Subclinical hypothyroidism and digit span test performance in children: a systematic review and meta-analysis

Nadhea Debrinita Surya¹, Nicolas Daniel Widjanarko¹, Theressa Kristiayu Permatasari¹, Yuliana Yosephine¹, Ellen Wijaya²

Abstract

Background The digit span (DS) test is one of the most commonly used neuropsychological tests to assess certain cognitive domains, i.e., short-term verbal memory, working memory, and attention. Subclinical hypothyroidism (SH) may be associated with a risk of cognitive impairment in children and adolescents.

Objective To evaluate the association between SH and DS test performance in children.

Methods Eligible studies evaluating SH and DS test performance were included in this systematic review and further assessed for risk of bias using the Newcastle Ottawa Scale. We carried out a meta-analysis using the random effects model to determine mean difference with 95% confidence interval (95%CI) for continuous data. This systematic review was conducted according to the PRISMA statement.

Results Out of 1,511 participants in the five included studies, 129 had SH and 1,382 were euthyroid. The quality of all studies were fair to good. Three studies were extracted for meta-analysis, with results showing a trend toward a poorer DS test performance in the SH group compared to controls, although this difference was statistically insignificant (IV -0.57; 95%CI -1.61 to 0.46; P=0.28). There was no significant heterogeneity among the included studies (I²=0%; P=0.69).

Conclusion No significant association was noted between SH and the domains of cognitive function assessed using the DS test. Several intrinsic and extrinsic factors and inability of the DS test to detect subtle impairment may limit its usefulness in children.

Keywords: subclinical hypothyroidism; cognitive function; digit span test; children

Digit span (DS) is a subtest of the Wechsler Adult Intelligence Scale (WAIS), Wechsler Intelligence Scale for Children (WISC), and Wechsler Memory Scale (WMS). DS is one of the most commonly used neuropsychological tests to assess certain cognitive domains in the Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5), i.e., short-term verbal memory, working memory, and attention in adults and children.¹⁻³ The superiority of DS may be due to the digits being sampled from a smaller pool, unlike other stimuli such as letters and words. As such, digits would be easier to recall compared to other stimuli.⁴ Subclinical hypothyroidism (SH) or mild thyroid failure is defined as a condition marked with elevated serum thyroid-stimulating hormone (TSH), compared to normal peripheral thyroid hormone levels. While 3-8% of the general population have SH, children have a reportedly lower prevalence (<2%).⁶ The role of SH in cognitive performance is

From School of Medicine and Health Sciences¹ and Department of Pediatrics, School of Medicine and Health Sciences², Universitas Katolik Atma Jaya, Jakarta, Indonesia.

Corresponding author: Ellen Wijaya, Department of Pediatrics, School of Medicine and Health Sciences, Universitas Katolik Atma Jaya, Jalan Pluit Raya No. 2, North Jakarta 14440, Indonesia. Telp: (+62) 669.3168. Email: ellen.wijaya@atmajaya.ac.id.

still controversial. Some studies suggest that untreated SH in children can interfere with central nervous system development due to its effect on growth, maturation, and myelination of nerve cells, and in later age could lead to several cardiovascular and metabolic complications. Other studies reported that mild hypothyroidism did not have a major effect on cognitive function in adults, however, there is growing evidence that SH may be associated with a risk of cognitive impairment in children and adolescents. Lack of initiative, inability to concentrate, impaired recall, and short-term memory deficits are common cognitive symptoms found in SH, suggesting that memory and executive function are the most affected domains. To the best of our knowledge, no systematic review has been done to evaluate the association between DS performance and SH in children. Hence, we aimed to investigate a possible association between SH and several domains of cognitive function in children based on their DS test performance.

Methods

This systematic review was designed and conducted based on guidelines from the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement. Studies were considered eligible if they met the following criteria: cross-sectional, case-control, or cohort studies that discussed DS test performance in children with SH (with well-defined diagnosis criteria), were published in English from 2000-2020, with populations consisting of children aged 1-18 years. Reviews, case-reports, case studies, non-human studies, and abstract-only articles were excluded.

