BRIEF COMMUNICATION

Healthy children and adolescents obtain some low scores across a battery of memory tests

BRIAN L. BROOKS,[1,2] GRANT L. IVERSON,[3,4] ELISABETH M.S. SHERMAN,[1,5,6]
AND JAMES A. HOLDNACK[7]

[1]Neurosciences Program, Alberta Children's Hospital, Calgary, Alberta, Canada
[2]Department of Pediatrics, University of Calgary, Calgary, Alberta, Canada
[3]Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada
[4]Research Department, British Columbia Mental Health & Addiction Services, Coquitlam, British Columbia, Canada
[5]Department of Pediatrics, University of Calgary, Calgary, Alberta, Canada
[6]Department of Clinical Neuroscience, University of Calgary, Calgary, Alberta, Canada
[7]Clinical Content Development, Pearson Assessment, San Antonio, Texas

Abstract

Obtaining some low memory scores across a battery of tests is common. The purpose of this study was to examine the prevalence of low scores on the Children's Memory Scale (CMS). Participants were 1000 children and adolescents between 5 and 16 years of age from the CMS standardization sample. Consistent with research on other batteries, having some low memory scores is common in healthy children and adolescents. The prevalence of low memory scores also increases with lower intelligence. Clinicians should be cautious when interpreting isolated low memory scores as sole evidence of memory impairment. Knowing the prevalence of low scores as a supplement to clinical judgment should reduce the likelihood of misdiagnosing memory problems. (*JINS*, 2009, *15*, 613–617.)

Keywords: Children, Pediatric, Memory, Psychometric, Base rate, Misdiagnosis

INTRODUCTION

Clinicians appreciate that healthy people can obtain some low scores on a test battery. This psychometric principle has been supported in the research literature when considering performance across batteries of cognitive tests (Axelrod & Wall, 2007; Binder et al., 2009; Crawford et al., 2007; Heaton et al., 1991, 2004; Iverson & Brooks, in press; Iverson et al., 2008a, 2008b; Schretlen et al., 2008).

The presence of low *memory* scores in healthy people (Brooks et al., 2007, 2008; Palmer et al., 1998) and the potential impact this has on identifying subtle or prodromal memory problems (de Rotrou et al., 2005) are of special relevance to neuropsychologists. Palmer et al. (1998) illustrated that 39.4% of healthy older adults obtained at least one score at or below 1.3 standard deviations (*SD*s) and 12.9% obtained at least one score at or below 2.0 *SD*s. In studies of two large standardization samples, Brooks et al. (2007, 2008) found that approximately one half of healthy older adults with below-average intelligence would meet the psychometric criterion for mild cognitive impairment (MCI; Petersen et al., 1999). Descriptive studies that highlight the likelihood of an isolated low memory score provide useful information on the potential to misdiagnose memory problems in older adults. For example, de Rotrou et al. (2005) reported that 48% of older adults, who were identified as having MCI at baseline based on the presence of a low memory score, had normal cognitive functioning at a 1-year follow-up.

The need to understand normal variability across a battery of neuropsychological measures, and thus the presence of some low scores in healthy people, should not be limited to adults and older adults. To our knowledge, information on the base rates of low scores on memory batteries in healthy children does not exist, even though performance on these batteries is used as the foundation for clinical inferences relating to memory problems. The purpose of this descriptive study was to illustrate that the principles of multivariate test

interpretation, which have been demonstrated in adults and older adults, are applicable to children and adolescents.

METHODS

Participants

Participants for the present study included 1000 healthy children and adolescents between 5 and 16 years of age (mean = 9.7, *SD* = 3.2) from the Children's Memory Scale (CMS; Cohen, 1997) standardization sample. An equal number of boys and girls were in each age group. Ethnicity included 68.4% White, 16.1% African American, 11.6% Hispanic, and 3.9% as "other." The sample was stratified by level of parent education (i.e., eighth grade or less, 9–11 years, high school graduate or equivalent, 1–3 years of technical school or college, and 4 or more years of college).

