

# FIABILITAT I VALIDESA DE LA VERSIÓ ESPANYOLA DEL TEST DE FRESNO. AVALUANT LES COMPETÈNCIES DELS RESIDENTS DE MEDICINA FAMILIAR I COMUNITÀRIA EN PRÀCTICA BASADA EN L'EVIDÈNCIA

**Autors:** Gemma Flores Mateo<sup>1</sup>, Josep M Argimon-Pallàs<sup>2</sup>, Josep Jiménez-Villa<sup>2</sup> i Enriqueta Pujol-Ribera<sup>1</sup>

<sup>1</sup> Tècnica de recerca de l'Institut d'Investigació en Atenció Primària Jordi Gol, Barcelona.

<sup>2</sup> Divisió d'Avaluació, Servei Català de la Salut, Barcelona.

**Correspondència:** [epujol@idiapjgol.org](mailto:epujol@idiapjgol.org)

**Correctors:** Mèrcia Benítez i Sebastià Juncosa

*Aquest treball va ser guardonat amb el primer premi dels VIII Ajuts a la Recerca de la CAMFiC, atorgats el novembre de 2006.*

**Publicat:** juny 2014

Gemma Flores Mateo, Josep M Argimon-Pallàs, Josep Jiménez-Villa i Enriqueta Pujol-Ribera (2014)

*Fiabilitat i validesa de la versió espanyola del test de Fresno. Avaluant les competències dels residents de medicina familiar i comunitària en pràctica basada en l'evidència.*

**Available at:** <http://pub.bsalut.net/butlleti/vol32/iss2/4>

*Aquest és un article Open Access distribuït segons llicència de Creative Commons (<http://creativecommons.org/licenses/by-nc-nd/2.5/es/>)*

---

<http://pub.bsalut.net/butlleti/vol32/iss2/4>

# RESUM

**Objectius:** Adaptar a l'espanyol i validar el test de Fresno per avaluar les competències en pràctica basada en l'evidència (PBE). Conèixer l'efectivitat d'una intervenció formativa als residents de medicina familiar i comunitària (MFiC) en aquestes competències.

## Subjectes i mètodes

**Disseny:** Adaptació transcultural del test a l'espanyol seguint recomanacions internacionals. Estudi transversal per avaluar la fiabilitat i la validesa del test. Estudi abans després per analitzar l'efectivitat de la intervenció formativa.

**Emplaçament:** Unitats docents de MFiC de Catalunya.

**Participants:** Tutors de MFiC (experts, n=56), residents segon any (principiants, n=202) i especialistes en MFiC (experiència intermèdia, n=17).

**Mesures:** edat, sexe i cursos de PBE previs a la intervenció formativa. El test de Fresno de coneixements i habilitats en PBE. Mesures de resultats principals: consistència interna, concordança intra i entre avaluadors; anàlisi dels ítems i capacitat de detectar canvis significatius.

**Intervenció formativa:** Curs amb 4 sessions interactives de 4 hores.

**Resultats:** van participar 275 persones. Van completar el pretest el 94,1% (190/202) dels residents i el 75,2% (152/202) van retornar els dos qüestionaris. Alfa de Cronbach 0,88 i 0,77, respectivament. Fiabilitat entre avaluadors: 0,95 pre-intervenció i 0,85 post-intervenció i intra-

avaluador 0,71 i 0,81, respectivament. La puntuació mitjana: residents 60,4, experts 149,8, i grup intermedi 110,4 (P<0,001). La mitjana de diferència entre abans i després de la intervenció va ser 47,7 (IC del 95%: 42,8-52,5). Mida de l'efecte 1,77 (IC del 95%: 1,57-1,95).

**Conclusions:** La versió espanyola del test de Fresno és una eina fiable i vàlida per avaluar coneixements i habilitats en PBE.

**Paraules clau:** Evidence-Based Practice; Family practice; professional competence; effectiveness; questionnaires; validation studies

# SUMMARY

**Objectives:** Adapt and validate the test of Fresno into Spanish in order to assess competencies in evidence-based practice (EBP). To assess the effectiveness of an educational intervention for family and community medicine (F&CM) residents in these skills.

## Subjects and methods

**Design:** Transcultural Adaptation of the test into Spanish following international recommendations. Cross-disciplinary study in order to assess the reliability and validity of the test. Before – after study in order to analyze the effectiveness of the training intervention.

**Location:** Educational Units of F&CM from Catalonia.

**Participants:** F&CM Tutors (experts, n = 56), second-year residents (beginners, n = 202) and specialists in F&CM (intermediate experience, n = 17).

**Measures:** age, gender and educational intervention prior to the PBE. The Fresno test knowledge and skills in PBE. Main results measures: internal consistency, concordance intra and between evaluators; analysis of the items and the ability to detect significant changes.

**Educational Intervention:** Course with 4 interactive sessions of 4 hours.

**Results:** participated 269 people. Completed the pre-test 94.1% (188/202) of residents and 75.2% (152/202) of them returned the two questionnaires. Alpha of Cronbach 0.88 and 0.77, respectively. Reliability between evaluators: 0.95 pre-intervention and 0.85 post-intervention and intra-evaluator 0.71 and 0.81, respectively. Average score: 60.4 for residents, 149.8 for experts, and 110.4 for intermediate group ( $P < 0.001$ ). The average difference between before and after the intervention was 47.7 (CI 95%: 42.8-52.5). The effect size 1.77 (CI 95%: 1.57-1.95).

**Conclusions:** The Spanish version of the test of Fresno is a reliable and valid tool for assessing knowledge and skills in EBP.

**Keywords:** Evidence-Based Practice; Family practice; professional competence; effectiveness; questionnaires; validation studies

## INTRODUCCIÓ

La pràctica basada en l'evidència (PBE) suposa prendre decisions integrant la millor evidència obtinguda de la recerca de qualitat, les preferències i circumstàncies de les persones (és a dir, les seves preocupacions, expectatives, influències culturals i característiques individuals) i

l'expertesa clínica, per tal de millorar l'efectivitat, la seguretat, l'equitat i en definitiva la qualitat de l'atenció sanitària<sup>1,2</sup>.

Diversos autors<sup>3,4</sup> i organitzacions com l'Institute of Medicine<sup>5</sup> proposen que la PBE és una de les competències clau del currículum de tots els professionals de la salut. No obstant això, fins ara, malgrat els esforços formatius en aquesta àrea<sup>6</sup>, no s'ha aconseguit l'aplicació habitual de la PBE a la capçalera del pacient o en la resolució dels problemes clínics diaris. Entre el 10 i el 40% de pacients no reben una atenció basada en la millor evidència disponible i entre un 11 i un 30% de tractaments són innecessaris o poden ser nocius pels pacients<sup>7-9</sup>.

