# ■ REVIEW ARTICLES

# Postoperative Cognitive Dysfunction after Noncardiac Surgery

## A Systematic Review

Stanton Newman, D.Phil., Dip. Psych., A.F.B.P.S., M.R.C.P. (Hon.),* Jan Stygall, M.Sc.,† Shashivadan Hirani, M.Sc.,‡ Shahzad Shaefi, M.B.B.S.,§ Mervyn Maze, F.R.C.A., F.R.C.P., F.Med.Sci., Ph.D.‖

    This article describes a systematic review on the research into postoperative cognitive dysfunction (POCD) in noncardiac surgery to ascertain the status of the evidence and to examine the methodologies used in studies. The review demonstrated that in the early weeks after major noncardiac surgery, a significant proportion of people show POCD, with the elderly being more at risk. Minimal evidence was found that patients continue to show POCD up to 6 months and beyond. Studies on regional *versus* general anesthesia have not found differences in POCD. Many studies were found to be underpowered, and a number of other methodologic difficulties were identified. These include the different types of surgery in studies and variations in the number and range of neuropsychological tests used. A particular issue is the variety of definitions used to classify individuals as having POCD.

FIFTY years ago, prompted by the number of anecdotal reports from his patients and their families regarding problems with cognitive function after surgery, Bedford[1] published a retrospective observational report of 251 older patients who underwent surgery with anesthesia. He noted that although minor degrees of dementia were common in this group of patients, 7% experienced extreme dementia, giving rise to his conclusion that "Operations on elderly people should be confined to unequivocally necessary cases."

This study encouraged investigators to conduct more rigorous prospective studies examining changes in cognitive performance from pre to post surgery as assessed by neuropsychological tests. The change in cognition, when "significant," is now commonly referred to as *postoperative cognitive dysfunction* (POCD). POCD is to be distinguished from postoperative delirium, which tends to be a transient and fluctuating disturbance of consciousness that tends to occur shortly after surgery, whereas POCD is a more persistent problem of a change in cognitive performance as assessed by neuropsychological tests.[2,3]

Until recently, the majority of the research in this field had focused on cardiac surgery, where studies have indicated that a proportion of patients have POCD manifesting as problems with memory, attention, concentration, speed of motor and mental response, and difficulties with learning.[4] The proportion found to have POCD after cardiac surgery varies as a result of a number of issues, including patient-related factors (*e.g.*, age), how soon after surgery the tests are administered, the tests used, and the analysis and criteria for determining deficits.[5] Although the causes of POCD in cardiac surgery are multifactorial, the use of cardiopulmonary bypass has often been cited as the major contributor to the problem. However, evidence is accumulating that off-pump cardiac surgery produces a similar effect on neuropsychological performance to that with the use of cardiopulmonary bypass.[6-8]

In contrast to cardiac surgery and other investigations of cognitive function and deterioration in diseases such as human immunodeficiency virus/acquired immunodeficiency syndrome and Alzheimer disease, the study of POCD in noncardiac surgery is in its infancy. Because the field is relatively new, a number of the studies on this topic are speculative and descriptive and often on small samples. Nonetheless, we believe that it is important to bring these together with the more recent research in a systematic fashion where the extent of the evidence can be assessed. Consequently, the aim of this article is to bring together the studies on this newer field in a systematic review to examine the evidence in relation to POCD in noncardiac surgery.

## Methods: Search Strategy and Selection Criteria

### Identifying Studies

    A review of citations from MEDLINE, EMBASE, PsychInfo, and the Cochrane Library (CDSR, DARE, CEN-

TRAL) was conducted without time limits until December 2005. Full-text articles were retrieved of any citations that were considered potentially relevant. Supplementary methods of retrieving studies included a review of relevant article bibliographies. Our search strategy was as follows: (surg* or operat* or anaesth* or anesthes* or postoperat* or postoperat*) and (neurocogniti* or cogniti* or neuropsycholog* or cerebr* or neurobehaviour* or microemboli*) and (effect* or outcome or decline or dysfunction or impairment or function or production). Journal articles were also searched by hand for relevant articles.

### Included Studies

Randomized controlled trials and observational studies were included subject to description of a study population of greater than 10 patients with an analysis of postoperative cognitive decline after surgery, as assessed by preoperative neuropsychological assessment and postoperative neuropsychological assessment at not less than 7 days after surgery. We have limited the article to studies that performed postoperative assessments after at least 7 days for two reasons: first, to avoid any confusion with delirium after surgery, and second, in an attempt to avoid the general effects of any anesthetic agents. We think it is unlikely that any anesthetic agent may effect neuropsychological assessment after 7 days, although this remains unproven. Articles were included if authors performed statistical analyses over time or between groups or made comparisons with normative data.

Exclusion criteria were surgery on the heart or the brain, including carotid artery surgery or angioplasty. In addition, we excluded noncardiac transplantation and surgery for thyroid disease because these are known to have a significant effect on the brain/cognition, but this is normally an improvement in cognition.[9–11] In addition, studies with unclear timing of test administration and/or articles describing the same or an overlapping patient sample as other articles already included in the review were excluded. Studies investigating only subjective reports of cognitive dysfunction or observational ratings of cognition were also excluded because the relation between these reports and formally assessed cognition is either not apparent or not clear.[12–15]

Articles retrieved were limited to the English language and peer-review publications. To assess the quality of the search strategy, eight studies that were known to be relevant to this field were sampled.[16–23] The search strategy was able to identify all these articles. Forty-six articles met the inclusion criteria. A further article by Abildstrom et al.[24] assessed a subgroup of the Moller et al.[16] study at 1–2 yr after surgery. These two studies have therefore been combined, and reference to Abildstrom et al.[24] only appears in the cohort studies at more than 1 yr.

## Results

Table 1 Describes the papers identified. The studies are divided into three categories:

1. Single-group and controlled studies: Twenty studies that examined a single group and a control group to estimate POCD and, in some cases, factors associated with POCD.
2. Comparison between general (GA) and regional anesthesia (RA): Seventeen studies compared RA and GA.
3. Comparisons between other techniques: Nine studies compared two groups in which a comparison was made that was considered to have a possible influence on the development of POCD.

### Study Details

**Study Design.** Eight of the cohort studies examined a single group and applied a definition of change to estimate the proportion of patients showing POCD (table 1). Twelve studies compared the performance of the group of interest with a control group (table 1). Of these 12, 10 compared the findings with a contemporaneously gathered control/comparison group, and 2 used data from a previously collected study.[25,26]

Although three studies from the cohort group[27–29] also compared the effects of different types of anesthesia on cognition, 17 studies were specifically designed to compare GA with RA (table 1). In 15 of these studies, patients were randomized. Nonsurgical control groups were also used in two studies.[30,31] Flatt et al.[30] used a group of 23 nonpatient individuals age and sex matched with the GA group, and Jones et al.[31] assessed 50 patients on the waiting list for major joint replacement. In the other technique comparison studies (table 1), 7 studies used random allocation to groups,[21,23,32–36] one used successive allocation,[37] and in one study, allocation to groups was not clearly stated.[38]

**Number of Participants.** In studies without a control group, the mean number of patients was 111 (range, 29–288). The largest samples were in those undergoing cataract surgery (mean, 254).

The mean number of patients in the 12 studies that used controls was 235 (range, 35–1,218). Of the RA/GA comparison studies, the mean number of participants was 100 (range, 20–428). The mean number of participants in the studies that compared different techniques was 169, with a range of 27–861.

**Type of Surgery and Anesthesia.** The type of surgery examined in the studies ranged from minor, such as cataract surgery, to major vascular and thoracic. Of the cohort studies, three of the eight studies that examined single group changes in performance over time included patients undergoing orthopedic surgery, a further two examined vascular and thoracic surgery, two assessed those undergoing cataract surgery, and one assessed abdominal surgery (table 1). In studies where the find-

**Table 1. Basic Information on Studies**

| Author | Year | Surgery | Anes | Power | Number F/U $\geq$ 7 Days | Interval, days |
|---|---|---|---|---|---|---|
| **Cohort studies** | | | | | | |
| No controls | | | | | | |
| Rodriguez[27] | 2005 | TKR | GA/RA | X | 2 | 7, 104 |
| Stockton[28] | 2000 | Cataract, urology general, orthopedic | GA/LA | X | 3 | 7, 183, 365 |
| Treasure[58]* | 1989 | Thoracic, major vascular | GA | X | 2 | 8, 56 |
| Ancelin[29] | 2001 | Orthopedic | GA/EA | X | 2 | 9, 84 |
| Goldstein[61] | 1993 | General, orthopedic | GA | X | 2 | 30, 300 |
| Chung[17] | 1990 | Cholecystectomy | GA | X | 1 | 30 |
| Grichnik[60] | 1999 | Thoracic, vascular | GA | X | 1 | 42–84 |
| Elam[39] | 1988 | Cataract | NR | X | 2 | 112, 365 |
| Controls | | | | | | |
| Shaw[59]† | 1987 | Major vascular | GA | X | 1 | 7 |
| Canet[26] | 2003 | Minor | GA | ✓ | 2 | 7, 84 |
| Johnson[56] | 2002 | Abdominal, orthopedic | GA | ✓ | 2 | 7, 84 |
| Dijkstra[53] | 1999 | Major mixed | GA | X | 2 | 7,84 |
| Moller[16] | 1998 | Mixed | GA | ✓ | 2 | 7, 84 |
| Billig[40] | 1996 | Cataract | GA/LA | X | 4 | 7, 42, 183, 365 |
| Rasmussen[25]‡ | 2000 | Abdominal | GA | ✓ | 2 | 7, 84 |
| Goldstein[54] | 1998 | General, ortho | GA | ✓ | 2 | 28, 300 |
| Iohom[70] | 2004 | Abdominal | GA | X | 1 | 42 |
| Farrag[52] | 2001 | Gynecology | GA | X | 2 | 84, 183 |
| Gilberstadt[55] | 1968 | Abdominal | GA | X | 3 | 183, 365, 548 |
| Hall[41] | 2005 | Cataract | NR | X | 1 | 365 |
| **Comparisons to GA** | | | | | | |
| Nonrandomized | | | RA type | | | |
| Berant[18] | 1995 | Mixed | SA/EA | ✓ | 2 | 7, 90 |
| Flatt[30] | 1984 | Plastics | LA | X | 1 | 42 |
| Randomized | | | | | | |
| Hughes[42] | 1988 | THR | SA | X | 1 | 7 |
| Karhunen[49] | 1982 | Cataract | LA | X | 1 | 7 |
| Casati[48] | 2003 | Orthopedic | EA | X | 1 | 7 |
| Riis[51] | 1983 | THR | EA, GA/EA | X | 2 | 7, 90 |
| Rasmussen[17] | 2003 | Mixed | SA/EA | ✓ | 2 | 7, 90 |
| Bigler[43] | 1985 | Orthopedic | SA | X | 2 | 7, 90 |
| Williams-Russo[19] | 1995 | TKR | EA | ✓ | 2 | 7, 182 |
| Campbell[20] | 1993 | Cataract | LA | ✓ | 1 | 14 |
| Asbjorn[22] | 1989 | TURP | EA | X | 1 | 21 |
| Chung[44] | 1989 | TURP | SA with sedation | X | 1 | 30 |
| Chung[45] | 1987 | TURP/pelvic floor | SA | X | 1 | 30 |
| Ghoneim[50] | 1988 | Mixed | SA/EA | X | 1 | 90 |
| Haan[46] | 1991 | TURP | SA | X | 1 | 90 |
| Jones[31] | 1990 | THR/TKR | SA | ✓ | 1 | 90 |
| Nielson[47] | 1990 | TKR | SA | X | 1 | 90 |
| **Technique comparisons** | | | | | | |
| Normotensive *vs.* hypotensive | | | | | | |
| Williams-Russo[21] | 1999 | THR | EA | ✓ | 2 | 7, 120 |
| Rollason[32] | 1971 | Retropubicprostatectomy | GA | X | 1 | 42 |
| Townes[38] nonrandomized | 1986 | Maxillofacial | GA | X | 1 | 180 |
| Intravenous *vs.* inhalation | | | | | | |
| Enlund[33] | 1998 | Orthognathic | GA | X | 1 | 28–56 |
| Hypoxemia | | | | | | |
| Moller[34] | 1993 | Mixed | GA or RA (EA/SA) | ✓ | 2 including subgroup | 2–16 subsample at 97 days |
| Casati[35] | 2005 | Abdominal | GA | ✓ | 1 | 7 |
| Prior[37] | 1982 | Suprapubicprostatectomy | GA/LA | X | 1 | 7 |
| Normocapnia *vs.* hypocapnia | | | | | | |
| Jhaveri[36] | 1989 | Cataract | GA/LA | X | 1 | 28 |
| Vitamins | | | | | | |
| Day[23] | 1988 | Orthopedic | NR | X | 3 | 7, 14, 84 |

* Study group reported here was controls for a study on cardiac surgery.   † Data on controls were not used to compare with patient group in this study. ‡ Previously gathered controls.   § Reported at the time of first assessment.

