Anesthesiology
73:815–820, 1990

Appraisal of the Quality of Assessment of Memory in Anesthesia and Psychopharmacology Literature

M. M. Ghoneim, M.D., M. A. Ali, M.D.,† R. I. Block, Ph.D.‡*

To test the hypothesis that there are important differences between studies on memory published in anesthesia literature and those published in the psychopharmacology literature, we compared the two from the period January 1978 through May 1988 to identify deficiencies in the design and methodologies used and to provide guidelines for future experiments. Eighty-eight articles in each discipline were reviewed. The sample sizes were larger in the articles in anesthesia journals than in those in psychopharmacology journals (medians 52.5 *vs.* 18 subjects, respectively). Most (85%) of the studies in the anesthesia literature used patients, who had a median age of 38.9 yr and included a median of 28 women among the subjects per study. In contrast, the majority (60%) of the studies in the psychopharmacology literature used healthy volunteers, who had a median age of 23.6 yr and included a median of only 3.5 females among the subjects per study. Characteristics more common in the psychopharmacology than in the anesthesia literature, respectively, were use of a control or placebo group (90% *vs.* 42%), double-blind design (80% *vs.* 47%), use of pre- and posttreatment memory measurements (64% *vs.* 23%), use of multiple memory tests with distinct equated stimuli (83% *vs.* 8%), relation of methodology to some theoretical model of memory (72% *vs.* 17%), and use of other behavioral tests (68% *vs.* 48%). Relative to the psychopharmacology literature, the anesthesia literature used pictures as stimuli for the memory tests more often (44% *vs.* 14%, respectively) and words less often (11% *vs.* 67%) and relied heavily on questions about recall of perioperative events (41% *vs.* 0%). There is room for improvement in both types of literature, and more so in the anesthesia literature. (Key words: Memory, amnesia. Publications: anesthesia literature, psychopharmacology. Psychological Responses.)

LEARNING AND MEMORY are crucial to normal human existence. All other cognitive functions would be virtually meaningless or impossible without the ability to record and recall previous experience. Studies of memory have therefore been extensive and have employed physiologic, anatomic, pharmacologic, behavioral, and neurologic techniques. We have been acquainted with both the anesthesia literature and the psychopharmacology literature for a number of years and have had the impression that there are important differences between studies on memory published in these two types of journals. This difference was intriguing because both disciplines deal almost exclusively with effects of drugs on memory. We hypothesized that: relative to the anesthesia literature, the psychopharmacology literature used a greater number of methods for measuring memory, more varied and better equated stimuli, and fewer pictorial and more verbal stimuli; measured the degree of amnesia more quantitatively; used methods that were more often related to a variety of theoretical models of memory; and used more control or placebo groups, double-blind designs, and healthy volunteers as subjects.

Researchers studying the effects of drugs on memory address several issues, which vary according to the degree of scientific sophistication employed. At the simplest level, they look for an effect, investigate its specificity, relate it to wider behavioral and pharmacokinetic profiles of drugs, and place the memory change within a theoretical framework. At a more complex level, researchers use drugs to create pharmacologic models for pathologic disorders of memory and to create probes to investigate the psychoneurobiology of normal memory. The concepts, and the possible pitfalls, of conducting sound research with drugs in humans have been well formulated. Good research on the effects of drugs on human memory would bridge the gap between cognitive psychology and clinical pharmacology and would benefit both disciplines. We compared the anesthesia and psychopharmacology literature to evaluate the closeness with which each has approached these goals. We performed a critical appraisal to identify deficiencies in the design and methodologies used and to provide guidelines for future experiments. This achieved, the research literature would better serve to advance the boundaries of knowledge of the involved disciplines.

