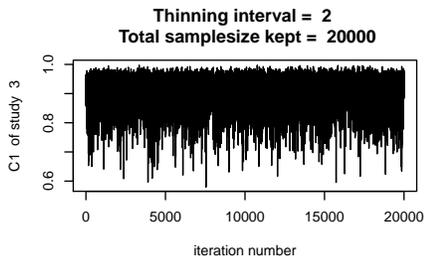
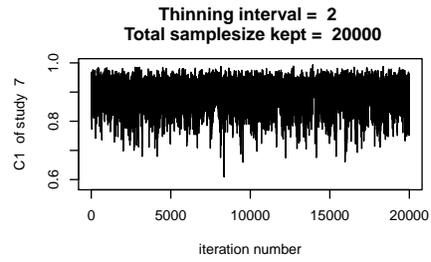
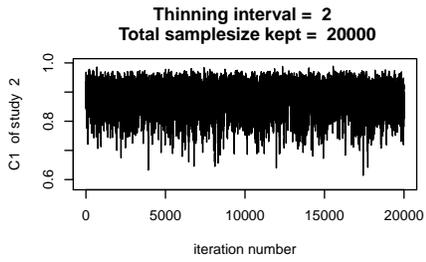
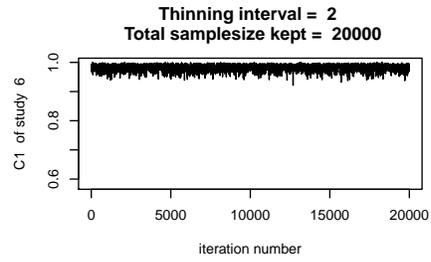
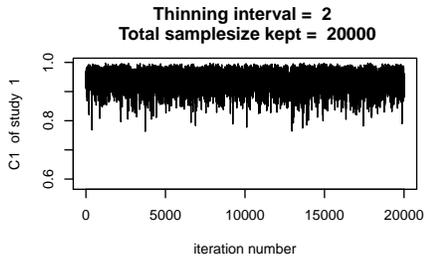
ANEXO B

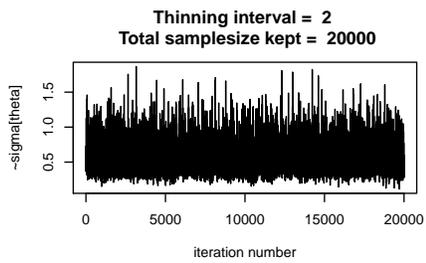
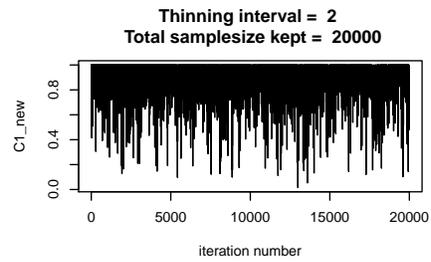
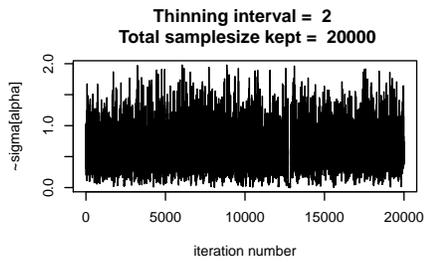
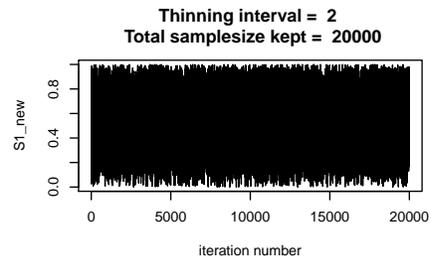
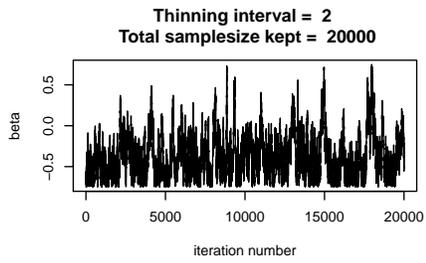
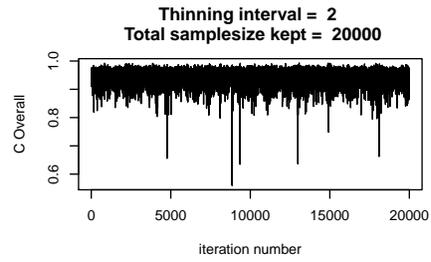
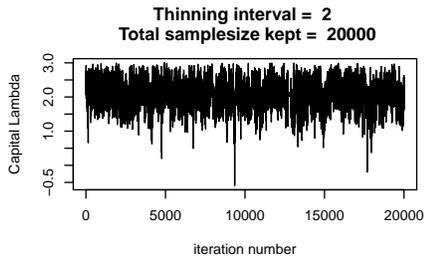
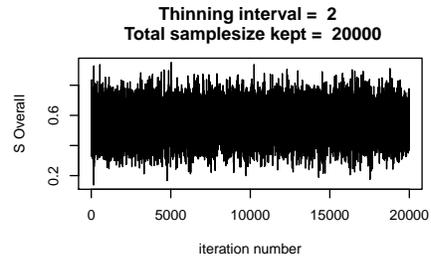
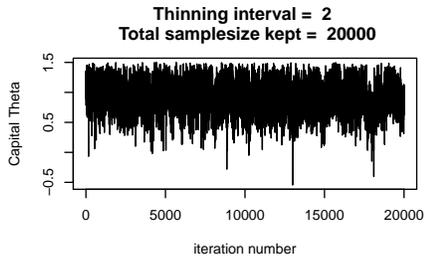
GRÁFICOS DE DISPERSIÓN, DISTRIBUCIÓN
MUESTRAL A POSTERIORI, DATOS MRI





