

The NFL Parties, appealing the Level 1.5 Neurocognitive Impairment diagnostic assessment on January 13, 2020, argued that the Claims Administrator was clearly erroneous in adopting it. Doc. 219432. According to the NFL Parties, Mr. ██████ clinicians performed an invalid assessment, as they failed to apply demographic adjustments to the assessment scores and offered insufficient consideration to the possibility of medication side effects.

On April 4, 2020, an AAP Consultant found that, while Mr. ██████ assessment was facially valid, the lack of demographic adjustments “clouds valid determination of the presence of cognitive impairment,” noting that had appropriate racial adjustments been applied, Mr. ██████ would not qualify for Level 1.5 Neurocognitive Impairment. Doc. 223145. The Consultant further added that “there is high probability that medication side-effects account for part or all the player’s functional impairments.” *Id.*

On April 12, 2020, an AAP Reviewer concurred with the finding against Level 1.5 Neurocognitive Impairment, but on different grounds. Doc. 223323. The Reviewer concluded that Mr. ██████ functional abilities were generally consistent with scores of at least 1.0 for the CDR categories of Community Affairs, Home/Hobbies, and Personal Care.” *Id.* However, the Reviewer concluded reversal was warranted given the absence of full demographic adjustments. *Id.* On that basis, the AAP Reviewer explained that the “lack of confirmation of the degree of dysfunction by a valid neuropsychological evaluation is not generally consistent with the Settlement Agreement’s criteria,” although “it is clear that this player has both cognitive and functional impairments.” *Id.*

Since the AAP Consultant, as well as NFL Parties, noted that Dr. ██████ did not provide a rationale for deciding against the use of racial adjustments during the testing regimen, the Claims Administrator directly corresponded with Dr. ██████ and inquired as to his rationale for not using a “full demographic correction for the ACS norms as recommended in the BAP Clinician’s Guide.” Doc. 225562. Dr. ██████ responded: “I remain unsure what you are talking about. He was done using standard norms like everyone else. Using different racial standards is indeed discriminatory and illegal. We stand by our scores and will gladly defend them in any legal proceeding.” *Id.* The Claims Administrator shared that response with the AAP Consultant, who “noted that it appeared that Dr. ██████ was not complying with the language and intent of the settlement around demographic adjustments and therefore avoiding responding to our inquiry.” *Id.*

DISCUSSION

Though the briefing moots several issues, at this point the keystone is the use of race-based adjustments for neuropsychological test results, and, more specifically whether “recommended” functionally means “required.” Essentially, may a neuropsychologist avoid using a Claimant’s race in interpreting his test scores when the clinician’s medical judgment is that such a practice is

not appropriate, or is even discriminatory?¹ The Claims Administrator here decided that a full demographic adjustment was not necessary to make results valid. Was that choice a clear error, as the NFL Parties and the latest AAP Reviewer assert? To make that determination, we first provide background on the role of demographic norms in neuropsychological assessment.²

From Raw Scores to T-Scores

Retired players assessed through the Settlement’s Baseline Assessment Program undergo a test battery consisting of twenty-two individual subtests. Sixteen of these, the “Wechsler tests,” are drawn from the fourth edition of the Wechsler Adult Intelligence Scale or the fourth edition of the Wechsler Memory Scale. The remaining six come from elsewhere.³ All twenty-two subtests ask a series of questions or require behavioral performances (e.g., block constructions, paper-and-pencil coding) that probe cognitive ability and processing speed. The clinician administers each individual test, yielding a set of raw scores.

Raw scores are noisy. To make them comparable to each other and interpretable, neuropsychologists convert them to common scales (i.e., scaled scores, z-scores, T-scores). These common scales derive from comparing a given raw score to those obtained from a sample of “normal” people with similar characteristics. The “normal” set generates a mean (average) score, which results in a *scaled score* of 10, a z-score of 0, or a T-score of 50. The clinician then can compare a given raw score to this distribution and determine the corresponding *standard score*. Using standard scores (i.e., scaled scores, z-scores, T-scores) permits the clinician to evaluate the test taker’s performance in different cognitive domains using common metrics.

