

COMUNICACIONES CARTEL

Viernes, 13 de septiembre (15:30 h)

C.2.5. Métodos de medición y análisis

Moderadora:
Carmen Íñiguez

REVISIÓN SISTEMÁTICA DE LA FIABILIDAD Y VALIDEZ DE LA COMPOSITE INTERNATIONAL DIAGNOSTIC INTERVIEW (CIDI)

B. Romera, G. Vilagut, J.M. Puigvert, C. Palacín, M. Ferrer, J.M. Haro y J. Alonso

Unidad de Investigación en Servicios Sanitarios, Instituto Municipal de Investigación Médica (IMIM), Barcelona. Unidad de Investigación y Desarrollo, Hospital Sant Joan de Déu-SSM, Sant Boi de Llobregat.

Antecedentes: Los estudios epidemiológicos de salud mental requieren instrumentos diagnósticos fiables, válidos y aceptables. La Composite International Diagnostic Interview (CIDI) es una entrevista diagnóstica estandarizada y estructurada diseñada para la evaluación de los trastornos mentales. La nueva versión de la CIDI, la CIDI-2000, permite el diagnóstico siguiendo los criterios de la Clasificación Internacional de Enfermedades (CIE-10) y del Manual Diagnóstico y Estadístico de los Trastornos Mentales (DSM-IV). Aunque la validez diagnóstica de este instrumento ha sido cuestionada, se usa cada vez más en estudios epidemiológicos internacionales. El objetivo de este estudio es analizar la fiabilidad y validez de la CIDI.

Métodos: Búsqueda bibliográfica en la base de datos Medline (1988-2002) con las palabras clave: Composite International Diagnostic Interview, CIDI y [Reliability, Validity (Specificity and Sensitivity)]. Se seleccionaron todos los artículos (originales o de revisión) de estudios sobre fiabilidad y validez de cualquier versión de la CIDI. Se excluyeron los estudios de prevalencia, los estudios donde la CIDI se utilizaba como "gold standard" y aquellos no publicados en inglés. Los datos recogidos fueron: versión de la CIDI; ámbito de estudio (clínico o poblacional); tamaño de la muestra; tipo de diseño; valores de fiabilidad y de validez. Para el análisis de la concordancia se consideró como valor suficiente $k > 0,40$ definidos por Landis y Koch (Biometrics 1977;33:159-74).

Resultados: Se identificaron 109 publicaciones relevantes. Hasta el momento se han revisado 47 (43,2%), de las cuales solamente 11 cumplían los criterios de selección. Dichas publicaciones correspondían a un total de 26 investigaciones: 13 de fiabilidad y 13 de validez. Tres de los 13 estudios de fiabilidad (23%) se realizaron con muestras de pacientes y los 10 restantes (77%) con muestras de población general. Cinco de los 13 estudios de validez (38,5%) se realizaron con muestras de pacientes y los 8 restantes (61,5%) con muestras de población general. De 33 posibles diagnósticos de la CIDI, se ha evaluado solo la fiabilidad de 16 (48,5%). Todos ellos, excepto uno, informaban de un valor de fiabilidad $k > 0,40$. La mayor fiabilidad se describió para el diagnóstico de Pánico ($k = 0,88$) en un estudio de fiabilidad test-retest de ámbito clínico, y para Agorafobia ($k = 0,99$) en un estudio de fiabilidad interobservadores de ámbito de población general. El análisis de la validez mostró mucha heterogeneidad tanto en los diseños como en la selección del procedimiento diagnóstico considerado como "gold standard". De los 33 posibles diagnósticos, se habían evaluado sólo 8 (24,24%), 3 de los cuales mostraron concordancia con el "gold standard" de $k > 0,40$. La máxima concordancia se encontró con el Present State Examination (PSE) para el diagnóstico de Fobia Específica ($k = 0,88$) en un estudio de ámbito clínico.

Conclusiones: La fiabilidad es aceptable para la gran mayoría de diagnósticos evaluados. La validez muestra resultados variables y de difícil interpretación. Muchos diagnósticos de la CIDI requieren un mayor estudio de su fiabilidad y validez.

Financiación: UE QLG5-1999-01042, FIS (00/0028-02), CICYT 2000-1800-CE, CIRIT (2001SGR 00405).

263

262

CONSISTENCIA INTERNA Y ESTRUCTURA FACTORIAL DE UN CUESTIONARIO DE SATISFACCIÓN DE LOS USUARIOS EN ATENCIÓN PRIMARIA DE SALUD

M.C. Cabezas, J. Baroni, C. Valero, M.P. González, A. Céspedes, M.L. Maldonado, L. Sampietro, J. Gené y C. Pujol

Unitat de Recerca en Atenció Primària, Fundació J. Gol i Gurina. Institut Català de la Salut, Barcelona. Servei d'Atenció Primària Baix Llobregat Centre, Institut Català de la Salut, Cornellà. DOSIPA, Institut Català de la Salut, Barcelona. Divisió atenció Primària, Institut Català de la Salut, Barcelona. Fundació J Gol i Gurina, Barcelona.

Antecedentes: Las encuestas de satisfacción de los usuarios constituyen un instrumento fundamental para evaluar la calidad de los servicios sanitarios. Se han desarrollado diversos cuestionarios que han sido utilizados en la práctica, pero sin haber llegado a un acuerdo y sin haberse realizado un análisis de su fiabilidad y validez.

Objetivo: Evaluar la fiabilidad y estructura factorial de un cuestionario de satisfacción de los usuarios en atención primaria de salud.

