Does Hydration Impact Memory: A Systematic Review

Kelsey Ellis-Lepard
Does Hydration Status Impact Memory: a Systematic Review
Kelsey Ellis-Lepard
I. Introduction

With 60% of the adult human body composed of water it makes sense that maintaining proper hydration levels is important to survival as well as living a healthy life. Water has many roles throughout the human body: thermoregulation, being a carrier, being a lubricant, and acting as a reaction medium. As a carrier, water is responsible for allowing exchanges between cells, interstitial fluid and capillaries. In addition, water regulates the blood volume and allows blood circulation. Since water is responsible for these tasks, many systems in the body as well as the brain are reliant upon proper hydration levels to function properly.8

As the brain is reliant upon proper hydration levels for optimal function, severe dehydration can ultimately lead to delirium, and death. Due to these consequences of dehydration, it is obvious that a deterioration of cognition occurs prior to coma and death.9 Although it is clear that this deterioration occurs, it is unknown what severity and duration of dehydration will cause initial decrements as well as which aspects of cognition are impacted.9

Cognition is defined as “the mental action or process of acquiring knowledge and understanding through thought, experience, and the senses.”6 Cognition is complex and includes factors of attention, learning, memory, and reasoning.1,9 There are multiple assessments for each of these specific functions, which results in a wide variety of test batteries to assess cognition as a whole. Due to the overwhelming amount of available test batteries as well as the disagreement on the precision, validity and sensitivity of the testing batteries, assessing cognitive performance can be complex and difficult.1,9 Instead of looking into cognition as a whole, this review will focus on analyzing the available literature on the impact of hydration on memory.

II. Review of Literature
Hydration

Being properly hydrated is a necessity—not only to survive, but to live a healthy life. It is obvious that severe dehydration impairs a person's cognitive abilities, but it has been shown that even mild dehydration impairs cognitive function.\textsuperscript{3,12} Currently the physiological cause of cognitive impairments when a person is dehydrated is not known. Armstrong et al.\textsuperscript{3} showed that when the body detects dehydration, higher-order cortical brain regions that regulate mood are signaled resulting in adverse mood and symptoms. These authors suggest that due to the critical role hydration plays in overall health, the brain uses adverse moods and other symptoms as a way of communicating that they need to rehydrate before more severe consequences occur due to their dehydration. On the other hand, Faraco et al.\textsuperscript{13} performed a study that suggests that dehydration results in a significant disruption in the ability of the neurons, endothelial cells, and smooth muscle cells to regulate cerebral blood flow. Using a Y maze, the cerebrovascular dysfunction that was brought on by dehydration (key mechanisms of cerebral circulation being impaired leading to vascular reserves being reduced) was linked to cognitive dysfunction.

Hydration Testing

When it comes to testing hydration status there are many methods that provide researchers with an idea of the subject's hydration status. Some of the ways of discovering hydration state is by a change in body mass, urinary and haematological indices.\textsuperscript{14}

Body mass change is the easiest way to assess the amount of water a person has lost during physical exercise. Water makes up roughly 60\% of the human body and no other component is lost at such a rate; because of this, it is commonly assumed that the mass lost is the amount of water lost through sweating during physical activity. Using changes in body mass to determine hydration is an easy way to get an estimate of how much water has been lost from the body since
it is simple, quick, and doesn’t require training or expertise to carry out this test. Differences in body mass are good for detecting the change in hydration status for short periods of time such as the length of time it takes to complete a workout, but due to factors such as food and fluid intake, bowel movements, and urine production, it is not a suitable method for tracking hydration status over a longer period of time. A limitation of using body weight is that there needs to be a three-day baseline in order to determine a standard mass for an individual.

Hydration levels can be assessed through urine in multiple ways: urine color, urine specific gravity (USG), and urine osmolality. When testing hydration using urine color, the color of the urine sample is ranked on a scale of colors (hydrationcheck.com) ranging from 1 to 8; 1 being the palest shade and 8 being the darkest. Testing urine color is a convenient method due to its ease; it is not expensive, is noninvasive, and like body weight changes, there is no need for expertise to execute this test. Also, similar to body weight changes, urine color can undergo changes from variables other than hydration such as diet, illness, and medications. Urine osmolality is another noninvasive way to test for hydration status. Osmolality is measured with a freezing-point or vapor pressure-depression osmometer and has a physiological range from 50-1400 mOsm/kg. Although osmolality increases with dehydration, it is hard to identify a universal value as a cutoff. Despite cultural differences amongst the populations, scientists have determined a cut point of 700-800 mOsm/kg as diagnostic for hypohydration. Urine osmolality and specific gravity are closely related, they are equal in accuracy, specificity, and sensitivity. Specific gravity measures the density of urine compared with that of an equal volume of distilled water and is measured using a refractometer. A normal urine sample will typically have a specific gravity value between 1.013 and 1.029, while anything over 1.020-1.025 is considered to be hypohydrated. For all urine indices, the urine sample provided should be the first urine
sample of the day (once the subject has woken up and before they eat breakfast) to ensure an accurate assessment of hydration status.\textsuperscript{19,20}