Diverses revisions sistemàtiques han analitzat els resultats de la formació en PBE en les competències dels participants<sup>10-14</sup>. En síntesi, aquestes revisions, han observat de forma consistent, que la formació en PBE comporta millores en els coneixements en metodologia de la recerca i estadística, així com, en les actituds envers la utilització de la literatura en la presa de decisions, però no han trobat canvis en les conductes. No obstant, aquests resultats han de ser interpretats amb cautela donades les limitacions en la validesa interna de la majoria d'estudis<sup>15</sup>. Entre les principals crítiques d'aquests estudis, cal destacar que es donen pocs detalls sobre com s'ha mesurat l'efectivitat de la intervenció formativa, sobre el procés de desenvolupament i validació dels qüestionaris utilitzats, sobre la forma d'administració i sobre el temps entre l'administració i la intervenció<sup>15</sup>.

Pel que fa als instruments disponibles per avaluar la formació en PBE cal destacar la revisió sistemàtica de Shaneyfelt et al<sup>16</sup>,

que va incloure 115 estudis (representatius de 104 instruments). Aquests autors van identificar diversos instruments de gran qualitat<sup>17-20</sup>, vàlids i fiables per avaluar la majoria de competències requerides per a la PBE (procediments d'aprenentatge, coneixements, habilitats, actituds i conductes en aquesta àrea) dels alumnes individuals i per determinar l'efectivitat dels programes de formació en PBE. Aquests instruments disposen de 3 o més proves de validesa, inclosa la capacitat de discriminar entre diferents nivells de competències (validesa discriminant). D'altra banda, han demostrat la seva fiabilitat entre avaluadors i comporten una avaluació externa dels resultats (no únicament autoavaluació). Donades les seves propietats psicomètriques sòlides, es justifica la seva utilització en avaluacions formatives i/o sumatives.

Entre els instruments d'alta qualitat, en el present estudi s'ha seleccionat test de Fresno<sup>17</sup>, dissenyat per avaluar totes les dimensions de la PBE. Aquest test, s'inicia amb a presentació de dos escenaris clínics que plantegen alguna incertesa clínica. Inclou set preguntes obertes sobre els escenaris de forma que el participant ha de: 1) formular una pregunta estructurada seguint l'estratègia PICO (problema, intervenció, comparació i resultats); 2) identificar el disseny d'investigació més apropiat per respondre la pregunta i realitzar la cerca a les bases de dades electròniques adients; 3) fer la lectura crítica dels articles per tal de determinar la seva pertinença i validesa, i analitzar la magnitud i la importància de les troballes de la investigació. També ha de respondre cinc preguntes que impliquen fer alguns càlculs.

El fet de plantejar preguntes obertes, enlloc de preguntes amb opció múltiple o

veritable-fals, exigeix als participants una resposta reflexiva i desenvolupament dels propis raonaments.

Tot i que el test de Fresno és un instrument de gran qualitat, amb proves de fiabilitat i validesa, aquest atributs no són fixes, sinó que s'han d'avaluar en relació amb la població específica i el context<sup>21</sup>. Així doncs, un instrument que ha demostrat tenir propietats psicomètriques satisfactòries en una població no necessàriament és adequat per ser utilitzat en altres<sup>22</sup>. La validació del qüestionari de Fresno en diferents idiomes, diferents grups de professionals, i entorns culturals, permetrà la seva aplicació generalitzada, així com les comparacions entre països i l'avaluació de diferents mètodes d'ensenyament.

L'objectiu general del present estudi és fer l'adaptació cultural del qüestionari de Fresno a l'espanyol i la seva posterior validació per assegurar l'equivalència de la versió en espanyol comparada amb la versió original en anglès.

Es van definir els següents objectius específics:

1. Fer l'adaptació transcultural del qüestionari en espanyol.
2. Avaluar la fiabilitat del test mitjançant l'anàlisi de la consistència interna i la fiabilitat entre i intra avaluadors
3. Avaluar la validesa de constructe del test mitjançant la comparació de les qualificacions obtingudes pels tutors de Medicina Familiar i Comunitària (MFIC) amb instrucció formal en PBE, dels residents de MFIC, i de metges especialistes en MFIC amb experiència intermèdia en PBE.

4. Analitzar la capacitat del qüestionari de detectar canvis significatius, comparant les puntuacions dels residents de MFIC abans i després d'un curs breu sobre PBE.

## PARTICIPANTS I MÈTODES

### Disseny

El test de Fresno es va administrar a tres grups de professionals: 56 metges tutors dels residents de MFIC, amb formació reglada en PBE i que participaven regularment en estudis experimentals (grup d'experts), 202 residents de segon any (grup de principiants) i 17 especialistes en MFIC amb menor experiència en PBE (experiència intermèdia).

Instrument de mesura: El test de Fresno ([annex 1](#)).

Tal com s'ha descrit anteriorment, el test de Fresno<sup>17</sup>, s'inicia amb la presentació de dos escenaris que plantegen alguna incertesa clínica. Inclou set preguntes obertes sobre aquests escenaris de forma que el participant ha de: 1) formular una pregunta estructurada seguint l'estratègia PICO (problema, intervenció, comparació i resultats); 2) identificar el disseny d'investigació més apropiat per respondre la pregunta i realitzar la cerca a les bases de dades electròniques adients; 3) fer la lectura crítica dels articles per tal de determinar la seva pertinença i validesa, i analitzar la magnitud i la importància de les troballes de la investigació. També ha de

respondre cinc preguntes que impliquen fer alguns càlculs.

La puntuació de les respostes a les preguntes obertes es fa mitjançant rúbriques d'avaluació estandarditzades, que especifiquen criteris explícits d'avaluació per a cada pregunta, amb l'objectiu de minimitzar la variabilitat en la mateixa. Amb aquest procediments s'estableixen quatre categories de resposta: insuficient, limitada, notable i excel·lent. Així, cada resposta té una puntuació numèrica que pot anar de 0 a 24 punts en algunes preguntes o 4 punts en altres. A mode d'exemple, en el cas d'una pregunta amb puntuació màxima de 24 punts, una puntuació < de 8 es considera resposta insuficient, entre 8 i 15 punts resposta limitada, entre 16 i 23 punts resposta notable i 24 resposta excel·lent. La puntuació màxima possible del test de Fresno és de 212 punts, que s'obtenen de sumar les puntuacions de tots els seus ítems. Les puntuacions numèriques de les preguntes permeten establir punts de tall per classificar les respostes en aprovat o insuficient. En el present estudi, dos avaluadors (membres de l'equip investigador) van puntuar les respostes de forma independent, després d'eliminar les dades d'identificació dels participants. Per homogeneïtzar la metodologia de puntuació van fer una prova pilot amb 10 qüestionaris. Els procediments complets de puntuació de Ramos et al es poden consultar al web de BMJ.