The standardization sample was recruited from 149 sites from the western, north central, northeastern, and southern regions of the United States. Children were excluded from the standardization sample if they were reading below their grade level, had repeated a grade, were receiving special education, were previously diagnosed with a neurological disorder, or had sustained an injury that would have put them at risk for having memory problems (Cohen, 1997). The treatment of participants and the collection of data were done in compliance with the Helsinki Declaration. Use of the archival CMS data was approved by the University of Calgary research ethics board.

A subsample of the CMS standardization group (*n* = 209) was administered the Wechsler Intelligence Scale for Children (Third Version) (WISC-III; Wechsler, 1991) as part of a linking study. The WISC-III linking sample ranged in age from 6 to 16 years (mean = 10.1, *SD* = 2.8), was 48.8% male, 77.0% White (11.0% African American, 9.6% Hispanic, 2.4% other), and had a mean WISC-III Full Scale Intelligence Quotient (FSIQ) of 105.3 (*SD* = 14.4, range = 70–146). Inclusion of the WISC-III linking data allows for stratification of the CMS data by level of intelligence (i.e., WISC-III FSIQ). Although the subsample with WISC-III FSIQ data had more people identified as Caucasian compared to the entire sample, $\chi^2(1) = 6.12$, $p = .01$, there were no statistically significant differences in age ($p = .08$), sex ($p = .75$), or performance on the CMS index scores (p values = .32–.71) for the WISC-III subsample compared to the total standardization sample.

Measures

There are six primary subtests on the CMS (i.e., Stories, Word Pairs, Dot Locations, Faces, Numbers, and Sequences). These six subtests contribute to eight age-adjusted index scores, six of which were included in the base rate of low scores analyses: Learning, Verbal Immediate, Verbal Delayed, Verbal Delayed Recognition, Visual Immediate, and Visual Delayed. The General Memory index and the Attention/Concentration index were not included in the analyses.

Analyses

Analyses involved examining performance on all six index scores, *simultaneously*. The cutoffs used for analyses of the CMS data included <16th percentile (<1 *SD*; index < 85), <10th percentile (index < 81), ≤5th percentile (index ≤ 76), and ≤2nd percentile (<2 *SD*s; index < 70). The prevalence of low CMS index scores was examined for the total sample (5–16 years; *N* = 1000) and for the three levels of intellectual abilities: below average (FSIQ = 70–89; *n* = 30), average (FSIQ = 90–109; *n* = 93), and above average (FSIQ = 110+; *n* = 86).

RESULTS

The base rates of low CMS index scores in children and adolescents are presented in Figures 1 and 2. Using the <1 *SD* cutoff score (Figure 1), one or more low index scores was found in 37.6% of the total sample and three or more low scores were found in 10.6% of the sample. One or more index scores <10th percentile was found in 30.2% (data not shown), one or more index scores ≤5th percentile was found in 22.4% (Figure 2), and one or more index scores <2 *SD*s was found in 12.4% of healthy children (data not shown). There were no substantial differences in the prevalence of low scores across the age bands. For example, the prevalence of one or more index scores ≤16th percentile included 47.1% of 5- to 8-year-olds, 43.2% of 9- to 13-year-olds, and 47.5% of 13- to 16-year-olds. These slight, but not substantial, variations were present across different cutoff scores and for different numbers of low scores.

