

8910
Pág 66

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE CIENCIAS QUIMICAS Y FARMACIA

"Análisis estadístico a seguir en la comparación de métodos diagnósticos cualitativos en la Escuela de Química Biológica: Evaluación de las tesis de Químicos Biólogos elaboradas de 1986 a 1997 y elaboración de un instructivo general para la comparación de métodos"



Presentado por:

Rosa Carolina Joo León

Para optar al título de:

Químico Biólogo

Guatemala, octubre de 1998.

PROPIEDAD DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
Biblioteca Central

06
T(1892)

c.4

JUNTA DIRECTIVA
FACULTAD DE CC.OO. Y FARMACIA

| | |
|------------|-------------------------------------|
| DECANA | LICDA. HADA MARIETA ALVARADO BETETA |
| SECRETARIO | LIC. OSCAR FEDERICO NAVE HERRERA |
| VOCAL I | DR. OSCAR MANUEL COBAR PINTO |
| VOCAL II | DR. RUBEN DARIEL VELASQUEZ MIRANDA |
| VOCAL III | LIC. RODRIGO HERRERA SAN JOSE |
| VOCAL IV | BR. HERBERTH RAUL AREVALO ALVARADO |
| VOCAL V | BR. MANOLA ANLEU FORTUNY |

Acto que dedico a:

Dios y la Virgen : Por ser los guías en mi camino.

Mis Padres : Por que gracias a su esfuerzo soy hoy lo que soy.

Mi Hermanita : Cuchi, que siempre la llevo en mi corazón.

Mi Familia : Tia Nena, Tere, Carmen, Tita, Vera, Tio Canche, Víctor, Alfredo, Juan, Marisol, Liseth, Mario, Susan, Ana Lucia, Alfredito, Johanita, Mariela y Juan Pablo.

Mis Tíos : Guillermo León y Liliana Flores.

Mis Amigas y amigos : Claudia Bonilla.

" La Mancha " : Ana María, Claudia, Delmy y Geraldina.

Karen, Klelia y Javier.

Y a todas las personas que han contribuido a lo largo de mi vida para que hoy pueda alcanzar esta meta.

Agradecimientos

Al Lic. Oscar Federico Nave Herrera, por su asesoría y orientación, las cuales han hecho posible este trabajo de tesis.

A Karen Contreras, por su gran apoyo durante la carrera.

Al Lic. Francisco Mendizabal Prem por su amable colaboración bibliográfica.

Al Lic. Jorge Luis de León por su asesoría estadística.

A la Licda. Mariane Morales por su ayuda en la exposición de ésta tesis.

I N D I C E

| | |
|---|----|
| I. Resumen | 1 |
| II. Introducción | 3 |
| III. Antecedentes | 5 |
| A. Bioestadística y el Químico Biólogo | 5 |
| B. Unidad de Informática y Biometría | 5 |
| 1. Funciones | 5 |
| 2. Trabajos de tesis anteriores | 6 |
| C. Prueba Diagnóstica de medición nominal y respuesta dicotómica. | 7 |
| D. Valoración de una Prueba de Oro | 8 |
| 1. Reproducibilidad y Exactitud | 9 |
| 2. Estándar de Oro | 10 |
| 3. Sensibilidad | 11 |
| 4. Especificidad | 11 |
| 5. Valor Predictivo Positivo | 12 |
| 6. Valor Predictivo Negativo | 12 |
| 7. Prevalencia de la enfermedad | 12 |
| 8. Tablas de Contingencia | 15 |
| E. Tipos de pruebas diagnósticas: | 16 |
| 1. Prueba Diagnóstica o de Tamizaje | 16 |
| 2. Prueba Confirmatoria | 17 |
| 3. Prueba de Exclusión | 17 |
| F. Selección de la muestra | 17 |
| G. Cálculo del tamaño de la muestra: | 21 |
| 1. Nivel de Confianza | 22 |
| 2. Varianza | 22 |
| 3. Limite de Error | 24 |
| H. Otros Indices Estadísticos: | 25 |
| 1. Prueba de Kappa | 25 |
| 2. Índice de Validez | 27 |
| III. Justificación | 28 |
| IV. Objetivos | 29 |
| V. Materiales y Métodos | 30 |
| VI. Resultados | 33 |
| VII. Discusión de resultados | 35 |
| VIII. Conclusiones | 43 |
| IX. Recomendaciones | 45 |
| X. Referencias | 46 |
| XI. Anexos | 48 |

I. RESUMEN

Con el fin de determinar los procedimientos estadísticos y detectar posibles problemas y errores cometidos en las tesis de la Escuela de Química Biológica, que compararon pruebas diagnósticas de medición nominal y respuesta dicotómica, se analizaron 17 tesis elaboradas durante el periodo comprendido de 1988 a 1997. Se revisaron aspectos como el uso de estándar de oro, cálculo y selección de muestra, prevalencia de la enfermedad, cálculo e interpretación de índices estadísticos y conclusiones.

Algunos problemas encontrados en los estudios fueron la falta de recursos económicos y materiales, así como también la contaminación de muestras, razones por las que dos tesis no lograron su objetivo de comparar este tipo de pruebas diagnósticas. En las tesis revisadas se detectó los siguientes errores: no se enfatizó que los valores de la prueba que se compara están dados relativamente en función de otra prueba al no utilizar el estándar de oro en la comparación; falta una descripción adecuada de los pacientes, no se mencionan criterios de inclusión y exclusión; limitaciones de algunos estudios a pacientes con enfermedad grave o avanzada; cálculo de valores predictivos sin mencionar la prevalencia de la enfermedad utilizada; falta de cálculo de los valores predictivos de la prueba a

pesar de contarse con el dato de la prevalencia de la enfermedad; generalización de resultados cuando el muestreo fue por conveniencia; falta de construcción de una Tabla de Contingencia y el uso inapropiado de Ji Cuadrada en la comparación de pruebas diagnósticas.

Con base en lo analizado de las tesis, se procedió a elaborar una guía para orientar al investigador que desee evaluar este tipo de pruebas diagnósticas antes de empezar su fase experimental. Esta guía hace referencia a temas como: uso de estándar de oro, muestreo (criterios de inclusión y exclusión, cálculo y selección de la muestra), procedimientos de análisis estadístico y cuidados en la interpretación de resultados. En general se refiere a aspectos a tomar en cuenta y condiciones a seguir al evaluar la capacidad diagnóstica de la prueba.

II. INTRODUCCION

Existe en la actualidad una proliferación de pruebas diagnósticas para diferentes enfermedades, la publicidad que se les hace, atribuye que una u otra es la mejor elección, olvidando muchas veces hacer referencia a la sensibilidad y especificidad de la prueba o los resultados de su comparación con la prueba de referencia o " estándar de oro ".

En la Escuela de Química Biológica se han realizado tesis en las que se comparan pruebas diagnósticas evaluándose métodos alternativos en función a la prueba de oro, u otra no necesariamente la de referencia. Se determinó que de 1988 a 1997, diecisiete tesis de un total de 283 tesis (un seis por ciento), tienen dentro de sus objetivos, comparar dos o más pruebas diagnósticas.

Para realizar tal comparación se deben tener en cuenta ciertas consideraciones como: si se utiliza estándar de oro u otra prueba, el muestreo a seguir (criterios de inclusión y exclusión, cálculo y selección de muestra), el análisis estadístico a aplicar, la interpretación de resultados.

En el presente estudio se analizaron las tesis de la Escuela de Química Biológica que compararon pruebas diagnósticas, cuya medición fue nominal y de respuesta dicotómica para expresar si un individuo padece o no la enfermedad. De acuerdo al análisis, se evaluaron los problemas confrontados con mayor frecuencia, detectándose los errores más frecuentes y se elaboró una guía general para orientar al investigador que desee comparar este tipo de pruebas.

III. ANTECEDENTES

A. Bioestadística y el Químico Biólogo

Día tras día el Químico Biólogo maneja una gran cantidad de información que recopila, valora, analiza e interpreta (1,2). Para ello, se apoya en la Bioestadística, la cual es una herramienta que sirve para la deducción de conclusiones y en la toma de decisiones razonables (1,3). Decisiones que van desde la aceptación o rechazo de un resultado de examen de laboratorio, hasta la elección de la prueba diagnóstica más adecuada. La Bioestadística tiene aplicación en el registro y empleo de datos, control de calidad, hasta validación de pruebas de laboratorio, investigaciones y publicaciones científicas (4,5).

B. Unidad de Informática y Biometría

1. Funciones

Para asesorar el diseño estadístico de los proyectos de investigación y en el análisis de datos resultantes, se encuentra en la Facultad de Ciencias Químicas y Farmacia la Unidad de Informática, que es una unidad de apoyo del Instituto de Investigaciones Químicas y Biológicas (IIQB). Esta Unidad se encarga de recabar y procesar la información procedente de las distintas unidades de la Facultad (6).

2. Trabajos de Tesis anteriores

Con el propósito de evaluar la investigación en la Escuela de Química Biológica y favorecer la producción de trabajos de tesis con mayor proyección a la sociedad, Bertha Rodas Mazariegos realizó su tesis " Evaluación de la investigación en la Escuela de Química Biológica según las tesis de 1979 a 1987 " . Luego de evaluar ciento sesenta y cinco tesis se determinó que el tipo de investigación más utilizado fue la descriptiva, seguida de la experimental, además se determinó que la estadística descriptiva fue más frecuente que la inferencial. Se concluye al final que más de la mitad de las tesis analizadas recomiendan darle continuidad al tema, por lo que estas tesis quedan inconclusas al no dárseles seguimiento con estudios analíticos o experimentales (7).