Thus, a crucial step in comparing test takers to others—in an attempt to accurately identify their cognitive state—is to decide which individual demographic characteristics to use in norming the raw score, and to settle on a technique to accomplish the adjustment.

Overall, demographic adjustments in neuropsychological testing strive to make the raw score meaningful when controlling for factors that are not unique to the individual’s cognition. Thus, adjusting for the test taker’s age makes sense in improving accuracy, as it would not be sensible to say that an older test taker experiences abnormal memory if the test takers in the comparison set are young. Adjusting for education similarly calibrates the scores by capturing the benefits of a formal learning experience on test taking skill, as well as on measures of cognitive abilities such as vocabulary and arithmetic skills. The manual for the Heaton norms, one set of normative data that we will describe in detail later, neatly describes the overall benefit of demographic-based adjustments:

¹ While gender also plays an operative role in both sets of demographic adjustments, the issue here specifically relates to the clinician’s use, or lack thereof, of racial adjustments, to account for Mr. [REDACTED] African American identity.

² The description in the text about such testing was gathered through consultation with the Settlement Program’s AAP Consultants, who provided invaluable perspective on how adjustments are applied and the use of the practice across the field of neuropsychology.

³ See Settlement Agreement, Exhibit 2, Section 4; Retired NFL Football Players’ Baseline Assessment Program: Neuropsychologists Handbook (the Clinician’s Interpretation Guide), at 4-6.

The ability to convert raw test scores to demographically corrected scores has a number of advantages for the neuropsychologist. First, it facilitates comparisons of an individual's test results with normal expectations based on that person's demographic characteristics. For example, the fact that a patient's raw test score is below average for the general adult population does not mean that it is a poor score for persons at all levels of age and education. The score may be above average in comparison to some adult samples and in the "impaired" range in comparison to others. Using the corresponding standard score, the clinician can establish more precisely what percentage of normal individuals with the same demographic characteristics perform at or below the patient's level of performance. Second, because standard scores in this normative system are more directly comparable, use of these scores facilitates the analysis of an individual's patterns of strengths and deficits across tests. Instead of trying to compare a time score on one test with an error score on another, comparisons are made between scores that have the same units of measurement. Similarly, in research involving groups of individuals, the use of demographically corrected scores facilitates comparisons of strengths and deficits both within and between groups.⁴

While incorporating measures for educational attainment and age may be seen as a common-sense way of standardizing scores, including race and ethnicity is more complicated. This technique may be most charitably seen as offering an imperfect proxy for socioeconomic status. Research has found that structural barriers and accompanying misfortune early in life may lead to diminished performance on cognitive tests, even after controlling for education and age.⁵ Some researchers attribute disparities in neurocognitive test scores to three underlying factors: acculturation, socioeconomic status, and quality of education.⁶ As a result, clinicians are tempted to use race as a representative heuristic for these experiences which contribute to lower cognitive test scores.

Moreover, standardized testing developed for and normed on largely White populations may bias others in ways difficult to control for without explicitly including race in post-hoc regressions.⁷ Race-specific norms that essentially adjust raw scores (upwards) for African American test takers thus result from a seemingly evidence-based and beneficent impulse: to correct for biased measuring techniques. Those who developed these adjustments believe them to be a key tool for avoiding misclassification of cognitive impairment:

⁴ ROBERT K. HEATON, S. WALDEN MILLER, MICHAEL J. TAYLOR, & IGOR GRANT, REVISED COMPREHENSIVE NORMS FOR AN EXPANDED HALSTEAD REITAN BATTERY: DEMOGRAPHICALLY ADJUSTED NEUROPSYCHOLOGICAL NORMS FOR AFRICAN AMERICAN AND CAUCASIAN ADULTS 4 (2004).

⁵ See Desiree A. Byrd, S. Walden Miller, Judy Reilly, Shirley Weber, Tamara L. Wall, & Robert K. Heaton, *Early Environmental Factors, Ethnicity, and Adult Cognitive Test Performance*, 20 CLINICAL NEUROPSYCHOLOGIST 243, 255 (2006).

