Evaluation of Results of Surgical Procedures in the Elderly

BERNARD S. LINN, M.D., MARGARET W. LINN, PH.D., NEIL WALLEN

Authorities are not in full agreement in regard to risk of surgery in the elderly. One hundred eight studies of surgery in the elderly over the past 40 years were reviewed. The purpose was not merely to tabulate results, but to identify differences existing between reports with regard to data reported that could affect results independent of the surgical management itself. Sources of variance that need to be taken into account in comparing mortality rates between studies, such as whether mortality was computed by number of patients or operations, differing lengths of follow-up for recording mortality, proportions of emergency versus elective operations, and types of surgical procedures, were documented. A nine-item confidence in results scale was used to classify studies into high and low confidence groups. Surgical specialties scored substantially higher than general surgical studies. More recent studies received higher scores than earlier studies. Although mortality rates varied widely depending on methods of their calculation, there appeared to be a trend toward increases in elective, but not emergency, mortality rates in general surgery since 1941 that should be examined more closely. One thing that cannot be answered clearly from these studies is the relative risk of surgery with age. Some control of variations between studies and standardization of reporting surgical deaths are required before risk of surgery in the elderly can be assessed more accurately.

From the Departments of Surgery and Psychiatry, VA Medical Center and University of Miami School of Medicine, Miami, Florida

IN TALKING WITH SURGEONS today about the risk of surgery in elderly patients, there appears to be a feeling among them that over the years this risk has decreased. This is epitomized in an article on surgery in 90-year-old patients which states that "less than half a century ago, surgery was considered hazardous for patients over 50 years. Gradually the concept of an upper age limit has faded, and major surgical procedures are now regularly performed on the very elderly. . . . Age in itself is no barrier to surgery."[4] Yet, in the late 1930's, Rankin and Johnson[3] concluded, similarly, that

> Gradually, experience has shown that old age is no longer the contraindication to surgery that it was at the beginning of the century, and more and more data are being accumulated to substantiate this fact . . . some persons 50 years of age seem older than others of 70. However, due credit must be given these people, for without sturdiness of fiber, they would never have reached the age of 70 years. The end result is a steady increase in the operability of patients in this group with a concomitant fall in surgical mortality.

In light of the foregoing, one can question whether times have really changed. What is actually known about risk of surgery in the elderly? Unfortunately, three contrasting conclusions emerge from reviews aimed at better defining changes in risk which may have occurred over the years.[1,2] One is that extreme caution and even reluctance to operate is warranted in elderly patients, since older age alone carries increased risk. A second is that persons who survive into very old age are biologically elite and by that very fact the surgical risk is not so great. A third conclusion has emerged between those extremes that acknowledge some increased risk, particularly with coexisting disease in the elderly, but suggests that with attention to pre- and postoperative care, even the most difficult operations in the elderly can have acceptable outcomes. Which conclusion is true? How much confidence can one place in the conclusions of any single study? Can one compare results among these studies, as is often done?

The purpose of this paper is to review a large number of studies of surgery in the elderly and not only to report results, but to identify differences in the types of data being reported, since such differences in the parameters reviewed could account for as much variation in results as differences in surgical management. Hence, for example, some studies describing themselves as dealing with surgery in the elderly, include patients in their fifties while others begin with patients 90 and above. Only some separate results from emergency as contrasted to elective surgery. Some compute mortality by number of operations per patient; some by patient; some do not say. Lengths of postoperative observation periods range from merely counting deaths in the operating

Reprint requests: Bernard S. Linn, M.D., (141) VA Medical Center, 1201 N. W. 16 Street, Miami, Florida, 33125

Submitted for publication: June 3, 1981.

0003-4932/82/0100/0090 $00.85

room to 90 days after operation. Some identify rates for a particular procedure, while some, which include a mix of general and specialty operative procedures, do not.

