Neuropsychological Diagnostic Accuracy:

Factors That Might Lead to a Wrong Diagnosis and How to Avoid Them

Exactitud del diagnóstico neuropsicológico: factores que pueden conducir a un diagnóstico incorrecto y cómo evitarlos

Jeniffer **Ortega Marín**

Resumen

Los falsos positivos y los falsos negativos en la evaluación neuropsicológica pueden ocurrir debido a diferentes razones y pueden tener consecuencias graves tales como la administración de un tratamiento innecesario que puede causar efectos adversos y la incapacidad de proporcionar un tratamiento oportuno a los pacientes. Por lo tanto, es importante saber qué tanto podemos confiar en el diagnóstico. Los resultados de las pruebas son insuficientes para este fin. Como punto de partida, debemos establecer hipótesis sobre la condición del paciente utilizando la información que hemos recopilado de varias fuentes. Adicionalmente, es necesario utilizar información sobre las propiedades psicométricas de las pruebas, los valores predictivos y las tasas base de los trastornos neurocognitivos y de las puntuaciones bajas. En ausencia de esta información, la interpretación de los resultados debe ser cautelosa. En este artículo se discuten varias estrategias que pueden contribuir a mejorar la precisión diagnóstica.

Abstract

False positives and false negatives in neuropsychological assessment can occur due to different reasons and they could have serious consequences such as the administration of an unnecessary treatment that may cause adverse effects and failure to provide timely treatment to patients. Therefore, it is important to know how confident we can be in the diagnosis. Tests scores are insufficient to this end. As a starting point, we need to establish hypotheses about the patient's condition using the information we have collected from several sources. We also need to use information about the psychometric properties of the tests, predictive values and base rates of neurocognitive disorders and low scores. Whenever this information is lacking the interpretation of performance should be cautious. Several strategies are discussed that may contribute to improve diagnostic accuracy.

Citar como:

Dra Jeniffer **Ortega Marín**, MA Psi

Source | Filiacion:
Universidad el Bosque

BIO:
Doctora en Psicología y Magíster en Neuropsicología Clínica. Docente de Psicología. Ha investigado sobre errores de metacognición visual usando trucos de magia que explotan los fenómenos de ceguera por desatención y ceguera al cambio.

City | Ciudad:
Bogotá DC [co]

e-mail:
jortegamar@unbosque.edu.co

Exactitud del diagnóstico neuropsicológico:
factores que pueden conducir a un diagnóstico incorrecto y cómo evitarlos

Factors That Might Lead to a Wrong Diagnosis and How to Avoid Them

Jeniffer **Ortega Marín**

Introduction

Neuropsychological assessment is a decision-making process in which we collect, analyze and integrate information from different sources (Hebben & Milberg, 2009). As will be explained in this article, tests scores alone cannot tell us whether a patient has a condition of interest. Instead, we have to make a decision about the patient's condition and always bear in mind that decision-making might be affected by heuristics and biases, which could potentially lead to a wrong diagnosis.

The aim of this article is to discuss the factors that might lead to a wrong diagnosis and explain the strategies that might be used to minimize the risk of getting false positives and false negatives. The diagnosis of a disorder in a healthy individual (i.e., false positive) will have serious consequences such as the administration of an unnecessary treatment that may cause adverse effects, the generation of stress in patients and their relatives, and the use of resources from the healthcare system that should be invested in patients who really need them. On the other hand, the patient who is incorrectly classified as a healthy person (i.e., false negative) will not receive the necessary treatment (Huizenga, Agelink van Rentergem, Grasman, Muslimovic, & Schmand, 2016)

Ultimately, we want to know how confident we can be that our neuropsychological diagnosis is correct. When the neuropsychological assessment involves the administration of a battery of tests, this question needs to consider the following key aspects: sensitivity, specificity, predictive values, base rates of neurocognitive disorders and base rates of impaired performance. Other key elements that concerns the neuropsychologist are heuristics and biases that might affect her decision-making process.

Theoretical Framework

Factors That Might Lead to a Wrong Diagnosis

A wrong diagnosis could result from factors related to the instruments and testing conditions, for example, low sensitivity and specificity as well as the lack or neglect of base rates. It could also result from factors related to the neuropsychologist, such as biases and heuristics that affect decision-making processes. The aim of this section is to describe the impact of such factors in the diagnostic accuracy of neuropsychological assessments.

Sensitivity and specificity

False positives and false negatives are related to the sensitivity and specificity of the tests. Sensitivity refers to the ability of an instrument to produce a positive result for a person who has the condition of interest. It can be expressed as the proportion of people with the condition who get a score below a cut-off point and can be calculated using formula 1. Specificity refers to the ability of an instrument to produce a negative result for a person who does not present the condition of interest and can be calculated using formula 2. It can be expressed as the proportion of healthy people who score in the normal range. Sensitivity and specificity may range from **0%** to **100%** (Kent & Hancock, 2016). If a test has a sensitivity and a specificity of **100%** it will correctly categorize all cases of ill and healthy people. Values less than **100%** mean that the test will incorrectly categorize individuals, that is, in some cases it will indicate that a healthy person has a condition (i.e., false positive) and in other cases it will indicate that a person who has a condition is healthy (i.e., false negative). A test with a sensitivity of **90%** will correctly categorize **9** out of **10** people who have a disease. Just by keeping these two properties in mind, we know that tests scores alone cannot be the sole basis for a neuropsychological diagnosis.

$$Sensitivity = TP/TP+FN \qquad [1]$$

$$Specificity = TN/TN+FP \qquad [2]$$

In Figure 1, the rows indicate the type of result that a patient can obtain. The test is positive when it shows the presence of a condition and is negative when it shows that a condition is absent. The columns indicate the true presence or absence of a condition and refers to the sensitivity and specificity of the test. When crossing the rows and columns we have four possible results. The true positives (TP) and the true negatives (TN) are the ideal results. However, because the tests do not have perfect sensitivity or specificity, false positives (FP) and false negatives (FN) will also be obtained (Hebben & Milberg, 2009).

