PNCQ
Programa Nacional
de Controle de Qualidade

Diseño de estudios sobre pruebas médicas

c

Curso de formación básica em investigación clínica y epidemiológica 2023-2024

GCIAMT

Amadeo Sáez-Alquezar

2024

PNCQ
Programa Nacional de Controle de Qualidade

Pruebas Médicas

- **Qué son y para que sirven.**
  - Detectar un factor de riesgo
  - Diagnosticar una enfermedad
  - Valorar el pronóstico de un paciente
- **Siguen un protocolo**
  - Hipótesis
  - Diseño
  - Método analítico
  - Aprobación por un comité de ética

PNCQ
Programa Nacional de Controle de Qualidade

Investigaciones clínicas / Pruebas médicas

Anatomía de la investigación

Protocolos

- Preguntas temas del estudio.
- El diseño.
- Los participantes.
- Las mediciones/determinaciones.
- Cálculo del tamaño de la muestra.

Fisiología de la investigación

Como funciona

- Proporcionando inferencias y deducciones válidas.
- Lo que ocurrió en la muestra en estudio.
- Como se pueden generalizar los hallazgos del estudio a las personas que no participaron del mismo.
- Reducir al máximo los errores aleatorios y sistemáticos que amenacen a las conclusiones basadas em las inferencias.

Hulley SB, Newman TB & Cummings SR, 2014

PNCQ
Programa Nacional de Controle de Qualidade

Pruebas Médicas

- **Qué son y para que sirven.**
  - Detectar un factor de riesgo
  - Diagnosticar una enfermedad
  - Valorar el pronóstico de un paciente
- **Pruebas médicas, centradas en parámetros *descriptivos*.**
  - **Sensibilidad, especificidad** y **cocientes de verosimilitudes**, con sus correspondientes **intervalos de confianza**
  - Difieren de los Estudios de observación que Identifican asociaciones estadísticamente significativas que representan relaciones causales.
    - La causalidad es muchas veces irrelevante.

PNCQ
Programa Nacional de Controle de Qualidade

Aspectos importantes para el diseño de las Pruebas médicas

- Utilidad
- Espectro de gravedad
- Sesgo de espectro
- Enmascaramiento
- Fuentes de variación
- Métodos de referencia
- Conceptos estadísticos:
  - X, DE, CV, Z escore, limites de confianza.
  - Precisión, Exactitud, Validez.
- Reproducibilidad
- Diseños
- Escalas de medición
- Variables categóricas / continuas
- Exactitud de las pruebas

- Variable predictiva / de respuesta
- Conceptos estadísticos:
  - Sensibilidad / Especificidad / VPP / VPN / curva ROC / cocientes de verosimilitudes, tablas de contingencia
  - Concordancia / Índice Kappa.
- Riesgo absoluto / cociente de riesgo
- Mejoría por reclasificación / Reglas de predicción clínica
- Resultados en las decisiones clínicas
- Errores:
  - Diseño y análisis / Tamaño de la muestra
  - Exclusión incorrecta / Resultados limítrofes
  - Sesgo de verificación / diferencial
- Ayuda por la Inteligencia Artificial

Sáez-Alquezar A

PNCQ
Programa Nacional de Controle de Qualidade

Determinación de la utilidad de una Prueba

- Para que una prueba sea útil debe superar una serie de preguntas de dificultad creciente que analizan su **reproducibilidad, exactitud, viabilidad** y, lo que es más importante, sus **efectos sobre las decisiones clínicas** y los **resultados**
- Las respuestas favorables a todas estas cuestiones son criterios necesarios, aunque no suficientes, para que merezca la pena realizar una prueba.
- Es poco probable que la Prueba sea útil:
  - Si ofrece resultados muy diferentes dependiendo de quién la realiza o dónde se realiza.
  - Si la prueba rara vez proporciona nueva información.
  - Aunque afecte a las decisiones, si estas no mejoran las respuestas clínicas de los pacientes.

PNCQ
Programa Nacional
de Controle de Qualidade

Determinación de la utilidad de una Prueba

Preguntas para determinar la utilidad de una Prueba Médica, posibles diseños para responderlas y Estadísticos para describir los resultados

Sáez-Alquezar A

PNCQ
Programa Nacional de Controle de Qualidade

Preguntas para determinar la utilidad de una Prueba Médica, posibles diseños para responderlas y Estadísticos para describir los resultados

¿Qué reproducibilidad tiene la prueba?	Estudios intra- e Inter observador, e intralaboratorio y entre laboratorios	Proporción de acuerdo, k, coeficiente de variación, media y distribución de diferencias
¿Qué exactitud tiene la prueba?	Diseños transversales, de casos y testigos o de cohortes, en los que se compara el resultado de una prueba con un método de referencia	Sensibilidad, especificidad, valor predictivo positivo y negativo, curvas de eficacia diagnóstica y cocientes de verosimilitudes
¿Con que frecuencia los resultados de las pruebas afectan a las decisiones clínicas?	Estudios de rendimiento diagnóstico, estudios de toma de decisiones clínicas antes y después de la prueba	Proporción anormal, proporción con resultados discrepantes, proporción de pruebas que producen cambios en las decisiones clínicas; coste por resultado anormal o por cambio de decisión

Newman, TB et al, 2013.

PNCQ Programa Nacional de Controle de Qualidade

Preguntas para determinar la utilidad de una Prueba Médica, posibles diseños para responderlas y Estadísticos para describir los resultados

¿Cuales son los costes, riesgos y aceptabilidad de la prueba?

Estudios prospectivos o retrospectivos

Costes medios, proporciones de los que sufren efectos adversos, proporciones de los dispuestos a que se les realice la prueba

Si se realiza la prueba ¿se mejora la respuesta clínica o se producen efectos adversos?

Ensayos aleatorizados, estudios de cohortes o de casos y testigos en los que la variable predictiva sea la realización de la prueba y las respuestas incluyan morbilidad, mortalidad o costes relacionados bien con la enfermedad o bien con su tratamiento

Cocientes de riesgo, cocientes de posibilidades, cocientes de riesgos instantáneos, número que es necesario tratar, incidencias y cocientes de respuestas deseables e indeseables

Newman, TB et al, 2013.

PNCQ
Programa Nacional de Controle de Qualidade

Determinación de la utilidad de una Prueba

- Cuando el uso de una prueba mejora los resultados de los pacientes estudiados.
  - Los estudios de si realizar una prueba mejora la evolución de los pacientes son los más difíciles de realizar.
- Posibles efectos de una prueba sobre los resultados clínicos suelen inferirse comparando la exactitud, la seguridad y los costes con los de las pruebas ya existentes.
- Al desarrollar una nueva prueba diagnóstica o pronóstica, puede que valga la pena considerar qué aspectos de la práctica actual son los que más necesitan una mejora.
  - Por ejemplo, ¿son las pruebas actuales poco fiables, caras, peligrosas o difíciles de realizar?

PNCQ
Programa Nacional de Controle de Qualidade

Aspectos generales de los estudios de pruebas médicas

Espectro de la gravedad de la enfermedad y de los resultados de la prueba.

En el contexto de las pruebas médicas y la evaluación de la gravedad de una enfermedad, el "espectro de la gravedad" se refiere a la gama completa de manifestaciones clínicas que pueden ocurrir como resultado de una enfermedad.

Esto incluye desde formas leves hasta formas graves o críticas de la enfermedad.

Ejemplo: en el caso del COVID-19, el espectro de la gravedad puede variar desde casos asintomáticos o leves hasta casos graves que requieren hospitalización e incluso pueden ser mortales.

PNCQ
Programa Nacional de Controle de Qualidade

Aspectos generales de los estudios de pruebas médicas

Espectro de la gravedad de la enfermedad y de los resultados de la prueba.

En las primeras fases del desarrollo de una prueba diagnóstica, puede ser razonable investigar si una prueba permite distinguir entre los pacientes con enfermedad manifiesta en estadios avanzados y los testigos sanos;

(Si la respuesta es no, se busca otra prueba)

Posteriormente, cuando la pregunta de la investigación aborde la utilidad clínica de la prueba, el espectro de enfermedad y ausencia de enfermedad debe ser representativo de las personas en las que se utilizará la prueba.

Ejemplo: una prueba que se haya desarrollado comparando pacientes con cáncer pancreático sintomáticos con testigos sanos se podría evaluar posteriormente en una muestra más difícil, pero clínicamente realista, como pacientes consecutivos con dolor abdominal o pérdida de peso de causa no determinada.

PNCQ
Programa Nacional de Controle de Qualidade

Aspectos generales de los estudios de pruebas médicas

Espectro de la gravedad de la enfermedad y de los resultados de la prueba.

El "sesgo de espectro", por otro lado, se refiere a una distorsión sistemática en la selección de los sujetos de estudio o pacientes que pueden sesgar los resultados de una investigación o evaluación clínica.

Este tipo de sesgo puede ocurrir cuando los sujetos seleccionados para un estudio no representan adecuadamente la diversidad de gravedad de la enfermedad en la población general.

Ejemplo: si un estudio clínico sobre COVID-19 solo recluta pacientes con casos leves o moderados de la enfermedad y excluye casos graves o críticos, habrá un sesgo de espectro y los resultados no serán aplicables a toda la población afectada por la enfermedad.

Es importante tener en cuenta el espectro completo de la enfermedad al diseñar estudios clínicos, realizar evaluaciones diagnósticas y desarrollar estrategias de tratamiento para garantizar que se capturen todas las variaciones de la enfermedad y se minimice el sesgo de espectro.

Sáez-Alquezar A

PNCQ
Programa Nacional de Controle de Qualidade

Aspectos generales de los estudios de pruebas médicas

Espectro de la gravedad de la enfermedad y de los resultados de la prueba.

- Sesgo de Espectro
  - Por un espectro inadecuado de resultados de la prueba.
  - Por un espectro inadecuado de la enfermedad
- **Ejemplo:** Considere un estudio de acuerdo Inter observador en radiólogos que interpretan mamografías.
  - Si se les pide que clasifiquen las imágenes como normales o anormales, su acuerdo será mucho mayor si las imágenes «positivas» que selecciona el investigador para su estudio se eligieron porque son claramente anormales, y las imágenes «negativas» se seleccionan porque están libres de cualquier anomalía sospechosa
- **Otros:** Examen microscópico de cortes histológicos o de frotis para detección de parásitos.

PNCQ
Programa Nacional de Controle de Qualidade

Aspectos generales de los estudios de pruebas médicas

Importancia del enmascaramiento

El enmascaramiento, también conocido como cegamiento o blinding en inglés, es una técnica utilizada en estudios de investigación médica para minimizar el sesgo y garantizar la validez de los resultados.
Consiste en ocultar cierta información a los participantes, investigadores o evaluadores que podrían influir en la interpretación de los resultados.

PNCQ® Programa Nacional de Controle de Qualidade

Aspectos generales de los estudios de pruebas médicas

Importancia del enmascaramiento

Existen varios tipos de enmascaramiento:

1.Enmascaramiento simple: En este caso, los participantes o los evaluadores no conocen la asignación de tratamiento, pero los investigadores sí conocen quién está recibiendo el tratamiento activo y quién el placebo.

2.Enmascaramiento doble: Tanto los participantes como los investigadores que interactúan con ellos desconocen quién está recibiendo el tratamiento activo y quién el placebo. Este tipo de enmascaramiento es común en ensayos clínicos aleatorizados.

3.Enmascaramiento triple: Además de ocultar la asignación del tratamiento a los participantes y los investigadores, también se oculta a los estadísticos que analizan los datos.