⁶ See Philip G. Gasquoin, *Race-Norming of Neuropsychological Tests*, 19 NEUROPSYCHOLOGY REV. 250, 254-55 (2009).

⁷ See Robert K. Heaton, Lee Ryan, & Igor Grant, *Demographic Influences and Use of Demographically Corrected Norms in Neuropsychological Assessment*, in NEUROPSYCHOLOGICAL ASSESSMENT OF NEUROPSYCHIATRIC AND NEUROMEDICAL DISORDERS, 127, 146-47 (Igor Grant & Kenneth M. Adams eds., 3d ed. 2009).

[T]he noted ethnic differences in test performance are reliably observed and, although they may not accurately reflect differences in the ultimate potential of the examinees, they should be considered when interpreting cognitive tests within a neurodiagnostic context. Failure to do so typically results in a substantial (sometimes up to three-fold) increase in the probability of misclassifying normal African Americans as having brain disorders, as compared to misclassification rates for Caucasians.⁸

In practice (i.e., outside the structures of the Settlement and its accompanying guidance), the sixteen Wechsler tests are scored using Pearson Advanced Clinical Solutions (“ACS”) software, and clinicians have different standardization options.

One option is to use tables provided in the individual test booklets. Such normative data are typically drawn from relatively small samples observed at the time the test was developed, and may therefore be decades old. Often these data are stratified only by age.

Alternatively, clinicians can click on a button in the ACS software which will apply identical scaling and standardization practices to each test taken. The software always norms the raw score by age, binned into thirteen groups from sixteen to ninety. The clinician using the ACS software then must make a choice as to how to proceed.

One button in the software simply stratifies further only by education level. The software categorizes education by number of years of schooling, from less than eight years all the way up to more than eighteen years.

Alternatively, the clinician may apply the “full demographic adjustments.” Despite its name, this “full” adjustment only includes corrections—in addition to age and education—for two additional categories. The first is race/ethnicity, divided into four groups: White, African American, Hispanic, or Asian.⁹ The second is an indicator for a binary gender variable.

The software programs for the Wechsler tests use data from ostensibly typical people in a census-stratified sample of 2,200 Americans who took the tests between 2007 and 2008. A significant problem is that characteristics are not independent—e.g., women may be more likely to complete college than men. The process by which the ACS software accounts for this problem is somewhat obscure, as the underlying regression analyses are not easily available. It is our understanding that it approaches the problem in a step-wise manner: first, by creating an

⁸ HEATON, *supra* note 4, at 4.

⁹ The Hispanic normative group performed only slightly better on cognitive testing as compared to the African American normative group (whose adjustments are discussed more at length throughout this opinion), and therefore the two groups’ adjustments are of a similar magnitude. By contrast, the Asian normative group performed fairly close to the overall mean, indicating that adjustments for Asian identity have only a marginal impact on the ultimate cognitive scoring.

adjustment for gender-by-education-by-age, and then by applying these categories to norms for each racial/ethnic group.

The remaining six tests listed in the Settlement are not part of the Wechsler scales and are not covered by the ACS software. The Settlement, as described below, notes that the Revised Comprehensive Norms for an Extended Halstead-Reitan Battery, also known as “Heaton norms,” may be used for normative adjustments. In practice, it appears that the Heaton method requires explicit use of racial norming.

