

ANÁLISIS DE LA EFECTIVIDAD DE LA DIAGNOSIS DEL DENGUE EN EL ESTADO DE GUERRERO, MÉXICO 2007: HECHOS ESTADÍSTICOS Y CONSENSO

Sira Allende *, Carlos Bouza¹*; Saúl López Silva** y Agustín S. Moreno***

* Universidad de La Habana, Cuba

**Laboratorio Estatal de Salud Pública "Dr. Galo Soberón y Parra", Guerrero, México

*** Universidad Autónoma de Guerrero, México

RESUMEN

En este trabajo se presenta un estudio de los criterios emitidos por la clínica respecto a las personas que acudieron a consulta, considerándose como infectados por el dengue, en el estado de Guerrero, México, durante el 2006. La clasificación de la clínica fue evaluada tomando los resultados de los exámenes de laboratorio como regla de oro. Para hacerlo se llevó a cabo un estudio con todas las personas consultadas que tenían los síntomas. El consenso de los médicos y el laboratorio fue evaluado al estudiar datos recogidos aplicando los índices de consenso más populares; Se constató una baja consensualidad en la predicción del dengue clásico mientras que los niveles para el hemorrágico fueron más adecuados. Este estudio se llevó a cabo analizando cada jurisdicción del estado. Este estudio del acuerdo entre clínica y laboratorio fue desarrollado por vez primera. Este junto con el análisis de datos brindaron una panorámica del comportamiento de esa epidemia.

ABSTRACT

This paper presents a study of the clinical criteria on the persons consulting the physicians considering, that they were affected by dengue, in the state of Guerrero, México, during 2006. The classification made by the clinic was evaluated with respect to the golden rule represented by the laboratory results. All the consulted persons with the symptoms were studied. The consensus between physicians and laboratory was evaluated by studying the collected data and computing the most popular indices of consensus. It was established that a low consensus level was present in the prediction of classical dengue while the levels for the hemorrhagic were more adequate. This study was developed in each state jurisdiction. This the first occasion in which this study is developed. It and the analysis of the data gave an overall idea of the behavior of the epidemic.

KEY WORDS: Consensus index, kappa, golden rule, efficiency

MSC 62P10

1. INTRODUCCION

En la medicina es frecuente que el médico, experto, trate de clasificar en un par de clases excluyentes a los pacientes sometidos a un tratamiento: mejora o no mejora. Asumiendo que la clasificación es realizada usando el método de "doble a ciegas", lo que implica independencia, se trabaja con una variable que solo toma dos posibles valores

$$Y_t = \begin{cases} + & \text{si } t \in A \\ - & \text{si } t \notin A \end{cases}$$

Se valora cada objeto t , usando dos criterios o dos expertos, y se obtienen dos valores de Y_t . Cada objeto se clasifica como "positivo" o "negativo". Al analizar los reporte de los dos expertos sobre el n objetos los resultados se pueden reflejar en una tabla como la tabla 1.

Son importantes el número de falsos positivos n_{21} y el de los falsos negativos n_{12} . En la practica medica es claro su uso pero también lo son los valores de las otras casillas cuando hacemos una valoración integral

¹ bouza@matcom.uh.cu

del proceso de clasificación. Esto lleva al problema del consenso en la clasificación. Diferentes índices han sido propuestos para evaluar este tipo de problemas. Ellos hacen uso de las proporciones

Tabla 1. Clasificación de n objetos por dos expertos.

Experto 1	Experto 2		Total
	Positivo	Negativo	
Positivo	n_{11}	n_{12}	n_{1+}
Negativo	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	n

$p_{ij} = \frac{n_{ij}}{n}$, $q = \frac{1}{2}(p_{2+} + p_{+2})$, $p = \frac{1}{2}(p_{+1} + p_{1+})$ donde $p_{i+} = \frac{n_{i+}}{n}$, $p_{+i} = \frac{n_{+i}}{n}$
de una otra forma.

En la Sección 2 se da una visión de los procedimientos de acopio de la información. En la siguiente presentamos un grupo casi exhaustivo de medidas de consenso. En la Sección 4 se brindan los resultados obtenidos al estudiar los resultados del consenso entre clínica y laboratorio usando diversos índices de consenso. Un análisis mas detallado por Jurisdicción se brinda en ella también.

2. DATOS DE LA EPIDEMIA ANALIZADOS Y MÉTODOS DE CLASIFICACIÓN DE PACIENTES DEL SISTEMA DE SALUD

Durante 2006, el LESP del estado de Guerrero recibió 16,604 muestras para confirmación serológica de casos de Dengue, de los cuales sólo consideramos para el estudio los recabados hasta septiembre del 2006, que hacen un total de 10890 casos. El proceso de clasificación de caso probable de DC o DH, se ajusta al protocolo adoptado por el sistema nacional de salud para el caso de enfermedades transmitidas por vector y, según este, el médico tratante debe hacer, por cada paciente sospechoso, un estudio epidemiológico de caso. El esquema del proceso de clasificación de los pacientes es el que aparece en la figura 2.1.