(<http://www.bmj.com/content/suppl/2003/02/10/326.7384.319.DC1>)

### **Variables de resultats principals i altres variables**

Com a mesures de resultats principals es va analitzar la consistència interna del test de Fresno, la concordança intra i entre avaluadors (intra-rater inter-rater reliability); anàlisi de la capacitat de discriminació i la dificultat dels ítems i la capacitat del test de detectar canvis significatius (responsiveness). També es va registrar edat, el sexe, formació prèvia en PBE i el temps necessari per respondre les preguntes del test.

### **Intervenció formativa**

La intervenció formativa consistia en 4 sessions interactives de 4 hores de durada, dissenyades per desenvolupar els coneixements i habilitats dels residents de segon any en PBE. El curs incloïa lectures breus, treball en grups petits i anàlisi de casos, seguint els passos proposats per la formació en PBE (plantejament de preguntes clíniques estructurades, cerca de la bibliografia, selecció i obtenció de l'evidència, avaluació crítica de revisions sistemàtiques, assaig clínics i estudis de diagnòstic, interpretació de la rellevància i la precisió dels resultats i aplicació de l'evidència a la pràctica clínica). Totes les sessions emfasitzaven l'aplicabilitat a la pràctica clínica diària i deixaven un temps de reflexió i debat sobre aquest tema. Els docents eren epidemiòlegs i/o especialistes en medicina familiar i comunitària de les UDD de MFiC. La sessió sobre cerca bibliogràfica es va fer amb la participació d'un documentalista.

### **Administració del test de Fresno**

En el grup de residents, el test de Fresno es va administrar abans i després de la intervenció formativa. Per evitar un

possible biaix de memòria, que podria sobreestimar l'efecte de la intervenció, els escenaris i preguntes de càlcul eren diferents en cada administració. El primer test es va administrar un mes abans de la intervenció formativa, coincidint amb una sessió sobre la recerca a l'atenció primària, dirigida a residents i professionals amb experiència intermèdia, en la que també es presentava el present projecte. El segon test es va administrar el darrer dia de la intervenció formativa. En el cas dels tutors, el test de Fresno es va administrar coincidint amb una reunió anual de tutors de MFiC.

### **Grandària de la mostra**

La grandària de mostra es va estimar en base a la capacitat de detectar una mínima diferència important entre el test preintervenció i postintervenció de 10 punts (desviació estàndard de 20), amb una potència estadística del 90%, una hipòtesi bilateral i un error alfa del 5%. Amb aquestes assumpcions i considerant unes possibles pèrdues del 20%, la grandària de la mostra eren 85 participants.

### **Anàlisi estadística**

Es va fer una anàlisi descriptiva (les variables categòriques es van descriure amb freqüències i percentatges, i les contínues amb la mitjana i la desviació estàndard). Mitjançant l'anàlisi de la variància per a dades independents es van comparar les puntuacions obtingudes pels tres grups de participants en base a la seva expertesa (tutors-experts, metges amb experiència intermèdia i residents-principiants).

El canvi en el coneixement i les habilitats entre després i abans del curs es va



calcular com la diferència entre les puntuacions, que es van comparar amb la *t* de Student per a dades aparellades. Així, una diferència positiva significa millora en coneixements i habilitats.

La capacitat de detectar canvis significatius es va analitzar amb la grandària de l'efecte (GE) i la resposta mitjana estandarditzada (RME), el càlcul de les quals es descriurà més endavant.

### **Propietats psicomètriques dels test de Fresno**

#### Consistència interna

Es va mesurar amb correlació de cada ítem amb la puntuació total del test<sup>21</sup> en la que valors superiors a 0,3 mostren evidència de consistència interna i amb l'alfa de Cronbach<sup>23</sup> amb un rang de valors entre 0 i 1, en el que valors superiors a 0,7 es consideren evidència de consistència interna, en el cas de comparacions de grup.

#### Fiabilitat entre i intra avaluadors (Reliability)

La fiabilitat entre observadors ens va mesurar en una mostra de 40 participants (20 respostes pre-intervenció i 20 post-intervenció) mitjançant el coeficient Kappa per les variables qualitatives i el coeficient de correlació intraclasse en el cas de les quantitatives<sup>24</sup>. La fiabilitat intra observadors es va mesurar amb els mateixos coeficients, analitzant 45 respostes.

### **Validesa de constructe**

El grau d'expertesa dels participants es va considerar un indicador clau per avaluar la validesa de constructe, sota el supòsit que una major expertesa s'hauria de traduir en una millor puntuació en el test de Fresno.

### **Capacitat de detectar canvis significatius (Responsiveness)**

La capacitat de detectar canvis significatius és un tipus de validesa de constructe i es va analitzar amb la grandària de l'efecte (GE) i la resposta mitjana estandarditzada (RME). La GE es va obtenir a partir de la diferència de mitjanes entre les puntuacions postintervenció i preintervenció, dividida per la desviació estàndard dels valors preintervenció). Valors positius reflecteixen millora<sup>25</sup> i es defineix com a "petita" (GE <0,2), "petita a moderada" (GE entre 0,2 i 0,5), "moderat a gran" (GE entre 0,51 i 0,79), "gran" (GE > 0,79), tal com suggereix Cohen<sup>26</sup>. La RME es va calcular com la mitjana del canvi de puntuacions dividida per la desviació estàndard d'aquest canvi. Com en el cas de la GE, valors positius indiquen millora<sup>27</sup>.

### **Dificultat dels ítems**

Es va avaluar la dificultat dels ítems, amb la proporció de respostes correctes de cada pregunta. Per analitzar aquest aspecte es va assignar un punt de tall "puntuació aprovat", per a cada pregunta, basat en la mateixa definició dels autors de test de Fresno<sup>17</sup>, que situaven aquest punt de tall en el puntuació del mig de la resposta potent<sup>16-23</sup>. A mode d'exemple, en el cas de preguntes amb puntuació de 0 a 24 el punt de tall era 19,5. Per tal que un test pugui ser utilitzat per participants amb diversos nivells d'expertesa cal que les preguntes tinguin un rang ampli de dificultats, que permetin diferenciar entre ells.