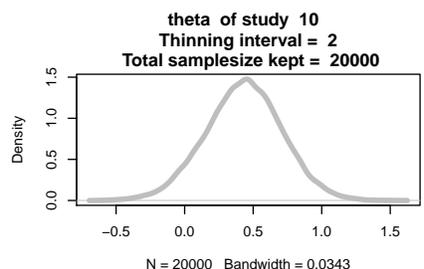
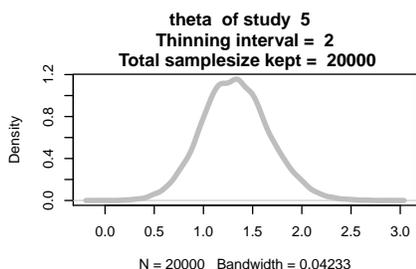
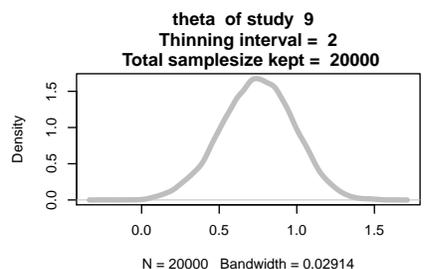
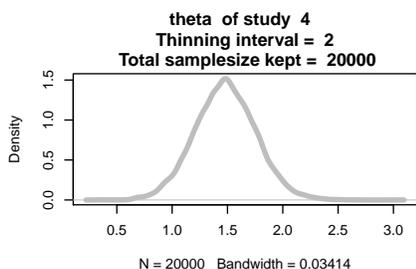
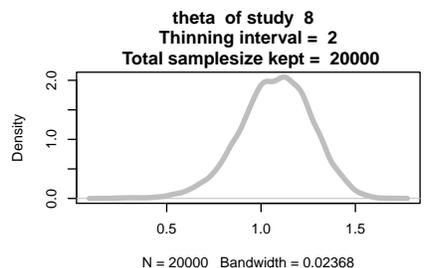
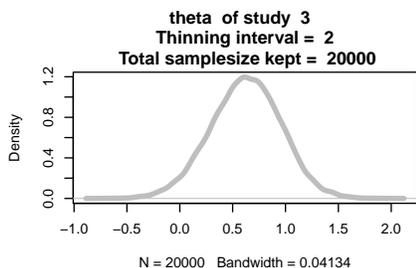
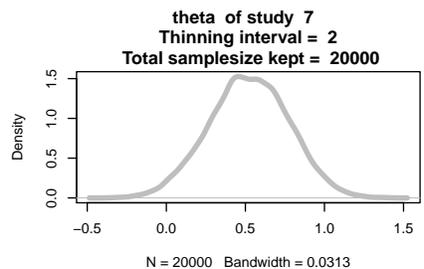
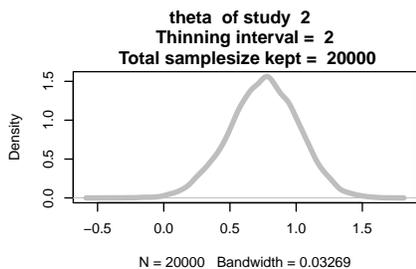
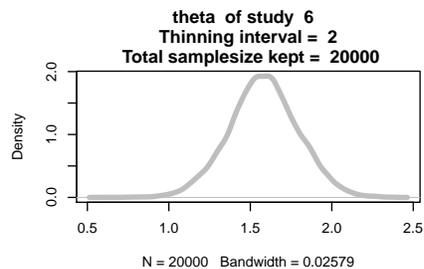
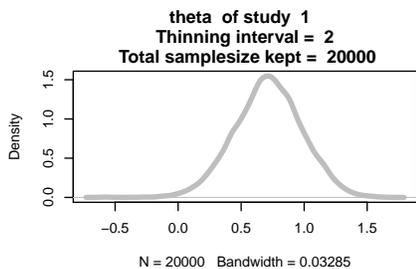


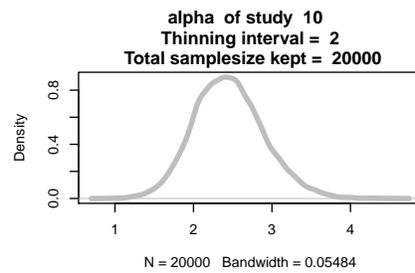
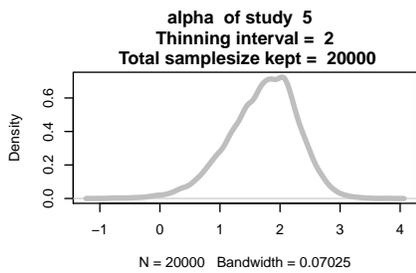
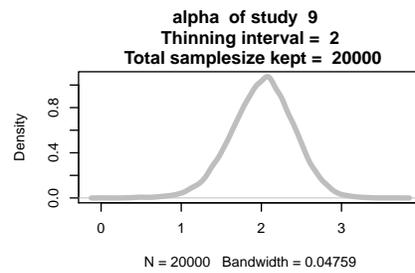
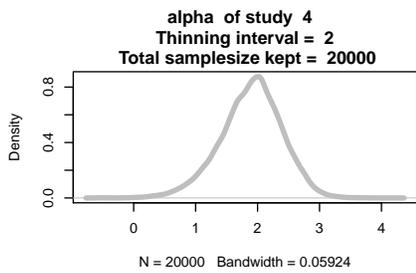
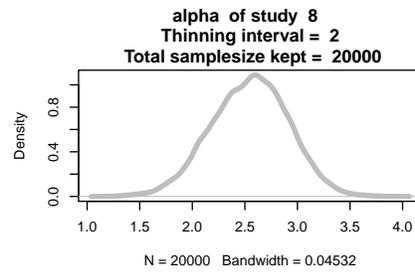
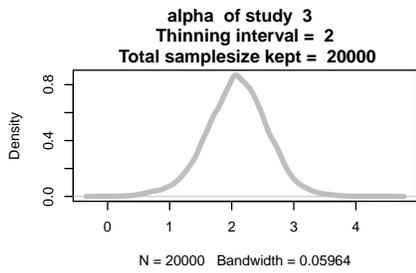
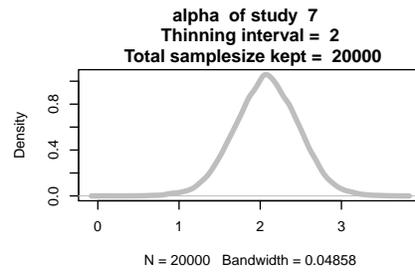
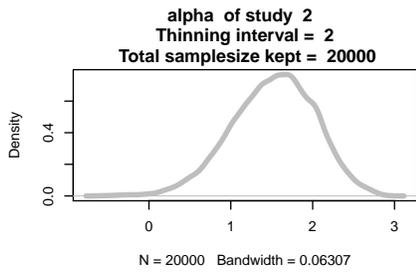
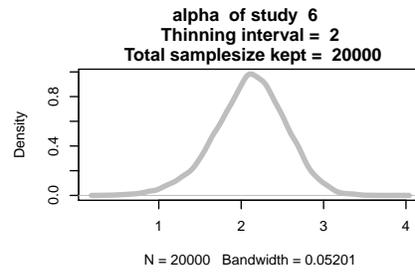
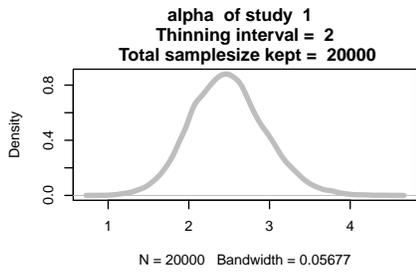