Anes = type of anesthetic; Comp = composite measures; Edu = education recorded; EA = epidural anesthesia; F/U = follow-up; GA = general anesthesia; IQ = intelligence quotient; LA = local anesthesia; Mood = mood assessed; $N_2O$ = nitrous oxide; NR = not reported; $O_2$ = oxygen; $P_{CO_2}$ = partial pressure of carbon dioxide; Power = power reported; RA = regional anesthesia; TCE = trichloroethylene; THA = total hip replacement; TKR = total knee replacement; TURP = transurethral prostatectomy; SA = spinal anesthesia.

**Table 1. Continued**

| Recruited | | % Male | Mean Age (SD), yr | Age Range, yr | Mood | IQ/Edu | Number of Tests | Domains |
|---|---|---|---|---|---|---|---|---|
| 37 | | 46 | 69.0 (9.0) | 45–82 | ✓ | ✓ | 9 | A, B, C, D |
| 288 | | 52 | 73 (7.0) | 60–93 | ✓ | ✓ | comp | H1 |
| 29 | | 62 | 60.0 (15.5) | 22–85 | ✓ | ✓ | 9 | B, C, D, E |
| 140 | | 33 | 72.6 (5.4) | 64–87 | ✓ | ✓ | 28 test scores | A, B, C, D |
| 82 | | 44 | 67.0 (7.0) | 55–82 | ✓ | ✓ | 1 + 2 comp | A, H1, H8 |
| 40 | | 29 | 53.3 | NR | X | ✓ | 4 + comp | B, C, H1 |
| 51 | | 66 | 60.3 (15.7) | 35–85 | X | ✓ | 6 | B, C |
| 219 | | 23 | 76.5 (4.8) | 70–94 | X | ✓ | comp | H4 |
| **Patients** | **Controls** | | | | | | | |
| 50 | 20 | 72 | 57.4 (6.4) | 41–68 | X | X | 3 + comp | A, B, C, D |
| 372 | Moller 1998 | 47 | 67.7 (NR) | 61–80 | ✓ | ✓ | 4 | B, C, D |
| 508 | 183 | 26 | Median 50.5 | 41–59 | ✓ | ✓ | 4 | B, C, D |
| 56 | 50 | 68.2 | 68.2 (NR) | 60–85 | ✓ | ✓ | 4 | B, C, D |
| 1,218 | 176+ 145 | 51§ | 68§ | 60–79§ | ✓ | ✓ | 6 | B, C, D |
| 108 | 48 | 31 | 75.9 (7.2) | NR | ✓ | ✓ | 2 + comp | A, H1 |
| 65 | Moller 1998 | 75 | Median 68 | NR | X | X | 4 | B, C, D |
| 172 | 190 | 44 | 66.5 (7.1) | 55–87 | ✓ | ✓ | 1 + 2 comp | A, H8 |
| 42 | 13 | 56 | Median 52 | 40–80 | ✓ | ✓ | 6 | A, B, C, E |
| 35 | 18 | 0 | 41.4 (5.2) | NR | X | ✓ | 5 + comp | B, H1 |
| 74 | 59 | 100 | 66.71 (4.18) | 54–75 | ✓ | ✓ | 15 + comp | A, B, C, D, E, F, G + H2 |
| 122 | 92* (no cataract) 87 (cataract) | 42 | 70.9 (6.8) | NR | ✓ | ✓ | 1 comp | H6 |
| **GA** | **RA** | | | | | | | |
| | 102 | NR | 69 (5.5) | 60–80 | X | X | 5 | B, F |
| 23 | 7 | 47 | 42.5 (NR) | 18–73 | ✓ | X | 6 | B, C |
| 15 | 15 | NR | 68 (NR) | NR | X | X | 1 | B |
| 30 | 30 | 0 | 73.5 (NR) | NR | ✓ | ✓ | 9 | B, E |
| 15 | 15 | 7 | 84 (NR) | 67–94 | X | X | comp | H1 |
| 10 | 10 | NR | > 60 | NR | X | X | 10 | B, C, D, G |
| 217 | 211 | 41 | Median 71 | 61–84 | ✓ | ✓ | 4 | B, C, D |
| 20 | 20 | 21 | 78.9 (2.3) | NR | X | X | comp | H4 |
| 128 | 134 | 30 | 69 (NR) | NR | ✓ | ✓ | 10 | A, B, C |
| 85 | 84 | 34 | 77.6 (7.5) | NR | X | ✓ | 4 | B, E |
| 20 | 20 | 100 | 68.8 (NR) | NR | X | X | 5 | B |
| 22 | 22 | 100 | 72 (1.3) | NR | ✓ | X | comp | H1 |
| 24 | 20 | 50 | 72.3 (NR) | 60–93 | ✓ | X | comp | H1 |
| 53 | 52 | 66 | 61 (2.0) | NR | ✓ | ✓ | 13 | B, C, D, E, F, G |
| 26 | 27 | 100 | 71.5 (5.5) | NR | ✓ | X | 4 + comp | B, C, D, H1 |
| 72 | 74 | 27 | > 60 | NR | ✓ | ✓ | 4 | B, C, E |
| 49 | 49 | NR | 69.1 (6.1) | NR | ✓ | ✓ | 5 + 2 comp | A, B, C, D, E |
| **Normotensive** | **Hypotensive** | | | | | | | |
| 118 | 117 | 50 | 72 (NR) | 50–88 | ✓ | ✓ | 9 | A, B, C |
| 13 | 14 | 100 | 65 (NR) | NR | ✓ | ✓ | 8 | A, B, C, D, F |
| 27 | 17 | 27 | 27.8 (NR) | NR | ✓ | X | 6 | A, B, C |
| **Propofol** | **Isoflurane** | | | | | | | |
| 16 | 16 | 53 | 36.1 (15.4) | NR | X | X | 3 | B |
| 861 whole group | | 44 | 52.2 (NR) | 18–82 | X | X | 1 + comp | B, C |
| 56 | 66 | 57 | 72.5 (5.0) | NR | X | X | comp | H1 |
| Air (LA) | 15 | 100 | Whole group 65.5 | 55–83 | X | ✓ | 1 | B |
| air–ether | 15 | | | | | | | |
| air–TCE | 15 | | | | | | | |
| $N_2O$–$O_2$ | 15 | | | | | | | |
| GA: $P_{CO_2}$ 5.3 kPa | 40 | 47 | 74.7 | 60.89 | X | X | 1 + comp | B, C |
| GA: $P_{CO_2}$ 2.7 kPa | 30 | | | | | | | |
| LA | 13 | | | | | | | |
| **Vitamins** | **Controls** | | | | | | | |
| 28 | 32 | 27 | 79.4 (7.95) | NR | X | X | 2 + comp | B, C, H4 |

ings of the index group were compared with a control group, the type of surgery varied and included three studies with patients undergoing abdominal surgery, two studies that included some orthopedic patients, two cataract surgery, one gynecologic, one vascular, and three that described surgery as minor or mixed. In addition to being minor, cataract surgery has the potential confounding variable of individuals having improved visual acuity that in turn may lead to improvements on some cognitive tests.[28,39–41]

General anesthesia alone was used in the majority (14 of 20; 70%) of the cohort studies, whereas one study combined GA and epidural anesthesia (EA),[29] two studies combined GA and local anesthesia (LA),[40] and one study combined GA, LA, and what they termed *neuroleptanalgesia*.[28] Two studies did not report the type of anesthesia used.[23,39] In the studies where different types of anesthesia were combined, it is not possible to attribute any findings to the effect of specific anesthetic techniques.

Specific comparisons between types of anesthesia were examined in 17 studies. Seven (41%) compared GA with spinal anesthesia (SA),[31,42–47] 3 (18%) with EA,[19,22,48] 3 with LA,[20,30,49] and 3 with SA/EA.[17,18,50] One study compared GA with EA and EA plus GA.[51] Again, the type of surgery investigated varied widely and included three transurethral resection of prostate,[22,44,46] one transurethral resection of prostate/pelvic floor repair,[45] two cataract,[20,49] six orthopedic,[19,31,42,47,48,51] one plastics,[30] and three mixed.[17,18,50]

Each of the nine technique comparison studies examined a different surgical group. In four of the studies, the participants underwent GA[32,33,35,38]; in one, EA[21]; and two studies used GA for study patients and examined patients receiving LA as a comparison group.[36,37] One included patients receiving GA or RA, and one did not report the type of anesthesia used.[23]

**Number and Timing of Assessments.** The timing of assessments is an important issue because early assessments may identify a transitory cognitive problem (*i.e.*, postoperative delirium), whereas those assessing patients at more remote times from the surgical intervention are able to establish POCD that may be persistent or permanent. In cardiac surgery, the timing of the assessments after surgery has been found to be one of the most significant factors in the number of patients found with POCD or the extent of postsurgical changes.[5] In the cohort studies reported here, 11 (55%) performed a follow-up within 10 days of surgery. At the other extreme, one study conducted the first follow-up approximately 1 yr after surgery. Fifteen studies (75%) conducted more than one follow-up; conversely, in the comparisons with GA and other comparison of techniques studies, the majority (69%) conducted only a single postoperative assessment, with 50% of these being performed during the first 10 days after surgery.

Whereas 35% of cohort studies examined patients 300 or more days postoperatively, the latest assessment in the RA *versus* GA studies was 6 months, and in the studies that compared different techniques, the latest assessment was 4 months after surgery. Conducting more than one follow-up assessment enables an evaluation to be made of the time course of the progression of POCD. This is, however, complicated by the confounding effect that learning may have where repeated assessments are performed.

**Age and Sex of Participants.** Early reports of cognitive change after surgery implicated "old people."[1] Both for this reason and because most surgical interventions occur in the latter years of life, the bulk of studies examined individuals with a mean or median age over 60 yr. It is also of note that age is the patient-related factor that has been found to be associated with greatest change in neuropsychological test performance in cardiac surgery.[5] In single-group cohort studies (table 1), the age for participants ranged between 22 and 93 yr, with all but one of the studies reporting a mean or median of 60 yr or greater for their sample. Where age was reported in the cohort studies with a control group (table 1), the majority of samples also had a mean or median age of over 60 yr (7 of 12), with the study on gynecologic surgery recruiting the youngest group (mean, 41.4; SD, 5.2).[52] With the RA *versus* GA studies (table 1), the age range, which was only documented in 5 studies (42%), was 18–93 yr, and, where reported, the mean age was over 60 yr in all but one of the studies. The youngest participants (with a mean age of 42.5 yr) were in a group of patients undergoing plastic surgery. In 66% of the studies that compared techniques, the participants' mean age exceeded 60 yr. Four studies (44%) reported an age range which was 18–89 yr.