The Heaton adjustments, originally developed in 1991 but most recently updated in 2004, offer a different approach than the ACS’s stepwise adjustment. Clinicians can either use software to make simultaneous adjustments for age, education, gender, and race, or they can use tables that the publisher provides.¹⁰ In using the printed tables, clinicians match the raw score to a scaled score in one appendix, and then use a table that pertains to a specific permutation of age, education, gender, and race to determine the adjusted score in another.¹¹ The computer program is more precise, as it uses actual age and educational values in calculating T-scores, rather than using the midpoints of age and education ranges.¹²

The Heaton adjustment process begins (like the ACS system) with normalizing the raw score, here to a scaled score with a mean of ten and a standard deviation of three. For all tests with sufficient sample sizes for both racial groups, the scaled scores result from a pooled sample, though for some tests, like the BDAE Complex Ideational Material, the conversion to scaled scores turns only on a White sample. Researchers gathered the comparison samples over twenty-five years across over multiple different centers across North America.¹³ However, almost all the African Americans in the sample were collected in San Diego, from a census-stratified sample in that place.¹⁴

The Heaton normalization process results from a regression that predicts individual scaled scores, based on the given raw scores, from assigned demographic characteristics. For reasons that are unclear, Heaton stratifies by race first, so that separate regressions result for the African American category and “Caucasian” examinee category (which, apparently, is meant to represent all who do not identify as African American). The individual characteristics are thus Race+(Age/Gender/Education). To our knowledge, no interaction effects were included in the regression models.

The Heaton norms thus require the use of a binary race identifier, but limit the clinicians’ ability to see how much race influences scoring for each test taker. However, such effects can be large. For example, Mr. ██████ obtained a raw score of 49/60 on the Boston Naming Test. Dr. ██████ adjusted for Mr. ██████ age, education level, and gender. However, from review of

¹⁰ HEATON, *supra* note 4, at 35.

¹¹ HEATON, *supra* note 4, at 450-51.

¹² HEATON, *supra* note 4, at 35.

¹³ HEATON, *supra* note 4, at 7.

¹⁴ HEATON, *supra* note 4, at 7.

the file, it seems that he then analyzed Mr. ██████ scores against the sample of White test takers, yielding a T-score of 29. Had Dr. ██████ used norms based on an African American-based test taker sample, Mr. ██████ would have achieved a T-score of 38, which is only 1.2 standard deviations below the population mean. To put it differently, it is our best understanding that Dr. ██████ did apply a racial adjustment—as he was required to in employing the Heaton norms—but assigned Mr. ██████ to the White category, instead of to the African American one. This decreased Mr. ██████ estimated score by almost a full standard deviation on this one test.

In summary, for the twenty-two tests that form the basis for Claimants' entitlement to relief under the Settlement, the available techniques for producing scaled scores do not necessarily require selecting an appropriate comparison group based on self-identified race, but using such racial adjustments is extremely convenient. In many cases, including in this Appeal, when adjustments are applied for race to Wechsler scores, and when African American Heaton norms are used for the other scores, African American players' T-scores are increased. In other words, using race-specific norms can be enormously consequential, and the adjustments may often make the difference in a clinician's determination of cognitive impairment and a determination of normal functioning for Retired NFL Players seeking benefits under the Agreement.

The impacts of demographic adjustments, especially for race, have led some prominent neuropsychologists to worry about their use. As James A. Holdnack and Lawrence G. Weiss—whose textbook the Agreement adopted in its test battery—have written:

[R]acial/ethnic status cannot be determined by any scientific means and it may not be possible to accurately classify an individual client. The clinician will need to use the examinee's own conceptualization of their race/ethnicity. The clinician will also need to determine if the individual's background is representative of the factors that can result in cognitive differences between groups. In other words, if the examinee grew up in a wealthy neighborhood, with highly educated parents, attended good schools, and is also Hispanic, do the adjustments made to the normative data accurately reflect the individual's background? It was hypothesized that socioeconomic disadvantage, health care disparities, poor educational experiences, and other potential discriminatory factors may account for the between group differences but if none of these factors is present for a specific individual, does it make sense to adjust for ethnicity? There is no simple answer to this question and in some cases it may be yes (e.g., level of acculturation in the family generally may be low, affecting language development, effects of racism limiting opportunities) or in some cases no (e.g., background is not inconsistent with other non-minority groups). The clinician must use his or her judgment as to when it is appropriate to adjust for racial/ethnic differences.¹⁵

¹⁵ James A. Holdnack & Larry G. Weiss, *Demographic Adjustments to WAIS-IV/WMS-IV Norms*, in WAIS-IV, WMS-IV, AND ACS: ADVANCED CLINICAL INTERPRETATION 171, 199-200 (James A. Holdnack, Lisa Whipple Drozdick, Lawrence G. Weiss, & Grant L. Iverson eds., 2013).