Is the condition present?		
	Yes (Sensitivity)	No (Specificity)
Positive	True positive (TP)	False positive (FP)
Negative	False negative (FN)	True negative (TN)

Figure 1. Sensitivity and specificity.

Base rates of neurocognitive disorders and predictive values

The situation gets more complex when it comes to identifying disorders that have a low prevalence or base rate. Let us say that Mr. OP takes the AZ test to see whether he has a certain type of neurocognitive disorder and gets a positive result. The test has a sensitivity of **70%** and a specificity of **57%**, thus correctly categorizing **7** out of **10** people who have the disorder. The patient informs his family about the diagnosis and begins preparing for the treatment and its side effects. However, his doctor tells him that this type of neurocognitive disorder rarely occurs in the population. The perspective changes completely when taking this information about base rate into account (Hebben & Milberg, 2009; Urbina, 2014). Using information about sensitivity, specificity and base rates, it is possible to obtain predictive values to know how confident we can be in the result. Figure 2 shows the expected results when administering the test to **100** people, of whom only **10** actually have the condition. Classification according to the AZ test would result in the correct diagnosis of about **58.3** people, whereas about **41.7** cases would be misclassified. What level of confidence could we have in the result obtained by Mr. OP? In other words, does he really have the disorder or is he a false positive? This question can be addressed by using predictive values.

The positive predictive value (PPV) refers to the probability that an individual has the condition given that the test result is positive, and the negative predictive value (NPV) is the probability that an individual does not have the condition given that the test result is negative (Glaros & Kline, 1988). A Bayesian analysis is applied to obtain the probabilities, which can be calculated using the following formulas:

$$PPV = TP/TP+FP \qquad [3]$$

$$NPV = TN/TN+FN \qquad [4]$$

Of the **45.7** cases classified as positive by the AZ test, **7** had cognitive impairment. Therefore, the PPV is **0.15**. Of the **54.3** cases classified as negative, **51.3** are healthy, thus the NPV is **0.94**. This means that we can be more confident in negative results than in positive ones. Although the result of Mr. OP test was positive, he is unlikely to have the disorder given that it is a very rare condition (it only occurs in **10%** of the population).

Is the disorder present?			
	Yes	No	Total
Positive	7	38.7	45.7
Negative	3	51.3	54.3
Total	10	90	100

Figure 2. The AZ test example.

Predictive values vary based on sensitivity, specificity, and base rates. In general, a high level of sensitivity and specificity is associated with a high level of predictive value (Glaros & Kline, 1988). Previously, we saw that the PPV would not be useful to confirm the result when the base rate is low. On the other hand, if the base rate is very high, we can be more confident in a positive result than in a negative result. With a base rate of **95%** we obtain a PPV of **0.97** and an NPV of **0.49**. However, in this case the neuropsychologist would be correct **95%** of the time just by assuming that all the people have the condition and a comprehensive assessment would be impractical.

In a study conducted with a sample of **95** patients, the Montreal Cognitive Assessment (MoCA) was used to identify cognitive impairment due to a cerebrovascular event. Using raw data, the instrument showed a sensitivity of **0.94** and a specificity of **0.42**. However, sensitivity and specificity may vary according to sociodemographic factors. Using scores corrected for age and education level, the MoCA showed a sensitivity of **0.67** and a specificity of **0.9** (Godefroy, et al., 2011)

If the MoCA is administered to patients who have suffered a cerebrovascular event and a cut-off point of **20** is used, how much confidence could we have in a positive outcome and a negative outcome? In the study by Godefroy et al. (2011) the PPV was **0.93** and the NPV was **0.57**. In other words, if a patient gets a score higher than **20**, we can be moderately sure that she is a healthy individual. Whereas if a patient gets a score below **20**, we can be confident that she has a cognitive impairment. In this case, we could proceed to perform a comprehensive assessment. Nevertheless, using a battery of tests is an intricate issue because it can increase the likelihood of getting false positives.

Base rate of low scores

Individuals may obtain low scores simply due to chance (Russell, Russell, & Hill, 2005). If we administer only one screening test, the cut-off point we choose for determining abnormality also sets the false positive rate (Decker, Schneider, & Hale, 2012). For instance, we may decide to categorize individuals as having an impaired performance when their score is below the 5th percentile. This means that our false positive rate is **0.05**. However, in the typical scenario we administer a battery of neuropsychological tests. As we add more tests, the probability of obtaining one or more low scores increases.

The percentage of people obtaining scores in the clinical range can be empirically derived in a battery of co-normed tests (Decker et al., 2012). In contrast, using a flexible battery requires a different approach. We could adjust the false positive rate using a Bonferroni correction (Blakesley, et al., 2009; Huizenga et al., 2016). For instance, if the false positive rate is **0.05** and we use **10** tests, the adjusted rate using the Bonferroni correction would be 0.05/10 = 0.005, which corresponds to a z-score of **-2.56**. However, this cut-off point is very stringent and it lowers sensitivity.