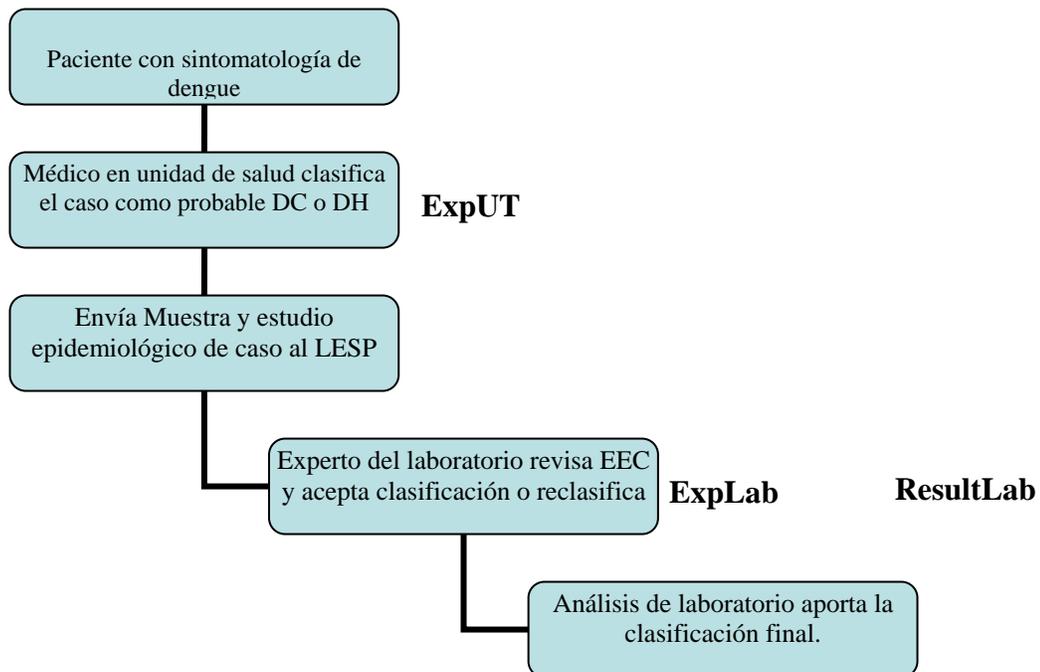


Figura 2.1, Esquema de clasificación de los pacientes con síntomas de dengue

Para analizar el consenso entre la clasificación emitida por el experto clínico que tuvo contacto con el paciente, el experto de laboratorio que recibe y analiza los datos del caso, y la clasificación final que otorga el resultado de laboratorio, se realizan las comparaciones de las clasificaciones entre ExpUT y ExpLab, ExpUT y ResultLab, ExpLab y ResultLab, y se generan, para cada uno de los casos, las tablas de contingencia correspondientes y se calculan los índices en referencia.

Para validar los resultados se hace una selección aleatoria de 50 muestras de tamaño 500, se trabaja con la distribución de medias y se estima la media poblacional para cada índice, además de su varianza y desviación típica, a partir de la cual construimos un intervalo de confianza para la media poblacional.

3. MEDIDAS DE CONSENSO

Es razonable que el interés se centre en un criterio asociado a las concordancias existentes entre aquellos considerados como positivos por al menos uno de los expertos. Este enfoque es el usual en la Taxonomía Numérica en la que n_{22} no es tomado en cuenta. Uno de esos índices es el propuesto por Dice (1945)

$$I_D = \frac{2p_{11}}{p_{1+} + p_{+1}} = \frac{p_{11}}{p} \quad (1)$$

que es la probabilidad empírica condicionada de concordancia entre los expertos dada la positividad del criterio asociado a la casilla “positivo-positivo”. El criterio opuesto establece el uso de

$$I'_D = \frac{2p_{22}}{p_{2+} + p_{+2}} = \frac{p_{22}}{q} \quad (2)$$

Para estos índices el valor cero determina un perfecto desacuerdo de los expertos y el valor 1 un perfecto acuerdo. Cuando el acuerdo observado es equivalente al de un experimento aleatorio es igual a 0,5.

Por su parte Rogot-Goldberg (1966) propusieron usar la suma de estos dos índices

$$I_{RG} = \frac{p_{11}}{p} + \frac{p_{22}}{q} = I_D + I'_D \quad (3)$$

Entonces este se mueve entre 0 y 2. El óptimo consenso determina un valor de 2 y la aleatoriedad en el acuerdo 1. Por su parte usando

$$I'_{RG} = \frac{I_{RG}}{4} \quad (4)$$

En este 0,5 es el valor del índice para el consenso perfecto y 0,25 el de uno aleatorio.

En estos índices se ha tomado como principio el uso de medidas de posición- si se usaran criterios de dispersión se puede usar el índice de Armitage

$$I_A^2 = \frac{n}{n-1} \left[p_{11} + p_{12} - (p_{11} - p_{22})^2 \right] \quad (5)$$

En el caso de desacuerdo perfecto este es cero y lo fuese el acuerdo toma el valor $n/n-1$. Armitage-Blendys-Smyl (1966) reescalán este índice proponiendo

$$I'_A = \frac{p_{11} + p_{12} - (p_{11} - p_{22})^2}{1 - (p - q)^2} \quad (6)$$

Uno de los más populares índices es de Kruskal (1954)

$$I_K = \lambda_r = \frac{(p_{11} + p_{22}) - q}{p} = \frac{2p_{11} - (p_{12} + p_{21})}{2p_{11} + (p_{12} + p_{21})} = 2I_D - 1 \quad (7)$$

Este se mueve en [-1,1]. Si el acuerdo es perfecto $\lambda_r = 1$ si es perfecto. Si $n_{11} = 0$ el índice toma el valor -1 lo que es un severo problema al hacer ciertos análisis si el objetivo del evaluador está en los objetos calificados de positivos por al menos un experto.