Although age had limited impact on the base rates, intellectual functioning had considerable influence on these base rates. Compared to children with above-average intelligence, those with below-average intelligence were 7.1 times more likely, 95% confidence interval for odds ratio = 2.9–17.6; $\chi^2(1) = 19.79$, $p < .001$, to have one or more scores <1 *SD* (Figure 1). When considering ≤5th percentile (Figure 2), 33.3% of children and adolescents with below-average intelligence had one or more low index scores compared to 3.5% with above-average intelligence. In other words, children with below-average intelligence were 13.8 times more likely than those with above-average intelligence to have one or more index scores ≤5th percentile, 95% confidence interval for odds ratio = 3.7–50.86; $\chi^2(1) = 19.91$, $p < .001$. When considering a more stringent cutoff, such as <2 *SD*s (data not shown), 20.0% of children and adolescents with below-average intelligence had one or more low index scores compared to 2.0% with above-average intelligence, $\chi^2(1) = 10.82$, $p < .001$; odds ratio = 10.5 (95% confidence interval = 2.2–48.2).

DISCUSSION

It is common for healthy people to have some low scores across a battery of tests (Axelrod & Wall, 2007; Binder et al., 2009; Brooks et al., 2007, 2008; Crawford et al., 2007;
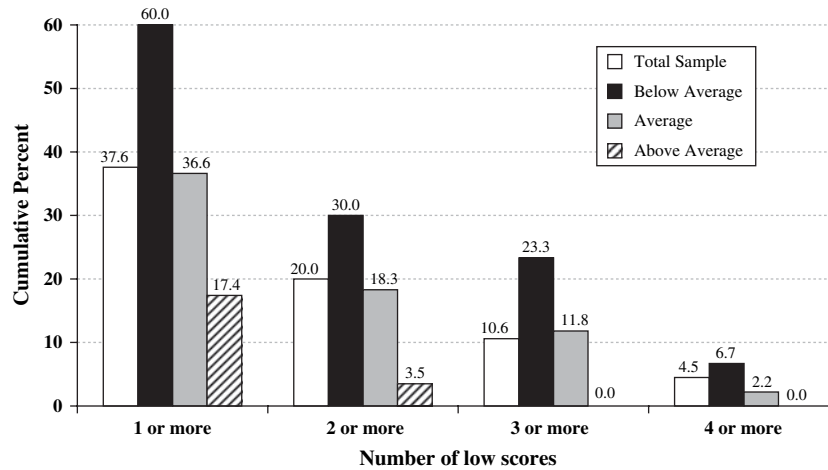
Fig. 1. Prevalence of low CMS index scores (<1 *SD*) by level of intelligence. Total sample *N* = 1000. Intelligence is based on WISC-III FSIQ scores and includes below average (FSIQ = 70–89; *n* = 30), average (FSIQ = 90–109; *n* = 93), and above average (FSIQ = 110+; *n* = 86). Analyses involved examining all index scores *simultaneously*. Analyses included six index scores (Learning, Verbal Immediate, Visual Immediate, Verbal Delayed, Verbal Delayed Recognition, and Visual Delayed). The Attention/Concentration and General Memory indexes were not included in the analyses. For index scores, <16th percentile or <1 *SD* is equal to an index <85. Standardization data are from the *Children's Memory Scale.* Copyright © 1997 by NCS Pearson, Inc. Used with permission. All rights reserved.

Heaton et al., 1991, 2004; Iverson & Brooks, in press; Iverson et al., 2008a, 2008b; Palmer et al., 1998; Schretlen et al., 2008). As a result, accumulating literature suggests that the psychometric interpretation of a test battery should include a multivariate approach. In other words, test scores should be interpreted simultaneously using empirical data because examining individual test scores can lead to overinterpretation of one or more isolated low scores (Brooks et al., 2007, 2008; de Rotrou et al., 2005; Palmer et al., 1998). To our knowledge, this is the first study that examines and presents the prevalence of low memory scores in a sample of healthy children and adolescents.