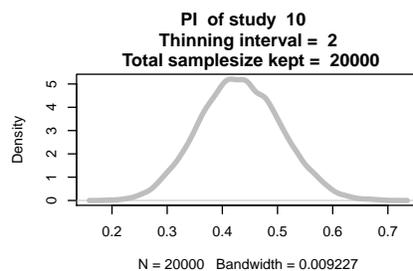
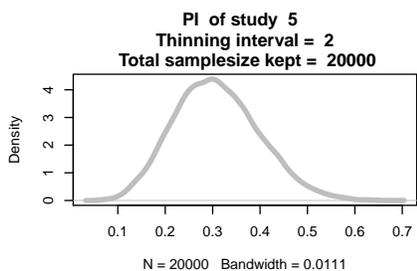
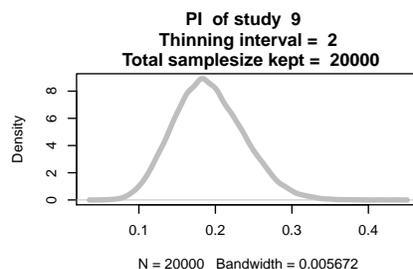
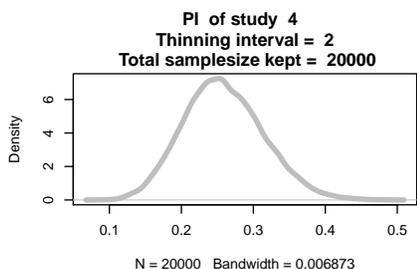
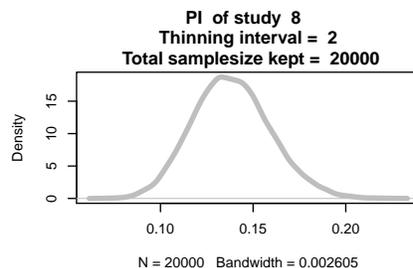
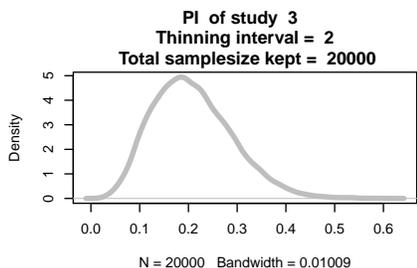
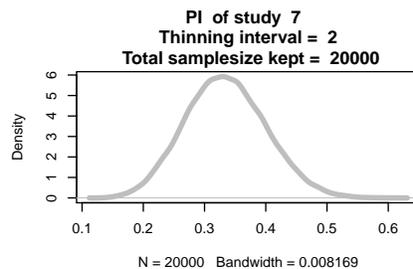
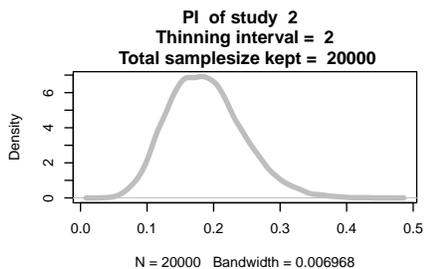
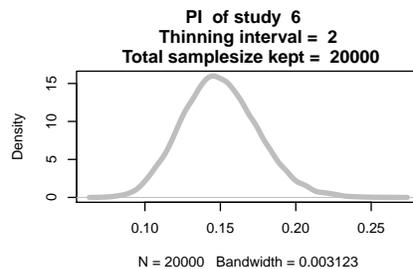
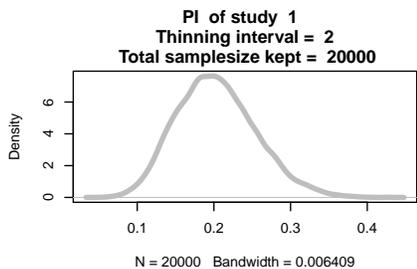


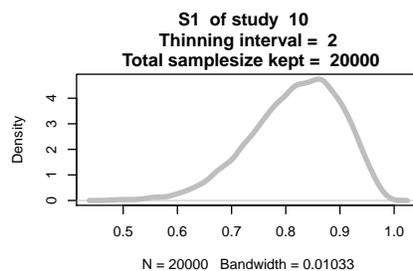
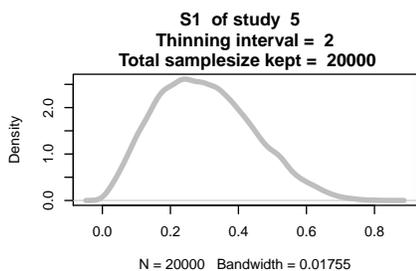
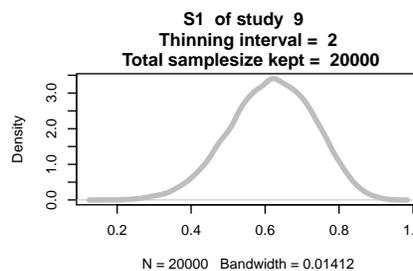
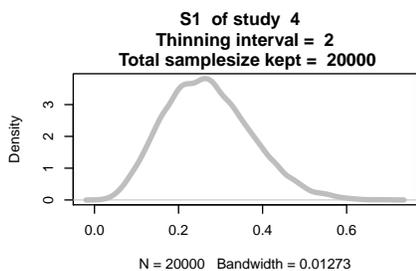
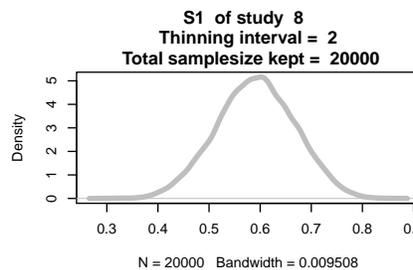
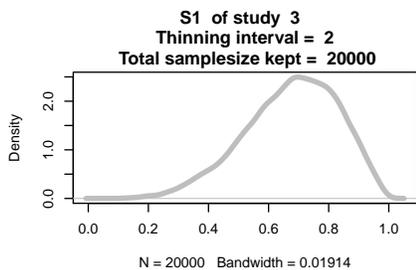
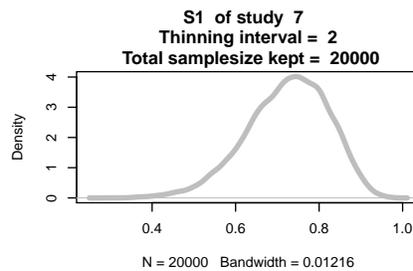
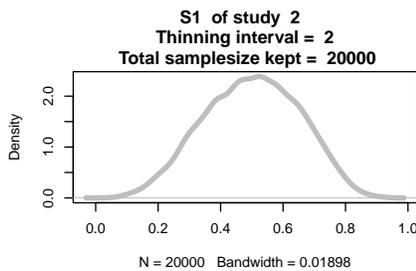
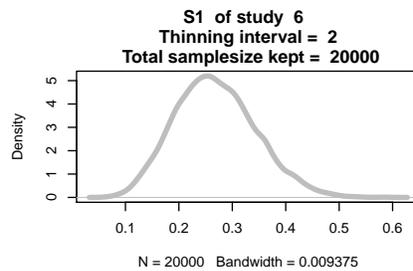
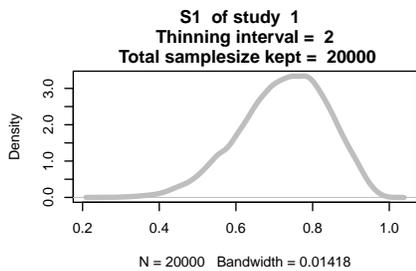
ANEXO C