Materials and Methods

We conducted a search of the literature to identify studies that involved learning and memory in anesthesia and psychopharmacology journals. All articles in the English language journals that contained the words anesthesia or anesthesiology and psychopharmacology in their titles were searched. Articles that were indexed under the key words "human," "learning," "memory disorders," "cognition," or "cognition disorders" were retrieved from Medline, the National Library of Medicine's bibliographic database. The period of time covered was January 1978 through May 1988. Letters, case reports, reviews, editorials and studies that did not include tests of memory were excluded (fig. 1). The same number of

* Professor of Anesthesia.

† Visiting Research Fellow.

‡ Assistant Professor of Anesthesia.

Received from the Department of Anesthesia, College of Medicine, University of Iowa, Iowa City, Iowa. Accepted for publication May 15, 1990.

Address reprint requests to Dr. Ghoneim: Department of Anesthesia, College of Medicine, University of Iowa, Iowa City, Iowa 52242.

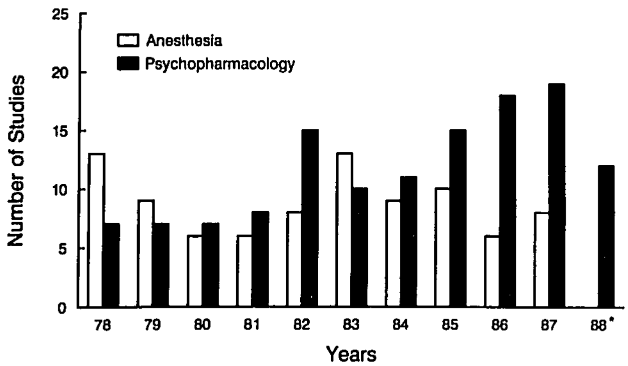

FIG. 1. Chronologic distribution of number of articles on learning and memory in anesthesia and psychopharmacology literatures for the period January 1978 through May 1988. *In this year the journals were reviewed from January through May only.

articles each year was randomly chosen from among the psychopharmacology journals to match the number in anesthesia journals. In the years where the numbers of psychopharmacology articles were less than those in anesthesia journals, we completed the numbers by picking psychopharmacology articles randomly from the closest years having a surplus of articles.

Each article was reviewed independently by two investigators, who identified and recorded the following information:§ 1) the type of subjects involved in the study (patients or healthy volunteers) and their number, age, weight, and sex distribution; 2) methods of measuring memory (*e.g.*, free recall or serial recall), type of stimuli (*e.g.*, words or pictures), number of times memory tests were given, pre- and posttreatment measurements, presence of amnesia and its quantitative measurement, and relation of the study to learning or awareness during general anesthesia; 3) relevance of the methods used to any theoretical model of memory, use of other behavioral tests, and presence of a control or placebo group; 4) for the active treatment groups, the types of drugs used and their dosages and methods of administration; and 5) the type of trial conducted (*e.g.*, double-blind or open), the statistical analyses done, and the major results and conclusions claimed by authors.

STATISTICAL ANALYSES

For quantitative variables (age, weight, and numbers of subjects, which were not normally distributed), Wilcoxon Rank Sum tests were used to compare the anesthesia and psychopharmacology literature. For each qualitative characteristic about which information was recorded, a contingency table comparing the numbers of articles in the anesthesia and psychopharmacology journals was analyzed. Mantel-Haenszel relative rate statistics from analyses stratified by year of publication (controlling, as a simple chi-squared test would not, for effects of year of publication) were used to determine whether statistically significant differences existed, except for a few characteristics which were noted in only one type of literature (anesthesia or psychopharmacology) and never in the other. For the latter characteristics, Fisher's Exact tests were used.

§ The list of articles analyzed along with the questionnaire used are available from the authors on request.

Results

In all, 88 articles published in anesthesia journals were compared with the same number of articles in psychopharmacology journals. We discovered after starting the analyses that 9 articles from Acta Anaesthesiologica Scandinavica had been dropped for unknown reasons from the Medline search. Although we had reviewed these articles and had established the similarity of their data to other articles in the same literature, the analyses were too advanced to include them.