Many studies specifically selected "older" participants over a specific age,[25,26,28,29,31,40,51,53–55] although the age cutoff varied between studies. One study[56] selected "patients between 40 and 59 yr of age to assess POCD in middle aged patients," and in one study[57] a specific comparison was made between two age groups. In comparing age groups, it is difficult to remove other confounders such as comorbidities or concurrent medications with the result that comparisons are not made on age alone. For example, Chung *et al.*[57] compared those younger than 60 yr to those 60 yr and older and documented that the latter group had more medical problems. Some other studies examined whether age had an influence on the extent of decline, whereas others controlled for age effects.[53]

Of all the studies in table 1, eight focused on single-sex surgical groups; six focused on males,[22,32,37,44,46,55] whereas two dealt with females,[49,52] of which one study[52] was designed to assess the differences found in surgical as opposed to physiologic menopause. Four studies did not report sex.[18,42,47,51]

**Assessment and Definition of POCD.** Many studies have chosen to define POCD using "individual change" scores. In this type of analysis, each participant acts as his or her own control, and a classification is made as to whether a particular participant showed evidence of sufficient decline to be defined as having POCD. The advantage of this approach is that it defines and categorizes individual performance. However, regardless of the definition chosen, these will be a statistically defined criterion that has no intrinsic meaning or reference to brain damage. The difficulty with these conventional definitions of POCD is that "sufficient decline" is variously defined; e.g., Treasure et al.[58] defined a decrease in performance equal to or greater than 1 SD from the preoperative score on two or more tests to indicate POCD, whereas Shaw et al.[59] regarded the same decline as indicative of POCD, but it only had to be present in one or more tests. Deficit was rated as being either 1 SD decline in 1 of 21 tests or 1 SD decline in 4 of 21 tests in one study[21] and as being 20% decline in 20% of the tests in another study.[60] Williams-Russo et al.[19,21] examined individual change that was based on establishing a clinically important difference score for each test and then converted the participant's raw within-subject change score to a −1, 0, or +1 score reflecting whether the observed change was worse than the clinically important difference, within one clinically important difference, or better than the clinically important difference. This score was then summed, and any participant with a score of −3 or less was defined as having a deficit. One study[57] also compared participants postoperatively with normative data to examine whether there was a significant difference. Five studies used a standardized cutoff score as an indication of decline on a screening measure.[23,40,44,46,61]

A number of studies used group change scores in neuropsychological tests to determine whether surgery or a comparison between two or more surgical groups resulted in differences.

Studies applied various statistical procedures (e.g., t tests, analysis of variance, analysis of covariance) to examine group differences. In addition, others used multivariate techniques such as multiple regression to explore variables that influence POCD or cognition after surgery.

**Number of Tests Used.** Because of the time constraints of the surgical environment, the neuropsychological assessments are limited when contrasted with a clinical neuropsychological assessment that would take approximately 2.5 h and attempt to cover most cognitive domains.[62] As a result, the tests selected end up being a compromise to fit within the restrictions imposed by the environment.

Establishing the number of tests used in studies is made more complex by the fact that in some studies, researchers used a comprehensive battery to assess a wide range of cognitive domains. In some cases, the tests

in these batteries (e.g., Mini-Mental State Examination [MMSE]) are accumulated to produce a single score. In others, a number of scores are produced.[29] Nineteen (41%) of the studies in table 1 used a comprehensive test battery either alone or in combination with other neuropsychological tests (see appendix for key). As can be seen from tables 2–4, there was a wide range in the number of tests used in studies.

**Domains and Types of Tests Selected.** When neuropsychological tests were first introduced into the study of POCD, they tended to be traditional "intelligence tests" or screening tests such as the MMSE. A problem with screening tests is that some are liable to show ceiling effects if cutoffs are applied.[40] The overall batteries tend to be highly reliable but are unlikely to have the sensitivity required to detect the subtle (but important) changes after surgery. For example, the Wechsler Adult Intelligence Scale has proved itself to be insensitive to assess change after cardiac surgery.[63] Recently, a number of tests have been specifically designed for repeated administration, and some have been computerized to improve the standardization and ease of administration.

Overall, in the studies reported here, 70 different neuropsychological tests have been used in this area along with 9 composite batteries (appendix). The domains assessed by these tests in the studies are displayed in table 1. The domain most assessed was memory and learning (B), where 33 of the studies applied specific tests. To this must be added those studies where composite batteries were used, because these also examine some aspects of memory and learning. Comparisons between studies are made extremely difficult because of the differences in the tests selected. Although different tests may assess a similar domain, their sensitivity to assess change is likely to differ.

**Dealing with Learning.** Despite attempts to restrict learning on repeated administration of neuropsychological tests, it is customary for some learning to be found. These can occur as a result of increased familiarity with the test structure and alterations in strategy in relation to the test. In studies of POCD, patients undergo at least two assessments, frequently with only a fairly short time separation. Many researchers have specifically selected tests that keep learning effects to a minimum, and parallel equivalent forms have also been used to reduce learning effects. Nonetheless, learning is apparent in most studies under review.[e.g., 40,55] In studies with two groups, the control group enables the impact of learning to be assessed. One approach by the multicenter International Study of Post-Operative Cognitive Dysfunction (ISPOCD) group[16,26,53,56] has been to analyze their data by comparing the mean of the neuropsychological change score from a healthy control group over three assessments corresponding to the assessment intervals of the surgical group. The mean of the control group

**Table 2. Outcomes of Cohort Studies without Controls and with Controls**

| Author | Year | Interval, days | Recruited Surgical Group | Completed Surgical Group | Tests | Definition of Decline | Decline Surgery | Decline Controls | Significant Difference |
|---|---|---|---|---|---|---|---|---|---|
| **7–21 days** | | | | | | | | | |
| No control | | | | | | | | | |
| Rodriguez[27] | 2005 | 7 | 37 | 29 | A5, A6, B5, B6, C1, C4, C5, D1, D3 | Decrease in performance ≥ 0.5 SD in 20% tests | 41% | | |
| Stockton[28] | 2000 | 7 | 288 | 274 | H1 | ANCOVA on MMSE | No classification | | |
| Treasure[58] | 1989 | 8 | 29 | 24 | B1, B22, C2, C4, C5, C7, C8, D2, E2 | ≥ 1 SD drop in ≥ 2 tests | 50% | | |
| Ancelin[29] | 2001 | 9 | 140 | 133 | H5 | > 1 SD drop in 1 of 21 summary scores | 71% | | |
| Controls | | | | | | | | | |
| Shaw[59] | 1897 | 7 | 50 | 48 | A1, C5, D2, H7 | ≥ 1 SD drop in 1 test | 31% | Controls not compared | NS |
| Canet[26] | 2003 | 7 | 372 | 323 | B28, C2, C11, D3 | Z scores (2 from 7) or combined ≥ 1.96 | 6.8% | 3.4% (from Moller 1998) | NS |
| Johnson[56] | 2002 | 7 | 508 | 463 | B28,C2,C11,D3 | Z scores (2 from 7) or combined ≥ 1.96 | 19.2% | 4.0% | $P = 0.001$ |
| Rasmussen[25] | 2000 | 7 | 65 | 52 | B28, C2, C11, D3 | Z scores (2 from 7) or combined ≥ 1.96 | 32.7% | 3.4% (from Moller 1998) | |
| Dijkstra[53] | 1999 | 7 | 56 | 48 | B28, C2, C11, D3 | Z scores (2 from 7) or combined ≥ 1.96 | 27% | 6% | $P = 0.048$ |
| Moller[16] | 1998 | 7 | 1,218 | 1,214 | B28, B29, C2, C8, C11, D3 | Z scores (2 from 7) or combined ≥ 1.96 | 25.8% | 3.4% | $P < 0.001$ |
| Billig[40] | 1996 | 7 | 108 | 108 | A3, A5, H1 | ANOVA MMSE and/or significant difference in other 2 tests | No change | No change | No change |
| **22–132 days** | | | | | | | | | |
| No control | | | | | | | | | |
| Goldstein[61] | 1993 | 30 | 82 | 62 | A3, H1, H8 | NR | NR | | |
| Chung[57] | 1990 | 30 | 40 | NR | B27, C3, C4, C5, H1 | Significant difference compared with norms | No difference, no decline | | |
| Stockton[28] | 2000 | 42 | 288 | 274 | H1 | MMSE 1 or > | 27% | | |
| Grichnik[60] | 1999 | 42–84 | 51 | 29 | B1, B5, B14, B27, C2, C5 | ≥ 20% decline in 20% of tests | 44.8% | | |
| Treasure[58] | 1989 | 56 | 29 | 24 | B1, B22, C2, C4, C5, C7, C8, D2, E2 | ≥ 1 SD drop in ≥ 2 tests | 50% | | |
| Ancelin[29] | 2001 | 84 | 140 | 98 | H5 | > 1 SD drop in 1 of 21 summary scores | 56% | | |
| | | | | | | > 1 SD drop in ≥ 4 of 21 | 11% | | |
| Rodriguez[27] | 2005 | 104 | 37 | 28 | A5, A6, B5, B6, C1, C4, C5, D1, D3 | Decrease in performance ≥ 0.5 SD in 20% tests | 18% | | |
| Elam[39] | 1988 | 112 | 219 | 164 | H4 | Change score *t* test | No decline (improved) | | |
| Controls | | | | | | | | | |
| Goldstein[54] | 1998 | 28 | 172 | NR | A3, H1, H8 | Comparison of MMSE change scores | No difference | | |
| Billig[40] | 1996 | 42 | 108 | Not clear | A3, A5, H1 | ANOVA MMSE and/or significant difference in other 2 tests | No change in MMSE; significant improvement in 2 tests | No change in MMSE; significant improvement in 2 tests | |
| Iohom[70] | 2004 | 42 | 42 | 40 | A5, B1, C2, C4, C5, E2 | RCI deficit in one or more domains | 53% | 23% | $P = 0.03$ |
| Canet[24] | 2003 | 84 | 372 | 323 | B28, C2, C11, D3 | Z scores (2 from 7) or combined ≥ 1.96 | 6.6% | 2.8% (from Moller 1998) | NS |
| Johnson[56] | 2002 | 84 | 508 | 422 | B28, C2, C11, D3 | Z scores (2 from 7) or combined ≥ 1.96 | 6.2% | 4.1% | NS |

*(continued)*

**Table 2. Continued**

| Author | Year | Interval, days | Recruited Surgical Group | Completed Surgical Group | Tests | Definition of Decline | Decline Surgery | Decline Controls | Significant Difference |
|---|---|---|---|---|---|---|---|---|---|
| Farrag[52] | 2001 | 84 | 35 | 35 | B4, B5, B6, B7, B27, H1 | Significant difference pre to post surgery per test | Significant decline MMSE | No decline | |
| Rasmussen[25] | 2000 | 84 | 65 | 53 | B28, C2, C11, D3 | Z scores (2 from 7) or combined $\geq$ 1.96 | 9.4% | 2.8% (from Moller 1998) | |
| Dijkstra[53] | 1999 | 84 | 56 | 48 | B28, C2, C11, D3 | Z scores (2 from 7) or combined $\geq$ 1.96 | 8% | 2% | NS |
| Moller[16] | 1998 | 84 | 1218 | 947 | B28, B29, C2, C8, C11, D3 | Z scores (2 from 7) or combined $\geq$ 1.96 | 9.9% | 2.8% | $P = 0.0037$ |
| **6 mo–1 yr** | | | | | | | | | |
| **No control** | | | | | | | | | |
| Stockton[28] | 2000 | 183 | 288 | 274 | H1 | No classification—MMSE score by ANCOVA | No classification | | |
| Goldstein[61] | 1993 | 300 | 82 | 54 | A3, H1, H8 | NR | NR | | |
| Stockton[28] | 2000 | 365 | 288 | 251 | H1 | MMSE score decline of 1 or > at 6 or 12 months | 35% | | |
| Elam[39] | 1988 | 365 | 219 | 164 | H4 | Change score $t$ test | No decline (improved) | | |
| **Controls** | | | | | | | | | |
| Billig[40] | 1996 | 183 | 108 | Not clear | A3, A5, H1 | ANOVA MMSE and/or significant difference in other 2 tests | No change in MMSE; significant improvement in 2 tests | No change in MMSE; no change | |
| Gilberstadt[55] | 1968 | 183 | 74 | ?63 Not clear | A4, C8, D4, E1, E2, E4, F2, G1, H2, H3, H7 | Comparison of tests | No decline | No decline | No difference |
| Farrag[52] | 2001 | 183 | 35 | 35 | B4, B5, B6, B7, B27, H1 | Significant difference pre to post tests | Significant decline in MMSE and all WMS | No change | |
| Goldstein[54] | 1998 | 300 | 172 | 108 | A3, H1, H8 | Comparison of MMSE change scores | No decline | No decline | No difference |
| Billig[40] | 1996 | 365 | 108 | 104 | A3, A5, H1 | $\geq$ 2 point difference in MMSE | 9%; no change | 12%; no change | NS |
| | | | | | | ANOVA MMSE and/or significant difference in other 2 tests | Significant improvement | Significant improvement | NS |
| Gilberstadt[55] | 1968 | 365 | 74 | ?63 Not clear | A4, C8, D4, E1, E2, E4, F2, G1, H2 H3, H7 | Comparison of tests | No decline | No decline | No difference |
| Hall[41] | 2005 | 365 | 122 | NR | H6 | Pre to post $t$ test ANOVA | Significant improvement | No change | |
| **> 1 yr** | | | | | | | | | |
| **Controls** | | | | | | | | | |
| Abildstrom[24] (Moller 1998) | 2000 | 365–730 | | 336 | B28, C2, C11, D3 | Z scores (2 from 7) or combined $\geq$ 1.96 | 10.4% | 10.6% | No difference |
| Gilberstadt[55] | 1968 | 548 | 74 | ?63 Not clear | A4, C8, D4, E1, E2, E4, F2, G1, H2, H3, H7 | Comparison of tests | No decline | No decline | No difference |
| Gilberstadt[55] | 1968 | 730 | 74 | ?63 Not clear | A4, C8, D4, E1, E2, E4, F2, G1, H2, H3, H7 | Comparison of tests | No decline | No decline | No difference |