The results of this descriptive study clearly demonstrate that having at least one low memory score is common in many healthy children and adolescents. However, it would be considered *uncommon* (i.e., prevalence rate of approximately 10% or less) to have four or more index scores <1 *SD*. There were not any substantial differences in the prevalence of low CMS scores across the age groups. In other words, prevalence rates were fairly consistent from 5 to 16 years of age. This is likely the result of using age-adjusted standard scores. As the cutoff for identifying cognitive problems becomes more stringent, the number of low scores below those cutoffs considered uncommon also declines. For
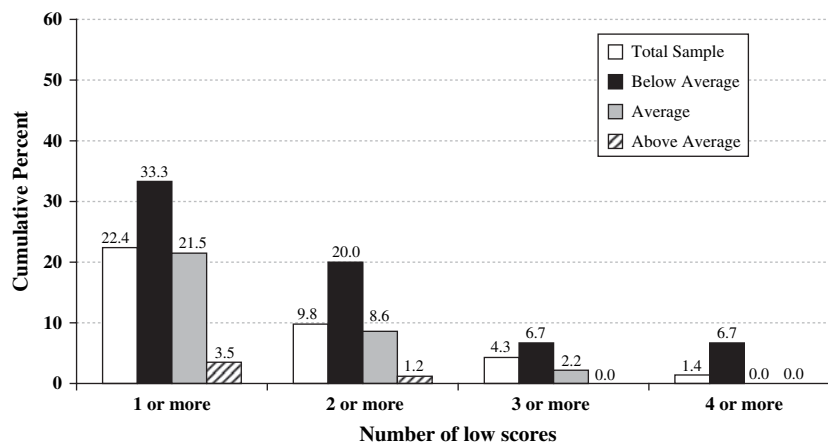

Fig. 2. Prevalence of low CMS index scores (≤5th percentile) by level of intelligence. Total sample *N* = 1000. Intelligence is based on WISC-III FSIQ scores and includes below average (FSIQ = 70–89; *n* = 30), average (FSIQ = 90–109; *n* = 93), and above average (FSIQ = 110+; *n* = 86). Analyses involved examining all index scores *simultaneously*. Analyses included six index scores (Learning, Verbal Immediate, Visual Immediate, Verbal Delayed, Verbal Delayed Recognition, and Visual Delayed). The Attention/Concentration and General Memory indexes were not included in the analyses. For index scores, ≤5th percentile is equal to an index ≤76. Standardization data are from the *Children's Memory Scale.* Copyright © 1997 by NCS Pearson, Inc. Used with permission. All rights reserved.

example, having one or more low index score is found in 37.6% when considering the 1 *SD* cutoff but is found in 22.4% when using the 5th percentile cutoff. Being able to determine the number of low CMS scores that would be considered uncommon is a strength of these analyses.

Interpretation of memory performance should be done in the context of overall level of intellectual functioning (Iverson & Brooks, in press). A person who is lower functioning cognitively will have more low scores, and be at greater risk for misdiagnosis of memory problems (i.e., false positives), than a person who is higher functioning (and at greater risk of having a missed diagnosis, i.e., false negatives). The number of low index scores (<1 *SD*) that would be considered uncommon (i.e., approximately 10% prevalence in the sample) was *four or more* for those with below-average intelligence, *three or more* in children with average intelligence, and *two or more* for those with above-average intelligence. Evidently, some caution should be exercised when interpreting test scores by level of intelligence in children because illness, injury, and/or developmental disorder may produce both intellectual and memory impairment in previously healthy children. Consequently, both low intelligence quotient (IQ) and a high prevalence low memory scores may be very clinically significant, depending on the presenting problem. Importantly, although multiple low memory scores may not be diagnostic of a specific disorder, the presence of such low scores, particularly in the context of normal intelligence, may indicate a relative weakness in mnemonic functions that is typical of some childhood developmental (e.g., memory for faces in developmental disorders; Williams et al., 2005) or neurological conditions (e.g., temporal lobe epilepsy).