We searched for eligible studies with major medical subject headings (MeSH) “subclinical hypothyroidism” as the independent variable and “Digit Span,” a subtest of WISC, as the dependent variable, in four main databases: PubMed, EBSCOHost, Google Scholar, and ProQuest. All steps were done systematically according to PRISMA 2009. Below is an example of keyword nomenclature used in our literature search using the PubMed Search Database: (("subclinical hypothyroidism") OR ("mild hypothyroidism") OR ("compensated hypothyroidism")) AND ((children) OR (pediatric)) AND ((learning) OR (memory) OR ("learning and memory") OR (attention) OR ("complex attention") OR ("executive function") OR ("digit span") OR ("Wechsler Intelligence Scale for Children") OR ("WISC") OR ("intelligence quotient") OR ("IQ").

All search results were downloaded using Mendeley version 1.19.4 (Elsevier BV, Amsterdam) as a reference manager. All five authors independently screened and reviewed titles and abstracts of the collected studies. Studies were excluded if the title and/or abstract exhibited irrelevance to cognitive function and SH. Full-text assessment was performed by all five reviewers and any disagreement was resolved by consensus. Data extraction was completed by all five reviewers. Detailed information recorded for each study included the first author’s name, type of study, publication year, number of participants, population baseline characteristics (age and sex), and diagnostic criteria for SH.

SH was diagnosed when there was a mild increase of TSH with normal free T4 (fT4) level from blood specimens, but the cut-off was determined by each study. DS, a subtest of WAIS, WISC, and WMS, was assessed by the examiners in each study, either blinded or unblinded. For the DS Forward test, the examiner read a group of numbers aloud, then the child repeated the numbers in the same order to the best of his or her ability. For the DS Backward test, the child repeated the numbers in reverse order to those read aloud by the examiner. Higher DS scores indicated better performance in cognitive domains. The DS scores of each study were presented as mean and standard deviation. Five reviewers independently appraised each included study using the Newcastle-Ottawa Scale (NOS) to assess the risk of bias according to the main categories of selection, comparability, and outcome (cohort or cross-sectional studies) or exposure (case-control studies). The assessment scores ranged from zero to nine stars, with higher scores representing better study quality. Discrepancies were resolved by discussion to achieve consensus among the five reviewers.

The main outcomes were tabulated as univariate and bivariate results. Univariate results are presented as demographic characteristics of each study, i.e., age and number of participants, country where the study was conducted, type of study, and SH diagnostic criteria. The DS test results for the SH and control
groups, as well as the P value in each study were shown in a bivariate data table. The results of the DS test for both the SH and control groups were tabulated and reported as continuous data in the form of mean and standard deviations with P values.

The meta-analysis was conducted on eligible studies using Review Manager version 5.4 (Cochrane, London). Narrative synthesis was performed for the remaining studies. Quantitative analysis was carried out with 95%CI. Heterogeneity analysis was done using the I2 test. An I2 of ≥50% was considered to indicate high heterogeneity, hence a random effects model was used because the results could be generalized to other random populations in new studies. Random effect size was represented as mean difference for continuous variable data with 95%CI.

Results

A total of 1,569 studies were obtained in our first literature search, out of which 1,391 remained after removing duplicates. Out of 1,382 articles, we screened the title and abstract. We excluded studies that failed to meet the population or exposure of interest criteria, such as studies with inappropriate design, participant age group, or year of publication. We performed full-text screening on the selected eight studies, out of which three were excluded (one study did not provide the full-text article and two studies failed to meet the population and exposure of interest criteria). Therefore, five studies were included in our qualitative synthesis and three studies with available data (mean difference and SD) were extracted for meta-analysis (Figure 1).

![Figure 1. PRISMA 2009 diagram of literature search and study selection](https://example.com/figure1.png)
Table 1 lists the characteristics of all five studies that were included in this systematic review; three were case-control and two were cross-sectional studies. Out of a total of 1,511 participants in the five included studies, 129 had SH and 1,382 were euthyroid. The SH was diagnosed by elevated TSH level and normal FT4 level, but the upper reference TSH limits differed between studies.

Assessment of the risk of bias in the five studies was performed using the NOS (Tables 2a and 2b). All studies were observational, consisting of four case-control and one cross-sectional study (Table 1). Of the four case-control studies, two were considered good quality, and the other two were considered fair quality. The cross-sectional study was rated as good quality.

Table 3 shows a summary of the primary outcomes. Almost all studies reported no significant difference in DS test scores between the SH and euthyroid groups. Although one study stated that DS test scores were lower in the SH group, two other studies stated the opposite. One study reported a significant difference in the two groups based on the study P value. Two out of the five studies did not present their data in the form of means and SDs: Cerbone et al. presented the DS test result in a histogram, while Ergur et al. presented their data as T-scores. Therefore, these two studies were not eligible for meta-analysis.