Base rate approaches have been proposed in scenarios where the tests are either correlated or uncorrelated. If the tests are uncorrelated, the percentage of individuals that will obtain one or more low scores will increase as more tests are used. This percentage can be calculated by means of the binomial distribution. If we use **10** tests and set the false positive rate to **0.05**, we get that **40.1%** of individuals will have one or more scores in the impaired range. Whereas if we use **20** tests, we get that **64.6%** of individuals will have one or more scores in the impaired range. This is a very straightforward procedure (for details, see Crawford, Garthwaite, & Gault, 2007). Nevertheless, neuropsychological tests are generally correlated and, in this case, using a Monte Carlo simulation method is a better approach (for details, see Crawford et al., 2007). To put this information into context, let us imagine that a co-normed battery of **12** tests with normal distribution was given to a patient who was suspected of having a mild neurocognitive disorder. According to the first diagnostic criteria in the DSM-5, the patient must present a modest cognitive impairment in at least one domain. This impairment should be documented through a neuropsychological assessment. The patient obtained a low score in one test and the

neuropsychologist thought that she met the first criterion. However, if one standard deviation was used as an indicator of impairment, about **1/6** of the scores obtained by the patient could be in the clinical range due to chance. Given that a battery of 12 tests was administered, it can be expected that two scores will be in the clinical range just by chance (Russell et al., 2005). Thus, the neuropsychologist capitalized on chance by concluding that the patient had an impairment.

Now let us imagine that a flexible battery of **10** tests was administered to a patient. The cut-off point was set to the **10**th percentile. The patient's scores in four tests was below this cut-off point. A study using a Monte Carlo simulation method estimated that **9.5%** of people obtained four or more scores at or below the **10**th percentile using the same battery. Therefore, the neuropsychologist can be confident that the patient has a neurocognitive disorder because relative few people have four or more scores at or below the cut-off point.

How common are low scores in healthy individuals? Mistridis et al. (2015) determined base rates of low performance using the Consortium to Establish a Registry for Alzheimer's Disease-Neuropsychological Assessment Battery (CERAD-NAB). In this study, **1081** adults participated between **49** and **92** years of age. In total, **60.6%** obtained scores at or below the **10**th percentile. Mistridis et al. (2015) provided information about the number of low scores that are required to determine probable cognitive impairment using different cut-off points. This diagnosis is established when the patient obtains a number of low scores that are observed in a percentage equal or less than **10%** of the sample. For example, if a patient gets four low scores on the battery and a cut-off at or below the **10**th percentile is used, cognitive impairment is considered likely since only **9.5%** of participants obtained four or more low scores. The less strict the cut-off point, the greater the number of low scores is required to establish the diagnosis. If the **25**th percentile is used as the cut-off point, six or more low scores are needed.

The neuropsychologist's requirements on sensitivity and specificity determines the choice of the cut-off point. A strict cut-off point (e.g., 1st percentile) would mean a greater specificity and therefore a lower number of false positives. However, the sensitivity would decrease, thereby increasing the number of false negatives. In the study by Godefroy et al. (2011) the MoCA had a sensitivity of **39%** and a specificity of **100%** when using the **15**th percentile as the cut-off point.

The base rate of low scores may differ according to the education level. Tanner-Eggen, Balzer, Perrig and Gutbrod (2015) administered a battery of neuropsychological tests to a sample of **569** healthy adults between **16** and **65** years of age and with an average education level of **13.8** years. In total, **40.9%** of the sample obtained three or more low scores using the **16**th percentile as the cut-off point. Interestingly, Tanner-Eggen et al. (2015) found that **72.6%** of people who had an education level of less than **12** years had three or more low scores, compared to **37.7%** of those with an education level of **12** years or more.

Additionally, Tanner-Eggen et al. (2015) reported base rates for the intra-individual standard deviation (ISD). The ISD is a measure of the variability of performance that indicates the level of dispersion of the scores. An ISD of zero indicates a flat profile, while an ISD equal or greater than one means that the performance is variable. The ISDs in this study ranged from **0.4** to **1.275** and only **17.8%** of the sample had a dispersion level equal or greater than **0.9**. Of particular relevance is that variability was greater with increasing age.

Another factor that may affect the base rate of low scores is the presence of mood disorders. Iverson, Brooks, Langenecker, and Young (2011) calculated the base rate of low scores for a group of healthy adults and a group of patients with mood disorders between the ages of **20** and **54** using a computerized cognitive screening battery. In total, **31.2%** of patients had two scores at or below the **5**th percentile compared to **8.2%** of controls. This finding suggests that it is not appropriate to use information about base rates of low scores or normative data determined in a sample of healthy individuals to establish whether a patient diagnosed with a mood disorder has a neurocognitive dysfunction.

Biases and heuristics

Wrong diagnoses might stem from unconscious factors that influence the decision-making process. Unknowingly, we could try to fit tests results into a particular diagnosis. This is due to the confirmation bias (Mercier, 2017; Spores, 2013) which consists of "the seeking or interpreting evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand" (Nickerson, 1998, p. 175). For instance, a neuropsychologist might interpret tests results in such a way that it helps the patient get a treatment while ignoring information that points to other alternatives. Conversely, the neuropsychologist's interpretation might result in a false negative if the patient is likely to be negatively affected by a clinical diagnosis. Because this may occur unconsciously, we should look upon our decision-making process critically and seek ways to improve our diagnostic accuracy.