En 1992, Froilán Alvarez Garrido revisó críticamente el uso de análisis de regresión en su tesis " Uso del análisis de regresión lineal en los trabajos de tesis *ad gradum* de la Escuela de Química Biológica ". En cientocuatro tesis se detectó aquellos trabajos cuyo planteamiento requería la aplicación de dicho análisis y si se realizó o no. De las tesis analizadas el 11.54 por ciento debió aplicar el análisis de regresión, pero de esa cantidad sólo se planteó en el 16.67 por ciento, teniéndose que replantear en una de ellas. Alvarez Garrido concluye en su tesis que este

análisis no ha sido aplicado correctamente en las tesis de la Escuela de Química Biológica (8).

Luego de un estudio piloto se determinó que durante el periodo de 1988 a 1997, en la Escuela de Química Biológica se han elaborado diecisiete tesis que comparan pruebas diagnósticas cualitativas, cuya escala de medición es nominal y cuya variable de respuesta es dicotómica (Centro de Documentación y Biblioteca, Facultad de Farmacia, Universidad de San Carlos de Guatemala).

C. Prueba Diagnóstica de medición nominal y respuesta dicotómica

Una situación común para el Químico Biólogo, es hayarse ante una serie de pruebas diagnósticas y no tener claro cómo seleccionar la más conveniente.

Una prueba de laboratorio es empleada para obtener un dato objetivo que ayude al diagnóstico y tratamiento del paciente, por lo que el fin de una prueba diagnóstica es predecir o identificar la verdadera condición de un paciente (9,10). Aunque el establecimiento de un diagnóstico es un proceso imperfecto que resulta siendo más bien una probabilidad que una certeza (11).

Una prueba diagnóstica de medición nominal, se refiere a la prueba que basa su medición en categorías o

clasificaciones, por ejemplo: escasa, regular o abundante cantidad. Estas mediciones están en escala nominal, que es la forma más simple de medición y la más débil de las escalas (10).

Al decir que la prueba diagnóstica de medición nominal, tiene respuesta dicotómica, se refiere a que el resultado puede clasificarse únicamente de dos formas o sólo entre dos opciones, por ejemplo: positivo o negativo, presencia o ausencia de la enfermedad. Como ejemplos de pruebas de este tipo se puede mencionar el examen de KOH y la prueba de embarazo, entre otras. Sin embargo en algunas pruebas resulta difícil poder clasificar un diagnóstico clínico dentro de dos opciones, ya que en algunas circunstancias una enfermedad no puede ser establecida como definitivamente presente o ausente, y el resultado podría parecer más bien como *probable o incierto* en lugar de un positivo o negativo. Un ejemplo de esta situación son los resultados que se obtienen de la técnica de Western Blot, los cuales pueden ser positivo, negativo o indeterminado. Esta simplificación de resultados en dos categorías, puede hacer que en algunas pruebas se pierda información (10,11).

D. Valoración de una Prueba Diagnóstica

El valor de una prueba diagnóstica reside en su habilidad para distinguir entre factores que comúnmente se confunden, especialmente cuando su pronóstico o terapia

difieren grandemente. La importancia el uso de pruebas diagnósticas radica no sólo en su valor predictivo en los casos equívocos, sino también en los casos sobre los cuales se tiene incertidumbre (10,12-14).

1. Reproducibilidad y Exactitud

Una prueba debe ser reproducible y exacta, y para cumplir esto debe producir los mismos resultados cuantas veces se aplique en las mismas condiciones. Hay **reproducibilidad** cuando la prueba produce resultados consistentes al repetirse en las mismas condiciones y ser interpretada de la misma forma sin conocer sus resultados previos. Debe tenerse presente que las condiciones del paciente y laboratorio bajo las cuales se realiza la prueba pueden no ser las mismas, y presentarse variaciones, como variaciones biológicas, instrumentales, de medición o técnicas (9,11).

La **exactitud** significa que la prueba producirá resultados semejantes al verdadero valor anatómico, fisiológico o bioquímico. La exactitud es una propiedad necesaria de una buena prueba, ya que con ella se elimina todo tipo de sesgos como efectos al azar y errores sistemáticos. Sin embargo que haya exactitud no implica que la prueba sea válida, ya que la validez de una prueba implica que la prueba es una medida apropiada del fenómeno estudiado (9,11).

2. Estándar de Oro

Es la que discrimina sano y enfermo o bien, positivo o negativo correspondientemente, indica el diagnóstico definitivo del paciente. Se le conoce también como prueba de oro, estándar dorado, ideal o método de referencia (1,9,10). El uso de un estándar de oro para determinar definitivamente a los que padecen la enfermedad, es un requisito cuando se va a probar la utilidad de una nueva prueba, ya que su utilidad se basa en su comparación con el estándar de oro. Por lo que se debe estar seguro de que el estándar de oro que se vaya a emplear es la mejor referencia posible, la mejor prueba disponible y la prueba generalmente aceptada para diagnosticar la enfermedad (9). Puede ser que en la práctica la nueva prueba sea de mayor utilidad que la de referencia (mayores beneficios para el paciente y su diagnóstico, menores riesgos y costos), aunque al comparar una nueva prueba con la de referencia, no se contempla precisamente la posibilidad de que la prueba nueva sea mejor que la de oro (9,12).

Luego de establecer el estándar de oro a utilizar, se puede determinar la capacidad de discriminación diagnóstica de la prueba a comparar, calculando sus índices estadísticos de sensibilidad, especificidad y valores predictivos (10,12).

3. Sensibilidad

La sensibilidad y especificidad fueron introducidos como índices estadísticos que describen la utilidad de una prueba diagnóstica por Yerushalmy, en 1947 (15).

El término Sensibilidad se refiere a la capacidad de la prueba para detectar la enfermedad, por tanto, debe ser positiva en pacientes que la padecen. Se le conoce también como positiva en los enfermos (PEE), tasa de positivas verdaderas o positividad en la enfermedad. Generalmente se expresa como un porcentaje, o sea la proporción de individuos con la enfermedad cuya prueba es positiva. Si hay alta sensibilidad, la tasa de falsos negativos será baja, o sea que no dará un resultado negativo en muchos pacientes con enfermedad (1,9,16,17).

4. Especificidad

Se refiere a la capacidad de la prueba de identificar correctamente a pacientes que no tienen la enfermedad. Se expresa también como un porcentaje, o sea la proporción de individuos sanos cuya prueba es negativa. Se le conoce también como negativa en los sanos (NES), negativa en la salud o específica de la salud. Si la especificidad es alta, los falsos positivos serán pocos, es decir que no dará un resultado positivo en muchas personas sin la enfermedad (1,9,16,17).

5. Valor predictivo positivo

Indica el porcentaje de pacientes con resultado positivo que de hecho padecen la enfermedad. Es la fracción numérica de los resultados positivos que son positivos verdaderos. Se le llama también Probabilidad postprueba de enfermedad. Su verdadera utilidad radica en indicar cuál es la probabilidad de que haya enfermedad si el resultado de la prueba es positivo (1,9,10,12,13,16).

6. Valor predictivo negativo

Indica el porcentaje de pacientes que no tienen la enfermedad y su resultado fue negativo. Es la fracción numérica de los resultados negativos que son negativos verdaderos. Se le conoce como Probabilidad postprueba de no enfermedad e indica la probabilidad de que no haya enfermedad si el resultado es negativo (1,9,10,12,16).

Los valores predictivos complementan la información que se obtiene de la sensibilidad y especificidad de la prueba.

7. Prevalencia

Se obtiene como la proporción de pacientes realmente enfermos, dentro del total de pacientes a los cuales se aplica la prueba en una población determinada (1). Se refiere a la frecuencia relativa de la enfermedad y es una estimación de la probabilidad de enfermedad antes de

realizar la prueba, por eso se le conoce también como "probabilidad anterior a la prueba" porque indica la probabilidad de la enfermedad en la población antes de efectuar la prueba (9,13). Este término está estrechamente relacionado con los valores predictivos (por eso a ellos también se les llama Probabilidades postprueba). Al cambiar la prevalencia de una enfermedad en la población a muestrear, los valores predictivos también cambian. El valor predictivo positivo cambiará en la misma dirección que la prevalencia, es decir que si la prevalencia de la enfermedad aumenta, también lo hará el valor predictivo positivo de la prueba. Con el valor predictivo negativo sucede lo contrario, si la prevalencia aumenta, el valor predictivo negativo disminuye (10,13,18,19). Esto se aprecia mejor en la Tabla 2 de los Anexos.

La sensibilidad y especificidad de la prueba se afectan también por la prevalencia de la enfermedad de la población en la que se aplique la prueba. Al aplicar una prueba en una población con proporción de enfermos alta (o sea una prevalencia de enfermedad alta), los valores de sensibilidad y especificidad serán mayores que si se aplica en una población donde la prevalencia de la enfermedad es baja.

Por lo tanto la capacidad de una prueba para detectar y predecir la presencia de una enfermedad cambia sustancialmente, según se aplique a grupos e individuos con

diferentes prevalencias. Esta es una de las razones por la cual una prueba puede ser útil para el diagnóstico en un grupo de pacientes que se sospecha la enfermedad, pero inútil para el tamizaje de la población general en la cual la sospecha de enfermedad es baja (9). Por eso es importante realizar la mejor estimación de la prevalencia de la enfermedad, antes de realizar la prueba.