The existing literature on the prevalence of isolated low memory scores across a battery of tests has focused on older adults (Brooks et al., 2007, 2008; de Rotrou et al., 2005; Palmer et al., 1998), particularly in the context of identifying memory changes consistent with prodromal dementia or MCI. The results of the present study, which considered similar analyses of base rates of low memory scores but in a pediatric standardization sample, were fairly similar (albeit slightly higher) to the results presented by Brooks et al. (2007) for the Neuropsychological Assessment Battery (NAB; Stern & White, 2003) Memory module and by Brooks et al. (2008) for the Wechsler Memory Scale–Third Edition (WMS-III; Wechsler, 1997). The similarities across the CMS, NAB Memory module, and WMS-III studies are also maintained when considering performance stratified by level of intelligence. The consistency in findings across different memory batteries with very different standardization samples suggests that the presence of low memory scores is not an artifact of any particular battery and is not attributable to a particular age group.

Iverson and Brooks (in press) suggested that clinicians and researchers should be familiar with the following five psychometric principles when interpreting multiple test scores. As suggested, low test scores (1) are common across all batteries, (2) depend on where cutoff scores are set, (3) depend on the number of tests administered, (4) vary by demographic characteristics of the examinee, and (5) vary by level of intelligence. The descriptive analyses presented in this article clearly illustrate principles 1, 2, and 5. It will be important for future research to examine how different demographic characteristics impact the prevalence of low scores and the interpretation of performance on any test battery. Although it is important for clinicians to understand that low scores are common and to consider these principles, it can be challenging to consider this information readily in everyday clinical practice unless easy-to-use interpretive tables and/or figures are readily available.

There are a few limitations to this study that warrant a brief discussion. First, like other studies involving a standardization sample, memory problems were not screened for *a priori* in the normative group. Although inclusion in the standardization was contingent on not having medical, neurological, or psychiatric conditions that could negatively impact memory performance, it is possible that some children and adolescents with memory problems were included. However, if a small proportion of the CMS standardization sample did have primary memory problems atypical of most healthy children, it is likely that this proportion would be quite small and unlikely to account for a large percentage of the prevalence rates presented in Figures 1 and 2. Second, the sample of children and adolescents who were part of the WISC-III linking study was small in size, consisted of relatively higher functioning youth, and the number of children with below-average intelligence was relatively small compared to the other intelligence groups.

The inclusion of the WISC-III (Wechsler, 1991) as the measure of intelligence warrants some discussion. Since the standardization and publication of the CMS in 1997, a newer version of the Wechsler Intelligence Scale for Children (WISC) has been published [e.g., Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV); Wechsler, 2003]. Although the correlation between the WISC-III and the WISC-IV FSIQ scores is high (*r* = .87; Wechsler, 2003), the WISC-IV yielded slightly lower scores compared to the WISC-III. However, given that the CMS sample was given the WISC-III approximately 6 years after it was normed and published in 1991, and it has been 6 years since the WISC-IV was normed and published (i.e., 2003), children's WISC-IV FSIQ scores today *might* be similar to the WISC-III FSIQ scores from the CMS standardization sample. Despite these similarities in time from publication of the respective WISC version, clinicians and researchers should use caution when interpreting the prevalence of low scores based on level of intelligence that is derived for a measure of intelligence other than the WISC-III. Users should also exercise some caution when interpreting the prevalence of low CMS scores for those with intelligence scores that are close to the cutoff for the different IQ ranges. It might be important to consider the base rates in the obtained classification as well as the base rates in the neighboring classification when drawing any conclusions on performance.

Knowing the prevalence of low scores in healthy children and adolescents is designed to supplement other psychometric interpretive methods (e.g., discrepancies between indexes) and clinical decision-making. The results of this study suggest that some caution is needed when interpreting isolated low CMS subtest scores as sole evidence of memory impairment. The goal is to use this information to reduce the likelihood of misdiagnosing memory problems. The lower prevalence of low scores in healthy children of above-average intelligence also holds potential in reducing the chances of a missed diagnosis of memory problems in children and adolescents who are higher functioning.

