

## Artículo de Investigación

### **¿Cómo se creó una metodología para aplicar el análisis estadístico implicative en las investigaciones clínicas?**

How was a methodology for applying the statistical implicative analysis in clinical investigations created?

Nelsa María Sagaró del Campo<sup>1\*</sup>. ORCID: <https://orcid.org/0000-0002-1964-8830>

Larisa Zamora Matamoros<sup>2</sup>. ORCID: <https://orcid.org/0000-0003-2210-0806>

<sup>1</sup>Universidad de Ciencias Médicas, Santiago de Cuba, Cuba.

<sup>2</sup>Universidad de Oriente, Santiago de Cuba, Cuba.

\*Autor para la correspondencia. Correo electrónico: [nelsa@infomed.sld.cu](mailto:nelsa@infomed.sld.cu)

## RESUMEN

**Fundamento:** Ante la necesidad del constante perfeccionamiento de los métodos de análisis e interpretación de los resultados en las investigaciones clínicas para identificar factores pronósticos se creó una metodología que permitió el empleo eficiente del análisis estadístico implicative en estas investigaciones.

**Objetivo:** Describir los métodos y procedimientos empleados en cada una de las fases de la investigación de obtención de la metodología propuesta.

**Metodología:** Se realizó una revisión documental de los métodos de validación apropiados, aprovechando las bases de datos bibliográficas de la Internet, lo cual se confrontó con la experiencia de las autoras para conformar el diseño y validación de la propuesta.

**Desarrollo:** La investigación se realizó en 5 etapas: la fundamentación teórica-práctica de la utilidad de este análisis, la elaboración de la propuesta, la validación por tres vías (de criterio, de contenido y un metaanálisis) y el diseño definitivo de la metodología.

**Conclusiones:** A pesar de la variedad de formas de validación empleadas, se continúa la aplicación de la metodología para detectar posibles desviaciones en el análisis e interpretación a solucionar en el futuro.

**DeCS:** ANÁLISIS ESTADÍSTICO; MÉTODOS DE ANÁLISIS/métodos; ESTUDIOS DE VALIDACIÓN; ESTUDIOS DE EVALUACIÓN COMO ASUNTO; METODOLOGÍA.

**Palabras clave:** Metodología; métodos; análisis estadístico implicative; validación; identificación de factores pronósticos; análisis estadístico; métodos de análisis; estudios de validación; estudios de evaluación como asunto.

## ABSTRACT

**Background:** Faced with the need to constantly improvement of the analytical methods and interpretation of results in clinical investigations to identify prognostic factors, a methodology that allowed the efficient use of statistical implicative analysis in these investigations was created.

**Objective:** To describe the methods and procedures used in each of the research phases to obtain the proposed methodology.

**Methodology:** A documentary review of the proper validation methods was carried out, taking advantage of the bibliographic databases from Internet, which was compared with the authors' experience to conform the design and validation of the proposal.

**Development:** The research was carried out in 5 stages: the theoretical-practical foundation of the usefulness of this analysis, the preparation of the proposal, the validation by three routes (criteria, content and a meta-analysis) and the final design of the methodology.

**Conclusions:** In spite of the variety of validation forms used, the application of the methodology continues to detect possible deviations in the analysis and interpretation to be solved in the future.

**MeSH:** STATISTICAL ANALYSIS; ANALYTICAL METHODS/methods; VALIDATION STUDIES; EVALUATION STUDIES AS TOPIC; METHODOLOGY.

**Keywords:** Methodology; methods; statistical implicative analysis; validation; identification of prognostic factors; statistical analysis; analysis methods; validation studies; evaluation studies as topic.

## INTRODUCCIÓN

Ante la necesidad del constante perfeccionamiento de los métodos de análisis e interpretación de los resultados de las investigaciones clínicas para identificar factores pronósticos se creó una metodología que permitió el empleo eficiente del análisis estadístico implicativo (ASI) en estas investigaciones.

Este tipo de análisis conocido por la sigla ASI de *Analyse Statistique Implicative* del idioma francés donde se originó, es una herramienta de la minería de datos basada en las técnicas estadísticas multivariadas, la teoría de la cuasi-implicación, la inteligencia artificial y el álgebra booleana, para modelar la cuasi-implicación entre los sucesos y variables de un conjunto de datos. <sup>(1)</sup>

Su creador, Régis Gras, <sup>(2-5)</sup> adoptó distintos enfoques teóricos para modelar la extracción y la representación del conocimiento en forma de reglas y metarreglas de inferencia que apuntan a la significación de un todo y no a la sumatoria de sus partes y su finalidad en principio fue resolver problemas de la didáctica de las matemáticas.

La aplicación empírica del ASI en las investigaciones clínicas de causalidad fue modelando las ideas de estas autoras para crear una metodología definitiva que permitiera la identificación de posibles factores causales en las investigaciones médicas, fue así que se evaluó su efectividad empleando la regresión logística binaria como estándar de oro en varios estudios para identificar factores pronósticos y de riesgo, los cuales constituyeron trabajos de terminación de la especialidad de los bioestadísticos graduados en la Universidad de Ciencias Médicas de Santiago de Cuba. <sup>(6-12)</sup>

También se realizaron estudios colaterales para apoyar la validez de los resultados encontrados con la aplicación de la metodología propuesta.

El objetivo de este trabajo es exponer los métodos empleados en cada una de las fases de la investigación de obtención de la metodología propuesta desde su concepción hasta la formalización de su empleo por los investigadores, para lo cual se realizó una revisión documental de los métodos de validación apropiados, aprovechando las bases de datos bibliográficas de la Internet, lo cual se confrontó con la experiencia de las autoras para conformar el diseño y validación de la propuesta.

## DESARROLLO

La presente investigación se desarrolló en las cinco etapas siguientes:

- 1. Fundamentación** teórica-práctica de la utilidad del análisis estadístico implicativo en los estudios de causalidad en las ciencias médicas que conllevó al reconocimiento de la necesidad de crear una metodología para su aplicación eficiente.
- 2. Elaboración** de una propuesta metodológica para la aplicación eficiente del análisis estadístico implicativo en los estudios de causalidad en las ciencias médicas.
- 3. Aplicación** de la propuesta metodológica. Se presenta como ejemplo de aplicación en el estudio de identificación de factores pronósticos de mortalidad en el cáncer de mama.
- 4. Validación** de la propuesta mediante tres formas diferentes:
  - a. Validación de criterio a través de estudios analíticos de tipo casos y controles clásicos o anidados en una cohorte, en los cuales se realizó la comparación de los resultados obtenidos tras la aplicación, sobre el mismo conjunto de datos, del análisis estadístico implicativo y la regresión logística binaria, seleccionada esta última, como estándar de oro, por ser la técnica habitualmente empleada en la identificación de factores pronósticos y de riesgo. Esta validación concluye con la evaluación del desempeño del análisis estadístico implicativo a través del cálculo de indicadores tales como: sensibilidad, especificidad, valores predictivos, cocientes de probabilidad, odds ratio diagnóstico, entre otros, con sus correspondientes intervalos del 95% de confianza.
  - b. Validación de contenido por tres vías:
    - I. Validez de apariencia, a través de la opinión de usuarios potenciales.
    - II. Validez racional, basada en la correspondencia de lo expresado en la literatura con lo encontrado tras la aplicación del análisis estadístico implicativo y verificado por una investigación documental.
    - III. Juicio de expertos para evaluar la calidad de las reglas derivadas del análisis estadístico implicativo.
  - c. Validación basada en la evidencia proporcionada por los 13 estudios realizados, en los cuales se ha empleado el análisis estadístico implicativo, según la metodología propuesta y se han comparado sus resultados con los de la regresión logística, según se describió en la primera opción de validación, a través de un metaanálisis.

## 5. Diseño definitivo de la metodología.

A continuación, se detalla cada una de estas etapas.

### Primera etapa: Fundamentación

En esta etapa se realizó una investigación documental con una exhaustiva revisión bibliográfica, donde se aplicaron los métodos teóricos de la investigación para comprender los fundamentos que sustentan los procedimientos de cada una de las etapas de la metodología propuesta y elaborar las recomendaciones para su empleo adecuado. A partir de esta revisión se elaboró un amplio marco teórico donde se abordó el estado del arte de las técnicas estadísticas para el estudio de la causalidad como objeto de estudio. En el mismo, partiendo de un análisis histórico lógico, se enunciaron cinco etapas en la evolución histórica de estas técnicas, se sistematizaron las técnicas para el estudio de las relaciones bivariadas y multivariadas, los métodos gráficos para la exploración, el análisis y la presentación de las relaciones multicausales, así como las técnicas que aseguran la validez de estas investigaciones, incluyendo el diseño y control de sesgos.

Luego se evidenciaron los hechos que hacen posible el empleo del análisis estadístico implicativo en la identificación de factores pronósticos y de riesgo, se realizó una comparativa entre esta técnica y la regresión logística binaria, como la mejor propuesta para dicha identificación, en la cual se tuvieron en cuenta 14 aspectos de orden teóricos y prácticos.<sup>(13,14)</sup>

También se revisaron las investigaciones de causalidad realizadas en las ciencias médicas en las cuales se empleó el análisis estadístico implicativo, lo que permitió determinar las regularidades y contradicciones que pautaron el diseño de la metodología y los cambios paulatinos en el análisis e interpretación de los resultados.

