Anesthesiology
77:1095–1104, 1992

Anesthetic Techniques during Surgical Repair of Femoral Neck Fractures

A Meta-analysis

Robert M. Sorenson, M.D., Nathan L. Pace, M.D.†*

Fracture of the hip typically occurs in older women. These patients usually have serious accompanying chronic illnesses. There is a difference of opinion as to the choice of regional *versus* general anesthesia for surgery in these patients. This meta-analysis compared survival of patients with traumatic femoral neck fractures undergoing operative repair during regional or general anesthesia. The data sources were articles comparing regional and general anesthesia from peer reviewed journals. Thirteen randomized controlled trials were found. Besides 1-month mortality, variables used were estimated operative blood loss and the incidence of deep venous thrombosis. For dichotomous outcomes, two effect measures were calculated: the difference in probabilities and the odds ratio. For blood loss, a continuous variable, the effect measure was the mean difference in blood loss. A random-effects Bayesian meta-analysis was used to combine study data, estimate parameters and create 95% confidence intervals. Only the incidence of deep venous thrombosis was clearly greater for patients receiving general anesthesia, being 31 percentage points higher than for patients receiving regional anesthesia. By the odds ratio, deep venous thrombosis was almost four times more likely following general anesthesia. There was no difference in estimated operative blood loss. By probability difference, mortality was a nonsignificant 2.7 percentage points less following regional anesthesia. By odds ratio effect measure, death was 1.5 times more likely following general anesthesia, but the lower bound of the 95% confidence interval was close to 1. Meta-analysis does not allow a conclusion that important differences in mortality exist between regional and general anesthesia for traumatic hip fracture surgery. (Key words: Anesthetic techniques: general; regional. Complications. Statistics: meta-analysis.)

THE SAFETY OF regional anesthesia (RA) *versus* general anesthesia (GA) has long been debated. In 1933 Nygaard published a study comparing spinal and general anesthesia;[1] Nygaard found fewer postoperative complications with spinal anesthesia than with open drop ether anesthesia. Since then, numerous additional investigators have compared outcome following RA and GA. These studies show conflicting results. Many of the studies have been limited by small study size, having low statistical power to reveal real differences. It remains unclear whether there is any patient benefit in choosing between RA and GA.

Proximal femoral neck fractures are a substantial health threat to older people; by one estimate there are more than 200,000 such fractures each year in the United States.[2] For at least three decades, the standard surgical approach for this fracture has been urgent open reduction and internal fixation.[3] This volume of surgical cases has generated considerable enthusiasm for the possible benefits of RA in decreasing operative blood loss, deep venous thrombosis, and mortality. We used meta-analysis to compare outcomes from published randomized controlled clinical trials (RCTs) that compared mortality and morbidity differences between GA and RA.

Methods

There are two major tasks in performing meta-analysis, data acquisition and data analysis. Our data acquisition followed accepted guidelines.[4–6] Although most published meta-analyses have used frequentist statistical methods, we used Bayesian statistical methods, implemented as the confidence profile method.[7,8]

DATA ACQUISITION

A Medline literature search was conducted of all reports making a comparison between RA (epidural and spinal) and GA since 1966. From these references, those involving RCTs in patients having surgical repair of femoral neck fractures were chosen. The bibliographies of each of the relevant reports were searched for other studies. The reference lists of current anesthesia textbooks were searched. No abstracts or reports of meeting presentations were included, and no attempt was made to obtain results of unpublished studies. Literature searching was stopped December 31, 1991.

The definitions of anesthesia were as follows.

GA: anesthesia with loss of consciousness. Tracheal intubation was not required to satisfy the criteria for GA. Neuroleptic anesthesia was included within general anesthesia. Combined general/regional anesthesia was not used in any study.

RA: spinal anesthesia (SAB) or epidural anesthesia (EPI).

* Resident.

† Professor of Anesthesiology, Adjunct Professor of Bioengineering.

Received from the University of Utah School of Medicine, Salt Lake City, Utah. Accepted for publication August 7, 1992. Presented in part at the Annual Meeting of the Association of University Anesthesiologists, Philadelphia, Pennsylvania, May 3–5, 1991, and at the annual meeting of the American Society of Anesthesiologists, San Francisco, California, October 26–30, 1991.

Address reprint requests to Dr. Pace: Department of Anesthesiology, University of Utah School of Medicine, 3C-205 Medical Center, Salt Lake City, Utah 84132.

Common techniques and drugs were always used. Continuous spinal anesthesia was not mentioned in any article.

The results of comparisons between the anesthetic techniques were tabulated. The outcome variables included mortality and various morbid events. Covariates such as age, male to female ratio, use of antithromboembolism prophylaxis, ASA physical status, elapsed time from injury to surgery, surgery duration, follow-up duration and methods of postoperative analgesia were tallied.

Varying definitions of morbidity and mortality were evident in the published reports. For each outcome variable, a definition was chosen that was sufficiently broad to allow inclusion of as many reports as possible, but also sufficiently focused to make data combining meaningful.

Death: A death was considered perioperative if it occurred within 30 days of the operation. There was no attempt to tally causation since only some studies reported this.

Estimated blood loss (EBL): blood loss during operation. This did not include postoperative blood losses, although some studies reported blood loss for both periods.

Deep venous thrombosis (DVT): a DVT of the lower extremity. Two diagnostic methods were reported. ^{125}I-fibrinogen leg scanning and lower extremity venography.

Other morbid events such as nausea and/or vomiting, cardiovascular events, urinary retention, neuropsychiatric problems, pulmonary embolism, and pneumonia, were reported, but were not tallied because of varying and unclear definitions and the absence of systematic and unbiased application of diagnostic tests to record these events.