Métodos: Cuestionario autoadministrado repartido durante una semana de 2001 en las salas de espera de 9 centros de atención primaria urbanos. Consta de 18 ítems medidos en una escala del 1 al 9, 2 con respuesta categórica y 2 preguntas cualitativas, además de la edad y el sexo del encuestado. Los 18 ítems forman una escala que recoge información sobre la satisfacción con 5 bloques de aspectos relevantes: aspectos organizativos, atención de los profesionales médicos, de enfermería y de atención al usuario y estado de las instalaciones. Después de comprobar las condiciones de aplicación (Bartlett, Kayser-Meyer-Olkin, y determinante de la matriz de correlaciones), se evalúa la validez factorial de la escala mediante análisis factorial, y la fiabilidad mediante un análisis de consistencia interna. Se ha utilizado SPSS versión 10.

Resultados: 1.780 personas han contestado el cuestionario, 64,2% mujeres. La edad media fue de 46,4 años (DE 15,5) con un rango entre 18 y 90. El porcentaje de no respuesta para todas las preguntas fue muy bajo, menor del 7% en todas las ocasiones. Las condiciones de aplicación del análisis factorial fueron muy buenas. La estructura conceptual del cuestionario se vio confirmada mediante un análisis factorial con extracción de factores por el método de máxima verosimilitud y rotación varimax. Los 5 factores explicaban un 72,82% de la varianza. El alfa de Cronbach mostró una consistencia óptima tanto para la escala global 0,94, como para los 5 factores: Organización 0,87, atención médica 0,92, atención de enfermería 0,96, atención de profesionales de atención al usuario 0,96, y estado de las instalaciones 0,74. La correlación ítem-total varió entre 0,61 y 0,92 (buena o excelente en todas las ocasiones).

Conclusiones: El cuestionario de satisfacción utilizado se muestra como un instrumento viable, fiable y válido para medir la satisfacción de los usuarios de centros de atención primaria según el modelo propuesto. Faltan los resultados de un estudio test-retest en marcha que valorará la estabilidad de los resultados.

CONSTRUCCIÓN DE UMBRALES DE ALERTA MEDIANTE MÉTODOS BOOTSTRAP

A.M. Alonso y J.J. Romo

Dpto de Matemáticas, Universidad Autónoma de Madrid, Madrid. Dpto de Estadística y Econometría, Universidad Carlos III de Madrid, Getafe.

Antecedentes: En vigilancia epidemiológica es necesario el pronóstico preciso de la ocurrencia de casos de determinadas enfermedades y la detección oportuna de valores inusuales en los casos notificados. Un enfoque clásico para investigar este problema se basa en modelos de series temporales, en particular, es usual la utilización de modelos ARIMA estacionales gaussianos. Sin embargo, este enfoque y por tanto sus pronósticos y umbrales de alerta pueden verse afectados por desviaciones del supuesto de normalidad de la verdadera distribución de las observaciones. Por ejemplo, Cardinal et al. (1999) ilustran este defecto en la serie de casos de meningitis meningocócica. Es frecuente observar en datos epidemiológicos distribuciones asimétricas, siendo algunas de sus posibles causas el reporte en lotes, o la conjunción de varios serogrupos como en la enfermedad meningocócica. En Alonso et al. (2002) hemos propuesto métodos bootstrap asociados al problema de predicción, en este trabajo consideramos la utilización de estos métodos para la construcción de umbrales de alerta en series de casos notificados de enfermedades infecciosas.

Métodos: Se extiende la metodología de pronóstico y construcción de intervalos de confianza basada en el sieve bootstrap propuesta por Alonso et al. (2002) a series temporales con componente estacional característica presente en las series de casos notificados. El procedimiento propuesto es independiente del modelo, por tanto, los intervalos de predicción que construyamos no necesitarán de la especificación de un modelo finito dimensional como en métodos precedentes. Se realiza un estudio de Monte Carlo para compararlo con métodos clásicos utilizando los modelos considerados por Zaidi et al. (1989), Cardinal et al. (1999) y Williamson y Hudson (1999). Por último, se ilustra su comportamiento en series de notificaciones de meningitis meningocócica.

Resultados: Para comparar los distintos intervalos de predicción, se utiliza la cobertura media de los intervalos, y la proporción de observaciones fuera del intervalo, siendo los principales resultados: (i) la cobertura media con el método Box-Jenkins es ligeramente superior a la cobertura media con sieve bootstrap para innovaciones normales; (ii) en procesos con innovaciones asimétricas, las proporciones de observaciones futuras fuera de los intervalos de predicción, muestran que BJ no aproxima correctamente a la distribución de X_{T+h} ; (iii) en los tres modelos, observamos el mejor comportamiento en cobertura media del método sieve bootstrap con respecto al BJ. Por último, en la serie de casos notificados de meningitis, el método bootstrap propuesto capta la asimetría presente en los datos.

Conclusiones: El método propuesto para construir umbrales de alerta tiene un comportamiento generalmente superior a los métodos clásicos y presenta un comportamiento satisfactorio en estudios de Monte Carlo y en datos reales bajo distintas distribuciones.

Bibliografía:

Alonso, A.M., Peña, D. y Romo, J. (2002) In: J. Stat. Plann. Inference, 100, 1-11.
Cardinal M., Roy R. y Lambert J. (1999) In: Stat. Medicine, 18, 2025-2039.
Williamson G.D. y Hudson G.W. (1999) In: Stat. Medicine, 18, 3283-3298.
Zaidi A.A., Schnell D.J. y Reynolds G.H. (1989) In: Stat. Medicine, 8, 353-362.

264

265

CALIDAD FORMAL DE LOS DATOS DE ACCIDENTES DE TRABAJO CON BAJA

F.G. Benavides, M.T. Giráldez, E. Castejón, N. Catot, M. Zaplana, J. Benach, D. Gimeno y J. Delclós

Antecedentes: Los datos de accidentes de trabajo con baja constituyen la fuente más utilizada para valorar la salud laboral en nuestro país. Sin embargo, existen muy pocos estudios que hayan valorado su calidad. La evaluación de la calidad formal no puede ser un sustituto de la evaluación de la fiabilidad y la exactitud de la información, pero proporciona indicadores para evaluar campañas para mejorar la calidad de nuestros sistemas de información y para identificar problemas más graves de la información que manejamos. El objetivo de este estudio es valorar la calidad formal de los datos de accidentes de trabajo con baja.

