Coronary artery bypass grafting (CABG) with the use of cardiopulmonary bypass (CPB) is associated with significant cerebral morbidity. The two main clinical manifestations of brain injury after CABG are stroke and cognitive decline. The incidence of postoperative stroke is consistently reported to be around 3%-12% and as high as 9% in patients older than 75 years. However, the reported incidence of cognitive decline varies widely. This variability is partly caused by methodologic problems: a multitude of definitions of cognitive decline are used, and a large number of neuropsychologic tests exist to assess the various cognitive domains. Moreover, the interval between operation and administration of the neuropsychologic tests may range from a few days to several years. Some literature reviews on cognitive deficits after CPB have been published, but the methodologic variability of the studies hampers comparison of the results.

At a consensus meeting in 1994, several guidelines for the assessment of neuropsychologic deficits after CPB were established. It was recommended that the neurologic and neuropsychologic state be assessed before the operation to provide accurate baseline information. A second important recommendation was that the analysis should be based on the individual change in performance from baseline to a particular time after the operation. In general, practice effects cause the overall group performance to improve after the operation. Accordingly, when the overall postoperative mean is compared with the preoperative mean, the decline of

Objective: Substantial, albeit scattered, evidence suggests that coronary artery bypass grafting may impair cognitive function. As methods and definitions differ greatly across studies, the reported incidence of cognitive decline after coronary bypass surgery varies widely as well. The aim of the present study was to systematically review those studies on cognitive decline that are relatively comparable and meet with certain quality criteria.

Methods: Four electronic databases and the references of several abstract books and earlier reviews were used to identify relevant literature. Stringent criteria, based in part on the 1994 consensus meeting on assessment of neurobehavioral outcomes after cardiac surgery, were used to assess the studies that were found. In total, 256 different titles were found, of which 23 met with the formulated selection criteria.

Results: Twelve cohort studies and eleven intervention studies were evaluated. A pooled analysis of six highly comparable studies yielded a proportion of 22.5%-95% confidence interval, 18.7%-26.4%) of patients with a cognitive deficit (a decrease of at least 1 standard deviation in at least two of nine or ten tests) 2 months after the operation.

Conclusions: Neurocognitive dysfunction is a frequently occurring complication of coronary artery bypass grafting. The etiologic contribution of cardiopulmonary bypass to this complication will remain unclear until a randomized trial that directly compares off-pump and on-pump bypass surgery is carried out.
some individuals is overshadowed by the improvement of others. Also, it was agreed that a late assessment (ideally after 3 months) should be included in the study design, because the patients’ performances are unstable in the immediate postoperative period.

Although these consensus guidelines were developed for the design of new research protocols, we found some of them also suitable to interpret studies carried out in the past. This systematic review is restricted to studies on cognitive decline 1 to 12 months after CABG, which were considered methodologically homogeneous according to well-defined and strict criteria. Both the primary selection of studies and the assessment and comparison of the studies that were included were largely based on recommendations of the 1994 consensus. As a result of the technical improvements of CPB and the alterations in anesthetic management, studies published before 1980 were considered less relevant and therefore were excluded.

Methods

Literature search. For the literature search, four electronic databases (MEDLINE, PsychLit, PubMed, and Current Contents) and references of four earlier published (abstract) books and three reviews were used. The precise search strategy is described in Table I. To assess the quality of the search strategy, we sampled ten studies that we already had on file and considered relevant. The search strategy was able to identify these articles. The total number of different titles found was 256 (Table I).

<table>
<thead>
<tr>
<th>Source</th>
<th>No. of titles found</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEDLINE electronic database 1980-1999: Medical Subject Heading (MeSH) standard terms: (heart surgery or coronary artery bypass or extracorporeal circulation or cardiopulmonary bypass; cognition disorders) and (cognitive symptoms or neuropsychology or neuropsychological tests)</td>
<td>161</td>
</tr>
<tr>
<td>PsychLit electronic database 1980-1999: standard terms: (heart surgery) and (neuropsychological assessment or cognitive assessment or cognition or concentration or attention)</td>
<td>10</td>
</tr>
<tr>
<td>Book: “Cardiac Surgery and the Brain” by Smith and Taylor: relevant references of relevant chapters</td>
<td>67</td>
</tr>
<tr>
<td>The references of three recent reviews on cardiac surgery and cognitive decline</td>
<td>91</td>
</tr>
<tr>
<td>Total number of different titles</td>
<td>256</td>
</tr>
</tbody>
</table>

In the numbers of titles found, articles published before 1980 and books are not included.