In addition, if we collect information solely about a patient's symptoms it is likely that our clinical judgment would rely on representativeness. This heuristic involves making judgments based on how much a person looks like a typical patient with a certain condition (Watkins, 2009). If a patient's symptoms were very similar to those experienced and reported by people with dementia, then we would be confident that our diagnosis of dementia is correct. However, the problem with representativeness is that it ignores other relevant information (e.g., premorbid condition) and therefore it might be misleading (Teigen, 2017; Watkins, 2009).

Decision-making is far from being a fully conscious process. The mind has many unconscious heuristics and biases (Kahneman, 2011) that may have an impact on our clinical judgments (Watkins, 2009). However, we can learn more about them and implement strategies to minimize diagnostic error.

In sum, there are several factors that might lead to a wrong diagnosis: **1**. Decontextualized interpretation of tests scores. **2**. Using tests with low sensitivity and specificity, **3**. Overlooking base rates of neurocognitive disorders and low scores. **4**. Biases and heuristics that affect decision-making. In the next section, several strategies are described, which could minimize the risk of getting false positives and false negatives.

Strategies to Improve Diagnostic Accuracy

In some contexts, it may occur that normative data are not available, that the base rates are unknown, and/or the tests have not been validated for a group of interest. We would face greater uncertainty and our diagnoses should be more conservative or provisional. Having all this information in mind we should critically examine how we are using tests and conducting neuropsychological assessments in places where little work has been done in standardizing batteries, validating tests and obtaining base rates.

Recently, normative data was obtained for a sample of Colombian adults over **50** years old using the NEURONORMA battery (Duarte, 2017; Espitia, 2017). However, the administration of flexible batteries, the use of normative data obtained from non-representative samples or from other cultures are still common practices. Moreover, base rates have not yet made inroads into clinical neuropsychology. It is uncertain whether neuropsychologists are acting appropriately to minimize the risk of getting false positives.

Several strategies can be useful to achieve greater diagnostic accuracy. For instance, establishing hypotheses before the administration of tests, collecting comprehensive information about the patient, conducting interdisciplinary assessments, learning information about the tests, performing serial neuropsychological testing and using several instruments to assess a single cognitive ability.

Use a hypothesis-driven assessment

Spores (2013) suggested that testing is analogue to conducting a scientific research. In this sense, the first step consists of establishing hypotheses about the patient's condition. Wright (2011) posited that each case:

> *Can be treated as a research study by (a) making hypotheses and testing them to rule out possibilities and incorporate others, and (b) using multiple tests and multiple methods, which provide more solid data and allow the assessor to be much more confident in his or her findings (p. 6).*

A hypothesis-driven assessment enables the clinician to use knowledge of brain-behavior relationships and the available information about the patient to generate several hypotheses about her condition. Subsequently, we would proceed as if doing a research and tests our hypotheses using a neuropsychological battery. It is important to keep in mind that we should collect the data after we have formulated the hypotheses and not the other way around. The reason for this is that we might be biased by the data. Moreover, we should always examine the hypothesis that the patient is healthy, although this may not be apparent from a single assessment.

Collect comprehensive information about the patient and adopt an interdisciplinary approach

The hypotheses are constructed based on the information obtained through interviews with the patient and relatives, medical reports and other sources of information (Wanlass, 2012). The more information we collect, the richer the context will be to guide diagnosis. Therefore, an interdisciplinary approach is advisable.

To give an example of a real case, Figure 3 shows the performance of a **79**-year-old female patient on the Rey Complex Figure at two different times.
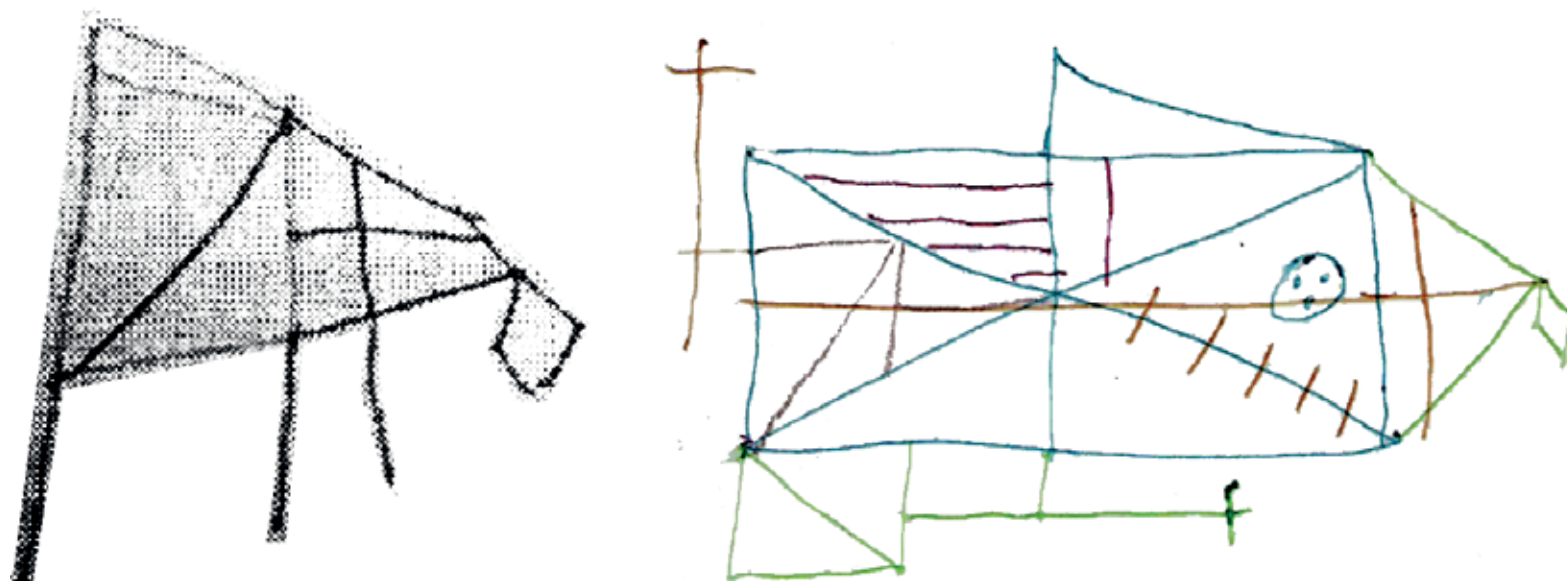
Figure 3. Rey Complex Figure Test of a 79-year-old female patient.