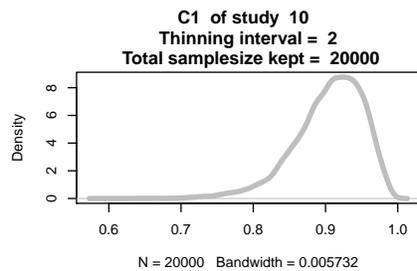
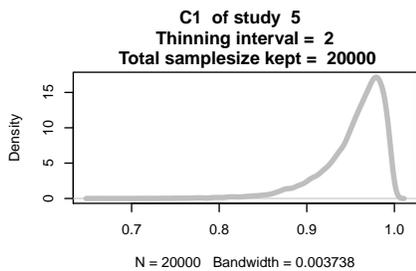
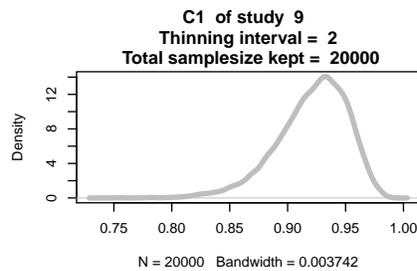
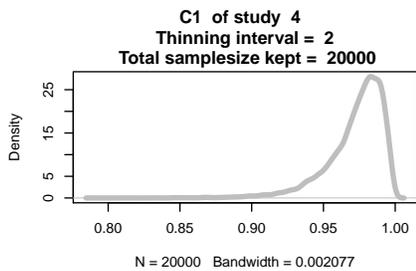
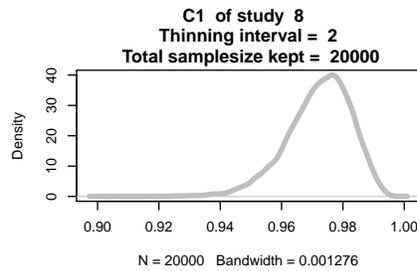
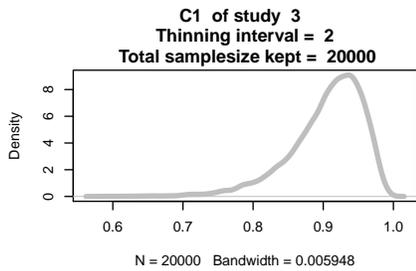
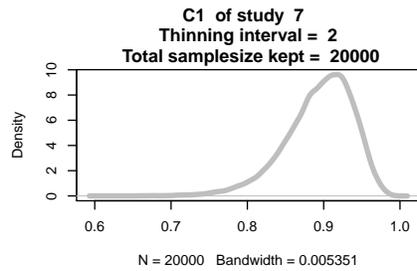
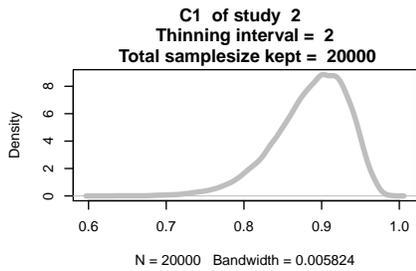
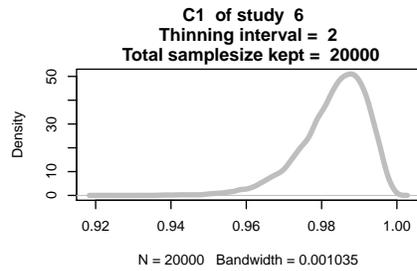
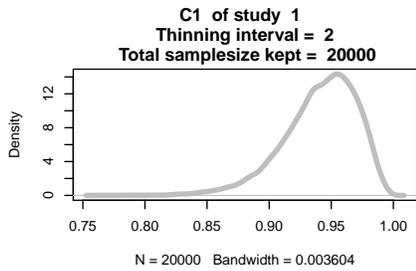
GRÁFICOS DE DENSIDAD DE LOS
PARÁMETROS INTER-ESTUDIO, DATOS
MRI

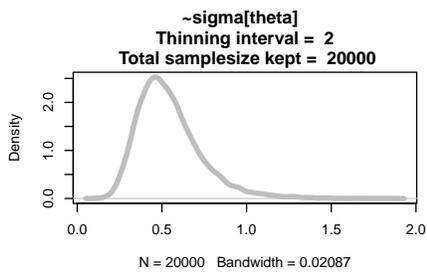
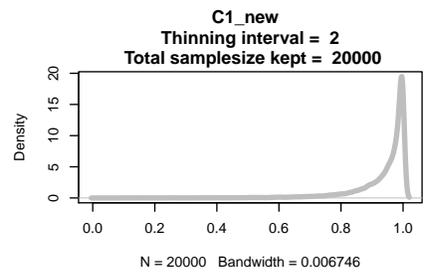
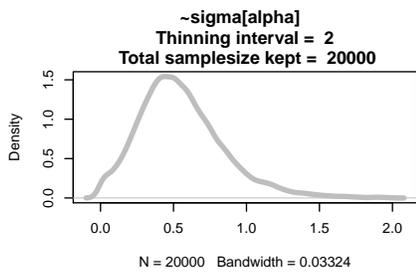
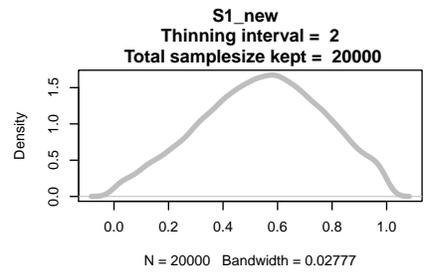
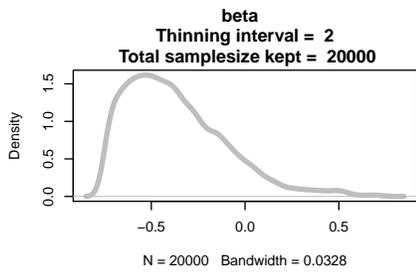
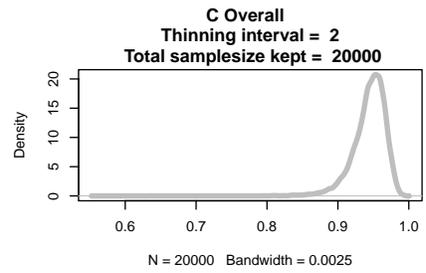
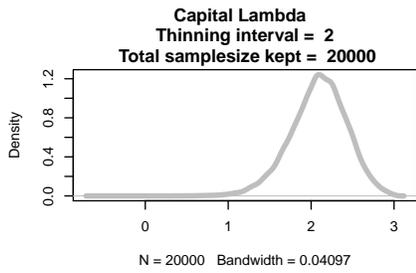
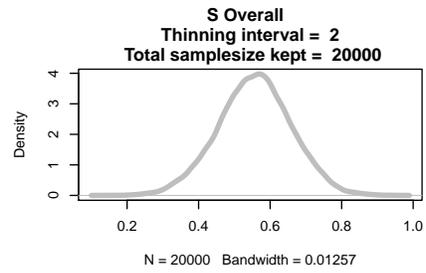
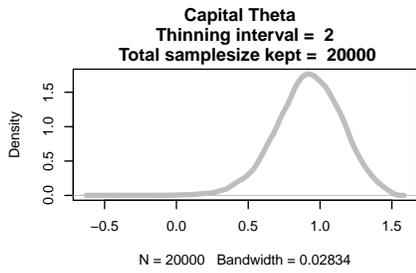












BIBLIOGRAFÍA

”(···) se enfrascó tanto en su lectura, que se le pasaban las noches leyendo de claro en claro, y los días de turbio en turbio, y así, del poco dormir y del mucho leer, se le secó el cerebro, de manera que vino a perder el juicio”.