El índice Scott (1955) es

$$I_S = \frac{4(p_{11}p_{22} - p_{12}p_{21}) - (p_{12} - p_{21})^2}{(p_{1+} + p_{+1})(p_{2+} + p_{+2})} \quad (8)$$

Podemos considerar que los expertos hacen su valoración usando un mecanismo aleatorio independiente. Esta hipótesis fija que

$$\Pr \{t \in C_i \text{ para } \text{experto1} \text{ y } t \in C_j \text{ para } \text{experto2}\} = \pi_{ij} = \pi_{i+} \pi_{+j}$$

donde π_{i+} es la probabilidad marginal de que el experto1 ubique un objeto en C_i y π_{+j} lo es de que lo haga experto2 en C_j . Al analizar un índice I^* sin usamos este modelo al calcular su esperanza condicional:

$$E(I^* | \pi_{ij} = \pi_{i+} \pi_{+j} \forall i, j = 1, \dots, k) = I_0 \quad (9)$$

entonces podemos considerar la descomposición

$$p_{ij} = p_{i+} p_{+j}$$

Como $1 - I_0$ es el máximo de las diferencias obtenibles respecto a una clasificación aleatoria una medida del acuerdo es el índice tipificado

$$M(I^*) = \frac{I^* - I_0}{1 - I_0} \quad (10)$$

Cuando el acuerdo observado es mayor que el atribuible a un fenómeno aleatorio $M(I^*) > 0$, si es menor $M(I^*) < 0$ y será igual a cero si son equivalentes. Es conveniente acotar el hecho de que si $I_0 = 0,5$ entonces $\text{Min} \{M(I^*)\} = -1$. Por tanto este es un índice que es deducido al aceptar la existencia de un modelo superpoblacional que fija que

$$E(n_{ij}) = E(n_{i+}) E(n_{+j})$$

Note que esta esperanza es calculada respecto al modelo utilizado para modelar el consenso y no tiene nada que ver con el diseño muestral utilizado.

El índice “kappa” es el mas utilizado y es definido por

$$\kappa = \frac{\sum_{i=1}^k p_{ii} - \sum_{i=1}^k p_{i+} p_{+i}}{1 - \sum_{i=1}^k p_{i+} p_{+i}} \quad (11)$$

Este fue propuesto por Cohen (1960).

Utilizando este marco podemos decir que el κ -índice es un estimador ingenuo de

$$K_{pob} = \frac{\sum_{i=1}^k \pi_{ii} - \sum_{i=1}^k \pi_{i+} \pi_{+i}}{1 - \sum_{i=1}^k \pi_{i+} \pi_{+i}} \quad (12)$$

En general el número de posible calificaciones de un objeto pueden ser $k \geq 2$. En particular no es recomendable hacer preguntas con más de 5 alternativas. Pero esta es una recomendación asociada a la práctica pues para valores relativamente “grandes” de k los entrevistados tienden a verse desestimulados a ser “exactos”.

Podemos analizar la esperanza de los índices más populares bajo este modelo de aleatoriedad.

Para (1)

$$E(I_D | \pi_{ij} = \pi_{i+} \pi_{+j} \forall i, j = 1, \dots, k) = \frac{P_{1+} P_{+1}}{p}$$

y el índice tipificado es

$$M(I_D) = \frac{2(p_{11} p_{22} - p_{12} p_{21})}{p_{1+} + p_{+2} + p_{+1} + p_{+2}}$$

que es igual a (3). Para I_{RG} tenemos

$$E(I_{RG} | \pi_{ij} = \pi_{i+} \pi_{+j} \forall i, j = 1, \dots, k) = \frac{P_{1+} P_{+1}}{p} + \frac{P_{2+} P_{+2}}{q}$$

por lo que nuevamente obtenemos que $M(I_{RG}) = \kappa$

Similarmente a partir de que

$$E(I_A | \pi_{ij} = \pi_{i+} \pi_{+j} \forall i, j = 1, \dots, k) = \frac{P_{1+} P_{+2}}{1 - (p - q)^2} + \frac{P_{1+} P_{+1}}{1 - (p - q)^2}$$

el índice tipificado correspondiente

$$M(I_A) = \kappa$$

Fleiss (1975) destacó el hecho de que el índice κ caracteriza la clase de índices

$$\mathfrak{I} = \left\{ I \mid I = \frac{I^* - I_{\Omega}}{\text{Max}(I^* - I_{\Omega})} \right\}$$

donde $I_{\Omega} = E(I^* | \Omega)$.

Estos hechos soportan la popularidad del índice propuesto por Cohen aunque en la práctica se utilice muchas veces uno de los introducidos en la sección 2. Razones históricas, entre las que incluyen la costumbre, hacen que en ciertas áreas se utilicen uno de ellos en vez del κ -índice.