Each study was reviewed at least three times to insure accurate data transcription. Each author reviewed the articles independently. It was occasionally necessary to calculate event counts from occurrence rates. Some studies had more than one general anesthetic group, *e.g.*, halothane and enflurane. These were combined into one GA group for analysis. The studies included in the meta-analysis were chosen on the basis of the design of each study and not on the results reported.

Data Analysis

Two types of health outcome were analyzed: 1) dichotomous variables (death and DVT) and 2) a continuous variable (EBL). For dichotomous variables, the outcome measure was the probability of outcome θ, ranging from 0.0 to 1.0. This is denoted θ_{RA} and θ_{GA}. For the continuous variable, the outcome measure was mean blood loss, μ_{RA} and μ_{GA}.

Two measures of the effect of intervention (effect measures) were estimated: the actual difference in probabilities, $\varepsilon_d = \theta_{GA} - \theta_{RA}$ and the odds ratio, $\varepsilon_{OR} = (\theta_{GA}/(1 - \theta_{GA}))/(\theta_{RA}/(1 - \theta_{RA}))$. The mean of ε_d is the difference in event probabilities between the RA and GA treatment interventions. A positive value indicates a higher mortality or morbidity for GA. All probabilities are reported as percentages, *i.e.*, 100 times θ_{GA}, 100 times θ_{RA}, and 100 times ε_d. The mean of ε_{OR} is the multiplicative ratio of probabilities for GA and RA. An odds ratio of 1 indicates equal outcome probabilities for GA and RA; an odds ratio greater than 1 indicates a greater mortality or morbidity for GA. For EBL, the effect measure was the actual difference in means, $\varepsilon_d = \mu_{GA} - \mu_{RA}$. The mean of ε_d is thus the mean difference in blood loss between GA and RA. A positive mean difference in blood loss indicates that blood loss with GA is greater.

A nonequal, random effects model for combining study evidence was performed; this is also known as hierarchical Bayesian meta-analysis. A random effects model assumes that each of the individual studies is a random sample from the spectrum of true values; this model allows for unknown, random variations in the true effects of anesthesia. It is expressed as a summary mean and 95% confidence interval. When the 95% confidence interval of the mean for ε_d includes 0, then the difference in probabilities for GA and RA is not statistically different from 0. When the 95% confidence interval of the mean odds ratio for ε_{OR} bounds 1, then odds for GA and RA are not statistically different from unity.

See the Statistical Methods Appendix for further details.

Results

Thirteen reports of RCTs comparing RA and GA for traumatic hip surgery were found (table 1).[9–21] These studies spanned the years from 1978 to 1987. These studies were performed in six different industrialized countries, but none came from the United States. Eleven studies used random allocation to the two study groups, but gave no details of the randomization process. One study specified the use of a random number table.[17] One study used an acceptable pseudo-random allocation method (assignment by birth date).[20] While the vast majority of patients had surgical repair by open reduction with internal fixation, in three studies some patients received hemiarthroplasty with prosthesis.[17–19] The large predominance of female patients and the advanced age of all patients is typical of clinical reports of traumatic hip injuries.[3] By ASA physical status, most patients had significant medical illnesses.

A variety of general anesthetics were administered including intravenous and inhalation techniques; isoflurane was not used. There was also no use of induced hypoten-

TABLE 1. Study Demographics I

Author Country	Year	No. of Patients (GA/RA)	Males/Females (%/%) ALL or GA RA	Mean Age (yr) ALL or GA/RA	ASA (%) ALL or GA RA I II III IV	GA Type	RA Type
McLaren[16] Scotland	1978	55 (29/26)	NL	76/76	0 0 21 27 76 69 3 4	Alt Fent	SAB ↑Bar Dib
Davis[9] New Zealand	1980	74 (37/37)	14/86	82	NL	Dia Fent	SAB ↑Bar Tet
McKenzie[14] Scotland	1980	100 (51/49)	NL	75/77	NL	Alt Hal	SAB ↑Bar Dib
White[19] South Africa	1980	40 (20/20)	8/92	80/78	5 10 50 30 35 50 10 10	Thio Fent Hal	SAB ↑Bar Dib
Davis[10] New Zealand	1981	132 (68/64)	13/87 17/83	78/81	NL	Dia Fent	SAB ↑Bar Tet
Wickstrom[20] Sweden	1982	169 (137/32)	0/100	83/80*	NL	Ket/ Fent/ Hal/ Enf	EPI Mep
McKenzie[15] Scotland	1984	148 (75/73)	NL	74/75	NL	Alt Hal	SAB ↑Bar Dib
Bigler[12] NL	1985	40 (20/20)	25/75 10/90	78/80	10 10 70 75 20 15 0 0	Dia Fent	SAB ↑Bar Bup
McKenzie[13] Scotland	1985	40 (20/20)	30/70 20/80	72/74	NL	Alt Hal	SAB ↑Bar Dib
Racle[17] France	1986	70 (35/35)	0/100	82/82	NL	Thio Fent Enf	SAB ±Bar Bup
Valentin[18] Denmark	1986	578 (297/281)	20/80 21/79	79/79	27 28 45 47 25 17 3 8*	Thio Enf Fent	SAB ±Bar Bup
Berggren[21] Sweden	1987	57 (29/28)	24/76 14/86	77/78	NL	Thio Hal	EPI Pril
Davis[11] New Zealand	1987	538 (279/259)	22/78	NL	I/II 43 III 48 IV 9	Thio Fent	SAB ↑Bar/ ±Bar Tet/ Dib/ Bup

NL = not listed; GA = general anesthesia; RA = regional anesthesia; ASA = ASA physical status score; Dia = diazepam; Fent = fentanyl; Alt = Althesin; Thio = thiopental; Ket = ketamine; Hal = halothane; Enf = enflurane; SAB = subarachnoid block; EPI = epidural block; Dib = dibucaine (cinchocaine); Tet = tetracaine (amethocaine); Bup = bupivacaine; Mep = mepivacaine; Pril = prilocaine; ↑Bar = hyperbaric; ±Bar = isobaric. Nitrous oxide was used in all studies during GA. In some studies alternative GA and RA conditions were allowed. These are indicated by a slash (/).