Some have gone further and concluded that demographic adjustment is a flawed method that should *never* be used. Since race-norming studies sample only neurologically intact participants, the false negative rate of using the adjustments is unknown.¹⁶ A neuropsychologist may reason that potential costs of misdiagnosis outweigh the equalizing benefits. As using African American-specific norms increases the rate of false negatives, there is a risk that some may be denied access to necessary benefits or compensation solely on the basis of race.¹⁷

Meanwhile, identity groups are neither psychologically nor experientially homogeneous.¹⁸ For example, consider the application of the Heaton norms, based on a geographically limited sample of African Americans, compared to a national sample of individuals described as “Caucasian.” Are the differences observed between those two samples equally applicable to football players of different ethnicities and self-identified races, whose lives and experiences may have commonalities not shared by members of the general population? To what extent, if at all, should clinicians use their discretion in deciding whether adjusting for race is appropriate given a player’s prior socioeconomic circumstances?

Concerns about error in high-stakes settings leads the *Clinical and Interpretative Manual for the Wechsler tests* to a clear warning:

Caution is warranted when using norms adjusted for racial/ethnic group differences, as scores may be misunderstood or inappropriately interpreted. No scientific models exist for accurately identifying race and ethnicity (i.e., these are primarily social/political constructs), and the underlying factors for which race/ethnicity serve as a “proxy” are poorly understood. Manly and Echemendia observed that using “race-based” normative data may result in minority examinees not receiving needed services (e.g., increasing their scores above cutoffs). In addition, the use of race/ethnicity in normative data ignores the underlying cultural, health, and educational factors that result in disparities in test performance; subsequently, adjustments are made based on group membership, which may not fully represent the experiences and characteristics of a specific individual.¹⁹

In light of this warning, we now turn to what the Settlement Agreement requires.

The Settlement’s Exhibit 2, Section 4 and the Guide

¹⁶ See Gasquoine, *supra* note 6, at 254.

¹⁷ See Jason Brandt, *2005 INS Presidential Address: Neuropsychological Crimes and Misdemeanors*, 21 CLINICAL NEUROPSYCHOLOGIST, 553, 555–68 (2007).

¹⁸ See Patrick F. McKay, *The Effects of Demographic Variables and Stereotype Threat on Black/White Differences in Cognitive Ability Test Performance*, 18 J. BUS. AND PSYCHOL. 1, 6-13 (2003).

¹⁹ N.C.S. PEARSON, *ADVANCED CLINICAL SOLUTIONS FOR WAIS-IV AND WMS-IV: CLINICAL AND INTERPRETIVE MANUAL* (2009) (cleaned up).

To recover under the Agreement, Mr. ██████████ Diagnosis of Level 1.5 Neurocognitive Impairment must have been “determined by and in accordance with the standardized neuropsychological testing protocol annexed in Exhibit 2.”²⁰ Exhibit 2, Section 4 states:

There are 5 domains of cognitive functioning. In each domain, there are several tests that contribute 3, 4, or 6 demographically-adjusted test scores for consideration. Test selection in the domains was based on the availability of demographically-adjusted normative data for Caucasians and African Americans. These domains and scores are set out below.

The basic principle for defining impairment on testing is that there must be a pattern of performance that is . . . 1.7-1.8 standard deviations (for Level 1.5 Impairment) . . . below the person’s expected level of premorbid functioning. Therefore, it is necessary to have more than one low test score in each domain. A user manual will be provided to neuropsychologists setting out the cutoff scores, criteria for identifying impairment in each cognitive domain, and statistical and normative data to support the impairment criteria.