### Segunda etapa: Elaboración

Basado en los resultados de la etapa anterior, se contextualizó el modo de analizar los datos, y se adaptó la interpretación de los resultados de este análisis, identificando las primeras seis etapas de la metodología, según su finalidad, algunas de ellas con subetapas, y dentro de cada etapa o subetapa, los procedimientos a desarrollar para lograr el objetivo. El diseño de la metodología estuvo condicionado por el reconocimiento de las principales adecuaciones del análisis estadístico implicativo para ser empleado en las investigaciones médicas de causalidad que, según las autoras de esta investigación, pueden dividirse en dos vertientes fundamentales:

- Adecuaciones en el análisis de los datos, entre las que se destacan: la duplicación de la variable dependiente, que tradicionalmente es única con dos categorías y la obtención de dos grafos implicativos en modo cono, uno para las relaciones causales de los factores que influyeron en el desenlace peor, tomando como cono la variable dependiente de peor desenlace, y otro para los que influyeron en el mejor desenlace, tomado como cono la variable que representa el mejor desenlace.<sup>(15)</sup>
- Adecuaciones en la interpretación de los resultados, entre las que se destacan: la identificación de los posibles factores causales (a partir del modo cono del grafo implicativo), la identificación de posibles variables confusoras (mediante la presencia de dos nodos adyacentes unidos por un enlace bidireccional en el grafo implicativo, un cambio significativo (superior al 10 %) en la intensidad implicativa en el análisis al incluir o no la supuesta variable confusora y la obtención de reglas absurdas (sin sentido desde el punto de vista médico) en el árbol cohesitivo) y la creación de categorías que permiten clasificar los factores identificados en cuatro tipos de causas: cuasi-componentes, cuasi-suficiente, cuasi-necesaria y cuasi-equivalente (cuasi-suficiente y cuasi-necesaria a la vez).

### Tercera etapa: Aplicación

Se escogió como ejemplo para la aplicación un estudio de casos y controles para identificar factores pronósticos, cuyo diseño se presenta a continuación.

#### 1. Diseño y contexto del estudio

Se realizó un estudio observacional analítico prospectivo de tipo casos y controles anidado en una cohorte para la aplicación eficiente del análisis estadístico implicativo en la identificación de los factores pronósticos de mortalidad por una entidad en un territorio y momento determinados.

La entidad seleccionada fue el cáncer de mama, dada su elevada incidencia, prevalencia y letalidad en el mundo, el país y la provincia donde se realiza el estudio. Para verificar que este era un problema de salud y tener una primera aproximación que permitiera pasar a un estudio analítico se realizaron estudios descriptivos, de tendencia, supervivencia, estimación de los años

de vida potencialmente perdidos, entre otros, los cuales constituyeron tesis de la especialidad de Bioestadística y otros publicados, como el de Rodríguez y colaboradores.<sup>(16)</sup>

El periodo de reclutamiento se extendió de junio de 2014 a abril de 2019. Este periodo se estimó en base a la ocurrencia habitual de fallecidas a causa de esta entidad para garantizar un tamaño de muestra acorde con los requerimientos de la regresión logística que se empleó luego.

## 2. Universo y muestra

El universo de estudio quedó conformado por todas las mujeres mayores de 18 años de edad con el diagnóstico clínico e histológico de cáncer de mama, procedentes de la provincia de Santiago de Cuba, que fueron atendidas en el Hospital Oncológico “Conrado Benítez”, en el periodo antes mencionado, constituyendo los casos todas las fallecidas durante el periodo de estudio y los controles una muestra de las vivas en igual momento.

## 3. Elección de casos y controles

Al estar los casos y controles anidados en una cohorte cada caso fue seleccionándose a medida que ocurría un fallecimiento (caso incidente) y para aumentar la comparabilidad entre los grupos se escogió un control por cada caso, en el mismo momento, empleando un muestreo simple aleatorio a partir del listado de historias clínicas de pacientes con fecha de diagnóstico de no más o menos un mes de diferencia con respecto al caso actual, dada la existencia de este registro en el Departamento de Registros Médicos del hospital, lo cual facilitó la localización de las historias seleccionadas. Los controles que fallecieron en el periodo de estudio convirtiéndose en casos fueron sustituidos por otro control.

## 4. Criterios de exclusión

Se admitieron como criterios de exclusión el ser paciente fallecida por causa ajena a la enfermedad en estudio o paciente con enfermedad concomitante como Diabetes Mellitus, que pudiera interferir en la validez de los resultados del estudio.

## 5. Tamaño de muestra

Se tomó un control por cada caso para una razón 1:1, por lo que se empleó la fórmula de tamaño muestral para casos y controles balanceados,<sup>(17)</sup> cuyos parámetros se estimaron a partir de los resultados de dos estudios pilotos realizados;<sup>(8,12)</sup>

$$n_i = \left( \frac{r+1}{r} \right) \frac{(\bar{p})(1-\bar{p}) \left( Z_{1-\beta} + Z_{1-\frac{\alpha}{2}} \right)^2}{(p_1 - p_2)^2}$$

donde:

- $n_i$ : sujetos necesarios en cada una de las muestras considerando la  $i$ -ésima covariable significativamente asociada con la variable dependiente
- $r$ : razón de controles por caso
- $Z_{1-\frac{\alpha}{2}}$ : valor del  $\left(1 - \frac{\alpha}{2}\right)$  percentil de la  $N(0,1)$  en función de la seguridad o confianza según el riesgo admitido.  $Z_{1-\frac{\alpha}{2}} = 1.96$  para  $\alpha = 0.05$
- $Z_{1-\beta}$ : valor del  $(1 - \beta)$  percentil correspondiente al poder estadístico deseado.  $Z_{1-\beta} = 0.84$  para  $\beta = 0.2$
- $p_1$ : proporción en el grupo control que presentó el evento de interés
- $p_2$ : proporción en el grupo de casos que presentó el evento
- $p_1 - p_2$ : tamaño del efecto buscado
- $\bar{p}$ : media de las dos proporciones

Para conformar el tamaño de muestra de este estudio se consideraron como eventos de interés las covariables significativamente asociadas con la variable dependiente, identificadas como factores pronósticos en los dos estudios pilotos. Se aplicó la fórmula anterior para cada una de dichas covariables y se tomó, a partir de los resultados obtenidos, el máximo valor de las  $n_i$ . Para viabilizar este procedimiento se programó una función en el lenguaje R que devuelve el tamaño de muestra ( $n_i$  máximo) a emplear en un estudio con estas características.

También se tuvo en cuenta que el número de casos excediera en 10 el número de variables a incluir en el modelo de regresión para evitar pérdida de precisión en la estimación de los coeficientes de regresión siguiendo el criterio de Freeman, citado en Sagaró y Zamora, <sup>(14)</sup> como el más apropiado. Teniendo en cuenta estos elementos se obtuvo que la muestra estuviera constituida por 140 casos y 140 controles.

## **6. Operacionalización de las variables**

La selección de variables se realizó a partir de la revisión de la literatura, consulta a expertos y resultados de los dos estudios pilotos, <sup>(8,12)</sup> empleando 25 covariables, relacionadas con la paciente, el tumor, el tratamiento y el diagnóstico, cuya operacionalización y codificación en la base de datos se muestra en el Anexo 1.

## **7. Recolección de la información**

Se realizó una extensa búsqueda bibliográfica en las bases de datos biomédicas alojadas en Internet para determinar el estado actual del problema en estudio. Para conocer los datos estadísticos de morbilidad y mortalidad se consultaron los anuarios estadísticos de salud de la Dirección Provincial de Estadística y de la Dirección Nacional de Registros Médicos y Estadísticas de Salud del Ministerio de Salud Pública.

La recolección del dato primario se realizó mediante la revisión de las historias clínicas de los casos y los controles que integraron la muestra del estudio, los informes de anatomía patológica y otros documentos con información de las variables en estudio. Para este fin, se confeccionó un formulario que recoge la información necesaria. (Anexo 2).

## **8. Procesamiento de la información**

Se aplicó el análisis estadístico implicative según lo previsto en las seis etapas propuestas: análisis exploratorio, transformación de los datos duplicando la variable dependiente, análisis principal con variable dependiente como principal y suplementaria, presentación de los resultados, interpretación y análisis *a posteriori*.

## **9. Síntesis y redacción**

Para comunicar los resultados del estudio y conformar el artículo científico con los resultados de esta etapa se empleó la guía STROBE (*STrengthening the Reporting of Observational studies in Epidemiology*), que establece las directrices de comunicación para estudios observacionales. <sup>(18)</sup>

### **Cuarta Etapa: Validación**

La validación se realizó mediante tres estudios, considerando en cada uno de ellos una forma diferente de validación: validación de criterio, validación de contenido y metaanálisis.