### **Capacitat de discriminació dels ítems**

La capacitat de discriminació dels ítems és la diferència en el percentatge de respostes correctes entre dos grups classificats segons nivell d'expertesa. Es pot

calcular ordenant els participants en base a la puntuació total i seleccionant els del quartil més alt i més baix. També es calcula per a cada ítem i es consideren acceptables valors al voltant del 20%.

El qüestionari es va respondre en paper i les dades van ser introduïdes electrònicament al final dels curs. L'anàlisi estadística es va fer amb Stata software versió 9.0 (STATA Corp, College Station, TX) i l'SPSS software versió 15.0 for Windows (SPSS, Chicago, IL, USA).

### Consideracions ètiques

L'estudi es va desenvolupar seguint normes de bona pràctica en recerca i els principis de la declaració de Helsinki, i va ser aprovat pel Comitè Ètic d'Investigació Clínica de l'IDIAP. Els participants van rebre informació oral i escrita sobre l'estudi i van acceptar voluntàriament la seva participació. Per mantenir la confidencialitat i l'anonimat, les dades es van guardar i analitzar amb una codificació que garantia aquests aspectes.

## RESULTATS

### Adaptació transcultural del test de Fresno

La [figura 1](#) enumera els procediments realitzats. La majoria de paraules es van traduir a l'espanyol sense dificultat. En base a les respostes de les entrevistes cognitives l'equip investigador va elaborar la versió final del test, clara, sense ambigüitats i completa. Les petites diferències es van resoldre en les reunions de consens. L'única

excepció va ser la paraula "evidència", que té un significat diferent en espanyol. Els traductors van utilitzar "prova" com una traducció més acurada d'evidència. No obstant això, donat que la paraula "Evidència" s'ha utilitzat àmpliament en espanyol, l'equip investigador va trobar més apropiat mantenir aquesta paraula.

### Descripció de la mostra i la participació

Van participar 275 persones: 56 experts, 17 experiència intermèdia i 202 residents.

La mitjana d'edat dels participants era 31 anys (desviació estàndard=8), el 73% eren dones. Entre els residents 190/202 (94,1%) van completar el pretest, i 158 (78,2%) retornar el post-test. En total, 152/202 residents (75,2%) van tornar els dos qüestionaris. Els 44 residents que no van retornar el segon qüestionari tenien menor puntuació que els que van retornar els dos (58,8 vs 63,9;  $P=0,03$ , respectivament).

### Consistència interna

El qüestionari va mostrar una consistència interna elevada. L'alfa de Cronbach va ser de 0,88 i 0,77 abans i després de la intervenció, respectivament. El rang de valors dels coeficients de correlació ítem-total van ser entre 0,36 per la pregunta 1 (formulació d'una pregunta) i 0,67 per a la pregunta 3 (cerca de l'evidència) en el qüestionari pretest. En el qüestionari post intervenció els coeficients van oscil·lar entre 0,25 (pregunta 11: el



millor disseny de l'estudi, diagnòstic) a 0,52 (pregunta 5: rellevància).

### Fiabilitat entre i intra avaluadors

La fiabilitat entre avaluadors va ser de 0,95 preintervenció i 0,85 post-intervenció i intra-avaluador 0,71 i 0,81, respectivament (taules 1 i 2).

Dificultat dels ítems i capacitat de discriminació

El rang de dificultat dels ítems va oscil·lar entre dificultat moderada i elevada (taula 3). La capacitat de discriminació dels ítems es mostra a la taula 4.

### Validesa de constructe

La puntuació mitjana (desviació típica) dels residents va ser 60,4 (DT= 25), la dels tutors/ experts 149,8 (DT=23,2) i la grup amb expertesa intermèdia 110,4 (DT=11,5), amb diferències de mitjanes i comparacions entre els grups estadísticament significatives ( $P < 0,001$  i  $P < 0,01$  respectivament). Els residents amb formació prèvia en PBE van obtenir una puntuació superior i estadísticament significativa (75,9 vs 57,6;  $P < 0,001$ ) comparats amb la resta de residents.

Capacitat de detectar canvis significatius (Responsiveness)

La diferència mitjana en les puntuacions dels residents entre el test pre i post-intervenció va ser de 47,7 (IC del 95%: 42,8-52,5) punts. La GE i la RME van ser grans i es presenten a la

taula 4. La taula 5 mostra el percentatge de residents amb puntuació d'aprobat abans i després de la intervenció formativa.

## DISCUSSIÓ

### Resum de les troballes

El present estudi mostra que la versió espanyola del test de Fresno és una eina fiable i vàlida per avaluar els coneixements i habilitats en PBE. El qüestionari va diferenciar els participants segons nivells d'expertesa en PBE, la consistència interna i la fiabilitat són elevades i la capacitat de detectar canvis significatius (responsiveness) va ser gran ( $> 1.6$ ) amb els anàlisis realitzats.

### Fiabilitat

Consistència interna

En el nostre estudi l'alfa de Cronbach va ser igual a la de l'instrument original (0,88 preintervenció i 0,77 postintervenció) i superior al valor 0,7 proposat per Streiner i Norman (24) àmpliament acceptat. D'altra banda, les correlacions de cada ítem amb la puntuació total del test van ser superiors a 0,30, excepte per a la pregunta 11 postintervenció. Aquest resultat indiquen una bona consistència interna, especialment si es té en compte que el test inclou ítems fàcils i difícils, la qual cosa tendeix a reduir-la.

### Fiabilitat entre i intra avaluadors

La fiabilitat entre avaluadors, essencial per a un instrument amb preguntes obertes, subjectes a la interpretació dels avaluadors, va ser molt bona, en ambdues mesures (taula 1). La fiabilitat intra avaluadors va ser superior a 0,7, però inferior a la fiabilitat entre avaluadors, amb valors particularment baixos en les preguntes sobre rellevància i validesa interna. Una possible explicació és que la fiabilitat intra avaluador va ser avaluada a l'inici del procés de qualificació, quan l'avaluador era menys expert i més inconsistent, mentre que la fiabilitat entre avaluadors es va avaluar a la meitat del procés de qualificació, després d'haver qualificat més de cent tests i per tant, amb una major expertesa. Aquests resultats tenen algunes implicacions: 1) els canvis en coneixements i habilitats en PBE poden ser fins a cert punt erronis en alguns casos, donada una menor concordança intra avaluadors en alguna pregunta, fet que podria resultar en un biaix en favor de la hipòtesi nul·la, en un test per avaluar una d'intervenció educativa; 2) és important continuar precisant i perfeccionant el procés de puntuació del test, 3) l'entrenament previ dels avaluadors és imprescindible i finalment, 4) la importància d'avaluar la fiabilitat entre i intra avaluadors. Aquesta darrera no va ser avaluada en la validació de l'instrument original <sup>17</sup>.