* Significant difference. $P < .05$.

sion. Eleven studies used SAB, one study used single injection EPI,[20] and one study used continuous EPI[21]; in the latter, the catheter was removed at the end of surgery. The SAB drugs used were hyperbaric or isobaric tetracaine, dibucaine, or bupivacaine. Spinal opioids were not administered in any study. No description of postoperative pain management was given in any report.

In most studies, patients had their operative repair within 3 or 4 days following injury; in some studies surgery was delayed for up to a week (table 2). Average surgery duration was 30–120 min. DVT prophylaxis was used infrequently; in only one study was subcutaneous heparin administered.[17] In only three studies were diagnostic tests for DVT systematically administered to all patients. While mention of EBL is made in nine reports, mean values were reported in only four. The duration of follow-up was at least 4 weeks in all studies.

Virtually all experiments have biases. In combining evidence, the analyst has three choices concerning these biases: 1) overlook biases as being too small to alter the results and accept a report at face value, 2) reject a report for having large biases, and 3) adjust the results of a report for potential biases. Bayesian statistics offers formal methods for adjusting probability distributions for such biases.[7]

Any factor that can cause the observed results to not accurately reflect the effect of intervention is a bias to internal validity. There were four instances in three studies in which a statistically significant difference was noted between patients receiving GA and RA. There was a slightly higher mean age (3 yr) for GA patients,[20] a few more ASA physical status 4 patients (~5%) for RA,[18] a slightly longer delay of surgery for RA patients,[20] and a mean surgery duration 15 min longer for RA patients.[13] The use of the study of Wickstrom *et al.*[20] with its significant difference in time to operation might present some concern as it is common surgical opinion to operate within 1 or 2 days. However, there is considerable evidence that within the first few days following injury there is no relationship between delay in surgery and outcome.[3,22–24] Clearly, the small differences in age, surgery duration, and ASA physical status are trivial. Thus, we accepted the studies at face value, considering these small differences to be immaterial to the mortality and morbidity results. Thus no adjustments were made for biases affecting internal validity.

There are also biases that affect the comparability of studies. This is the problem of "mixing apples and oranges." Formal adjustments also can be made for these biases. These 13 studies were similar in most regards (*e.g.*, patients and surgery). More than 2,000 patients were included. Because of multiple general anesthetic groups in some studies reporting mortality, there were unequal numbers of GA and RA patients. In spite of the similarities, the mortality in the control groups of the 13 studies was not homogeneous (table 3). Combined perioperative mortality rate for patients receiving GA was 11.7% with a 95% confidence interval of 7.9–16.3%. Since a random effects model was used, pooling of studies is appropriate even if inhomogeneity is present.[25] These studies were combined at face value without adjustments. Nevertheless, because of concerns that SAB and EPI are dissimilar, meta-analysis was performed for SAB studies, EPI studies, and all studies.

Table 4 lists the 95% confidence intervals for variables comparing RA *versus* GA; mortality was analyzed thrice,

TABLE 2. Study Demographics II

Author	Time to Repair (days) ALL or GA/RA	Mean Surgery Duration (min) ALL or GA/RA	DVT Dx Test	DVT Px	EBL Listed	Follow-up Interval
McLaren[16]	NL	NL	None	NL	No	4 weeks
Davis[9]	≤3	NL	Scan	None	No	1 month
McKenzie[14]	1.9/2.1	77/78	None	NL	Yes	4 weeks
White[19]	≤7	58/58	None	NL	No	4 weeks
Davis[10]	1.0/1.1	104/104†	Scan	None	Yes	4 weeks
Wickstrom[20]	2.5/3.5*	NL	None	Dextran	No	4 yr
McKenzie[15]	1.9/2.0	77/82	None	NL	Yes	1 yr
Bigler[12]	≤2	59/67	None	None	No	90 days
McKenzie[13]	1.8/2.0	79/94*	X-ray	None	Yes	1 yr
Racle[17]	≤1	116/125	None	Heparin	No	3 months
Valentin[18]	≤3	NL GA = RA	None	TED	Yes‡	2 yr
Berggren[21]	≤3	31/35	None	Dextran	No	1 yr
Davis[11]	NL	NL	None	NL	No	3 ≤ and ≤ 30 months

NL = not listed; GA = general anesthesia; RA = regional anesthesia; DVT Px = deep venous thrombosis prophylaxis; DVT Dx test = diagnostic text for deep venous thrombosis; TED = antithromboembolism stockings; Dextran = Dextran infusion; Scan = ^{125}I-fibrinogen leg scan; x-ray = lower extremity venography.

* Significant difference, $P < .05$.
† Anesthesia time.
‡ Only histogram plotted.