La importancia del enmascaramiento : reduce la posibilidad de sesgo tanto en la selección de los participantes como en la interpretación de los resultados. Al mantener a los participantes y a los investigadores ciegos respecto a la asignación del tratamiento, se minimiza la probabilidad de que los resultados sean influenciados por las expectativas o preferencias preconcebidas.

El enmascaramiento es especialmente crucial en estudios de pruebas médicas, donde se están evaluando la eficacia y seguridad de tratamientos nuevos o existentes. Ayuda a garantizar que los resultados reflejen de manera precisa el verdadero impacto del tratamiento y facilita la comparación entre grupos de tratamiento y grupos de control de manera imparcial.

Sáez-Alquezar A

PNCQ Programa Nacional de Controle de Qualidade

Aspectos generales de los estudios de pruebas médicas

Importancia del enmascaramiento

- Los investigadores deben ocultar otra información sobre el paciente al que se está estudiando a las personas que interpreten los resultados de la prueba. **Ejemplos:**
- a) Estudio de la contribución de la ecografía al diagnóstico de la apendicitis. Los profesionales que interpreten las ecografías no deben conocer los resultados de la historia clínica ni de la exploración física
- b) Igualmente, el anatomopatólogo que realice la determinación final de quién tiene, o no, apendicitis (método de referencia con el que se compararán los resultados de la ecografía) no debe saber los resultados de la exploración ecográfica.

El enmascaramiento evita que el sesgo, las ideas preconcebidas y la información de otras fuentes distintas a la prueba afecten a estos juicios.

Sáez-Alquezar A

PNCQ
Programa Nacional de Controle de Qualidade

Aspectos generales de los estudios de pruebas médicas

Fuentes de variación, posibilidad de generalizar y esquema de muestreo

- Para algunas preguntas de la investigación, las diferencias entre los pacientes son la principal fuente de variación en los resultados de una prueba.
- **Ejemplo:** algunos lactantes con bacteriemia tendrán leucocitosis, mientras que otros no.
  - No se espera que la proporción de lactantes con bacteriemia y con leucocitosis varíe mucho según el laboratorio que realice el hemograma.

PNCQ
Programa Nacional de Controle de Qualidade

Aspectos generales de los estudios de pruebas médicas

Fuentes de variación, posibilidad de generalizar y esquema de muestreo

- Los resultados de muchas pruebas dependen de la persona que realiza las pruebas o del entorno en que se han realizado.
- **Ejemplo:** la sensibilidad, la especificidad y la fiabilidad Inter observador para la interpretación de las mamografías depende de la capacidad y de la experiencia del lector, así como de la calidad del equipo.

Cuando la exactitud puede variar de un lector a otro o de un equipo a otro o de un centro a otro, es útil estudiar a diferentes lectores, equipos e instituciones para poder evaluar la homogeneidad de los resultados.

PNCQ
Programa Nacional de Controle de Qualidade

Aspectos generales de los estudios de pruebas médicas

Método de referencia para el diagnóstico

- **Enfermedades que tienen un método de referencia.**
  - indica la presencia o ausencia de la enfermedad en estudio.
  - estudio anatomopatológico de una muestra de una biopsia para diagnosticar cáncer.
- **Enfermedades, donde el método de referencia es una definición.**
  - Enfermedad coronaria como una obstrucción del 50% de al menos una arteria coronaria principal observada mediante angiografía coronaria.
- **Enfermedades que necesitan un número mínimo de síntomas, signos o alteraciones analíticas para cumplir los criterios que definen la enfermedad.**
  - Enfermedades reumatológicas

Sáez-Alquezar A

PNCQ
Programa Nacional de Controle de Qualidade

Aspectos generales de los estudios de pruebas médicas

Método de referencia para el diagnóstico

- Por supuesto, si cualquier síntoma, signo o prueba de laboratorio utilizados para diagnosticar una enfermedad se utilizan como parte del método de referencia, un estudio que los compare con ese método de referencia puede hacer que parezcan falsamente buenos.
- Esto se denomina sesgo de incorporación, porque se incorpora la prueba que se está estudiando al método de referencia; evitar este sesgo es uno de los motivos que se han señalado previamente para el enmascaramiento

PNCQ
Programa Nacional de Controle de Qualidade

Aspectos generales de los estudios de pruebas médicas

Método de referencia para el diagnóstico

- Es muy importante considerar si el método de referencia es realmente de referencia.
- Si el método de referencia es imperfecto, puede hacer que una prueba parezca peor de lo que realmente es (si en realidad la prueba tiene mejor rendimiento que el método de referencia) o mejor de lo que realmente es (si la prueba índice comete los mismos errores que el método de referencia).

PNCQ
Programa Nacional de Controle de Qualidade

Algunos conceptos estadísticos

- Media (M)
- Desviación Estándar (DE)
- Coeficiente de Variación (CV)
- Z Escore
- Precisión
- Exactitud
- Validez

Conceptos de Estadística

Media: Tendencia central de distribución.
Está relacionadas con la exactitud

$$X = \Sigma xn \ / \ n$$

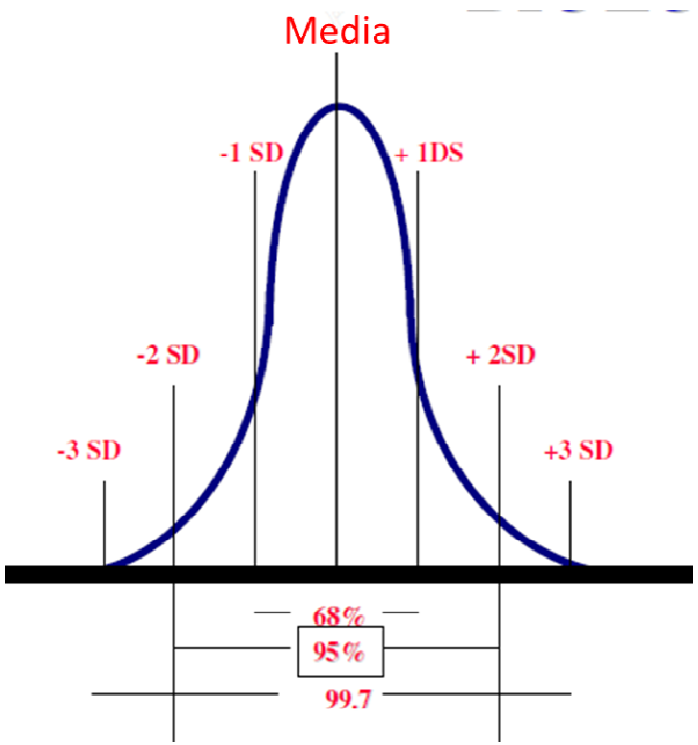
Desviación Estándar: Describe la dispersión al rededor de la media.
Medida de precisión

$$DE = \sqrt{\frac{\Sigma (Xn - \bar{X})^2}{n - 1}}$$

Coeficiente de variación (CV) %: Desviación estándar como % de la media

$$CV = (DE / X)\ 100$$

Distribución Normal



Unidades Desviación Estándar

Imaginemos N determinaciones por un ensayo Cualitativo, en una misma muestra de suero.
Los valores em cada determinación son un índice que corresponde a la lectura / valor de corte

$$\frac{\text{Índice (L/Co) - M}}{\text{DE}} = Z$$

Se obtiene a partir de la variabilidad analítica de una serie de controles.
"Z Score"

Sáez-Alquezar A

Ajuste de los SCI (+)

SCI (+)							
N	L	CO	L/CO	Media	L/CO - Média	DE	Z
1	0,711	0,187	3,80	3,913	0,111	0,240	-0,46
2	0,75	0,187	4,01	3,913	0,097	0,240	0,41
3	0,685	0,187	3,66	3,913	0,250	0,240	-1,04
4	0,701	0,187	3,75	3,913	0,165	0,240	-0,69
5	0,723	0,187	3,87	3,913	0,047	0,240	-0,20
6	0,695	0,176	3,95	3,913	0,036	0,240	0,15
7	0,684	0,176	3,89	3,913	0,027	0,240	-0,11
8	0,772	0,176	4,39	3,913	0,473	0,240	1,97
9	0,719	0,176	4,09	3,913	0,172	0,240	0,72
10	0,697	0,176	3,96	3,913	0,047	0,240	0,20
11	0,651	0,171	3,81	3,913	0,106	0,240	-0,44
12	0,745	0,171	4,36	3,913	0,443	0,240	1,85
13	0,719	0,171	4,20	3,913	0,291	0,240	1,21
14	0,687	0,171	4,02	3,913	0,104	0,240	0,43
15	0,653	0,171	3,82	3,913	0,095	0,240	-0,39
16	0,702	0,171	4,11	3,913	0,192	0,240	0,80
17	0,761	0,192	3,96	3,913	0,050	0,240	0,21
18	0,643	0,192	3,35	3,913	0,564	0,240	-2,35
19	0,687	0,192	3,58	3,913	0,335	0,240	-1,40
20	0,712	0,192	3,71	3,913	0,205	0,240	-0,85
Media	0,70485	0,192	3,91				
DE	0,034053	0,192	0,240				
CV (%)			6,1				

Z = (L/CO – Media) / DE

Test de Grubbs para detectar outliers

N	Z crítico
10	2,29
11	2,34
12	2,41
13	2,46
14	2,51
15	2,55
16	2,59
17	2,62
18	2,65
19	2,68
20	**2,71**

Z = (Valor – Media) / Desviación Estándar

PNCQ
Programa Nacional de Controle de Qualidade

Conceptos de estadística

- **PRECISIÓN**: mejor concordancia en determinaciones repetitivas.
- **IMPRECISIÓN:** desviación estándar o coeficiente de variación de los resultados de un conjunto de mediciones en replicado, cuando el procedimiento está operando en condiciones estables.