QUANTITATIVE VARIABLES

The sample sizes were larger for the anesthesia journal articles than for the psychopharmacology journal articles (medians of 52.5 *vs.* 18, respectively, $P < 0.001$). The anesthesia journal articles involved substantial numbers of women, whereas the articles in psychopharmacology journals did not (medians of 28 *vs.* 3.5, respectively, $P < 0.001$). Anesthesia and psychopharmacology journal articles did not differ significantly in numbers of men (medians of 10 *vs.* 12). The anesthesia journal articles on average involved older participants than did psychopharmacology journal articles (medians of 38.9 *vs.* 23.6 yr, $P < 0.05$). The minimum and maximum weights of participants in the anesthesia journal articles (medians of 45.8 *vs.* 85 kg, respectively) did not differ significantly from those of participants in the psychopharmacology journal articles (medians of 52 *vs.* 82.5 kg).

QUALITATIVE CHARACTERISTICS

Types of Stimuli and Methods of Measuring Memory

Relative to the psychopharmacology journal articles, those in anesthesia journals used pictures as stimuli for the memory tests more often, and words and other stimuli less often (table 1). With respect to methods of measuring memory (table 1), the anesthesia journal articles, in marked contrast to those in psychopharmacology journals, relied heavily on general questions about perioperative events (41% of anesthesia *vs.* 0% of psychopharmacology articles). The psychopharmacology journal articles used

TABLE 1. Types of Stimuli and Methods of Measuring Memory

	Articles (%)		Relative Rates	95% Confidence Intervals
	Anesthesia	Psychopharmacology		
Types of Stimuli				
Words	11	67	0.2	0.1–0.3*
Pictures	44	14	3.4	2.0–5.8*
Numbers	16	18	0.8	0.4–1.6
Paragraphs	0	6	0.0	—
Others	24	55	0.5	0.3–0.7*
Methods of Measuring Memory				
Free recall	47	63	0.7	0.6–1.0*
Serial recall	2	7	0.4	0.1–1.5
Cued recall	5	11	0.4	0.1–0.4*
Recognition	35	35	1.0	0.7–1.5
Digit span	10	22	0.5	0.2–1.0*
Paired associates	3	9	0.4	0.1–1.4
Questions about events before, during and after surgery	41	0	7.5	3.1–18.3†
Questions about personal historic information	2	0	3.5	0.4–31.7
Others	19	47	0.4	0.3–0.7*

The percentages of articles cover all years of publication. The relative rates and 95% confidence intervals are based on analyses stratified by year of publication.

* $P < 0.05$ by Mantel-Haenszel relative rate statistics.
† $P < 0.05$ by Fisher's exact test.

free recall, cued recall, digit span, and other formal memory assessment techniques more often than did anesthesia journal articles.

Other Qualitative Characteristics

Tables 2 and 3 present the results for other qualitative characteristics. For certain characteristics, the necessary information (*e.g.*, the type of trial) was not provided in some articles or was presented too vaguely to categorize. For these characteristics (table 3), an additional Mantel-Haenszel relative rate statistical contrast was computed to compare the frequency of such unstated or ambiguous information between anesthesia and psychopharmacology journal articles. To illustrate: with respect to type of trial, the Mantel-Haenszel contrast shown in the two middle columns of table 3 is statistically significant, and indicates that double-blind trials (compared to single-blind or open trials) were more common in the psychopharmacology than in the anesthesia literature. In computing these relative rates and confidence intervals, we omitted articles with unstated or ambiguous information. The Mantel-Haenszel contrast shown in the two right-hand columns of table 3 is not statistically significant, and indicates that the frequency of unstated or ambiguous information concerning type of trial was comparable in the anesthesia and psychopharmacology journal articles (13 and 10%, respectively).