Selection of the literature. The studies were independently judged by a clinician and a psychologist using the following selection criteria:

Inclusion criteria:

- Primary research on cognitive decline after CABG, including a preoperative neuropsychologic assessment to provide baseline information
- Postoperative neuropsychologic assessment between 1 and 12 months after the operation
- Data analysis based on the individual change in performance from baseline to a time after the operation

Exclusion criteria:

- Studies published before 1980
- Studies including open chamber or valvular procedures
- Studies with unclear timing of test administration
- Articles describing the same or an overlapping patient sample as other articles already included in the review

These selection criteria were applied on the 256 titles found. Thirty-four articles could be excluded on the basis of the title only and 111 articles were excluded on the basis of the abstract. For the remaining 111 titles, the complete article was studied. The majority of these publications were written in English. Many articles were excluded because data of patients undergoing CABG were mixed with data of patients having valve replacement, because data analysis was based on comparison of mean group performance before and after the operation, or because only an early postoperative neuropsychologic assessment (ie, less than 1 month) was performed. Seven studies were excluded because the same patient series had (partly) been used in other studies already included in the review and their inclusion would have overemphasized the results of these series. Finally, one article published in Japanese was rejected because a translation could not be obtained. Only 23 articles matched all the selection criteria.

Literature processing. The 23 articles meeting the selection criteria were processed independently by a psychologist and a clinician using a three-page standard form. This form was filled out to systematically assess the research question, study design, neuropsychologic tests used, statistics, main conclusions, and other items. The forms were the basis for Tables II and III in the “Results” section.

The neuropsychologic sections were independently judged by two trained neuropsychologists, blinded for author and journal names. Again, several recommendations of the Statement of Consensus were taken into account.
Neuropsychologic test batteries were rated with respect to (1) the cognitive domain of the tests, (2) the sensitivity of the tests, (3) the availability of parallel forms of the tests, and (4) the overall balance of the cognitive domains assessed in the battery. So that the comparability of the studies could be further improved, (5) a recommended core battery was taken as the basis for our assessment of quality of the test batteries used. This core battery minimally includes the Rey Auditory Verbal Learning Task, the Trail Making Tests A and B, and the Grooved Pegboard. These tests are widely used, easy to conduct, well-normalized, and sensitive to cerebral damage.

**Analysis.** Studies were classified as (1) cohort studies, aiming to determine the incidence of cognitive decline at a certain moment after the operation and/or to identify determinants of cognitive decline, or (2) intervention (controlled) studies, investigating the cerebral protective effect of intraoperative interventions.

To determine whether a pooled analysis could be carried out, we assessed the comparability of the studies entered into the review in terms of precise timing of neuropsychologic testing, comparability of tests, and definition of decline.

**Results**

**Cohort studies.** The twelve cohort studies meeting the selection criteria are presented in Table II. Some of these studies exclusively aimed to determine the incidence of neuropsychologic decline, whereas others were designed to identify determinants of cognitive decline. The studies in Table II are ordered by year of publication. The mean age of the patients studied increased over time, but no clear time trend could be observed in the incidences of cognitive deficits, which varied from 4% to 47%.

None of the studies designed as a correlation study was able to identify determinants of cognitive decline in the primary analysis. However, in a multivariate analysis, Tardiff and associates found a significant association between short-term memory dysfunction after CABG and a variant form of the apolipoprotein E gene, especially in patients with lower educational levels. This gene encodes the APOE protein, which is responsible for repair of neuronal injury and probably involved in the development of Alzheimer disease.

**Intervention studies.** The eleven intervention studies included are shown in Table III. Ten of them were randomized trials. In two studies, the initial data analysis demonstrated that the intervention studied led to a statistically significantly ($P < .05$) decreased risk of cognitive decline. Hammon and colleagues compared the results in patients operated on in 1991 and 1992 with results in patients operated on in the next 2 years. In the second time period, the surgical team had adopted the use of epiaortic scanning and had increased the use of the single crossclamp technique and left ventricular venting. This combined strategy reduced the incidence of cognitive decline by 11% ($P = .01$). The other study
with a significant result in the initial analysis showed a beneficial effect of the use of an arterial line filter.\textsuperscript{35}

Many intervention studies had insufficient statistical power to detect clinically meaningful differences. In some of these studies, post hoc use of another definition of cognitive decline led to significant results. One example is the remacemide trial by Arrowsmith and coworkers.\textsuperscript{33} Individual cognitive decline was less frequent in the treated group, but the study lacked power to reach statistical significance. Redefining cognitive decline as “overall postoperative change” made the favorable effect of remacemide statistically significant.