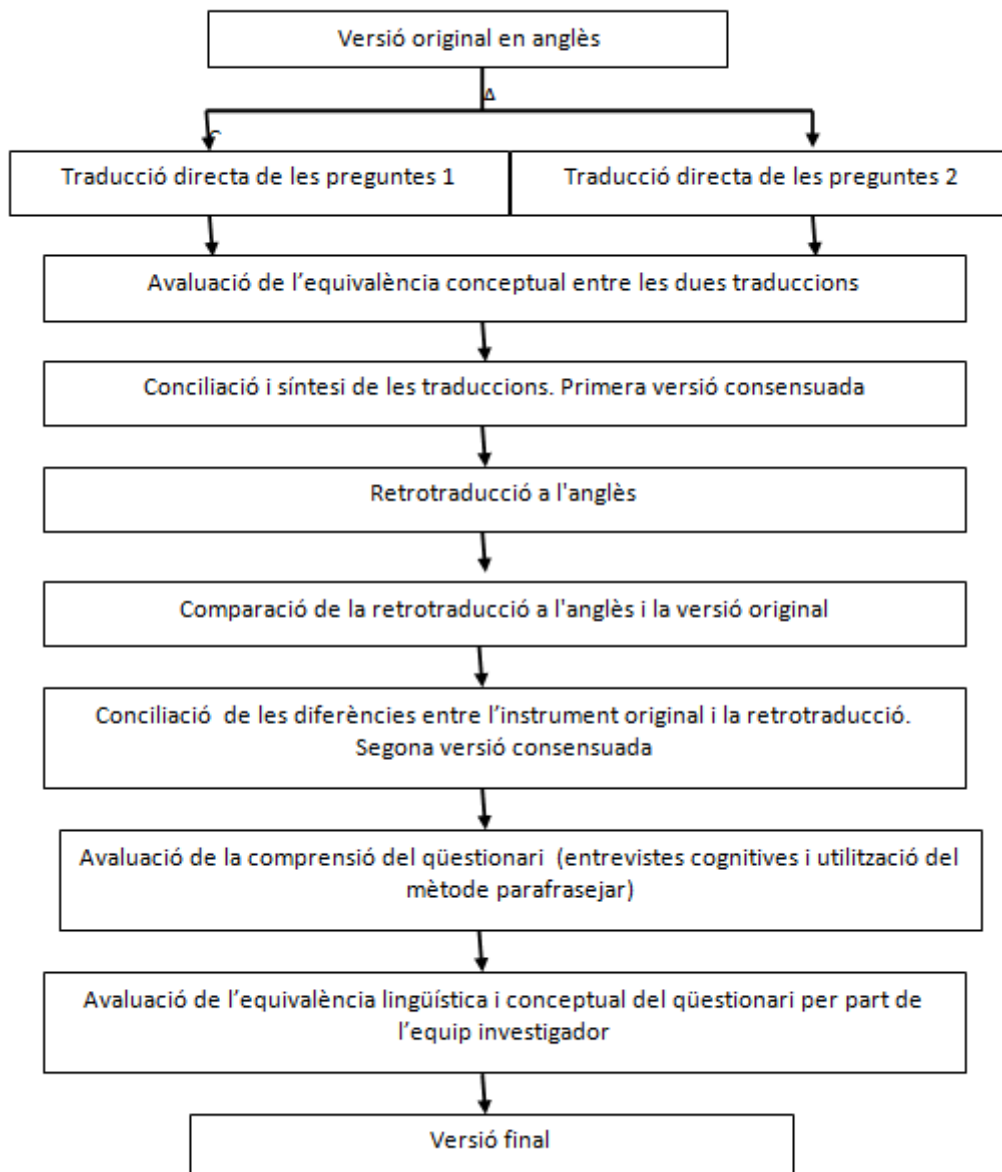
### Validesa

#### Validesa de constructe

En el present estudi les puntuacions del test de Fresno han sigut superiors en el grup de participants d'experts, seguit dels d'experiència intermèdia i inferiors en els principiants. A més, els residents amb instrucció prèvia en PBE també van tenir una puntuació més alta que la resta dels residents. Aquests resultats confirmen la validesa discriminant del test de Fresno.

Capacitat de detectar canvis significatius o importants (Responsiveness)

En el present estudi la capacitat de detectar canvis significatius del test de Fresno es va analitzar mitjançant la distribució de les seves puntuacions. Es van obtenir dues mesures: la grandària de l'efecte (ES) per quantificar la magnitud del canvi i la mitjana de resposta estandarditzada (MRE) que mostra els canvis estadísticament significatius<sup>28</sup>. Els valors d'aquestes mesures van ser de 1,77 i 1,65 respectivament, amb resultats similars en el subgrup de residents amb formació prèvia en PBE. Per tant, la capacitat de detectar canvis significatius de la versió espanyola del Fresno és molt alta i el test és sensible al canvi en el context d'un estudi abans-després. A més, donat que es va obtenir una bona resposta en els dos grups de diferent expertesa en PBE, podria ser utilitzat avaluar professionals principiants o amb major experiència.

**Figura 1. Procés d'adaptació transcultural del test de Fresno a l'espanyol**

**Taula 1. Anàlisi de la fiabilitat dels ítems del test de Fresno. Coeficient de Correlació Intraclasse**

Pregunta	Entre avaluadors		Intraavaluadors	
	Preintervenció	Postintervenció	Preintervenció	Postintervenció
P1 Plantejar la pregunta	0,97	0,86	0,78	0,90
P2 Fonts d'informació	0,95	0,87	0,65	0,74
P3 Cerca	0,92	0,92	0,65	0,83
P4 Disseny de l'estudi	0,87	0,73	0,71	0,84
P5 Rellevància	0,88	0,90	0,46	0,34
P6 Validesa Interna	0,89	0,76	0,44	0,66
P7 Magnitud de l'efecte	0,99	0,79	0,72	0,85
P1 a P7	0,95	0,85	0,71	0,81

P: Preguntes

**Taula 2. Anàlisi de la dificultat dels ítems del test de Fresno**

	Preintervenció	Postintervenció
P1 Plantejar la pregunta	0.29	0.33
P2 Fonts d'informació	0.20	0.21
P3 Cerca	0.15	0.35
P4 Disseny de l'estudi	0.38	0.51
P5 Rellevància	0.29	0.38
P6 Validesa Interna	0.21	0.58
P7 Magnitud de l'efecte	0.10	0.20
P8 Sensibilitat	0.54	0.84
Especificitat	0.44	0.77
Valor Predictiu Positiu	0.44	0.79
Valor Predictiu Negatiu	0.39	0.78
Quocient de Probabilitat Positiu	0.17	0.46
P9 Reducció Absoluta de Risc	0.40	0.84
Reducció Relativa de Risc	0.24	0.66
Nombre Necessari a Tractar	0.25	0.62
P10 Interval de Confiança	0.29	0.56
P11 Millor disseny, estudi diagnòstic	0.21	0.78
P12 Millor disseny, estudi pronòstic	0.24	0.68

P: Preguntes.