ACKNOWLEDGMENTS

REFERENCES

Axelrod, B.N. & Wall, J.R. (2007). Expectancy of impaired neuropsychological test scores in a non-clinical sample. *International Journal of Neuroscience*, *117*(11), 1591–1602.

Binder, LM., Iverson, GL., Brooks, BL. (2009). To err is human: "abnormal" neuropsychological scores and variability are common in healthy adults. *Archives of Clinical Neuropsychology*, *24*(1), 31–46.

Brooks, B.L., Iverson, G.L., Holdnack, J.A., & Feldman, H.H. (2008). Potential for misclassification of mild cognitive impairment: A study of memory scores on the Wechsler Memory Scale-III in healthy older adults. *Journal of the International Neuropsychological Society*, *14*(3), 463–478.

Brooks, B.L., Iverson, G.L., & White, T. (2007). Substantial risk of "Accidental MCI" in healthy older adults: Base rates of low memory scores in neuropsychological assessment. *Journal of the International Neuropsychological Society*, *13*(3), 490–500.

Cohen, M.J. (1997). *Children's Memory Scale, manual*. San Antonio, TX: The Psychological Corporation.

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.

Heaton, R.K., Grant, I., & Matthews, C.G. (1991). *Comprehensive norms for an extended Halstead-Reitan Battery: Demographic corrections, research findings, and clinical applications*. Odessa, FL: Psychological Assessment Resources, Inc.

Heaton, R.K., Miller, S.W., Taylor, M.J., & Grant, I. (2004). *Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults professional manual*. Lutz, FL: Psychological Assessment Resources.

Iverson, G.L. & Brooks, B.L. (in press). Improving accuracy for identifying cognitive impairment. In M.R. Schoenberg & J.G. Scott (Eds.), *The black book of neuropsychology: A syndrome-based approach*.

Iverson, G.L., Brooks, B.L., & Holdnack, J.A. (2008a). Misdiagnosis of cognitive impairment in forensic neuropsychology. In R.L. Heilbronner (Ed.), *Neuropsychology in the courtroom: Expert analysis of reports and testimony*, (pp. 243–266). New York: Guilford Press.

Iverson, G.L., Brooks, B.L., White, T., & Stern, R.A. (2008b). Neuropsychological Assessment Battery (NAB): Introduction and advanced interpretation. In A.M. Horton, Jr. & D. Wedding (Eds.), *The neuropsychology handbook*, (pp. 279–343). New York: Springer Publishing, Inc.

Palmer, B.W., Boone, K.B., Lesser, I.M., & Wohl, M.A. (1998). Base rates of "impaired" neuropsychological test performance among healthy older adults. *Archives of Clinical Neuropsychology*, *13*(6), 503–511.

Petersen, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G., & Kokmen, E. (1999). Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology*, *56*(3), 303–308.

de Rotrou, J., Wenisch, E., Chausson, C., Dray, F., Faucounau, V., & Rigaud, A.S. (2005). Accidental MCI in healthy subjects: A prospective longitudinal study. *European Journal of Neurology*, *12*(11), 879–885.

Schretlen, D.J., Testa, S.M., Winicki, J.M., Pearlson, G.D., & Gordon, B. (2008). Frequency and bases of abnormal performance by healthy adults on neuropsychological testing. *Journal of the International Neuropsychological Society*, *14*(3), 436–445.

Stern, R.A. & White, T. (2003). *Neuropsychological assessment battery*. Lutz, FL: Psychological Assessment Resources.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—Third edition*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997). *Wechsler Memory Scale—Third edition*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth edition*. San Antonio, TX: The Psychological Corporation.

Williams, D.L., Goldstein, G., & Minshew, N.J. (2005). Impaired memory for faces and social scenes in autism: Clinical implications of memory dysfunction. *Archives of Clinical Neuropsychology*, *20*(1), 1–15.