4. ESTUDIO DEL CONSENSO ENTRE LA CLASIFICACIÓN REALIZADA POR LOS MÉDICOS DEL SISTEMA DE SALUD CONTRA LA REGLA DE ORO (GOLD STANDARD)

4.1 Resultados Preliminares

Los datos de la población de sospechosos de dengue en el estado de Guerrero, durante el 2006, se organizaron por municipio, Región o Jurisdicción Sanitaria y a nivel estatal sin considerar Municipios o regiones. El objetivo de esta sección es establecer cuan consensuadas fueron las clasificaciones de los especialistas respecto al estándar, aceptando como estándar el resultado del laboratorio (ResultLab), a partir del cual, un caso “sospechoso” pasa a la categoría de “confirmado”. El resultado del laboratorio se comparará con la clasificación realizada a nivel de municipio, región o jurisdicción sanitaria y a nivel estatal, por distintos especialistas a nivel de unidad tratante (ExpUT) y especialistas del laboratorio estatal de salud (ExpLab).

Tabla 4.1: Resumen de la clasificación por los ExpUT y ExpLab

ExpUT	ExpLab		Total
	DC	DH	
DC	8443	298	8741
	96.59	3.41	100
	98.69	38.45	93.69
DH	112	477	589
	19.02	80.98	100
	1.31	61.55	6.31
Total	8555	775	9330
	91.69	8.31	100
	100	100	100

La clasificación de las personas que acudieron a consulta por los médicos y el laboratorio aparece en la Tabla 4.1. En ella presentamos el resumen en la clasificación hecha por los expertos de UT y del Laboratorio, en la cual se observa que de 8741 el ExpUT clasificó como DC, coincidió con el ExpLab en 8443 que representa el 96.59% de los casos y 298 le fueron reclasificados como DH, esto representa solo el 3.41%. De los 589 clasificados como DH por el ExpUT, existe coincidencia en la clasificación con el ExpLab en 477 casos, es decir, en el 80.98% de los casos y solo 112 (19.02%) le son reclasificados como DC, entonces, el ExpLab clasifica 8555 casos (91.69%) como DC y 775 (8.31%) como DH.

Tabla 4.2: Resumen de la clasificación por los ExpUT y ResultLab

ExpUT	ResultadoLab		Total
	Positivo	Negativo	
DC	1658	6627	8285
	20.01	79.99	100
	82.28	97.38	93.93
DH	357	178	535
	66.73	33.27	100
	17.72	2.62	6.07
Total	2015	6805	8820
	22.85	77.15	100

	100	100	100
--	-----	-----	-----

En las tablas 4.2 y 4.3 se tienen las comparaciones de los EUT y el experto del laboratorio con la regla de oro, es decir, el resultado del examen en el laboratorio.

ExpLab	ResultadoLab		Total
	Positivo	Negativo	
DC	1681	6874	8555
	19.65	80.35	100
	87.42	92.80	91.69
DH	242	533	775
	31.23	68.77	100
	12.58	7.20	8.31
Total	1923	7407	9330
	20.61	79.39	100
	100	100	100

Tabla 4.3: Resumen de la clasificación por los ExpLab y ResultLab

En la tabla 4.1 se tenía que el ExpUT clasificaba 8741 casos como DC y en la tabla 4.2 solo aparecen 8285 de ellos, 456 casos menos. La razón de esta disminución se debe a que la muestra de suero que se envió al laboratorio era inadecuada para hacer la clasificación a partir de ella. De acuerdo a estos resultados, de los 8285 casos válidos de DC, solo el 20.01% de los casos que reportó como DC, resultaron en DC y el 79.99% resultaron negativos. De los 535 casos válidos de DH clasificados por el ExpUT, el 66.73% resultaron positivos y el 33.27% resultaron negativos. Se tiene una positividad general de solo el 22.85%, es decir, 77.15% de los casos fueron mal clasificados por el ExpUT.

En la tabla 4.3 se observa que el experto del laboratorio clasifica correctamente 19.65% de los casos de DC y 80.35% es mal clasificado. De los 775 casos que clasifica como DH, el 31.23% resultó en positivo, mientras que el 79.39% resultó en negativo, en general se tiene una positividad en la clasificación del ExpLab de 20.61%

Pasemos a revisar el caso general de clasificación y el resultado IgM. Esto aparece en la tabla 4.4 en la que se observa que existe una positividad general en la clasificación de dengue, sin distinguir entre clásico y hemorrágico, de 21.55%, 17.64% para DC y 71.83% para DH, es decir, se clasifica mejor el dengue tipo hemorrágico.

Tabla 4.4: Resumen de la clasificación general y resultado IgM

ClasificGral.	Resultado IgM		Total
	Negativo	Positivo	
DC	7850	1681	9531
	82.36	17.64	100
	97.41	75.93	92.78
DH	209	533	742
	28.17	71.83	100
	2.59	24.07	7.22
Total	8059	2214	10273
	78.45	21.55	100

	100	100	100
--	-----	-----	-----

4.2 Resultados a nivel de Estado

La tabla 4.5 brinda los resultados obtenidos en el estado para los diversos índices de consenso

Tabla 4.5 Comparación de Clasificación de dengue Vs Resultado laboratorio para el Estado de Guerrero

Clasificación	Kappa	I_D	I'_D	I_{RG}	I'_{RG}	I_A^2	I'_A	I_K	I_S
General	-0.098	0.286	0.047	0.334	0.083	0.907	0.926	-0.428	-0.666
Dengue Clásico	0.030	0.161	0.300	0.461	0.115	0.065	0.066	-0.678	-0.539
Dengue Hemorrágico	0.826	0.989	0.836	1.825	0.456	0.160	0.690	0.978	0.825

Los especialistas del laboratorio consideran como los índices más significativos para su estudio los de Dice y Rogot-Goldberg. Se observa que el consenso en la clasificación de dengue es muy bueno en el dengue tipo hemorrágico para los índices más notables. κ , I_D , I'_D , I_{RG} y I'_{RG} están cerca del óptimo que es el valor 1 para los tres primeros, de 2 para el segundo y de 0,5 para el último. Al hacer un análisis similar para el dengue tipo clásico resulta que el consenso es muy malo.