Error Aleatorio

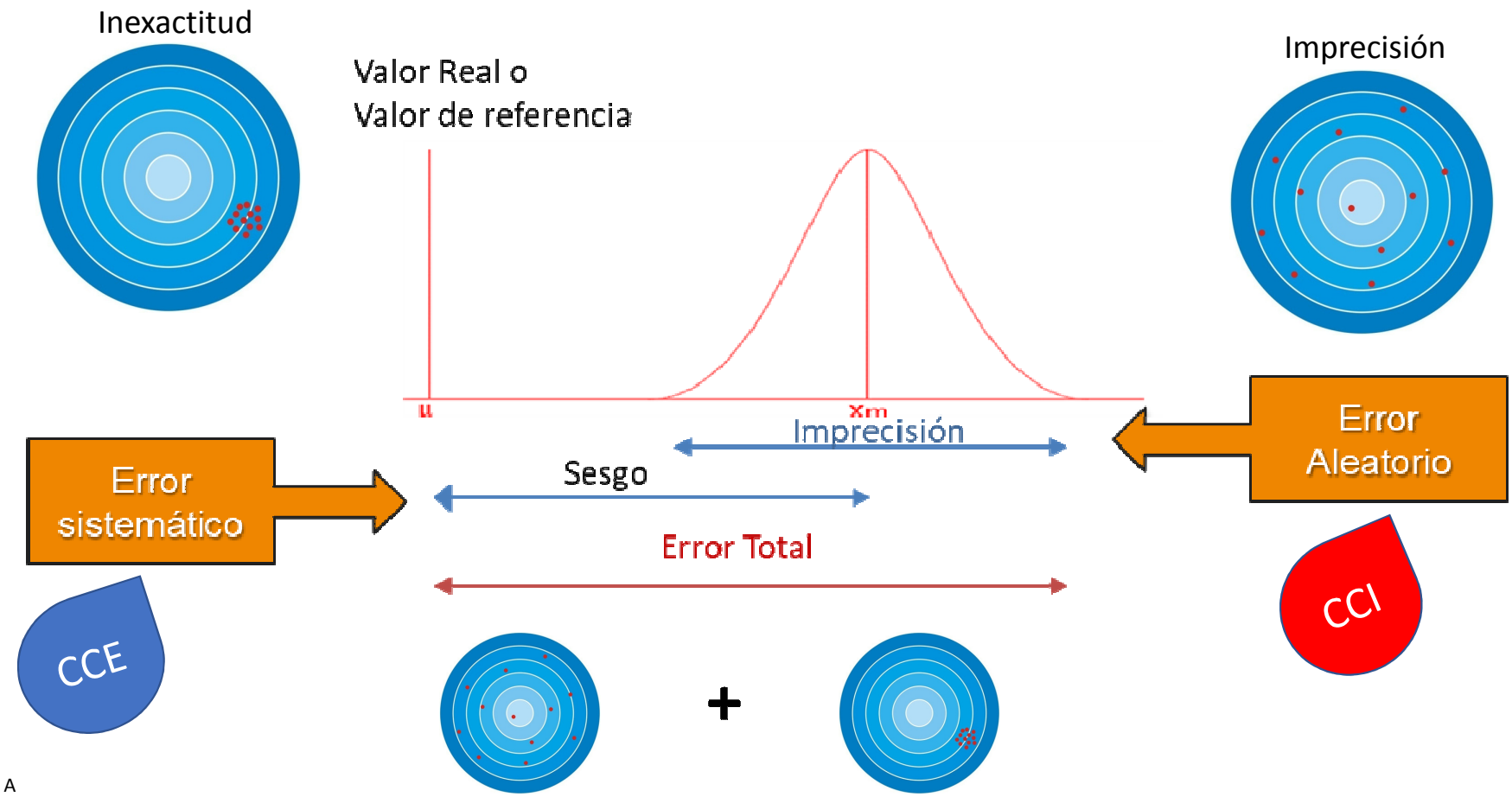
- **EXACTITUD** : concordancia entre la mejor estimativa de un valor y el valor verdadero.
- **INEXACTITUD** diferencia numérica entre la media de un conjunto de mediciones en replicado y el valor verdadero. Esta diferencia (positiva o negativa) puede ser expresada en la unidad de medición o como porcentaje del valor verdadero.

Error Sistemático

AS4

PNCQ Programa Nacional de Controle de Qualidade

Errores observados en el laboratorio

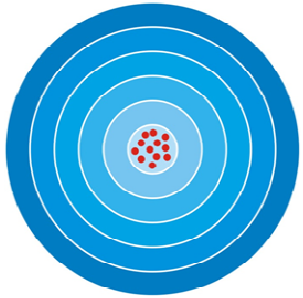


Sáez-Alquezar A

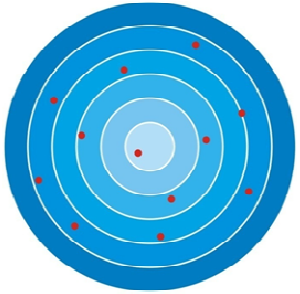
Slide 28

AS4 Amadeo Saez; 11/07/2022

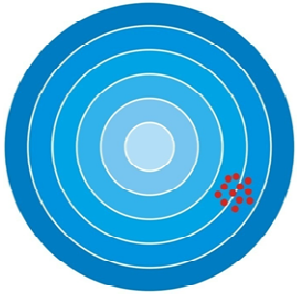
Metodologías: Precisión / Exactitud



Buena Precisión
Buena exactitud



Mala Precisión
Mala exactitud



Buena Precisión
Mala exactitud

Conceptos estadísticos (Precisión)

La precisión: es el grado en que una variable es reproducible (con casi el mismo valor cada vez que se mida).

*Una balanza puede medir el peso corporal con gran precisión.
*Una entrevista para medir la calidad de vida puede producir valores variando de observador u ocasión a otro.
*La precisión tiene una gran influencia en la potencia de un estudio.
*Cuanto más precisa la medida, mayor será la potencia estadística con un tamaño de la muestra concreto para calcular valores medios y comprobar hipótesis.

La precisión (reproducibilidad, fiabilidad o consistencia) es una función del **error aleatorio** (variabilidad por el azar); cuanto mayor sea el error, menos precisa será la medida.

Existen tres fuentes principales de error aleatorio al realizar medidas:

- **Variabilidad del observador:** se debe al observador, e incluye factores como escoger palabras en una entrevista o tener habilidad para usar un instrumento mecánico.
- **Variabilidad del instrumento:** se debe al instrumento, e incluye factores ambientales cambiantes (p. ej., temperatura), el desgaste de los componentes mecánicos, lotes diferentes de reactivos, etc.
- **Variabilidad del participante:** se debe a la variabilidad biológica intrínseca de los participantes en el estudio, que no se relaciona con las variables en estudio, como la variabilidad debida a la hora del día de las mediciones o el tiempo desde la última medicación

PNCQ
Programa Nacional de Controle de Qualidade

Conceptos estadísticos (Precisión)

Evaluación de la precisión

La precisión se valora como la reproducibilidad de mediciones repetidas.

*Comparando mediciones realizadas por la misma persona (reproducibilidad intra observador)
*Comparando mediciones realizadas por personas diferentes (reproducibilidad entre observadores).
*Se puede evaluar para un instrumento o entre diversos instrumentos.

La reproducibilidad de las variables continuas se expresa, a menudo, como la desviación típica intraobservador o el coeficiente de variación (desviación típica intraobservador dividida por la media).
Para las variables categóricas se utiliza a menudo la concordancia porcentual, el coeficiente de correlación intraclase y el estadístico kappa.

Conceptos estadísticos (Exactitud)

- **EXACTITUD:** La exactitud de una variable es el grado en que representa el valor verdadero.
  - La exactitud y la precisión no siempre están unidas pero en la mayoría de los casos andan lado a lado. Medidas para mejorar la precisión acaban mejorando también la exactitud.
  - La exactitud depende del error sistemático (sesgo); cuanto mayor sea el error, menos exacta será la variable.

PNCQ Programa Nacional de Controle de Qualidade

Conceptos estadísticos (Exactitud)

- Cada una de las tres clases principales de error de medición observadas sobre la precisión tiene su equivalente para la exactitud:
  - **Sesgo del observador:** Es una deformación, consciente o inconsciente, de la percepción o la notificación de la medida por el observador. Puede representar errores sistemáticos en el modo en que se utiliza un instrumento, como la tendencia a redondear a la baja las determinaciones de la presión arterial, o utilizar preguntas dirigidas en la entrevista a un participante.
  - **Sesgo del instrumento:** Puede deberse al funcionamiento defectuoso de un instrumento mecánico. Una balanza que no ha sido calibrada recientemente puede haberse desviado a la baja, produciendo continuamente lecturas de pesos corporales bajas.
  - **Sesgo del participante:** Es la deformación de la medición por parte del participante en el estudio, por ejemplo, al notificar un suceso (sesgo de respuesta o de recuerdo). Las pacientes con cáncer de mama que creen que el alcohol es una causa de su cáncer, por ejemplo, pueden exagerar el consumo que refieren

PNCQ
Programa Nacional de Controle de Qualidade

Conceptos estadísticos (Exactitud)

- **La exactitud** de una medición se evalúa mejor comparándola, cuando es posible, con un **«criterio de referencia»:**
- La decisión sobre qué abordaje de medición a usar debe considerar como criterio de referencia trabajos previos en ese campo.
- **Para las mediciones en una escala dicotómica, la exactitud en comparación con el criterio de referencia se puede describir con la sensibilidad y la especificidad.**
- Para las mediciones en escalas categóricas con más de dos opciones de respuesta, se puede calcular el porcentaje de valores correctos con cada método

Precisión y Exactitud de las mediciones

Resumen

	PRECISIÓN	EXACTITUD
DEFINICIÓN	Grado en una variable tiene casi el mismo valor cuando se mide varias veces	Grado en que una variable se aproxima del valor verdadero
MEJOR MODO DE EVALUACIÓN	Comparación entre repetidas medidas	Comparación con un patrón de referencia
UTILIDAD PARA EL ESTUDIO	Aumenta la potencia para detectar efectos	Aumenta la validez de las conclusiones
AMENAZADO POR	Error aleatorio (azar) al que contribuyen: Observador Participante Instrumento	Error sistemático (sesgo) al que contribuyen: Observador Participante Instrumento

PNCQ
Programa Nacional de Controle de Qualidade

Conceptos estadísticos (Validez)

- **VALIDEZ:** La validez es similar a la exactitud, aunque nos gusta pensar que añade una dimensión cualitativa a la consideración del grado en que una medición representa el fenómeno de interés.

Ejemplo: las mediciones de la creatinina y la cistatina C en la sangre, dos productos químicos excretados por los riñones, podrían tener la misma exactitud (<1% de la concentración verdadera), pero la cistatina C puede ser más válida como medición del funcionamiento renal, porque la concentración de creatinina también depende de la cantidad de músculo.

La validez a menudo no se puede evaluar con un patrón de referencia, particularmente en el caso de mediciones dirigidas a fenómenos subjetivos y abstractos como el dolor y la calidad de vida.

Los científicos sociales han creado constructos cualitativos y cuantitativos para abordar la validez de estos abordajes de medición.

PNCQ
Programa Nacional de Controle de Qualidade

Conceptos estadísticos (Validez)

- **Validez del contenido:** Examina hasta qué punto la evaluación representa todos los aspectos de los fenómenos en estudio; por ejemplo, incluyendo preguntas sobre la capacidad funcional social, física, emocional e intelectual para evaluar la calidad de vida.
- **Validez aparente:** Parece inherentemente razonable, como la medición del dolor en una escala de 10 puntos o la clase social por los ingresos del hogar.
- **Validez de constructo:** Es el grado en el que un dispositivo de medición específico concuerda con un constructo teórico; por ejemplo, una prueba de cociente intelectual debe distinguir entre personas que, según la teoría u otras medidas, tienen diferentes niveles de inteligencia.
- **Validez predictiva:** Es la capacidad que tiene la medición de predecir un resultado; por ejemplo, en qué medida un cuestionario diseñado para evaluar la depresión predice la pérdida de trabajo o el suicidio.
- **Validez relacionada con el criterio:** Es el grado en que una nueva medida se relaciona con medidas existentes y aceptadas.

PNCQ
Programa Nacional de Controle de Qualidade

ESTUDIOS DE REPRODUCIBILIDAD DE LAS PRUEBAS

- Los resultados de las pruebas pueden variar:
  - cuándo o dónde se realizaron, o quién las realizó.
- La variabilidad intraobservador describe la falta de reproducibilidad de los resultados:
  - cuando el mismo observador o laboratorio realiza la prueba en la misma muestra en momentos diferentes.
- Ejemplo: si un radiólogo observa la misma radiografía en dos ocasiones, ¿qué porcentaje de veces estará de acuerdo consigo mismo en la interpretación, asumiendo que no recuerda su interpretación previa?
  - La variabilidad interobservador describe la falta de reproducibilidad entre dos o más observadores: si otro radiólogo observa la misma radiografía, ¿qué probabilidad hay de que esté de acuerdo con el primer radiólogo?