From the performance on the first assessment, the neuropsychologist inferred that the patient had a "severe impairment in the integration and orientation of objects" and lacked "appropriate organization and planning strategies". Her performance was also low in other tests. In the clinical history, it was briefly mentioned that that the patient had been treated for depression. Surprisingly, this important detail was ignored at the time of interpreting the results. In the second assessment that was carried out in a different institution, the patient's performance in the same tests was within normal limits. This example illustrates the risk of using neuropsychological tests in a decontextualized manner, it highlights the importance of longitudinal assessments, the convenience of using multiple tests to assess the same construct and the relevance of an interdisciplinary approach. Moreover, it shows that overlooking variability might lead to a wrong diagnosis.

The risk of getting a false positive is huge when more attention is paid to tests results at the expense of overlooking relevant contextual information. To begin with, the patient had two characteristics that made her likely to have a variable cognitive profile: her age and depressive symptoms. Representativeness could easily lead us to think that a **79**-year-old person with memory complaints has dementia. That is why hypotheses play a crucial role in clinical decision-making. Another important mistake was to give the diagnosis based on a single neuropsychological assessment instead of approaching the case more cautiously by doing more assessments. To make matters worse, the neuropsychologist wrote in her report that the patient's symptoms were compatible with a degenerative condition of moderate severity despite the fact that she indicated in the clinical history that the patient could perform activities of daily living independently. She did not provide a measure of functionality in her report. However, the second assessment revealed that the patient was able to perform basic and instrumental activities of daily living by herself and consequently she did not meet diagnostic criteria for a major neurocognitive disorder. Typically, patients and relatives do not have a way of telling whether a neuropsychologist made a mistake. It only becomes apparent during the course of a longitudinal assessment when a different neuropsychologist looks at the clinical history and tests results more carefully. It would be a relief to know that most neuropsychologists in our healthcare system take precautions to minimize the risk of misdiagnosis.

The following case illustrates the importance of knowing as much we can about the patients' lives and not just about their functional and cognitive impairments. An **84**-year-old female patient was referred to the neuropsychology service in a psychiatric institution. At the time of the assessment, the patient had depressive symptoms and her performance indicated that she had mild deficits in verbal learning and naming. Because depression might account for these deficits,

the neuropsychologist stated in her conclusion that she needed a psychiatric and psychological assessment, then she would assess the patient again in six months to establish the diagnosis. For many patients and relatives, it would be a relief to know that conditions other than dementia might explain functional problems. However, the daughter of this patient was outraged. She angrily complained to the neuropsychologist that she needed a diagnosis of dementia to declare her mother as mentally incompetent. If the daughter succeeded, this would allow her to administer her mother's assets. Because the neuropsychologist refused to change her conclusion based on this information, the daughter said she would seek an opinion from another specialist. Obviously, she did not care about what was really going on with her mother. This case highlights the fact that some elderly patients are in a vulnerable position and we have the responsibility to be meticulous about contextual information.

Learn background information about the tests and keep a critical perspective

We should know about the psychometric properties of the tests and the availability of normative data (Thambirajah, 2005). Whenever this information is lacking the interpretation of performance should be cautious. For those who are keen on neuropsychological testing and get very excited when a new instrument is available in the market, it might be eye-opening to read "A friendly critique of neuropsychology: Facing the challenges of our future" (Ruff, 2003). In addition, here is an anecdote that always shocks my students: some time ago, I bought a battery of tests to assess executive functions. There were several papers about it and I was very excited with the purchase until I discovered some inconsistencies in the manual. I reached out to the first author of the battery who happens to be a recognized neuropsychologist and author. His revelation shocked me to the core, but at least he was honest. This was his reply:

Unfortunately, I did not have anything to do with the construction and development of the processes that you mention (surprising as it may seem to you). The publishing company is responsible for the editing (I'm just the intellectual author -in abstract- of the battery). The publishing company is responsible for the concrete material. In fact, I know that some modifications were made due to this and other errors, although it might seem surprising to you, up to this date I do not know what they modified exactly (because nobody has informed me) (Author, personal communication, November 12, 2012).

Factors That Might Lead to a Wrong Diagnosis and How to Avoid Them

Professionals might feel attracted to an instrument because the authors are respected clinicians or academics, but behind the scenes, their contribution might have been minimal. This is a disturbing ethical issue.