To summarize, the three most popular urine indices are urine color, urine osmolality, and urine specific gravity. Armstrong\textsuperscript{21} compared the three methods (Table 1).

\textbf{Table 1}\textsuperscript{15}

<table>
<thead>
<tr>
<th>Testing Method</th>
<th>Cost</th>
<th>Time Required to Perform Test</th>
<th>Expertise Required to Perform Test</th>
<th>Likelihood of an Adverse Event Occurring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urine Osmolality</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Urine Color</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Urine Specific Gravity</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

1: low 2: moderate 3: high

As shown in Table 1, all three urine indices provide accurate measurements for hydration status, but all do have the potential to be affected by variables such as severe dehydration, severe illness, alcohol consumption, and rapid rehydration.\textsuperscript{14,15}

Hydration status can also be assessed through haematological indices. For this test, changes in haemoglobin and haematocrit concentration can be examined, which represent change in plasma volume, for an indication of hydration status. This method for determining hydration status is extremely reliable if there are standard baseline reliable levels of haemoglobin and haematocrit available and they are accurate. The downfall of this method is that there must be baseline data and it can be affected by factors such as posture and use of a tourniquet when gathering the blood sample.\textsuperscript{14} Another haematological method to assess hydration status is
through plasma osmolality, which is measured using a freezing point depression osmometer. This method provides a measurement of hydration status with superior accuracy when in a controlled laboratory setting. Some limitations of using this method to test for hydration status is that plasma osmolality levels change due to a number of variables and that it is considered invasive for the participant due to needing a blood sample.

Memory

Memory is defined as the faculty by which the mind stores and remembers information. Despite over a hundred years of research, scientists are perplexed by the mechanism of memory. In a forum, Poo et al. contributed different ideas as to what exactly memory is. One theory is that an ensemble of brain cells is activated when a person is learning. These brain cells are induced with persistent physical and chemical changes. Relevant recall cues later in life will result in the reactivation of these cells which will then result in the retrieval of that memory. An opposing theory suggests that synapses in the brain are the building blocks of memory rather than brain cells. This theory suggests that stimuli which mimic learning events cause a change, and addition of dendritic spines, at the synaptic level, and since these spines are altered when learning occurs and they are located in the hippocampus, which deals with memory, then this is the key to understanding memory and its mechanism.

III. Methods

Data Sources

The following terms: hydration, memory, dehydration, and cognitive performance were used to search the electronic database, PubMed, SPORTDiscus, and CINAHL. The search was limited to research articles in English with human participants. The articles identified, and their sources were cross-referenced to find other potentially relevant studies. This search resulted in
24 research articles related to the impact of hydration on memory. Articles included had a crossover design, review articles were not included in this review.

Of the 24 articles identified, 12 met the inclusion criteria of: participants aged 18-30 years, hydration was controlled, there was control of medication that impacts fluid balance, memory was a main outcome measure. In order to be considered for inclusion, the investigators had to control the hydration level of the participants. The investigators induced the desired level of hydration through varied methods (heat, exercise, or a combination). Hydration also had to be quantified through a percentage change in body mass. Medication also needed to be controlled for possible inclusion. Participants taking medications that could possibly alter fluid-electrolyte balance, thermoregulation, or inhibit sweating were not included. Women taking oral contraceptives were included.

**Quality Assessment**

The 12 articles that met inclusion criteria were then evaluated on the Physiotherapy Evidence Database (PEDro) Scale (Appendix). Articles were evaluated independently by 2 investigators who then met to settle any discrepancies. The articles ranged from 4 to 9 on the PEDro scale. Articles that received a score of 4 or less showed a lack of blinding and lack of random allocation. Seven articles were left focusing on the impact of hydration on memory with a PEDro rating of 5 or greater (Figure 1).
IV. Results

The primary focus of this review was to analyze the available literature that controlled and quantified hydration levels and included memory as a main outcome, to determine if there is an impact of varying hydration status on memory (Table 2).