The other statistically significant contrasts shown in tables 2 and 3 can be summarized as follows:

1) The following characteristics were more common in the psychopharmacology than in the anesthesia articles: use of a control or placebo group; use of pre- and posttreatment memory measures; use of multiple memory tests with distinct, equated stimuli; methodology related to a theoretical model of memory; and inclusion of other behavioral tests.
2) Adjustment of dosages of drugs according to weight was more common in the anesthesia than in the psychopharmacology journal articles.
3) The anesthesia journal articles more often used patients, whereas articles in psychopharmacology journals more often used healthy volunteers.
4) Unstated or ambiguous information regarding the presence of amnesia was more frequent in the anesthesia than in the psychopharmacology journal articles.

Discussion

We did not do an exhaustive search for every article that addresses the effects of drugs on human learning and memory. Apart from anesthesia and psychopharmacology journals, a few articles are published in a variety of specialty journals, such as journals of clinical pharmacology, oral surgery, psychology, neuropsychology, and psychiatry. What constitutes a psychopharmacology journal is not well defined, so we restricted the review to journals with the word "psychopharmacology" in their titles. Another limitation to our study that applies to meta-analysis is the possibility that some published studies do not include crit-

TABLE 2. Qualitative Characteristics

	Articles (%)		Relative Rates	95% Confidence Intervals
	Anesthesia	Psychopharmacology		
Types of subjects				
Patients	85	40	2.2	1.7–2.9*
Healthy volunteers	15	60		
Control or placebo group				
Yes	42	90	0.5	0.4–0.6*
No	58	10		
Dosages adjusted to weight				
Yes	45	27	1.7	1.1–2.6*
No	55	73		
Pre- and posttreatment memory measures				
Yes	23	64	0.4	0.3–0.5*
No	77	36		
Number of memory tests and equating of stimuli				
Tested multiple times with distinct, equated stimuli	8	83	0.1	0.1–0.2*
Other†	92	17		
Theoretical model of memory				
Yes	17	72	0.2	0.2–0.4*
No	83	28		
Other behavioral tests				
Yes	48	68	0.7	0.5–0.9*
No	52	32		

The percentages of articles, relative rates, and 95% confidence intervals are shown in the same format as indicated in table 1.
* $P < 0.05$ by Mantel-Haenszel relative rate statistics.
† Includes single tests, multiple tests with the same stimuli, or multiple tests with distinct but unequated stimuli.

ical information concerning demographics of subjects, design, and methods used.[1] Meta-analysis may be sensitive also to publication bias, since more positive studies, as compared to negative ones, are submitted and eventually published.[2] This problem, however, is unlikely to be of concern to this review. Finally, it would have been desirable to blind the reviewers regarding the origin of the articles. It is not feasible, however, to blind photocopies of articles of both literatures even to a novice.

There were several important demographic differences between the two literatures. Most of the studies in the anesthesia literature used patients, with an average age

TABLE 3. Qualitative Characteristics with Some Unstated or Ambiguous Information

	Articles (%)		Contrast of *A vs. B*		Contrast for Unstated or Ambiguous	
	Anesthesia	Psychopharmacology	Relative Rates	95% Confidence Intervals	Relative Rates	95% Confidence Intervals
Type of trial						
A) Double-blind	47	80	0.6	0.5–0.7*	1.0	0.9–1.1
B) Single-blind or open	40	9				
Unstated or ambiguous	13	11				
Amnesia present in some test						
A) Yes	78	86	1.1	0.9–1.2	0.8	0.7–1.0*
B) No	8	12				
Unstated or ambiguous	14	3				
Statistical analysis						
A) Done and reported	92	94	0.4	0.04–5.0	3.5	0.8–15.5
B) Done but not reported	1	3				
Unstated or ambiguous	7	2				

The percentages of articles, relative rates, and 95% confidence intervals are shown in the same format as indicated in table 1. In computing relative rates and confidence intervals for the contrast of *A versus B*, articles with unstated or ambiguous information were omitted.
* $P < 0.05$ by Mantel-Haenszel relative rate statistics.