TABLE 3. Survival after Surgery

Reference	GA Mortality (95% Confidence Interval)	RA Mortality (95% Confidence Interval)	Mortality Difference (95% Confidence Interval)
McLaren[16]	31.7% (9/29) (16.6 to 49.0%)	5.6% (1/26) (0.4 to 16.6%)	26.1%* (7.0 to 43.8%)
Davis[9]	19.7% (7/37) (9.5 to 33.6%)	9.2% (3/37) (2.3 to 20.1%)	10.5 (−5.1 to 25.8%)
McKenzie[15]	17.8% (13/75) (10.1 to 27.1%)	11.5% (8/73) (5.3 to 19.6%)	6.3% (−4.9 to 17.4%)
McKenzie[14]	16.3% (8/51) (7.7 to 27.4%)	11.0% (5/49) (4.0 to 20.9%)	5.3% (−7.9 to 18.4%)
Racle[17] ‡	14.3% (5/35) (5.7 to 28.5%)	11.8% (4/35) (4.0 to 24.9%)	2.8% (−13.0 to 18.4%)
Davis[10]	13.8% (9/68) (6.8 to 22.8%)	5.4% (3/64) (1.3 to 12.0%)	8.4% (−1.4 to 18.1%)
McKenzie[13]	11.9% (2/20) (2.1 to 28.3%)	2.4% (0/20) (0.0 to 11.7%)	9.5% (−5.6 to 24.3%)
Valentin[18]	8.2% (24/297) (5.5 to 11.8%)	6.2% (17/281) (3.7 to 9.3%)	2.0% (−2.2 to 6.2%)
Bigler[12]	7.1% (1/20) (0.5 to 21.1%)	7.1% (1/20) (0.5 to 21.1%)	0.0% (−15.2 to 15.2%)
Wickstrom[20]	6.9% (9/137) (3.3 to 11.7%)	7.6% (2/32) (1.3 to 18.5%)	−0.7% (−10.5 to 9.1%)
Davis[11]	5.9% (16/279) (3.5 to 9.9%)	6.7% (17/259) (4.0 to 10.1%)	−0.8% (−4.9 to 3.3%)
White[19]	2.4% (0/20) (0.0 to 11.7%)	2.4% (0/20) (0.0 to 11.7%)	0.0% (−9.0 to 9.0%)
Berggren[21]	1.7% (0/29) (0.0 to 8.2%)	5.2% (1/28) (0.4 to 15.5%)	−3.5% (−12.6 to 5.6%)
Summary	11.7%† (103/1097) (7.9 to 16.3%)	7.0% (62/944) (5.4 to 9.2%)	

GA = general anesthesia; RA = regional anesthesia.
* Probability difference ≠ 0, $P < .05$.
† Reject homogeneity, $P < .05$.
‡ 3-month survival reported.

for EPI only, for SAB only, and for all RA combined. The meta-analysis results using probability differences shows the mortality rate for patients receiving GA was 2.7 percentage points higher than that of patients receiving RA when combining all studies. In the meta-analysis restricted to studies using SAB, the probability difference was even slightly greater at 3.8%. However, the lower bound of the 95% confidence interval for the treatment effect was less than 0 for both meta-analyses, indicating that there is insufficient precision of the estimate to claim statistical significance for a lower mortality with RA. Of the 13 studies, only one individually had a significant effect measure (table 3). Meta-analysis using the odds ratios reveals somewhat different results for mortality (table 4). For all studies combined and for SAB studies only, a patient receiving GA was about 1.5 times more likely to die than a patient receiving RA. The 95% confidence intervals for mortality had lower bounds (1.02 and 1.05) only marginally above 1.

The probability difference for DVT was 31% greater for patients receiving GA; the lower bound of the 95% confidence interval was well away from 0 (table 4). The

TABLE 4. Treatment Efficacy

Outcome Event	No. of Studies	No. of Patients (GA/RA)	Difference (95% Confidence Interval)	Odds Ratio (95% Confidence Interval)
Mortality				
Epidural studies[20,21]	2	226 (166/60)	−2.2% (−10.6 to 6.3%)	0.75 (0.10 to 5.50)
Spinal studies[9–19]	11	1815 (931/884)	3.8% (−0.3 to 7.8%)	1.53 (1.05 to 2.22)*
All studies	13	2041 (1097/944)	2.7% (−0.7 to 6.1%)	1.49 (1.02 to 2.16)*
Morbidity				
Deep venous thrombosis[9,13,20]	3	246 (125/121)	31.3% (16.7 to 44.6%)*	3.99 (2.04 to 7.77)*

* Probability difference ≠ 0 or odds ratio ≠ 1, $P < .05$.

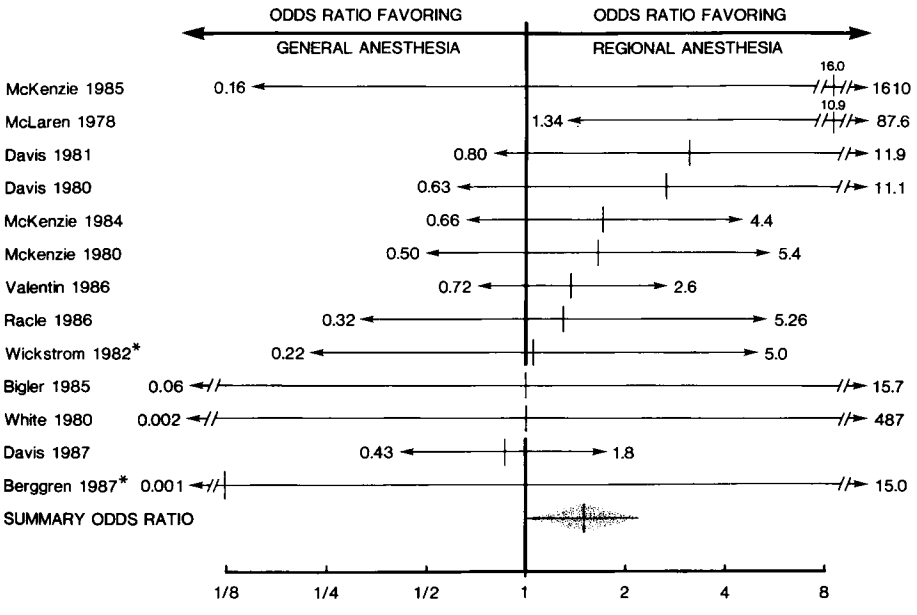

FIG. 1. Interval plot of the 95% confidence interval of the odds ratios for mortality (general anesthesia *vs.* regional anesthesia). For each study the vertical line is the geometric mean odds ratio and the left and right arrows are the lower and upper bound of the 95% confidence interval. The numerical values of the lower and upper bound are also displayed. The heavy vertical line is the odds ratio = 1, the odds ratio of no treatment effect. The diamond is the summary odds ratio and 95% confidence interval for the meta-analysis of the 13 studies. *A study using epidural anesthesia.

summary odds ratio for DVT was clear cut. A patient receiving GA was almost four times more likely to develop a DVT than a patient receiving RA.