#### **Primer estudio: Validación de criterio**

Para analizar las potencialidades de la técnica como un instrumento o medio para identificar posibles causas se decidió comparar la misma con un criterio externo. En este caso se consideró como la técnica estándar más empleada para estos fines según la literatura la regresión logística binaria; por lo que se aplicaron ambas técnicas al mismo conjunto de datos y se compararon sus resultados. La aplicación del análisis estadístico implicative se realizó siguiendo las seis etapas antes mencionadas y la aplicación de la regresión logística se detalla a continuación.

#### **1. Aplicación de la regresión logística**

Para la aplicación de la regresión logística binaria se verificó el cumplimiento de los supuestos de la prueba, eliminando del modelo las variables correlacionadas biológica o estadísticamente y las consideradas intervinientes.

El diagnóstico de la multicolinealidad se realizó a través del factor de inflación de la varianza, creando nuevas combinaciones siempre que fue posible u omitiendo la variable menos importante según la literatura y los expertos. Se comprobó el ajuste del modelo por la prueba de Hosmer y Lemeshow. Se empleó el método por pasos hacia adelante Wald ya que fue el que mostró un mejor ajuste de los datos al modelo.

Como criterios para el método por pasos se tomó una probabilidad de entrada de 0.05 y una de eliminación de 0.10; con un punto de corte de 0.5 y un máximo de 20 iteraciones. Se incluyó la constante en el modelo. Se interpretó la significación de los factores a través de la prueba de

Wald y del intervalo de confianza de las razones de ventajas representadas por las exponenciales de los coeficientes de regresión ( $\text{Exp}[\beta_i]$ ), donde  $\beta_i$  es el coeficiente asociado a la covariable  $i$ -ésima) con un 95 % de confianza.

## 2. Comparación del análisis estadístico implicativo con la regresión logística

Dada la respuesta dicotómica de estos procedimientos (identifica o no identifica) se construyó una tabla de contingencia de "2x2", en la cual se consideraron como individuos u objetos de estudio a las covariables empleadas en el mismo y como resultado la identificación de la variable como factor causal o no según ambas técnicas. Se calcularon los indicadores de eficacia de una prueba diagnóstica para el análisis estadístico implicativo como la sensibilidad, la especificidad, los valores predictivos positivo y negativo y las razones de verosimilitud positiva y negativa, los índices de Youden, de Validez y de Kappa y el odds ratio de diagnóstico, así como sus intervalos de confianza del 95 %, según la expresión de Wilson.<sup>(19,20)</sup> El índice de Youden se estimó según su forma clásica definida en 1950 y según la versión de Chen de 2015.<sup>(21)</sup>

Para el cálculo de los intervalos de confianza se programó una función en lenguaje R, teniendo en cuenta que se trabajó con un tamaño de muestra pequeño, representado por una cantidad de variables inferior a 30, y que los programas disponibles emplean fórmulas para su cálculo con aproximación a la distribución normal, no aplicable para muestras pequeñas.

### Segundo estudio: Validación de contenido

Esta validación se realizó por tres vías. Para evaluar si la técnica está a tono con los requerimientos actuales de la investigación se empleó la validez de apariencia y para evaluar si los resultados obtenidos de su aplicación, específicamente las reglas, reflejan verazmente lo que ocurre en la práctica médica se empleó la validez racional y el juicio de expertos.

#### Validez de apariencia

Esta vía se fue a través de la opinión de los usuarios potenciales de la metodología propuesta acerca de las características que justifican el empleo de la misma en las investigaciones médicas de causalidad. La misma se llevó a cabo en cuatro fases: preparatoria, de consulta, de análisis y conclusiva, las cuales se exponen a continuación.

1. **Preparatoria:** se definieron los objetivos y diseño de la investigación, se seleccionaron los usuarios potenciales, se creó el instrumento a aplicar y se definió cómo aplicarlo.

#### Diseño de la investigación

Se realizó un estudio observacional con la aplicación de técnicas cualitativas de tipo criterio de expertos, funcionando como expertos los usuarios potenciales

#### Universo de estudio y muestra

El universo estuvo constituido por los profesionales de la Bioestadística y la Informática de la provincia de Santiago de Cuba con cinco años y más de experiencia. Luego de la localización y confección del listado de posibles usuarios que actuaran como expertos se realizó un proceso de selección a través de la clasificación por aspectos con su puntuación correspondiente, quedando seleccionados los 30 profesionales de mayor puntuación entre los clasificados como muy competentes y con disposición a cooperar en la investigación.

#### Procedimiento para la selección de los usuarios

Al tratarse de usuarios que iban a funcionar como expertos, para su selección, se realizó una clasificación de los mismos (biograma sintético) en base a los siguientes aspectos:

- Especialidad: Bioestadística (10 puntos) e Informática (2 puntos)
- Años de experiencia: de 5 a 10 - (5 puntos), de 11 a 20 - (10 puntos) y más de 20 - (20 puntos)
- Dedicado al análisis estadístico de datos: Sí (10 puntos) o No (2 puntos)

Solo se seleccionaron aquellos que sumaron 25 puntos o más, para un total de 16 usuarios. Los aspectos para la clasificación y puntuación a otorgar se decidieron sobre la base de la experiencia de las autoras.

#### Características del instrumento

El instrumento a aplicar (Anexo 3) constó de dos preguntas, la primera sobre su grado de acuerdo con las consideraciones que se exponen acerca del análisis estadístico implicativo sobre si: aporta información interesante, es útil para la investigación, ayuda en la toma de decisiones, sus resultados son acertados y sus gráficos permiten una mejor comprensión del

problema estudiado. Estas afirmaciones fueron presentadas a través de una escala Likert de 5 puntos de acuerdo desde el acuerdo total al desacuerdo total, pidiéndoles que marcaran con una "X" la más conveniente.

En la segunda pregunta, con el objetivo de comparar el análisis estadístico implicativo y la regresión logística binaria, se les presentó un conjunto de aspectos y se le pidió que otorgaran una puntuación del 1 al 5 referente a cada técnica en todos los aspectos, siendo el 5 la mejor puntuación. Los aspectos considerados para cada técnica fueron si: está más tono con las exigencias de la investigación actual, el procesamiento computacional es fácil de realizar, aporta gran cantidad de información, el resultado es comprensible, brinda una excelente presentación desde el punto de vista gráfico.

2. **De consulta:** en la cual se realizó un taller con la participación de los usuarios potenciales de la metodología seleccionados en la etapa previa que permitió obtener la información requerida.

Para la obtención de la información se desarrolló el taller antes mencionado, en el cual se explicó a los usuarios las características de la investigación y lo relacionado con el análisis estadístico implicativo, su aplicación en las ciencias médicas y otros aspectos necesarios para poder responder a las preguntas del instrumento. Al final se entregó el instrumento donde estos profesionales emitieron su juicio.

3. **De análisis:** en la cual se procesaron las respuestas al instrumento y se resumieron e interpretaron los resultados obtenidos.

La primera pregunta se procesó a través del cálculo del índice de validez de contenido V, coeficiente propuesto por Aiken <sup>(22)</sup> entre 1980 y 1985, cuya fórmula es:  $V = \frac{\bar{x}-l}{k}$ , donde  $\bar{x}$  es la media de las calificaciones dada por los usuarios en la muestra,  $l$  es la calificación más baja posible, y  $k$  es el rango de los valores posibles de la escala Likert utilizada (diferencia entre calificación más baja y más alta). La calificación de cada ítem fue desde 5 puntos para el total acuerdo hasta 1 punto para el total desacuerdo. A partir del valor de este coeficiente se interpretó cada aspecto considerado sobre el análisis estadístico implicativo, según los estándares de Cicchetti de 1994, como aceptable cuando dicho valor estuvo por encima de 0.50 o por el más conservador de Charter de 2003 por encima de 0.70. En caso contrario el aspecto fue calificado de no aceptable. La significación de este coeficiente se probó a través de los intervalos de confianza al 95 %, los cuales se calcularon a través de las siguientes expresiones:

$$L = \frac{2nkV + z_{1-\frac{\alpha}{2}}^2 - z \sqrt{4nkV(1-V) + z_{1-\frac{\alpha}{2}}^2}}{2 \left( nk + z_{1-\frac{\alpha}{2}}^2 \right)}$$

$$U = \frac{2nkV + z_{1-\frac{\alpha}{2}}^2 + z \sqrt{4nkV(1-V) + z_{1-\frac{\alpha}{2}}^2}}{2 \left( nk + z_{1-\frac{\alpha}{2}}^2 \right)}$$

donde:

- L: límite inferior del intervalo
- U: límite superior del intervalo
- $z_{1-\frac{\alpha}{2}}$ : valor del  $1 - \frac{\alpha}{2}$  percentil de la distribución normal estándar
- V: valor del coeficiente V de Aiken
- $n$ : número de jueces
- $k$ : es el rango de los valores posibles de la escala Likert utilizada

A partir de estos intervalos se consideró un coeficiente estadísticamente significativo, en dependencia del estándar empleado, cuando el límite inferior estuvo por encima de 0.50 (Cicchetti) o de 0.70. (Charter).

