

Some Thoughts on How to Interpret the Results of Studies on Diagnostic Tests: The DINO Study

To the Editor:

We read the study by Huerta et al¹ with considerable interest. We feel that validating the precision of diagnostic questionnaires is important from an epidemiological point of view because of their advantages (comfort, time, and low cost) compared with biometric and clinical methods, but we would like to make some comments on the study in order to contribute to the scientific debate and the exchange of ideas:

1. Although the authors state that “the interviewers are well qualified,” we believe that the use of a procedure (for example, the Kappa index) to establish concordance among them would have helped make the study more reliable.²

2. As for the prevalence rates for the different diseases, the authors state that “all of the reported prevalence rates are lower than the estimates given by the reference model.” However, in the case of diabetes, the 95% confidence intervals (CI) for the specific model and questionnaire overlap, so it cannot be stated that the specific resulting measurements are different from each other.

3. The statement in the article, based on the Kappa index calculation, is that “we have reached a degree of concordance between the diagnosis in question and the proper reference model for diabetes, with moderate AHT and low hyperlipaemia.” This is not correct; although a lesser-known use of the Kappa index is evaluating discrepancies in diagnostic test, the indispensable requirement for using the index in this way is for the interpretation of the tests to be subject to the human factor (ie, radiographies, electrocardiograms, anatomical pathology samples, etc),³ which does not occur in this case, since the measurements for the reference model are biometric, calculated by automatic devices, calibrated, and therefore each measurement is not subject to human variability. Furthermore, the declared diagnosis (the sum of the data contributed by several researchers and not previously subjected to any test of concordance among observers) is compared with data from the test of reference (which is objective from the viewpoint of variability for each ascertainment), and for this reason that comparison cannot be made. In addition, the sensitivities for all three diseases are below 70%. We know from mathematical models that with sensitivities below 70% the Kappa index will necessarily be <0.4 because low sensitivity itself leads to discordance.²

4. Given that the purpose of the study is to evaluate the usefulness of the diagnostic questionnaire in question, using diagnoses from biometric tests as a standard, we believe that in addition to calculating sensitivity, specificity and predictive values, the coefficients of positive and negative probability (CPP and CPN respectively) should have been calculated as well. This is because the sensitivity and specificity values, despite completely defining validity (the degree to which a test measures what it is supposed to measure) for the diagnostic test, have the disadvantage of not providing relevant information when it comes to making a clinical decision when faced with a certain test result. The CPP and CPN express a unified summary of sensitivity and specificity, and therefore do not depend on the disease prevalence in each location and permit comparison between different studies, unlike predictive values, which are only valid for the location at which they were calculated.^{4,5} In this study, the CPP and CPN with 95% CI for each disease (CPP for diabetes, 161.46 [72.21-361.06]; AHT, 15.41 [10.84-21.9]; hyperlipaemia, 10.98 [7.07-17.05]; CPN for diabetes, 0.30 [0.24-0.38]; AHT, 0.52 [0.48-0.57]; hyperlipaemia, 0.68 [0.64-0.71]) indicate that the questionnaire is useful for diagnosing diabetes and hypertension (relevant CPP >10 with 95% CI) but they are not useful for ruling them out (CPN are not <0.1). With regard to hyperlipaemia, it will be necessary to design more studies (coefficients of probability are statistically significant, but not clinically conclusive).

5. When we refer to diagnostic test studies, parameters $\geq 80\%$ are considered to be optimum or high, and for this reason, we cannot claim the sensitivities of the study as such.

To conclude, we must be careful and methodical with all details involved in a research study in order to contribute efficiently to its comprehensibility and usefulness to the researchers and clinics for which it is intended.