Armstrong et al\textsuperscript{3} conducted a study in which memory was tested after three conditions: exercise-induced dehydration plus diuretic (DD), exercise-induced dehydration plus placebo (DN), and euhydration plus placebo (EU). The DD and DN conditions resulted in a minimum loss of 1% body mass. Memory was tested using two test batteries: a matching to sample test and a repeated acquisition test. This study found no significant differences in memory when comparing the dehydrated trials to the euhydrated trials.\textsuperscript{3}
<table>
<thead>
<tr>
<th>Article</th>
<th>Participants</th>
<th>Methods</th>
<th>Dehydration Level</th>
<th>Dehydration Method</th>
<th>Heat Protocol (if applicable)</th>
<th>Exercise Protocol (if applicable)</th>
<th>Difference in Memory</th>
<th>Memory Testing</th>
<th>PEDro Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Armstrong, L.E.</td>
<td>n = 25 females</td>
<td>randomized, double blind, crossover</td>
<td>21%</td>
<td>exercise + placebo, exercise diuretic, euhydration + placebo</td>
<td>N/A</td>
<td>Treadmill, 40 min, 5.6 km/h</td>
<td>No significant difference</td>
<td>Matching to sample</td>
<td>9</td>
</tr>
<tr>
<td>Sharma, V.M.</td>
<td>n = 8 males</td>
<td>randomized, crossover study</td>
<td>1%, 2%, 3%</td>
<td>continuous work in heat chamber</td>
<td>TN: 27°C, 50%</td>
<td>15 steps per minute on</td>
<td>TN: decrement in 2% (-18.72%) &amp; 3% (-18.09%)</td>
<td>Concentration test</td>
<td>5</td>
</tr>
<tr>
<td>Gan, C. (2001)</td>
<td>n = 7 males</td>
<td>randomized, crossover study</td>
<td>up to 2.8%</td>
<td>passive heat exposure, exercise</td>
<td>alternate 45°C, 70% and 50°C, 20%</td>
<td>Treadmill, 65% VO2</td>
<td>-11.27% difference in short-term memory</td>
<td>Long-term memory</td>
<td>5</td>
</tr>
<tr>
<td>Ganio, M.S.</td>
<td>n = 26 males</td>
<td>randomized, double blind, crossover</td>
<td>21%</td>
<td>exercise + placebo, exercise diuretic, euhydration + placebo</td>
<td>N/A</td>
<td>Treadmill, 60 min, 5.6 km/h</td>
<td>No significant difference</td>
<td>Matching to sample</td>
<td>9</td>
</tr>
<tr>
<td>Patel, A.V.</td>
<td>n = 24 males</td>
<td>crossover study</td>
<td>avg: 2.5%</td>
<td>fluid restriction, exercise</td>
<td>N/A</td>
<td>Stationary bike, 45 min, 65-70% HR max (thermoregulatory)</td>
<td>-5.80% difference in matching to sample test</td>
<td>Starnberg memory</td>
<td>5</td>
</tr>
<tr>
<td>Breton, D.</td>
<td>n = 101 (52 males)</td>
<td>randomized, non crossover</td>
<td>avg: 72%</td>
<td>heat exposure, fluid replacement</td>
<td>4 hours in 30°C, 53%</td>
<td>N/A</td>
<td>Decreased recall in no drink group</td>
<td>Word-list recall</td>
<td>7</td>
</tr>
<tr>
<td>Ely, B.R.</td>
<td>n = 32 males</td>
<td>randomized, crossover study</td>
<td>4%</td>
<td>exercise, heat exposure</td>
<td>3 hours in 50°C, 3.5% grade, 30 min rest</td>
<td>Treadmill, 30 min, 1.34 m/s, N/A</td>
<td>No significant difference in memory</td>
<td>Matching to sample</td>
<td>5</td>
</tr>
<tr>
<td>van den Heuvel, A.M.</td>
<td>n = 8 males</td>
<td>crossover study</td>
<td>0%, 3%, 5% at 36.5°C &amp; 38.5°C body temp</td>
<td>passive heat exposure</td>
<td>Normalhemia: 34-35°C Hyperthermia: 40-41°C</td>
<td>N/A</td>
<td>No significant difference in memory</td>
<td>Working memory task</td>
<td>5</td>
</tr>
</tbody>
</table>
Sharma et al.\textsuperscript{11} examined the differences in memory for various levels of dehydration (1\%, 2\%, and 3\%) in a thermoneutral, hot dry, and hot humid environment. Hydration levels were produced via participants completing continuous work until the targeted hydration level was reached. Memory was assessed with a concentration test in which series of numbers were read aloud and stopped at unpredictable intervals. When stopped, the participant was required to recall the last five digits in reverse order. These researchers identified that memory was significantly impaired at 2 and 3\% dehydration in all environments.\textsuperscript{11}