Para la segunda pregunta se emplearon dos métodos considerados por las autoras, llamados A, basado en el promedio y B, basado en el coeficiente de concordancia de Kendall, los cuales se exponen a continuación.

Método A: se consideró como medida de resumen el promedio del puntaje dado en cada aspecto para cada técnica, con vistas a clasificar el aspecto a comparar, considerándolo excelente en caso de un promedio por encima de 4.5; bueno uno de 4.0 a 4.5; regular de 3.0 a 3.9 y malo, un puntaje promedio inferior a 3.0. Para identificar si existieron diferencias



significativas entre ambas técnicas se empleó la prueba de rangos con signos de Wilcoxon con una significación del 5 %

Se obtuvo además la estimación por intervalos de confianza del 95 % de las puntuaciones por aspectos a comparar en ambas técnicas, la cual se presentó en gráfico de barras de error apiladas.

Método B: Teniendo en cuenta que el análisis estadístico implicativo debe ser tan eficiente como la regresión logística se buscó el grado de acuerdo entre las opiniones de los usuarios sobre ambas técnicas para lo cual se empleó el coeficiente de concordancia de Kendall (W) al tratarse de datos ordinales probando la hipótesis de que hay concordancia significativa entre los rangos o calificaciones otorgadas por los usuarios que actúan como jueces. La interpretación del coeficiente también permitió conocer la fuerza de la concordancia. <sup>(23)</sup>

**4. Conclusiva:** se elaboró el informe ilustrado con tablas y gráficos.

#### **Validez racional**

Con el objetivo de verificar la correspondencia de lo expresado en la literatura con los resultados derivados de la aplicación del análisis estadístico implicativo se realizó una investigación documental, en la cual se desarrollaron las etapas propias de este tipo de investigación:

- 1. Diseño o plan de la investigación:** en esta etapa se elaboró un plan con todos los procedimientos a ejecutar. Primeramente, se acopiaron todas las reglas obtenidas a partir de los 3 estudios desarrollados sobre factores pronósticos en cáncer de mama, se revisaron sus interpretaciones según el enfoque del supuesto factor causal al desenlace y viceversa y se seleccionaron 23 reglas a criterio de las autoras, teniendo en cuenta la no duplicidad y la variabilidad en los contenidos de las mismas. Algunas de las reglas escogidas verifican la misma relación, pero fueron interpretadas de manera diferente con el objetivo de determinar cuál de las formas de interpretación es mejor entendida.
- 2. Búsqueda de información documental:** se realizó una amplia revisión bibliográfica aprovechando las bases de datos biomédicas alojadas en internet, con el empleo de los términos factores pronósticos y cáncer de mama en la estrategia de búsqueda en los idiomas español, inglés, francés y portugués.
- 3. Registro de información mediante técnica de fichero:** se elaboraron 23 fichas resumen donde se plasmó todo lo encontrado referente a las variables que se relacionan en cada regla. Las reglas se agruparon en tres categorías según su correspondencia con la literatura:
  - Correspondencia total: cuando la literatura plantea que la regla siempre se verifica en la práctica médica.
  - Correspondencia posible: cuando la regla puede verificarse o no en la práctica médica, pero nunca resulta absurda.
  - No correspondencia: cuando la regla nunca se verifica en la práctica médica porque es absurda.
- 4. Análisis de la información:** se aplicaron métodos teóricos para analizar los artículos encontrados sobre el tema e identificar congruencias e incongruencias con las reglas analizadas y se redactó una síntesis con los aspectos referentes con cada una de estas reglas que permitió el debate en torno a los resultados encontrados con la aplicación del análisis estadístico implicativo.
- 5. Elaboración del informe final:** se elaboró informe donde se presentaron los resultados mediante tablas y diagramas de barras simples y la discusión en relación con la veracidad de cada una de las reglas.
- 6. Presentación de resultados:** se publicó un artículo científico y se elaboró un trabajo de tesis.

#### **Juicio de expertos**

Esta vía, al igual que la anterior, tuvo por objetivo evaluar la calidad de las reglas derivadas de la aplicación del análisis estadístico implicativo, para lo cual se emplearon las 23 reglas antes seleccionadas.

#### **Diseño de la investigación**

Se realizó un estudio observacional con la aplicación de técnicas cualitativas de tipo criterio de expertos, siguiendo las cuatro fases empleadas para determinar la validez de apariencia antes presentadas.

#### **Universo de estudio y muestra**

El universo de estudio estuvo constituido por un grupo de expertos integrado por profesionales a cargo de la atención del cáncer de mama en la provincia de Santiago de Cuba con 5 años y

más de experiencia y que reunieran los requisitos de un buen experto como creatividad, colectivismo y espíritu crítico.

La muestra quedó integrada por los 33 expertos de mayor puntuación entre los clasificados como muy competentes y con disposición a cooperar en la investigación, según un proceso de selección.

### **Procedimiento para la selección de los expertos**

La selección se llevó a cabo luego de la localización y confección del listado de posibles expertos y obtención del consentimiento para su participación. Fueron considerados diferentes aspectos a los que se les otorgó determinada puntuación, en base a criterios de la autora. No se empleó la autovaloración para calcular el coeficiente de competencia experta en base a los coeficientes de conocimiento y argumentación porque en la experiencia de las autoras el método provoca un aumento en la tasa de abandono de los expertos en la fase posterior de la investigación.

Para la clasificación de los expertos en la atención a pacientes con cáncer de mama se tuvo cuenta los siguientes aspectos con su puntuación:

- Los años de experiencia como especialista de cualquier nivel:
  - 5 a 10 (2 puntos)
  - 11 a 20 (5 puntos)
  - más de 20 (10 puntos)
- El nivel de especialización:
  - Especialista básico: especialista de primer o segundo grado en Medicina General Integral (MGI), Medicina Interna o Ginecobstetricia. (2 puntos)
  - Especialista oncólogo: especialista de primer o segundo grado en Oncología. (5 puntos)
  - Superespecialista: especialista de primer o segundo grado en Oncología, Medicina Interna o Ginecobstetricia que está dedicado a la investigación y atención directa de pacientes con cáncer de mama por años y por lo cual son reconocidos entre sus colegas y pacientes. (10 puntos)

Se consideraron competentes para participar en el estudio solo aquellos expertos los que totalizaron 10 o más puntos. Esta indagación se hizo previa a la aplicación del instrumento y se corroboró por la respuesta del propio experto a los datos generales que pide el instrumento.

### **Obtención de la información**

La obtención de la información se realizó de manera individual, se explicó a cada experto las características de la investigación y de la técnica evaluada para aplicar luego una encuesta con dos preguntas, donde los expertos emitieron su juicio. (Anexo 4).

Con la primera pregunta se evaluó la validez del grafo implicativo en modo cono para identificar posibles factores causales, en este caso en particular, factores pronósticos de mortalidad en cáncer de mama. En esta pregunta se le presentaron dos opciones, los factores encontrados por la regresión logística binaria y los encontrados por el análisis estadístico implicativo, respectivamente, a seleccionar la que consideraran más apropiada. A los expertos no se les dio a conocer que técnica fue la que aportó los resultados a seleccionar. Los factores presentados se tomaron del estudio de Pardo,<sup>(12)</sup> donde hubo mayores diferencias entre los hallazgos de cada técnica, que en el presente trabajo.

Con la segunda pregunta se evaluó la validez de las reglas que se forman con la aplicación del análisis estadístico implicativo. En ella se les presentó, mediante una escala Likert de cinco puntos u opciones de respuesta, las 23 reglas antes seleccionadas y se les pidió que consideraran con qué frecuencia se cumplen las mismas en la práctica médica. Los cinco puntos de la escala fueron: se cumple siempre, se cumple frecuentemente, se cumple pocas veces, casi nunca se cumple, nunca se cumple.

### **Procesamiento de los datos**

En la primera pregunta se emplearon el número y el porcentaje para totalizar cuantas respuestas fueron a favor de los resultados del análisis estadístico implicativo y cuantas a favor de los resultados de la regresión logística binaria. Para decidir si hubo diferencias significativas en la selección de una u otra opción se aplicó la prueba de diferencia de proporciones con un nivel de significación del 5 %.

La segunda pregunta se procesó por el método Delphi. A partir de las respuestas de los expertos se conformó la tabla de frecuencias absolutas para las categorías de la escala

asumida, para cada una de las reglas; se determinaron las frecuencias relativas y relativas acumuladas, y se buscó la imagen de las frecuencias relativas acumuladas a través de la inversa de la función de distribución normal, para lo cual se tuvo en cuenta que eran más de 30 expertos.<sup>(24)</sup>

Se calcularon los promedios por filas (P), el promedio general (N) y la diferencia entre estos (N-P). Luego se establecieron los puntos de corte que se obtienen al dividir la suma de los valores correspondientes a cada columna entre el número de reglas en este caso (promedio relativo).