Métodos: A partir de los ficheros de datos correspondientes a los años 1989 (704.570 registros), 1993 (590.043) y 2000 (1.041.493), proporcionados por el Ministerio de Trabajo y Seguridad Social, fueron analizados algunos indicadores de calidad formal: cumplimiento de las variables ($n = 34$) y coherencia entre la antigüedad en la empresa y la edad, la edad y el año del accidente, el año del accidente y el año del fichero de datos, la antigüedad en el puesto de trabajo y la antigüedad en la empresa. Resultados. La mayoría de las variables estaban cumplimentadas. Si bien el tamaño de la empresa y el motivo del alta presentaban un porcentaje elevado de no cumplimentación: 12,3% y 18,5%, respectivamente, en 1989; 16% y 18% en 1993; y 22,1% y 15% en 2000. Además, los accidentes con baja en los que el motivo del alta no estaba cumplimentado eran superiores entre los accidentes graves: 39,5% en 1989, 37,2% en 1993 y 36% en 2000; y en estos se observó un incremento en el porcentaje de no cumplimentación a lo largo del año, siendo muy superior entre los accidentes graves ocurridos en diciembre: 61,4% en 1989, 61,2% en 1993 y 58,4% en 2000. Respecto a la incoherencia, el mayor porcentaje lo encontramos en la relación entre una antigüedad en el puesto mayor que la antigüedad en la empresa: 3,8% en 1989, 3,5% en 1993 y 6,1% en 2000. Asimismo, en el fichero de 1989 se encontraban 9.523 (1,4%) accidentes con baja correspondiente a un año diferente, en 1993 se encontraron 11.658 (2%) y 25.248 (2,4%) en 2000. En un menor porcentaje se encontraron 552 accidentes con baja correspondientes a trabajadores con más de 70 años de edad en 1989, 348 en 1993 y 1.136 en 2000.

Conclusiones: En general, y aunque la calidad formal de los datos de accidentes de trabajo con baja parece elevada, los indicadores estudiados indican un empeoramiento de la calidad formal. Y si bien, el porcentaje de no cumplimentación del motivo de alta de los accidentes graves ha ido disminuyendo, es llamativo que una tercera parte de éstos siga dejándose en blanco. El control de calidad de la cumplimentación del parte de accidente con baja y la relación de altas o fallecimientos es exigible a las instituciones responsables, especialmente a las Mutuas de Accidentes de Trabajo, ya que lo son en exclusiva de la cumplimentación de la relación de altas o fallecimientos.

267

ANÁLISIS DE LA VARIACIÓN ESPACIAL DEL CÁNCER DE RECTO-ANO EN GALICIA EN EL PERÍODO 1980-99

A. Figueiras, E. Lado Lema, M. Vacariza, J.A. Ferreira, R. Iglesias, M.C. Almiron, M.D. Gómez, J.J. Gestal, X. Hervada Vidal, E.M. Molanes López, M.I. Santiago Pérez y M.E. Lado Lema

Universidad de Santiago de Compostela, Área de Medicina Preventiva y Salud Pública, A Coruña. Dirección Xeral de Saúde Pública, Santiago de Compostela.

Objetivos: El análisis de la variación espacial de una enfermedad y su representación en un mapa se ha convertido en un tema importante en la investigación epidemiológica. Una medida habitualmente representada en el mapa es la razón de mortalidad estandarizada (RME), ya sea bruta o suavizada mediante técnicas bayesianas. En la construcción del mapa, el método de cloropetas es uno de los más comunes. Este método implica categorizar primero las áreas, según determinados percentiles de la distribución de las RME y, a continuación colorearlas según los colores asignados a sus categorías. Esta categorización en percentiles es bastante arbitraria ya que, no hay ningún criterio que determine si lo mejor es utilizar terciles, cuartiles o tal vez quintiles, pero además tampoco hay garantías de que esta clasificación facilite y permita la detección de áreas de alto y bajo riesgo. El objetivo de este estudio es el de ilustrar, mediante una aplicación práctica, el uso de un método bayesiano empírico y no paramétrico que facilite la construcción del mapa, eliminando la arbitrariedad presente en el método clásico de cloropetas, y proporcionando una estimación del verdadero número de grupos de riesgo en la población y de sus verdaderos riesgos relativos.

Métodos: El método supone que la población en estudio está formada por k subpoblaciones o grupos de riesgo con diferentes riesgos relativos $\theta_1, \dots, \theta_k$ y de modo que cada uno de ellos representa una proporción p_1, \dots, p_k del total de áreas. Hemos aplicado este método al estudio de la variación geográfica de la mortalidad en Galicia por cáncer de recto-ano a nivel de municipios desde 1980 a 1999. Los datos de mortalidad se obtuvieron del Instituto Nacional de Estadística para el período 1980-88 y del Registro de Mortalidad de Galicia para el período 1989-99, analizando la causa 154 de la CIE-9^{rev} (1980-98) y C19-C21 de la CIE-10^{rev} (1999). Los datos de población por grupos de edad, sexo y municipio se obtuvieron de los padrones de 1986 y 1996 y de los censos de 1981 y 1991. El intervalo de estudio se dividió en 4 períodos de 5 años cada uno (80-84, 85-90, 91-94, 95-99). Se ajustó el modelo para cada período y género.