Miguel de Cervantes Saavedra
El Ingenioso Hidalgo Don Quijote de la Mancha

- [1] Alexandersson, A. et al. (2004). Graphing confidence ellipses: An update of ellip for stata 8. *Stata Journal*, 4:242–256.
- [2] Arends, L., Hamza, T., Houwelingen, J., Heijnenbrok-Kal, M., Hunink, M., and Stijnen, T. (2008). Bivariate random effects meta-analysis of roc curves. *Medical Decision Making*.
- [3] Avilés Merens, R., Morales Morejón, M., Sao Avilés, A., and Cañedo Andalia, R. (2004). La colaboración cochrane en cuba. los metanálisis: aproximaciones útiles para su comprensión. acimed [internet]. [citado 30 ago 2014]; 12 (4):[aprox. 25 p.].
- [4] Bailey, K. R. (1987). Inter-study differences: How should they influence the interpretation and analysis of results? *Statistics in Medicine*, 6(3):351–358.
- [5] Baker, F. (1992). Item response theory. statistics: Textbooks and monographs n. 129. *Item Response Theory: STATISTICS: Textbooks and Monographs n. 129*.
- [6] Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of mathematical psychology*, 12(4):387–415.
- [7] Beck, J. R. and Shultz, E. K. (1986). The use of relative operating characteristic (roc) curves in test performance evaluation. *Archives of pathology & laboratory medicine*, 110(1):13–20.

- [8] Begg, C. B. and Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, pages 1088–1101.
- [9] Begg, C. B. and McNeil, B. (1988). Assessment of radiologic tests: control of bias and other design considerations. *Radiology*, 167(2):565–569.
- [10] Bipat, S., Glas, A. S., van der Velden, J., Zwinderman, A. H., Bossuyt, P. M., and Stoker, J. (2003). Computed tomography and magnetic resonance imaging in staging of uterine cervical carcinoma: a systematic review. *Gynecologic oncology*, 91(1):59–66.
- [11] Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., Moher, D., Rennie, D., De Vet, H. C., and Lijmer, J. G. (2003). The stard statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Annals of internal medicine*, 138(1):W1–12.
- [12] Boyko, E. J. (1994). Ruling out or ruling in disease with the most sensitive or specific diagnostic test short cut or wrong turn? *Medical Decision Making*, 14(2):175–179.
- [13] Burgueño, M., García-Bastos, J., and González-Buitrago, J. (1995). Las curvas roc en la evaluación de las pruebas diagnósticas. *Med Clin (Barc)*, 104(17):661–70.
- [14] Casares, L. M. (2001). Metaanálisis: claves para interpretar una herramienta de investigación controvertida. *Hipertensión y Riesgo Vascular*, 18(5):232–240.
- [15] Castillo, I. I. (2009). Evaluación de resultados clínicos (ii): Las medidas de la significación clínica o los tamaños del efecto. *Norte de Salud mental*, 8(34).
- [16] Chew, V. (1966). Confidence, prediction, and tolerance regions for the multivariate normal distribution. *Journal of the American Statistical Association*, 61(315):605–617.
- [17] Chu, H., Chen, S., and Louis, T. A. (2009). Random effects models in a meta-analysis of the accuracy of two diagnostic tests without a gold standard. *Journal of the American Statistical Association*, 104(486):512–523.
- [18] Chu, H. and Cole, S. R. (2006). Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of clinical epidemiology*, 59(12):1331–1332.
- [19] Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, 10(1):101–129.
- [20] Cooper, H. M. (1982). Scientific guidelines for conducting integrative research reviews. *Review of educational research*, 52(2):291–302.
- [21] de Dios, J. G. and Santamaría, A. B. (2007). Revisión sistemática y metanálisis (i): conceptos básicos. *Evidencias en pediatría*, 3(4):21.
- [22] Deeks, J. J. (2001). Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. *British Medical Journal*, 323(7305):157.
- [23] Deeks, J. J., Macaskill, P., and Irwig, L. (2005). The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of clinical epidemiology*, 58(9):882–893.

- [24] DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, pages 837–845.
- [25] Dendukuri, N., Hadgu, A., and Wang, L. (2009). Modeling conditional dependence between diagnostic tests: a multiple latent variable model. *Statistics in medicine*, 28(3):441–461.
- [26] Dendukuri, N. and Joseph, L. (2001). Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics*, 57(1):158–167.
- [27] Dendukuri, N., Schiller, I., Joseph, L., and Pai, M. (2012). Bayesian meta-analysis of the accuracy of a test for tuberculous pleuritis in the absence of a gold standard reference. *Biometrics*, 68(4):1285–1293.
- [28] DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled clinical trials*, 7(3):177–188.
- [29] Devillé, W. L., Buntinx, F., Bouter, L. M., Montori, V. M., De Vet, H. C., Van der Windt, D. A., and Bezemer, D. P. (2002). Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC medical research methodology*, 2(1):1.
- [30] Doebler, P. (2015). Mada: Meta-analysis of diagnostic accuracy. r package version 0.5. 7.
- [31] Douglas, J. (1993). Confidence regions for parameter pairs. *The American Statistician*, 47(1):43–45.
- [32] Dudewicz, E. and Mishra, S. (1988). Modern mathematical statistics, vol. 19 of wiley series in probability and mathematical statistics.
- [33] Dukic, V. and Gatsonis, C. (2003). Meta-analysis of diagnostic test accuracy assessment studies with varying number of thresholds. *Biometrics*, 59(4):936–946.
- [34] Dwamena, B. (2007). Midas: computational and graphical routines for meta-analytical integration of diagnostic accuracy studies in stata. *Ann Arbor: Division of Nuclear Medicine, Department of Radiology, University of Michigan Medical School*.
- [35] Egger, M. and Smith, G. D. (1997). Meta-analysis. potentials and promise. *BMJ: British Medical Journal*, 315(7119):1371.
- [36] Egger, M. and Smith, G. D. (1998). Bias in location and selection of studies. *BMJ: British Medical Journal*, 316(7124):61.
- [37] Fagan, T. J. (1975). Letter: nomogram for bayes theorem. *The New England journal of medicine*, 293(5):257–257.
- [38] Fahey, M. T., Irwig, L., and Macaskill, P. (1995). Meta-analysis of pap test accuracy. *American journal of epidemiology*, 141(7):680–689.
- [39] Fisher, N. and Switzer, P. (2001). Graphical assessment of dependence: Is a picture worth 100 tests? *The American Statistician*, 55(3):233–239.
- [40] Galindo-Villardón, M. (1991). Meta-análisis: el uso de los métodos estadísticos en revisiones de estudios de investigación relacionados. *Cirugía española*, 54(3):201–3.