approaching middle age, an adequate representation of both sexes, and a median of 52.5 subjects. On the other hand, more of the studies in the psychopharmacology literature used healthy volunteers, with an average age of 23.6 yr, a preponderance of males, and a median of 18 subjects. It is often lamented that most studies with psychoactive drugs use young healthy volunteers even though the people who take these drugs are patients who suffer from disease(s), who may be older, who may be of different gender, who may be taking other drugs, and whose abilities to metabolize and/or excrete drugs may be impaired.[3,4] Although this criticism is certainly true and although studies in patients seem logical in evaluating drugs' efficacies, studies in healthy volunteers remain useful. Healthy subjects allow better control of experimental conditions than sometimes is feasible in patients and may be more useful for assessing the behavioral effects of drugs unconfounded by therapeutic changes in patients[5] or by drug–drug interactions. There probably is a need for data from both types of subjects. The use of relatively large samples in studies published in the anesthesia literature is a desirable feature, particularly when investigating human behavioral effects, which show considerable variability. This alleviates the concern of overlooking a therapeutic improvement because of small sample size[6] and may facilitate the detection of uncommon side-effects of drugs.

Less than half of the studies in the anesthesia literature used a control or placebo group and a double-blind procedure, whereas the use of these methods increased to 90 and 80%, respectively, in the psychopharmacology literature. Practice on experimental tasks, environmental influences, fatigue, and a host of other factors can change behavior over time. In the absence of a control group, overlooking a therapeutic effect or finding one where none exists are dangers that need to be avoided. Double-blind designs offer the best chance of obtaining unbiased assessments. It should be remembered, however, that ethical issues may prohibit the use of placebos in some patients and that it may not be possible to "blind" patients to some treatments, such as general *versus* regional anesthesia. In such cases, a single-blind design should be used, although it would eliminate only one source of bias.

A closely related factor in the design of behavioral experiments is pre- and posttreatment measurements. Comparisons of treatment and control groups alone are inadequate unless it can be established that the groups are equivalent before treatment. However, only 23% of the studies published in anesthesia literature, in contrast to 64% of those published in psychopharmacology journals, followed this type of design. Only about 27% of studies in both literatures tested the effects of more than a single dose of drug—an astonishing finding given the fundamental relation in pharmacology between the dose of drug administered and magnitude of response obtained. Adjustment of the drug dose according to weight helps to reduce variability in response to drugs and is preferable to the administration of a fixed dose. Studies in the anesthesia literature seem to follow this procedure of adjustment more often; 45% of the studies followed such a strategy compared to 27% in the psychopharmacology literature.

An important and often difficult problem for the investigator lacking formal training in psychology or experience in the topic being researched is the selection of appropriate memory tasks and interpretation of their findings. It is important that the choice of tasks involves a theoretical or analytic perspective, since testing should go beyond simply looking for a decrement or enhancement of memory. At the least, the investigator should make distinctions among the types of memory involved and try to explain the specific memory changes produced (*e.g.*, the short-term and long-term memory dichotomy, the distinction between episodic and semantic memories, the delineation between declarative and implicit memories, and the dissociation of encoding from retrieval). Sound theoretical underpinnings for research with drugs are important if this research is to advance the frontiers of knowledge about the psychoneurobiology of memory. Only 17% of studies in the anesthesia literature adopted a theoretical model of memory, compared to 72% in psychopharmacology, and the types of models used in anesthesia articles were restricted mostly to the anterograde–retrograde memory distinction. Unfortunately, application of models and theories of cognitive psychology by clinicians has generally been difficult for a variety of reasons.[7]

The types of stimuli used to test memory also were different. Most studies in the anesthesia literature used pictures, whereas those in psychopharmacology used words. There are extensive normative data on verbal materials[8,9] with measures such as language frequency of usage,[10,11] image-evoking ability,[12] concreteness, meaningfulness, familiarity, and emotionalism. Use of normative data allows control of level of difficulty of the stimuli and ensures approximate comparability across repeated tests. In contrast, normative data on pictorial stimuli had been lacking until the 1980 study by Snodgrass and Vanderwart.[13] These authors standardized pictures on variables of central relevance to memory and cognitive processing—name agreement, familiarity, and visual complexity. However, we did not locate studies in the anesthesia literature that have taken advantage of these norms. The absence of comparability across repeated tests and among different investigators must be of concern.