PNCQ
Programa Nacional de Controle de Qualidade

ESTUDIOS DE REPRODUCIBILIDAD DE LAS PRUEBAS

- Muchas veces, el nivel de reproducibilidad (o su ausencia) es la principal pregunta de la investigación.
- En otros casos, la reproducibilidad se estudia con el fin de mejorar la calidad, ya sea como parte de la asistencia clínica o como parte de un estudio de investigación.
- Cuando la reproducibilidad es baja (porque la variabilidad intraobservador o interobservador es grande), no es probable que una determinación sea útil, y puede que necesite mejorarse o abandonarse

PNCQ
Programa Nacional de Controle de Qualidade

ESTUDIOS DE REPRODUCIBILIDAD DE LAS PRUEBAS

- Los estudios de reproducibilidad en sí mismos abordan la precisión, no la exactitud ni la validez.
  - Todos los observadores pueden estar de acuerdo entre sí y estar equivocados.
- Cuando se dispone de un método de referencia, los investigadores de la reproducibilidad intraobservador e interobservador pueden comparar las observaciones realizadas en los sujetos con un método de referencia para determinar la exactitud.

PNCQ
Programa Nacional de Controle de Qualidade

Diseños

- El diseño básico para evaluar la reproducibilidad de las pruebas supone comparar los resultados de pruebas de más de un observador o que se han realizado en más de una ocasión.
  - En pruebas que conllevan varias etapas, y en las que diferencias en cualquiera de esas fases podrían afectar a la reproducibilidad, el investigador deberá decidir la amplitud del objetivo del estudio.
  - **Ejemplo:** medir el acuerdo interobservador de anatomopatólogos en un conjunto de portaobjetos de frotis de citologías cervicovaginales en un solo hospital puede sobrestimar la reproducibilidad general de los frotis de las citologías cervicovaginales, porque no se detectaría la variabilidad en el modo de obtención de la muestra y en la preparación del frotis.

PNCQ
Programa Nacional de Controle de Qualidade

Diseños

- Como un investigador debe aislar los pasos que podrían producir desacuerdos entre observadores depende, en parte, de los objetivos de su estudio.
- La mayor parte de los estudios deben calcular la reproducibilidad de todo el proceso de la prueba, porque es lo que determina si merece la pena realizarla.
- Por otro lado, un investigador que esté desarrollando o mejorando una prueba puede desear centrarse en los pasos específicos que son problemáticos para mejorar el proceso.
  - En cualquier caso, el investigador debe mencionar el proceso exacto para obtener el resultado de la prueba en el manual de instrucciones y describirlo después en la sección de metodología cuando comunique los resultados del estudio.

Análisis – Escalas de medición

ESCALAS DE MEDICIÓN				
Tipo	**Variable características**	**Ejemplo**	**Estadístico descriptivo**	**Potencia estadística**
CATEGÓRICA				
Dicotómica	Dos categorías	Estado vital	Recuentos, proporciones	Baja
Nominal	Categorías no ordenadas	Raza, tipo sanguíneo	Igual	Baja
Ordinal	Categorías ordenadas con intervalos que no se pueden cuantificar	Grado de dolor, clase social	Igual + medianas	Intermedia
NUMÉRICA				
Continua o discreta	Espectro ordenado con intervalos cuantificables	Peso, número de cigarrillos/ día	Anterior + medias, desviaciones típicas	Elevada

PNCQ
Programa Nacional de Controle de Qualidade

Análisis
Variables categóricas

- La medida más sencilla del acuerdo interobservador es el porcentaje de observaciones en las que los observadores están exactamente de acuerdo. Sin embargo, cuando las observaciones no están distribuidas homogéneamente en las distintas categorías (p. ej., cuando la proporción de resultados «anormales» en una prueba dicotómica no es próxima al 50%), puede ser difícil interpretar el acuerdo porcentual, porque no tiene en consideración el acuerdo que podría deberse simplemente a que dos observadores tienen algunos conocimientos sobre la prevalencia de la alteración.
  - El acuerdo porcentual también es una medida subóptima cuando una prueba tiene más de dos resultados posibles que están ordenados intrínsecamente (p. ej., normal, limítrofe, anormal), porque el desacuerdo parcial (p. ej., normal/limítrofe) cuenta igual que el desacuerdo completo (normal/anormal)

PNCQ
Programa Nacional de Controle de Qualidade

Análisis

Variables categóricas

- Una medida mejor del acuerdo interobservador, denominada kappa (k) determina la magnitud del acuerdo más allá del que cabría esperar por el conocimiento por los observadores de la prevalencia de la alteración , y puede tener en consideración el acuerdo parcial.
- El valor de k oscila entre –1 (desacuerdo perfecto) y 1 (acuerdo perfecto). Un valor de 0 indica que no hay más acuerdo de lo que cabría esperar por las estimaciones de los observadores de la prevalencia de cada nivel de alteración. En general, se considera que valores de k superiores a 0,8 son muy buenos; se considera que los valores entre 0,6 y 0,8 son buenos.

PNCQ Programa Nacional de Controle de Qualidade

Escala de concordancia Índice kappa

Kappa	Concordancia	% de datos confiables
<0,00	Ninguna	
0,00 – 0,20	Débil	0 – 4%
0,21 – 0,40	Sufrible	4 – 15%
O,41 – 0,60	Regular	15 – 35%
0,61 – 0,80	Buena	35 – 63%
0,81 – 0,90	Óptima	64 – 81%
0,91 – 1,00	Perfecta	82 – 100%

Sáez-Alquezar A

PNCQ
Programa Nacional de Controle de Qualidade

Índice Kappa

El índice Kappa lleva em consideración las proporciones de las concordancias esperadas y de las concordancias observadas.

$$K = \frac{Po - Pe}{1 - Pe}$$

Po = Proporción de la concordancia observada

$$Po = a + d / N$$

Pe = Proporción de la concordancia esperada

$$Pe = \frac{[(a+b)(a+c)] + (c+d)(b+d)}{N^2}$$

Sáez-Alquezar A

Concordancia

	Analista 1		Total
	(+)	(-)	
Analista 2			
(+)	18 (a)	12 (b)	30
(-)	2 (c)	88(d)	90
Total	20	100	120

Resultados concordantes: (18 +88) = 106

Resultados discordantes: (2 +12) = 14

Concordancia: 106 / 120 = 88,3%

K = 65%

PNCQ
Programa Nacional de Controle de Qualidade

Análisis
Variables continuas

- Las medidas de la variabilidad interobservador para variables continuas dependen del diseño del estudio.
- Ejemplo: Estudios que miden el acuerdo entre dos máquinas o métodos (p. ej., temperaturas obtenidas con dos termómetros diferentes).
- La mejor forma de describir los datos de un estudio de este tipo es obtener los datos de los pares de mediciones (cada par está formado por dos mediciones realizadas casi al mismo tiempo en el mismo sujeto) y describir la diferencia media entre sus pares con alguna medida de la dispersión de las variables, como la desviación típica o con qué frecuencia la diferencia supera un umbral clínicamente importante.
- **Por ejemplo**, si se cree que una diferencia clínicamente importante de la temperatura corporal es de 0,3 °C, un estudio que compare temperaturas de termómetros timpánicos y rectales podría calcular la diferencia media (± desviación típica) entre las dos técnicas, y describir con qué frecuencia las dos determinaciones difieren en más de 0,3 °C3

Análisis
Variables continuas

- En otros estudios se examina la variabilidad entre métodos de análisis, observadores o instrumentos de estudio en un gran grupo de técnicos, laboratorios o máquinas diferentes. Estos resultados suelen resumirse usando el coeficiente de variación (CV), que es la desviación típica de todos los resultados obtenidos de una sola muestra dividida por el valor medio.
- Muchas veces se comparan los CV de dos o más métodos o instrumentos diferentes; el que tiene el menor CV es el más preciso (aunque puede no ser el más exacto)

Concordancia entre Pruebas

Tabla de contingencia cuando la prueba de comparación no es referencia			
	Prueba de comparación		
Prueba evaluada	(+)	(-)	total
(+)	289	12	301
(-)	11	220	231
total	300	232	532

Concordancia (%) = 100% [(a + d) / n]

Concordancia (C) (%) = 100% [(289 + 220) / 532] = 95,7%

Concordancia para positividad (%) = 100% [a / (a + c)]

C p/ positividad (%) = 100% [289 / (289 + 11)] = 96,3%

Concordancia para negatividad (%) = 100% [d / (b + d)]

C p/ negatividad (%) = 100% [220 / (12 + 220)] = 94,8%

PNCQ
Programa Nacional de Controle de Qualidade

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

- ¿En qué medida la prueba proporciona la respuesta correcta?
- Por supuesto, se asume que hay un **método de referencia** para mostrar cuál es la respuesta correcta.

PNCQ
Programa Nacional de Controle de Qualidade

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

Diseños - Muestreo

- En el diseño de un estudio diagnóstico de **casos y testigos**, se muestrea por separado a los que tienen y los que no tienen la enfermedad, y se comparan los resultados de la prueba en los dos grupos.
- Los estudios de pruebas en los que se obtienen muestras por separado de los que tienen y los que no tienen la enfermedad están sujetos a sesgo en la medición o la notificación del resultado de la prueba, porque su medición necesariamente se produce después de la medición del estado de la enfermedad.
- Además, los estudios con este esquema de muestreo habitualmente no pueden usarse para calcular los **valores predictivos**

PNCQ Programa Nacional de Controle de Qualidade

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

Diseños - Muestreo

- Una muestra **consecutiva** de pacientes a los que se evalúa para un determinado diagnóstico proporcionará, generalmente, resultados más válidos e interpretables, incluyendo los valores predictivos.
- **Ejemplo:** Tokuda y cols. observaron que la intensidad de los escalofríos (como sentir frío o sacudidas en todo el cuerpo bajo una manta gruesa) era un buen factor predictivo de bacteriemia en una serie de 526 pacientes adultos y febriles consecutivos en el servicio de urgencias.
- Como se incluyó a los participantes antes de conocerse si tenían o no bacteriemia, el espectro de pacientes de este estudio debe ser razonablemente representativo de los pacientes que acuden a los servicios de urgencias con fiebre.

PNCQ Programa Nacional de Controle de Qualidade

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

Diseños - Muestreo

- Un método de muestreo que denominamos **prueba en tándem** se usa a veces para comparar entre sí dos pruebas (supuestamente imperfectas). Ambas pruebas se realizan en una muestra representativa de pacientes, y el método de referencia se aplica selectivamente a los pacientes con resultados positivos en una o en ambas pruebas.
- El método de referencia también debe aplicarse a una muestra aleatoria de pacientes con resultados negativos concordantes para asegurarse de que realmente no tienen la enfermedad. Este diseño, que permite al investigador determinar qué prueba es más exacta sin el gasto de usar el método de referencia en todos los pacientes con resultados negativos en la prueba, se ha utilizado en estudios que comparan diferentes métodos de citología cervical

PNCQ Programa Nacional de Controle de Qualidade

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

Diseños - Muestreo

- Los estudios de pruebas pronósticas necesitan **diseños de cohortes**.
- Diseño prospectivo: la prueba se realiza al principio y se sigue a los pacientes para observar quién presenta la respuesta de interés.
- Diseño de cohortes retrospectivo: se puede utilizar cuando se dispone de una nueva prueba, como la viremia en pacientes infectados por el VIH, si se dispone de una cohorte definida previamente con muestras de sangre almacenadas
- La viremia puede determinarse en la sangre almacenada para observar si predice el pronóstico.

PNCQ Programa Nacional de Controle de Qualidade

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

Diseños - Variable predictiva: resultado de la prueba.

- Aunque es más fácil pensar en los resultados de una prueba diagnóstica como positivos o negativos, muchas pruebas tienen resultados categóricos, ordinales o continuos.
- A fin de aprovechar toda la información disponible en la prueba, los investigadores generalmente deben describir los resultados de las pruebas ordinales o continuas, en lugar de dicotomizarlos como «normal o anormal».
- La mayor parte de las pruebas son más indicativas de enfermedad si son muy anormales que si son levemente anormales, y tienen un intervalo limítrofe en los que no proporcionan mucha información.