Pulsatile blood flow during CPB did not improve neurocognitive outcome in 316 patients,\textsuperscript{36} but in the same series, alpha-stat blood gas management reduced the incidence of cognitive deficits. This reduction was significant in patients with a bypass time of more than 90 minutes. The protective effect of alpha-stat blood gas management was also found by Patel and associates.\textsuperscript{37} The observed difference became statistically significant when a more stringent definition of neuropsychologic deficit (decline on three tests instead of two tests) was used.

The three studies on the influence of perfusion temperature\textsuperscript{39–41} failed to demonstrate a clear advantage of hypothermia compared with normothermia, but the sample sizes of the studies were relatively small, and in the study of Heyer and colleagues\textsuperscript{41} only 26% of the study patients were available for follow-up.

**Neuropsychology.** The selected studies included neuropsychologic tests as the measure of cognitive decline. Of the 23 selected studies, six included the core battery as recommended by the Statement of Consensus.\textsuperscript{11} In addition, four studies used tests similar to the recommended core battery (see Tables II and III). Accordingly, ten studies were at least partially comparable and met with widely accepted quality criteria.

Four study groups\textsuperscript{5,22,30,37} used several definitions of cognitive decline in their samples. The definition used most frequently (in nine studies) was a postoperative deterioration of at least 1 standard deviation (of the population’s performance at baseline) compared with preoperative testing in at least two tests. In one study the same definition was used with a minimum of three tests showing deterioration, whereas one other study took deterioration in one test as criterion. Comparability in terms of definition of cognitive decline was thus limited to nine studies.

**Pooled analysis.** Four cohort studies\textsuperscript{23–25,28} and two intervention studies\textsuperscript{33,35} were comparable not only in definition of decline and use of the core battery or comparable tests, but also in timing of test administration.

These six studies in total included 505 patients, of whom 448 completed 2 months of follow-up (Table IV). We combined the results of these studies in a weighted average (the sum of the proportions of patients with cognitive decline per study times number of patients per study, divided by the total number of patients). For the two intervention studies, the weighted mean of the two treatment groups was used. On average, 22.5% of the CABG patients had a decline of at least 1 standard deviation in at least two of a total of nine or ten tests 2 months after their operation ($P < .0001$; 95% confidence interval, 18.7%-26.4%).

**Discussion**

This study systematically reviewed reports on cognitive decline after CABG. The literature search was extensive. In an attempt to select studies that could be compared, we formulated stringent selection criteria. A limitation of using these selection criteria is that some informative aspects of research were rejected. The relatively small number of the studies that could be included in the review underlines the lack of comparability of studies carried out until now.

The reported incidences of cognitive deficits varied widely. At least in part, this can be explained by the differences in timing of test administration and definitions of cognitive deficit. A high loss to follow-up in 3 studies\textsuperscript{5,24,41} may also have influenced the incidences found, as loss to follow-up is seldom random.\textsuperscript{6} Most of the included studies were comparable in terms of neuropsychologic tests used, which is probably encouraged by the Statement of Consensus that was held in 1994.\textsuperscript{11}

Within the six studies used for the pooled analysis, there was still a considerable variation in the reported incidence of cognitive decline (10%-38% 2 months after the operation). It is not possible to conclude from these studies that the incidence of cognitive deficits has decreased in the past 10 years. Improved outcome as a result of better anesthesiologic and perfusion management may be offset by the increasing age of the patients, which in itself is associated with an increased risk for cerebral complications.\textsuperscript{36,42} With the six selected studies, we calculated an average incidence of cognitive deficits of 22.5% 2 months after CABG. This figure must be interpreted with caution because, even within these six studies, methodologic differences were present. For example, the cutoff value of 1 standard deviation, used to define a deficit, varied per study, because the baseline performances of the six patient series are inevitably not the same. A formal meta-analysis, which is typically performed with a series of
methodologically comparable randomized trials, includes the use of more advanced weighing factors, testing of homogeneity of the effect estimates of the different studies, and a sensitivity analysis. However, the methodology of meta-analysis of uncontrolled observational studies is subject to debate, and its use in this review would have unjustly suggested a high level of objectiveness and precision.

The clinical meaning of a decline of 1 standard deviation in two tests is relative, because the figure calculated, 22.5%, does not indicate the percentage of patients who are significantly disabled after CABG. Several authors emphasize that, in most patients, the deficit does not matter to the patient in functional terms. Apparently, many activities of daily life do not require the level of performance called for during neuropsychologic testing.