Dada la consistencia entre los resultados desarrollamos los análisis a otros niveles usando solo los de Dice y el kappa.

Los resultados en la Tabla 4.6e permite estudiar el consenso entre las clasificaciones de dengue clásico y hemorrágico hechas por los especialistas en las unidades tratantes, expertos del laboratorio y resultado de laboratorio.

Tabla 4.6: Clasificación general de dengue clásico y hemorrágico

Clasificadores	I_D	I'_D	Kappa
ExpUT-ExpLab	0.9763(B)	0.6994(R)	0.6762(R)
ExpUT-ResultLab	0.3219(N)	0.0485(N)	-0.0721(N)
ExpUT-ResultLabDC-DH	0.9198(B)	0.7107(R)	0.6341(R)
ExpLab-ResultLab	0.2862(N)	0.0475(N)	-0.0977(N)
ExpLab-ResultLabDC-DH	1.0000(B)	1.0000(B)	1.0000(B)

Como se ve para el caso general de clasificación como DC o DH es necesario fijar las características de ese índice respecto al modelo considerado para la comparación con una regla de oro. En este caso el resultado del laboratorio representa esta regla. El valor del Índice Kappa es de aproximadamente -0.098, este debe variar entre 0 y 1 si el azar (la casualidad) modela el consenso entre la valoración del médico y la prueba de laboratorio. Si el valor es próximo a cero pero positivo se considera que fue el azar lo que caracteriza los acuerdos observados. En el caso de ser cero esto indica que no hay un método para predecir el acuerdo. Si es menor que cero lleva a afirmar que la falta de consenso observada, como es este caso, entre las clasificaciones a nivel estatal, que es negativa, es debida a una causa no explicada por la aleatoriedad. Esto sugiere que son mal aplicadas sistemáticamente las metodologías de clasificación de los síntomas.

Estos resultados indican que es alto el consenso entre unidad tratante y laboratorio estatal en clasificar positivos como positivos (I_D), pero que es regular para detectar los negativos (I'_D) y en general (kappa). Para el caso general de clasificación como DC o DH es necesario fijar las características de ese índice respecto al modelo considerado para la comparación con una regla de oro. En este caso el resultado del laboratorio representa esta regla. El valor del Índice Kappa es de -0.098 para el experto del laboratorio y de -0.0721 para el experto de unidad tratante, este debe variar entre 0 y 1 si el azar (la casualidad) modela el consenso entre la valoración del médico y la prueba de laboratorio. Si el valor es próximo a cero pero positivo se considera que fue el azar lo que caracteriza los acuerdos observados. En el caso de ser cero esto indica que no hay un método para predecir el acuerdo. Este análisis nos lleva a la valoración cualitativa que aparece entre paréntesis en las tablas 4.6 y 4.7. Es de notar el bajo nivel de consenso en los pares ExpUT-ResultLab y

ExLab-ResultLab en los cuales se puede afirmar que no hay consenso bajo ninguno de los criterios analizados.

Los resultados por tipo de dengue se muestran en las tablas 4.6 y 4.7.

Tabla 4.6 Clasificación de dengue clásico

Clasificadores	I _D	I _D '	Kappa
ExpUT-ExpLab	0.9757(B)	0.6943(R)	0.6706(R)
ExpUT-ResultLab	0.2976(N)	0.1248(N)	0.0132(N)
ExpLab-ResultLab	0.2999 (N)	0.1604(N)	0.0303(N)

En el dengue tipo clásico, tabla 4.6, el consenso en la clasificación es malo tanto para el experto de laboratorio como el experto en la unidad tratante (Kappa = 0.0303 y 0.0132, respectivamente) pero, debemos observar que el consenso en la clasificación de ambos expertos sigue siendo de 0.6706, es decir, no existen diferencias significativas en la clasificación que hacen ambos expertos. Los resultados para ExpUT-ResultLab y ExpLab-ResultLab son malos para todos los criterios de consenso.

Tabla 4.7 Clasificación de dengue Hemorrágico.

Clasificadores	I _D	I _D '	Kappa
ExpUT-ExpLab	0.6990(R)	0.9772(B)	0.6761(R)
ExpUT-ResultLab	0.5955(R)	0.9744(B)	0.570(R)
ExpLab-ResultLab	0.8362 (A)	0.9891(B)	0.8260(A)

Se observa en la tabla 4.7 que el consenso en la clasificación del tipo de dengue es bueno en el dengue tipo hemorrágico para ExpUT-ExpLab y ExpLab-ResultLab, dado que en ningún caso los índices están por debajo de 0.5. Sin embargo es malo para ExpUt-ResultLab para el consenso en detecta lo positivos y en general. Observemos, además que se tiene un consenso moderado en la clasificación de dengue entre los expertos de unidad tratante y experto de laboratorio (Kappa =0.6761).