Por último, se determinó la categoría o el grado de adecuación de cada regla con la opinión de los expertos consultados, que se asignó según los siguientes valores:

- Los valores de N-P menores o iguales que el valor del primer punto de corte quedaron clasificados como: "Se cumple siempre".
- Los valores mayores que el valor del primer punto de corte y menores o iguales que el segundo se clasificaron como: "Se cumple frecuentemente".
- Siguiendo este procedimiento para los restantes puntos de cortes se estableció entre los puntos de corte 2 y 3 como: "Se cumple a veces".
- Entre los puntos de corte 3 y 4 como: "Casi nunca se cumple".
- Los valores mayores que el punto de corte 4 quedaron como: "Nunca se cumple".

Una vez concluidas la validación racional y el juicio de experto se efectuó una triangulación metodológica al comparar ambos resultados buscando la posible concordancia entre la literatura y los expertos (dos métodos diferentes para evaluar la veracidad de las reglas obtenidas) para lo cual se estableció una correspondencia entre las categorías establecidas en ambos métodos formando las tres nuevas categorías siguientes:

- I: si pertenece a la categoría "correspondencia total" de la validez racional o a las categorías "se cumple siempre" y "se cumple frecuentemente" del juicio de expertos.
- II: si pertenece a la categoría "correspondencia posible" de la validez racional o a las categorías "se cumple a veces" y "casi nunca se cumple" del juicio de expertos.
- III: si pertenece a "no correspondencia" de la validez racional o a "nunca se cumple" del juicio de expertos.

Se emplearon como medidas de resumen el número y el porcentaje de reglas en cada categoría y como medida del acuerdo el índice de Kappa ponderado y la prueba de 5 % de significación de la concordancia.

Esta triangulación permitió encontrar contradicciones literatura-experto que facilitaron la identificación de problemas en la redacción de las reglas, los cuales fueron reparados, pasando a la próxima ronda del criterio de expertos, en la cual se les presentó el resultado de la ronda anterior y las nuevas reglas, modificadas en su redacción. Nuevamente se procesó el juicio emitido por el método Delphi y se identificó la concordancia literatura-expertos. El criterio de parada para las rondas de expertos estuvo dado por la obtención de una concordancia excelente (superior a 0.75) entre los criterios de los expertos y lo planteado en la literatura.

### **Tercer estudio: Metaanálisis**

Con el objetivo de evaluar la efectividad de la metodología propuesta, basada en la evidencia de la efectividad que reportan las 13 investigaciones que la han empleado, se realizó un metaanálisis, donde se utilizaron métodos estadísticos para combinar los resultados de estos estudios, lo cual permitió combinar los resultados procedentes de todos los estudios y obtener un solo indicador para cada uno de los efectos (sensibilidad, especificidad y otros), aumentando la potencia estadística y la precisión de estos indicadores.

Para llevar a cabo el mismo se siguieron los procedimientos propuestos por el Manual Cochrane de revisiones sistemáticas de intervenciones Versión 5.1.0 actualizado en 2011 adaptados a las condiciones de este trabajo.<sup>(25)</sup>

### **Procedimientos llevados a cabo**

1. Establecimiento de la pregunta que se desea responder: la pregunta a responder fue si el efecto de la metodología propuesta es suficientemente grande para asimilar su introducción en los estudios de identificación de factores pronósticos y de riesgos.
2. Cuantificación de los efectos: el efecto de la metodología se cuantificó mediante los índices de rendimiento de un procedimiento diagnóstico al considerar el análisis estadístico implicative

como tal en la identificación de factores pronóstico o de riesgo dentro del conjunto de las covariables investigadas, tomando como principal indicador el odds ratio de diagnóstico.

3. Localización de los estudios de investigación: se realizó una búsqueda exhaustiva de los trabajos originales sobre el tema en las bases de datos biomédicas alojadas en la internet y las redes sociales, siendo los únicos trabajos encontrados con aplicación del análisis estadístico implicativo para la identificación de factores pronósticos y de riesgo, los realizados en la Universidad de Ciencias Médicas de Santiago de Cuba bajo la dirección de la Dra. Zamora.
4. Criterios de inclusión y exclusión de los estudios: se incluyeron todos los trabajos publicados y no publicados con diseño caso control para identificar factores de riesgo o factores pronósticos donde se aplica el análisis estadístico implicativo comparativamente con la regresión logística binaria.
5. Búsqueda de información y datos relevantes de cada estudio: no fue necesario crear una estrategia de búsqueda pues se conoce la localización de todos los artículos al formar parte de la producción científica de las autoras y previamente se había identificado la no existencia de otros estudios similares.
6. Evaluación de la calidad de los estudios incluidos: se verificaron criterios para clasificar el riesgo de sesgo en bajo, moderado o alto, a partir de una propuesta de la Guía STARD (*Standart for Reporting of Diagnostic Accuracy*) de EQUATOR (normas consolidadas para la publicación de estudios de evaluación de medios diagnósticos).<sup>(26)</sup>
7. Análisis de la heterogeneidad de los estudios: la homogeneidad de los estudios se garantiza por el tipo de diseño y otros factores que se han mantenido constantes en los mismos. No obstante, se midió con el test de homogeneidad de Cochran, los índices  $H^2$  e  $I^2$  y los gráficos de L'Abbé y de Galbraith previendo el tratamiento de la heterogeneidad con análisis en subgrupos y metarregresión.<sup>(27)</sup>
8. Combinación de resultados: se calculó el promedio de forma ponderada y no ponderada, ponderando por la inversa de la varianza con modelo de efecto fijo o usando un modelo de efectos aleatorios, para producir los resultados intermedios. Los odds ratio de diagnóstico se combinaron en un estimador global mediante la técnica de Mantel-Haenszel. Las parejas de proporciones de verdaderos positivos y falsos positivos permitieron conformar una curva SROC (acrónimo de *Summary Receiver Operating Characteristic*), a partir de pares de valores  $(D_i, S_i)$  donde:
  - $S_i$  es la suma de las transformaciones logit de la proporción de verdaderos positivos y falsos positivos para el estudio  $i$ -ésimo y se define como el 'umbral' de la prueba.
  - $D_i$  es la diferencia de las transformaciones logit de la proporción de verdaderos positivos y falsos positivos para el estudio  $i$ -ésimo.A partir de los pares antes mencionados se ajustó un modelo de regresión lineal simple, esto es,  $D_i = a + bS_i + e_i$ , donde  $a$  y  $b$  son los respectivos coeficientes de regresión.<sup>(28)</sup>
9. Identificación del sesgo de publicación: se realizaron las técnicas habituales como gráfico en embudo o funnel plot donde se representó el tamaño muestral de cada trabajo frente al tamaño del efecto detectado y las técnicas estadísticas de Begg y de Egger.<sup>(29)</sup>
10. Análisis de sensibilidad: se realizó el análisis separando los estudios de factores de riesgo y los de factores pronósticos y eliminando en cada paso un estudio diferente.
11. Actualización y mejora de la revisión: permanecerá vigente aún después de concluido este trabajo a raíz de los nuevos cambios.

### Presentación de los resultados

Los resultados se presentaron mediante los gráficos de bosque para cada uno de los efectos medidos, en el cual se muestran las estimaciones del efecto individual de cada estudio, el valor global obtenido al combinar todos los resultados y su intervalo de confianza. La redacción del artículo científico se hizo acorde con la declaración PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*), establecida como guía de literatura crítica para este tipo de estudio.<sup>(30)</sup>

### Quinta Etapa: Diseño definitivo

En el diseño definitivo de la metodología, tras la aplicación del análisis estadístico implicativo, se identificaron dos nuevas etapas que se describen a continuación.

- **Selección de las variables a incluir en el modelo de regresión logística binaria:** Elegir primero las variables que se enlazan con el cono de los dos grafos implicativos:

- con mayor intensidad implicativa
  - con enlace continuo (relación directa)
- **Discusión de los resultados:** Contrastar con los resultados obtenidos por la regresión logística y comparar con los estudios anteriores

### **Sistemas informáticos y procesadores estadísticos empleados**

Para el procesamiento y presentación de los resultados en las diferentes etapas de la investigación se emplearon los siguientes sistemas:

- Microsoft Excel del paquete Office 2016, para confeccionar la base de datos en formato .xlsx para exportar al SPSS y en formato .csv como condición necesaria para la aplicación del CHIC y para el procesamiento de los instrumentos aplicados en la validación de contenido y programar el cálculo del coeficiente de Aiken y sus intervalos, el procesamiento por el método Delphi y otras fórmulas necesarias.
- IBM SPSS Statistics 24.0, para obtener el modelo basado en la regresión logística.
- CHIC 5.0, para obtener el modelo basado en el análisis estadístico implicativo, así como los gráficos ilustrativos de la técnica.
- R 3.2.0 para efectuar el metaanálisis y la presentación de los resultados numéricos y gráficos del mismo, con el empleo de las librerías apropiadas (meta, metafor, compute.es, copas) y para la programación de las funciones que calculan los indicadores del desempeño del análisis estadístico implicativo comparado con la regresión logística binaria, teniendo en cuenta que todos los programas que calculan los intervalos de confianza de estos indicadores lo hacen con la aproximación a la normal, válido para muestras grandes y en este caso se necesitó la transformación de Wilson, y el RStudio 1.0.44 como entorno de desarrollo integrado sobre el cual se corrió el R.