Muchas veces se asume una prevalencia ficticia al incluir en la muestra en la que se ensayará la prueba, la mitad de personas con la enfermedad y la otra mitad sin la enfermedad (una prevalencia del 50%). Esta es usualmente una prevalencia más alta de la que se observa en la práctica clínica (10,12). Por lo que los valores predictivos de la prueba a ensayar como se mencionó anteriormente, se verán directamente afectados por ésta prevalencia ficticia, obteniéndose datos irreales. A esto se debe que algunas pruebas al asumirse una prevalencia ficticia, resulten con altos valores predictivos en la fase experimental, y luego no sean tan eficaces en discriminar los verdaderos positivos en la práctica común. Sin embargo una prevalencia de 50 por ciento, es útil para la comparación de pruebas en un laboratorio de referencia, en donde el número de resultados positivos es mayor que en un laboratorio común o de tamizaje, o también cuando se quiere estimar la prevalencia de una enfermedad que se sabe es bastante común en la población.

B. Tablas de Contingencia de 2 X 2

Las relaciones entre las condiciones clínicas y los resultados de una prueba diagnóstica pueden ordenarse por medio de una tabla de contingencia de doble entrada o de 2X2 como sigue:

| Prueba en estudio | Prueba de Oro Enfermos | Prueba de oro Sanos |
|-------------------|---|---|
| Positivos | a = No. enfermos y positivos. Verdaderos positivos. | b = No. de sanos y positivos. Falsos positivos. |
| Negativos | c = No. enfermos y negativos. Falsos Negativos. | d = No. de sanos y negativos. Verdaderos Negativos. |

En las columnas se muestran las condiciones verdaderas del paciente (o sea los resultados según la prueba o estándar de oro) y en las filas los resultados de la prueba diagnóstica a comparar. Las cuatro celdas interiores indican si el paciente ha sido correcta (celdas a y d) o falsamente clasificado (celdas b y c) (10). En estas tablas se registran las frecuencias observadas, y se utilizan para determinar la sensibilidad, especificidad, valores predictivos, exactitud de la prueba, así como prevalencia de la enfermedad que detecta (1,9,10,12). De esta forma se obtiene que:

$$\text{Sensibilidad} = \frac{a}{a + c}$$

$$\text{Especificidad} = \frac{d}{b + d}$$

Valor predictivo positivo = $a / a + b$

Valor predictivo negativo = $d / c + d$

Exactitud = $a + d / a + b + c + d$

Prevalencia = $a + c / a + b + c + d$

E. Tipos de Pruebas:

El fin por el cual se emplea una prueba debe tenerse presente, ya que indica las características que debe tener la prueba a utilizar.

Existen tres tipos de pruebas diagnósticas de acuerdo al fin por el que se emplean: prueba diagnóstica, prueba confirmatoria y prueba de exclusión. Estas pueden utilizarse para uno, dos o los tres fines que se mencionan a continuación (10,12,20).

1. Prueba Diagnóstica

Su fin es detectar la presencia de una enfermedad. En los estudios de tamizaje se utilizan estas pruebas para detectar una enfermedad en pacientes aparentemente sanos y asintomáticos. Cuando se necesita una prueba diagnóstica, se requiere que posea alta sensibilidad y correr el riesgo de poseer una tasa alta de falsos positivos (usualmente al tener un resultado positivo en una prueba diagnóstica, se procede a una prueba confirmatoria) (10).

2. Prueba Confirmatoria

Se utiliza para establecer si una enfermedad está presente cuando hay fuertes sospechas y se desea verificarlo. Esta prueba debe ser de alta sensibilidad, con pocos o ningún falso positivo (10).

3. Prueba de Exclusión

Su objetivo es descartar la presencia de una enfermedad cuando hay sospechas. Este tipo de prueba debe poseer una sensibilidad aún mayor que la de la prueba diagnóstica (sensibilidad cercana a 1), y si existe el riesgo de falsos negativos, ante un resultado negativo en esta prueba, se procede a una prueba confirmatoria (10).

De esta forma una prueba positiva de alta sensibilidad debe ser seguida de una prueba de alta especificidad. Puede así establecerse la clasificación de negativo para la enfermedad cuando la prueba diagnóstica fue negativa, o positivo si y sólo si las pruebas diagnósticas y confirmatorias fueron positivas (10).

F. Selección de la Muestra

Para el cálculo del tamaño de la muestra se deben tener en cuenta varios factores como el tipo de estudio a realizar, la magnitud de los errores tipo I y II a trabajar si se realizará una prueba de hipótesis, o el nivel de confianza para una estimación. Para ello el investigador

puede consultar con un estadístico y determinar así el número suficiente de pacientes que debe incluir en su estudio, para demostrar con una probabilidad razonable, la existencia de una diferencia estadísticamente significativa entre las muestras del estudio, y si dicha diferencia existe realmente en la población de la cual se extrajo las muestras. Así también, para fines de estimación de un parámetro (prevalencia) se sabrá la confiabilidad de la muestra y el nivel de exactitud o inexactitud de la estimación (9).

Es común que el investigador incluya en su muestra sólo a pacientes con síntomas claros de la enfermedad, pero de esta forma lo que se está haciendo es limitar el estudio a los individuos que tienen una enfermedad grave o en fase muy avanzada. Al considerar sólo los casos avanzados, se eliminan los casos dudosos o en etapas iniciales de la enfermedad, por eso en la práctica el rendimiento de una prueba puede diferir a los resultados obtenidos cuando se comparó con el estándar de oro. Por esto, puede ser engañoso hacer generalizaciones sobre la utilidad de una prueba cuando en el estudio de ella se incluyó sólo a individuos con enfermedades claramente definidas (9,12).

Por lo tanto es importante definir con cuidado los criterios de inclusión para la selección de los pacientes o muestras, ya que definirán la población a la que está dirigida la prueba. Se debe tratar de abarcar todo el

espectro médico de la enfermedad: cuadros leves, intermedios y severos, pacientes con enfermedades con las que se puede confundir pero que comúnmente se asocian, sintomáticos y asintomáticos (9,12,16).

De esta forma para evaluar la capacidad diagnóstica de la prueba se trabaja con individuos que se sospecha están sanos y con una prevalencia baja. Si se evalúa la capacidad confirmatoria o de exclusión de la prueba, los individuos con los que se trabaja son pacientes cuyas condiciones clínicas hacen sospechar que padecen de la enfermedad y que se desea confirmar o excluir su padecimiento, la prevalencia en este caso debe ser alta (10).

Feinstein propone cuatro tipos de grupos para evaluar la capacidad diagnóstica de una prueba para una enfermedad X (10):

Grupo 1: Pacientes con la enfermedad X y asintomáticos.

Grupo 2: Pacientes con la enfermedad X y sintomáticos que abarquen el espectro médico de la enfermedad.

Grupo 3: Pacientes sin la enfermedad X y con manifestaciones similares a los incluidos en el Grupo 2.

Grupo 4: Pacientes que padecen de otras enfermedades que producen trastornos similares a la enfermedad X, debido a que se localizan en la misma región o por que producen disfunciones paraclínicas similares (10,21).

La sensibilidad de una prueba que se quiera utilizar para tamizaje dependerá entonces de su capacidad de identificar a los pacientes del Grupo 1. Si la prueba quiere ser utilizada como prueba de exclusión su sensibilidad se medirá en la detección adecuada de los pacientes de los Grupos 1 y 2. La especificidad de la prueba dependerá en su habilidad para dar pocos resultados falsos positivos en los pacientes de los Grupos 3 y 4.

Algunas pruebas para ser positivas necesitan que la enfermedad en el paciente haya alcanzado cierto nivel patológico y mostrar así ciertos síntomas que detecta la prueba, en este tipo de pacientes sospechosos la prueba será positiva y de alta sensibilidad. Pero si la enfermedad está presente pero no ha alcanzado el nivel patológico necesario para que la prueba sea positiva, la prueba será negativa (falso negativo), y en este tipo de pacientes la prueba tendrá baja sensibilidad. Esto podría ser una de las explicaciones de los falsos negativos. Otras causas de los falsos resultados, pueden ser el padecimiento de enfermedades alternas y/o ingesta de medicamentos (10).

Se aconseja realizar un estudio ciego en donde el investigador o quien aplica el instrumento de diagnóstico no debe saber el verdadero estatus de enfermedad (positivo o negativo) del paciente. Así se evita el llamado sesgo de sospecha (12,13).

6. Cálculo del número de muestra

El número de muestras (n) necesario para la comparación de pruebas diagnósticas de medición nominal y respuesta dicotómica, depende de lo que se desea conocer de la prueba.

Si se va a determinar sólo la sensibilidad y especificidad, el tamaño y tipo de muestra puede ser seleccionado *a priori* (por conveniencia), siendo el investigador quien determine cuántos pacientes incluirá en su estudio, aunque éste tipo de muestreo carece de validez externa, y no puede hacerse generalizaciones de sus resultados. Por otro lado, si además de la sensibilidad y especificidad se va a evaluar los valores predictivos positivo y negativo, entonces en el cálculo interviene la prevalencia de la enfermedad que detecta la prueba, ya que como se mencionó antes, la prevalencia tiene efectos directos sobre los valores predictivos de la prueba, no así sobre la sensibilidad y especificidad (9,10).