This language suggests that the Agreement’s drafters believed that demographically adjusting normative data was an important technique in accurately identifying whether the Claimant demonstrated a pattern 1.7-1.8 standard deviations below their own expected level of premorbid function. Indeed, both “standard deviations” and “expected level” are concepts that require some kind of comparison to an appropriate normative sample. However, the Agreement does not explicitly require adjustments based on *race*, but rather a broader category of “demographic” adjustment as defined by a user manual. In early 2017, Class Counsel and NFL Parties, with input from the BAP Administrator, developed that user manual, the Retired NFL Football Players’ Baseline Assessment Program: Neuropsychologist Handbook (the “Guide”).

The Guide represents the Parties’ joint understanding of the Agreement, and resulted from a collaborative and iterated process. The Guide provides that, “in general,” the procedure to characterize impairment has ten steps. The first is to “administer and score all tests.” The second reads:

Convert test scores to demographically-corrected T-scores via ACS software (*use of the full demographic correction is recommended*) or Revised Comprehensive Norms for an Extended Halstead-Reitan Battery.”²¹

Thus, the Guide suggests that clinicians “in general” employ a “full demographic correction” for the ACS norms, or some variation of the Heaton norms, in adjusting the raw scores to “demographically-corrected T-scores.” This is consistent with the Settlement, which explicitly

²⁰ See Settlement Agreement, Exhibit A-1, 2(a)(ii).

²¹ Retired NFL Football Players’ Baseline Assessment Program: Neuropsychologist Handbook (the Clinician’s Interpretation Guide), at 12 (emphasis added). To avoid test manipulation, the Guide is not a public document, and we therefore quote from it sparingly.

states that tests were selected based on their susceptibility to norming using samples with sufficient African American participants. But, contrary to the AAP Reviewer's views in this case, the language leaves significant room for clinical discretion.

For the ACS software, the Guide only *recommends* the “full demographic correction,” explicitly leaving open the possibility of merely using education and age.

Likewise, while the plain text of the Guide contemplates the use of Heaton norms—which require application of a binary race adjustment—it does not specify that the clinician must apply the African American sample adjustment to African American players. In other words, while “in general” Heaton norms may need to be applied, it is unclear whether the clinician can exercise discretion in deciding whether to apply the White or African American norms, or even, in particular cases, apply adjustments that result in “demographically-corrected” T-scores from the manuals themselves.

This discretionary language is admirable given the cautions in the Wechsler *Clinical and Interpretative Manual*, which provides the basis for much of the Settlement's processes. The Settlement's drafters were obviously concerned not to predetermine which particular kinds of demographic adjustment (i.e., for race, or socioeconomic status) were the gold standard, or were universally appropriate. The Guide expresses a set of default best practices, but expressly does not mandate them. The Guide's explicit refusal to require racial adjustments finds support in the views of the Appeals Advisory Panel Consultants, who have expressed significant concerns about the accuracy and appropriateness of these norms, and the possibility that they are particularly inapt when applied to the Settlement Program's population of Retired Players.

Thus, a careful reading of the Settlement and its associated documents leads us to conclude that it is inappropriate to deny a claim solely because the clinician chose to reject the Guide's recommendation to use African American normative samples in interpreting raw scores.

Here, the Claims Administrator approved the claim, though Dr. [REDACTED] apparently applied neither the ACS “full demographic adjustment” nor the African American-based Heaton norms. But the terminal AAP Reviewer disagreed. As the Reviewer wrote:

The Settlement is clear in recommending full demographic adjustments and the argument that adjusting for certain demographic variables (e.g. ethnoracial status) but not others (e.g. age and education) is discriminatory is not compelling. Failure to adjust for extraneous variables clouds valid determination of the presence of cognitive impairment. Consequently, full demographic adjustments should have been applied in order to be a valid assessment and when they are applied the results no longer meet the Settlement Criteria for Level 1.5 Neurocognitive Impairment.

Doc. 223323.