### **Limitaciones del estudio**

Los factores pronósticos clásicos se pueden obtener con un estudio histopatológico convencional. Sin embargo, desde hace algunos años, van apareciendo nuevos marcadores pronósticos relacionados con la estructura y el ciclo celular o mutaciones genéticas que requieren tecnologías de avanzada para su detección, como las técnicas inmunohistoquímicas.

Al no disponer de todos los resultados de los biomarcadores se decidió definir una variable explicativa que evaluara el tener o no la determinación, pues en caso contrario se hubiese tenido que descartar algunos casos y controles que no tenían el resultado de estos biomarcadores, con la consecuente reducción del tamaño de muestra y la imposibilidad de aplicar la regresión logística. Solo en la sexta etapa de análisis *a posteriori* se trabajó con los casos que tenían esta determinación para ver la influencia del subtipo molecular como factor pronóstico. Este inconveniente pudiera haber interferido en la precisión de los factores pronósticos identificados, pero no influyó en la demostración de la eficacia de la metodología propuesta que es el objetivo principal de este trabajo.

### **Bioética Médica**

El presente estudio se articuló sobre los principios fundamentales que regulan la conducta ética médica, dispuestos en la Declaración de Helsinki de la Asociación Médica Mundial.<sup>(31)</sup> El diseño de la investigación e inclusión de variables fue consultado con el jefe provincial del programa de cáncer y los facultativos relacionados con la atención a las pacientes objeto de estudio. No se recogió el consentimiento informado por no tratarse de un experimento. Los datos obtenidos fueron estrictamente confidenciales y no serán utilizados con fines ajenos a esta investigación. Antes de iniciar la investigación se elaboró el proyecto, el cual fue aprobado por las autoridades competentes de los centros donde se desarrolló el estudio, los consejos científicos y los comités de ética de la investigación en salud de los niveles correspondientes.

### **Control de la calidad**

Para tener certeza de la calidad de los datos empleados se determinó la calidad de todas las historias clínicas, informes de anatomía patológica y registros involucrados en el estudio como parte de un trabajo de terminación de especialidad en Bioestadística defendido en la Universidad de Ciencias Médicas de Santiago de Cuba, en noviembre de 2019. En los casos de historia defectuosa se sustituyó la unidad de análisis por otra cuya historia tuviera una calidad demostrada.

Se recomienda consultar los artículos de la bibliografía que explican en detalles y ejemplifican cómo aplicar este análisis de manera eficiente según las consideraciones de las autoras, lo cual hará posible que se sumen investigadores a su empleo.

## **CONCLUSIONES**

Como parte de la elaboración de la metodología propuesta se emplearon técnicas para su validación y se continúa la aplicación de la misma para detectar posibles desviaciones en el análisis e interpretación a solucionar en el futuro.

## REFERENCIAS BIBLIOGRÁFICAS

1. Valls Pla X, Gregori Huerta P. Diseño de un paquete R para el Análisis Estadístico Implicativo [Tesis para Máster en Matemática Computacional Internet]. España: Universitat Jaume I; 2014. [citado 10 Ene 2019]. Disponible en: [http://repositori.uji.es/xmlui/bitstream/handle/10234/107441/TFM\\_2013\\_vallsX.pdf?sequence=1&isAllowed=y](http://repositori.uji.es/xmlui/bitstream/handle/10234/107441/TFM_2013_vallsX.pdf?sequence=1&isAllowed=y)
2. Gras R, Suzuki E, Guillet F, Spagnolo F. Statistical Implicative Analysis, Theory and Applications [Internet]. Nueva York: Springer; 2008. [cited 2019 Jan 10]. Available from: [https://books.google.com/cu/books?id=\\_AltCQAAQBAJ&hl=es&source=gbs\\_book\\_other\\_versions](https://books.google.com/cu/books?id=_AltCQAAQBAJ&hl=es&source=gbs_book_other_versions)
3. Guillet F, Gras R, Régnier JC, Marinica C, Couturie R. L'analyse Statistique Implicative. Méthode exploratoire et confirmatoire à la recherche de causalités [Internet]. Toulouse: Cepadues; 2013. [cited 2019 Jan 10]. Available from: [https://www.researchgate.net/publication/236005381\\_L'Analyse\\_Statistique\\_Implicative\\_Methode\\_exploratoire\\_et\\_confirmatoire\\_a\\_la\\_recherche\\_de\\_causalites/link/5a7ad747aca2722e4df584c4/download](https://www.researchgate.net/publication/236005381_L'Analyse_Statistique_Implicative_Methode_exploratoire_et_confirmatoire_a_la_recherche_de_causalites/link/5a7ad747aca2722e4df584c4/download)
4. Gras R. Genese et developpement de l'analys estatistique implicative: retrospective Historique. Educ Matem Pesq São Paulo [Internet]. 2014 [cited 2019 Jan 20];16(3):645-61. Available from: [https://revistas.pucsp.br/emp/article/view/21541/pdf\\_1](https://revistas.pucsp.br/emp/article/view/21541/pdf_1)
5. Régnier JC, Gras R, Lahanier-Reuter D, Marinica C, Guillet F. L'Analyse Statistique Implicative. Des Sciences dures aux Sciences Humaines et Sociales [Internet]. 3ra éd. Toulouse: Cépaduès; 2017. [cited 2019 Jan 20]. Available from: <http://critiquelittera.info/lanalyse-statistique-implicative-des-sciences-dures-aux-sciences-humaines-et-sociales-pdf-et-epub-rgis-gras.pdf>
6. García Mederos Y, Zamora Matamoros L, Sagaró del Campo N. Análisis estadístico implicativo en la identificación de factores de riesgo en pacientes con cáncer de pulmón. Medisan [Internet]. 2015 [citado 20 Ene 2019];19(8):947 Disponible en: <http://scielo.sld.cu/pdf/san/v19n8/san03198.pdf>
7. Moraga Rodríguez A, Zamora Matamoros L, Sagaró del Campo NM, Moraga Rodríguez A, Rodríguez Griñán A. Análisis estadístico implicativo para la identificación de factores pronósticos de la mortalidad por cáncer de pulmón. Medisan [Internet]. 2016 [citado 15 Ene 2019];20(3):344-53. Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1029-30192016000300010](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192016000300010)
8. Moraga Rodríguez A, Zamora Matamoros L, Sagaró del Campo NM, Moraga Rodríguez A, Rodríguez Griñán A. Análisis estadístico implicativo para la identificación de factores pronósticos de la mortalidad por cáncer de mama. Medisan [Internet]. 2017 [citado 15 Ene 2019];21(4):413. Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1029-30192017000400003](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192017000400003)
9. Moraga Rodríguez A, Zamora Matamoros L, Sagaró del Campo NM, Moraga Rodríguez A, Rodríguez Griñán A. Análisis estadístico implicativo para la identificación de factores pronósticos de la mortalidad por cáncer de próstata. Medisan [Internet]. 2018 [citado 15 Ene 2019];22(1):48-56. Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1029-30192018000100007](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192018000100007)
10. Paez Candelaria Y, Sagaró del Campo NM, Zamora Matamoros L. Análisis estadístico implicativo en la determinación de factores pronósticos del estado nutricional del paciente grave al egreso. Medisan [Internet]. 2018 [citado 20 Ene 2019];22(6):431-40. Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1029-30192018000600007](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192018000600007)
11. Galano Vázquez K, Sagaró del Campo NM, Zamora Matamoros L, Lambert Matos Y, Mingui Carbonell E. Análisis estadístico implicativo en la identificación de factores pronósticos de mortalidad del cáncer renal. Rev. inf. cient. [Internet]. 2018 [citado 14 May 2019];98(2). Disponible en: <http://www.revinfscientifica.sld.cu/index.php/ric/article/view/2268>
12. Pardo-Santana S, Sagaró-del-Campo NM, Zamora-Matamoros L, Viltre-Castellanos DM. Utilidad del análisis estadístico implicativo para identificar factores pronósticos en pacientes con cáncer de mama. Rev electron Zoilo [Internet]. 2019 [citado 2 Abr 2019];44(4):[aprox.9p.]. Disponible en: [http://revzoilomarinaldo.sld.cu/index.php/zmv/article/view/1869/pdf\\_608](http://revzoilomarinaldo.sld.cu/index.php/zmv/article/view/1869/pdf_608)
13. Sagaró Del Campo NM, Zamora Matamoros L. ¿Por qué emplear el análisis estadístico implicativo en los estudios de causalidad en salud? Rev Cubana Informát Méd [Internet]. 2019