Els valors representen la proporció de participants que assoleixen la puntuació "aprovat" per a cada ítem.

**Taula 3. Anàlisi de la capacitat de discriminació dels ítems del test de Fresno**

	Preintervenció	Postintervenció
P1 Plantejar la pregunta	0.32	0.29
P2 Fonts d'informació	0.15	0.22
P3 Cerca	0.16	0.29
P4 Disseny de l'estudi	0.36	0.46
P5 Rellevància	0.24	0.33
P6 Validesa Interna	0.24	0.59
P7 Magnitud de l'efecte	0.07	0.19
P8 Sensibilitat	0.47	0.82
Especificitat	0.44	0.83
Valor Predictiu Positiu	0.42	0.80
Valor Predictiu Negatiu	0.36	0.82
Quocient de Probabilitat Positiu	0.14	0.80
P9 Reducció Absoluta de Risc	0.28	0.48
Reducció Relativa de Risc	0.21	0.83
Nombre Necessari a Tractar	0.31	0.67
P10 Interval de Confiança	0.24	0.62
P11 Millor disseny, estudi diagnòstic	0.17	0.54
P12 Millor disseny, estudi pronòstic	0.20	0.68

P= Preguntes

Els valors representen la diferència en la proporció dels participants que contesten correctament entre els del quartil de puntuació més alta y la proporció de participants que contesten correctament entre el quartil de puntuació més baixa.

**Taula 4. Capacitat de detectar canvis significatius (*Responsiveness*) del test de Fresno segons grup**

	Total de residents (IC 95%)	Residents amb formació en PBE (CI 95%)
Grandària de l'efecte*	1.77 (1.57-1.95)	1.78 (1.38-2.17)
Mitjana de resposta estandaritzada**	1.65 (1.47-1.82)	1.60 (1.24-1.96)

**Taula 5. Percentatge de residents amb puntuació d'aprobat segons el test de Fresno abans i després de la intervenció formativa i interval de confiança de la diferència.**

	Pre intervenció	Post intervenció	IC 95%
P1 Plantejar la pregunta	22.9	28.3	-2.1-13.1
P2 Fonts d'informació	12.3	22.3	3.3-16.1
P3 Cerca	04.8	22.6	12.0-23.6
P4 Disseny de l'estudi	22.9	37.7	6.7-22.5
P5 Rellevància	14.4	23.9	2.7-16.3
P6 Validesa Interna	28.2	38.2	2.1-18.4
P7 Magnitud de l'efecte	1.6	14.5	8.4-17.7
P8 Sensibilitat	42.2	82.3	32.2-47.6
Especificitat	34.0	78.5	36.9-52.4
Valor Predictiu Positiu	32.6	81.5	41.5-56.5
Valor Predictiu Negatiu	25.9	79.9	46.4-61.1
Quocient de Probabilitat Positiu	8.2	47.8	32.5-46.6
P9 Reducció Absoluta de Risc	21.3	81.0	52.7-66.7
Reducció Relativa de Risc	06.5	63.9	52.1-64.3
Nombre necessari a tractar	10.6	60.3	42.7-56.9
P10 Interval de Confiança	13.2	48.3	27.7-42.6
P11 Millor disseny, estudi de diagnòstic	24.0	44.0	11.7-27.8
P12 Millor disseny, estudi de pronòstic	24.9	34.9	12.0-28.3

Abreviatures: P: Preguntes. IC=Interval de confiança.

### Punts forts i limitacions de l'estudi

El present estudi té diversos punts forts a destacar. En primer lloc, la validació d'un qüestionari és un procés longitudinal, continu i dinàmic que persisteix durant la seva utilització repetida. La validació del test Fresno en espanyol, en una població amb diversos graus d'experiència en PBE i diferent a la estudiada per Ramos et al<sup>17</sup>, aporta noves evidències de validesa que permeten proposar la generalització de la seva utilització, la comparacions entre els països i l'avaluació de diferents mètodes d'ensenyament. En segon lloc, l'estudi

actual va mesurar la capacitat de test per detectar canvis significatius, aspecte no contemplat en la validació original<sup>17</sup>. Això serà de gran utilitat per estimar la mida de la mostra necessària en futurs assajos clínics que vulguin mesurar l'efectivitat d'una intervenció educativa. En tercer lloc, aquest estudi es va realitzar en condicions que reflecteixen la pràctica habitual, amb múltiples professors a qui es va aconsellar no modificar les seves sessions. La participació de diversos professors dificulta la normalització de la intervenció formativa i la manca d'estandardització pot disminuir la capacitat de detectar diferències, però



resulta coherent amb l'enfocament pragmàtic de l'estudi <sup>29,30</sup>.

Entre les limitacions potencials d'aquest estudi cal considerar les següents. Tot i que el test va ser validat en una mostra amb un rang ampli d'expertesa en PBE, hi va haver un predomini dels participants principiants i el percentatge d'especialistes en MFiC amb experiència intermèdia va ser baix (5%). Per tant, no podem descartar que una mostra amb una estructura diferent no hagués donat uns altres índexs psicòmètrics. A més, les diferències obtingudes entre el pretest i el posttest podrien ser atribuïbles sobretot a la millora dels coneixements i habilitats dels residents. És possible que la restricció de la mostra, als subjectes amb respostes completes, també hagi influït en les propietats psicòmètriques del test. Els 44 residents que no van retornar el segon qüestionari, podrien estar menys compromesos, menys informats i menys segurs que els qui el van retornar.

Altres autors han adaptat i validat el test de Fresno en altres poblacions i han confirmat la seva fiabilitat i validesa <sup>31,32</sup>. En el nostre cas, per raons logístiques, només es van incloure residents de MFiC. Si bé es podria suggerir que entre els residents hi ha més similituds que diferències pel que fa a la PBE, la utilització del test de Fresno en residents d'altres especialitats requeriria proves de validesa i fiabilitat específiques.