We performed meta-analysis of the three studies with complete and sufficient data (Figure 2). The random effects analysis showed a lower mean DS test result in the SH group, but the difference with the euthyroid group was not statistically significant (IV = -0.57; 95%CI -1.61 to 0.46; P = 0.28). We did not find significant heterogeneity between the three studies (I² = 0%; P = 0.69). The funnel plot to analyze for an association between SH and DS test performance in children with SH was symmetrical, as shown in Figure 3, suggesting no evidence of publication bias.

Discussion

The DSM-5 divides cognitive function into several domains: complex attention, executive function, learning and memory, language, perceptual motor, and social cognition. In several cross-sectional and case-control studies, children with SH tended to have impaired cognitive function. The domains of memory, attention, and executive function were the most frequently affected. Learning is an acquisition of knowledge and skill, while memory is the expression of what a person have accepted. Attention is a complex concept that refers to ability to obtain and maintain focus to a task; maintaining attention is essential to learning and doing everyday tasks. Executive function refers to a mental process of playing with ideas, staying focused, thinking before acting, meeting unanticipated challenges, and resisting temptation. Without minimizing the contributions of other cognitive domains, the domains of memory, complex attention, and executive function are very important in the process of one's cognitive, social, and psychological development.

The DS test is superior to word span tests because digits are more frequently used compared to other verbal materials. In a previous study, forward and backward DS tests had AUC scores of 0.82 and 0.79, respectively, suggesting excellent diagnostic accuracy. But the sensitivities of the forward and backward DS test were considerably low (41% and 51%, respectively), although the specificities were high (93% and 91%, respectively). Nevertheless, the DS test has good utility in detecting subtle cognitive performance in relatively high-functioning older children and adolescents.

In our meta-analysis, we found no significant association between SH and DS test performance in children. A previous study reported that SH patients had altered intrinsic resting-state functional connectivity within the somatomotor network and right frontoparietal attention network. Hence, SH might be associated with attenuated motor, working memory, attention, and executive cognitive functions. As mentioned, the DS test is able to assess these three cognitive domains. It is easy to use in children. The DS test is performed by the test-giver reading a group of numbers to the subject, which they are required to recall, sequence, and vocalize either in consecutive order in a forward span, or in reverse order in a backward span. The forward DS test can be used to reflect a subject’s attention efficiency and capacity, whilst the backward DS test relies on executive tasks, specifically working memory. DS is considered to be superior to word span tests because digits are more frequently used compared to other verbal materials. In a previous study, forward and backward DS tests had AUC scores of 0.82 and 0.79, respectively, suggesting excellent diagnostic accuracy. But the sensitivities of the forward and backward DS test were considerably low (41% and 51%, respectively), although the specificities were high (93% and 91%, respectively). Nevertheless, the DS test has good utility in detecting subtle cognitive performance in relatively high-functioning older children and adolescents.

In our meta-analysis, we found no significant association between SH and DS test performance in children. A previous study reported that SH patients had altered intrinsic resting-state functional connectivity within the somatomotor network and right frontoparietal attention network. Hence, SH might be associated with attenuated motor, working memory, attention, and executive cognitive functions. As mentioned, the DS test is able to assess a subject’s executive function, learning and memory, and
<table>
<thead>
<tr>
<th>Author</th>
<th>Type of study</th>
<th>Number of participants</th>
<th>Population age, years</th>
<th>Country</th>
<th>SH diagnostic criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atli et al.15 (2019)</td>
<td>Case-control</td>
<td>20/20</td>
<td>12.8 ± 3.5/13.6 ± 2.7</td>
<td>Turkey</td>
<td>4.94-20 μIU/L (not stated in article)</td>
</tr>
<tr>
<td>Capalbo et al.16 (2020)</td>
<td>Case-control</td>
<td>34/34</td>
<td>9.1±2.6/9.2±2.8</td>
<td>Italy</td>
<td>5.0-9.9 mIU/L 0.75-1.7 ng/dL</td>
</tr>
<tr>
<td>Cerbone et al.17 (2011)</td>
<td>Case-control</td>
<td>36/36</td>
<td>9.7±0.6/9.5±0.5</td>
<td>Turkey</td>
<td>4.2-10 mIU/L 9-26 pmol/L</td>
</tr>
<tr>
<td>Ergur et al.13 (2012)</td>
<td>Case-control</td>
<td>17/17</td>
<td>9.92±3.21/9.66±1.36</td>
<td>Turkey</td>
<td>5-25mIU/L (not stated in article)</td>
</tr>
<tr>
<td>Wu et al.18 (2006)</td>
<td>Cross-sectional</td>
<td>22,1,275</td>
<td>13-16/13-16</td>
<td>United States</td>
<td>&gt; 4.6 mIU/L ≥ 4.5-13.2 μg/dL</td>
</tr>
</tbody>
</table>