Perform serial neuropsychological testing

Patients may exhibit a certain degree of variability in their performance (Fellows & Schmitter-Edgecombe, 2015). Neuropsychologists can apply several techniques to compare test results when more than one assessment has been carried out. For instance, simple discrepancy scores result from subtracting the score obtained in the second assessment from the score obtained in the first assessment. A discrepancy score of zero indicates that there has been no change, a positive discrepancy score could be attributed to the practice effect, intervention or regression to the mean, and a negative score could signal that the patient's condition has gotten worse if other explanations have been ruled out. Variability in either direction can be completely normal, and therefore it is necessary to estimate whether it is significant. Reliable change indexes (RCI) can be used to compare test-retest scores and determine whether there has been a significant change over time (Duff, 2012; Frerichs & Tuokko, 2005; Hinton-Bayre, 2010; Pardo & Ferrer, 2013; Sánchez-Benavides, et al., 2016; Stein, Luppa, Brähler, König, & Riedel-Heller, 2010).

Use more than one instrument to assess a single cognitive ability

Using several measures to test the same hypothesis can give us more confidence in the result (Sohlberg & Mateer, 2011; Woods, Ludicello, Cobb Scott, & Grant, 2009; Wright, 2011). For instance, if the patient's scores are in the clinical range in all the tests that measure processing speed, we can be more confident that her processing speed is impaired.

Sometimes there might be pressure to indicate a diagnosis based on a single assessment. Whenever there is a strong indication that full criteria for a diagnosis will be met, but we need to collect more information before making a firm diagnosis, we can indicate our uncertainty by using the specifier "provisional" (American Psychiatric Association, 2013). For instance, a provisional diagnosis might be used when a patient is undergoing a depressive episode at the time of the assessment and the history and neuropsychological assessment strongly suggest that he may have a neurocognitive disorder. Especially in these cases, it is very valuable to administer multiple tests to assess the same cognitive ability and to perform longitudinal assessments in order to detect extreme variability in performance.

Discussion and conclusions

Clinical diagnosis is a decision-making process that involves forming and testing hypotheses about a patient's condition (Jurado & Pueyo, 2012; Vakil, 2012; Wright, 2011). The ability of the neuropsychologist to analyze and integrate information from all sources will be determinant to reach a diagnosis. One important source is tests scores. Some students and professionals believe that they will find something if

they use many tests. Indeed, they are more likely to find something. However, the more tests they use, the more likely they are to get false positives (Brooks, Sherman, Iverson, Slick, & Strauss, 2011).

As mentioned earlier, the sensitivity and specificity of the tests are not perfect, which may give rise to false positives and false negatives. Moreover, it is common for some patients with no neuropsychological impairment to perform poorly on one or more tests simply due to chance or normal variability in performance (Binder, Iverson, & Brooks, 2009). The history and current status of the individual provides the necessary context for interpreting test results (Wright, 2011). Accordingly, a neuropsychological assessment that poorly describes the patient's history and current symptoms could be considered a bad practice.

Additionally, knowledge of predictive values is very useful to determine how confident we can be in test results. However, it is necessary to bear in mind the following aspects: first, it may be difficult to determine the base rates for neurocognitive disorders and we need this information to calculate predictive values (Slick, 2006). Nevertheless, it is possible to use imprecise base rates when administering tests that have a high level of sensitivity and specificity (Glaros & Kline, 1988). Second, it is possible for a healthy person to obtain one or more scores in the clinical range. Therefore, to increase diagnostic accuracy it would be valuable to know the base rates of low scores in the general population. Given the practical value of this information, more studies should be conducted to obtain base rates. In the absence of this information, normative data (e.g., the NEURONORMA data for Colombians over **50** years) should be used cautiously in both clinical and research settings.

There are many neuropsychological instruments available in the market and the number is growing (Ruff, 2003). However, in some contexts this flood of tests is probably being used in the absence of normative data and base rates. Accordingly, more effort should be exerted to obtain this information. Neuropsychological reports should state clearly whether normative information and base rates were used to interpret tests results because this would highlight the degree of uncertainty in the interpretation. The neuropsychologist should also highlight other information that was relevant to make the diagnostic decision.

The following key points emerge from the aforementioned studies: **1**. Sensitivity and specificity are influenced by sociodemographic factors (Godefroy, et al., 2011). Therefore, interpretation of scores in the absence of normative data for the target population should be conservative. **2**. Low scores are common in healthy individuals (Mistridis, et al., 2015) and even more common in patients with low education level (Tanner-Eggen et al., 2015) and mood disorders (Iverson et al., 2011). Longitudinal assessments can minimize the risk of misinterpretation in the absence of base rates for low scores and lack of normative data. **3**. Variability in performance is more common in older adults than in younger adults (Tanner-Eggen et al., 2015). Thus, variability indicators become essential during the assessment. If we consider these findings, our selection of neuropsychological tests should be based on their psychometric properties, availability of normative data and base rates. Moreover, we should be careful not to interpret low scores and average scores as indicators of cognitive impairment and normal cognitive functioning, respectively.

Nevertheless, we cannot completely eliminate false positives and false negatives. They may occur even if we use strategies to improve accuracy, but to a lesser extent. An interdisciplinary approach is vital in this process as it provides a wider perspective about the patient's condition. In the case of the **79**-year-old female patient,

the neuropsychologist that conducted the first assessment in **2013** overlooked the information related to her mood disorder. It seemed that the diagnosis was based solely on tests scores. Conducting an interdisciplinary hypothesis-driven assessment would have led the neuropsychologist to consider the possibility that the symptoms were explained by the mood disorder. Because the patient's performance during the first assessment was probably affected by her depressive symptoms, the neuropsychologist could have made a provisional diagnosis and justify the need to conduct a serial neuropsychological testing. In fact, the **2015** assessment revealed that the previous diagnosis was wrong.