ANCOVA = analysis of covariance; ANOVA = analysis of variance; MMSE = Mini-Mental State Examination; NR = not reported; NS = not significant; RCI = reliable change index; WMS = Wechsler memory scale.

changes were used as an estimation of learning and were subtracted from the surgical participants' change score, and the result was divided by the control group SD to obtain a Z score for each test. This calculation allows for each test to be analyzed separately and also enables the scores be combined into a total neuropsychological

**Table 3. Outcomes of Studies Comparing General and Regional Anesthesia**

| Author | Year | Interval, days | Recruited GA | Completed GA | Recruited RA | Completed RA | Tests | Definition Decline | Decline GA | Decline RA | Significant Difference |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **7-21 days** | | | | | | | | | | | |
| Nonrandomized | | | | | | | | | | | |
| Berant[18] | 1995 | 7 | | 40 | | 48 | B27, B21, B33, F2, F3, adapted tests | ANOVA of individual tests | | | No difference |
| Randomized | | | | | | | | | | | |
| Ghoneim[50] | 1998 | 1–7 | 53 | 53 | 52 | 52 | B2, B19, B22, B27, C2, C3, C7, C8, D3, E1, F1, G2 | Significant decline MANOVA | Significant decline | Significant decline | NS |
| Williams-Russo[19] | 1995 | 7 | 128 | 128 | 134 | 143 | A5, A6, B16, B27, B25, B23, B24, C2, C4, C5 | CID per test | Significant decline 4 tests | Significant decline 4 tests | NS |
| Hughes[42] | 1988 | 7 | 15 | 15 | 15 | 15 | B30 | Comparison of means | Unclear | Unclear | NS |
| Bigla[43] | 1985 | 7 | 20 | 19 | 20 | 19 | H4 | t test between groups | Improved | Improved | No difference |
| Riis[51] | 1983 | 7 | 10 | 10 | 10GA/LA 10 | 10:10 | B3, B15, B27, B20, B21, C3, C4, C5, D2, G2 | Z score and ANOVA of individual tests | Improved | Improved; GA/LA improved | No difference |
| Karhunen[49] | 1982 | 7 | 30 | 30 | 30 | 30 | B4, B5, B6, B7, B10, B11, B18, B27, E1 | t test between groups | Decline in Luria memory | Decline in Luria memory | Significantly less decline in GA |
| Rasmussen[17] | 203 | 7 | 217 | 188 | 211 | 176 | B28, C2, C11, D3 | Z scores (2 from 7) individual tests or combined ≥ 1.96 | ITT 19.7% PP 21.2% | 18.3% 18.9% | NS P = 0.04 |
| Campbell[20] | 1993 | 14 | 85 | 77 | 84 | 80 | B27, B31, B32, E3 | Significant decline MANOVA | No decline | No decline | |
| Asbjorn[22] | 1989 | 21 | 20 | 20 | 20 | 20 | B2, B15, B27, B33, B34 | Compared Z scores | Improved | Improved | No difference |
| **22–132 days** | | | | | | | | | | | |
| Nonrandomized | | | | | | | | | | | |
| Flatt[30] | 1984 | 42 | 23 | 23 | 7 | 7 | B1, B7, B27, C3, C8, C9 | Mean total score (ANOVA) | Improved | Improved | Greater improvement in GA, P ≤ 0.05 |
| Berant[18] | 1995 | 90 | | 40 | 7 | 48 | B27, B21, B33, F2, F3, adapted tests | ANOVA of individual tests | | | No difference |
| Randomized | | | | | | | | | | | |
| Chung[45] | 1987 | 30 | 24 | 24 | 20 | 20 | H1 | Compared mean difference | No decline | No decline | |
| Chung[44] | 1989 | 30 | 22 | 22 | 22 | 22 | H1 | ≥ 2 point drop in MMSE | No decline | No decline | |
| Rasmussen[17] | 2003 | 90 | 217 | 175 | 211 | 165 | B28, C2, C11, D3 | Z scores (2 from 7) individual tests or combined ≥ 1.96 | ITT 20.4% PP 19.7% | 20.2% 21.0% | NS NS |
| Haan[46] | 1991 | 90 | 26 | 22 | 27 | 27 | B13, C2, D6, H1 | ≥ 2 point drop in MMSE | No decline | No decline | |
| Jones[31] | 1990 | 90 | 72 | 64 | 74 | 65 | B26, C8, C12, E3 | Compared mean difference | RT improved | No decline | Significantly improved RT, P = 0.05 |
| Neilson[47] | 1990 | 90 | 49 | 39 | 49 | 25 | A5, C4, C5, C12, E1, H3 (short form), H7 | ANOVA individual tests | No decline | No decline | |
| Ghoneim[50] | 1988 | 90 | 53 | 53 | 52 | 52 | B2, B19, B22, B27, C2, C3, C7, C8, D3, E1, F1, G2 | Significant decline MANOVA | No decline | No decline | |

**Table 3. Continued**

| Author | Year | Interval, days | Recruited GA | Completed GA | Recruited RA | Completed RA | Tests | Definition Decline | Decline GA | Decline RA | Significant Difference |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Bigler[43] | 1985 | 90 | 20 | 19 | 20 | 19 | H4 | Compared means | No decline | Improved | NS |
| Riis[51] | 1983 | 90 | 10 | 10 | 10 | 10 | B3, B15, B27, B20, B21, C3, C4, C5, D2, G2 | Z score and individual test ANOVA | No decline | No decline; no decline GA/RA | Improved attention and memory, $P = 0.01$ |
| **6 months** | | | | | | | | | | | |
| Williams-Russo[19] | 1995 | 183 | 128 | 117 | 134 | 114 | A5, A6, B16, B27, B25, B23, B24, C2, C4, C5 | CID per test | Improved | Improved | Significant improvement in 7 tests; no difference between groups |

ANOVA = analysis of variance; CID = clinically important difference; GA = general anesthesia; ITT = intention to treat; LA = local anesthesia; MANOVA = multivariate analysis of variance; MMSE = Mini-Mental State Examination; NS = not significant; PP = per protocol; RA = regional anesthesia; RT = reaction time.

score because the difference in dispersion of scores on each test is removed by scoring them in SD units from the mean. However, the authors did apply a cutoff score to define POCD by rating the participants as having POCD when Z scores on two individual tests or the combined Z score reached 1.96 or more (the higher the score the more deterioration).

**Controlling for Alternative Explanations: Education, Intelligence, and Mood.** In a number of studies, either an estimate of general intelligence was performed before surgery or level of education was recorded. This was done to examine either whether individuals with high or low intelligence or education are particularly susceptible to the negative effects of surgery or, in studies with more than one group, to ensure the groups are balanced on this potential confounder or whether there is a need to use education/intelligence quotient (IQ) as a control in the analyses. The "cognitive reserve" hypothesis suggests that individuals with relatively low intelligence should be more susceptible to an equivalent brain injury than individuals with higher intelligence or education (*e.g.*, Elkins *et al.*[64]). On the basis of this hypothesis, it would be expected that a higher rate of POCD should occur in those with lower intelligence or limited education. There is little evidence to support an association of general intelligence with decline after cardiac surgery, although there is some evidence to suggest that higher levels of education *protect* against decline after cardiac surgery.[65] As displayed in table 1, 18 of the cohort studies (90%) assessed either education or performed an assessment of IQ before surgery. However, these assessments were conducted in only 41% (7 of 17) of the RA *versus* GA studies and 33% (3 of 9) of the intervention studies.

Two related factors are used to justify the need to make assessments of mood in studies of cognitive change after surgery. The first is that mood changes may occur from before to after surgery[55] and that mood, in particular depression and anxiety, has been found to correlate in some studies with performance on some neuropsychological tests.[62] In the articles under review, 70% (14 of 20) of the cohort studies, 59% (10 of 17) of the GA *versus* RA studies, and 33% of the studies that compared different techniques assessed mood. Both education/IQ and mood were assessed in 70% (14 of 20) of the cohort studies, but the percentage decreases to 29% (5 of 17) in the GA *versus* RA studies and 22% (2 of 9) of the technique comparison studies.

Because of the diversity of types of assessments of both cognitive function and mood or psychiatric state, it would be unlikely that a clear picture would emerge from the studies performed. In addition, many researchers have selected neuropsychological tests that are not particularly susceptible to the effects of, or changes in, mood.

**Table 4. Outcomes of Studies Comparing Different Techniques**

| Author | Year | Interval, days | Recruited Normotensive | Completed Normotensive | Recruited Hypotensive | Completed Hypotensive | Tests | Definition Decline | Decline Normotensive | Decline Hypotensive | Significant Difference |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Normotensive vs. hypotensive** | | | | | | | | | | | |
| Williams-Russo[21] | 1999 | 7 | 118 | 118 | 117 | 117 | A5, A6, B27, B25, B23, B24, C2, C4, C5 | Individual test change score (CID) | Significant decline in 2 tests | Significant decline in 2 tests | NS |
| | | 120 | 118 | 113 | 117 | 113 | | | Improved or no decline | Improved or no decline | NS |
| Rollason[32] | 1971 | 42 | 13 | 13 | 14 | 14 | A1, B2, B7, B16, C7, D5, F2 | Comparison of ranking of tests | Improved | Improved | NS |
| Townes[38] | 1986 | 180 | 27 | ?227 | 17 | ?17 | A1, B3, C4, C5, C9, C14 | Comparison of mean change | No decline | No decline | NS |
| **Intravenous vs. inhalation** | | | Propofol | Propofol | Isoflurane | Isoflurane | | | Propofol | Isoflurane | |
| Enlund[33] | 1998 | 28–56 | 16 | 15 | 16 | 14 | B12, B18, B27 | t test comparison pre to post | Significant decline B18; significant improvement B12 | Significant decline B18; significant improvement B12 | NS |
| **Hypoxemia** | | | Pulse oximetry | Pulse oximetry | No pulse oximetry | No pulse oximetry | | | Pulse oximetry | No pulse oximetry | |
| Moller[34] | 1993 | 2–16 | 861 whole group | 358 | | 378 | C8, H7 | Pre to post difference scores | | | NS |
| | | 97 | | | | | | | | | NS |
| Casati[35] | 2005 | 7 | 131 whole group | 56 | | 66 | H1 | MMSE ≥ 2 points | 35% | 54% | NS |
| | | | | | | | | | 0% desaturated only | 66% desaturated only | P = 0.001 |
| Prior[37] | 1982 | 7 | Air (LA) / air–ether / air–TCE / N₂O–O₂ | 15 / 15 / 15 / 15 | 15 | 15 | B17 | t test comparisons | All improved | | NS |
| **Normocapnia vs. hypocapnia** | | | | | | | | | | | |
| Jhaver[36] | 1989 | 28 | GA:5.3kPa / GA: 2.7kPa / LA | 40 / 30 / 13 | | | C8, H9 | Comparison of tests | No decline any group | No decline any group | NS |
| **Vitamin** | | | Vitamin | Vitamin | Control | Control | | | Vitamin | Control | |
| Day[23] | 1988 | 7 | 28 | 26 | 32 | 32 | B26, C10, H4 | ≥ 3 point drop in AMT | No decline | No decline | NS |
| | | 14 | | ?226 | | ?232 | | | No decline | No decline | NS |
| | | 84 | | 20 | 27 | 27 | | | No decline | No decline | NS |

AMT = Abbreviated Mental Test; CID = clinically important difference; GA = general anesthesia; LA = local anesthesia; MMSE = Mini-Mental State Examination; $N_2O$ = nitrous oxide; NS = not significant; $O_2$ = oxygen; TCE = trichloroethylene.