4.3 Resultados por Jurisdicción Sanitaria y municipio

Un análisis a nivel de Jurisdicción Sanitaria o región fue llevada a cabo también. Cada jurisdicción sanitaria coincide con cada una de las siete regiones del estado, de manera que los resultados pueden presentarse para cualquiera de los dos casos. La tabla 4.8 brinda los resultados de los índices de consenso.

Se observa que el consenso en la clasificación de dengue que hacen los expertos del laboratorio, dentro de jurisdicciones sanitarias o regiones, fue muy bueno en el dengue tipo hemorrágico. Los valores de Kappa fluctúan entre 0.763 y 0.922. Si eliminamos la Tierra Caliente, donde no hubo casos. Fue la mejor la clasificación realizada por la Jurisdicción sanitaria 4 de la montaña de Guerrero. El consenso para los expertos de unidad tratante, es inferior a la clasificación que hacen los expertos del laboratorio. Note que los únicos valores del índice kappa altos son los que corresponden a la región norte y la región Acapulco, con valores de 0.603 y 0.562, respectivamente. Para el dengue tipo clásico el consenso en la clasificación fue muy malo. Tomando el índice Kappa se tiene que este varió entre 0.001 y 0.194. Esto es soportado por el hecho de que un valor cercano a cero es evidencia de falta de consenso en la clasificación. Podemos observar, además, que otros índices nos indican una buena clasificación dentro de cada jurisdicción para el dengue hemorrágico, y muy mala, para el dengue clásico.

El análisis a nivel de municipio permitió establecer que nuevamente el consenso en la clasificación de dengue clásico es malo dentro de todos los municipios. Son los mejores los de Acapulco Juárez, Apaxtla de Castrejón y José Azueta, los que presentamos en la tabla 4.9.

4.4 Estudio de los errores e intervalos de confianza

Un procedimiento estadístico para validar los resultados, consiste en hacer un remuestreo para computar los errores estándar de la distribución de medias de los índices. Esto nos permite estimar un intervalo confidencial. Utilizamos el método de Bootstrap, ver por ejemplo Hole-Fleiss (1993).

Tabla 4.8 Clasificación de dengue Clásico y Hemorrágico por jurisdicciones sanitarias

Jurisdicción	Clasificadores	Dengue Clásico			Dengue Hemorrágico		
		I _D	I _D '	Kappa	I _D	I _D '	Kappa
1: T. caliente	ExpUT-ExpLab	0.9891(B)	0.0000(N)	-0.0043(N)	0.0000(N)	0.9988(B)	0.0000(N)
	ExpUT-ResultLab	0.1651(M)	0.0055(N)	0.0005(N)	0.0000(N)	0.9988(B)	0.0000(N)
	ExpLab-ResultLab	0.1447(M)	0.0376(N)	0.0030(N)	0.0000(N)	0.6667(R)	0.0000(N)
2: Norte	ExpUT-ExpLab	0.9918(B)	0.7188(R)	0.7107(R)	0.7188(R)	0.9918(B)	0.7107(A)
	ExpUT-ResultLab	0.2457(M)	0.0495(N)	0.0017(N)	0.6122(R)	0.9911(B)	0.6035(R)
	ExpLab-ResultLab	0.2500(M)	0.0746(N)	0.0110(N)	0.7945(R)	0.9941(B)	0.7889((R)
3: Centro	ExpUT-ExpLab	0.9908(B)	0.5313(R)	0.5222(R)	0.5313(R)	0.9908(B)	0.5222(R)
	ExpUT-ResultLab	0.1518(M)	0.0364(N)	-0.0058(N)	0.3529(M)	0.9894(B)	0.3436(M)
	ExpLab-ResultLab	0.1545(M)	0.0360(N)	0.0031(N)	0.7660(R)	0.9968(B)	0.7629(R)
4: Montaña	ExpUT-ExpLab	0.9904(B)	0.1429(M)	0.1335(M)	0.1481(M)	0.9908(B)	0.1391(M)
	ExpUT-ResultLab	0.4810(M)	0.0120(N)	-0.0061(N)	0.0870(N)	0.9912(B)	0.0782(N)
	ExpLab-ResultLab	0.4793(M)	0.0334(N)	0.0106(N)	0.9231(B)	0.9992(B)	0.9223(B)
5: C. Grande	ExpUT-ExpLab	0.9879(B)	0.3214(M)	0.3106(M)	0.3214(M)	0.9879(B)	0.3106(M)
	ExpUT-ResultLab	0.2989(M)	0.0217(N)	-0.0013(N)	0.1463(M)	0.9885(B)	0.1350(M)
	ExpLab-ResultLab	0.2704(M)	0.0431(N)	0.0069(N)	0.7692(R)	0.9964(B)	0.7659(R)
6: C. Chica	ExpUT-ExpLab	0.9907(B)	0.1818(M)	0.1732(M)	0.1818(M)	0.9907(B)	0.1732(M)
	ExpUT-ResultLab	0.1264(M)	0.0250(N)	-0.0047(N)	0.1111(M)	0.9913(B)	0.1051(M)
	ExpLab-ResultLab	0.1270(M)	0.0166(N)	0.0011(N)	0.7692(R)	0.9985(B)	0.7678(R)
7: Acapulco	ExpUT-ExpLab	0.9177(B)	0.7498(R)	0.6700	0.7504(R)	0.9180(B)	0.6710(R)
	ExpUT-ResultLab	0.4140(M)	0.4008(M)	0.1064(M)	0.6543(R)	0.9080(B)	0.5624(R)
	ExpLab-ResultLab	0.4749(M)	0.5135(R)	0.1943(M)	0.8449(A)	0.9534(B)	0.7998(R)

Se tomaron 50 muestras aleatorias de tamaño 500 y se estimaron las medias y varianzas de cada uno de los índices, además del error estándar y un intervalo de confianza para la media de los índices. Los resultados estadísticos fueron los dados en las tablas 4.10 - 4.12.