- [citado 20 Jun 2019];11(1):88-103. Disponible en: <http://scielo.sld.cu/pdf/rcim/v11n1/1684-1859-rcim-11-01-88.pdf>
14. Sagaró Del Campo NM, Zamora Matamoros L. Análisis estadístico implicativo versus Regresión logística binaria para el estudio de la causalidad en salud. *Multimed* [Internet]. 2019 [citado 20 Jun 2019];23(6):1416-40. Disponible en: <http://scielo.sld.cu/pdf/mmed/v23n6/1028-4818-mmed-23-06-1416.pdf>.
  15. Sagaró Del Campo NM, Zamora Matamoros L. ¿Cómo aplicar el análisis estadístico implicativo en los estudios de causalidad en salud? *Revista Electrónica Dr. Zoilo E. Marinello Vidaurreta*. 2020[citado 26 dic 2019];45(1). Disponible en: <http://revzoilomarinello.sld.cu/index.php/zmv/article/view/196>
  16. Rodríguez Sarria YM, Delisle Ureña G, Sagaró del Campo NM, Escalona Fonseca M. Factores pronósticos y supervivencia de mujeres con cáncer de mama en Santiago de Cuba. *Medisan* [Internet]. 2018 [citado 10 May 2019];22(5):485. Disponible en: <https://www.medigraphic.com/pdfs/medisan/mds-2018/mds185d.pdf>
  17. Gastón Duffau T. Tamaño muestral en estudios biomédicos. *Rev Chil Pediatr* [Internet]. 1999 [citado 8 May 2019];70(4):314-24. Disponible en: [https://scielo.conicyt.cl/scielo.php?script=sci\\_arttext&pid=S0370-41061999000400009](https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0370-41061999000400009)
  18. Erik von Elm E, Altman DG, Egger M, Pocock SJ, Gotsche PC, Vandenbroucke JP. Declaración de la Iniciativa STROBE (Strengthening the Reporting of Observational studies in Epidemiology): directrices para la comunicación de estudios observacionales. *Gac Sanit* [Internet]. 2008 [citado 8 May 2019];22(2):144-50. Disponible en: [http://scielo.isciii.es/pdf/gsv/v22n2/articulo\\_especial.pdf](http://scielo.isciii.es/pdf/gsv/v22n2/articulo_especial.pdf)
  19. Olvera Porcel MC. Coeficiente Kappa promedio: un nuevo parámetro para evaluar y comparar el rendimiento de test diagnósticos binarios. [Tesis Doctoral en Estadística e Investigación Operativa Internet]. Granada: Universidad de Granada; 2015. [citado 10 May 2019]. Disponible en: <https://hera.ugr.es/tesisugr/24935256.pdf>
  20. Montero Alonso MA. Intervalos de confianza y contrastes de hipótesis para parámetros de test diagnósticos binarios. [Tesis doctoral Internet]. Granada: Universidad de Granada;2010. [citado 10 May 2019]. Disponible en: [https://www.researchgate.net/publication/46593163\\_Intervalos\\_de\\_confianza\\_y\\_contrastes\\_de\\_hipotesis\\_para\\_parametros\\_de\\_tests\\_diagnosticos\\_binarios/link/02bfe50f1e8bbd49bc00000/download](https://www.researchgate.net/publication/46593163_Intervalos_de_confianza_y_contrastes_de_hipotesis_para_parametros_de_tests_diagnosticos_binarios/link/02bfe50f1e8bbd49bc00000/download)
  21. Estrada Álvarez JM. El índice de Youden y su aplicación a la determinación del punto de corte en un test cuantitativo. [Tesis de Máster en Estadística Aplicada Internet]. Granada: Universidad de Granada; 2016. [citado 10 May 2019]. Disponible en: <https://docplayer.es/80683909-El-indice-de-youden-y-su-aplicacion-a-la-determinacion-del-punto-de-corte-en-un-test-cuantitativo.html>
  22. Merino Soto C. Confidence interval for difference between coefficients of content validity (Aiken's V): A SPSS syntax. *An. psicol.* [Internet]. 2018[cited 2019 Apr 20];34(3):587-90. Disponible en: <https://revistas.um.es/analesps/article/view/analesps.34.3.283481/232241>
  23. Picado-Alvarado F. Análisis de concordancia de atributos. *Tecnología en Marcha* [Internet]. 2008 [citado 6 May 2019];21(4):29-35. Disponible en: [https://revistas.tec.ac.cr/index.php/tec\\_marcha/article/view/222/219](https://revistas.tec.ac.cr/index.php/tec_marcha/article/view/222/219)
  24. Linstone HA, Turoff M. *The Delphi Method Techniques and Applications* [Internet]. Southern California: University of Southern California; 2002. [citado 6 May 2019]. Disponible en: <https://web.njit.edu/~turoff/pubs/delphibook/delphibook.pdf>
  25. Higgins JPT, Green S. *Manual Cochrane de Revisiones Sistemáticas de Intervenciones*, versión 5.1.0 [Internet]. Barcelona: Centro Cochrane Iberoamericano; 2012. [Actualizada Mar 2011; citado 6 May 2019]. Disponible en: [https://es.cochrane.org/sites/es.cochrane.org/files/public/uploads/Manual\\_Cochrane\\_510\\_reduit.pdf](https://es.cochrane.org/sites/es.cochrane.org/files/public/uploads/Manual_Cochrane_510_reduit.pdf)
  26. González Rodríguez MP, Velarde Mayol C. Lista de comprobación de estudios sobre precisión de pruebas diagnósticas: declaración STARD. *Evid Pediatr* [Internet]. 2012 [citado 6 May 2019];8:43. Disponible en: <https://evidenciasenpediatria.es/files/41-11628-RUTA/43Fundamentos.pdf>
  27. Molina Arias M. El metaanálisis de pruebas diagnósticas. *Rev Pediatr Aten Primaria* [Internet]. 2015 [citado 20 Abr 2019];17(67):281-5. Disponible en: [http://scielo.isciii.es/pdf/pap/v17n67/23\\_lectura-critica.pdf](http://scielo.isciii.es/pdf/pap/v17n67/23_lectura-critica.pdf)



28. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc* [Internet]. 2009 [cited 2019 Apr 20];172(1):137-59. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2667312/>
29. Bolaños Díaz R, Calderón Cahua M. Introducción al meta-análisis tradicional. *Rev Gastroenterol Peru* [Internet]. 2014 [citado 20 Abr 2019];34(1):45-51. Disponible en: <http://www.scielo.org.pe/pdf/rgp/v34n1/a07v34n1.pdf>
30. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* [Internet]. 2009 [cited 2019 Jul 21];6(7):e1000100. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2707010/>
31. World Medical Association. Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* [Internet]. 2000 [cited 2019 Apr 20];284(23):3043-5. Available from: <https://jamanetwork.com/journals/jama/article-abstract/193387>

## ANEXOS

### Anexo 1. Operacionalización de las variables para la identificación de factores pronósticos de mortalidad por cáncer de mama.

Variable (Nombre en la bases de datos)	Categoría	Definición operacional de las categorías
<b>Variable de respuesta:</b>		
Estado	Viva Fallecida	Según el estado vivo o fallecido por causa de la enfermedad al término del estudio.
<b>Variables explicativas:</b>		
<b>Factores dependientes del paciente</b>		
1. Menor de 50 años (edad)	Sí No	Según tenga o no menos de 50 años cumplidos al momento del diagnóstico.
2. Premenopausia (menopau)	Sí No	Según se realizó o no el diagnóstico en la etapa premenopáusica.
<b>Factores dependientes del tumor</b>		
3. Tipo histológico (tipo histol)	Peor Mejor	Los tipos ductal infiltrante, lobulillar, mixto corresponden a la categoría peor, y los tipos papilar, medular o mucinoso a la mejor.
4. Grado de diferenciación histológica (grado histol)	Peor Mejor	Según la modificación de Elston-Ellis del sistema de graduación Bloom-Richardson, b se consideraron los de mayor grado (pobremente diferenciados) como peor y los grados I y II (bien y moderadamente diferenciados) como mejor.
5. Grado de diferenciación nuclear (grado nucle)	Peor Mejor	Los grados III, IV corresponden a peor y I, II al mejor.
6. Estadio (estadio)	Peor Mejor	De acuerdo con la clasificación TNM de la <i>American Joint Committee on Cancer (AJCC)</i> <sup>c</sup> se consideraron: - como peor los estadios siguientes: • IIIA (T0N2M0 o T1N2M0 o T2N2M0 o T3N1M0 o T3N2M0) • IIIB (T4N0M0 o T4N1M0 o T4N2M0) • IIIC (Cualquier TN3M0) • IV (Cualquier T, cualquier N y M1) - como mejor: • 0 (TisN0M0) • I (T1N0M0) • IIA (T0N1M0 o T1N1M0 o T2N0M0) • IIB (T2N1M0 o T3N0M0)
7. Tamaño tumoral (tamaño)	Peor Mejor	Se consideró peor un tumor mayor o igual a 5 cm (T <sub>3</sub> o T <sub>4</sub> ) y mejor uno menor de 5 cm (T <sub>0</sub> , T <sub>1a</sub> , T <sub>1</sub> o T <sub>2</sub> ).
8. Afectación ganglionar (ganglio)	Sí No	Según la presencia (N <sub>1</sub> a N <sub>3c</sub> ) o no (N <sub>0</sub> ) de metástasis de los ganglios linfáticos regionales.
9. Metástasis (meta)	Sí No	Según la evidencia (M <sub>1</sub> ) o no (M <sub>0</sub> ) de metástasis a distancia.
10. Invasión sanguínea (inv. sangre)	Sí No	Según presente o no invasión sanguínea.
11. Invasión linfática (inv. linfa)	Sí No	Según presente o no invasión linfática.
12. Invasión perineural (inv. neura)	Sí No	Según presente o no invasión perineural.
13. Invasión muscular (inv. muscul)	Sí No	Según presente o no invasión muscular.
14. Invasión del tejido adiposo (inv. adipo)	Sí No	Según presente o no invasión del tejido adiposo.
15. Invasión de piel y pezón (piel pezón)	Sí No	Según presente o no invasión de piel y pezón.
16. Necrosis (necrosis)	Sí No	Según presente o no necrosis.
17. Desmoplasia (desmopla)	Sí No	Según presente o no desmoplasia.
18. Histiocitosis ganglionar (histiocito)	Sí No	Según presente o no histiocitosis ganglionar.
19. Reacción inflamatoria (reac. infla)	Sí No	Según presente o no una reacción inflamatoria.
20. Margen quirúrgico libre (Margen)	Sí No	Según presente o no distancia entre el borde del tumor y la tinta china.
<b>Factores dependientes del tratamiento</b>		
21. Cirugía (cirugía)	Sí No	Según reciba o no tratamiento quirúrgico.
22. Radioterapia (radio)	Sí No	Según reciba o no radioterapia.
23. Quimioterapia (quimio)	Sí No	Según reciba o no quimioterapia.
24. Hormonoterapia (hormona)	Sí No	Según reciba o no hormonoterapia.
<b>Factores dependientes del diagnóstico</b>		
25. Biomarcadores (biomar)	Sí No	Según se hayan obtenido o no al diagnóstico el resultado de los biomarcadores tumorales con la posterior terapia personalizada.