**Table 2a. Risk of bias assessment of the included case-control studies**

<table>
<thead>
<tr>
<th>Author</th>
<th>Adequate case definition</th>
<th>Representativeness of the cases</th>
<th>Selection of controls</th>
<th>Definition of controls</th>
<th>Comparability of cases and controls on the basis of the design or analysis</th>
<th>Ascertainment of exposure</th>
<th>Same method of ascertainment for cases and controls</th>
<th>Non-response rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atli et al.15 (2019)</td>
<td>Yes, with independent T validation</td>
<td>Potential for selection biases or not stated</td>
<td>No description</td>
<td>No history of disease (endpoint)</td>
<td>Study controls for age</td>
<td>Structured interview blinded to case/control status</td>
<td>Yes</td>
<td>Same rate for both group</td>
</tr>
<tr>
<td>Capalbo et al.16 (2020)</td>
<td>Yes, with independent T validation</td>
<td>Potential for selection biases or not stated</td>
<td>Community controls</td>
<td>No history of disease (endpoint)</td>
<td>Study controls for age, sex, socio-economic status, and geographic region</td>
<td>Structured interview blinded to case/control status</td>
<td>Yes</td>
<td>Same rate for both group</td>
</tr>
<tr>
<td>Cerbone et al.17 (2011)</td>
<td>Yes, with independent validation</td>
<td>Potential for selection biases or not stated</td>
<td>Community controls</td>
<td>No history of disease (endpoint)</td>
<td>Study controls for age, sex, pubertal, socio-economic status, and geographic region</td>
<td>Structured interview blinded to case/control status</td>
<td>Yes</td>
<td>Same rate for both group</td>
</tr>
</tbody>
</table>

**Table 2b. Risk of bias assessment of the included cross-sectional studies**

<table>
<thead>
<tr>
<th>Author</th>
<th>Representativeness of the sample</th>
<th>Sample size</th>
<th>Non-respondents</th>
<th>Ascertainment of the exposure (risk factors)</th>
<th>Comparability</th>
<th>Assessment of outcome</th>
<th>Statistical test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wu et al.18 (2006)</td>
<td>Truly representative of the average in the target population</td>
<td>22,1,275</td>
<td>Compatibility between respondent and non-respondent characteristics was established, and the response rate was satisfactory</td>
<td>Validated measurement tool</td>
<td>Data/results adjusted for relevant predictors/risk factors/confounders</td>
<td>Unblinded assessment using objective validated laboratory methods</td>
<td>Statistical test not described</td>
</tr>
</tbody>
</table>
Table 3. Results of studies included in the systematic review

<table>
<thead>
<tr>
<th>Author</th>
<th>DS results for SH group</th>
<th>DS results for euthyroid group</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td></td>
</tr>
<tr>
<td>Atli et al.¹⁵</td>
<td>10 (3.57)</td>
<td>9.78 (3.82)</td>
<td>NS (&gt;0.05)</td>
</tr>
<tr>
<td>Capalbo et al.¹⁶</td>
<td>9.3 (2.3)</td>
<td>10.1 (2.6)</td>
<td>NS (&gt;0.05)</td>
</tr>
<tr>
<td>Wu et al.¹⁸</td>
<td>9.8 (25.14)</td>
<td>8.49 (4.01)</td>
<td>NS (&gt;0.01)</td>
</tr>
<tr>
<td>Cerbone et al.¹⁷</td>
<td>N/A* (??)</td>
<td>N/A* (N/A*)</td>
<td>NS (&gt;0.05)</td>
</tr>
<tr>
<td>Ergur et al.¹³</td>
<td>N/A** (N/A**)</td>
<td>N/A** (N/A**)</td>
<td>0.001</td>
</tr>
</tbody>
</table>

*data presented in histogram; **data presented in T-score; NS=not significant; N/A=not available

Figure 2. Forest plot of the mean difference in DS test performance between SH and euthyroid subjects. The mean differences in individual studies are represented by green squares; the pooled mean difference is represented by the black square.