PNCQ Programa Nacional de Controle de Qualidade

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

Diseños - Variable de respuesta: enfermedad (o su respuesta)

- La variable de respuesta en el **estudio de una prueba diagnóstica** es la presencia o la ausencia de la enfermedad, que en situación óptima se determina mediante un método de referencia.
- Si es posible, la evaluación de la respuesta no debe verse influida por los resultados de la prueba diagnóstica que se está estudiando. La forma óptima de realizarlo es ocultando la información a las personas que realizan la prueba de referencia, para que no conozcan los resultados de la prueba índice.

PNCQ Programa Nacional de Controle de Qualidade

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

Diseños - Variable de respuesta: enfermedad (o su respuesta)

- A veces, particularmente en el caso de pruebas de cribado, la aplicación uniforme del método de referencia no es ética ni viable.
- **Ejemplo:** Estudio de la exactitud de la mamografía según las características del radiólogo que las interpreta.
  - En mamografías positivas>>>>>pruebas adicionales, finalmente con evaluación anatomopatológica como método de referencia.
  - No es razonable realizar biopsias mamarias a mujeres cuyas mamografías son negativas.
  - Se han de encontrar otras soluciones, aunque puedan crear un sesgo.

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

Diseños - Variable de respuesta: enfermedad (o su respuesta)

- La variable de respuesta en un **estudio de una prueba pronóstica** supone lo que les ocurre a los pacientes con una enfermedad, como el tiempo que viven, qué complicaciones presentan o qué tratamientos adicionales precisan.
  - Es importante el enmascaramiento, especialmente si los médicos que atienden a los pacientes pueden tomar decisiones basadas en los factores pronósticos que se están estudiando.

Resultados de los estudios de la exactitud de Pruebas Diagnósticas.

Sensibilidad

Especificidad

Valores Predictivos

Curvas ROC

Cocientes de Verosimilitudes

Conceptos Estadísticos

Tabla de contingencia

Prueba comparativa que o de referencia

ENSAYO EVALUADO	Presente	Ausente	Total
Positivo	Verdadero Positivo a	Falso Positivo b	a + b
Negativo	Falso Negativo c	Verdadero Negativo d	c + d
Total	a + c	b + d	n

Sensibilidad = a / (a+c)

Especificidad = d / (b+d)

N = (a+b+c+d)

Valor predictivo positivo
VPP = a / (a+b)

Valor predictivo negativo
VPN = d / (c+d)

Sáez-Alquezar A

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

Análisis - Sensibilidad, especificidad y valores pronóstico positivos y negativos

La **sensibilidad (S)** de una prueba se define como la proporción de participantes con la enfermedad en los que la prueba da la respuesta correcta (positiva)

$$\text{Sensibilidad} = \frac{\text{Verdaderos Positivos}}{\text{Verdaderos positivos + Falsos Negativos}}$$

la **especificidad (E)** es la proporción de participantes sin la enfermedad en los que la prueba da la respuesta correcta (negativa).

$$\text{Especificidad} = \frac{\text{Verdaderos Negativos}}{\text{Falsos Positivos + Verdaderos Negativos}}$$

Tabla de contingencia – valor referencial conocido

PRUEBA COMPARATIVA (REFERENCIA)
(MUESTRA ESTÁNDAR)

ENSAYO EN EVALUACIÓN	Presente	Ausente
Positivo	Verdadero Positivo VP 51	Falso Positivo FP 9
Negativo	Falso Negativo FN 0	Verdadero Negativo VN 2.364

a b c d

Sensibilidad = a / (a+c)

51 / 51 = 100%

Especificidad = d / (b+d)

2.364 / 2.364 + 9
2.364 / 2.373 = 99,7%

Sáez-Alquezar A

Tabla de contingencia – valor referencial conocido

PRUEBA COMPARATIVA (REFERENCIA)
(MUESTRA ESTÁNDAR)

ENSAYO EN EVALUACIÓN	Presente	Ausente
Positivo	Verdadero Positivo VP 101	Falso Positivo FP 03
Negativo	Falso Negativo FN 0	Verdadero Negativo VN 1.176

a b c d

Sensibilidad = a / (a+c)

101 / 101+0 = 100%

Especificidad = d / (b+d)

1.176 / 1.176 + 03

1.176 / 1.179 = 99,8%

Sáez-Alquezar A

Tabla de contingencia – valor referencial conocido

PRUEBA COMPARATIVA (REFERENCIA)
(MUESTRA ESTÁNDAR)

ENSAYO EN EVALUACIÓN	Presente	Ausente
Positivo	Verdadero Positivo VP 105	Falso Positivo FP 10
Negativo	Falso Negativo FN 02	Verdadero Negativo VN 1.210

a b c d

Sensibilidad = a / (a+c)

105 / 105 + 02 =
105 / 107 = 98,1%

Especificidad = d / (b+d)

1.210 / 10 + 1.200 =
1.210 / 1.220 = 99,2%

Sáez-Alquezar A

PNCQ
Programa Nacional de Controle de Qualidade

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

Análisis - Sensibilidad, especificidad y valores pronóstico positivos y negativos

El **valor predictivo positivo (VPP)** es la proporción de personas con un resultado positivo en la prueba que tienen la enfermedad o,
es la proporción de individuos infectados (+) entre aquellos con resultados Positivos

El **VPP** aumenta con la prevalencia

El **valor predictivo negativo (VPN)** es la proporción de personas con un resultado negativo en la prueba que no
tienen la enfermedad, o
es la proporción de individuos sanos (-) entre aquellos con resultados Negativos

El **VPN** disminuye con la prevalencia

PNCQ
Programa Nacional de Controle de Qualidade

Estudios e poblaciones con baja prevalencia

- Candidatos a donación de sangre
  - Utilizar pruebas con alta sensibilidad
    - Evitar Resultados Falsos Negativos
  - Utilizar pruebas con la mejor especificidad
    - Evitar Resultados Falsos Positivos

El **VPP** aumenta con la prevalencia

El **VPN** disminuye con la prevalencia

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

Análisis - Curvas de eficacia diagnóstica

- Muchas pruebas diagnósticas proporcionan resultados ordinales o continuos.
  - Puede haber varios valores de sensibilidad y especificidad, dependiendo del valor de corte elegido para definir una prueba positiva.
- Esta concesión mutua entre sensibilidad y especificidad puede representarse usando una técnica gráfica desarrollada originalmente en electrónica: **curvas de eficacia diagnóstica o curvas ROC *(receiver operating characteristic)***.

PNCQ Programa Nacional de Controle de Qualidade

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

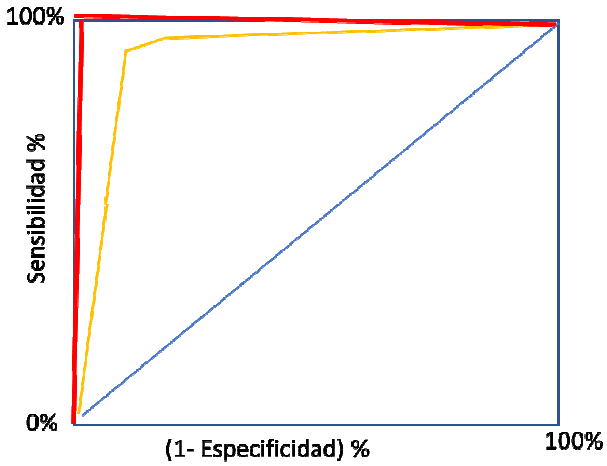
Análisis - Curvas de eficacia diagnóstica

- Sensibilidad (S) X Especificidad (E)
- Las curvas ROC son una forma de representar la relación entre la S y la E de una prueba cuantitativa por medio de valores continuos de "cutoff" (valor de corte o punto de corte)
- Para construir una curva ROC se monta un diagrama que represente la sensibilidad en función de la proporción de resultados Falsos positivos (1-Especificidad) para un conjunto de valores de "cutoff"
- Cuando se tiene una variable continua, resultado de la aplicación de una prueba diagnóstica cuantitativa, y se quiere transformar en una variable dicotómica, del tipo enfermo/no enfermo, hay que utilizar un determinado valor en la escala continua que discrimina entre estas dos clases. Este valor se denomina "punto de corte" o "cutoff".

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

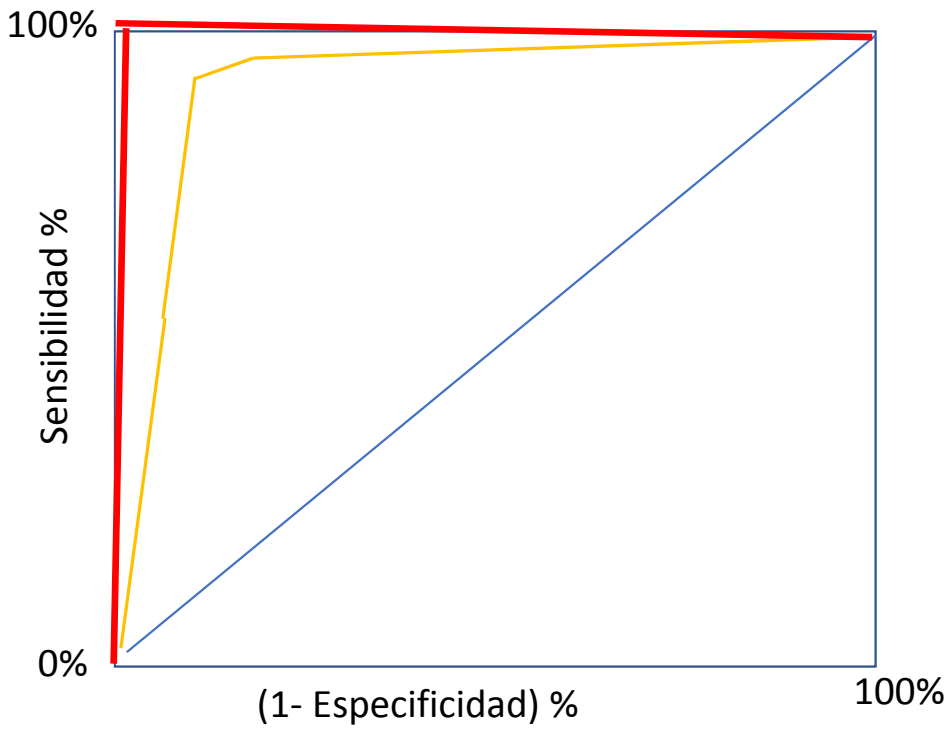
Análisis - Curvas de eficacia diagnóstica

- El investigador selecciona varios valores de corte y determina la sensibilidad y la especificidad en cada punto. Representa después la sensibilidad (tasa de positivos verdaderos) en el eje de ordenadas (Y) en función de 1 – especificidad (tasa de positivos falsos) en el eje de abscisas (X).
- Una prueba ideal es una que alcance la esquina superior izquierda de la gráfica (un 100 % de positivos verdaderos y ningún positivo falso).
- Una prueba sin valor sigue la diagonal desde la esquina inferior izquierda hasta la superior derecha: en cualquier valor de corte, la tasa de positivos verdaderos es la misma que la de positivos falsos.
- El área bajo la curva ROC, que, por tanto, oscila desde 0,5 para una prueba fútil hasta 1 para una prueba perfecta, es un resumen útil de la exactitud general de una prueba, y puede utilizarse para comparar la exactitud de dos o más pruebas.