Finally, we should keep in mind that the neuropsychological assessment is a decision-making process and, as such, interpretation of test findings might be influenced by heuristics and biases. The following strategies might be adopted to avoid the representative heuristic: use an interdisciplinary hypothesis-driven assessment, collect comprehensive information about the patient and perform a serial neuropsychological testing. To avoid the confirmation bias it is useful to adopt an interdisciplinary hypothesis-driven assessment and to use more than one instrument to assess a single cognitive ability.

Acknowledgments

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders*. Washington, DC: Author.

Binder, L., Iverson, G., & Brooks, B. (2009). To err is human: "abnormal" neuropsychological scores and variability. *Archives of Clinical Neuropsychology*, 24, 31-46. doi:*10.1093/arclin/acn001*

Blakesley, R., Mazumdar, S., Dew, M., Houck, P., Tang, G., Reynolds III, C., & Butters, M. (2009). Comparisons of methods for multiple hypothesis testing in neuropsychological research (Vol. 23). *Neuropsychology*. doi:doi:*10.1037/a0012850*

Brooks, B., Sherman, E., Iverson, G., Slick, D. J., & Strauss, E. (2011). sychometric foundations for the interpretation of neuropsychological test results. In J. Glenn, & M. Schoenberg, *The little black book of neuropsychology: A syndrome-based approach* (pp. 893-992). New York: Springer. doi:*https://psycnet.apa.org/doi/10.1007/978-0-387-76978-3_16*

Crawford, J. R., Garthwaite, P. H., & Gault, C. B. (2007). Estimating the percentage of the population With abnormally low scores (or abnormally large score differences) on standardized neuropsychological test batteries: A generic method with applications. *Neuropsychology,* 21(4), 419-430. doi:*10.1037/0894-4105.21.4.419*

Decker, S., Schneider, J., & Hale, J. (2012). Estimating base rates of impairment in neuropsychological test batteries: a comparison of quantitative models. *Archives of Clinical Neuropsychology*, 27, 69–84. doi:*10.1093/arclin/acr088*

Duarte, L. (2017). L*enguaje, memoria y habilidades visuoconstructivas en el envejecimiento normal: Datos normativos con la Batería Neuronorma Colombia* (Tesis de maestría). Bogotá, Colombia: Universidad Nacional de Colombia. Retrieved from *http://www.bdigital.unal.edu.co/56804/*

Duff, K. (2012). Evidence-based indicators of neuropsychological change in the individual patient: Relevant concepts and methods. *Archives of Clinical Neuropsychology*, 27(3), 248-261. doi:*10.1093/arclin/acr120*

Espitia, A. (2017). *Funciones ejecutivas en el envejecimiento normal: Datos normativos con la batería Neuronorma.Colombia* (Tesis de maestría). Bogotá, Colombia: Universidad Nacional de Colombia. Retrieved from *http://www.bdigital.unal.edu.co/56861/*

Fellows, R., & Schmitter-Edgecombe, M. (2015). Between-domain cognitive dispersion and functional abilities in older adults. *Journal of Clinical and Experimental Neuropsychology*, 37(10), 1013-1023. doi:*10.1080/13803395.2015.1050360*

Frerichs, R., & Tuokko, H. (2005). A comparison of methods for measuring cognitive change in older adults. *Archives of Clinical Neuropsychology*, 20(3), 321-333. doi:*10.1016/j.acn.2004.08.002*

Glaros, A., & Kline, R. (1988). Understanding the accuracy of tests with cutting scores: The sensitivity, specificity, and predictive value model. *Journal of Clinical Psychology*, 44(6), 1013-1023. doi:*10.1002/1097-4679(198811)44:6<1013::AID-JCLP2270440627>3.0.CO;2-Z*

Godefroy, O., Fickl, A., Roussel, M., Auribault, C., Bugnicourt, J. M., Lamy, C., . . . Petitnicolas, G. (2011). Is the Montreal Cognitive Assessment superior to the Mini-Mental State Examination to detect poststroke cognitive impairment? A study with neuropsychological evaluation. *Stroke*, 42, 1712-1716. doi:*10.1161/STROKEAHA.110.606277*

Hebben, N., & Milberg, W. (2009). *Essentials of neuropsychological assessment* (Second ed.). Hoboken, New Jersey: John Wiley & Sons. Retrieved from *https://www.wiley.com/en-us/Essentials+of+Neuropsychological+Assessment-p-9780471215899*

Hinton-Bayre, A. (2010). Deriving reliable change statistics from test–retest normative data: Comparison of models and mathematical expressions. *Archives of Clinical Neuropsychology*, 25(3), 244-256. doi:*10.1093/arclin/acq008*

Huizenga, H., Agelink van Rentergem, J., Grasman, R., Muslimovic, D., & Schmand, B. (2016). Normative comparisons for large neuropsychological test batteries: User-friendly and sensitive solutions to minimize familywise false positives. J*ournal of Clinical and Experimental Neuropsychology*, 38(6), 611-629. doi:*10.1080/13803395.2015.1132299*

Iverson, G., Brooks, B., Langenecker, S., & Young, A. (2011). Identifying a cognitive impairment subgroup in adults with mood disorders. *Journal of Affective Disorders*, 132, 360–367. doi:*10.1016/j.jad.2011.03.001*