En general, el tamaño de la muestra se obtiene por medio del siguiente cálculo:

$$n = \frac{NC^2 * \sigma^2}{\Delta^2}$$

Donde NC : Nivel de Confianza

σ^2 : Varianza

Δ : Límite de Error

1. Nivel de Confianza

Es la probabilidad que el estimador contenga al parámetro de la población, antes de seleccionar la muestra (22,23). El nivel de confianza se calcula bajo el supuesto de que la distribución del promedio de la variable que se estima, representa una forma simétrica que se asemeja a la curva normal. El NC cuando se desea hacer una estimación está dado por:

$$NC = Z_{1 - (\alpha/2)}$$

Donde Z es un valor proveniente de la curva normal que representa la probabilidad " $1 - \alpha/2$ " de que el estimador contenga el parámetro poblacional, previo a seleccionar la muestra (entonces " α " representa la probabilidad de que el estimador no contenga el parámetro) (22,23).

Mientras más seguridad se quiera tener de que el estimador contiene el parámetro o sea un nivel de confianza elevado, el número de muestras n a incluir en el estudio, será mayor o más grande. Por lo regular se trabaja con un NC de 90% a 95% .

2. Varianza

Es una medida de cuánto se alejan los datos del valor medio, representa la dispersión de los datos respecto a su media. En este caso, el término varianza se refiere a la

" variabilidad " del estadístico en la muestra (del estimador) (1,22).

Cuando el propósito es la comparación de este tipo de pruebas diagnósticas, la varianza se obtiene por medio de la prevalencia de la enfermedad, por medio de las siguientes fórmulas:

$$\text{Prevalencia (p)} = \frac{\text{Número de casos positivos}}{\text{Total de casos}}$$

$$\text{Complemento (q)} = \frac{\text{Número de casos negativos}}{\text{Total de casos}}$$

O bien $(q) = 1 - p$ Ya que p y q son complementarias.

$$\sigma^2 = p q$$

Si la varianza de una población aumenta, también aumentará el número de muestras a incluir en el estudio, pero si la varianza es pequeña, el número de muestras será menor.

Frecuentemente se trabaja con una " prevalencia esperada ", la cuál establece el supuesto de que la

prevalencia de la enfermedad es del 50 por ciento (que también puede expresarse como una frecuencia relativa = 0.50) en la población. Esto dará siempre el máximo de muestra posible, y abarcará cualquier otra alternativa de prevalencia que fuese la real. El inconveniente con esta prevalencia es que se obtiene una varianza grande que hará que el n sea mayor que si se tomara una varianza menor.

Otras formas de obtener la prevalencia de la enfermedad para poder calcular n es por medio de un estudio piloto, en el que se estime "tentativamente" un valor de prevalencia. También se puede consultar la literatura nacional sobre investigaciones recientes en la población a la que se destina la prueba.

3. Limite de Error

Es la diferencia mínima en la que el estimador se puede alejar del valor real, es decir el margen máximo con el cual se desea brindar la estimación (22,23). El límite de error también se expresa en porcentaje o frecuencia relativa, dependiendo de cómo se halla expresado la prevalencia, ya que deben operarse en las mismas dimensionales. Su selección depende de los propósitos de la prueba (si es de tamizaje, confirmatoria o de exclusión).

Si se permite un limite de error pequeño el n será mayor que si el limite de error es grande.

H. Otros índices Estadísticos

Con el fin de querer reducir la información de las Tablas de Contingencia de 2X2, se puede caer en el error de querer combinar los términos de sensibilidad y especificidad que se obtienen de dicha Tabla, a un sólo índice estadístico, la desventaja de utilizarlos es que se pierde información, ya que al obtener un solo dato numérico éste ya no dice nada acerca de cómo es la sensibilidad y especificidad de la prueba (10,24). Existen varios índices de asociación que combinan la sensibilidad y especificidad en uno, entre estos índices se puede mencionar: la Q de Yule, la Y de Yule, la C de Pearson, la T de Tschuprow, el Índice de Validez y la J de Youlden. Estos dos últimos son los que aparecen con más frecuencia en la literatura (10).

1. Prueba de Kappa

Esta prueba puede utilizarse si se dispone de una prueba de oro o no. Es un índice estadístico de concordancia entre las pruebas, una medida del grado de acuerdo entre ellas. También utiliza la información contenida en una Tabla de Contingencia de 2X2 (1,24). Para poder determinar este índice se utiliza la siguiente fórmula:

$$K = \frac{4(ad - bc) - (b-c)^2}{(2a+b+c) * (2d+b+c)}$$

Esta fórmula puede resumirse a una fórmula general como sigue:

$$K = \frac{D - C}{1 - C}$$

D: Concordancia observada

C: Concordancia esperada

Existen varios criterios para interpretar el valor numérico que se obtiene al final de la fórmula:

Criterios de Interpretación de Fleiss (24):

| Valor Numérico | Interpretación |
|----------------|--------------------------------|
| < 0.40 | No hay acuerdo |
| 0.40 - 0.75 | Acuerdo Intermedio (aceptable) |
| > 0.75 | Buen Acuerdo (excelente) |

Criterios de Interpretación de Landis y Koch (25):

| Valor Numérico | Interpretación del Grado de Concordancia |
|----------------|--|
| < 0.20 | Deficiente |
| 0.20 - 0.40 | Regular |
| 0.41 - 0.60 | Moderado |
| 0.61 - 0.80 | Buena |
| 0.81 - 1.00 | Muy Buena |

2. El Índice de Validez

Llamado también Índice de Eficiencia de una prueba. Indica el porcentaje de clasificaciones correctas. Este se calcula como:

$$(a + d) / N * 100$$

Representa el total de clasificaciones correctas de la prueba, dividido por el número total de pruebas. Otra forma de calcularlo es:

$$(\text{sensibilidad} * \text{prevalencia}) + \text{especificidad}(1 - \text{prevalencia})$$

Por lo que se puede obtener valores altos de este índice al trabajar con prevalencias altas o bajas, de acuerdo a los valores de sensibilidad y especificidad de la prueba (10,26).

Weiss aconseja no utilizar este índice, porque a veces da resultados erróneos y debido a que no encaja en el proceso de toma de decisiones como lo hacen los valores predictivos (26).

IV. JUSTIFICACION

El tipo de prueba a utilizar en cada contexto es importante ya que las conclusiones de los resultados obtenidos dependerán de las características de la prueba que se haya utilizado.

La evaluación de una prueba diagnóstica por medio de su comparación con un estándar de oro u otra prueba, permite obtener índices estadísticos que describen las características de la prueba comparada. De esta forma es posible determinar la prueba más conveniente a utilizar en una investigación, laboratorio o paciente.

Dicha evaluación de pruebas diagnósticas es de gran utilidad cuando se debe decidir cuál utilizar entre dos o más pruebas, o cuando se ha desarrollado una nueva y se desea evaluar su utilidad tal como lo han hecho varias tesis de la Escuela de Química Biológica, las cuales han comparado pruebas con un estándar de oro o respecto a otra prueba.

Es importante entonces que la Escuela de Química Biológica establezca un protocolo o guía que aplique el investigador que vaya a comparar métodos diagnósticos cualitativos, en el cual se unifique la información que determine los criterios y elementos necesarios a seguir.

V. OBJETIVOS

A. Analizar las tesis de Químicos Biólogos elaboradas en 1988 a 1997 en las que se comparen métodos diagnósticos cualitativos.

B. Determinar los problemas y errores frecuentemente cometidos en las tesis que comparan pruebas diagnósticas de medición nominal y respuesta dicotómica, en cuanto al uso de estándar de oro, cálculo y selección de muestra, prevalencia de la enfermedad, análisis estadístico aplicado e interpretación de los resultados.

C. Elaborar una guía que oriente a la persona que desee evaluar dos o más pruebas diagnósticas de medición nominal o bien con degradación de la variable, y de respuesta dicotómica.

VI. MATERIALES Y METODOS

A. Universo y muestra

1. Universo: Tesis *ad gradum* de la Escuela de Química Biológica. (N= 583)
2. Muestra: Tesis *ad gradum* de la Escuela de Química Biológica del período de 1988-1997 (n= 284) en las que se comparen métodos diagnósticos (n= 17).

B. Materiales

1. Recursos Humanos:

Tesista: Rosa Carolina Joo León
Asesor: Lic. Oscar Federico Nave Herrera
Colaboradores: Lic. Jorge Luis de León Arana
Lic. Francisco Mendizabal Prem

2. Recursos Institucionales

- Centro de Documentación y Biblioteca (CEDOBF), Facultad de Ciencias Químicas y Farmacia, Universidad de San Carlos de Guatemala.

- Unidad de Informática, Instituto de Investigaciones Químicas y Biológicas, Facultad de Ciencias Químicas y Farmacia, Universidad de San Carlos de Guatemala.

- Oficina Fase IV, Facultad de Ciencias Médicas, Universidad de San Carlos de Guatemala.

- Biblioteca Central, Universidad de San Carlos de Guatemala.

- Biblioteca del Instituto de Nutrición de Centroamerica y Panamá.

- Biblioteca de la Universidad del Valle de Guatemala.

- Biblioteca de la Universidad Francisco Marroquín.

