Memory Bias in Observer-Performance Literature

Tamara Miner Haygood, PhD, MD; Samantha Smith

Department of Diagnostic Radiology, UT M.D. Anderson Cancer Center, Houston, Texas, USA;
Tamara.Haygood@mdanderson.org

Rationale

Observer-performance studies using human observers often test the ability of the observer to accomplish a certain task such as detection of pulmonary nodules or breast masses under a variety of different conditions.. When the images being interpreted between one condition and the next are similar, the results can be affected by an observer's recognition of an image. There is relatively little advice available in the radiology literature on how to prevent recognition memory. We studied what published authors have done in this regard.

Methods

We searched AJR online using the terms "observer performance" and "observer study" to identify articles reporting observer-performance experiments. We identified 50 articles dating between 1973 and 2016. We extracted from each the number of reported experiments and five factors related to the authors' approach to recognition memory. These included 1) separate or other viewing of the tested conditions, 2) random or other ordering for tested condition, 3) random or other ordering for cases within sessions, 4) availability of clinical information.

Results

In these 50 articles, 57 separate experiments were reported.

Based on how the tested conditions were presented to viewers, these experiments fit into three categories: 1) Sequential-viewing experiments in which the second tested condition would normally be used as an adjunct to the first, for example, computer-assisted detection in mammography. There were 11 of this type of experiment. 2) Experiments in which memory for images was irrelevant such as alternative forced-choice or multi-point ranking experiments. Sixteen experiments were in this category. 3) Experiments in which image memory remained a potential source of bias, of which there were 30. These experiments are considered in this report.

Among these 30 experiments, tested conditions were presented in a counter-balanced order in 8 experiments, in the same order for each reader in 9 experiments, in random order in 3 experiments, in unique orders in 3 experiments. 7 reports did not indicate in what order the tested conditions were presented.

In the 40 experiments using separate viewing of tested condition, individual images were ordered randomly in 17 and in a pseudo-random fashion in 3. Twenty reports did not indicate how they were ordered.

Within a reading session, 13 experiments presented cases in random order, 2 presented them in a pseudo-random order, and 15 did not indicate how they were ordered.

18 experiments used a time lapse between viewings. Time lapses ranged from 1 to 730 days, 73.7 days average, 21 days median. 11 reports did not indicate if a time lapse was used, and one used sequential viewing.

Readers were blinded to patient information in 10 experiments. Three reports said there was no clinical information given but did not specifically say that patient identifying information was unavailable. In 15 reports there was no indication of whether patient information was available. 2 experiments used phantoms.

Conclusions

In the spirit of providing full details of how an experiment was conducted, many of these published papers would have benefitted from a more precise discussion of how they avoided memory bias.