*Findings*

**Cohort Studies.** In table 2, the cohort studies are divided into five periods according to the time of postoperative assessment and, within these periods, by the design of the study (no controls and with controls).

The number of those recruited and completing the follow-up assessment is indicated in table 2. It shows that overall attrition rates, where reported, were lower, with shorter follow-ups: 5.4% for assessments between 7 and 21 days; 19% for assessments between 22 and 132 days; and for the few studies reporting attrition at the times beyond 6 months, 17%. It is unclear whether this attrition is selective and that participants with certain characteristics were lost to follow-up. Selective attrition raises questions regarding the validity of findings, and this is particularly pertinent when considering POCD because there is a need to establish whether those lost to follow-up are more or less likely to have had POCD. Some reports on POCD after cardiac surgery have suggested that there may be selective attrition, with sicker patients being more likely to be unavailable for follow-up[66] (see also Newman and Stygall[5]). Long follow-up studies in cardiac surgery have shown that attrition is higher among those with lower IQ[67] and lower education.[68]

*7- to 21-Day Assessments.* Of the eight cohort studies conducted without a control group, four conducted the first assessment within 21 days of surgery. Three reported a decline in performance ranging from 41%[27] to 71%.[28] However, the classification of POCD varied between these studies, *i.e.*, Rodriguez[27] adopted a decrease in performance of $\geq 0.5$ SD in 20% of tests, Treasure[30] $>$ 1 SD drop in two or more tests, and Ancelin[29] $> 1$ SD drop in 1 of 21 summary scores. POCD was not defined in one study.[28]

In the cohort studies with a control group, seven conducted an assessment at 7 days after surgery. One study reported no change in performance, and the other six described decline occurring in 6.8%[26] to 31%.[59] Three definitions of decline were used: $\geq 1$ SD drop in one test,[59] a comparison between preoperative and postoperative scores by analysis of variance,[40] and the remainder defined decline as Z scores of 2 from seven tests or a combined score of 1.96 or greater. Of those that made specific comparisons of the prevalence of POCD, only one study[26] found no significant difference in performance between the control and surgical groups.

These early findings showed a tendency for the studies with the least stringent definitions to report a greater proportion of patients with POCD,[27,29] greater deficits occurring in the more severe forms of surgery,[58,59] and the most minor forms of surgery showing no[40] or minimal POCD.[26] It is also of note that the controlled studies produced deterioration rates between 3.4% and 6% in the control group. Only in the case of minor surgery were differences between the control and the study

group found to be not significant. This relatively clear pattern of results attests to the robustness of POCD soon after surgery, given all the differences in methods of assessment between the studies (*e.g.*, number, type, and sensitivity of the neuropsychological tests). Differences in rates of early POCD between minor and major surgery are reinforced by comparing the ISPOCD studies of Canet[26] on minor surgery with those of Moller[16] and Rasmussen,[25] who both assessed major surgery using similar methodologies and neuropsychological tests. This indicated that major surgery produced between 26% and 33% POCD compared with 7% for minor surgery.

It is of note that two studies used the MMSE either alone[28] or with two other neuropsychological tests.[40] In both of these studies, no overall differences were found between preoperative and early postoperative performance, with the exception of the oldest age group (85 yr and older) in the Stockton[28] study. These findings further suggest that screening tests such as the MMSE do not have the sensitivity to examine for POCD.

The findings do suggest that older people are more likely to have early POCD. Two of the single-group studies reported that older patients were more susceptible to early decline,[28,29] and one of the controlled studies[26] found that age over 70 yr was a risk factor for early POCD. Further support that older age is associated with early POCD comes from a comparison between ISPOCD group studies where an identical methodology was used. Johnson[56] examined a middle-aged sample (40–59 yr) and found POCD in 19.2% and Moller[16] and Rasmussen[25] found POCD in 25.8% and 32.7% of their samples who were older than 60 yr (see also Rasmussen *et al.*[69]).

In addition, there is some suggestion in two studies that patients who may have been sicker or requiring more extensive surgery may be more likely to have POCD. In one study,[26] those selected by the hospital to undergo inpatient rather than outpatient surgery were more likely to show POCD. In the other study,[27] an association was found between postoperative complications and cognitive dysfunction at 7 days. It is possible that these increased rates of POCD in those with complications may reflect this and the additional medication to deal with the complications.

*22-Day up to 6-Month Assessments.* The majority of studies reported no evidence of POCD, or no decline or an improvement in neuropsychological performance. Where reported, POCD prevalence in the surgical groups ranged between 6.2% and 56%. Ignoring the one study with a high incidence of POCD,[70] the other studies produced POCD rates of between 6.2% and 9.4% in the surgical group and between 2% and 4% in the control groups studied. In most cases, the scores in the surgical group were greater than those found in the controls, but in only two studies did this reach significance.

By examining the eight studies that performed an as-

sessment at both this and the previous time point, it is possible to assess the changing rates of POCD with increasing intervals after surgery. In all but one of these,[58] POCD decreased from the first to the second time point. The percentage decrease in POCD ranged between 3% and 71%. Besides the study by Treasure,[58] the lowest change occurred in the study by Canet[26] on minor surgery that did not find any differences in the surgery and control groups at 7–21 days postoperatively. The four other ISPOCD studies using the same methodology showed POCD rates at 12 weeks after surgery of between 62% and 71% lower than were found at 7 days. These data provide clear evidence that rates of POCD decline from the acute phase (7 days) to longer periods after surgery.

Although a number of studies examined the possible effect of education and/or IQ on the occurrence of POCD, few effects were found. In one study,[29] those with low educational attainment had more POCD, and another study[60] found that those with lower education showed greater declines after surgery. Whereas no relation was found between mood and cognitive decline in two studies,[30,24] depression or the risk of depression before surgery was found to be associated with decline in a number of studies.[28,29]

Significant differences in POCD between inpatient and outpatient treatment reported soon after surgery by Canet[26] was not apparent at this later assessment, but the inpatient group had higher POCD than the controls. The report by Rodriguez[27] of an association of POCD and postoperative complications at the early assessment was found to persist at this later assessment.

*Assessments > 6 Months.* Eight studies reported assessments at 6 months or longer after surgery, with one study[55] having four assessments over this period and two other studies assessing patients on two occasions.[28,40] The bulk of studies reported no decline or an improvement from before surgery. Importantly, none of the studies with a control found any difference from the control group.

Abildstrom,[24] who examined a subset of the ISPOCD study of Moller at 1–2 yr after surgery, found no differences between the elderly group undergoing surgery and the controls. The authors estimated that POCD persists to this time period in only approximately 1% of patients. They did identify age as a risk factor and, in common with work on the long-term impact of cardiac surgery,[71] showed that an early deterioration increased the likelihood of long-term POCD. One difficulty identified by the authors is that only 3 of the 35 patients with POCD at 1–2 yr had POCD at the earlier assessment points.

**General *versus* Regional Anesthesia.** One hypothesis regarding POCD after noncardiac surgery is that the mechanism of damage occurs through the use of GA. Consequently, the use of alternative methods of anesthe-

sia for the same procedure should result in a reduction or a removal of POCD. A number of studies have considered this issue, and their findings are displayed in table 3, organized by whether random allocation to groups was used and the time of the assessment after surgery.

*7- to 21-Day Assessments.* One nonrandomized study and nine randomized studies performed follow-up assessments at 7–21 days after surgery. At this time, one study[49] of patients undergoing cataract surgery found differences between GA and LA, but it is unclear whether the analysis in this study took account of the higher preoperative scores of the LA group. This would increase the likelihood for this group to show a greater decline for statistical reasons.

The study by Rasmussen and the ISPOCD investigators[17] examined patients aged 60 yr and older undergoing a range of surgeries requiring a hospital stay of at least 4 days. The ISPOCD investigators' analysis protocol included accounting for learning by subtracting from the performance of the GA and RA groups the changes in performance of healthy controls that were collected in an earlier study. The investigators' intention-to-treat analysis showed a higher incidence of POCD in GA (19.7%) compared with RA (12.5%), which just failed to reach significance ($P < 0.06$). However, a further per-protocol analysis that excluded 56 participants showed the difference between GA (21.2%) and RA (12.7%) to be statistically significant ($P < 0.04$).

*22-Day up to 6-Month Assessments.* None of the 12 studies (2 nonrandomized) that assessed patients between 1 and up to 6 months after surgery found differences between the performance of those undergoing GA or RA. Rasmussen,[17] who had reported differences at 7 days, assessed participants 3 months postoperatively and detected cognitive dysfunction in approximately 20% of their sample at 3 months, but with no differences between RA and GA (intention-to-treat GA 20.4%/RA 20.2% and per-protocol GA 19.7/RA 21%). Of the remaining 11 studies, some reported no decline in both RA and GA groups,[44–47,50] whereas others reported some improvements in performance.[30,31,43,51] The study with the longest follow-up[19] of 6 months found modest improvement from earlier declines in both anesthetic (GA and EA) groups.

The evidence suggests that using RA as an alternative to GA does not result in any reduction in POCD. The one large well-designed study that on early changes suggested a better outcome in RA on a per-protocol analysis, did not show any differences at the 3-month assessment.[17]

**Studies Comparing Different Techniques.**

*Normotensive* versus *Hypotensive.* Hypotensive anesthesia offers advantages of a dry surgical field and potential reductions in blood loss. However, it has been suggested that hypotensive surgery may increase the

likelihood of ischemic damage to the brain. Three studies have examined the effects of deliberate hypotensive anesthesia on POCD[21,32,38] and satisfied the inclusion criteria of this systematic review. GA was used in two studies,[32,38] and EA was used in one.[21] Each study examined a different form of surgery (joint replacement, prostatectomy, maxillofacial), and the specification of "hypotensive" differed between studies. The times of follow-up assessments also varied with each study. None of the studies found any differences in cognition between hypotensive and normotensive anesthesia.

*Intravenous* versus *Inhalation Anesthesia.* Enlund *et al.*[33] compared the effect of isoflurane or propofol on neuropsychological performance after major orthognathic surgery. At 4–8 weeks after surgery, they detected a significant decline compared with baseline in the Luria verbal learning test and a significant improvement in the Taylor-Rey-Osterreith (copying), with no differences between groups.

*Hypoxemia.* Various techniques have been introduced to reduce the occurrence of hypoxemia during surgery. In a randomized study, Moller *et al.*[34] found that using a pulse oximetry in and after surgery to identify and indicate the need to intervene to reduce instances of hypoxemia did not affect neuropsychological performance at discharge (2–16 days after surgery). Forty patients who had poor memory performance were followed up 3 months later, and at that point, their median scores had returned to baseline. Casati *et al.*,[35] using a decrease of 2 or more points in the MMSE as a definition of decline, found no difference between groups when comparing those undergoing surgery using pulse oximetry and those without oximetry. However, when comparing those patients who had an intraoperative episode of desaturation, a decline of cognitive function was observed in 10 patients in the control group only ($P = 0.001$). Prior *et al.*[37] assessed 60 prostatectomy patients before and 7 days after surgery. The participants were divided into four groups: (1) extradural and air, (2) air and ether, (3) air–trichloroethylene, and (4) nitrous oxide–oxygen. Improvement was detected in all groups; however, there was no difference between groups.