Cian et al.\textsuperscript{5} performed a test of long-term memory (30 minutes) as well as the use of a digit span test to assess short-term memory. The participants took each test under 5 conditions: a control session, two with dehydration via controlled passive heat exposure, and two with dehydration via exercise. Each of the dehydrated conditions consisted of two trials, one with and one without fluid replacement. Participants were dehydrated up to 2.8\%. Short-term memory was significantly reduced across all of the dehydrated trials when compared to the euhydrated trial. There was no significant difference between the methods used for dehydration.\textsuperscript{5}

Ganio et al.\textsuperscript{7} tested memory via a matching to sample and repeated acquisition test in three conditions: exercise-induced dehydration plus diuretic (DD), exercise-induced dehydration plus placebo (DN), and euhydrated plus placebo (EU). Participants were dehydrated to a minimum of 1\%, and researchers did not find a significant difference in memory between the dehydrated trials and euhydrated trial.\textsuperscript{7}

Patel et al.\textsuperscript{10} tested participants in both euhydrated and dehydrated conditions. Dehydration was induced to an average of 2.5\% via exercise. A matching to sample test was used as well as a repeated acquisitions test to assess memory. There was a significant difference
between the euhydrated and dehydrated trials for the matching to sample assessment, however there was no significant difference in memory between trials for the repeated acquisition test.\textsuperscript{10}

Benton et al\textsuperscript{4} dehydrated their participants to an average of .72\% through the use of passive heat exposure and fluid restriction. Memory was assessed using a word-list recall test. There was a significant difference in the number of words remembered in the dehydrated group as opposed to the euhydrated group.\textsuperscript{4}

Van den Heuvel et al\textsuperscript{2} assessed memory at 0, 3, 5\% dehydration at both 36.5 and 38.5\textdegree C body temperatures using passive heat exposure. Memory was assessed using an n-back task, a test of working memory. Although there was a significant increase in reaction time for the dehydrated groups, there was not a significant difference in the number of correct answers for the memory assessment.\textsuperscript{2}

Ely et al\textsuperscript{24} used a matching to sample test to measure memory at 4\% dehydration and euhydration in 10 \textdegree C, 20 \textdegree C, 30 \textdegree C, and 40 \textdegree C environments. Dehydration to 4\% was accomplished by a combination of exercise and heat exposure. There was no significant difference in memory found at any of the environments.\textsuperscript{24}

\textbf{V. Discussion}

Half of the studies reviewed found that memory was impaired when a participant was dehydrated. Since the results are split exactly in half and the methods were varied, it is better to compare similar studies to determine if there is a relationship between hydration status and memory. Secher and Ritz\textsuperscript{25} wrote a review which also yielded inconclusive findings on the relationship between memory and hydration.

The isolation of hydration status can be challenging due to the methods used to induce dehydration. Both exercise and heat are commonly used to induce dehydration in participants.
However, the issue then becomes the inability to determine if the dehydration, fatigue from exercise, or the heat stress was the reason for the impairments in cognition.

Exercise has been shown to increase the production of brain-derived neurotrophic factor (BDNF) in the hippocampus, an area important for learning and memory. BDNF along with other neurotrophic factors play a large role in the process of information being transmitted across the cell synapse. Due to the resulting levels of BDNF in the hippocampus, exercise has been shown to increase an individual’s cognitive abilities. The studies that exclusively used exercise and similar testing batteries, found that at 1% dehydrated there wasn’t a significant change in memory, however a study where an average of 2.5% dehydration was induced saw a significant deficit in memory. Since exercise has been shown to increase cognitive abilities, it seems that the decrement in memory was a result of the dehydration as opposed to the stress of exercise on the body.