- [41] Gasparrini, A., Armstrong, B., and Kenward, M. (2012). Multivariate meta-analysis for non-linear and other multi-parameter associations. *Statistics in medicine*, 31(29):3821–3839.
- [42] Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409.
- [43] Gelman, A. and Rubin, D. B. (1992). A single series from the gibbs sampler provides a false sense of security. *Bayesian statistics*, 4:625–631.
- [44] Gisbert, J. and Bonfill, X. (2004). ¿ cómo realizar, evaluar y utilizar revisiones sistemáticas y metaanálisis? *Gastroenterología y hepatología*, 27(3):129–149.
- [45] Glas, A. S., Lijmer, J. G., Prins, M. H., Bossel, G. J., and Bossuyt, P. M. (2003). The diagnostic odds ratio: a single indicator of test performance. *Journal of clinical epidemiology*, 56(11):1129–1135.
- [46] Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational researcher*, 5(10):3–8.
- [47] Gluud, C. and Gluud, L. L. (2005). Evidence based diagnostics. *Bmj*, 330(7493):724–726.
- [48] Green, B. F. and Hall, J. A. (1984). Quantitative methods for literature reviews. *Annual review of psychology*, 35(1):37–54.
- [49] Griner, P. F., Mayewski, R. J., Mushlin, A. I., and Greenland, P. (1981). Selection and interpretation of diagnostic tests and procedures. principles and applications. *Annals of internal medicine*, 94(4 Pt 2):557.
- [50] Guyatt, G. H., Tugwell, P., Feeny, D., Haynes, R., and Drummond, M. (1986). A framework for clinical evaluation of diagnostic technologies. *CMAJ: Canadian Medical Association Journal*, 134(6):587.
- [51] Habbema, J. D. F., Eijkemans, R., Krijnen, P., and Knottnerus, J. A. (2002). Analysis of data on the accuracy of diagnostic tests. *The evidence base of clinical diagnosis*, pages 117–144.
- [52] Hanley, J. A. (1998). Receiver operating characteristic (roc) curves. *Encyclopedia of biostatistics*.
- [53] Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36.
- [54] Harbord, R. M., Deeks, J. J., Egger, M., Whiting, P., and Sterne, J. A. (2007). A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics*, 8(2):239–251.
- [55] Harbord, R. M., Egger, M., and Sterne, J. A. (2006). A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Statistics in medicine*, 25(20):3443–3457.
- [56] Hasselblad, V. and Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological bulletin*, 117(1):167.
- [57] Hedges, L. and Olkin, I. (1985). Statistical methods for meta-analysis: Academic press. *Orlando, FL*.
- [58] Higgins, J. and Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*, 21(11):1539–1558.

- [59] Higgins, J. P., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *Bmj*, 327(7414):557–560.
- [60] Hilden, J. (1991). The area under the roc curve and its competitors. *Medical Decision Making*, 11(2):95–101.
- [61] Holling, H., Böhning, W., and Böhning, D. (2012). Meta-analysis of diagnostic studies based upon sroc-curves: a mixed model approach using the lehmann family. *Statistical Modelling*, 12(4):347–375.
- [62] Honest, H. and Khan, K. S. (2002). Reporting of measures of accuracy in systematic reviews of diagnostic literature. *BMC health services research*, 2(1):1.
- [63] Ioannis, N. (2009). Bayesian modeling using winbugs. hoboken.
- [64] Irwig, L., Macaskill, P., Glasziou, P., and Fahey, M. (1995). Meta-analytic methods for diagnostic test accuracy. *Journal of clinical epidemiology*, 48(1):119–130.
- [65] Irwig, L., Tosteson, A. N., Gatsonis, C., Lau, J., Colditz, G., Chalmers, T. C., and Mosteller, F. (1994). Guidelines for meta-analyses evaluating diagnostic tests. *Annals of Internal medicine*, 120(8):667–676.
- [66] Kardaun, J. W. and Kardaun, O. J. (1990). Comparative diagnostic performance of three radiological procedures for the detection of lumbar disk herniation. *CONTRIBUTIONS TO A RATIONAL DIAGNOSIS AND TREATMENT OF LUMBAR DISK HERNIATION*, page 49.
- [67] Khan, K. S., Bachmann, L. M., and ter Riet, G. (2003). Systematic reviews with individual patient data meta-analysis to evaluate diagnostic tests. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 108(2):121–125.
- [68] Knottnerus, J. and Muris, J. (2003). Assessment of the accuracy of diagnostic tests: the cross-sectional study. *Journal of clinical epidemiology*, 56(11):1118–1128.
- [69] Koelemay, M. J., Nederkoorn, P. J., Reitsma, J. B., and Majoie, C. B. (2004). Systematic review of computed tomographic angiography for assessment of carotid artery disease. *Stroke*, 35(10):2306–2312.
- [70] Kotz, S., Balakrishnan, N., and Johnson, N. L. (2000). Bivariate and trivariate normal distributions. *Continuous Multivariate Distributions: Models and Applications, Volume 1, Second Edition*, pages 251–348.
- [71] Kriston, L., Hoolzel, L., Weiser, A.-K., Berner, M. M., and Haarter, M. (2008). Meta-analysis: are 3 questions enough to detect unhealthy alcohol use? *Annals of internal medicine*, 149(12):879–888.
- [72] Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, pages 159–174.
- [73] Le, C. T. (2006). A solution for the most basic optimization problem associated with an roc curve. *Statistical methods in medical research*, 15(6):571–584.
- [74] Leeflang, M. M., Deeks, J. J., Gatsonis, C., and Bossuyt, P. M. (2008). Systematic reviews of diagnostic test accuracy. *Annals of internal medicine*, 149(12):889–897.