3. Recursos Materiales:

- 17 Tesis de la Escuela de Química Biológica que comparan pruebas diagnósticas de medición nominal y respuesta dicotómica. Disponibles en el Centro de Documentación y Biblioteca (CEDOBF), Facultad de Ciencias Químicas y Farmacia, Universidad de San Carlos de Guatemala.

- Papeletas para recabar la información necesaria de cada tesis.

- 200 hojas en blanco, papel bond 60 grs.

- Lapicero y lápiz

- 2 Folders.

- Computadora.

- Impresora de computadora.

- Libros, documentos, folletos, revistas, etc, que contengan información pertinente.

4. Metodología

a. Revisión de Literatura sobre el tema.

b. Revisión de tesis y recopilación de la información necesaria por medio de una boleta.

GUATEMALA

- c. Análisis de la información recopilada en la boleta.
- d. Revisión del análisis realizado junto al asesor.
- e. Determinación de las consideraciones generales a tomar en cuenta para comparar métodos diagnósticos y elaboración de la guía.

5. Diseño de la Investigación

Tipo de Estudio: Descriptivo - Prospectivo.

El presente estudio analizó los problemas y errores cometidos por los 17 tesisistas de la Escuela de Química Biológica que compararon métodos diagnósticos cualitativos en 1988 a 1997, revisando los aspectos como: uso de estándar de oro, cálculo y selección de muestra, prevalencia de la enfermedad, análisis estadístico que se aplicó, interpretación de los resultados y las conclusiones. Luego se elaborará una Guía General para la comparación de estos métodos.

VII. RESULTADOS

Por medio de la boleta de recolección de datos (Anexo 2) se revisaron las dieciséis tesis obteniéndose los siguientes resultados:

T A B L A 1 : RESULTADOS OBTENIDOS EN LA REVISIÓN DE TESIS DE QUÍMICA BIOLÓGICA QUE COMPARARON PRUEBAS DIAGNÓSTICAS .

| ASPECTO REVISADO: | FRECUENCIA: | PORCENTAJE: |
|---|-------------|-------------|
| 1. Número de pruebas comparadas: | | |
| 2 pruebas | 8 | 47.1 % |
| 3 pruebas | 8 | 47.1 % |
| más de 3 pruebas | 1 | 5.9 % |
| 2. Tipo de metodología de las pruebas comparadas en cada tesis: | | |
| Serológicas | 6 | 39.7 % |
| Serológica - Tinción | 5 | 32.2 % |
| Serológica - Microbiológica | 2 | 10.0 % |
| Tinciones | 2 | 10.0 % |
| Microbiológica - Tinción | 1 | 4.0 % |
| Microbiológica - Serológica - Tinción | 1 | 4.0 % |
| 3. Uso de Estándar de Oro: | | |
| Si | 8 | 47.1 % |
| No, se utilizó otra prueba | 9 | 52.9 % |
| 4. Describe el tipo de pacientes que incluyó: | | |
| Si | 15 | 88.2 % |
| No | 2 | 11.8 % |
| 5. Forma de obtención del tamaño de muestra: | | |
| Por fórmula general para n | 8 | 47.1 % |
| Muestreo por conveniencia | 7 | 41.2 % |
| No indicó | 2 | 11.9 % |
| 6. Tipo de prevalencia utilizada: | | |
| Con base en estudios previos | 5 | 29.4 % |
| Prevalencia ficticia | 1 | 5.9 % |
| No indicó | 11 | 64.7 % |
| 7. Tesis que lograron su objetivo de comparar pruebas: | | |
| Si | 15 | 88.2 % |
| No | 2 | 11.8 % |

| ASPECTO REVISADO: | FRECUENCIA: | PORCENTAJE: |
|--|-------------|-------------|
| 8. Presenta Tabla de Contingencia: | | |
| Si | 9 | 60.0 % |
| No | 6 | 40.0 % |
| 9. Cálculo de sensibilidad y especificidad: | | |
| Si | 10 | 66.7 % |
| No | 5 | 33.3 % |
| 10. Cálculo del valor predictivo positivo: | | |
| Si | 6 | 40.0 % |
| No | 9 | 60.0 % |
| 11. Cálculo del valor predictivo negativo: | | |
| Si | 5 | 33.3 % |
| No | 10 | 66.7 % |
| 12. Conclusiones acordes a los resultados obtenidos: | | |
| Si | 15 | 88.2 % |
| No | 2 | 11.8 % |

VIII. DISCUSION DE RESULTADOS

Después de revisar las diecisiete tesis, en las que todas incluyeron dentro de sus objetivos la comparación de pruebas diagnósticas, se determinó que por lo regular se compara dos o tres pruebas, en su mayoría serológicas.

La mayoría no utilizó un estándar de oro como referencia, podría deberse a la falta de recursos económicos, material, equipo, complejidad de la prueba y/o a las molestias que puede implicar para el paciente la práctica de una prueba. La comparación de pruebas entre sí, y no respecto a un estándar de oro, fue entonces lo más frecuente, posiblemente por ser menos costoso, más accesible o sencillo de llevar a cabo. Sin embargo, se debe tener cuidado al concluir sobre los resultados, y debe especificarse que la prueba que se tomó de referencia no fue la que generalmente se acepta para diagnosticar la enfermedad. Los resultados que se obtienen de esta comparación no son iguales a los que se obtendrían al comparar con el estándar de oro, ya que la capacidad diagnóstica de una Prueba X, se comparó contra otra Prueba Y, que a su vez tiene sus propias características respecto al estándar de oro. Los datos sólo pueden aplicarse al estudio realizado, haciendo la salvedad de lo indicado. Además se debe enfatizar que los valores de la prueba que se compara

están dados relativamente en función de otra, lo cual no se observó en algunas de las tesis.

Las pruebas diagnósticas serológicas fueron las más comparadas, posiblemente debido a la gran variedad de metodologías que actualmente se basan en principios inmunológicos. También muchas de las tesis compararon los resultados obtenidos de diversas tinciones con los de pruebas serológicas .

La mayoría de tesis tuvo como propósitos principales la determinación de la sensibilidad y especificidad de la(s) prueba(s) comparada(s), así como también la comparación de una prueba con el estándar de oro. Esto indica interés en establecer la utilidad de otras pruebas en el diagnóstico de enfermedades.

Dos tesis no describieron el tipo de pacientes que incluyeron en su estudio, esto no permite saber las características de la muestra empleada. De las tesis restantes, algunas únicamente señalaron la edad, sexo y ocupación de los pacientes, lo cual no es suficiente al comparar pruebas diagnósticas. Es importante que se describa el tipo de pacientes para establecer a qué tipo de población está dirigida la prueba que se está comparando. Lo ideal es incluir a pacientes que abarquen todo el espectro médico de la enfermedad: cuadros leves,

intermedios, severos, pacientes con enfermedades con las que se puede confundir la enfermedad, pero que comúnmente se asocian, sintomáticos y asintomáticos. Feinstein propone 4 grupos a evaluar, e indica la característica de la prueba que se evalúa con cada grupo de los pacientes que propone (10). Otras tesis incluyeron sólo a pacientes con diagnóstico previo de la enfermedad, pero de esta forma lo que se está haciendo es limitar el estudio de la prueba a los individuos que tienen una enfermedad grave o en fase muy avanzada. Esto sería correcto si la prueba se va a emplear sólo en pacientes de este tipo, lo cual no estaba indicado en las tesis. Al considerar sólo los casos avanzados, se eliminan los casos dudosos o en etapas iniciales de la enfermedad, por eso en la práctica el rendimiento de una prueba puede diferir a los resultados obtenidos cuando se comparó con el estándar de oro. Entonces, se confirma lo expuesto por Riegelman y Hirsch, que " puede ser engañoso hacer generalizaciones sobre la utilidad de una prueba cuando en el estudio de ella se incluyó sólo a individuos con enfermedades claramente definidas " (9).

Previo a la selección de los pacientes debe establecerse claramente los criterios de inclusión y exclusión a seguir, estos deben de conocerse para tener una idea del tipo de muestra con que se trabajó.

El tamaño de muestra (n) se obtuvo por medio de la fórmula general, en donde la mitad de ellas utilizó un Nivel de Confianza de 95 por ciento y la otra mitad de 90 por ciento, con un límite de error de 10 por ciento, que son los valores con los que regularmente se trabaja. Las tesis cuya variable es de amplia distribución trabajaron con un nivel de confianza de 95 por ciento, esto permitió tener mayor seguridad de que el estimador iba a contener al parámetro de la población. Se observó también que las tesis que trabajaron bajo un nivel de confianza de 90 por ciento, manejaron una variable menos dispersa en la población, por lo tanto un nivel de confianza de este valor fue adecuado. La varianza fue establecida mayormente por medio de la prevalencia de la enfermedad. Dicha prevalencia fue obtenida por estudios previos o utilizando la prevalencia ficticia (50:50). Al utilizar una prevalencia ficticia se abarca cualquier posible prevalencia que fuese la verdadera en la población, aunque la desventaja de su empleo es que aumentará el tamaño de muestra, y por lo tanto la cantidad de recursos para llevar a cabo el estudio.