Sáez-Alquezar A

Curvas ROC



VP (sensibilidad) y FP (1-especificidad)

Sáez-Alquezar A

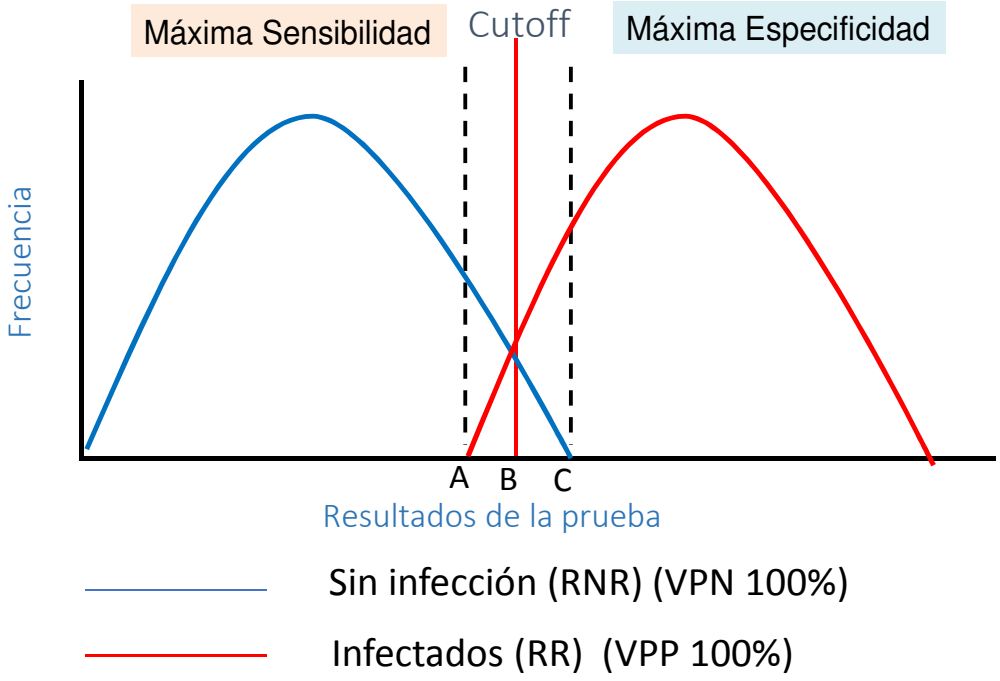
- La exactitud de una prueba diagnóstica será proporcional al área por debajo de la curva ROC

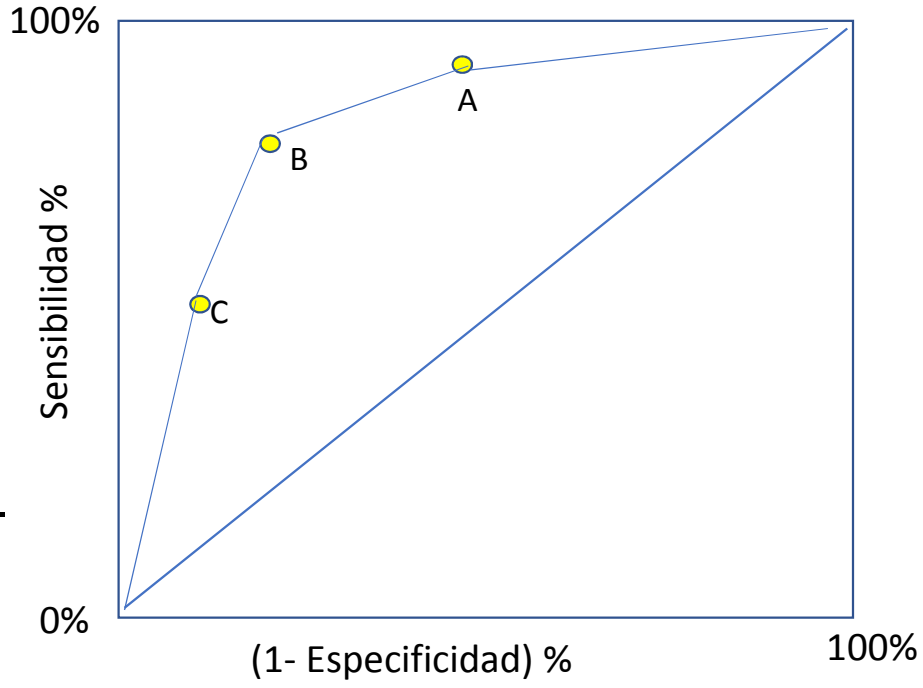
- En la comparación de pruebas diagnósticas, aquella que presente área mayor, debajo de la curva, será la que tiene mejor exactitud.

- Optimización de la sensibilidad en función de la especificidad: el punto (Cutt off) más próximo del canto superior izquierdo del gráfico

- Discriminación perfecta (rojo) -100%
- Discriminación nula (azul) – 50%

Valores de corte para las pruebas serológicas cualitativas



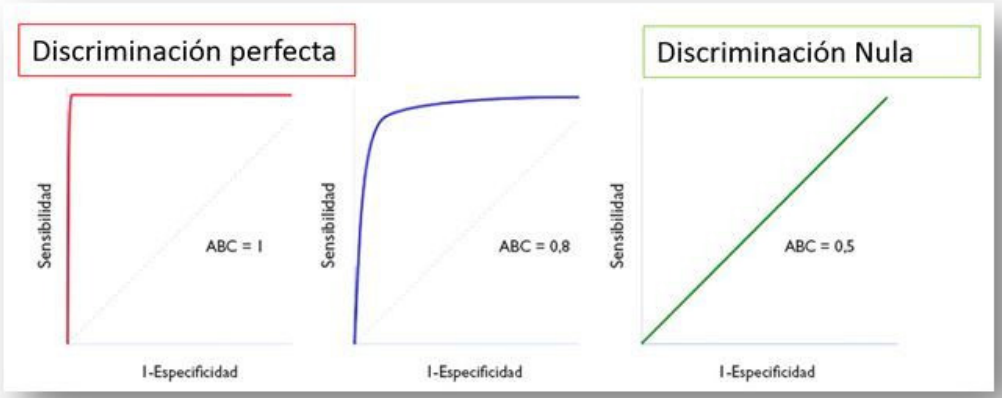


Curva ROC para un conjunto de valores de cut off

Sáez-Alquezar A

Curvas ROC

- Ejemplos de Curvas ROC en que se visualiza la validez de una prueba diagnóstica, oscilando entre dos puntos limites:
  - La discriminación perfecta: en que la probabilidad de identificar pacientes sanos y enfermos es del 100%
  - La discriminación nula: en que la prueba no es capaz de identificar más que al 50% de sanos o enfermos.
  - Se puede analizar el área bajo la curva, de manera que la prueba con mayor área será la mejor.



Sáez-Alquezar A

PNCQ
Programa Nacional de Controle de Qualidade

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

Análisis - Cocientes de verosimilitudes

- Aunque la información en una prueba diagnóstica con resultados continuos u ordinales puede resumirse usando curvas ROC de sensibilidad y especificidad, existe una forma mejor.
- Los cocientes de verosimilitudes permiten que el investigador se beneficie de toda la información de una prueba.
- Para cada resultado de una prueba, el cociente de verosimilitudes es la proporción de la posibilidad de ese resultado en alguien con la enfermedad con respecto a la posibilidad de ese resultado en alguien sin la enfermedad.

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

Análisis - Cocientes de verosimilitudes

$$\text{Cociente de verosimilitudes} = \frac{P\,(\text{resultado} \mid \text{enfermedad})}{P\,(\text{resultado} \mid \text{ausência de enfermedad})}$$

P se lee «probabilidad de» y el símbolo «|» se lee «dado».
Así, *P* (resultado|enfermedad) es la probabilidad del resultado dada la enfermedad, y *P* (resultado|ausencia de enfermedad) es la probabilidad de ese resultado dada la ausencia de enfermedad.
El cociente de verosimilitudes es una razón de estas dos probabilidades[4]

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

Análisis - Cocientes de verosimilitudes

- Cuanto **mayor** sea el cociente de verosimilitudes, mejor será la prueba para *confirmar* la enfermedad;
  - Un cociente de verosimilitudes superior a 100 es muy alto (y poco habitual en las pruebas).
- Cuanto **menor** sea el cociente de verosimilitudes (cuanto más cercano sea a 0), mejor será la prueba para *descartar* la enfermedad.
- Un cociente de verosimilitudes de 1 (en el intervalo de 0,8 a 1,25) ofrece poca información útil.

PNCQ Programa Nacional de Controle de Qualidade

EJEMPLO DE CÁLCULO DE LOS COCIENTES DE VEROSIMILITUDES DE UN ESTUDIO DEL HEMOGRAMA COMPLETO PARA PREDECIR LAS INFECCIONES GRAVES EN RECIÉN NACIDOS PREMATUROS

- **Ejemplo:** cómo calcular el cociente de verosimilitudes, de acuerdo con los resultados de un estudio de hemograma completo en recién nacidos con riesgo de infecciones graves.
- Un recuento leucocitario inferior a 5 000 leucocitos/ml fue más frecuente en lactantes con infecciones graves que en otros lactantes.
- El cálculo del cociente de verosimilitudes cuantifica esto: el 19 % de los lactantes con infecciones tenía recuentos leucocitarios menores de 5 000 leucocitos/ml, en comparación con solo el 0,52 % de los que no tenían infecciones.
- Entonces, el cociente de verosimilitudes es 19 %/0,52 % = 36.

RECUENTO LEUCOCITARIO	INFECCIÓN GRAVE		COCIENTE DE VEROSIMILITUDES
	SÍ	NO	
< 5 000	46 19 %	347 0,52 %	36
5 000-9 999	53 22 %	5 103 7,6 %	2,9
10 000-14 999	53 22 %	16 941 25 %	0,86
15 000-19 999	45 18 %	21 168 31 %	0,58
≥ 20 000	48 20 %	23 818 35 %	0,56
Total	245 100 %	67 377 100 %	

Sáez-Alquezar A

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

Análisis - Riesgo absoluto, cocientes de riesgo, diferencias de riesgo y cocientes de riesgos instantáneos.

- El análisis de estudios de **pruebas pronósticas** es similar al de otros estudios de cohortes.
- Si se realiza un seguimiento durante un período de tiempo establecido (p. ej., 3 años) a todos los que participan en un estudio de una prueba pronóstica, con pocas pérdidas durante el seguimiento, los resultados pueden resumirse con riesgos absolutos, cocientes de riesgo y diferencias de riesgo.
- Especialmente, cuando el seguimiento es completo y dura poco, los resultados de los estudios de las pruebas pronósticas se resumen a veces como los de las pruebas diagnósticas, usando la sensibilidad, la especificidad, el valor predictivo, los cocientes de verosimilitudes y las curvas ROC.
- Por otro lado, cuando se sigue a los participantes del estudio durante períodos de tiempo variables, es preferible realizar una técnica de análisis de supervivencia que tenga en consideración la duración del período de seguimiento y que estime los cocientes de riesgos instantáneos

Estudios del valor de Pruebas Pronósticas.

Cocientes de Riesgo

Cocientes de Riesgo Instantáneos

Cocientes de riesgo y cocientes de riesgo instantáneo son conceptos utilizados en epidemiología y estadística para medir la relación entre la exposición a un factor de riesgo y la aparición de una enfermedad o evento adverso.

Mejoría por la Reclasificación

Métrica utilizada para comparar la capacidad predictiva de diferentes modelos.