Interval plots are a visual method of data inspection for odds ratios. In an interval plot the odds ratio and its 95% confidence interval are displayed for each study (fig. 1 and 2). By plotting the odds ratios on a log scale, the precision of each study is better indicated. Figure 1 shows the inconsistent individual odds ratios and the barely significant summary odds ratio for mortality. Figure 2 shows the consistent benefit for RA lowering the incidence of DVT.

Only five studies reported intraoperative EBL; of these, one displayed a histogram and could not be used in meta-analysis (table 2). In addition, four studies made nonquantitative comments that EBL was minimal or equivalent[19,21] or listed similar blood transfusions to RA and GA patients.[12,17] In two studies the blood loss of the RA group was greater than the blood loss of the GA group (105 and 16 ml greater, respectively),[13,15] and in two studies the blood loss of the GA group was greater than the blood loss of the RA group (164 and 3 ml greater, respectively).[10,14] The homogeneity statistic for these studies was significant. The meta-analysis summary mean difference showed an intraoperative EBL ~ 18 ml higher for those receiving GA. However, the 95% confidence interval was wide, from −99 ml to 116 ml.

Discussion

Among physicians there seems to be a prevalent and implicit impression that RA is safer than GA. This impression is reflected in the frequent admonition from internists and other consultants to choose regional anesthesia for the surgical patient with serious medical problems. Anesthesiologists are divided in their preference for RA and GA. Both advocates of RA and advocates of GA can cite personal experience and published studies supporting their positions. One method to resolve such questions is through literature review. Yet literature re-

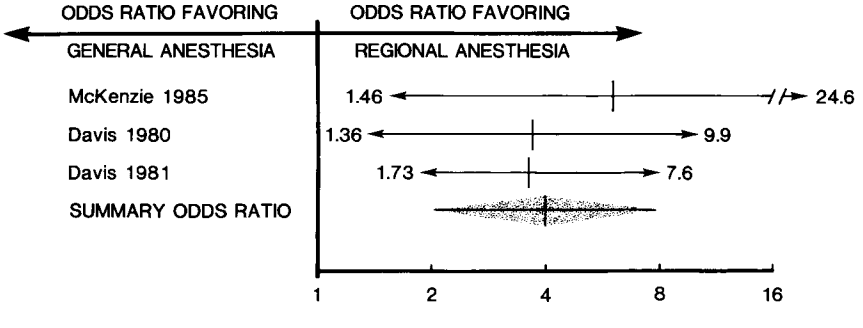

FIG. 2. Interval plot of the 95% confidence interval of the odds ratios for the incidence of deep venous thrombosis (general anesthesia *vs.* regional anesthesia). For each study the vertical line is the geometric mean odds ratio and the left and right arrows are the lower and upper bound of the 95% confidence interval. The numerical values of the lower and upper bound are also displayed. The heavy vertical line is the odds ratio = 1, the odds ratio of no treatment effect. The diamond is the summary odds ratio and 95% confidence interval for the meta-analysis.

views can also lead to opposite conclusions about the superiority of RA *versus* GA. Scott and Kehlet concluded that mortality and morbidity is probably decreased by using RA for surgery below the umbilicus.[26] An earlier paper by Kehlet in 1984 was more skeptical about the evidence.[27] Journals of anesthesia have published match ing pro/con editorials contrasting RA and GA.[§,‖]

The standard approach for resolving such a controversy would be to conduct a large, multicenter, double-blinded, randomized, controlled clinical trial that would simultaneously control for covariates such as age, gender, ASA physical status, type of surgical procedure, and the presence of other pre-existing conditions. Such a study would preferably include various treatment groups, *i.e.*, several types of GA and both SAB and EPI. Only once has a large study comparing types of anesthesia been attempted. The recently reported Multicenter Study of General Anesthesia included more than 17,000 patients having many different surgical procedures, but was unable to distinguish differences in death, myocardial infarction, and stroke between halothane, enflurane, isoflurane, and fentanyl anesthesia.[28,29] For lack of comparable resources of funding, researchers, and enthusiasm, such a comparative study of RA *versus* GA is unlikely to be attempted in the near future.

Data analysis can be divided into three categories: primary analysis, the original analysis of an experiment; secondary analysis, a reanalysis of research data with better statistical tools or with different questions; and meta-analysis, a systematic approach for summarizing quantitatively through statistical methods the results of previously published studies. Meta-analysis is a set of techniques for quantitatively pooling published results to give tentative answers about controversies for which the evidence is incomplete and inconsistent. By pooling evidence, there is an increase in statistical power for detecting differences. Both frequentist and Bayesian meta-analysis extensively use confidence limits and confidence intervals respectively to discuss estimation of treatment effect. Estimating the magnitude of treatment effect should be foremost as this can translate more easily into policy choices. Although used only occasionally by anesthesia researchers,[30] meta-analyses are widely used in other specialties of medicine as well as in education and psychology.

There is considerable criticism that much of the meta-analysis being published is performed improperly and/or is basically illegitimate.[31–33] However, meta-analysis has reached a stage of general acceptance in the statistical and clinical literature. As in most scientific endeavors, real problems remain to be clarified.[34] These concerns focus on the process of literature searching, study abstraction, and study presentation; the criteria for study combinability, the effect measure chosen, and the statistical model for combining estimates; the control and adjustment for potential bias; and the statistical methods for estimation, subgroup analyses, the use of confidence intervals, publication bias, and the presentation of results.