Álvaro Campillo-Soto, Ramón Lirón-Ruiz,
Juan Gervasio Martín-Lorenzo,
and José Luis Aguayo-Albasini

Servicio de Cirugía General y Digestiva, Hospital General Universitario
J.M. Morales Meseguer, Murcia, Spain

REFERENCES

1. Huerta JM, Tormo MJ, Egea-Caparrós JM, Ortolá-Devesa JB, Navarro C. Validez del diagnóstico referido de diabetes, hipertensión e hiperlipemia en población adulta española. Resultados del estudio DINO. *Rev Esp Cardiol*. 2009;62:143-52.
2. Saturno PJ. La construcción de criterios para evaluar la calidad. In: Saturno PJ, editor. *Evaluación y mejora de la calidad en Servicios de Salud. Conceptos y métodos*. Murcia: Consejería de Sanidad y Consumo de la Región de Murcia; 2000. p. 111-32.

3. Sackett DL. El examen clínico. In: Sackett DL, Haynes RB, Guyatt GH, Tugwell A, editors. Epidemiología clínica. Ciencia básica para la medicina clínica. Boston: Panamericana; 1997. p. 34-61.
4. Escrig-Sos J, Martínez-Ramos D, Miralles-Tena JM. Pruebas diagnósticas: nociones básicas para su correcta interpretación y uso. Cir Esp. 2006;79:267-73.
5. Díaz Guzmán J. Investigación clínica. Diagnóstico. In: Diploma Superior en Metodología de la Investigación. 3.^a ed. Madrid: Instituto de Salud Carlos III; 2008. p. 1-55.

Response

To the Editor:

It was with interest that we read and discussed the remarks Campillo et al made with regard to our article.¹ We would like to thank the authors for their comments, particularly those referring to the Kappa index, which assist us in understanding and contextualising our objectives.

The authors mention the overlapping confidence intervals for the resulting and validated diabetes prevalence rates. On this point, we do not agree with their conclusions due to the following: *a*) the overlap is only marginal; *b*) each confidence interval excludes the alternative score; and *c*) each difference is statistically significant (McNemar's Test, $P < 10^{-6}$).

With respect to the observation on the correctness of using the Kappa index as an agreement measurement between the information collected by the questionnaire and the corresponding biometric models, it is true that a study with these characteristics is not the normal setting for applying the index. However the κ value is not the most important result in this analysis, although excessive attention may well have been drawn to it in the text. Rather, the most important topic for discussion is the different approach used in a population-based cardiovascular study like this one (based on

interviews and on a small, highly select number of biometric tests performed a single time to reach a diagnosis, which are collectively valid although they may be individually inexact) and our focus on clinical use in a hospital setting with the possibility of requesting and repeating a high number of tests. Our objective was to evaluate the reliability of these diagnoses based on polled individuals' responses within the context of cardiovascular risk factor studies, which are typically included in transversal population-based and follow-up studies; furthermore, these studies examine the confusion variables necessary for adjustment, while the diagnosis is often of no interest. This is a common approach on the pages of the *Revista Española de Cardiología*, which maintains a permanent "Epidemiology and Prevention" section in its publication.

In keeping with our approach, the absolute validity of the questionnaire is defined by sensitivity and specificity values, while other indexes, whether apparent or not, may provide complementary information of varying degrees of interest for epidemiologically or clinically-minded readers. We feel that one of the study's strong points is having contributed enough data to be able to calculate alternative indexes.

We thank Campillo et al for their remarks which allowed us to clarify a distinction that might not otherwise have been observed.

José M. Huerta, María José Tormo, and Carmen Navarro
Servicio de Epidemiología, Consejería de Sanidad
y Consumo de la Región de Murcia, Murcia, CIBER Epidemiología y
Salud Pública (CIBERESP), Spain

REFERENCES

1. Huerta JM, Tormo MJ, Egea-Caparrós JM, Ortolá-Devesa JB, Carmen Navarro. Validez del diagnóstico referido de diabetes, hipertensión e hiperlipidemia en población adulta española. Resultados del estudio DINO. Rev Esp Cardiol. 2009;62:15-22.