Resultados: En el caso de la población total gallega en el 1º período se identificaron 3 grupos de riesgo con riesgos de magnitud 0,64, 1,21 y 2,32 y un porcentaje de representación en el total de áreas del 58%, 39% y 3%, respectivamente. En el 2º período se identificaron 3 grupos de riesgo con riesgos de magnitud 0, 0,98 y 1,83 y un porcentaje del 9%, 85% y 6%, respectivamente. En el 3º período se identificaron 3 grupos de riesgo con riesgos de magnitud 0,56, 1,04 y 1,53 y un porcentaje del 31%, 58% y 11%, respectivamente. En el 4º período se identificaron 2 grupos de riesgo con riesgos de magnitud 0,70 y 1,18 y un porcentaje del 45% y 55%, respectivamente. Los municipios identificados como áreas de alto riesgo en cada período han sido: Lugo, O Corgo, Quiroga y Leiro en el 1º, Lugo, Ourense, y Sanxenxo en el 2º, Lugo, O Saviñao, Monforte de Lemos, O Barco y Ourense en el 3º, mientras que en el 4º se observó un mapa bastante uniforme.

Conclusiones: Los resultados confirman una evolución a igualar los riesgos por municipios, en ambos sexos. Se destaca el municipio de Lugo por ser el único identificado como área de alto riesgo, en los 3 períodos de mayor riesgo. Este método permite la identificación de territorios de riesgo y la cuantificación del riesgo de cada uno. Uno de sus inconvenientes es que no tiene en cuenta la correlación espacial con las áreas contiguas.

266

FIABILIDAD DE LOS REGISTROS DE SUEÑO AMBULATORIO EN UNA POBLACIÓN ESCOLAR SANA MEDIANTE UN APARATO PROTOTIPO PORTÁTIL

M. Saez, J.J. Mazuecos, C. Campos, F. Caro y E.M. Andres

Universidad de Zaragoza, Departamento de Métodos Estadísticos, Zaragoza. Universitat de Girona, Departamento de Economía, Girona. Universidad de Zaragoza, Escuela de Ciencias de la Salud, Zaragoza. Hospital Miguel Servet, Zaragoza. Hospital Miguel Servet, Zaragoza.

Introducción: Muchas de las enfermedades de los distintos pacientes se ven reflejadas de alguna forma en su sueño, de forma que, una persona con algún tipo de trastorno físico y/o psíquico ve alterado sus parámetros de sueño. Hasta la actualidad, la mayoría de los registros de sueño se obtenían mediante la dependencia de una estructura hospitalaria. Hoy en día se comienza a utilizarse registros de sueño ambulatorio, es decir, que no necesita de la hospitalización del paciente para recoger los datos. El objetivo de este estudio es demostrar que la realización de los estudios en edad infantil (hasta los 15 años) se puede realizar de un modo válido y fiable mediante la utilización de un aparato portátil y sin la dependencia de una estructura hospitalaria. Para ello nos proponemos estudiar la validez de los registros de sueño con monitor portátil permitiendo valorar las fases de sueño lento, su duración (al igual que se obtiene con los registros polisomnográficos de los laboratorios del sueño) y además cuantificaremos la intensidad (potencia) de sueño lento. Para este estudio se ha cogido una muestra aleatoria e independiente de treinta y cuatro niños sanos (sin ningún tipo de trastorno) de edad escolar (entre tres y quince años) y se les ha tomado dos registros de sueño, en dos meses y días de la semana distintos (para evitar el efecto que podría producir este hecho en el sueño del niño).

Metodología: Para valorar la fiabilidad de los registros realizaremos un estudio de la variabilidad interindividual. Para ello, primeramente, hemos realizado un contraste de normalidad de las variables que incluyen los parámetros de sueño utilizando el estadístico de Lillieford; ya que si la variable sigue una distribución normal, para ver si existe diferencia entre los valores de ambos registros recogidos para cada uno de los individuos, utilizaremos el estadístico t -student (contrastando previamente la igualdad de varianzas con un test de Levene); y si no fuese normal, tendríamos que utilizar el test no paramétrico de Friedman. También hemos estudiado la variabilidad intersujeto. Para ello, realizamos un modelo ANOVA para el caso de unidades repetidas para detectar si existe efecto de la variable sexo. También observamos si había efecto del sexo y lugar de instalación con un modelo ANOVA bifactorial. Finalmente observamos que, aunque podría modelizarse el comportamiento de uno de los parámetros más importantes de sueño (potencia máxima delta) con un modelo de regresión polinómico, existía un cambio estructural según variaba la edad del individuo. Calculamos un modelo lineal para cada uno de los dos períodos hallados y realizamos el test de permanencia estructural de Chow.

Resultados: Estudio intrasujeto: después de realizar el test Lillieford tenemos que las variables latencia de sueño delta, tiempo de delta total, las potencias de sueño delta en cada uno de los canales, la potencia máxima de sueño delta y el porcentaje de delta en la noche son continuas y siguen una distribución normal con varianzas desiguales. Además no podemos rechazar la hipótesis de que ambos registros son iguales. En el otro grupo de variables, aquellas que no siguen una distribución normal, tiempo de sueño el día anterior, tiempo de sueño en el registro, tiempo de registro y tiempo delta en el primer ciclo de sueño, obtenemos que tampoco podemos rechazar la igualdad entre ambos registros. Estudio intersujeto: con los modelos Anova para unidades repetidas obtenemos que el sexo no es significativo en ninguno de los parámetros de sueño delta. Con los modelos ANOVA bifactorial obtenemos que sólo es significativo el lugar de colocación de los electrodos en los minutos de delta total y la interacción sexo-lugar de colocación en la variable minutos delta en el primer ciclo ($p < 0,05$). Finalmente ajustamos la potencia máxima delta con un modelo de regresión lineal, cuadrático y cúbico sin obtener resultados muy distintos (Multiple R: 0,76146, 0,77633, 0,77784 para los modelos lineales, cuadráticos y cúbicos, respectivamente). Localizamos un posible cambio estructural en los ocho años, para ello ajustamos dos modelos lineales y realizando el test de Chow obtuvimos un valor del estadístico de 2,518 con el que, al compararlo con una distribución F obteníamos que era significativo al 5% este cambio del modelo.