Our meta-analysis followed published guidelines.[6] This project was started with a well defined goal: a comparison of mortality between RA and GA. The literature search was extensive and did not rely solely on computer queries against journal databases. The patient characteristics of each trial have been displayed (tables 1 and 2) to allow independent reader judgment about the generalizability of these studies; these studies are reasonably homogenous. Only RCTs were acceptable for combining. Two effect measures were used to check whether results vary with different assumptions; unfortunately, the results are close, but not identical.

Attention to the quality of the studies was of great concern. The two main aspects of quality are randomization and blinding; blinding is not feasible in a comparison of RA and GA, and only RCTs were used for meta-analysis. There are two different approaches for handling potential biases. In the first, methods for assessing the quality of RCTs have been constructed; numerous study characteristics (*e.g.*, randomization method and blinding) are numerically graded and a summary quality score tallied.[35] It has been recommended that the quality score be used to weight the results of each study when combining them. An actual benefit of using quality scores has not been demonstrated. To the contrary, the use of quality scores has been demonstrated to not change the summary effect sizes or the variance of the effect sizes.[36] No quality scores were calculated in this meta-analysis. A second approach for potential biases is distinct for Bayesian meta-analysis. Formal adjustments for any bias may be made during estimation. Although there were several differences in study design, these differences were considered inconsequential and the studies were accepted at face value.

Although there was inconsistent evidence of study inhomogeneity, equal effects were not assumed and a random effects model was used to allow for variation in the true effects of anesthesia. Confidence intervals have been listed for all results to allow easy inspection of the magnitude of the summary effect sizes. Visual graphs of the odds ratios illustrates the inconsistency and consistency of two of the outcome measures (mortality and DVT incidence). Bayesian meta-analysis was chosen for estimation of parameters because: 1) it allows formal incorporation

§ Yeager MP: Pro: Regional anesthesia is preferable to general anesthesia for the patient with heart disease. Journal of Cardiothoracic Anesthesia 3:793–796, 1989.

‖ Beattie C: Con: Regional anesthesia is not preferable to general anesthesia for the patient with heart disease. Journal of Cardiothoracic Anesthesia 3:797–800, 1989.

of bias adjustments;[7] 2) Bayesian confidence intervals have the interpretation mistakenly ascribed by the statistically inexperienced to frequentist confidence limits[37]; and 3) the results of additional studies may be added to Bayesian parameters without the logical difficulties of the frequentist multiple-look phenomenon.[34]

Meta-analysis does contain areas where bias and error may intrude, but the structured literature search, formal combining of results, and statistical rigor allows for a systematic analysis. A properly conducted meta-analysis is likely to include all available studies and is likely to quantitatively amass and report the data used for analysis. There is less room for interpretation and bias in the results obtained; the results of a meta-analysis should be reproducible by others.

Thirteen studies comparing GA to RA (SAB and EPI) have failed to show a consistent treatment effect for 1-month perioperative mortality. The probability difference among RCTs ranged from 27% to −4%. McLaren *et al.* reported a significant mortality difference between GA (31%) and RA (4%).[16] By contrast, Berggren *et al.* reported a nonsignificant mortality difference favoring GA, GA (0 %) and RA (4%).[21] Two moderate-size studies used in this meta-analysis, each including more than 500 patients, found no significant difference in mortality between RA and GA.[11,18] The summary odds ratio for all studies and for SAB studies does favors RA, having a lower bound of the 95% confidence interval assuring statistical significance (table 4). Even then, these significant lower bounds are only slightly greater than 1, 1 being the odds ratio of no treatment effect. It is important to notice that only 1 of the 13 reports has an individual 95% confidence interval that does not include 1 (fig. 1). If there is a real treatment effect of RA on mortality, it is not clearly discernible from the past level of research effort. There is a much lower rate of DVT for RA. By contrast to the interval plot of mortality, all three reports on DVT have individual 95% confidence intervals that do not cross 1 (fig. 2). There was clearly no superiority of RA for lowering operative blood loss.

This meta-analysis does not address issues such as patient fear and comfort, nor does it consider the preferences, work environments, patient volume, and skills of individual anesthesiologists. While these areas are difficult to research and document, they are none the less important as factors in selection of an anesthetic technique. When faced with indistinguishable outcomes from alternative anesthetics, there is no reason that the choice for RA or GA should not be by the preference of the patient and the predilection of the anesthesiologist.

This meta-analysis has provided provisional answers from the existing literature pending more definitive and conclusive clinical trials. Continued investigations into the benefits and risks of these techniques should be considered. This will require the performance of additional RCTs comparing RA and GA. As no RCT has been performed in the United States (table 1), it is to be hoped that such additional research will also be performed in North America. Possible themes for additional RCTs might include the following: 1) Currently available studies have insufficient consistency in reporting covariates. Is it possible that there is a subset of patients (distinguished perhaps by gender, age, or ASA physical status) for whom GA or RA is clearly superior? 2) Further study concerning the implications of lower DVT rates following RA might be merited. Why doesn't this decreased morbidity translate into better outcome? None of studies observing systematically for DVT have used prophylaxis with subcutaneous heparin.[38] Would the favorable effect of RA on DVT disappear if this standard measure was used? 3) To create a larger pool of studies for meta-analysis, SAB and EPI results were combined. Could the more rapid onset of sympathetic block with spinal anesthesia or the higher circulating blood levels of local anesthetics with epidural anesthesia make one or the other superior? 4) Similarly, various types of GA were combined for purposes of analysis. Perhaps the more frequent hypertension/tachycardia with narcotic anesthetic techniques or the more frequent hypotension with inhalation anesthetics have consequences[29] making one particular general anesthetic inferior or superior to RA. 5) The current popularity of SAB and EPI is due in part from their use to administer spinal opioids for postoperative pain relief. It is hoped that better pain management will lower stress and favorably influence morbidity and mortality. None of these studies used spinal opioids. It will be essential to distinguish the benefits of RA for surgery from the benefits of spinal opioids for postoperative pain relief. 6) The significant odds ratio raises the possibility that more evidence would provide greater statistical power to cleanly show benefit of RA on mortality. Assuming that the mortality rate for GA is 11.7%, that the actual difference in probabilities for mortality is 2.7% (all studies) and 3.8% (SAB studies only), and that the observed variance is a reasonable estimation, it can be calculated that 11,200 and 4,200 patients, respectively, would have to be studied in a RCT to have 80% power to show a difference. This suggests that it will be difficult for any single future study to clarify this ambiguity since only 2,000 patients have been studied to date in all reported RCTs. However, future studies could be combined with existing studies in a new meta-analysis. If the true benefit of RA is a 2.7% reduction in mortality rate, it can be estimated that a study with 1,000 patients (500 GA *vs.* 500 RA) would provide statistical confirmation when combined with the 13 reported studies. Until such confirmation is available, we believe a conservative statement of unproven efficacy should be attached to claims of RA superiority.