*Normocapnia* versus *Hypocapnia.* The neuropsychological effects of hypocapnia were investigated in one study.[36] Comparing patients undergoing cataract surgery who received either ventilation to a mean arterial carbon dioxide tension ($Pa_{CO_2}$) of 4.9 kPa, hyperventilation to a mean $Pa_{CO_2}$ of 2.9 kPa, or LA, no decline was found in neuropsychological performance in any group after surgery, and no difference was found between groups.

*Vitamins.* The cognitive effect of the intravenous administration of vitamins (B complex and C) given to patients undergoing surgery for a fractured femur was compared with that of randomized nonsupplemented controls by Day *et al.*[23] They assessed participants on three occasions, 7, 14, and 84 days postoperatively, and found no decline and no difference between groups on any assessment.

## General Discussion and Conclusions

This article reviewed studies of postoperative cognitive decline after noncardiac surgery. However, a major difficulty in trying to compare investigations or establish an incidence of POCD was the diversity in participants, types of surgery and anesthesia, methods of assessment, definition of POCD, and mode of analysis. Despite all the diversity, the findings in cohort studies present relatively clear evidence of POCD 1 week after major surgery. In the large well-designed studies (largely the ISPOCD group), the data suggest that POCD is only evident after major surgery. This conclusion is supported by the reanalysis of the ISPCOD data set that attempted to control for the variability in performance by taking account of improvements and deteriorations after surgery.[69]

At periods between 22 and 132 days, only two studies found evidence of greater declines than control groups. Although this data are persuasive that well-controlled studies are able to demonstrate POCD at this later time, it is of note that further analysis has indicated that at this time the number of participants showing significant improvements in their performance is similar to those showing declines. On the basis of this, Rasmussen and Siersma[69] suggest that the findings may reflect random variation rather than POCD.

One area that requires further examination is the possibility that symptoms such as pain and/or some types of postoperative medication may lead to poorer neuropsychological performance. It is possible that these factors may also lead to larger declines in the days after surgery when pain and the use of medication may be at its greatest and to less POCD at later assessment times.

In interpreting these findings, it is important to recognize that the numbers of participants in many studies were well below what may be considered adequate to assess POCD in noncardiac surgery. This is not surprising in a new field, but it is instructive to consider the numbers required for an adequately powered study. Assuming 80% power and an $\alpha$ of 0.05, where one group showed 10% of patients to have POCD and the second group to have twice that proportion (20%), the sample size for each group would need to be 199, assuming groups of equal size. If the background incidence of POCD was 50% and the index group had an incidence of 60%, the numbers per group would need to be 388. Therefore, many of the studies are underpowered (table 1). Only five studies recruited groups of 200 or more. In the cohort studies, 40% (8 of 20) recruited 50 or fewer. In the studies comparing different anesthetic methods with GA, this percentage is 70% (12 of 17), and in the studies that compared different techniques, it is 66% (6 of 9).
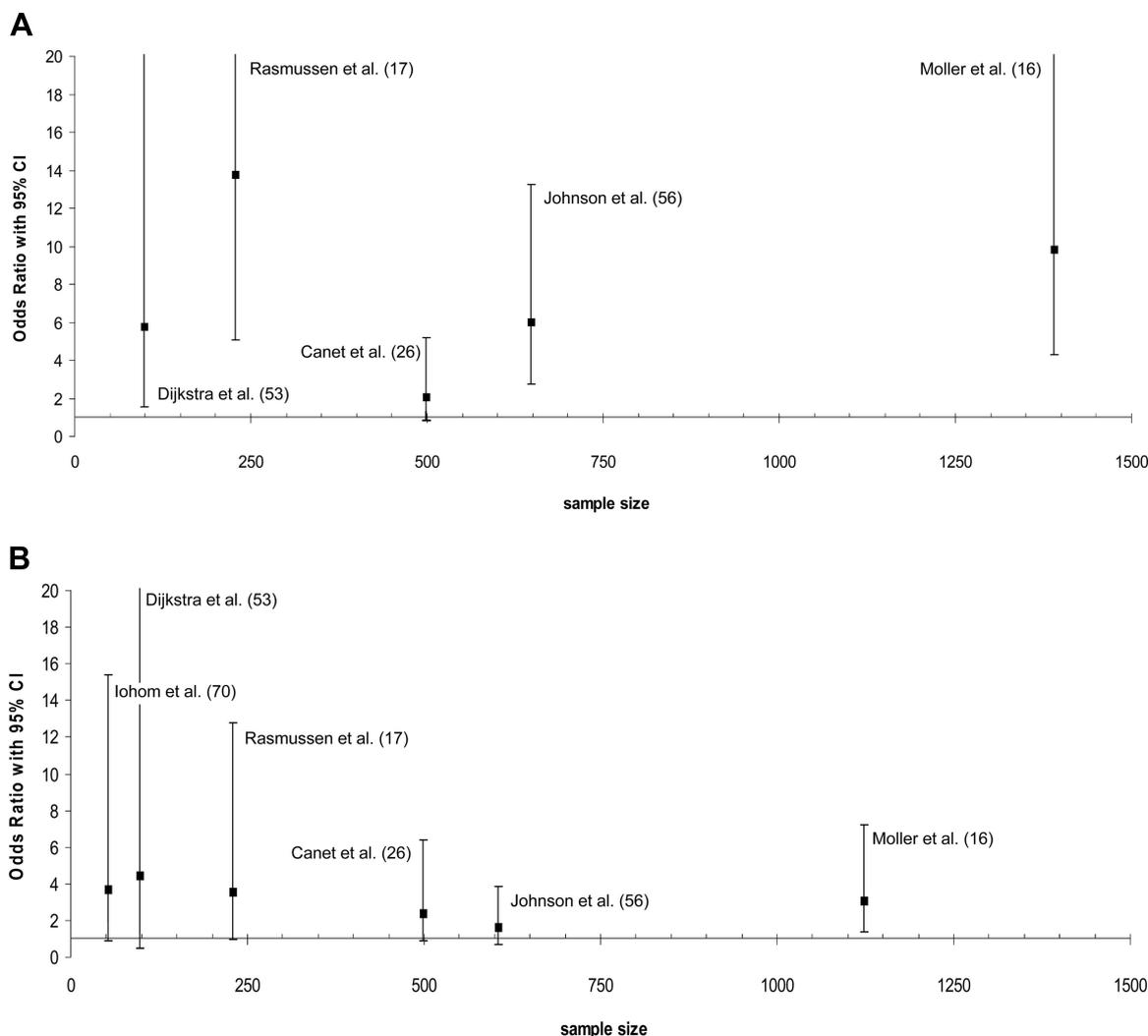
**Fig. 1.** (*A*) Odds ratios and 95% confidence intervals (CIs) for studies with controls examining follow-ups between 7 and 21 days. (*B*) Odds ratios and 95% CIs for studies with controls examining follow-ups between 22 days and 6 months.

The importance of sample size in relation to the timing of the assessments is provided by a consideration of the odds ratios and 95% confidence intervals of the cohort studies with controls that provided data on the percentage of participants with POCD. Figure 1A shows these findings for the early assessments. It is apparent that the effects at this early period are sufficiently large to produce 95% confidence intervals that do not cross with unity even with relatively small samples. The one study that does cross the line of unity (indicating no difference to controls) is where minor surgery was examined. In contrast, in studies with postoperative assessments from 22 days up to 6 months, it was only in the largest study, that assessed patients at 84 days, where the confidence intervals did not cross with unity (fig. 1B). These figures demonstrate that to get a clear signal in this area, studies with large samples are required.

The studies that compared GA and RA as well as those comparing other techniques provide little evidence as to what may be responsible for any changes in cognition observed after surgery. It is only in orthopedic surgery that a putative mechanism has been identified in the form of microemboli (probably fat) that have been identified through transcranial Doppler studies of the middle cerebral artery during surgery.[72] The one study that did use transcranial Doppler in orthopedic surgery did not find any relation between the numbers of microemboli and changes in neuropsychological performance.[27] However, great caution must be exercised in interpreting the studies that varied aspects of anesthesia or surgery, because the majority were significantly underpowered.

The research has concentrated mainly on an older age group, with only nine studies[30,33,34,38,52,56,59,60,70] examining participants with a mean age of less than 60 yr. The evidence suggests that older participants are more likely to show POCD. In a large study with a control group (ISPOCD2 group), Johnson *et al.*[56] compared patients aged 40–60 yr with a previous group aged over 60 yr and concluded that the younger group showed signifi-

cantly less POCD at both 7 days ($P = 0.0064$) and 3 months ($P = 0.026$).

Designs that used a single group and examined change in performance over time cannot control for the influence of extraneous variables, especially the learning that occurs in many neuropsychological tests. This is an important consideration because any learning and resultant improvement in performance will have the effect of reducing the prevalence of POCD. The use of a design that involves a control group makes it possible to control for alternative explanations and specifically for any improved performance through learning. These designs do, however, raise the question of what constitutes an appropriate control group for this type of study. In this review, some studies have selected healthy individuals as controls[53] or relatives of the individuals under study,[52] whereas others selected patients with other conditions to those under study,[54] patients with the same condition but who did not undergo surgery as determined by the surgeon,[55] or patients who elected to have or not have surgery. Each of these approaches to select a control group has its strengths and weaknesses. In the case of healthy controls, the ability to learn is controlled for, but it is assumed that the patients under study would evidence an equivalent rate of learning. Studies using patients with another condition as controls assume equivalence between patients with different conditions, whereas those that use controls with the same condition are able to control for the illness. The problem for the latter is the ethical difficulty of random allocation to receive surgery or not to receive surgery. Where group allocation is determined by the surgeon, clinical factors may introduce bias. Whether the patients elect for surgery or not, it is likely that patient-related factors would differ between groups. The ethical issues of randomizing to receive or not receive surgery are obvious in cohort studies, but in studies comparing different techniques, this is more easily achieved (tables 3 and 4).

Neuropsychological assessments have been found to have sufficient sensitivity to be able to detect small and subtle cognitive changes that may occur after surgical procedures or medical treatment.[62] By necessity, the neuropsychological batteries chosen for investigations into the impact of surgery on the brain are often a compromise, balancing of the time constraints imposed by the clinical environment with the selection of sensitive and reliable tests. Ideally, to gain information on cognitive change these tests should be comprehensive and assess more than one domain. Where the definition of POCD involves a deterioration on a specified number of tests, conducting more tests will increase the probability of finding deficits, not only because of the number of tests used but also because more domains will be assessed.[73,74] The number of tests used in the studies under review was large, and these differences make it difficult to compare studies. Not only is the number of tests important, but whether they were drawn from separate domains can influence the findings. For example, one study[22] used five tests but only examined memory. It is likely that tests from the same domain are likely to show a greater correlation than tests drawn from different domains. Seventy different neuropsychological tests were used in these studies (appendix). Studies used anywhere between 1 and 13 tests. Six studies[28,39,41,43–45] used only a generic screening test such as the MMSE or Abbreviated Mental Test, and seven further studies[23,40,46,52,54,57,61] used these assessments in conjunction with other tests; however, five of these studies[26,40,46,54,61] based their primary definition of decline solely on the generic measure. Nineteen (46.3%) explored three or more domains. Memory was assessed in all studies (including batteries), with 4 studies assessing memory only.