Tabla 4.9 Clasificación de dengue Clásico y Hemorrágico por Municipio

Municipio	Clasificadores	Dengue Clásico			Dengue Hemorrágico		
		I _D	I _D '	Kappa	I _D	I _D '	Kappa
Acapulco de Juárez	ExpUT-ExpLab	0.9166(B)	0.7437(R)	0.6630(R)	0.7444(R)	0.9169(B)	0.6640(R)
	ExpUT-ResultLab	0.4092(M)	0.3941(M)	0.1021(M)	0.6542(R)	0.9094(B)	0.5636(R)
	ExpLab-ResultLab	0.4679(M)	0.5080(M)	0.1884(M)	0.8450(A)	0.9541(B)	0.8005(A)
Apaxtla	ExpUT-ExpLab	0.9792(B)	0.6667(R)	0.6458(R)	0.6667(R)	0.9792(B)	0.6458(R)
	ExpUT-ResultLab	0.6849(R)	0.1481(M)	0.0464(N)	0.4000(M)	0.9684(B)	0.3697(M)
	ExpLab-ResultLab	0.7123(R)	0.2759(M)	0.1626(M)	0.8571(A)	0.9895(B)	0.8468(A)
José Azueta	ExpUT-ExpLab	0.9528(B)	0.3478(M)	0.3101(M)	0.3478(M)	0.9528(B)	0.3101(M)
	ExpUT-ResultLab	0.4965(M)	0.0928(N)	0.0203(N)	0.1714(M)	0.9539(B)	0.1285(M)
	ExpLab-ResultLab	0.5188(R)	0.2491(M)	0.0950(N)	0.8000(A)	0.9824(B)	0.7831(R)

En la Tabla 4.10 se confirma que el índice kappa para el caso general fue negativo, lo que indica que la clasificación general de dengue es mala y el signo negativo es debido a una causa no explicada por la aleatoriedad. Esto confirma la hipótesis de que son mal aplicadas las metodologías de clasificación de los síntomas.

En el caso de la tabla 4.11 con el dengue clásico, se confirma que es mala la clasificación hecha por los expertos, toda vez que el índice kappa es muy próximo a cero (0.030), lo cual indica falta de consenso en la clasificación. Explicación similar se hace para los otros índices

Tabla 4.10. Comparación de índices para la clasificación Dengue Clásico Vs Dengue Hemorrágico.

Índices	Media	Desviación típica	Error típico	Intervalo de conf. para la media al 95%	
				Límite inferior	Límite superior
kappa	-0.0927	0.0252	0.0036	-0.0998	-0.0855
ID	0.2937	0.0274	0.0039	0.2860	0.3015
ID'	0.0486	0.0141	0.0020	0.0446	0.0526

Los resultado de la Tabla 4.12 confirma que la clasificación hecha por los expertos para el dengue tipo hemorrágico es buena (kappa = 0.82) con un error típico muy cercano a cero (0.0089).

5. CONCLUSIONES

LA existencia de pocos casos confirmados con relación a los casos sospechosos de dengue, se traduce en un alto costo para identificar los casos reales. Por ello podemos concluir que los médicos tratantes no clasifican adecuadamente los casos de dengue clásico, aunque sí clasifican mejor el hemorrágico. Esto puede ser motivado por el temor a equivocarse en la clasificación, toda vez que lo hacen fundamentalmente mediante el reconocimiento clínico de la enfermedad. Para ello se basan en su experiencia con la enfermedad y la aplicación del manual de dengue, el cual obliga a que hagan un reporte escrito, tomen una muestra de suero y sea enviada al Laboratorio Estatal de Salud para su estudio y en su caso, confirmación del diagnóstico.

Tabla 4.11 Comparación de índices para la clasificación Dengue Clásico Vs No Dengue Clásico.

Índices	Media	Desviación típica	Error típico	Intervalo de conf. para la media al 95%	
				Límite inferior	Límite superior
kappa	0.03059	0.00561	0.00079	0.02900	0.03217
ID	0.30696	0.02759	0.00390	0.29916	0.31476
ID'	0.15753	0.02461	0.00348	0.15057	0.16449

La identificación de los errores en las etapas del proceso de clasificación de la enfermedad, nos permitirá orientar en que sentido deben mejorarse las actividades de capacitación dentro de la atención primaria.

Tabla 4.12 Comparación de índices para la clasificación Dengue Hemorrágico Vs No Dengue Hemorrágico.

Índices	Media	Desviación típica	Error típico	Intervalo de conf. para la media al 95%	
				Límite inferior	Límite superior
kappa	0.8168	0.0627	0.0089	0.7990	0.8345
ID	0.8431	0.1472	0.0208	0.8015	0.8848
ID'	0.9781	0.0767	0.0108	0.9564	0.9998

REFERENCIAS

[1] AGRESTI, A. (1988): A model for agreement between ratings on the ordinal scale. **Biometrics**, 44, 539-548.