Conclusión: En resumen podemos concluir diciendo que los valores de las variables de sueño tomados en ambos registros son similares, lo que afirma que el aparato de registro ambulatorio da medidas fiables y tiene la ventaja de que no depende de una estructura hospitalaria. Además, ni el sexo ni el lugar de colocación de los electrodos son factores que influyen en los resultados del registro. Finalmente comentar que la potencia de sueño delta es distinta en edades tempranas (hasta 8 años) al resto de edad infantil (de 8 a 15 años).

268

ESTIMACIÓN DE LA SUPERVIVENCIA CON CENSURAS INFORMATIVAS

V. Navarro, X. Puig y J.R. González

Unidad de Investigación Clínica, Institut Català d'Oncologia, L'Hospitalet del Llobregat. Servei d'Informació i Estudis, Departament de Sanitat i Seguretat Social. Generalitat de Catalunya. Servei de prevenció del Càncer, Institut Català d'Oncologia, L'Hospitalet del Llobregat.

Antecedentes y objetivo: En el análisis de supervivencia es común encontrarse con observaciones censuradas por la derecha, bien por finalización del estudio bien por pérdida de contacto con los sujetos, llamadas censuras administrativas y abandonos respectivamente. La estimación de la curva de supervivencia por el método clásico de Kaplan-Meier asume la hipótesis que la distribución de la censura es independiente a la distribución de los tiempos de eventos del suceso de estudio, hipótesis que a menudo no se cumple. Por ello el objetivo de este trabajo es presentar el estimador propuesto por Frangakis y Rubin, para estimar correctamente la curva de supervivencia.

Métodos: Dada una cohorte, o un ensayo clínico, que ha sido objeto de seguimiento hasta una fecha final de estudio, se requiere estimar la curva de supervivencia. Si en nuestro estudio aparecen abandonos, probablemente sean censuras relacionadas con tiempos de eventos (informativas). Procederemos entonces en una segunda etapa a la realización de una muestra aleatoria simple de éstos, para recuperar la distribución de sus tiempos de supervivencia. A continuación estamos en condiciones de estimar correctamente la curva de supervivencia mediante el estimador propuesto por Frangakis y Rubin, basado en una nueva función de verosimilitud, que incorpora toda la información recogida en las dos etapas.

Resultados: Como ilustración presentamos las diferencias en la estimación de la curva de supervivencia, en diferentes tiempos y bajo ciertas condiciones de no independencia entre la distribución de eventos y censuras, mediante Kaplan-Meier y método Frangakis-Rubin, sobre datos simulados.

Conclusiones: El estimador Kaplan-Meier asume la hipótesis de independencia entre la distribución de la censura y la de los eventos, y en el caso de no cumplirse dicha hipótesis se convierte en una mala estimación de la curva de supervivencia. En esta situación se hace necesario la realización de una muestra aleatoria simple de los abandonos y utilizar el método de Rubin y Frangakis para la correcta estimación de la función de supervivencia. Este método está limitado a la calidad de la información recogida en la segunda etapa, por ello suele ser preferible centrar los esfuerzos en la búsqueda intensa de la información de los abandonos sobre una muestra aleatoria.

269

ANÁLISIS COMBINADO DE EFECTOS MÚLTIPLES: EFECTO DE LA CONTAMINACIÓN POR CO EN LA MORTALIDAD CONTROLANDO POR EL EFECTO DE OTRO CONTAMINANTE

S. Pérez-Hoyos, C. Iñiguez, F. Ballester, M. Sáez, J.V. García González, I. Galán, por Grupo EMECAM

*Unitat d'Epidemiologia i Estadística, Escola Valenciana d'Estudis per a la Salut (EVES), València. Grup de Recerca en Estadística, Economia Aplicada i Salut, (GRECS) Universitat de Girona, Girona. Consejería de Salud y Servicios Sanitarios, Gobierno Principado de Asturias, Gijón. Servicio Epidemiología, Consejería de Sanidad de la CAM, Madrid.***Antecedentes:** Las técnicas de meta-análisis univariantes se han utilizado para analizar el efecto de un contaminante en la salud a partir de los resultados observados en diferentes ciudades participantes en estudios multicéntricos. Sin embargo los contaminantes están presentes en el aire como una mezcla y no como compuestos aislados.**Objetivos:** Comparar diversos métodos para analizar el efecto combinado de la contaminación por CO en la mortalidad controlando por el efecto de otros contaminantes en cinco ciudades españolas del proyecto EMECAM.**Métodos:** Se examinó la relación entre el número de defunciones diarias por todas las causas (excepto las externas), enfermedades del aparato circulatorio y enfermedades respiratorias y los niveles diarios de CO en las cinco ciudades del proyecto EMECAM (Madrid, Barcelona, Valencia, Gijón y Oviedo) con datos disponibles del contaminante. Siguiendo una metodología estandarizada, basada en el proyecto EMECAS, y utilizando modelos autoregresivos generalizados aditivos de Poisson, se estimó el efecto del CO y sus retardos hasta orden 5 en cada ciudad. Se seleccionó el retardo de mayor efecto y se introdujo el mismo retardo de un segundo contaminante (NO₂, SO₂ y partículas en suspensión). Para evaluar el efecto combinado se efectuaron cuatro tipos de meta-análisis. En primer lugar, asumiendo la no existencia de correlación entre los efectos, se ajustó un modelo de análisis combinado de efectos fijos y otro de efectos aleatorios para cada uno de los dos efectos por separado (CO y segundo contaminante). En segundo lugar se consideró la existencia de correlación entre los efectos. Se asume que el vector Y de los dos efectos (CO y otro contaminante) en cada ciudad sigue una distribución normal multivariante con media el vector de efectos combinados y con varianza Si⁻¹D, donde Si es la matriz de varianzas de los efectos en cada ciudad y D es la varianza que mide la existencia de heterogeneidad entre ciudades. El efecto combinado se obtiene por mínimos cuadrados generalizados. Para el ajuste de estos modelos se utilizaron macros propias programadas en S-Plus al no estar disponibles en programas comerciales.**Resultados:** De los distintos retardos ajustados se seleccionó el retardo 0 (efecto del mismo día) por ser el más significativo en la mayoría de causas y ciudades. Para las tres causas analizadas se observó un efecto combinado significativo del CO al controlar por el segundo contaminante utilizando un modelo de efectos fijos o aleatorios por separado, con un riesgo relativo de 1.017 por mg con SO₂ y NO₂, y 1.015 por mg con partículas para el total de causas por ejemplo. Al tener en cuenta las correlaciones entre los efectos de los contaminantes no se encuentran grandes diferencias con el modelo separado a excepción del efecto con el SO₂ para el total de causas que pasa de 1.017.**Conclusiones:** El ajuste de modelos multivariantes de análisis combinado permite controlar la existencia de correlación entre los efectos a la hora de proporcionar un dato sobre el efecto de CO, teniendo en cuenta que los contaminantes no se encuentran aislados en la atmósfera.