Jurado, M., & Pueyo, R. (2012). Doing and reporting a neuropsychological assessment. *International Journal of Clinical and Health Psychology*, 12(1), 123-141. Retrieved from *https://www.redalyc.org/pdf/337/33723707009.pdf*

Kahneman, D. (2011). *Thinking fast and slow*. London: Penguin Group. Retrieved from *https://www.penguin.co.uk/books/563/56314/thinking--fast-and-slow/9780141033570.html*

Kent, P., & Hancock, M. (2016). Interpretation of dichotomous outcomes: sensitivity, specificity, likelihood ratios, and pre-test and post-test probability. *Journal of Physiotherapy*, 62(4), 231-233. doi:*10.1016/j.jphys.2016.08.008*

Mercier, H. (2017). Confirmation bias - Myside bias. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory* (Second ed., pp. 99-114). New York: Routledge.

Mistridis, P., Egli, S. C., Iverson, G. L., Berres, M., Willmes, K., Welsh-Bohmer, K. A., & Monsch, A. U. (2015). Considering the base rates of low performance in cognitively healthy older adults improves the accuracy to identify neurocognitive impairment with the Consortium to Establish a Registry for Alzheimer's DiseaseNeuropsychological Assessment Battery. *European Archives of Psychiatry and Clinical Neuroscience*, 265, 07–417. doi:*10.1007/s00406-014-0571-z*

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175-220. doi:*10.1037/1089-2680.2.2.175*

Pardo, A., & Ferrer, R. (2013). Significación clínica: falsos positivos en la estimación del cambio individual. *Anales de Psicología*, 29(2), 301-310. doi:*10.6018/analesps.29.2.139031*

Ruff, R. M. (2003). A friendly critique of neuropsychology: facing the challenges of our future. *Archives of Clinical Neuropsychology,* 18(8), 847-864. doi:*10.1016/j.acn.2003.07.002*

Russell, E. W., Russell, S. L., & Hill, B. (2005). The fundamental psychometric status of neuropsychological batteries. *Archives of Clinical Neuropsychology*, 20(6), 785-794. doi:*10.1016/j.acn.2005.05.001*

Sánchez-Benavides, G., Peña-Casanova, J., Casals-Coll, M., Gramunt, N., Manero, R. M., Puig-Pijoan, A., . . . Blesa, R. (2016). One-year reference norms of cognitive change in spanish old adults: Data from the NEURONORMA sample. *Archives of Clinical Neuropsychology*, 31(4), 378-388. doi:*10.1093/arclin/acw018*

Slick, D. J. (2006). Psychometrics in neuropsychological assessment. In E. Strauss, E. M. Sherman, & O. Spreen (Eds.), *A Compendium of neuropsychological tests: Administration, norms, and commentary* (Third ed., pp. 3-32). New York: Oxford University Press.

Sohlberg, M. M., & Mateer, C. A. (2001). *Cognitive rehabilitation: An integrative neuropsychologcial approach*. New York: The Guilford Press.

Spores, J. M. (2013). *Clinician's guide to psychological assessment and testing: With forms and templates for effective practice*. New York: Springer.

Stein, J., Luppa, M., Brähler, E., König, H.-H., & Riedel-Heller, S. G. (2010). The assessment of changes in cognitive functioning: Reliable change indices for neuropsychological instruments in the elderly – A systematic review. *Dementia and Geriatric Cognitive Disorders*, 29(3), 275-286. doi:*10.1159/000289779*

Tanner-Eggen, C., Balzer, C., Perrig, W. J., & Gutbrod, K. (2015). The neuropsychological assessment of cognitive deficits considering measures of performance variability. *Archives of Clinical Neuropsychology*, 30, 217–227. doi:*10.1093/arclin/acv008*

Teigen, K. H. (2017). *Judgments by representativeness. In R. F. Pohl (Ed.), Cognitive illusions: Intriguing phenomena in thinking, judgment and memory* (Second ed.). Oxon, United Kingdom: Routledge.

Thambirajah, M. S. (2005). Chapter 8: Psychological measurements: Their uses and misuses. In *Psychological Basis of Psychiatry* (pp. 181-201). New York: Churchill Livingstone.

Urbina, S. (2014). *Essentials of psychological testing* (Second ed.). Hoboken, New Jersey: Wiley.

Vakil, E. (2012). Neuropsychological assessment: Principles, rationale, and challenges. *Journal of Clinical and Experimental Neuropsychology*, 34(2), 135-150. doi:*10.1080/13803395.2011.623121*

Wanlass, R. L. (2012). *The neuropsychology toolkit: Guidelines, formats, and language*. New York: Springer.

Watkins, M. W. (2009). Errors in diagnostic decision making and clinical judgment. In T. B. Gutkin, & C. R. Reynolds (Eds.), *Handbook of School Psychology* (Fourth ed., pp. 210-229). Hoboken, New Jersey: Wiley.

Woods, S. P., Ludicello, J. E., Cobb Scott, J., & Grant, I. (2009). Neuropsychological methods in mental disorders research: illustrations from methamphetamine dependence. In S. J. Wood, N. B. Allen, & C. Pantelis (Eds.), *The neuropsychology of mental illness* (pp. 194-205). Cambridge, United Kingdom: Cambridge University Press.

Wright, A. J. (2011). *Conducting psychological assessment: A guide for practitioners*. Hoboken, New Jersey: John Wiley & Sons.