- [2] AGRESTI, A. and J. LANG (1993): Quasi-symmetric L class models with application to rater agreement. **Biometrics**, 49, 131-139.
- [3] ARMITAGE, P.; L.M. BLENDYS and H.C. SMYLLIE (1966): The measurement of observer disagreement in the recording of signs. **J. Royal Sta. Soc. Series A**, 129, 96-109.
- [4] BLECH, D.A. and H.C. KRAEMER (1989): 2x2 kappa coefficient: measures of agreement and assessment. **Biometrics**, 45, 269-287.
- [5] BOUZA, C.N. (1987): A ratio estimator of the kappa index of agreement between two observers. **Biometrical J.**, 29, 1011-1015.
- [6] CICHETTI, D. V. and A.R. FEINSEIN (1990): High agreement but low kappa: II. Resolving the paradoxes. **J. of Clinical Epidemiology**. 43, 551-558.
- [7] COHEN, J. (1960): A coefficient of agreement for nominal scales. **Educational and Psychological Measurement**. 20, 37-46.
- [8] COHEN, J. (1968): Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. **Psychological Bull.**, 70, 213-220.
- [9] DONNER, A.; M. ELIASZUO and N. KLAR (1996): Testing the homogeneity of kappa stratified. **Biometrics**, 52, 176-183.
- [10] DONNER, A.; M. ELIASZUO and N. KLAR (1992): A goodness of fit approach to inference procedures for the kappa stratified: confidence interval construction, significance tests and sample size estimates. **Statistics in Medicine**, 11, 1511-1519.
- [11] FLEISS, J. (1975): Measuring agreement between two judges on the presence in the absence of a trait. **Biometrics**, 31, 651-659.
- [12] FLEISS, J. (1971): Measuring a nominal scale agreement among many raters. **Psychological Bull.** 76, 378-382.
- [13] FLEISS, J. (1965): Estimating the accuracy of dichotomous judgements. **Psychometrika**, 30, 469-479.
- [14] FORMAN, A. (1993): Fixed latent class models for the analysis of sets of 2-way contingency tables. **Biometrics**, 49, 511-521
- [15] GÓMEZ H., RAMOS B. TAPIA R. (1995): El riesgo de transmisión del dengue: un espacio para la estratificación. **Salud. Pública. Mex.** ; 37:supl:88-97.
- [16] HOLE, C.A. and J.L. FLEISS (1993): Interval estimates under 2 study designs for kappa with binary classification. **Biometrics**, 49, 523-524.
- [17] KRAEMER, H.C. (1992): Measurement of reliability for categorical data in medical research. **Statistical Methods in Medical Res.** 1, 183-200.
- [18] KRAUTH, J. (1984): "Multivariate Behandlungstabilität. Klinischer Skalen", **Psychol. Beitr.** 23, 438-457.
- [19] LAU, T.S. (1993): "Higher order kappa type statistics for the dichotomous attribute in multiple ratings", **Biometrics**, 49, 535-542.
- [20] LEE, J.J. and Z.N. TU (1994): A better confidence of interval for the kappa on measuring agreement between two raters with binary outcome. **J. of Computational and Graphical Stat.** 3,301-321.

- [21] LIPSITZ, S.R.; N.M. LAIRD and T.A. BREMEN (1994): Simple moments estimators of the kappa coefficient and its variance. **Applied Stat.** 43, 309-323.
- [22] OMS. (2006): Dengue and Dengue Haemorrhagic fever. FACT Sheet No. 117. Revised April 2006. <http://rhone.b3e.jussieu.fr/DegueNet/pages/f-info/html>
- [23] OPS. (1995): Dengue y Dengue Hemorrágico en las Américas: guías para su prevención y control . **Organización Panamericana de la Salud. Publicación Científicas 548**, Washington , DC. ,
- [24] ROGOT, E. and I.D. GOLBERG (1966): A proposed index for measuring agreement in test-retest studies. **J. Chronic Dis.**19, 991-1006.
- [25] ROSNER, B. (1982): Statistical methods in ophthalmology: an adjustment for the intraclass correlation between eyes. **Biometrics**, 38, 899-905.
- [26] SCHOUTEN, H.J.A. (1986): The kappa coefficient of agreement among observers. **Psychometrika**, 51,453- 466.
- [27] SCOTT, W.A. (1955): Reliability of content analysis: the case of nominal scale coding. **Public Opinion Quarterly**.
- [28] SHOUKRI, M.M.; S.W. MARTIN and I.U.H. MIAAN (1995): Maximum likelihood estimation of the kappa coefficient from models of matched binary responses. **Statistics in Medicine**, 124, 83-99.
- [29] SSM: (2006): **Manual para la Vigilancia, Diagnóstico, Prevención y Control del Dengue**. Secretaría de Salud. México. , México, DF.
- [30] TANNER, M.A. and M.A. YOUNG (1985): Modeling agreement among raters. **J. Amer. Stat. Ass.** 80, 175-180.
- [31] UBERSAX, J.C. (1993): Statistical modeling of expert ratings in medical treatment appropriateness. **J. Amer. Stat. Ass.** 88, 321-427.
- [32] VON EYE, A. and S. SORENSEN (1991): Model chance when measuring interrater agreement with kappa. **Biometrical J.**, 33,781-787.

Received February 2007

Revised October 2007