Dos tesis no mencionan nada acerca de la prevalencia utilizada, debiendo describirla para así poder interpretar los valores predictivos positivo y negativo de la prueba, ya que la prevalencia está relacionada directamente con esos valores. Se observó que en algunas tesis no siempre se calculó los Valores Predictivos de la prueba, a pesar de

contar con la prevalencia de la enfermedad. La información que estos valores indican es importante, ya que el Valor Predictivo Positivo dice cuál es la probabilidad de que un paciente padezca la enfermedad cuando su resultado fue positivo, y por su lado el Valor Predictivo Negativo indica la probabilidad de no padecer la enfermedad si su resultado fue negativo.

El muestreo por conveniencia no es lo más adecuado en un estudio, ya que no tiene validez externa y no se puede generalizar sus resultados. Cuando se utiliza este tipo de muestreo sólo se puede evaluar la sensibilidad y especificidad de la prueba, sin embargo dos tesis cuyo muestreo fue por conveniencia, evaluaron además los valores predictivos de la prueba y en sus conclusiones no indicaron que su estudio sólo posee validez interna, y que sus resultados no pueden generalizarse. Por lo anterior, sus conclusiones son inapropiadas.

Una forma práctica y sencilla de presentar los resultados obtenidos en la comparación de las pruebas es por medio de una tabla de Contingencia de 2x2, como aparece en la mayor parte de las tesis revisadas, aunque no en todas. Esta tabla facilita el análisis de los resultados ya que a través de ella es posible calcular diversos índices como la sensibilidad, especificidad y valores predictivos de una prueba, prevalencia de la enfermedad etc. Debe entonces,

incluirse en el trabajo para saber cómo se calcularon dichos índices y para poder verificar o calcular otros que no se hayan determinado. Las guías para leer artículos acerca de pruebas diagnósticas aconsejan que si no hay una tabla de Contingencia o no es posible construir una con los resultados que se presentan, no vale la pena seguir leyendo el trabajo o artículo (12,13).

Pocas fueron las tesis que midieron el grado de acuerdo entre las pruebas: Prueba de Kappa. Las que lo hicieron utilizaron estándar de oro, aunque la prueba de Kappa puede aplicarse también cuando no se dispone del estándar de oro. Para su interpretación las tesis utilizaron adecuadamente los criterios de Fleiss (24). Lo útil de esta prueba es que el dato numérico que al final se obtiene, permite por medio de criterios ya establecidos, tener una idea rápida y general de la prueba que se compara y de su situación respecto otra prueba o respecto al estándar de oro, por medio de un valor numérico único que luego se interpreta. Se ha expuesto que la desventaja de utilizar índices que reducen la información a un sólo dato numérico, es que no se puede saber después las características de sensibilidad, especificidad, etc. que puede ser de importancia para conocer mejor la prueba. Además existen varios criterios de interpretación para el valor obtenido de la prueba de Kappa, los cuales difieren en su categorización, y que por lo tanto poseen discrepancias a la hora de indicar el grado de

acuerdo de las pruebas para ciertos valores numéricos. Los criterios de Landis y Koch tienen cinco categorías, mientras que los de Fleiss poseen tres categorías.

Otra prueba utilizada para determinar si existe o no asociación en los resultados obtenidos entre las pruebas comparadas fue la de Ji cuadrada, la cual no es adecuada ya que esta prueba se utiliza para saber si hay diferencia (o grado de asociación) entre medidas nominales de grupos independientes, pero en la comparación de pruebas diagnósticas, las muestras de los pacientes que se utilizan para cada una de las prueba, son las mismas, por lo tanto no se consideran tales muestras como grupos independientes, sino como grupos dependientes. Las mismas muestras que se analizan por el estándar de oro se analizan también por la(s) prueba(s) a comparar (recordando que debe ser un estudio ciego), por lo que se debe utilizar la prueba de Kappa y no la de Ji Cuadrada (1).

La interpretación de los resultados para su consecuente conclusión fue adecuada en la mayoría de las tesis, pero no en todas, debido a los aspectos mencionados anteriormente. Debido a esto algunos estudios tienen algunas conclusiones que no tienen validez.

En dos tesis no se pudo alcanzar el objetivo propuesto de comparar pruebas, una de ellas por limitaciones técnicas

de equipo, y otra por problemas de contaminación bacteriana en los sueros con los que iba a trabajar.

En resumen, los problemas y errores encontrados en las tesis revisadas son: no se enfatizó que los valores de la prueba que se compara están dados relativamente en función de otra prueba, al no haber utilizado el estándar de oro en la comparación; falta de descripción adecuada de los pacientes, no se mencionan criterios de inclusión y exclusión; limitaciones de los estudios a pacientes con enfermedad grave o avanzada; cálculo de valores predictivos sin mencionar la prevalencia de la enfermedad utilizada; falta de cálculo de los valores predictivos de la prueba a pesar de contarse con el dato de la prevalencia de la enfermedad; generalización de los resultados al haber muestreado por conveniencia; falta de construcción de una Tabla de Contingencia y el uso inapropiado de Ji Cuadrada en la comparación de pruebas diagnósticas.

Luego de determinar los anteriores errores, se elaboró una guía general que pueda dar la información al investigador que desee comparar este tipo de pruebas, para que así esté consciente de los factores y condiciones que debe tomar en cuenta dependiendo de los objetivos de su estudio.

IX. CONCLUSIONES

A. En las tesis revisadas se detectó los siguientes errores: no se enfatizó que los valores de la prueba que se compara están dados relativamente en función de otra prueba, al no utilizar el estándar de oro en la comparación; falta de descripción adecuada de los pacientes, no se mencionan criterios de inclusión y exclusión; limitaciones de los estudios a pacientes con enfermedad grave o avanzada; cálculo de valores predictivos sin mencionar la prevalencia de la enfermedad utilizada; falta de cálculo de los valores predictivos de la prueba a pesar de contarse con el dato de la prevalencia de la enfermedad; generalización de los resultados al haber muestreado por conveniencia; falta de construcción de una Tabla de Contingencia y el uso inapropiado de Ji Cuadrada en la comparación de pruebas diagnósticas.

B. Por medio de un muestreo por conveniencia se puede determinar únicamente los índices de sensibilidad y especificidad de una prueba. Con este tipo de muestreo el estudio carece de validez externa y no se puede generalizar sus resultados.

C. Si además de la sensibilidad y especificidad se va a establecer los valores predictivos de la prueba, entonces debe tomarse en cuenta y especificarse la prevalencia de la enfermedad en la población.

D. La prevalencia de la enfermedad afecta directamente a los valores predictivos. Si la prevalencia cambia, también ellos lo hacen.

E. Se elaboró una guía general que unifica la información y criterios que puedan orientar al investigador que desee comparar pruebas diagnósticas de medición nominal y respuesta dicotómica.

X. RECOMENDACIONES

Antes de realizar una comparación de pruebas diagnósticas debe tenerse claro los objetivos y alcance de los resultados que se pretende obtener, para así planificar adecuadamente los aspectos que llenen los requisitos con los que se cumplan los objetivos del estudio. Estos aspectos, como el uso de estándar de oro, la selección de los pacientes, la prevalencia de la enfermedad, interpretación de los índices estadísticos, etc., algunas veces se desconocen por lo que exhorta al investigador a leer la guía elaborada y la presente tesis, previo a la realización de su estudio.

XI. REFERENCIAS

1. Dawson-Saunders B, Trapp R. Bioestadística Médica. México: Editorial El Manual Moderno, 1993. 380p. (p.265-292)
2. Todd-Sanford-Davidsohn. Diagnóstico y Tratamiento Clínicos por el Laboratorio. 8a edición. España: Salvat Editores S.A., 1988. XXIV + 1774p. (p.55)
3. Murray RS. Estadística. México: McGraw-Hill, 1992. 339p. (p. 201-203)
4. McDonough JA, Sada E, et al. Microplate and dot immunoassays for the serodiagnosis of tuberculosis. J Lab Clin Med 1992;119(4):321-440
5. Arienti HM, Guignard SI, et al. Comparación de dos métodos serológicos para el diagnóstico de hidatidosis. Bol OPS 1996;121(3):221-225
6. Reglamento del Instituto de Investigaciones Químicas y Biológicas. USAC. Facultad de Ciencias Químicas y Farmacia. 1990. 6p. (p. 1-5)
7. Rodas BR. Evaluación de la investigación en la Escuela de Química Biológica según las tesis de 1979 a 1987. Guatemala: USAC (tesis de graduación, Facultad de Ciencias Químicas y Farmacia) 1990. 43p.
8. Alvarez FD. Uso del Análisis de regresión lineal en los trabajos de tesis *ad gradum* de la Escuela de Química Biológica. Guatemala: USAC (tesis de graduación, Facultad de Ciencias Químicas y Farmacia) 1992. 73p.
9. Riegelman RK, Hirsch RP. Cómo estudiar un estudio y probar una prueba: lectura crítica de la literatura médica. OPS. Publicación Científica 531, 1992. 256p (p. 95-131)
10. Feinstein AR. Clinical Biostatistics. USA: The C.V. Mosby Company, 1977. 468p (p. 214-225)
11. Fletcher RH, Fletcher SW, Wagner EH. Clinical Epidemiology: The Essentials. USA: Williams & Wilkins, 1982. 223p (p. 41-58)
12. Haynes RB. Clinical Epidemiological Rounds: How to read clinical journals: To learn about a diagnostic test. Trad. Lic. Francisco Mendizabal. J Can Med 1981;124.
13. Sackett DL, Haynes RB, Tugwell P. Clinical Epidemiology. USA: Library of Congress, Catalog Card No. 84-81905, 1985.