PNCQ
Programa Nacional de Controle de Qualidade

Cociente de Riesgo

Medida de la asociación entre la exposición a un factor de riesgo y la ocurrencia de un resultado específico, como una enfermedad.
Se calcula dividiendo la probabilidad de que ocurra el resultado en el grupo expuesto por la probabilidad de que ocurra en el grupo no expuesto.

$$CR = \frac{\text{Probabilidad de enfermedad en el grupo expuesto}}{\text{Probabilidad de enfermedad en el grupo no expuesto}}$$

CR = 1 sugiere que no hay asociación entre la exposición y el resultado;
CR >1 sugiere un mayor riesgo asociado con la exposición;
CR <1 sugiere un menor riesgo asociado con la exposición.

Cociente de Riesgo Instantáneo

- Medida de la tasa instantánea de riesgo en un punto específico en el tiempo.
- Se utiliza comúnmente en el contexto de análisis de supervivencia y modelos de riesgos proporcionales. En esencia, es una medida de cómo cambia el riesgo en función del tiempo.
- En análisis de supervivencia, por ejemplo, el cociente de riesgo instantáneo describe la razón instantánea de cambio en el riesgo de un evento en un momento dado, condicionado al hecho de que el evento no ha ocurrido antes de ese momento.

Mejora de las Pruebas médicas por reclasificación

Evaluación de modelos de predicción

Se refiere a una métrica utilizada para comparar la capacidad predictiva de diferentes modelos.
Cuando se desarrollan modelos predictivos, especialmente en el campo de la medicina y la salud, es crucial evaluar su capacidad para clasificar correctamente a los individuos en diferentes categorías de riesgo o de diagnóstico. La mejora por reclasificación ayuda a determinar si un nuevo modelo predictivo mejora la clasificación de los individuos en comparación con un modelo de referencia.
El proceso de reclasificación implica volver a clasificar a los individuos en categorías de riesgo utilizando un nuevo modelo y comparar esta clasificación con la de un modelo de referencia. Se pueden calcular varias medidas de mejora por reclasificación, pero una de las más comunes es la categorización correcta neta (Net Reclassification Improvement, NRI).
La NRI compara la proporción de individuos correctamente reclasificados hacia arriba en riesgo con la proporción de individuos incorrectamente reclasificados hacia abajo en riesgo entre el nuevo modelo y el modelo de referencia.
En resumen, la mejora por reclasificación es una herramienta útil para evaluar la utilidad clínica y predictiva de un nuevo modelo en comparación con un modelo existente o de referencia, especialmente en el campo de la medicina y la salud.

Sáez-Alquezar A

PNCQ Programa Nacional de Controle de Qualidade

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

Análisis - Mejoría neta por la reclasificación (MNR)

- Para nuevas pruebas o biomarcadores cuya finalidad es la predicción de episodios futuros de la enfermedad, es importante cuantificar lo que añaden las nuevas pruebas a los modelos de predicción existentes
- Ejemplo, Shepherd y cols, encontraron que la adición del volumen fibroglandular calculado en la mamografía (es decir, la cantidad estimada de tejido mamario con riesgo de neoplasia maligna) a un modelo que incluía los factores de riesgo clínicos tradicionales mejoraba la predicción de la aparición posterior de un carcinoma de mama o un carcinoma ductal *in situ* con un MNR del 21 % (p = 0,0 001).

PNCQ Programa Nacional de Controle de Qualidade

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

Estudios para elaborar reglas de predicción clínica

- Los modelos matemáticos para elaborar reglas de predicción generalmente suponen el uso de una técnica multivariada para seleccionar variables predictivas candidatas y combinar sus valores para generar una predicción.
- Entre las variables candidatas deben estar todas las **variables predictivas** conocidas y creíbles que se puedan medir de manera fácil, fiable y económica.
- Un modelo multivariado, como el análisis de **regresión logística** o el **modelo de Cox (de los riesgos proporcionales)**, puede cuantificar la contribución independiente de las variables predictivas candidatas a la predicción de la variable de respuesta.
- Las que tengan una asociación más fuerte y constante con la variable de respuesta se pueden incluir en la regla, y se pueden asignar puntos a diferentes valores de las variables predictivas dependiendo de los coeficientes del modelo.
- Por ejemplo, Wells y cols. Utilizaron análisis de regresión logística en 40 posibles variables predictivas clínicas de embolia pulmonar para elaborar una puntuación predictiva basada en tan solo siete variables

PNCQ Programa Nacional de Controle de Qualidade

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

Estudios para elaborar reglas de predicción clínica

TABLA 12-4. EJEMPLO DE UNA REGLA DE PREDICCIÓN CLÍNICA (PARA LA EMBOLIA PULMONAR) DERIVADA DE UN ANÁLISIS DE REGRESIÓN LOGÍSTICA (14)

CARACTERÍSTICA CLÍNICA	PUNTOS
Embolia pulmonar o trombosis venosa profunda previa	+ 1,5
Frecuencia cardíaca > 100 latidos por minuto	+ 1,5
Cirugía o inmovilización reciente (en los 30 días previos)	+ 1,5
Signos clínicos de trombosis venosa profunda	+ 3
Un diagnóstico alternativo es menos probable que la embolia pulmonar	+ 3
Hemoptisis (expectoración de sangre)	+ 1
Cáncer (tratado en los 6 meses previos)	+ 1
PROBABILIDAD CLÍNICA ESTIMADA DE EMBOLIA PULMONAR (15)	**PUNTUACIÓN TOTAL**
Baja (probabilidad ~1 %-2 %)	0-1
Intermedia (probabilidad ~16 %)	2-6
Elevada (probabilidad ~40 %)	≥ 7

Esta puntuación, actualmente popular, se utiliza para asignar una probabilidad previa de embolia pulmonar a fin de guiar las decisiones diagnósticas posteriores y la interpretación de sus resultados

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

ESTUDIOS DEL EFECTO DE LOS RESULTADOS DE LAS PRUEBAS EN LAS DECISIONES CLÍNICAS

Una prueba puede ser exacta, pero si la enfermedad es muy rara, la prueba puede ser positiva con una frecuencia tan baja que apenas merece la pena realizarla.
Otras pruebas pueden no afectar a las decisiones clínicas, porque no aportan nueva información aparte de la que ya se conocía.
Ejemplo: La historia clínica y la exploración física.

Los diseños de estudios en esta sección abordan el **rendimiento** de las pruebas diagnósticas y sus efectos sobre las decisiones clínicas.

PNCQ Programa Nacional de Controle de Qualidade

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

ESTUDIOS DEL EFECTO DE LOS RESULTADOS DE LAS PRUEBAS EN LAS DECISIONES CLÍNICAS

Tipos de estúdios.

Estudios de rendimiento diagnóstico. Estos estudios permiten responder preguntas como:

- Cuando se solicita una prueba para una determinada indicación, ¿con qué frecuencia es anormal?
- ¿Puede un resultado anormal de una prueba predecirse a partir de otra información disponible en el momento de realizar la prueba?
- ¿En qué grupos de pacientes tiene la prueba la máxima y la mínima utilidad?
- ¿Qué ocurre con los pacientes con resultados anormales? ¿Los beneficios superan a los perjuicios?

Los estudios de rendimiento diagnóstico calculan la proporción de pruebas positivas entre pacientes con una indicación particular para la realización de la prueba. Lamentablemente, mostrar que una prueba es positiva a menudo no es suficiente para indicar que la prueba debe realizarse.

- Sin embargo, un estudio de rendimiento diagnóstico en el que se muestra que una prueba casi siempre es negativa puede ser suficiente para cuestionar su uso para esa indicación.

PNCQ Programa Nacional de Controle de Qualidade

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS ESTUDIOS DEL EFECTO DE LOS RESULTADOS DE LAS PRUEBAS EN LAS DECISIONES CLÍNICAS

Tipos de estúdios.

- **Por ejemplo**, Siegel y cols. (18) estudiaron el rendimiento de los coprocultivos en pacientes hospitalizados con diarrea.
- Aunque no se realizaron coprocultivos a todos los pacientes con diarrea, parece razonable suponer que los que la sufren, si es que hay alguno, tienen más probabilidad de tener un cultivo positivo que los que no.
- En total, solo 40 (2 %) de 1 964 coprocultivos fueron positivos. Además, ninguno de los resultados positivos correspondía a los 997 pacientes que habían estado en el hospital durante más de 3 días.
- Como es poco probable que un coprocultivo negativo afecte al tratamiento en estos pacientes con una probabilidad baja de diarrea bacteriana, los autores concluyeron que los coprocultivos tienen poca utilidad en pacientes con diarrea que han estado en el hospital más de 3 días.

PNCQ
Programa Nacional de Controle de Qualidade

ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS ESTUDIOS DEL EFECTO DE LOS RESULTADOS DE LAS PRUEBAS EN LAS DECISIONES CLÍNICAS

Tipos de estúdios.

- **Estudios de toma de decisiones clínicas de antes/después.** Estos diseños abordan directamente
- El efecto del resultado de una prueba sobre las decisiones clínicas.
- El diseño suele conllevar una comparación entre lo que los médicos hacen (o dicen que harían) antes y después de obtener los resultados de una prueba diagnóstica.
- **Por ejemplo**, Carrico y cols, estudiaron prospectivamente la utilidad de la ecografía abdominal en 94 niños con dolor abdominal inferior agudo.
- Pidieron a los médicos que habían solicitado las ecografías que registraran su impresión diagnóstica y cuál sería su tratamiento si no dispusieran de una ecografía.
- Tras realizar las ecografías y proporcionar a los médicos los resultados, les preguntaron de nuevo.
- Observaron que la información ecográfica cambió el plan inicial del tratamiento en el 46 % de los pacientes.
- Por supuesto la alteración de una decisión clínica no garantiza que un paciente se vaya a beneficiar, y algunas decisiones alteradas realmente podrían ser perjudiciales.

Estudios de viabilidad, costes y riesgos de las pruebas

- Otra área importante para la investigación clínica se relaciona con los aspectos prácticos de las pruebas diagnósticas.
- ¿Cuáles son los efectos médicos de las pruebas de cribado falsamente positivas en recién nacidos y los efectos psicológicos sobre los padres?
- ¿Qué proporción de colonoscopias se complican con perforación colónica?

Estudios de viabilidad, costes y riesgos de las pruebas

- **Aspectos del diseño**
- Los estudios de **viabilidad, costes** y **riesgos** de las pruebas suelen ser descriptivos.
- El esquema de muestreo es importante, porque las pruebas a menudo varían entre las personas o las instituciones que las realizan y en los pacientes a los que se les realizan.

PNCQ
Programa Nacional de Controle de Qualidade

Estudios del efecto de las pruebas en las respuestas

- **Análisis**
- Los resultados de estos estudios suelen poder resumirse con estadísticos descriptivos simples, como medias y desviaciones típicas, medianas, recorridos y distribuciones de frecuencia.
- Las variables dicotómicas, como la incidencia de efectos adversos, pueden resumirse con proporciones y sus intervalos de confianza (IC) del 95 %.
- **Ejemplo:** Rutter y cols, encontraron perforaciones en 21/43 456 colonoscopias, lo que corresponde al 0,48 por 1 000, con un intervalo de confianza del 95 % desde 0,30 hasta 0,74 por 1 000.
- Generalmente no hay líneas claras que dividan las pruebas en las que son y no son viables, o en las que tienen o no un riesgo inaceptable de efectos adversos. Por este motivo es útil, en la etapa de diseño del estudio, especificar los criterios para decidir que la prueba es aceptable.
  - ¿Qué tasa de seguimiento será insuficiente?
  - ¿Qué incidencia de complicaciones será demasiado alta?