Appendix: Statistical Methods

For dichotomous variables, the outcome measure was the probability of outcome. This is denoted θ_{RA} and θ_{GA}; θ_{RA} and θ_{GA} are posterior probability distributions that combine any prior knowledge and the results of each study by methods of maximum likelihood. No prior knowledge was assumed; a noninformative prior distribution, a beta distribution, was used. Then θ_{RA} and θ_{GA} are conditional beta distributions of the form $\pi(\theta|s,f) = \theta^{(s+0.5)}(1 - \theta)^{(f+0.5)}/\text{Beta}(s + 0.5, f + 0.5)$, where s = the number of patients in whom the outcome occurred, $s + f$ = the number of patients in the sample, and the beta function $\text{Beta}(s + 0.5, f + 0.5) = \Gamma(s + 0.5)\Gamma(f + 0.5)/\Gamma(s + f + 1)$ with Γ being the gamma function $\Gamma(n) = (n - 1)!$. The mean value of $\pi(\theta|s,f)$ has the exact solution $(s + 0.5)/(s + f + 1)$.

The Bayesian estimation of parameters is considerably different from frequentist estimation. In frequentist estimation of rates, θ_{RA} and θ_{GA} would be derived directly from the observed proportions of the sample data ($\theta = x/n$, x representing the number of patients in whom the outcome occurred and n representing the number of patients in the sample). The mean value of $\pi(\theta|s,f)$ is the Bayesian rate analogous to the frequentist observed proportions and can be calculated as $(x + 0.5)/(n + 1)$.

Two measures of the effect of intervention (effect measures) were estimated: the actual difference in probabilities, $\varepsilon_d = \theta_{GA} - \theta_{RA}$, and the odds ratio, $\varepsilon_{OR} = (\theta_{GA}/(1 - \theta_{GA}))/(\theta_{RA}/(1 - \theta_{RA}))$. These parameters ($\varepsilon_d$ and ε_{OR}) are not single numbers, but are joint posterior probability distributions that combine any prior information and the results of each study by maximum likelihood estimation. Both ε_d and ε_{OR} were estimated because of uncertainty about the choice of the appropriate model.[39] The mean of ε_d is the difference in event probabilities between GA and RA. The mean of ε_{OR} is the multiplicative ratio of probabilities for GA and RA. For the continuous variable EBL, the outcome measure was estimated analogously to ε_d for mortality and DVT incidence. The estimation of all joint posterior probability distributions used noninformative prior distributions.

The hierarchical Bayesian meta-analysis model assumes that each of the individual studies is a random sample from the spectrum of true values; this model allows for unknown, random variations in the true effects of anesthesia. It combines the evidence of the effect measures (ε_d or ε_{OR}) from the individual studies with a noninformative prior distribution to estimate the parameters of the underlying distribution (mean and variance).

A collection of studies is considered homogenous if all are attempting to estimate the same true values; this assumes an equal effects model. Homogeneity among studies was checked by a frequentist χ^2 statistic. However, this statistical test is known to have low statistical power for rejecting the null hypothesis, and the random effects model was performed regardless of the homogeneity test. The alpha significance level was set at 0.05. The odds ratios were graphed on a log scale.[40] The software used was FAST*PRO (PC version 1.3 (Academic Press, Boston, MA)). Bayesian statistical methods and their software implementation in FAST*PRO are quite complicated in theory and computation; full details are available in Bayesian textbooks[37] and by the authors of FAST*PRO.[7]

References

1. Nygaard KK: Routine spinal anesthesia in provincial hospital: With comparative study of postoperative complications following spinal and general ether anesthesia. Acta Chir Scand 78:379–446, 1936
2. Zindrick MR, Daley RJ, Hollyfield RL, Jobski R, Kinzler GM, Schwartz CM, Wood WS: Femoral neck fractures in the geriatric population: The influence of perioperative health upon the selection of surgical treatment. J Am Geriatr Soc 33:104–108, 1985
3. Kenzora JE, McCarthy RE, Lowell JD, Sledge CB: Hip fracture mortality: Relation to age, treatment, preoperative illness, time of surgery, and complications. Clin Orthop 186:45–56, 1984
4. L'Abbe KA, Detsky AS, O'Rourke K: Meta-analysis in clinical research. Ann Intern Med 107:224–233, 1987
5. Thacker SB: Meta-analysis: A quantitative approach to research integration. JAMA 259:1685–1689, 1988
6. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC: Meta-analyses of randomized controlled trials. N Engl J Med 316:450–455, 1987
7. Eddy DM, Hasselblad V, Shachter R: Meta-analysis by the confidence profile method, The Statistical Synthesis of Evidence. Boston, Academic Press, 1992
8. Eddy DM, Hasselblad V, Shachter R: A Bayesian method for synthesizing evidence: The confidence profile method. Int J Technol Assess Health Care 6:31–55, 1990
9. Davis FM, Quince M, Laurenson VG: Deep vein thrombosis and anaesthetic technique in emergency hip surgery. Br Med J 281: 1528–1529, 1980
10. Davis FM, Laurenson VG: Spinal anaesthesia or general anaesthesia for emergency hip surgery in elderly patients. Anaesth Intensive Care 9:352–358, 1981
11. Davis FM, Woolner DF, Frampton C, Wilkinson A, Grant A, Harrison RT, Roberts MTS, Thakaka R: Prospective, multi-center trial of mortality following general or spinal anaesthesia for hip fracture surgery in the elderly. Br J Anaesth 59:1080–1088, 1987
12. Bigler 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 40:672–676, 1985
13. McKenzie PJ, Wishart HY, Gray I, Smith G: Effects of anaesthetic technique on deep vein thrombosis: A comparison of subarachnoid and general anaesthesia. Br J Anaesth 57:853–857, 1985
14. McKenzie PJ, Wishart HY, Dewar KMS, Gray I, Smith G: Comparison of the effects of spinal anaesthesia and general anaesthesia on postoperative oxygenation and perioperative mortality. Br J Anaesth 52:49–54, 1980
15. McKenzie PJ, Wishart HY, Smith G: Long-term outcome after repair of fractured neck of femur: Comparison of subarachnoid and general anaesthesia. Br J Anaesth 56:581–585, 1984
16. McLaren AD, Stockwell MC, Reid VT: Anaesthetic techniques