La fiabilitat entre i intra avaluadors, la consistència interna i la capacitat de discriminació d'un test estan molt lligades a la població a la que s'aplica i no es pot suposar que aquestes propietats es mantenen en poblacions de característiques diferents. La utilització d'un instrument en diferents poblacions contribueix a augmentar les seves proves de validesa. El test de Fresno s'hauria aplicar amb nous escenaris clínics, nous exemples de preguntes numèriques, i buscant una resposta i correcció més simple i àgil. La recerca futura podria contemplar el desenvolupament i la validació d'una versió més curta, utilitzant la versió llarga com a estàndard de referència, o amb respostes d'opció múltiple. Tot i que la fiabilitat intra avaluadors de la versió espanyola va ser bona, caldria millorar el sistema de puntuació per augmentar aquesta fiabilitat. També és important confirmar si la fiabilitat millora com efecte de la formació contínua i la familiarització creixent dels avaluadors amb els matisos del procés de qualificació del test Fresno. Això permetria establir una mesura del temps d'entrenament recomanable per realitzar aquests procediments.

## Conclusions

Les troballes d'aquest estudi indiquen que la versió espanyola del test de Fresno és un instrument vàlid i fiable per avaluar els coneixements i habilitats en PBE dels residents de MFiC de parla hispana. També un bona eina

d'avaluació, ja que és capaç de detectar els canvis rellevants en els coneixements i habilitats després d'una intervenció educativa. Per tant, podria

ser utilitzat com a variable principal de resultats en estudis aleatoritzats i controlats.

## BIBLIOGRAFIA

1. Haynes RB, Devereaux PJ, Guyatt GH. Physicians' and patients' choices in evidence based practice. *BMJ* 2002 Jun 8;324(7350):1350.
2. Ubbink DT, Guyatt GH, Vermeulen H. Framework of policy recommendations for implementation of evidence-based practice: a systematic scoping review. *BMJ Open* 2013;3(1).
3. Dawes M, Summerskill W, Glasziou P, Cartabellotta A, Martin J, Hopayian K, et al. Sicily statement on evidence-based practice. *BMC Med Educ* 2005 Jan 5;5(1):1.
4. Guyatt G, Cook D, Haynes B. Evidence based medicine has come a long way. *BMJ* 2004 Oct 30;329(7473):990-1.
5. Crossing the quality chasm: a new health system for the 21st century. Washington, DC: National Academies Press, 2001. 2001.
6. Barzansky B, Etzel SI. Educational programs in US medical schools, 2004-2005. *JAMA* 2005 Sep 7;294(9):1068-74.
7. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Lancet* 2003 Oct 11;362(9391):1225-30.
8. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The quality of health care delivered to adults in the United States. *N Engl J Med* 2003 Jun 26;348(26):2635-45.
9. Shrank WH, Asch SM, Adams J, Setodji C, Kerr EA, Keesey J, et al. The quality of pharmacologic care for adults in the United States. *Med Care* 2006 Oct;44(10):936-45.
10. Coomarasamy A, Khan KS. What is the evidence that postgraduate teaching in evidence based medicine changes anything? A systematic review. *BMJ* 2004 Oct 30;329(7473):1017.
11. Flores-Mateo G, Argimon JM. Evidence based practice in postgraduate healthcare education: a systematic review. *BMC Health Serv Res* 2007;7:119.
12. Green ML. Graduate medical education training in clinical epidemiology, critical appraisal, and evidence-based medicine: a critical review of curricula. *Acad Med* 1999 Jun;74(6):686-94.
13. Norman GR, Shannon SI. Effectiveness of instruction in critical

appraisal (evidence-based medicine) skills: a critical appraisal. *CMAJ* 1998 Jan 27;158(2):177-81.

14. Parkes J, Hyde C, Deeks J, Milne R. Teaching critical appraisal skills in health care settings. *Cochrane Database Syst Rev* 2001;(3):CD001270.

15. Hatala R, Guyatt G. Evaluating the teaching of evidence-based medicine. *JAMA* 2002 Sep 4;288(9):1110-2.

16. Shaneyfelt T, Baum KD, Bell D, Feldstein D, Houston TK, Kaatz S, et al. Instruments for evaluating education in evidence-based practice: a systematic review. *JAMA* 2006 Sep 6;296(9):1116-27.

17. Ramos KD, Schafer S, Tracz SM. Validation of the Fresno test of competence in evidence based medicine. *BMJ* 2003 Feb 8;326(7384):319-21.

18. Fritsche L, Greenhalgh T, Falck-Ytter Y, Neumayer HH, Kunz R. Do short courses in evidence based medicine improve knowledge and skills? Validation of Berlin questionnaire and before and after study of courses in evidence based medicine. *BMJ* 2002 Dec 7;325(7376):1338-41.

19. Taylor R, Reeves B, Mears R, Keast J, Binns S, Ewings P, et al. Development and validation of a questionnaire to evaluate the effectiveness of evidence-based practice teaching. *Med Educ* 2001 Jun;35(6):544-7.

20. Taylor RS, Reeves BC, Ewings PE, Taylor RJ. Critical appraisal skills training for health care professionals: a randomized controlled trial. *BMC Med Educ* 2004 Dec 7;4(1):30.

21. Norman GR, Streiner DL. *Health Measurement Scales: A Practical Guide to their Development and Use*. 4th edition ed. Oxford University Press; 2008.

22. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess* 1998;2(14):i-74.

23. Bland JM, Altman DG. Cronbach's alpha. *BMJ* 1997 Feb 22;314(7080):572.

24. Streiner DL, NGR. *Health measurement scales: A practical guide to their development and use*. (3rd ed.) ed. 2003.

25. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Control Clin Trials* 1991 Aug;12(4 Suppl):142S-58S.

26. Cohen J. *Statistical power analysis for the behavioral sciences*. ed. 2nd ed. New Jersey: 1988.

27. Bland JM, Altman DG. *Statistics Notes: Validating scales and indexes*. *BMJ* 2002 Mar 9;324(7337):606-7.

28. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-

related quality of life instruments: guidelines for instrument evaluation. *Qual Life Res* 2003 Jun;12(4):349-62.

29. Glasgow RE. What does it mean to be pragmatic? Pragmatic methods, measures, and models to facilitate research translation. *Health Educ Behav* 2013 Jun;40(3):257-65.

30. Foster N, Little P. Methodological issues in pragmatic trials of complex interventions in primary care. *Br J Gen Pract* 2012 Jan;62(594):10-1.