Table III. Intervention studies

<table>
<thead>
<tr>
<th>First author</th>
<th>Year of publication</th>
<th>No. of patients</th>
<th>Mean age (y)</th>
<th>NP-test timing (mo)</th>
<th>Lost to follow-up (%)</th>
<th>No. of tests used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fish</td>
<td>1987</td>
<td>2 × 50</td>
<td>58</td>
<td>2</td>
<td>26</td>
<td>10</td>
</tr>
<tr>
<td>Arrowsmith</td>
<td>1998</td>
<td>87 + 84</td>
<td>59</td>
<td>2</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>Sellman</td>
<td>1993</td>
<td>3 × 20</td>
<td>59</td>
<td>1/6</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>Pugsley</td>
<td>1994</td>
<td>53 + 52</td>
<td>55</td>
<td>2</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Markin</td>
<td>1995</td>
<td>4 × 79</td>
<td>61</td>
<td>2</td>
<td>24</td>
<td>5</td>
</tr>
<tr>
<td>Patel</td>
<td>1996</td>
<td>2 × 35</td>
<td>57</td>
<td>1.5</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Gold</td>
<td>1995</td>
<td>2 × 124</td>
<td>66</td>
<td>6</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>Mora</td>
<td>1996</td>
<td>54 + 55</td>
<td>63</td>
<td>1.5</td>
<td>23</td>
<td>5</td>
</tr>
<tr>
<td>Regragui</td>
<td>1996</td>
<td>31 + 36 + 29</td>
<td>59</td>
<td>1.5</td>
<td>27</td>
<td>7</td>
</tr>
<tr>
<td>Heyer</td>
<td>1997</td>
<td>46 + 53</td>
<td>64</td>
<td>1.5</td>
<td>74</td>
<td>5</td>
</tr>
<tr>
<td>Hammon</td>
<td>1997</td>
<td>192 + 203</td>
<td>61</td>
<td>1.5</td>
<td>?</td>
<td>11</td>
</tr>
</tbody>
</table>

NP-test timing, Time between CABG and postoperative neuropsychologic test administration; Definition of cognitive deficit: 1 (≥1 SD in ≥1 test); 2 (≥1 SD in ≥2 tests); 3 (≥1 SD in ≥3 tests); 4 (≥0.5 SD in a domain); 5 (≥20% in ≥2 tests); 6 (≥20% in ≥20%); 7 (other); Comparability: ++, tests of “core battery” are used (Rey Auditory Verbal Learning Task, Trail Making Tests A and B, Grooved Pegboard); +, comparable tests were used; –, tests were not comparable with core battery.

Table IV. The four cohort studies and two intervention studies that were comparable

<table>
<thead>
<tr>
<th>First author</th>
<th>Year of publication</th>
<th>No. of patients</th>
<th>Mean age (y)</th>
<th>NP-test timing (mo)</th>
<th>No. of tests used</th>
<th>No. of patients who completed follow-up</th>
<th>Proportion of patients with cognitive deficit (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treasure</td>
<td>1989</td>
<td>76</td>
<td>55</td>
<td>2</td>
<td>10</td>
<td>67</td>
<td>37</td>
</tr>
<tr>
<td>Harrison</td>
<td>1989</td>
<td>78</td>
<td>55</td>
<td>2</td>
<td>10</td>
<td>47</td>
<td>36</td>
</tr>
<tr>
<td>Pugsley</td>
<td>1994</td>
<td>105</td>
<td>55</td>
<td>2</td>
<td>10</td>
<td>100</td>
<td>17.45*</td>
</tr>
<tr>
<td>Toner</td>
<td>1996</td>
<td>61</td>
<td>60</td>
<td>2</td>
<td>10</td>
<td>61</td>
<td>38</td>
</tr>
<tr>
<td>Arrowsmith</td>
<td>1998</td>
<td>171</td>
<td>59</td>
<td>2</td>
<td>9</td>
<td>159</td>
<td>10.45*</td>
</tr>
<tr>
<td>Braeckden</td>
<td>1998</td>
<td>14</td>
<td>64</td>
<td>2</td>
<td>10</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>505</td>
<td></td>
<td></td>
<td></td>
<td>448</td>
<td>22.5</td>
</tr>
</tbody>
</table>

The four cohort studies and two intervention studies that were comparable in definition of cognitive decline (a decline in performance at least 1 standard deviation in at least two tests), use of the core battery (Rey Auditory Verbal Learning Task, the Trail Making Tests A and B, and the Grooved Pegboard) or comparable neuropsychologic tests, and timing of test administration (2 months after operation). These six studies include 505 patients, of whom 448 completed 2 months of follow-up. To calculate a weighted average, the proportion of patients with cognitive decline per study was multiplied with the number of patients per study who completed follow-up. The sum of the multiplications was then divided by the total number of patients who completed follow-up. On average, the proportion of patients with cognitive decline was 22.5% (P < .0001; 95% confidence interval, 18.7%-26.4%).