- [75] Light, R. and Pillemer, D. (1984). Summing up: the science of review research.
- [76] Light, R. and Smith, P. (1971). Accumulating evidence: Procedures for resolving contradictions among different research studies. *Harvard Educational Review*, 41(4):429–471.
- [77] Lijmer, J. G., Bossuyt, P. M., and Heisterkamp, S. H. (2002). Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Statistics in medicine*, 21(11):1525–1537.
- [78] Lijmer, J. G., Mol, B. W., Heisterkamp, S., Bossel, G. J., Prins, M. H., van der Meulen, J. H., and Bossuyt, P. M. (1999). Empirical evidence of design-related bias in studies of diagnostic tests. *Jama*, 282(11):1061–1066.
- [79] Lipsey, M. W. (2009). Identifying interesting variables and analysis opportunities. *The handbook of research synthesis and meta-analysis*, 2:147–158.
- [80] Littenberg, B. and Moses, L. E. (1993). Estimating diagnostic accuracy from multiple conflicting reports a new meta-analytic method. *Medical Decision Making*, 13(4):313–321.
- [81] Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. (2009). The bugs project: Evolution, critique and future directions. *Statistics in medicine*, 28(25):3049–3067.
- [82] Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). Winbugs—a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing*, 10(4):325–337.
- [83] Macaskill, P. (2004). Empirical bayes estimates generated in a hierarchical summary roc analysis agreed closely with those of a full bayesian analysis. *Journal of clinical epidemiology*, 57(9):925–932.
- [84] Macaskill, P., Gatsonis, C., Deeks, J., Harbord, R., and Takwoingi, Y. (2010). Chapter 10: analysing and presenting results. *Cochrane handbook for systematic reviews of diagnostic test accuracy version*, 1(0).
- [85] Macaskill, P., Walter, S. D., and Irwig, L. (2001). A comparison of methods to detect publication bias in meta-analysis. *Statistics in medicine*, 20(4):641–654.
- [86] Martín Rodero, H. (2011). Medicina basada en la evidencia y meta-análisis: una aproximación conceptual, histórica y metodológica.
- [87] Martínez-González, M., De Irala-Estévez, J., and Guillén-Grima, F. (1999). What is an odds ratio? *Medicina clínica*, 112(11):416.
- [88] McCullagh, P. (1980). Regression models for ordinal data. *Journal of the royal statistical society. Series B (Methodological)*, pages 109–142.
- [89] Metz, C. E. (1978). Basic principles of roc analysis. In *Seminars in nuclear medicine*, volume 8, pages 283–298. Elsevier.
- [90] Metz, C. E., Herman, B. A., and Shen, J.-H. (1998). Maximum likelihood estimation of receiver operating characteristic (roc) curves from continuously-distributed data. *Statistics in medicine*, 17(9):1033–1053.

- [91] Midgette, A. S., Stukel, T. A., and Littenberg, B. (1993). A meta-analytic method for summarizing diagnostic test performances receiver-operating-characteristic-summary point estimates. *Medical Decision Making*, 13(3):253–257.
- [92] Molinero, L. (2008). Metaanálisis, una guía rápida para lectores y usuarios. *Hipertensión y riesgo vascular*, 25(3):108–120.
- [93] Moses, L. E., Shapiro, D., and Littenberg, B. (1993). Combining independent studies of a diagnostic test into a summary roc curve: data-analytic approaches and some additional considerations. *Statistics in medicine*, 12(14):1293–1316.
- [94] Normand, S.-L. T. (1999). Tutorial in biostatistics meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in medicine*, 18(3):321–359.
- [95] Ortega, Z., Martín-Vallejo, J., Mencía, A., Galindo-Villardón, M., and Pérez-Mellado, V. (2016). Introducing meta-partition, a useful methodology to explore factors that influence ecological effect sizes. *PloS one*. To appear.
- [96] Oxman, A. D., Cook, D. J., Guyatt, G. H., Bass, E., Brill-Edwards, P., Browman, G., Detsky, A., Farkouh, M., Gerstein, H., Haines, T., et al. (1994). Users' guides to the medical literature: Vi. how to use an overview. *Jama*, 272(17):1367–1371.
- [97] Patrick, D. L., Cheadle, A., Thompson, D. C., Diehr, P., Koepsell, T., and Kinne, S. (1994). The validity of self-reported smoking: a review and meta-analysis. *American journal of public health*, 84(7):1086–1093.
- [98] Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., and Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *Jama*, 295(6):676–680.
- [99] Phillips, B., Stewart, L. A., and Sutton, A. J. (2010). "cross hairs" plots for diagnostic meta-analysis. *Research Synthesis Methods*, 1(3-4):308–315.
- [100] Pita Fernández, S. and Pértegas Díaz, S. (2003). Pruebas diagnósticas: sensibilidad y especificidad. *Cad Aten Primaria*, 10:120–4.
- [101] Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2004). Gllamm manual. uc berkeley division of biostatistics working paper series. *University of California, Berkeley*.
- [102] Rabe-Hesketh, S., Skrondal, A., Pickles, A., et al. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, 2(1):1–21.
- [103] Ransohoff, D. F. and Feinstein, A. R. (1978). Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *New England Journal of Medicine*, 299(17):926–930.
- [104] Reid, M. C., Lachs, M. S., and Feinstein, A. R. (1995). Use of methodological standards in diagnostic test research: getting better but still not good. *Jama*, 274(8):645–651.
- [105] Reitsma, J. B., Glas, A. S., Rutjes, A. W., Scholten, R. J., Bossuyt, P. M., and Zwinderman, A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of clinical epidemiology*, 58(10):982–990.

- [106] Riley, R. D., Abrams, K., Lambert, P., Sutton, A., and Thompson, J. (2007a). An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Statistics in medicine*, 26(1):78–97.
- [107] Riley, R. D., Abrams, K. R., Sutton, A. J., Lambert, P. C., and Thompson, J. R. (2007b). Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Medical Research Methodology*, 7(1):1.
- [108] Riley, R. D., Thompson, J. R., and Abrams, K. R. (2008). An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics*, 9(1):172–186.
- [109] Robert, R. (1984). *Meta-analytic procedures for social research*, Beverly hills.
- [110] Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological bulletin*, 86(3):638.
- [111] Rousseeuw, P. J., Ruts, I., and Tukey, J. W. (1999). The bagplot: a bivariate boxplot. *The American Statistician*, 53(4):382–387.
- [112] Rücker, G., Schwarzer, G., and Carpenter, J. (2008). Arcsine test for publication bias in meta-analyses with binary outcomes. *Statistics in medicine*, 27(5):746–763.
- [113] Rutter, C. and Gatsonis, C. (1995). Regression methods for meta-analysis of diagnostic test data. *Academic Radiology*, 2:S48–56.
- [114] Rutter, C. M. and Gatsonis, C. A. (2001). A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in medicine*, 20(19):2865–2884.
- [115] Sackett, D. and Haynes, R. (2002). The architecture of diagnostic research. *British Medical Journal*, 324(7336):539.
- [116] Sackett, D., Haynes, R., Guyatt, G., and Tugwell, P. (1991). The selection of diagnostic tests. *Clinical epidemiology*, 2nd ed. Boston/Toronto/London: Little, Brown and Company, pages 47–57.
- [117] Sackett, D. L., Haynes, R. B., Tugwell, P., et al. (1985). *Clinical epidemiology: a basic science for clinical medicine*. Little, Brown and Company.
- [118] Sadatsafavi, M., Shahidi, N., Marra, F., FitzGerald, M. J., Elwood, K. R., Guo, N., and Marra, C. A. (2010). A statistical method was used for the meta-analysis of tests for latent tb in the absence of a gold standard, combining random-effect and latent-class methods to estimate test accuracy. *Journal of clinical epidemiology*, 63(3):257–269.
- [119] Sáenz, A. and Ausejo, M. (2000). Guía para leer e interpretar una revisión sistemática. *Rev Soc Mad Med Fam Comunit*, 2:29–36.
- [120] Scheidler, J., Hricak, H., Kyle, K. Y., Subak, L., and Segal, M. R. (1997). Radiological evaluation of lymph node metastases in patients with cervical cancer: a meta-analysis. *Jama*, 278(13):1096–1101.
- [121] Schiller, I., Dendukuri, N., and Schiller, M. I. (2015). Package 'hsroc'.