271

PROBLEMAS EN LA SELECCIÓN DE UNA MUESTRA ADECUADA PARA ESTUDIAR LA PREVALENCIA DE RESISTENCIAS A FÁRMACOS ANTITUBERCULOSOS

A. Díaz, M. Díez, M.J. Bleda, por el grupo de trabajo de ResTB

*Unidad de Investigación en Tuberculosis, Centro Nacional de Epidemiología. ISCIII, Madrid.***Antecedentes y objetivos:** España ocupa el segundo lugar en incidencia de tuberculosis (TB) en la Unión Europea y en años recientes se han producido brotes de TB multirresistente. Pese a ello, no existe, en el ámbito nacional, información representativa sobre prevalencia de resistencia a fármacos antituberculosos de primera línea. Por ello se puso en marcha el Estudio de Resistencias a Fármacos Antituberculosos. En esta comunicación describimos las dificultades halladas para seleccionar una muestra adecuada.**Métodos:** Para poder comparar nuestros datos internacionalmente se siguió la metodología propuesta por la Organización Mundial de la Salud y la Unión Internacional contra la TB y las Enfermedades Respiratorias, que pone énfasis en tres aspectos: k control de calidad de los antibiogramas; b) separación correcta de los casos de TB iniciales y recurrentes; c) selección, mediante muestreo por conglomerados proporcional a la carga diagnóstica de los centros, de una muestra representativa de casos. Por razones logísticas se recomienda incluir sólo casos diagnosticados en centros públicos que, obviamente, deben tener un cultivo positivo para que el antibiograma pueda realizarse.**Resultados:** Al no existir en España centros específicos para TB se presentaron los siguientes problemas para seleccionar la muestra: a) falta de un listado de los centros que realizan cultivo de micobacterias: según estudios previos, el 95% de los cultivos de TB se realizan en centros hospitalarios, razón por la cual se decidió limitar el muestreo a estos. A partir del Catálogo Nacional de Hospitales, se identificaron los hospitales generales y geriátricos públicos (o privados con utilización pública), y mediante la información obtenida a través del Centro Nacional de Microbiología, de las Comunidades Autónomas (CCAA), o de la consulta telefónica con los hospitales, se averiguó qué hospitales realizaban cultivos o cual era su laboratorio de referencia en los que no disponían de este recurso; b) falta de información sobre carga diagnóstica en los centros: una vez obtenido el listado de 198 hospitales que configuran el marco muestral, se estimó la carga diagnóstica de cada uno mediante el Conjunto Mínimo Básico de Datos (CMBD); cuando no existía CMBD, el número de casos se imputó realizando modelos de regresión lineal (estratificados por tasa de TB de las CCAA) a partir del número de camas; c) diferencias en la carga diagnóstica de los centros: tras estos procesos, se observó que el 15% de casos de TB se diagnosticaban en hospitales con < 20 casos/año, lo que hacía peligrar la factibilidad del estudio puesto que, o bien no se podía garantizar la consecución del tamaño mínimo de casos elegido para el conglomerado, o, si se disminuía aquél, aumentaba desproporcionadamente el número de hospitales. Esto se solventó mediante un muestreo bietápico, estratificando por carga diagnóstica y con selección no equiprobable de los casos en cada estrato.**Conclusiones:** Al no existir en España centros específicos para la TB y faltar información sobre la carga diagnóstica en cada centro, la selección de una muestra representativa de casos resulta difícil, lo que hace necesario recurrir a un muestreo de diseño complejo con diferentes fases.