If pictorial stimuli are used, the subject should name the picture during its presentation. This involves two steps—picture recognition and name retrieval—and ensures that the subject is alert enough to execute the processes needed. Without naming, further measures of acquisition and recall would be suspect. In 46% of the studies in the anesthesia literature that used pictorial stimuli, however, this requirement seems to have been ignored.

Researchers of works published in anesthesia journals very commonly asked their patients 1 day after surgery to recall events after premedication, in the operating room, and in the recovery room. This procedure may be reasonable, as long as recall of meaningful events on the day of surgery is combined with laboratory-derived tests. As an assessment of memory, recall of perioperative events lacks sensitivity and standardization, does not allow appropriate statistical analyses, and in many situations may be regarded only as an anecdotal report which needs corroboration. However, 27% of studies in the anesthesia literature used this method as the only assessment of memory.

It would be very rare for a drug-induced central nervous system depression to be manifested as an impairment of a single behavioral skill. Usually other functions are impaired as well. Therefore, in most cases investigators should use a range of several tasks to establish a profile of effects for a drug and to allow any dysfunction to be specified in detail. Investigators in psychopharmacologic literature used a wider range of tasks than did those in anesthesia literature.

In conclusion, investigators involved in research dealing with assessment of memory published in the anesthesia literature need to use more placebo controls, double-blind procedures, and pretreatment measurements in the design of their experiments. They should not consider memory as a unitary process or as an all-or-none phenomenon. Studies should be based on a theoretical model of memory. If pictorial materials are used as stimuli, they should be appropriately equated and subjects should be required to name the pictures during their presentations. Recall of events on the day of operation is not by itself an adequate assessment of memory. Use of a wider battery of behavioral tasks is advisable. Investigators publishing in the psychopharmacology literature also can improve the quality of their work: they should perform studies more often in patients, use larger samples which include women, and adjust doses of drugs to body weight.

References

1. Feinstein AR: Scientific standards in epidemiologic studies of the menace of daily life. Science 242:1257–1263, 1988
2. Begg CB, Berlin JA: Publication bias and dissemination of clinical research. J Natl Cancer Inst 81:107–115, 1989
3. Linnoila M: Benzodiazepines and performance, The Benzodiazepines: From Molecular Biology to Clinical Practice. Edited by Costa E. New York, Raven Press, 1983, pp 267–278
4. Johnson LC, Chernik DA: Sedative-hypnotics and human performance. Psychopharmacology 76:101–113, 1982
5. Wittenborn JR: Effects of benzodiazepines on psychomotor performance. Br J Clin Pharmacol 7:61S–67S, 1979
6. Freiman JA, Chalmers TC, Smith H Jr, Kuebler RR: The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. N Engl J Med 299:690–694, 1978
7. Weingartner H, Strupp BJ: Cognitive analysis of neuropharmacological treatments. Drug Dev R 4:577–589, 1984
8. Bradshaw JL: A guide to norms, ratings and lists. Mem Cogn 12: 202–206, 1984
9. Brown AS: Catalog of scaled verbal material. Mem Cogn 4:1S–45S, 1976
10. Thorndike EL, Lorge I: The Teacher's Word Book of 30,000 Words. New York, Teachers College Press, 1944
11. Kucera H, Francis W: Computational Analysis of Present-Day American English. Providence, Brown University Press, 1967
12. Paivio A, Yuille JC, Madigan SA: Concreteness, imagery, and meaningfulness of 925 nouns. J Exp Psychol (Suppl) 76:1–25, 1968
13. Snodgrass JG, Vanderwart M: A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. J Exp Psychol [Hum Learn Mem] 6:174–215, 1980