The effects and mechanisms of heat stress is unclear, but there are a few proposed theories. One theory proposed suggests that the relationship between cognition and heat stress functions as an inverted-U relationship. This relationship follows the arousal theory and says that there is an optimal level of arousal (temperature), as the body approaches that optimal temperature cognitive function will increase but once the body surpasses the optimal temperature cognitive performance will begin to decline. It has been found that simple cognitive tasks (match to sample, choice reaction time) are less likely to be altered by heat stress as opposed to complex cognitive tasks (working memory, spatial span test). Through the use of the attention network test (ANT) and function magnetic resonance imaging (fMRI), investigators have observed that with both simple and complex cognitive tasks after heat stress there is increased activation in certain areas of the brain as a compensatory response to the heat and no
impairments to the regions of the brain responsible for completing the tasks. The impairment of complex cognitive processes without impairment of the brain regions is attributed to the decrease in dopamine synthesis, a neurotransmitter that plays a role in complex task performance, as a result of the increased plasma serotonin (5-hydroxytryptamine) from the heat stress.²⁸ The studies that exclusively used heat exposure as the method for dehydration,²,⁴ there was not a clear agreement. Of these studies, one showed a significant difference in memory while the other did not. The studies had varying dehydration methods and test batteries. The inconsistent results could be attributed to the variable methods or the sensitivity of the testing batteries.

Lastly, of the studies that used a combination of heat and exercise-induced dehydration⁵,¹¹,²⁴ two showed a significant decrement in memory when dehydration level of 2% or greater, however one showed no significant difference between euhydrated and dehydrated trials. These studies used different heat and exercise protocol as well as testing protocol. Following the theory proposed above, the study that showed no significant change in memory used a test battery that is classified as a simple cognitive task. This could be a reason for the lack of significant difference as the theory proposes that simple cognitive tasks are less likely to be subject to decrements from heat stress.

As of now there is not sufficient evidence to say that a lack of hydration causes a decrement in memory. From the comparison of these studies, there seems to be a relationship amongst hydration and memory, it is just unclear what that relationship is, and what exactly is impacting memory. At what level of dehydration is memory altered? Is the heat stress or fatigue from exercise impacting memory as opposed to the dehydration? Are the tests being used accurately assessing memory?
VI. Conclusion

Future research should be conducted in order to discover the true relationship between memory and hydration. Studies should focus on one method of dehydration (either heat or exercise) and keep a consistent measure of memory. With these methods there should be less confounding variables and a trend could be seen on how hydration levels play into a person’s memory.
References


### PEDro scale

1. Eligibility criteria were specified  
   - Yes

2. Subjects were randomly allocated to groups (in a crossover study, subjects were randomly allocated an order in which treatments were received)  
   - Yes

3. Allocation was concealed  
   - Yes

4. The groups were similar at baseline regarding the most important prognostic indicators  
   - Yes

5. There was blinding of all subjects  
   - Yes

6. There was blinding of all therapists who administered the therapy  
   - Yes

7. There was blinding of all assessors who measured at least one key outcome  
   - Yes

8. Measures of at least one key outcome were obtained from more than 85% of the subjects initially allocated to groups  
   - Yes

9. All subjects for whom outcome measures were available received the treatment or control condition as allocated or, where this was not the case, data for at least one key outcome was analysed by “intention to treat”  
   - Yes

10. The results of between-group statistical comparisons are reported for at least one key outcome  
    - Yes

11. The study provides both point measures and measures of variability for at least one key outcome  
    - Yes

---

The PEDro scale is based on the Delphi list developed by Verhagen and colleagues at the Department of Epidemiology, University of Maastricht (Verhagen, A.P. et al. (1998). The Delphi list: a criteria list for quality assessment of randomised clinical trials for conducting systematic reviews developed by Delphi consensus. *Journal of Clinical Epidemiology, 51*(12):1235-41). The list is based on "expert consensus" not, for the most part, on empirical data. Two additional items not on the Delphi list (PEDro scale items 8 and 10) have been included in the PEDro scale. As more empirical data comes to hand it may become possible to "weight" scale items so that the PEDro score reflects the importance of individual scale items.

The purpose of the PEDro scale is to help the users of the PEDro database rapidly identify which of the known or suspected randomised clinical trials (ie RCTs or CCTs) archived on the PEDro database are likely to be internally valid (criteria 2-9), and could have sufficient statistical information to make their results interpretable (criteria 10-11). An additional criterion (criterion 1) that relates to the external validity (or “generalisability” or “applicability” of the trial) has been retained so that the Delphi list is complete, but this criterion will not be used to calculate the PEDro score reported on the PEDro web site.