for surgical correction of fractured neck of femur: A comparative study of spinal and general anaesthesia in the elderly. Anaesthesia 33:10–14, 1978
17. Racle JP, Benkhadra A, Poy JY, Gleizal B, Gaudray A: Etude comparative de l'anesthesie generale et de la rachianesthesie chez la femme agee dans la chirurgie de la hanche. Ann Fr Anesth Reanim 5:24–30, 1986
18. Valentin N, Lomholt B, Jensen JS, Hejgaard N, Kreiner S: Spinal or general anaesthesia for surgery of the fractured hip? A prospective study of mortality in 578 patients. Br J Anaesth 58: 284–291, 1986
19. White IWC, Chappell WA: Anaesthesia for surgical correction of fractured femoral neck: A comparison of three techniques. Anaesthesia 35:1107–1110, 1980
20. Wickstrom I, Holmberg I, Stefansson T: Survival of female geriatric patients after hip fracture surgery: A comparison of 5 anaesthetic methods. Acta Anaesthesiol Scand 26:607–614, 1982
21. 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 66:497–504, 1987
22. Davis TRC, Sher JL, Porte BB, Checketts RG: The timing of surgery for intertrochanteric femoral fractures. Injury 19:244–246, 1988
23. Eiskjaer S, Ostgard SE: Risk factors influencing mortality after bipolar hemiarthroplasty in the treatment of fracture of the femoral neck. Clin Orthop 270:295–300, 1991
24. White BL, Fisher WD, Laurin CA: Rate of mortality for elderly patients after fracture of the hip in the 1980's. J Bone Joint Surg [Am] 69A:1335–1340, 1987
25. Berlin JA, Laird NM, Sacks HS, Chalmers TC: A comparison of statistical methods for combining event rates from clinical trials. Stat Med 8:141–151, 1989
26. Scott NB, Kehlet H: Regional anaesthesia and surgical morbidity. Br J Surgery 75:299–304, 1988
27. Kehlet H: Does regional anaesthesia reduce postoperative morbidity? Intensive Care Med 10:165–167, 1984
28. Forrest JB, Rehder K, Goldsmith CH, Cahalan MK, Levy WJ, Strunin L, Bota W, Boucek CD, Cucchiara RF, Dhamee S, Domino KB, Dudman AJ, Hamilton WK, Kampine J, Kotrly KJ, Maltby JR, Mazloomdoost M, MacKenzie RA, Melnick BM, Motoyama E, Muir JJ, Munshi C: Multicenter study of general anesthesia: I. Design and patient demography. ANESTHESIOLOGY 72:252–261, 1990
29. Forrest JB, Cahalan MK, Rehder K, Goldsmith CH, Levy WJ, Strunin L, Bota W, Boucek CD, Cucchiara RF, Dhamee S, Domino KB, Dudman AJ, Hamilton WK, Kampine J, Kotrly KJ, Maltby JR, Mazloomdoost M, MacKenzie RA, Melnick BM, Motoyama E, Muir JJ, Munshi C: Multicenter study of general anesthesia: II. Results. ANESTHESIOLOGY 72:262–268, 1990
30. Pace NL: Prevention of succinylcholine myalgias: A meta-analysis. Anesth Analg 70:477–483, 1990
31. Wachter KW: Disturbed by meta-analysis? Science 241:1407–1408, 1988
32. Goodman SN: Have you ever meta-analysis you didn't like? Ann Intern Med 114:244–246, 1991
33. Thompson SG, Pocock SJ: Can meta-analyses be trusted? Lancet 2:1127–1130, 1991
34. Chalmers TC: Problems induced by meta-analysis. Stat Med 10: 971–980, 1991
35. Chalmers TC, Smith H Jr., Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A: A method for assessing the quality of a randomized control trial. Controlled Clin Trials 2:31–49, 1981
36. Emerson JD, Burdick E, Hoaglin DC, Mosteller F, Chalmers TC: An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. Controlled Clin Trials 11:339–352, 1990
37. Berger JO: Statistical Decision Theory and Bayesian Analysis. 2nd edition. New York, Springer-Verlag, 1985
38. Collins R, Scrimgeour A, Yusuf S, Peto R: Reduction in fatal pulmonary embolism and venous thrombosis by perioperative administration of subcutaneous heparin. N Engl J Med 318: 1162–1173, 1988
39. DerSimonian R, Laird N: Meta-analysis in clinical trials. Controlled Clin Trials 7:177–188, 1986
40. Galbraith RF: A note on graphical presentation of estimated odds ratios from several clinical trials. Stat Med 7:889–894, 1988