270

DETECCIÓN DE INDIVIDUOS DUPLICADOS EN UNA BASE DE DATOS SIN INFORMACIÓN NOMINAL

X. Puig, V. Navarro, R. de Luis y R. Gispert

*Servei d'Informació i Estudis, Departament de Sanitat i Seguretat Social, Barcelona. Unitat d'Investigació Clínica, Institut Català d'Oncologia, Hospital de Llobregat.***Antecedentes y objetivo:** En epidemiología es común encontrarse con bases de datos, en las que no existe la certeza de la no recurrencia, y en donde es difícil la identificación de registros duplicados por falta de variables identificativas. El objetivo de este trabajo es presentar un método, basado en las fechas de nacimiento, que permite validar o rechazar la hipótesis de la existencia de duplicados así como estimar el porcentaje que suponen.**Métodos:** Dado que la distribución de nacimientos no es uniforme a lo largo de un año, resulta acertado estimar la distribución teórica de probabilidad de duplicados de forma empírica a partir de una base de datos suficientemente grande en la que se tenga la certeza de que no contiene duplicados, como puede ser el caso de un registro de defunciones. Así podemos estimar el número esperado de observaciones con fechas coincidentes en la base de datos a explorar, bajo la hipótesis nula de no duplicación de registros. A continuación, mediante la comparación entre observados y esperados, con un test de homogeneidad usando el estadístico Chi-cuadrado, nos permite aceptar o rechazar la hipótesis nula de registros no duplicados, y en caso de rechazarla estimar el porcentaje de duplicados.**Resultados:** Presentamos la estimación de la distribución teórica calculada de forma empírica a partir de todas las defunciones ocurridas en Cataluña el año 1999. Para su validación calculamos los esperados para el año 1998 y aceptamos la hipótesis nula de no duplicados ($p = 0,993$). También utilizando la base de datos del conjunto mínimo básico de datos correspondientes a las altas hospitalarias, escogiendo un hospital con 5.611 altas, del período 1994-2000, del cual se sabe mediante el número de historia clínica que existen 28,1% de altas de individuos con repeticiones, y en este caso rechazamos la hipótesis nula ($p < 0,0001$), y estimamos el número de duplicados en un 27,53% (IC95% 26,21-28,85). A continuación repetimos el test eliminando las repeticiones de la base de datos, y aceptamos la hipótesis nula ($p = 0,899$).**Conclusiones:** En la mayoría de análisis se requiere independencia entre observaciones, la existencia de duplicados viola esta hipótesis, con lo que en caso de sospecha se puede validar la hipótesis, si se dispone al menos de la fecha de nacimiento. Este método requiere de la disponibilidad de las fechas de nacimiento, y de que la calidad de su notificación sea aceptable, y en consecuencia el método también puede ser utilizado para detectar fechas asignadas por defecto en caso de missings. Aunque presentamos una metodología basada en las fechas de nacimiento puede ser adaptada a otras variables en función del estudio.

272

ESTIMACIÓN DEL NÚMERO ESPERADO DE CASOS FALLECIDOS EN UNA COHORTE EN FUNCIÓN DE LOS GRUPOS CONTROL UTILIZADOS

R. Clèries, K. Langohr, J. Ribes, G. Gómez y F.X. Bosch

*Institut Català d'Oncologia (ICO); Departament d'Estadística i Investigació Operativa de la Universitat Politècnica de Catalunya (UPC).***Antecedentes:** Se ha estimado el riesgo de fallecer por diferentes causas en una cohorte de 2208 donantes de sangre portadores del Virus de la Hepatitis B (VHB) durante el período 1972-1996. Para dicha estimación se han utilizado dos grupos control: la población general de Catalunya (PGC) y un grupo de 15659 donantes de sangre (GDS) no portadores del VHB. Para la evaluación de dicho riesgo se ha utilizado la Razón de Mortalidad Estandarizada (RME) que compara el número de muertos observados (O) con el número de muertos esperados (E). Cuando la PGC era el grupo control los E se han calculado mediante las tasas de mortalidad de Catalunya, de las que no se disponía en el período 1972-1974, utilizándose para ese período las tasas de 1975. Cuando el grupo control era el GDS se han calculado los E a partir de las tasas de mortalidad en el GDS.**Objetivos:** 1) Estimación de las tasas de mortalidad en la PGC para cada causa de fallecimiento observada en la cohorte en el período 1972-1974 mediante un modelo estadístico. 2) Cálculo de E para cada causa de fallecimiento observada en la cohorte a partir de un modelo de supervivencia ajustado al GDS. 3) Comparar los resultados iniciales con los que se obtienen a partir de las estimaciones en 1) y 2).**Materiales y métodos:** Para la estimación de las tasas de mortalidad en la PGC se ha utilizado un modelo de Poisson (MP). En este modelo el número de muertos es una función del tamaño de población y del estrato, definiéndose este último a partir de las variables sexo, grupo de edad y período de calendario. Para cada causa de mortalidad en la cohorte se realizan varios ajustes de MP teniendo en cuenta los datos de la mortalidad de los años posteriores a 1974. Seleccionaremos el MP que en el ajuste haya utilizado un mayor número de años y que no presente sobredispersión. Por otro lado, cuando se realiza la comparación de la mortalidad de la cohorte con el GDS, a éstos se les ajusta un modelo de supervivencia teniendo en cuenta la edad, sexo y año de entrada en el estudio. Mediante la aplicación de este modelo a la cohorte se puede calcular la mortalidad esperada.**Resultados:** La utilización de tasas de mortalidad estimadas para los años 1972-1974 en la PGC no muestra diferencias estadísticamente significativas respecto a los resultados iniciales en el cálculo de los fallecimientos esperados. La RME para esta comparación es significativa sólo en las patologías referentes al hígado. Cuando el grupo control ha sido el GDS no se observan diferencias estadísticamente significativas entre los resultados iniciales y los obtenidos mediante el modelo de supervivencia.**Conclusiones:** 1) Los métodos utilizados no cambian las conclusiones iniciales del estudio. A pesar de esto, no podemos afirmar que el aumento de la varianza de la RME no afectase a los resultados en el caso de que el período en el que no se dispone la mortalidad fuese más largo o que la población de referencia fuese más pequeña. 2) Ajustar un modelo de supervivencia para el cálculo de E tiene sentido si el grupo control tiene un número de eventos (fallecimientos) elevado. Si no es así, es mejor ajustar el modelo de supervivencia a la cohorte, estimando los muertos esperados en el grupo control y calculando, en tal caso, el inverso de la RME.

273

CLASIFICACIÓN CON DATOS INCOMPLETOS: ESTIMACIÓN DEL ESTADIO DE UN TUMOR MEDIANTE REDES NEURONALES

R. Cleries, J. Ribes, V. Moreno, M. Navarro, E. Guino, L. Pareja y J.M. Borrás

Institut Català d'Oncologia (ICO); Ciutat Sanitària i Universitària de Bellvitge (CSUB).

Introducción: Se ha estimado la Supervivencia Relativa a los 5 años (SR) de 1.528 pacientes diagnosticados de cáncer colorrectal del Registro Hospitalario de Tumores (RHT) del ICO/CSUB en el período 1993-1997 según el estadio (Clasificación TNM) en el momento del diagnóstico. En 1169 casos (76,5%) la información del estadio fue completa, en 194 casos (12,7%) incompleta y en 165 casos (10,8%) no se pudo recuperar. De los casos con estadio incompleto se pudo revisar la historia clínica y completar la información referente al TNM.