The PEDro scale should not be used as a measure of the “validity” of a study’s conclusions. In particular, we caution users of the PEDro scale that studies which show significant treatment effects and which score highly on the PEDro scale do not necessarily provide evidence that the treatment is clinically useful. Additional considerations include whether the treatment effect was big enough to be clinically worthwhile, whether the positive effects of the treatment outweigh its negative effects, and the cost-effectiveness of the treatment. The scale should not be used to compare the "quality" of trials performed in different areas of therapy, primarily because it is not possible to satisfy all scale items in some areas of physiotherapy practice.
Notes on administration of the PEDro scale:

**All criteria** Points are only awarded when a criterion is clearly satisfied. If on a literal reading of the trial report it is possible that a criterion was not satisfied, a point should not be awarded for that criterion.

**Criterion 1** This criterion is satisfied if the report describes the source of subjects and a list of criteria used to determine who was eligible to participate in the study.

**Criterion 2** A study is considered to have used random allocation if the report states that allocation was random. The precise method of randomisation need not be specified. Procedures such as coin-tossing and dice-rolling should be considered random. Quasi-randomisation allocation procedures such as allocation by hospital record number or birth date, or alternation, do not satisfy this criterion.

**Criterion 3** Concealed allocation means that the person who determined if a subject was eligible for inclusion in the trial was unaware, when this decision was made, of which group the subject would be allocated to. A point is awarded for this criteria, even if it is not stated that allocation was concealed, when the report states that allocation was by sealed opaque envelopes or that allocation involved contacting the holder of the allocation schedule who was “off-site”.

**Criterion 4** At a minimum, in studies of therapeutic interventions, the report must describe at least one measure of the severity of the condition being treated and at least one (different) key outcome measure at baseline. The rater must be satisfied that the groups’ outcomes would not be expected to differ, on the basis of baseline differences in prognostic variables alone, by a clinically significant amount. This criterion is satisfied even if only baseline data of study completers are presented.

**Criteria 4, 7-11** Key outcomes are those outcomes which provide the primary measure of the effectiveness (or lack of effectiveness) of the therapy. In most studies, more than one variable is used as an outcome measure.

**Criterion 5-7** Blinding means the person in question (subject, therapist or assessor) did not know which group the subject had been allocated to. In addition, subjects and therapists are only considered to be “blind” if it could be expected that they would have been unable to distinguish between the treatments applied to different groups. In trials in which key outcomes are self-reported (eg, visual analogue scale, pain diary), the assessor is considered to be blind if the subject was blind.

**Criterion 8** This criterion is only satisfied if the report explicitly states both the number of subjects initially allocated to groups and the number of subjects from whom key outcome measures were obtained. In trials in which outcomes are measured at several points in time, a key outcome must have been measured in more than 85% of subjects at one of those points in time.

**Criterion 9** An intention to treat analysis means that, where subjects did not receive treatment (or the control condition) as allocated, and where measures of outcomes were available, the analysis was performed as if subjects received the treatment (or control condition) they were allocated to. This criterion is satisfied, even if there is no mention of analysis by intention to treat, if the report explicitly states that all subjects received treatment or control conditions as allocated.

**Criterion 10** A between-group statistical comparison involves statistical comparison of one group with another. Depending on the design of the study, this may involve comparison of two or more treatments, or comparison of treatment with a control condition. The analysis may be a simple comparison of outcomes measured after the treatment was administered, or a comparison of the change in one group with the change in another (when a factorial analysis of variance has been used to analyse the data, the latter is often reported as a group x time interaction). The comparison may be in the form hypothesis testing (which provides a “p” value, describing the probability that the groups differed only by chance) or in the form of an estimate (for example, the mean or median difference, or a difference in proportions, or number needed to treat, or a relative risk or hazard ratio) and its confidence interval.

**Criterion 11** A point measure is a measure of the size of the treatment effect. The treatment effect may be described as a difference in group outcomes, or as the outcome in (each of) all groups. Measures of variability include standard deviations, standard errors, confidence intervals, interquartile ranges (or other quantile ranges), and ranges. Point measures and/or measures of variability may be provided graphically (for example, SDs may be given as error bars in a Figure) as long as it is clear what is being graphed (for example, as long as it is clear whether error bars represent SDs or SEs). Where outcomes are categorical, this criterion is considered to have been met if the number of subjects in each category is given for each group.