- [122] Scholten, R., Opstelten, W., van der Plas, C. G., Bijl, D., Deville, W., Bouter, L. M., et al. (2003). Accuracy of physical diagnostic tests for assessing ruptures of the anterior cruciate ligament: a meta-analysis. *Journal of family practice*, 52(9):689–696.
- [123] Schwarzer, G., Antes, G., and Schumacher, M. (2002). Inflation of type i error rate in two statistical tests for the detection of publication bias in meta-analyses with binary outcomes. *Statistics in medicine*, 21(17):2465–2477.
- [124] Schwarzer, G., Antes, G., and Schumacher, M. (2007). A test for publication bias in meta-analysis with sparse binary data. *Statistics in medicine*, 26(4):721–733.
- [125] Seqc (2010). Evaluación de Pruebas Diagnósticas. <http://bit.ly/1U5kAdA/>. [Online; acceso 20-Mayo-2016].
- [126] Sergas (2011). Pruebas Diagnósticas. <http://sergas.es/busca?k=PRUEBAS+DIAGNOSTICA/>. [Online; acceso 21-Mayo-2016].
- [127] Shapiro, D. (1995). Issues in combining independent estimates of the sensitivity and specificity of a diagnostic test. *Academic radiology*, 2:S37.
- [128] Sheps, S. B. and Schechter, M. T. (1984). The assessment of diagnostic tests: a survey of current medical research. *Jama*, 252(17):2418–2422.
- [129] Siadat, M. and Shu, J. (2004). Proportional odds ratio model for comparison of diagnostic tests in meta-analysis. *BMC medical research methodology*, 4(1):1.
- [130] Song, F. (1999). Exploring heterogeneity in meta-analysis: is the l’abbe plot useful? *Journal of clinical epidemiology*, 52(8):725–730.
- [131] Song, F., Khan, K. S., Dinnes, J., and Sutton, A. J. (2002). Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. *International journal of epidemiology*, 31(1):88–95.
- [132] Sorribas, A., March, J., and Trujillano, J. (2002). A new parametric method based on s-distributions for computing receiver operating characteristic curves for continuous diagnostic tests. *Statistics in medicine*, 21(9):1213–1235.
- [133] Stengel, D., Bauwens, K., Sehouli, J., Ekkernkamp, A., and Porzsolt, F. (2003). Original paper: A likelihood ratio approach to meta-analysis of diagnostic studies. *Journal of medical screening*, 10(1):47–51.
- [134] Sutton, A. J. and Higgins, J. (2008). Recent developments in meta-analysis. *Statistics in medicine*, 27(5):625–650.
- [135] Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240(4857):1285–1293.
- [136] Tatsioni, A., Zarin, D. A., Aronson, N., Samson, D. J., Flamm, C. R., Schmid, C., and Lau, J. (2005). Challenges in systematic reviews of diagnostic technologies. *Annals of internal medicine*, 142(12_Part_2):1048–1055.
- [137] Thompson, J. (2014). Stata vs r. <http://staffblogs.le.ac.uk/bayeswithstata/2014/11/13/stata-vs-r/>. [Online; acceso 30-Mayo-2016].

- [138] Thompson, S. G. and Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in medicine*, 18(20):2693–2708.
- [139] Tosteson, A. N. A. and Begg, C. B. (1988). A general regression methodology for roc curve estimation. *Medical Decision Making*, 8(3):204–215.
- [140] Vallejo, F. J. M. (1996). *Métodos estadísticos en meta-análisis*. PhD thesis, Universidad de Salamanca.
- [141] van Houwelingen, H. and Senn, S. (1999). Investigating underlying risk as a source of heterogeneity in meta-analysis by sg thompson, tc smith and sj sharp, *statistics in medicine*, 16, 2741–2758 (1997). *Statistics in Medicine*, 18(1):110–115.
- [142] Van Houwelingen, H. C., Arends, L. R., and Stijnen, T. (2002). Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in medicine*, 21(4):589–624.
- [143] Van Houwelingen, H. C., Zwinderman, K. H., and Stijnen, T. (1993). A bivariate approach to meta-analysis. *Statistics in medicine*, 12(24):2273–2284.
- [144] Walter, S. (2002). Properties of the summary receiver operating characteristic (sroc) curve for diagnostic test data. *Statistics in medicine*, 21(9):1237–1256.
- [145] Walter, S., Irwig, L., and Glasziou, P. (1999). Meta-analysis of diagnostic tests with imperfect reference standards. *Journal of clinical epidemiology*, 52(10):943–951.
- [146] Walter, S. and Jadad, A. (1999). Meta-analysis of screening data: a survey of the literature. *Statistics in medicine*, 18(24):3409–3424.
- [147] Weiss, R. E. (1996). Bayesian model checking with applications to hierarchical models. *Unpublished Technical Report*.
- [148] Whiting, P., Harbord, R., and Kleijnen, J. (2005). No role for quality scores in systematic reviews of diagnostic accuracy studies. *BMC Medical Research Methodology*, 5(1):1.
- [149] Whiting, P., Rutjes, A., Dinnes, J., Reitsma, J., Bossuyt, P., Kleijnen, J., et al. (2004a). *Development and validation of methods for assessing the quality of diagnostic accuracy studies*. Gray Pub.
- [150] Whiting, P., Rutjes, A. W., Reitsma, J. B., Bossuyt, P. M., and Kleijnen, J. (2003). The development of quadas: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC medical research methodology*, 3(1):1.
- [151] Whiting, P., Rutjes, A. W., Reitsma, J. B., Glas, A. S., Bossuyt, P. M., and Kleijnen, J. (2004b). Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Annals of Internal Medicine*, 140(3):189–202.
- [152] Whiting, P. F., Weswood, M. E., Rutjes, A. W., Reitsma, J. B., Bossuyt, P. N., and Kleijnen, J. (2006). Evaluation of quadas, a tool for the quality assessment of diagnostic accuracy studies. *BMC medical research methodology*, 6(1):1.
- [153] Zhou, X.-H. (1998). Correcting for verification bias in studies of a diagnostic test’s accuracy. *Statistical methods in medical research*, 7(4):337–353.