14. Griner PF, Mayewski RJ, Mushlin AI. Selection and interpretation of diagnostic tests and procedures: Principles and applications. *Ann Intern Med* 1981;296:1497-1501
15. Yerushalmy, J. Statistical Problems in assessing methods of medical diagnosis with special reference to X-ray techniques. *Pub Health Rep* 1947;62:1432-1449
16. Ahlbom A, Norell S. Introduction to Modern Epidemiology. USA: Epidemiology Resources Inc., 1990. 105p (p. 24-26)
17. Bennett BM. On comparisons of sensitivity, specificity and predictive value of a number of diagnostic procedures. *Biometric* 1972;28:793-800
18. Vecchio TJ. Predictive value of a single diagnostic test in unselected populations. *N Engl J Med* 1966;274:1171-1173
19. Sunderman FW. Laboratory suggestions: Probability computations for clinical interpretations of screening tests. *Am J Clin Pathol* 1971;55:105-111
20. Nissen-Meyer S. Evaluation of screening tests in medical diagnosis. *Biometrics* 1964;20:730-755
21. Feinstein AR. Clinical Judgement. USA: Robert E. Krieger Publishing Co., 1974.
22. Matute J. Representatividad y Confiabilidad de una muestra. Guatemala: Boletín Nutrición al Día, 1990;4(1):17-42.
23. Matute J. Cuántas repeticiones tengo que hacer en mi ensayo ?. Guatemala: Boletín Nutrición al Día, 1990;4(2):29-50.
24. Fleiss JL. Statistical Methods for Rates and Proportions. 2nd edition. USA: John Wiley & Sons, 1993. XVII+321p. (p. 217-222)
25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. USA: *Biometrics*, 1977;33:671-679.
26. Weiss NS. Clinical Epidemiology: The Study of the Outcome of Illness. USA: Oxford University Press, Inc. 1986. 235p (p 14-47)
27. Fescina R, Simini F, Belitzky R. Difusión: Evaluación de los procedimientos diagnósticos, aspectos metodológicos. *Salud Perinatal*. 1985;2(5):39-45

XII. A N E X O S

TABLA 2

Valores Predictivos de una prueba con sensibilidad de 80 %
y especificidad de 90 %.

| Prevalencia (%) | 1 | 10 | 50 | 90 |
|-------------------------|------|------|------|------|
| Valores Predictivos (%) | | | | |
| Positivo: | 7.5 | 47.1 | 88.9 | 98.6 |
| Negativo: | 99.8 | 97.6 | 81.8 | 33.3 |

Fuente: Riegelman RK, Hirsch RP. Cómo estudiar un estudio y probar una prueba: lectura crítica de la lectura médica. OBG: Publicación Científica 531. 1992. p. 120.

BOLETA DE RECOLECCION DE DATOS

Tesis: "Comparación de pruebas diagnósticas nominales y de respuesta dicotómica,"

No. de Tesis: _____

Pruebas a comparar: _____

Metodología: _____

a. _____

b. _____

Estándar de oro: SI, inciso: _____ No, se utilizó la prueba: _____

| | SI | NO |
|---|-------|-------|
| 1. Establece en sus objetivos la comparación de pruebas | _____ | _____ |
| 2. Indica el propósito de la comparación | _____ | _____ |
| a. Tipo de prueba según su fin | _____ | _____ |
| 3. Describe el tipo de pacientes que incluyó | _____ | _____ |
| 4. Cálculo correcto del tamaño de muestra: | _____ | _____ |
| a. Nivel de Confianza: _____ | | |
| b. Varianza: _____ | | |
| c. Límite de error: _____ | | |
| d. Forma de cálculo de la prevalencia: _____ | | |
| 5. Construcción correcta de Tabla de Contingencia | _____ | _____ |
| 6. Cálculo correcto de los índices de: | | |
| a. Sensibilidad | _____ | _____ |
| b. Especificidad | _____ | _____ |
| c. Valor predictivo positivo | _____ | _____ |
| d. Valor predictivo negativo | _____ | _____ |
| e. Prueba de Kappa | _____ | _____ |
| f. Otros: _____ | _____ | _____ |
| 7. Conclusiones acordes a los resultados | _____ | _____ |
| 8. Comentarios: _____ | _____ | _____ |

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE CIENCIAS QUIMICAS Y FARMACIA
ESCUELA DE QUIMICA BIOLOGICA
TRABAJO DE TESIS

ANALISIS ESTADISTICO A SEGUIR EN LA COMPARACION DE
METODOS DIAGNOSTICOS CUALITATIVOS: GUIA PARA LA
COMPARACION DE PRUEBAS.

Presentada por:

Rosa Carolina Joo León

A. INTRODUCCION

El gran desarrollo técnico de los últimos años ha incorporado a la práctica clínica una serie de pruebas diagnósticas. Esta diversidad de pruebas puede crear confusión por lo que puede desearse evaluar las características de una prueba.

Antes de iniciar la comparación de pruebas diagnósticas, el investigador debe tener claro los objetivos de su trabajo y el uso posterior a la comparación que le dará a la prueba.

Esta guía se refiere al análisis estadístico a seguir en la comparación de pruebas diagnósticas cuya medición sea nominal y de respuesta dicotómica, o sea la prueba cuyo resultado puede tomar una de dos opciones y se expresa en categorías o clasificaciones. Por ejemplo positivo o negativo, presencia o ausencia de la enfermedad, reactivo o no reactivo. Si se desea profundizar en los temas, se puede consultar la tesis: " Análisis estadístico a seguir en la comparación de métodos diagnósticos cualitativos en la Escuela de Química Biológica : Evaluación de las tesis de Químicos Biólogos elaboradas de 1988 a 1997 y elaboración de un instructivo general para la comparación de métodos " , elaborada por Rosa Carolina Joo León.

B. Estándar de oro

Como en toda comparación, debe establecerse un patrón de referencia. El estándar de oro es la prueba que discrimina sano y enfermo, o bien, positivo o negativo correspondientemente. Si se va a utilizar otra prueba que no es el estándar de oro, se debe tener presente que las características que se obtendrán de la prueba que se va a comparar, serán en base a las características que la prueba utilizada como referencia tiene respecto al estándar de Oro. Se debe enfatizar en el estudio que los valores de la prueba que se compara están dados relativamente en función de otra prueba, que no es es estándar de oro.

Se aconseja realizar un estudio ciego, en donde el investigador o quien aplica el instrumento diagnóstico no conoce los resultados previos de las muestras luego de correr el estándar de oro. De esta forma se evita el " sesgo de sospecha " (12,13).

C. Selección de la muestra

Es importante definir previamente a la selección de la muestra, los criterios de inclusión y/o exclusión de pacientes que delimitarán la población de interés, y se definirá la población a la que está dirigida la prueba, que idealmente, debe abarcar todo el espectro de la enfermedad que diagnóstica la prueba. Debe haber cuadros leves, intermedios y severos de la enfermedad, pacientes con

enfermedades con las que se puede confundir, pero que comúnmente se asocia, sintomáticos y asintomáticos.

D. Cálculo del tamaño de la muestra

Para el cálculo del tamaño de la muestra se deben tener en cuenta varios factores como el tipo de estudio a realizar, la magnitud de los errores tipo I y II a trabajar, si se realizará una prueba de hipótesis, o el nivel de confianza para una estimación. Para ello el investigador puede consultar con un estadístico y determinar así el número suficiente de pacientes que debe incluir en su estudio.

En términos generales, si únicamente se va a determinar la sensibilidad y especificidad de una prueba, se puede seleccionar el tamaño de la muestra *a priori* (por conveniencia), aunque este tipo de muestreo carece de validez externa y sus resultados no pueden generalizarse.

Ahora si se va a calcular además los valores predictivos de la prueba, se debe tomar en cuenta la prevalencia de la enfermedad, y el tamaño de muestra puede calcularse con la siguiente fórmula:

$$n = \frac{NC \cdot \sigma^2}{\Delta^2}$$

Donde: NC = Nivel de Confianza

σ^2 = Varianza, que puede sustituirse en la fórmula por $\sigma^2 = p q$, en donde p es la prevalencia de la enfermedad en la población (dado en frecuencia relativa), y $q = 1 - p$.

Δ = Límite de error.

Se debe tomar en cuenta que el tamaño de muestra aumentará al aumentar el nivel de confianza y la varianza, y al disminuir el límite de error en el estudio. Contrariamente, el tamaño de muestra disminuirá si se trabaja con un nivel de confianza y varianza menores, y con un límite de error mayor.

1. Formas de Calcular la Prevalencia de la Enfermedad

La prevalencia de una enfermedad se obtiene de forma general con la siguiente fórmula:

$$\text{Prevalencia} = p = \frac{\text{Número de personas positivas}}{\text{Número total de personas expuestas}}$$

Otras formas de obtener la prevalencia de la enfermedad para poder calcular n es por medio de un estudio piloto, en el que se estime "tentativamente", un valor de prevalencia. Se puede consultar la literatura nacional sobre investigaciones recientes en la población a la que se destina la prueba, o bien utilizar una prevalencia ficticia

del 50 por ciento, la cual dará el máximo de muestra posible.

E. Cálculo de los Índices Estadísticos de la Prueba

Los resultados obtenidos de las pruebas a comparar pueden tabularse en una Tabla de Contingencia de 2X2.