The extent of decline in neuropsychological scores necessary to be defined as POCD in the studies reviewed has made comparisons of the percentage of individuals with POCD across studies particularly complex. This is especially important because the numbers identified by different techniques show little agreement. For example, Mahanna *et al.*[75] used five different criteria to define neuropsychological deficits after cardiac surgery and found a sixfold difference in the incidence of deficits (3.4–19.4%). Where two surgical groups or a surgical and a control group are compared within studies with the same criteria for POCD, the relative incidence of POCD can be established. However, the use of conventional cutoffs even in studies with two or more groups results in detailed continuous measures being reduced to a binary decision of POCD or no POCD. This is especially problematic because the point of demarcation is arbitrary and, if increased or decreased, may lead to different findings. It also applies to studies where it has been found that individuals assessed at different times may move from having POCD at one point to not having it at a later point and *vice versa* (*e.g.*, Rasmussen and Siermsa[69]). Small changes for those on the boundary of one category are likely to lead to significant shifts in the individuals identified as having POCD. At a more general level, there are a host of difficulties in making binary classifications of continuous data in general[76] and in examining POCD in particular.[63]

An alternative approach in studies where more than one surgical group is assessed is to consider postoperative cognitive change and to examine differences in scores between groups without applying a cutoff. This approach recognizes that some learning with repetition should be expected with most neuropsychological tests and assumes that this is the background against which the impact of surgery needs to be considered. Even when parallel forms are used to minimize learning, participants have been found to develop different strategies that can lead to an enhancement of their accuracy or

speed on that test. One measure of cerebral damage is the inability to demonstrate learning on neuropsychological tests with repeated administrations. In this way, the presence or extent of learning can be used as the index of the relative success of an intervention to reduce the impact of surgery on neuropsychological function. This approach accepts that the retention of learning ability, coupled with a reduction in the extent of deficits, may lead to the intervention group showing greater learning than the surgical control group. In this approach, both learning and deterioration are taken into account in examining group differences in performance.[77] The use of group scores, however, does not enable individual differences in change in cognitive performance to be considered.

## Conclusion

Overall, the research in this review has demonstrated that in the early weeks after major noncardiac surgery, a significant proportion of people show POCD, with the elderly being more at risk. Although the research here is generally negative, there is a little evidence that a reduced proportion of patients continue to show POCD up to 6 months after major surgery, although it has been suggested that this finding may be due to random variation. None of the studies have elucidated the possible mechanisms for any cognitive changes.

The research area suffers from a large number of underpowered studies and a range of other methodologic difficulties. These include the differences in surgery, participants, the diversity, number, and range of neuropsychological tests used with varying sensitivity to change and learning, and the variety of definitions used to classify individuals as having POCD. These differences make it difficult to compare across studies. To overcome some of the methodologic issues, it would be useful to recognize the arbitrariness of any definition of POCD and the difficulties that a binary definition introduces into a continuous measure of cognition. It may be useful to consider whether the term *postoperative cognitive dysfunction* has outlived its usefulness and acknowledge a need to examine cognition and cognitive change as a continuous measure such that changes in scores may be analyzed.

Given the difficulty of funding adequately powered studies, it is useful to consider whether it is timely to establish a consensus that specifies a limited number of tests to be used in all studies and the value of pooling data across studies to increase power in secondary analyses.

## References

1. Bedford PD: Adverse cerebral effects of anaesthesia on old people. Lancet 1955; 2:259-63
2. Bryson GL, Wyand A: Evidence-based clinical update: General anesthesia and the risk of delirium and postoperative cognitive dysfunction. Can J Anaesth 2006; 53:669-77
3. Fong HK, Sands LP, Leung JM: The role of postoperative analgesia in delirium and cognitive decline in elderly patients: A systematic review. Anesth Analg 2006; 102:1255-66
4. van Dijk D, Keizer AMA, Diephuis JC, Durand C, Vos LJ, Hijman R: Neurocognitive dysfunction after coronary artery bypass surgery: A systematic review. J Thorac Cardiovasc Surg 2000; 120:632-9
5. Newman SP, Stygall J: Neuropsychological outcome following cardiac surgery, The Brain and Cardiac Surgery. Edited by Newman SP, Harrison MJG. Amsterdam, Harwood Academic Publishers, 2000, pp 21-49
6. van Dijk D, Jansen EWL, Hijman R, Nierich AP, Diephuis JC, Moons KGM, Lahpor JP, Borst C, Keizer AMA, Nathoe HM, Grobbee DE, De Jaegere PPT, for the Octopus Study Group: Cognitive outcome after off-pump and on-pump coronary artery bypass graft surgery. JAMA 2002; 287:1405-12
7. Samuels MA: Can cognition survive heart surgery? Circulation 2006; 113:2784-6
8. Jensen BO, Hughes P, Rasmussen LS, Pedersen PU, Steinbruchel DA: Cognitive outcomes in elderly high-risk patients after off-pump *versus* conventional coronary artery bypass grafting: A randomized trial. Circulation 2006; 113:2790-5
9. Griva K, Hansraj S, Thompson D, Jayasena D, Davenport A, Harrison M, Newman SP: Neuropsychological performance after kidney transplantation: A comparison between transplant types and in relation to dialysis and normative data. Nephrol Dial Transplant 2004; 19:1866-74
10. Trzepacz PT, McCue M, Klein I, Levey GS, Greenhouse J: A psychiatric and neuropsychological study of patients with untreated Graves' disease. Gen Hosp Psychiatry 1988; 10:49-55
11. Fahrenfort JJ, Wilterdink AM, van der Veen EA: Long term residual complaints and psychosocial sequelae after remission of hyperthyroidism. Psychoneuroendocrinology 2000; 25:201-11
12. Newman S, Klinger L, Venn G, Smith P, Harrison M, Treasure T: Subjective reports of cognition in relation to assessed cognitive performance following coronary artery bypass surgery. J Psychosom Res 1989; 33:227-33
13. Vingerhoets G, de Soete G, Jannes C: Subjective complaints *versus* neuropsychological test performance after cardiopulmonary bypass. J Psychosom Res 1995; 39:843-53
14. Berggren D, Gustafson Y, Eriksson B, Bucht G, Hansson L-I, Reiz, S, Winblad B: Postoperative confusion after anesthesia in elderly patients with femoral neck fractures. Anesth Analg 1987; 66:497-504
15. Crul BJ, Hulstijn W, Burger IC: Influence of the type of anaesthesia on post-operative subjective physical well-being and mental function in elderly patients. Acta Anaesthesiol Scand 1992; 36:615-20
16. Moller JT, Cluitmans P, Rasmussen LS, Houx P, Rasmussen H, Canet J, Rabbitt P, Jolles J, Larsen K, Hanning CD, Langeron O, Johnson T, Lauven, PM, Kristensen PA, Biedler A, van Beem H, Fraidakis O, Silverstein JH, Beneken JE, Gravenstein JS: Long-term postoperative cognitive dysfunction in the elderly: ISPOCD1 study. Lancet 1998; 351:857-61
17. Rasmussen LS, Johnson T, Kuipers HM, Kristensen VD, Siersma VD, Vila P, Jolles J, Papaioannou A, Abildstrom H, Silverstein JH, Bonal JA, Raeder J, Nielsen IK, Korttila K, Munoz L, Dodds C, Hanning CD, Moller JT: Does anaesthesia cause postoperative cognitive dysfunction? A randomised study of regional *versus* general anaesthesia in 438 elderly patients. Acta Anaesthesiol Scand 2003; 47:260-6
18. Berant A, Kaufman V, Leibovitz A, Habot B, Bahar M: Effects of anesthesia in elective surgery on the memory of the elderly. Arch Gerontol Geriatr 1995; 20:205-13
19. Williams-Russo P, Sharrock NE, Mattis S, Szatrowski TP, Charlson ME: Cognitive effects after epidural *versus* general anesthesia in older adults: A randomized trial. JAMA 1995; 274:44-50
20. Campbell DN, Lim M, Muir MK, O'Sullivan, G, Falcon M, Fison P, Woods R: A prospective randomised study of local *versus* general anaesthesia for cataract surgery. Anaesthesia 1993; 48:422-8
21. Williams-Russo P, Sharrock NE, Mattis S, Liguori GA, Mancuso C, Peterson MG, Hollenberg J, Ranawat C, Salvati E, Sculco T: Randomised trial of hypotensive epidural anesthesia in older adults. ANESTHESIOLOGY 1999; 91:926-35
22. Asbjorn J, Jakobsen BW, Pilegaard HK, Blom L, Ostergaard A, Brant MR: Mental function in elderly men after surgery during epidural analgesia. Acta Anaesthesiol Scand 1989; 33:369-73
23. Day JJ, Bayer AJ, McMahon M, Pathy MS, Spragg BP, Rowlands DC: Thiamine status, vitamin supplements and postoperative confusion. Age and Aging 1988; 17:29-34
24. Abildstrom H, Rasmussen LS, Rentowl P, Hanning CD, Rasmussen H, Kristensen PA, Moller JT: Cognitive dysfunction 1-2 years after non-cardiac surgery in the elderly. Acta Anaesthesiol Scand 2000; 44:1246-51
25. Rasmussen LS, Christiansen M, Rasmussen H, Kristensen PA, Moller JT: Do blood concentrations of neurone specific enolase and S-100 beta protein reflect cognitive dysfunction after abdominal surgery? Br J Anaesth 2000; 84:242-4
26. Canet J, Raeder J, Rasmussen LS, Enlund M, Kuipers HM, Hanning CD, Jolles J, Korttila K, Siersma VD, Dodds C, Abildstrom H, Sneyd JR, Vila P, Johnson T, Munoz CL, Silverstein JH, Nielsen IK, Moller JT: Cognitive dysfunction after minor surgery in the elderly. Acta Anaesthesiol Scand 2003; 47:1204-10
27. Rodriguez RA, Tellier A, Grabowski J, Fazekas A, Turek M, Miller D,

Wherrett C, Villeneuve PJ, Giachino A: Cognitive dysfunction after total knee arthroplasty. J Arthroplasty 2005; 20:763–71

28. Stockton P, Cohen-Mansfield J, Billig N: Mental status change in older surgical patients: Cognition, depression, and other comorbidity. Am J Geriatr Psychiatry 2000; 8:40–6

29. Ancelin ML, de Roquefeuil G, Ledesert B, Bonnel F, Cheminal JC, Ritchie K: Exposure to anaesthetic agents, cognitive functioning and depressive symptomatology in the elderly. Br J Psychiatry 2001; 178:360–6

30. Flatt JR, Birrell PC, Hobbes A: Effects of anaesthesia on some aspects of mental functioning of surgical patients. Anaesth Intensive Care 1984; 12:315–24

31. Jones MJ, Piggott SE, Vaughan RS, Bayer AJ, Newcombe RG, Twining TC, Pathy J, Rosen M: Cognitive and functional competence after anaesthesia in aged over 60: Controlled trial of general and regional anesthesia for elective hip or knee replacement. Anesth Analg 1987; 66:497–504

32. Rollason WN, Robertson GS, Cordiner CM, Hall DJ: A comparison of mental function in relation to hypotensive and normotensive anaesthesia in the elderly. Br J Anaesth 1971; 43:561–6

33. Enlund M, Mentell O, Flenninger A, Horneman G, Ronquist G: Evidence of cerebral dysfunction associated with isoflurane or propofol based anaesthesia for orthognathic surgery, as assessed by biochemical and neuropsychological methods. Ups J Med Sci 1998; 103:43–59

34. Moller JT, Svennild I, Johannessen NW, Jensen PF, Espersen K, Gravenstein JS, Copper JB, Djernes M, Johansen SH: Perioperative monitoring with pulse oximetry and late postoperative cognitive dysfunction. Br J Anaesth 1993; 71:340–7

35. Casati A, Fanelli G, Pietropaoli P, Proietti R, Tufano R, Danelli G, Fierro G, DeCosmo G, Servillo G: Continuous monitoring of cerebral oxygen saturation in elderly patients undergoing major abdominal surgery minimizes brain exposure to potential hypoxia. Anesth Analg 2005; 101:740–7

36. Jhaveri RM: The effects of hypocapnic ventilation on mental function in elderly patients undergoing cataract surgery. Anaesthesia 1989; 44:635–40