Figure 3. Funnel plot of mean difference values and standard errors (SE) in the three included studies.
complex attention, thus, this might explain findings by Wu et al.\textsuperscript{18} reporting a significant difference in DS test performance between SH and euthyroid subjects. Using positron emission tomography (PET), Bauer et al.\textsuperscript{26} found that patients with untreated SH have lower regional glucose metabolism in the anterior cingulate cortex of the brain. This specific brain area has a crucial role in several cognitive processes: empathy, impulse control, emotion, and decision-making. Similarly, Zhu et al.\textsuperscript{27} reported that working memory, assessed using the digit n-back task, was impaired in adult SH patients compared to the euthyroid group. Further investigations revealed abnormal findings detected by functional magnetic resonance imaging (fMRI), predominantly in the frontal brain areas responsible for decision-making. Executive function, including decision-making ability, is also part of cognitive function measured with DS test, and showed different results in patients with SH.

There are several possible reasons for the lack of a significant association between SH and DS test performance in our study. First, although neuropsychological tests, including the DS test, might detect subtle impairment in cognitive function domain, cognitive dysfunction associated with SH might be too insignificant to be detected by neuropsychological test. SH has been theorized to be in the initial, mild part of the hypothyroidism spectrum, which may result in subtle cognitive attenuation.\textsuperscript{28} Cognitive dysfunction becomes more apparent if SH progresses into overt hypothyroidism, since thyroid hormone plays an important role in normal brain development by regulating neurogenesis, myelination, dendrite proliferation, and synapse formation.\textsuperscript{29} A previous study noted that hypothyroidism, especially congenital hypothyroidism, can lead to cognitive and behavioral deficits in children and adolescents.\textsuperscript{13} In SH, fT4 level is within normal range, while TSH level is above the normal range. We suggest that it is possible that the DS test scores did not differ significantly because fT4 was within normal range in both the SH and euthyroid groups. In addition, neuropsychological test scores can also be influenced by unfavorable extrinsic environmental factors and intrinsic factors of the test-taker, such as enthusiasm, task engagement, stress, and motivation to perform the test.\textsuperscript{15}

Given that children with SH have a higher serum TSH concentration, Cui et al. found that there was no association between TSH concentration and urine iodine concentration with IQ performance in a pediatric population.\textsuperscript{30} In addition, a study reported no association between TSH concentration and any cognitive or behavioral outcome. However, a strong association was found when comparing mood and attention problems between subjects with low and normal fT4.\textsuperscript{31} Another study reported higher mean serum free T3 (fT3) and TSH in children with attention deficit/hyperactivity disorder (ADHD) compared to a control group.\textsuperscript{32} However, the correlation between TSH level and ADHD remains unclear.

There were several limitations in our study. The number of included studies was limited and had relatively small sample sizes. Moreover, studies were restricted to English language and grey literature searching was not performed. Furthermore, we did not control for other factors that may affect thyroid function, such as the presence of autoantibodies or iodine level. However, despite the limitations, our study was the first systematic review and meta-analysis to evaluate for an association between SH and DS test performance in children. Additional large, high-quality studies are needed for a more robust analysis. We recommend future studies with continuous follow-up to investigate a potential long-term association between SH and DS test performance and cognitive function.

In conclusion, we found no significant association between SH and the domains of cognitive function assessed using the DS test, including learning and memory, complex attention, and executive function. In the future, larger, high-quality studies are needed, as well as cohort studies to further investigate the association between SH and DS test performance in children, for eventual systematic reviews and meta-analyses. SH in children should also be routinely monitored to avoid progression and any other adverse effects.

\textbf{Conflict of interest}

None declared.
Acknowledgements

We would like to thank our colleagues from the Atma Jaya Catholic University of Indonesia, School of Medicine and Health Sciences, for their help, guidance, and support.

Funding acknowledgement

The authors received no specific grants from any funding agency in the public, commercial, or not-for-profit sectors.

References

14. 9.5.2 Identifying and measuring heterogeneity [Internet]. [cited 2021 Jan 27]. Available from: https://handbook-5-1.cochrane.org/chapter_9/9_5_2_identifying_and_measuring_heterogeneity.htm.
learning.