31. McCluskey A, Bishop B. The Adapted Fresno Test of competence in evidence-based practice. *J Contin Educ Health Prof* 2009;29(2):119-26.

32. Tilson JK. Validation of the modified Fresno test: assessing physical therapists' evidence based practice knowledge and skills. *BMC Med Educ* 2010;10:38.

# ANNEX 1. VERSIÓ ESPANYOLA DEL TEST FRESNO SOBRE PRÀCTICA BASADA EN L'EVIDÈNCIA

Se ruega rellenar todo el cuestionario en una misma sesión. Hay 7 preguntas de respuesta corta, 2 preguntas que requieren una serie de cálculos sencillos, y tres preguntas más de respuesta abierta. Prevea alrededor de 30 minutos para realizar la prueba.

Responda las preguntas 1- 4 basándose en los siguientes escenarios clínicos:

- Acude a su consulta una madre con un lactante de 3 meses que presenta desde hace dos semanas al anochecer, un llanto intenso e inconsolable, gritando y encogiéndose las piernas como si le doliera el vientre. Tras una exploración física completamente normal, le diagnostica de un cólico del lactante. La madre le comenta que otras madres le han aconsejado que le administre un té con extractos de manzanilla, verbena, hinojo y menta balsámica para mejorar los síntomas, pero usted no conoce mucho sobre las propiedades analgésicas de dicho té en el cólico del lactante.
- Acude a su consulta un paciente con acúfenos crónicos. Después un estudio completo se descarta cualquier patología subyacente que explique los acúfenos. El paciente refiere que le dificultan la conciliación del sueño y le solicita una benzodiacepina, pero usted no conoce la evidencia que existe sobre el uso de las benzodiacepinas en el tratamiento de los acúfenos crónicos y su seguridad.

1. Escriba una pregunta clínica para cada escenario de manera que le ayude a realizar una búsqueda bibliográfica y a elegir el mejor artículo de entre los que encuentre.

2. ¿Qué fuentes de información podrían consultar los clínicos para encontrar una respuesta a estas preguntas? Cite todos los posibles tipos o categorías de fuentes de información que conozca. Describa las ventajas y los inconvenientes más importantes de cada una.

3. Elija uno de los escenarios clínicos ¿Qué tipo de estudio (diseño) cree que sería más adecuado para contestar esta pregunta? ¿Por qué?

4. Describa los campos y palabras o términos para realizar una búsqueda bibliográfica en PubMed para localizar estudios que le ayuden a contestar la pregunta clínica. Explique los argumentos en que basa esta estrategia. Describa cómo podría limitar su búsqueda si fuera necesario, y explique sus argumentos.

Las preguntas 5-7 consideran una revisión crítica de la literatura dividida entre la importancia, validez, y la magnitud del tamaño del efecto. Estas pueden ser unas subdivisiones arbitrarias del proceso de la revisión crítica.

5. Cuando encuentra un artículo original ¿qué características del estudio consideraría Vd. relevantes para determinar si le sería útil en su práctica clínica? Incluya ejemplos. (Esta pregunta no se refiere a los aspectos de validez del estudio, ni a la magnitud de los resultados)

6. Cuando encuentra un artículo original, ¿qué características del estudio consideraría usted para determinar la validez de los hallazgos? Haga comentarios si lo considera oportuno. (Ya se ha considerado la relevancia, y la pregunta 7 se referirá a cómo determinar la relevancia de los hallazgos, por tanto en esta pregunta céntrese en la validez del estudio)

7. Cuando encuentra un artículo original, ¿qué aspectos valoraría usted para determinar la magnitud de los resultados? Incluya ejemplos y comentarios si lo cree oportuno. (Ya se ha considerado la relevancia y la validez para esta pregunta, considere cómo determinar el tamaño y el significado del efecto comentado en el estudio.)

8. Un estudio reciente sobre la precisión de los criterios de Centor (adenopatías cervicales, exudado amigdalár, fiebre y ausencia de síntomas catarrales) para el diagnóstico de la faringoamigdalitis estreptocócica, incluyó a 140 pacientes mayores de 14 años con “dolor de garganta” como principal síntoma, 34 de los cuales presentaron frotis positivo (Gold Standard). Entre los diagnosticados 27 presentaban los 4 criterios de Centor positivos. Entre los individuos que no se diagnosticaron de faringoamigdalitis estreptocócica, 8 presentaron los cuatro criterios de Centor positivos. Basándose en este estudio:

Calcule o escriba la fórmula:

En base a estos resultados, la sensibilidad del Centor es \_\_\_\_\_

En base a estos resultados, la especificidad del Centor es \_\_\_\_\_



En base a estos resultados, el valor predictivo positivo del Centor es \_\_\_\_\_

En base a estos resultados, el valor predicativo negativo Centor es \_\_\_\_\_

En base a estos resultados, el cociente de probabilidad positivo \_\_\_\_\_

9. Un estudio de casos y controles reciente mostró que el 10,9% de los pacientes infectados por el HIV vacunados con la vacuna antipneumocócica 23-valente polisacárida, presentaron una neumonía neumocócica al año del seguimiento, mientras que en el grupo de no vacunados la sufrió un 23,8%. Basándose en estos resultados:

Calcule o escriba la fórmula:

La reducción absoluta de riesgo para sufrir una neumonía neumocócica es \_\_\_\_\_

La reducción relativa de riesgo de sufrir una neumonía neumocócica es \_\_\_\_\_

El número que se necesita para tratar (NNT) evitar una neumonía neumocócica es \_\_\_\_\_

10. Una revisión sistemática evaluó el efecto de los antiagregantes orales comparado con el del placebo en la prevención primaria de la enfermedad cardiovascular en pacientes con alto riesgo de enfermedad cardiovascular. Los resultados mostraron que el riesgo relativo de eventos de cardiovasculares era de 0,89 para los individuos que tomaban antiagregantes. Esto indica que los antiagregantes son protectores de sufrir un evento cardiovascular, pero nos preguntamos si esta diferencia es importante en términos estadísticos; para ello consultamos el intervalo de confianza. Indique un ejemplo de un intervalo de confianza que apoyaría la conclusión de que el riesgo relativo de eventos cardiovasculares estaba disminuido entre los individuos que tomaban antiagregantes.

0.89 [95% CI, \_\_\_\_\_]

11. ¿Qué diseño es el mejor para evaluar la eficacia de un nuevo tratamiento?
  
12. ¿Qué diseño es el mejor para evaluar un el pronóstico de una enfermedad?