37. Prior FN, Chander P: Air as a vaporizing gas: Cognitive functions in elderly patients undergoing anaesthesia. Br J Anaesth 1982; 54:1207–12

38. Townes BD, Dikmen SS, Bledsoe SW, Hornbein TF, Martin DC, Janesheski JA: Neuropsychological changes in a young, healthy population after controlled hypotensive anesthesia. Anesth Analg 1986; 65:955–9

39. Elam JT, Marshall J, Graney MJ, Applegate WB, Miller ST, Freeman JM, Wood TO, Gettlefinger TC: Functional outcome one year following cataract surgery in elderly persons. J Gerontol 1988; 43:M122–6

40. Billig N, Stockton P, Cohen-Mansfield J: Cognitive and affective changes after cataract surgery in an elderly population. Am J Geriatr Psychiatry 1996; 4:29–38

41. Hall TA, McGwin G, Owsley C: Effect of cataract surgery on cognitive function in older adults. J Am Geriatr Soc 2005; 53:2140–4

42. Hughes D, Bowes JB, Brown MW: Changes in memory following general or spinal anaesthesia for hip arthroplasty. Anaesthesia 1988; 43:114–7

43. Biglar D, Adelhoj B, Petring OU, Pederson NO, Busch P, Kalhke P: Mental function and morbidity after acute hip surgery during spinal and general anaesthesia. Anaesthesia 1985; 40:672–6

44. Chung FF, Chung A, Meier RH, Lautenschlaeger E, Seyone C: Comparison of perioperative mental function after general anaesthesia and spinal anaesthesia with intravenous sedation. Can J Anaesth 1989; 36:382–7

45. Chung FF, Meier RH, Lautenschlager E, Carmichael FJ, Chung A: General or spinal anesthesia: Which is better in the elderly? ANESTHESIOLOGY 1987; 67:422–7

46. Haan J, van Kleef JW, Bloem BR, Zwartendijk J, Lanser JBK, Brand R, van der Does IGF, Krul EJT, Elshove HM, Moll AC, Buruma OJS, Roos RAC: Cognitive function after spinal or general anesthesia for transurethral prostatectomy in elderly men. J Am Geriatr Soc 1991; 39:596–600

47. Nielson WR, Gelb AW, Casey JE, Penny FJ, Merchant RN, Manninen PH: Long-term cognitive and social sequelae of general *versus* regional anesthesia during arthroplasty in the elderly. ANESTHESIOLOGY 1990; 73:1103–9

48. Casati A, Aldegheri G, Vinciguerra F, Marsan A, Fraschina G, Torri G: Randomized comparison between sevoflurane anaesthesia and unilateral spinal anaesthesia in elderly patients undergoing orthopaedic surgery. Eur J Anaesth 2003; 20:640–6

49. Karhunen U, Jonn G: A comparison of memory function following local and general anaesthesia for extraction of senile cataract. Acta Anaesthesiol Scand 1982; 26:291–6

50. Ghoneim MM, Hinrichs JV, O'Hara MW, Mehta MP, Pathak D, Kumar V, Clark CR: Comparison of psychologic and cognitive functions after general or regional anesthesia. ANESTHESIOLOGY 1988; 69:507–15

51. Riis J, Lomholt B, Haxholdt O, Kehlet H, Valentin N, Danielsen U, Dyrberg V: Immediate and long-term mental recovery from general *versus* epidural anesthesia in elderly patients. Acta Anaesthesiol Scand 1983; 27:44–9

52. Farrag AK, Khadr EM, Abdel-Aleem H, Rageh TA: Effect of surgical menopause on cognitive functions. Dement Geriatr Cogn Disord 2002; 13:193–8

53. Dijkstra JB, Houx PJ, Jolles J: Cognition after major surgery in the elderly: Test performance and complaints. Br J Anaesth 1999; 82:867–74

54. Goldstein MZ, Young BL, Fogel BS, Benedict RHB: Occurrence and predictors of short-term mental and functional changes in older adults undergoing elective surgery under general anesthesia. Am J Geriatr Psychiatry 1998; 6:42–52

55. Gilberstadt H, Aberwald R, Crosbie S, Schuell H, Jimenez E: Effect of surgery on psychological and social functioning in elderly patients. Arch Intern Med 1968; 122:109–15

56. Johnson T, Monk T, Rasmussen LS, Abidstrom H, Houx P, Korttila K, Kuipers HM, Hanning CD, Siersma VD, Kristensen D, Canet J, Ibanaz MT, Moller JT: Postoperative cognitive dysfunction in middle-aged patients. ANESTHESIOLOGY 2002; 96:1351–7

57. Chung F, Seyone C, Dyck B, Chung A, Ong D, Taylor A, Stone R: Age-related cognitive recovery after general anesthesia. Anesth Analg 1990; 71:217–24

58. Treasure T, Smith PL, Newman S, Schneidau A, Joseph P, Ell P, Harrison MJ: Impairment of cerebral function following cardiac and other major surgery. Eur J Cardiothorac Surg 1989; 3:216–21

59. Shaw PJ, Bates D, Cartlidge NE, French JM, Heaviside D, Julian DG, Shaw DA: Neurologic and neuropsychological morbidity following major surgery: Comparison of coronary artery bypass and peripheral vascular surgery. Stroke 1987; 18:700–7

60. Grichnik KP, Ijsselmuiden AJ, D'Amico TA, Harpole DH Jr, White WD, Blumenthal JA, Newman MF: Cognitive decline after major noncardiac operations: A preliminary prospective study. Ann Thorac Surg 1999; 68:1786–91

61. Goldstein MZ, Fogel BS: Cognitive change after elective surgery in nondemented older adults. Am J Geriatr Psychiatry 1993; 1:118–25

62. Lezak MD: Neuropsychological Assessment, 3rd edition. New York, Oxford University Press, 1995

63. Newman SP: Analysis and interpretation of neuropsychological tests in cardiac surgery. Ann Thorac Surg 1995; 59:1351–5

64. Elkins JS, Longstreth WT Jr, Manolio TA, Newman AB, Bhadelia RA, Johnston SC: Education and the cognitive decline associated with MRI-defined brain infarct. Neurology 2006; 67:435–40

65. Newman MF, Croughwell ND, Blumenthal JA, White WD, Lewis JB, Smith LR, Frasco P, Towner EA, Schell RM, Hurwitz BJ, Reves JG: Effect of aging on cerebral autoregulation during cardiopulmonary bypass: Association with postoperative cognitive dysfunction. Circulation 1994; 90(pt 2):ll-243–9

66. Borowicz LM, Goldsborough MA, Selnes OA, McKhann GM: Neuropsychologic change after cardiac surgery: A critical review. J Cardiothorac Vasc Anesth 1996; 10:105–12

67. Stygall J, Fitzgerald G, Steed L, Mulligan K, Arrowsmith JE, Humphries S, Harrison MJ, Newman SP: Cognitive change five years after coronary artery bypass surgery. Health Psychol 2003; 22:579–86

68. Selnes OA, Royall RM, Grega MA, Borowicz LM Jr, Enger C, Quaskey SA, McKhann GM: Cognitive changes 5 years after coronary artery bypass grafting. Arch Neurol 2001; 58:598–604

69. Rasmussen LS, Siersma VD, ISPOCD Group: Postoperative cognitive dysfunction: True deterioration *versus* random variation. Acta Anesthesiol Scand 2004; 48:1137–43

70. Iohom G, Szarvas S, Larney V, O'Brien J, Buckley E, Butler M, Shorten G: Perioperative plasma concentrations of stable nitric oxide products are predictive of cognitive dysfunction after laparoscopic cholecystectomy. Anesth Analg 2004; 99:1245–52

71. Newman MF, Kirchner JL, Philips-Bute B, Gaver V, Grocott H, Jones RH, Mark DB, Reves JG, Blumenthal JA: Longitudinal assessment of neurocognitive function after coronary-artery bypass surgery. N Engl J Med 2001; 344:395–402

72. Edmonds CR, Barbut D, Hager D, Sharrock NE: Intraoperative cerebral arterial embolization during total hip arthroplasty. ANESTHESIOLOGY 2000; 93:315–8

73. Keith JR, Puente AE, Malcolmson KL, Tartt S, Coleman AE, Marks HF Jr: Assessing postoperative cognitive change after cardiopulmonary bypass surgery. Neuropsychology 2002; 16:411–21

74. Lewis MS, Maruff P, Evered LA, Scott DA: Detection of postoperative cognitive decline after coronary artery bypass surgery is affected by the number of neuropsychological tests in the assessment battery. Ann Thorac Surg 2006; 81:2097–104

75. Mahanna E, Blumenthal JA, White WD, Croughwell ND, Clancy CP, Smith R, Newman MF: Defining neuropsychological dysfunction after coronary artery bypass grafting. Ann Thorac Surg 1996; 61:1342–7

76. Altman DG, Royston P: The cost of dichotomizing continuous variables. BMJ 2006; 332:1080

77. Arrowsmith JE, Harrison MJ, Newman SP, Stygall J, Timberlake N, Pugsley WB: Neuroprotection of the brain during cardiopulmonary bypass: A randomized trial of remacemide during coronary artery bypass in 171 patients. Stroke 1998; 29:2357–62

## Appendix: Measures Used

### Verbal and Language Skills

A1 Wechsler Adult Intelligence Scale–Revised (WAIS-R) Vocabulary
A2 WAIS-R Information
A3 WAIS-R Similarities
A4 Speed of writing
A5 Controlled oral word association test
A6 Boston Naming Test
A7 Alphabet

## *Memory and Learning*

B1 Rey Auditory Verbal Learning Test
B2 Continuous paired associate leaning
B3 Bushke verbal selective reminding test
B4 Wechsler Memory Scale (WMS)–Logical memory
B5 WMS–Visual reproduction
B6 WMS–Associative Learning
B7 WMS–Mental control
B8 WMS–Information
B9 WMS–Digit Total
B10 WMS–Personal and current information
B11 WMS–Orientation
B12 Taylor-Rey-Osterreith test battery
B13 Word list
B14 Randt memory test
B15 Prose passage/story recall
B16 Benton visual retention test
B17 Chandigarh memory scale
B18 Luria memory test
B19 Delayed recall test
B20 Visual Gestalt learning
B21 Picture recognition
B22 Recognition memory task
B23 Mattis-Kovner verbal recall
B24 Mattis-Kovner verbal recognition
B25 Benton visual recognition test
B26 Object learning test
B27 WAIS Digit Span
B28 Visual verbal leaning test
B29 Memory scanning test
B30 Unknown or unclear memory test (self-devised)
B31 Rivermead behavioral memory
B32 Fuld object memory
B33 Free recall task
B34 Baibizet and Cany visual recognition

## *Attention, Concentration, and Perception*

C1 Attention and Concentration Index
C2 WAIS-R digit–symbol or similar
C3 Symbol digit modalities test
C4 Trailmaking test A
C5 Trailmaking test B
C6 Unclear vigilance task
C7 Letter or symbol cancellation
C8 Reaction time tests
C9 Digit vigilance

C10 Ishihara color plates
C11 Concept shifting tasks (trails)
C12 Flicker fusion threshold
C13 Two-point discrimination
C14 Visual search

## *Visual and Spatial Skills*

D1 Hooper test (visual organization)
D2 WAIS-R block design
D3 Stroop color word interference
D4 Line drawings
D5 Bender-Gestalt test
D6 WAIS-R object assembly

## *Visuomotor and Manual Skills*

E1 Finger tapping
E2 Purdue pegboard
E3 Digit/words copying tests
E4 Steadiness

## *Numerical*

F1 Arithmetic
F2 Serial sevens subtraction
F3 Counting

## *Executive Functions*

G1 Maze test
G2 Card sort test

## *Composite Measures*

H1 Mini-Mental Status Examination
H2 Shipley Hartford examination
H3 Wechsler Adult Intelligence Scale–Revised
H4 Abbreviated Mental Test
H5 Examen Cognitif per Ordinateur
H6 Mattis Organic Mental Screening Examination
H7 Wechsler Memory Scale
H8 Iowa Battery of Mental Decline
H9 Rivermead Memory Scale