PNCQ
Programa Nacional de Controle de Qualidade

Estudios del efecto de las pruebas en las respuestas

- El mejor modo de determinar el valor de una prueba médica es observar si los pacientes a los que se ha realizado presentan una mejor respuesta clínica (p. ej., viven más o con mejor calidad de vida) que aquellos a los que no se les ha realizado.
- Los ensayos aleatorizados son el diseño ideal para hacer esta determinación, pero los ensayos de pruebas diagnósticas son, a veces, difíciles de realizar.
- El valor de las pruebas suele, por tanto, calcularse a partir de estudios de observación.
- La diferencia clave entre los diseños descritos en esta sección y los diseños experimentales y de observación comentados en otras secciones de este libro es que la variable predictiva, en este caso, es *la realización de la prueba*, en lugar de un tratamiento, factor de riesgo o resultado de una prueba.

PNCQ
Programa Nacional de Controle de Qualidade

Errores en el diseño y el análisis de los estudios de pruebas diagnósticas

Tamaño de la muestra inadecuado

- Si la variable de respuesta de un estudio de una prueba diagnóstica es frecuente, es probable que se pueda obtener una muestra de tamaño adecuado.
- Cuando la enfermedad o la variable de respuesta son infrecuentes, puede ser necesario un número muy grande de personas.
- Muchas pruebas de laboratorio, por ejemplo, no son caras, y un rendimiento del 1 % o menos podría justificar su realización, especialmente si pueden diagnosticar una enfermedad grave pero tratable.

Ejemplo: Sheline y Kher revisaron retrospectivamente pruebas analíticas sistemáticas en el momento del ingreso, entre ellas una prueba de VDRL *(Venereal Disease Research Laboratory)* para detectar sífilis, en 252 pacientes psiquiátricos, y observaron que las pruebas analíticas identificaron a un paciente con sífilis no sospechada previamente. Si los síntomas de este paciente psiquiátrico se debieran realmente a la sífilis, sería difícil argumentar que no mereció la pena el gasto de 3 186 dólares en las pruebas de VDRL para llegar al diagnóstico. Pero si la prevalencia real de sífilis no sospechada fuera próxima al 0,4 % observado en este estudio, un estudio de este tamaño de muestra fácilmente podría no haber encontrado casos.

PNCQ
Programa Nacional de Controle de Qualidade

Errores en el diseño y el análisis de los estudios de pruebas diagnósticas

Exclusión incorrecta

Cuando se calculan proporciones, es inadecuado excluir participantes del numerador sin excluir participantes similares del denominador.

Ejemplo: En un estudio de análisis de laboratorio habituales en pacientes del servicio de urgencias con convulsiones de nueva aparición, 11 de 136 pacientes (8 %) tenían una alteración de laboratorio corregible (p. ej., hipoglucemia) como causa de la convulsión.

En 9 de 11 pacientes, sin embargo, se sospechó la alteración por la historia clínica o la exploración física.

Los autores, por lo tanto, comunicaron que solo 2 de 136 pacientes (1,5 %) presentaban alteraciones no sospechadas por la historia clínica o la exploración física.

Pero, si se excluye del numerador a todos los pacientes con alteraciones sospechadas, se debería haber excluido a los mismos pacientes también del denominador. El denominador correcto para esta proporción es, por tanto, no todos los 136 pacientes a los que se estudió, sino solo aquellos en los que no se sospechó la existencia de alteraciones analíticas por la historia clínica o la exploración física.

Errores en el diseño y el análisis de los estudios de pruebas diagnósticas

Resultados limítrofes o no interpretables

A veces una prueba no puede dar respuesta alguna, como sucede si el análisis falla, la muestra de la
prueba se estropea o el resultado de la prueba se encuadra en una zona que no se sabe si es positiva
o negativa.
No suele ser válido ignorar estos problemas, pero el modo de manejarlos dependerá de
la pregunta específica de la investigación y del diseño del estudio. En los estudios que tratan sobre el coste o la incomodidad de las pruebas, los intentos fallidos de realizar las pruebas son, claramente,
resultados importantes.
Se debe considerar que los pacientes con estudios de imagen «no diagnósticos» o con resultados limítrofes en una prueba han tenido ese resultado específico en la prueba.
En efecto, esto puede transformar una prueba dicotómica (positiva, negativa) en una ordinal: positiva, indeterminada y negativa.
Pueden trazarse las curvas ROC y calcularse los cocientes de verosimilitudes para los resultados «indeterminados», así como para los positivos y los negativos.

PNCQ Programa Nacional de Controle de Qualidade

Errores en el diseño y el análisis de los estudios de pruebas diagnósticas

Sesgo de verificación: aplicación selectiva de un solo patrón de referencia

Una estrategia de muestreo habitual para los estudios de pruebas médicas consiste en estudiar (prospectiva o retrospectivamente) pacientes a los que se estudia para detectar una enfermedad y en los que también se aplica el método de referencia para el diagnóstico.

Esto puede causar un problema si los hallazgos que se estudian también se usan para decidir a quién se le realiza el método de referencia.

Ejemplo: considere un estudio de factores predictivos de fracturas en niños que acuden al servicio de urgencias con lesiones en el tobillo, en el que solo se incluyó a niños a los que se hicieron radiografías de sus lesiones de tobillo. Si los que tenían un hallazgo concreto (p. ej., tumefacción del tobillo) tenían más probabilidad de que se les hiciera una radiografía, esto podría afectar a la sensibilidad y la especificidad de la tumefacción del tobillo como prueba para diagnosticar una fractura.

Este sesgo, denominado **sesgo de verificación,** se ilustra numéricamente a seguir.

PNCQ
Programa Nacional de Controle de Qualidade

Ejemplo numérico del sesgo de verificación

Considere dos estudios en los que se examina la tumefacción del tobillo como factor predictivo de fracturas en niños con lesiones de tobillo.

Ejemplo 01: El primer estudio es una **muestra consecutiva** de 200 niños.
En este estudio se hizo una radiografía a todos los niños con lesiones del tobillo.
La sensibilidad y la especificidad de la tumefacción del tobillo son del 80 % y del 75 %, como se muestra en la tabla a seguir.

Ejemplo 02: En el segundo estudio se usa una muestra **seleccionada**, en la que solo se hizo una radiografía a la mitad de los niños sin tumefacción del tobillo.
Por lo tanto, los números de la fila «Sin tumefacción» se reducirán a la mitad.
Esto eleva la sensibilidad aparente desde 32/40 (80 %) hasta 32/36 (89 %), y reduce la especificidad aparente desde 120/160 (75 %) hasta 60/100 (60 %), como se muestra seguir.

PNCQ Programa Nacional de Controle de Qualidade

TUMEFACCIÓN DEL TOBILLO COMO FACTOR PREDICTIVO DE FRACTURA USANDO UNA MUESTRA CONSECUTIVA.

Se hizo una radiografía a todos los niños con lesiones del tobillo.

	Fractura	Sin Fractura
Tumefacción	Verdadero Positivo 32	Falso Positivo 40
Sin tumefacción	Falso Negativo 8	Verdadero Negativo 120
	a	b
	c	d
TOTAL	a + c = 40	b + d = 160

Sensibilidad = a / (a+c)

Sensibilidad = 32 / 40 = 80%

Especificidad = d / (b+d)

Especificidad = 120 / 160 = 75%

Sáez-Alquezar A

SESGO DE VERIFICACIÓN: LA TUMEFACCIÓN DEL TOBILLO COMO FACTOR PREDICTIVO DE FRACTURA USANDO UNA MUESTRA SELECCIONADA.

Solo se hizo una radiografía a la mitad de los niños sin tumefacción del tobillo.

	Fractura	Sin Fractura
Tumefacción	Verdadero Positivo 32	Falso Positivo 40
Sin Tumefacción	Falso Negativo 4	Verdadero Negativo 60
	a	b
	c	d
TOTAL	a + c = 36	b + d = 100

Sensibilidad = a / (a+c)

Sensibilidad = 32 / 36 = 89%

Especificidad = d / (b+d)

Especificidad = 60 / 100 = 60%

Eleva la sensibilidad aparente

Reduce la especificidad aparente

Sáez-Alquezar A

Errores en el diseño y el análisis de los estudios de pruebas diagnósticas

Sesgo de verificación diferencial: diferentes métodos de referencia para los que tienen pruebas positivas y negativas

Otra estrategia es utilizar un método de referencia diferente para aquellos pacientes en los que no está indicado el método de referencia habitual.

Ejemplo: se podría incluir a los pacientes con lesiones de tobillo en los que no se ha realizado una radiografía para contacto telefónico varias semanas después de la lesión, y considerar que no han tenido fracturas si se han recuperado sin problemas. Sin embargo, esto puede producir **sesgo de verificación diferencial**, también denominado **sesgo del doble método de referencia**. Este sesgo se puede producir siempre que el método de referencia difiera en las personas que tengan un resultado positivo y negativo de la prueba. En el estudio ya mencionado de mamografía, el método de referencia en las pacientes con mamografía positiva fue la biopsia, mientras que en las que tenían una mamografía negativa fue el seguimiento para ver si se manifestaba un cáncer en el año siguiente. Tener diferentes métodos de referencia para la enfermedad es un problema si los métodos de referencia no siempre dan los mismos resultados, como ocurriría si un cáncer de mama que se podría detectar mediante biopsia en caso de una

PNCQ
Programa Nacional de Controle de Qualidade

INTELIGENCIA ARTIFICIAL (IA)

La inteligencia artificial (IA) tiene el potencial de mejorar significativamente el desempeño y la planificación de las pruebas médicas en varias formas y puede ser una herramienta valiosa para mejorar los resultados de las pruebas médicas en enfermedades que tienen un método de referencia establecido y en aquellas donde la referencia es una definición

Sáez-Alquezar A

PNCQ Programa Nacional de Controle de Qualidade

Aplicaciones de la Inteligencia Artificial que pueden ayudar a mejorar el desempeño de las pruebas médicas

Acciones	Materiales y métodos	Objetivos
Interpretación de imágenes médicas	*Radiografías, Resonancia magnética, tomografía computadorizada.*	*Detección de patrones y anomalías. Mejorar la precisión del diagnóstico*
Selección de pacientes para las pruebas	*Identificar a los pacientes más propensos a beneficiarse de pruebas específicas*	*Optimizar el uso de recursos médicos. Realizar las pruebas de manera más eficiente*
Planificación de procedimientos médicos	*Proporcionando información detallada sobre los pacientes con sugerencias sobre la mejor manera de abordar un procedimiento.*	*Aumentar la precisión y reducir los riesgos asociados con intervenciones médicas.*
Modelos predictivos por Análisis de datos clínicos y de laboratorio	*Análisis de grandes conjuntos. Identificación de factores de riesgo. Elaboración de algoritmos.*	*Predicción temprana de enfermedades. Identificación de poblaciones de alto riesgo. Planificación de estrategias preventivas*
Optimización de terapias	*Análisis de datos clínicos y genéticos*	*Terapias específicas. Mejorar la eficacia de los tratamientos*

Sáez-Alquezar A

PNCQ Programa Nacional de Controle de Qualidade

La OMS esboza las cuestiones que cabe tener en cuenta a fin de regular la inteligencia artificial para la salud
19 de octubre de 2023

- Transparencia y documentación.
- Gestión de riesgos.
- Validación externa de los datos.
- Calidad de los datos.
- Privacidad y protección de los datos.
- Colaboración.
  - Organismos reguladores, pacientes, profesionales de la salud, representantes de las industrias y los asociados gubernamentales.

Regulatory considerations on artificial intelligence for health - **World Health Organization 2023**
https://creativecommons.org/licenses/by-nc-sa/3.0/igo
Under the terms of this licence, you may copy, redistribute and adapt the work for non-commercial purposes.

PNCQ
Programa Nacional
de Controle de Qualidade

Muchas gracias por su atención

amadeo62@gmail.com