<sup>a</sup> Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. *Br J Cancer*. 1957;11:359-77.

<sup>b</sup> Elston CW. Grading of invasive carcinoma of the breast In: Page DL, Anderson TJ, eds. *Diagnostic Histopathology of the Breast*. 1st ed. Edinburgh: Churchill Livingstone; 1987. pp 300-11.

<sup>c</sup> Edge S, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A, AJCC. *Cancer Staging Manual*, 7th Ed, New York; 2009.

#### Codificación en la base de datos:

Se codificó como 1 las fallecidas, los "Sí" y los "Peor" y 0 en caso contrario.

**Estudio de factores pronósticos de mortalidad en cáncer de mama**

HC \_\_\_\_\_

Viva ( )                      Fallecida ( )

1. Mayor de 50 años      Si ( )    No ( )
2. Premenopausia        Si ( )    No ( )
3. Tipo histológico \_\_\_\_\_
4. Grado de diferenciación histológica I ( ) II ( ) III ( ) IV ( )
5. Grado de diferenciación nuclear I ( ) II ( ) III ( ) IV ( )
6. Estadio (clasificación TNM) 0 ( ) I ( ) II ( ) III ( ) IV ( )
7. Tamaño tumor primario ..... < 5 ( ) ≥ 5 ( )
8. Afectación ganglionar ..... Si ( ) No ( )
9. Metástasis ..... Si ( ) No ( )
10. Invasión sanguínea ..... Si ( ) No ( )
11. Invasión linfática ..... Si ( ) No ( )
12. Invasión perineural ..... Si ( ) No ( )
13. Invasión muscular ..... Si ( ) No ( )
14. Invasión adiposa ..... Si ( ) No ( )
15. Invasión de piel y pezón ..... Si ( ) No ( )
16. Necrosis ..... Si ( ) No ( )
17. Desmoplasia ..... Si ( ) No ( )
18. Histiocitosis ganglionar ..... Si ( ) No ( )
19. Reacción inflamatoria ..... Si ( ) No ( )
20. Margen quirúrgico libre ..... Si ( ) No ( )
21. Cirugía ..... Si ( ) No ( )
22. Quimioterapia ..... Si ( ) No ( )
23. Radioterapia ..... Si ( ) No ( )
24. Hormonoterapia ..... Si ( ) No ( )
25. Biomarcadores ..... Si ( ) No ( )

**Universidad de Ciencias Médicas Santiago de Cuba**

**Encuesta de validación del análisis estadístico implicative (ASI) para bioestadísticos e informáticos**

**Encuesta**

Especialidad: Bioestadística ( ) Informática ( ) Años de experiencia: \_\_\_\_

Provincia: \_\_\_\_\_ ¿Se dedica usted al análisis estadístico de datos? Si ( ) No ( )

Después de haber recibido la información pertinente sobre el ASI necesitamos que nos responda con sinceridad lo siguiente:

1. Sobre el ASI valore su grado de acuerdo con los aspectos que se relacionan a continuación, marcando con una X según considere.

Aspectos a considerar	Grado de acuerdo				
	Totalmente de acuerdo	De acuerdo	Sin opinión	En desacuerdo	Totalmente en desacuerdo
1. Aporta información interesante					
2. Es útil para la investigación					
3. Ayuda en la toma de decisiones					
4. Los resultados son acertados					
5. Sus gráficos permiten una mejor comprensión del problema estudiado					

2. Con el objetivo de comparar el ASI y la regresión logística binaria (RLB) a continuación se le presenta un conjunto de aspectos para que otorgue a cada técnica una puntuación del 1 al 5, según considere, siendo el 5 la mejor puntuación y el 1 la peor.

Aspecto a comparar	RLB	ASI
1. Está a tono con las exigencias de la investigación actual		
2. El procesamiento computacional es fácil de realizar		
3. Aporta gran cantidad de información		
4. El resultado es comprensible		
5. Brinda una excelente presentación desde el punto de vista gráfico		

**Anexo 4.** Instrumento para la validación de contenido por juicio de expertos.

**Universidad de Ciencias Médicas Santiago de Cuba**

**Encuesta de validación del análisis estadístico implicativo (ASI) en Oncología**

Especialidad: \_\_\_\_\_ Años de experiencia: \_\_\_\_ Provincia: \_\_\_\_\_

**1. Encierre en un círculo el resultado más lógico según la práctica médica**

- A. El grado nuclear III-IV, el estadio avanzado de la enfermedad, la presencia de metástasis y la ausencia de histiocitosis son factores de mal pronóstico, mientras que estar en etapa premenopáusica es un factor de buen pronóstico,
- B. El grado nuclear III-IV, el estadio avanzado de la enfermedad, la presencia de metástasis, no recibir tratamiento quirúrgico ni hormonoterapia son factores de mal pronóstico,

**2. Analice las siguientes reglas (R) obtenidas en un estudio de identificación de factores pronósticos de mortalidad en el cáncer de mama y considere como se cumplen las mismas en la práctica médica,**

Regla	Nivel de cumplimiento en la práctica médica				
	Se cumple siempre	Se cumple frecuentemente	Se cumple a veces	Casi nunca se cumple	Nunca se cumple
R1: Todo el que tiene metástasis está en un estadio avanzado					
R2: Todo el que está en un estadio avanzado tiene metástasis					
R3: Un mayor grado de diferenciación histológica coincide con un mayor grado de diferenciación nuclear					
R4: Un tamaño tumoral mayor de 5 cm conlleva a tener necrosis					
R5: Un paciente que no es tributario de tratamiento quirúrgico debe tener metástasis					
R6: La paciente que no recibe quimioterapia tampoco recibe radioterapia					
R7: La que tenga histiocitosis no debe tener desmoplasia					
R8: Si hay invasión perineural es porque hay invasión sanguínea					
R9: La invasión linfática lleva casi siempre a la necrosis					
R10: Con un tamaño tumoral mayor de 5 cm es posible que haya invasión sanguínea					
R11: La afectación ganglionar lleva a la invasión sanguínea					
R12: En un estadio muy avanzado de la enfermedad no se indica hormonoterapia					
R13: En el periodo premenopáusico es probable que no haya margen quirúrgico					
R14: Los tipos histológicos más agresivos cursan con invasión neural					
R15: Los que usan radioterapia también usan quimioterapia					
R16: Hay equivalencia entre el grado nuclear y el grado histológico					
R17: Los que tienen desmoplasia también tienen histiocitosis					
R18: Los biomarcadores son un importante factor pronóstico					
R19: Los que se operan tienen invasión del tejido adiposo					
R20: Los que reciben quimioterapia y radioterapia tienen invasión muscular					
R21: La mayoría con invasión sanguínea tiene invasión ganglionar					
R22: Los que tienen mayores grados histológico y nuclear tienen necrosis					
R23: Los que se realizan biomarcadores reciben tratamiento hormonal					

**Conflicto de interés**

Los autores declaran no tener conflicto de interés en esta investigación.

**Recibido:** 11/11/19

**Aprobado:** 14/01/20



Esta obra está bajo una [licencia de Creative Commons Atribución-NoComercial 4.0 Internacional](https://creativecommons.org/licenses/by-nc/4.0/)