We conclude that the Reviewer misunderstood what the Settlement requires. The Reviewer's analysis consequently does not support a conclusion that the Claims Administrator's decision was clearly erroneous. Clinicians evaluating self-identified African American players *may* identify those players by that race and apply the full demographic adjustments under the ACS software and apply African American Heaton norm adjustments. But failure to do so is not itself a reason to deny a claim. The most that we can say is that *in general both such adjustments are presently recommended.*

Clinicians' discretion to adjust regarding racial norming has wide but appreciable limits. First, as the NFL Parties argued in their briefing, since full demographic adjustments are presently generally recommended, when a clinician does not use them, it is reasonable to require that the clinician explain why. The Claims Administrator may fairly worry that the clinician decided to adjust, or not, as a way of achieving a financial result for a particular player, instead of as an exercise of medical judgment. Thus, the Claims Administrator may require clinicians to show that their decision to avoid the recommendation was consistent with their ordinary practice. That might be accomplished in one of two ways.

First, a clinician may show that she consistently, in her non-Settlement practice, uses the limited ACS demographic adjustment and only the White norms for the Heaton tests. Such consistent practice would significantly assuage any worry that the adjustment was a one-time-only technique to achieve a particular result.

Second, she might adopt Dr. Holdnack and Dr. Weiss's argument that, due to individualized characteristics of the Claimant, racial norming using African American norms would not result in a "correct" demographic adjustment. That conclusion might then lead the clinician to use individual test norms, or another normative sample, if it is well-validated. Or it might lead the clinician to use the White Heaton sample regardless of the Claimant's contrary self-reported race and ethnicity. In either event, the Claims Administrator would certainly be justified in probing these answers to assure completeness and accuracy. In the broadest strokes, the Claims Administrator ought to permit the departure from the Guide's "general" recommendations when it is satisfied that clinicians' judgments are not outcome oriented.

In the medical records that Dr. [REDACTED] prepared, there is no clear indication as to how any demographic adjustments factored into Mr. [REDACTED] T-scores. In corresponding with the Claims Administrator, Dr. [REDACTED] indicated that he used "standard norms," expressed that using racial norms is an act of discrimination, and vowed to defend his scores "in any legal proceeding." That is not the same as saying that he does not use race-specific norms as an ordinary part of his clinical practice. And, in fact, his overall message comes close to saying the opposite. Dr. [REDACTED] statements are also obscure. Though he abjured "different racial standards," it is our understanding that he did apply norms based solely on a White Heaton normative sample.²² Dr.

²² Dr. [REDACTED] [REDACTED] has opined (positively) about the Heaton demographic adjustments:

[REDACTED]

█ additionally failed to provide a thoughtful and articulated analysis turning on Mr. █ individual background. He instead bluntly dismissed the Claims Administrator's well-founded authority to investigate the sufficiency of the records leading to the Diagnosis.

In sum, significant questions remain as to what system of adjustments Dr. █ used in assessing Mr. █ and how reflective that approach is of his general practice. The result of Dr. █ refusal to fully cooperate is that there is insufficient information in the file to know whether his choices regarding racial norms resulted from his consistent practice to never select African American normative samples, a judgment about Mr. █ as an individual, or something else. His report is lamentably devoid of the sort of detail which could enable the Claims Administrator to defer to it.

It is therefore appropriate for this Appeal to be remanded to the Claims Administrator, which may seek more clarity from Dr. █ as to exactly what he did, and how it relates to his normal practice. If he does not cooperate in developing a sufficiently clear record, the Claim may be denied on that basis, as the test scores would not sufficiently support the Diagnosis.

The remainder of this Appeal asks whether there were alternative explanations for Mr. █ functional impairment and whether his scoring was invalid on other grounds. The NFL Parties may raise these arguments again if Dr. █ explanation for his choices is sufficient. They may also decide, having been convinced by the AAP's analysis, not to raise them a second time in a later proceeding. In either case, concerns of economy and clarity make further discussion unnecessary.

CONCLUSION

The Appeal is granted in part. Mr. █ claim is remanded to permit the Claims Administrator to seek further documentation.

Date: August 20, 2020



David A. Hoffman, Special Master



Wendell E. Pritchett, Special Master

█

█