* Intervention studies, in which the weighted mean of the proportions of patients with a deficit in the two treatment groups was used.
However, a small proportion of patients with intellectual dysfunction or memory deficits become sufficiently disabled to prevent return to employment.\textsuperscript{1,22}

The discrepancy between decline in test performance and functional decline is also expressed by the methodologic difficulties of defining a cognitive deficit. Mahanna and colleagues\textsuperscript{5} demonstrated the enormous influence of the definition of cognitive deficit that is chosen by applying five different definitions on the same patient sample. Depending on the definition used, the incidence of cognitive deficit ranged from 1.1% to 34% at 6 weeks and from 3.4% to 19.4% at 6 months postoperatively.

Most authors defined deficit as a certain deterioration in one or more tests, which may seem accurate. However, some tests comprise more than one test variable, and from most studies it was not clear whether test deficit meant deterioration in one or all of the variables of one test. This obviously creates a multitude of possible outcomes and is therefore a factor complicating the interpretation of the incidence data. A deterioration of 1 standard deviation in postoperative functioning compared with preoperative functioning was the most frequently used cutoff value. This is arbitrary and does not necessarily reveal real cognitive change, since it does not take into account the reliability of the change scores. Practice effects were almost always mentioned but not included in analyses, which may have resulted in an underestimation of incidence figures. There is emerging recognition of the importance of defining real change in test-retest scores as opposed to artifactual change resulting from low test reliability and susceptibility to practice effects. Recent research is increasingly focused on the use of reliable change indices. These indices define the range in which an individual score is likely to fluctuate because of the imprecision of the measure, providing statistically based cutoff scores for cognitive change on each measure.\textsuperscript{44,45}

For example, a 90% reliable change interval is calculated on the basis of the correlation and the standard error of the difference between baseline and follow-up scores of control subjects. When the difference between an individual patient’s postoperative and preoperative scores falls outside this interval, he or she is considered to have a statistically significant change in performance on this particular neuropsychologic measure.\textsuperscript{44}

The single-case analysis technique, recommended in the consensus statements,\textsuperscript{11,46} uses the patient as his or her own control and defines a cognitive deficit as a 20% decrease in at least 20% of the tests. This method also has some drawbacks. In the first place, reducing the continuous test scores to a dichotomous outcome measure (presence of a 20% decrease or not) is a “costly” way of data handling that reduces statistical power and may have made several randomized studies fail to
reach statistically significant results. The 20% decrease rule is as arbitrary as the 1 standard deviation decrease rule. The problem may be overcome by refraining from “dichotomizing” data and just calculating how much the patient’s performance deviates from the expected (baseline or control group) performance. Second, due to floor effects it may be difficult to demonstrate a deficit in patients with a low preoperative test performance. In the third place, as demonstrated by Browne and colleagues, “regression toward the mean” may strongly influence single-case definitions of cognitive deterioration. High baseline performers may be wrongly classified as impaired and low baseline performers may not show a deficit although deterioration had actually occurred. This problem can be overcome by using group mean analysis, which is free from the influence of regression to the mean (in contrast to analysis by single-case definitions). As discussed before, comparison of mean group performance before and after surgery does have disadvantages, especially if large practice effects are present. However, if a suitable control group is available (randomized trial), comparison of group means allows for the control of both practice effects and regression to the mean.

Conclusions

A pooled analysis of six highly comparable studies yielded a proportion of 22.5% of patients with a cognitive deficit 2 months after CABG. Although this percentage is partly determined by the definition of cognitive deficit that was used, it demonstrates that cognitive dysfunction is a frequently occurring complication of CABG. The etiologic contribution of CPB to this complication will remain unclear until a randomized trial that directly compares off-pump CABG with on-pump CABG is carried out. To improve comparability of future studies, we advocate that researchers use the guidelines of the 1994 consensus meeting. However, the recommended single-case analysis technique has some drawbacks and may be replaced by other analysis techniques, especially when a control group is included in the study design.

We thank Professor D. E. Grobbee, Professor C. J. Kalkman, Professor R. S. Kahn, and Doctor E. Buskens for revising the manuscript.

REFERENCES