1. Tabla de Contingencia

| Prueba en estudio | Prueba de Oro Enfermos | Prueba de oro Sanos |
|-------------------|---|---|
| Positivos | a = No. enfermos y positivos. Verdaderos positivos. | b = No. de sanos y positivos. Falsos positivos. |
| Negativos | c = No. enfermos y negativos. Falsos Negativos. | d = No. de sanos y negativos. Verdaderos Negativos. |

En las columnas se muestran las condiciones verdaderas del paciente (o sea los resultados según la prueba de oro) y en las filas los resultados de la prueba diagnóstica a comparar. Las cuatro celdas interiores indican si el paciente ha sido correcta (celdas a y d) o falsamente clasificado (celdas b y c) (10). En estas tablas se registran las frecuencias observadas.

2. Índices estadísticos

$$\text{Sensibilidad} = \frac{a}{a + c}$$

$$\text{Especificidad} = \frac{d}{b + d}$$

Valor predictivo positivo = $a / a + b$

Valor predictivo negativo = $d / c + d$

Exactitud = $a + d / a + b + c + d$

Prevalencia = $a + c / a + b + c + d$

a. **Sensibilidad:** Se refiere a la capacidad de la prueba para detectar la enfermedad. Es la tasa de positivas verdaderas, se expresa como porcentaje (1,9,16,17).

b. **Especificidad:** Se refiere a la capacidad de la prueba de identificar correctamente a pacientes que no tienen la enfermedad. Es la proporción de individuos sanos cuya prueba es negativa. Se expresa como porcentaje (1,9,16,17).

c. **Valor Predictivo Positivo:** Indica el porcentaje de pacientes con resultado positivo que de hecho padecen la enfermedad. Es la fracción numérica de los resultados positivos que son positivos verdaderos. Su utilidad radica en indicar cuál es la probabilidad de que haya enfermedad si el resultado de la prueba es positivo (1,9,10,12,13,16).

d. **Valor Predictivo Negativo:** Indica el porcentaje de pacientes que no tienen la enfermedad y su resultado fue negativo. Es la fracción numérica de los resultados negativos que son negativos verdaderos. Indica la

probabilidad de que no haya enfermedad si el resultado es negativo (1,9,10,12,13,16).

e. **Exactitud:** Significa que la prueba producirá resultados semejantes al verdadero valor anatómico, fisiológico o bioquímico. La exactitud es una propiedad necesaria de una buena prueba, ya que con ella se elimina todo tipo de sesgos como efectos al azar y errores sistemáticos (9,11).

f. El Índice de Validez

LLamado también Índice de Eficiencia de una prueba. Indica el porcentaje de clasificaciones correctas. Este se calcula como:

$$(a + d) / N * 100$$

F. Cuidados en la interpretación

1. Al no utilizar el estándar de oro en la comparación, debe indicarse que los valores de la prueba se calcularon en función de otra prueba que no es el estándar de oro.

2. Si no se abarcó todo el espectro médico de la enfermedad, significa que el estudio de la prueba se limitó a cierto tipo de pacientes. Debe establecerse en el estudio los criterios de inclusión y exclusión seguidos, para determinar el tipo de pacientes al que se dirigió la prueba.

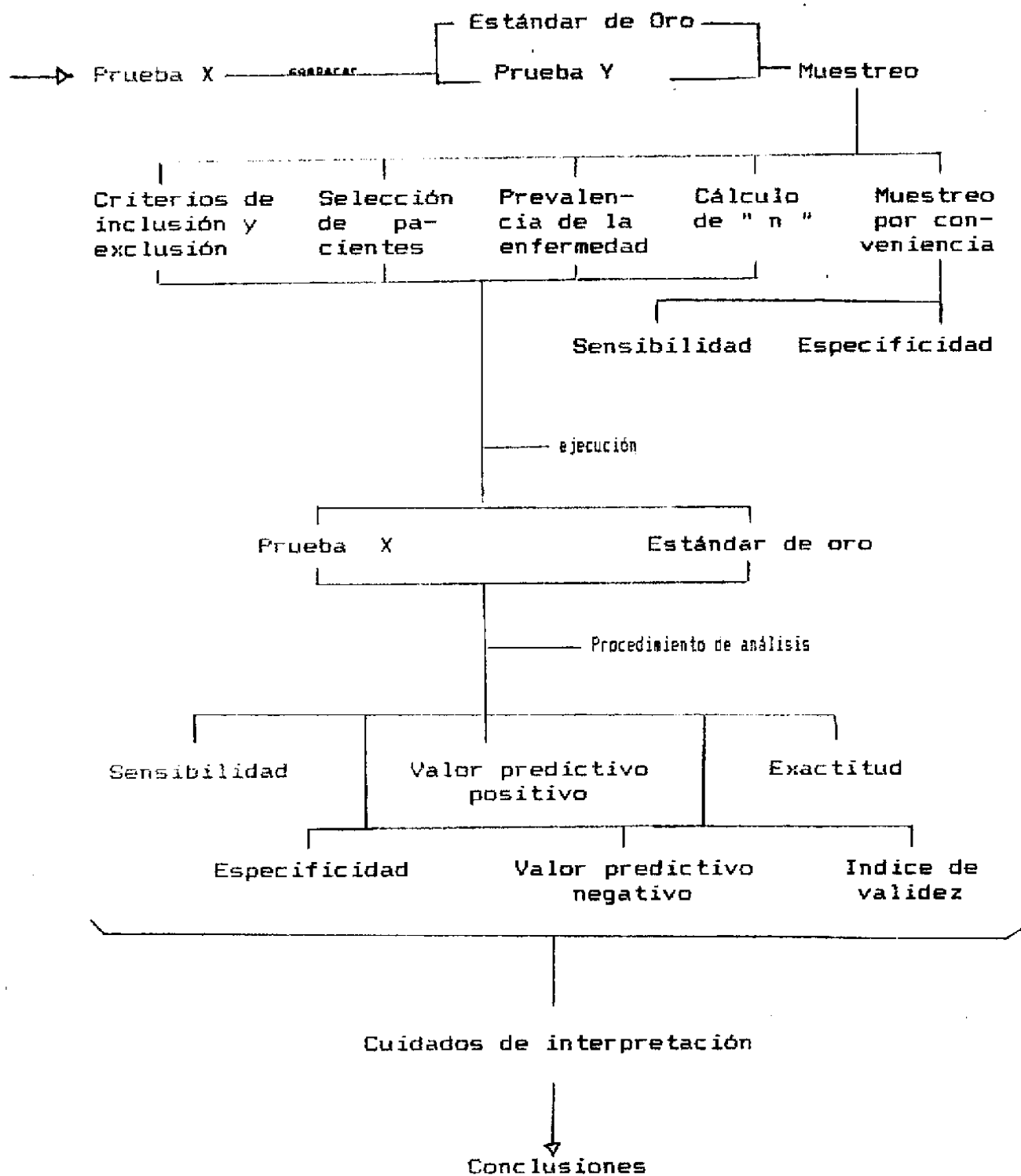
3. Al utilizar el muestreo por conveniencia el estudio sólo tiene validez interna, y se puede calcular únicamente la sensibilidad y especificidad de la prueba. No puede generalizarse los resultados.

4. La prevalencia de la enfermedad afecta directamente los valores predictivos de la prueba; si la prevalencia cambia, también cambian ellos.

G. Pasos a seguir en la comparación de pruebas diagnósticas de medición nominal y respuesta dicotómica

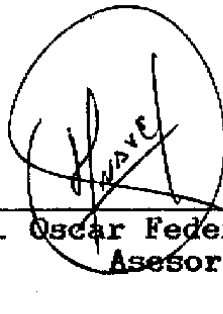
1. Establecer el estándar de oro o la prueba con la cual se hará la comparación.
2. Muestreo:
 - a. Criterios de inclusión y exclusión
 - b. Selección de pacientes
 - c. Cálculo del tamaño de muestra
 - d. Muestreo por conveniencia (consideraciones especiales)
3. Prevalencia de la enfermedad:
 - a. Formas de cálculo:
 - i. Estudio Piloto
 - ii. Literatura Nacional reciente
 - iii. Prevalencia ficticia
4. Procedimiento de análisis estadístico:
 - a. Tabla de Contingencia
 - b. Sensibilidad
 - c. Especificidad
 - e. Valor Predictivo Positivo
 - f. Valor Predictivo Negativo
 - g. Exactitud
 - h. Índice de Validez
5. Cuidados en la interpretación.

H. Esquematización de los pasos a seguir

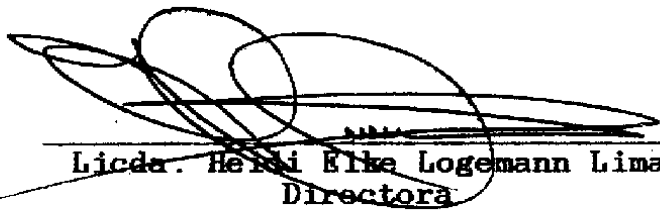




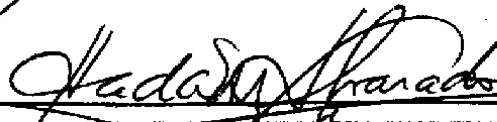
Rosa Carolina Joo León
Tesisista



Lic. Oscar Federico Nave
Asesor



Licda. Heidi Elke Logemann Lima
Directora



Licda. Hada Marieta Alvarado Beteta
Decana