Objetivos: 1) Determinar el estadio de los 194 pacientes con datos incompletos mediante un modelo de clasificación estadística y estimar la SR de los 1528 casos según el estadio. 2) Comparar los resultados de la clasificación de los 194 pacientes con los que se obtienen a partir de la revisión de historias clínicas. 3) Comprobar si la estimación de la SR a los 5 años es suficientemente robusta mediante un método de remuestreo.

Materiales y métodos: Para determinar el tipo de modelo estadístico que clasifique los datos incompletos se ha utilizado una red neuronal (RN) que permita definir para cada caso la probabilidad de que éste pertenezca a un estadio determinado, según las variables: TNM, edad, sexo, fecha de diagnóstico, fallecimiento y tiempo de seguimiento. El ajuste del modelo y la validación del mismo se realiza dividiendo la muestra en dos grupos, el de entrenamiento, que sirve para ajustar el modelo, y el de prueba. El modelo escogido será el que presente la menor tasa de datos mal clasificados en la muestra de prueba. Una vez escogido el modelo se hace la predicción del estadio, en forma de probabilidad, en los 194 casos con datos incompletos. Para evaluar la robustez de la estimación de la SR se ha aplicado el método bootstrap, mediante el cual se puede comprobar si la estimación de la SR obtenida mediante la clasificación es sesgada. Para aplicar este método efectuamos un muestreo con reemplazamiento de los individuos en cada estadio, repitiendo este proceso B = 400 veces. Para cada una de las 400 submuestras, se calcula la SR obteniendo un estimador bootstrap de la SR (media de la SR en las 400 submuestras).

Resultados: A partir del modelo de RN escogido se espera una tasa de datos mal clasificados cercana al 10%. La comparación de la clasificación de los 194 casos mediante RN con la clasificación real nos muestra que la tasa de datos mal clasificados es del 29%. A pesar de ello, los resultados de la SR por estadio que se obtienen entre la clasificación real y la obtenida mediante la RN no son estadísticamente diferentes, destacando las diferencias de la estimación puntual en el estadio IV en las mujeres (SR mediante RN: 5%, SR mediante clasificación real: 9%). Este grupo es el de menor tamaño, pero a pesar de ello, la estimación bootstrap de la SR nos indica que ésta no es sesgada, observándose este mismo hecho en los otros estadios.

Conclusiones: 1) A pesar de que la clasificación mediante RN no cambia las conclusiones del estudio se observa una tasa de datos mal clasificados elevada y esto podría afectar en el caso de que el número de casos sin clasificar fuese mayor. 2) El método bootstrap ha permitido reafirmar las conclusiones obtenidas en la estimación de la SR.

274

ESTUDIOS DE FIABILIDAD EN DOS PUBLICACIONES CIENTÍFICAS (JAMA Y MEDICINA CLÍNICA) DURANTE EL ÚLTIMO QUINQUENIO

P. Serrano, M. Ordobás y S.F. Rodríguez

Escuela Universitaria de Enfermería Puerta de Hierro UAM.; Servicio de Epidemiología. Instituto Salud Pública de la Comunidad de Madrid.

Antecedentes y objetivo: Garantizar la fiabilidad es un elemento imprescindible en la medición de la validez. El presente estudio ha pretendido valorar las estrategias metodológicas utilizadas para la medición de la fiabilidad externa en los estudios originales de dos publicaciones científicas [JAMA y Medicina Clínica (Med Clin)] que tuvieran como objetivo medir la fiabilidad externa.

Métodos: Estudio descriptivo retrospectivo para un período de 5 años (mayo de 1997 a diciembre de 2001). Se revisaron los artículos originales publicados durante el período descrito en las revistas JAMA y Med Clin. Para la selección de los artículos se utilizó como criterio el que estuviera manifestada la intención de los autores, a través del objetivo del estudio, de medir la fiabilidad externa, expresado con alguno de los siguientes términos: concordancia/agreement, discordancia/disagreement, fiabilidad/reliability, reproducibilidad/reproducibility, consistencia/consistence, repetibilidad/repeatability, precisión/accuracy. Cada artículo seleccionado fue revisado y discutido por todo el grupo investigador. Las variables medidas fueron: escala utilizada en la medición de la variable sobre la que se desea determinar la fiabilidad (cuantitativa/cualitativa), índice estadístico utilizado para ello, tipo de fiabilidad determinada (entre mediciones/ entre lectores/ambas). Para el tratamiento de los datos se llevó a cabo un análisis basado en índices de estadística descriptiva.

Resultados: Inicialmente fueron seleccionados 16 artículos de la revista JAMA y 5 de Med Clin. De JAMA fueron excluidos 9 de ellos, dado que en su objetivo, si bien figuraban alguno de los términos seleccionados, la pretensión era la medición de la validez (predictiva y/o de criterio). La relación según el año de publicación para Med Clin y JAMA fue la siguiente: 1997, 0/4; 1998, 1/1; 1999, 1/0; 2000, 1/0; 2001, 2/2. La escala utilizada para medir la variable fue: en 4 artículos cuantitativa, en 6 cualitativa, y en 2 ambas escalas. El tipo de concordancia medida fue: en 8 artículos medición entre lectores, en 3 entre mediciones y en 1 ambas. El indicador más utilizado fue el Coeficiente de Correlación Intraclase, seguidos de Kappa y Spearman.

Conclusiones: Existe una escasa producción de artículos originales cuyo objetivo principal sea la medición de la fiabilidad externa. Siendo ambas publicaciones semanales, el número de originales por revista es mayor en JAMA, pero no es mayor la proporción de los artículos seleccionados. Se aprecia un mayor predominio de escalas cualitativas para la medición de variables. Existe una amplia variedad de estrategia metodológicas para medir la fiabilidad, no siendo todas igual de idóneas.