After reviewing the literature, the authors will go beyond simply tabulating the results by classifying the studies according to the level of confidence that can be placed in them based on the data used. Each study will be given a rating based on how well it succeeded in defining criteria and reporting results in terms of variations such as those mentioned above. This will then permit a tabulation of the degrees of variation among the reports as well as a characterization of differences between reports with high as contrasted to low confidence ratings. It should be pointed out that the confidence ratings used here are a first step indicating whether, on retrospective review, a basic minimal data base was collected and tabulated, in such a way as to permit comparisons with other studies. There is no attempt made (nor would there be a way) to retrospectively assess the separate but equally important issue related to actual reliability of the data which were collected.

Methods

A Medlars Library search was used to identify published journal articles describing surgery in the elderly. This number was expanded from bibliographies of those found through the computerized search so that articles cover the years from 1930 through the early part of 1980. Only articles which actually described *results* of a series of elderly patients undergoing surgery were included for review. Case reviews and opinion type articles were excluded. A total of 108 articles met these criteria.

Each article was reviewed critically by one of the authors and identifying data and results recorded onto forms for computer analysis. The following items were transcribed whenever available, and if the data were not included in the articles, this was indicated. Background information included year of publication, journal, type of surgery (general or specialty of mixed or single types of procedures), number of operated patients, number of operations, and whether mortality rates were computed by one or both of these methods. The number of days patients were followed during which deaths were included in the mortality rate, lower age limit used, per cent male composition, and criteria for selection of the group (*e.g.* were only patients with a previous record of a certain condition included, etc?) were recorded. It was also noticed whether any type of comparison group was used such as overall operative mortality rate in the institution or mortality rate of a younger age group.

The mortality rate was recorded according to elective,

```
:__: TYPE OF SURGERY REPORTED (0-1)

     0= No
     1= Yes - one type or multiple types defined

:__: AGE (0-3)

     0= Under age 65 included
     1= Lower limit at least age 65
     2= Range given
     3= Mean given

:__: NUMBER OF PATIENTS OPERATED (0-1)

     0= Not given
     1= Provided

:__: NUMBER OF OPERATIONS PERFORMED (0-1)

     0= Not given
     1= Provided

:__: COMPUTATION OF MORTALITY (0-3)

     0= Not given
     1= By operation
     2= By patient
     3= Both

:__: MORTALITY REPORTED (0-1)

     0= Not clear.  Overall rate includes both emergency & elective
     1= Clearly defines mortality for group(s) studied

:__: POST OPERATIVE FOLLOW-UP (0-2)

     0= Not defined
     1= To hospital discharge
     2= Specific equal time for all patients

:__: COMPARISON GROUP USED (0-3)

     0= None used
     1= Within the elderly group studied
     2= With the overall rate of the institution
     3= With younger age group of the same institution

:__: REASONS FOR DEATHS REPORTED (0-1)

     0= Not
     1= Yes
```

FIG. 1. Confidence in results scale.

emergency, and overall rates as computed by operation or patient or both whenever available. If rates were given for subgroups, but not overall group, the overall mortality rate was computed if sufficient information was included in the article to calculate the rate. Likewise, if deaths after emergency and elective procedures were reported separately, an overall mortality rate was calculated for both procedures together, and if possible, all three rates recorded.

Postoperative complication rates for elective and emergency procedures, whether mortalities were included in the complication rate, leading types of postoperative complications, and any factors reported to significantly affect mortality were also recorded.

In addition to abstracting these data from the articles, another person completed a "confidence in results scale" for each study. The nine-item scale is shown in Figure 1. Each study was rated as to whether essential information about the outcome of surgery was included as well as whether a basically good design was used. For

TABLE 1. *Characteristics of the 108 Journal Articles Describing Surgery in the Elderly*

Variables	Number	Per Cent
Year published		
1931-40	6	5
1941-50	12	11
1951-60	20	19
1961-70	31	29
1971-80	39	36
Number conducted in U.S.	86	80
Authors university affiliated	56	52
Lower age included in study		
Under 60	5	5
60-64	31	29
65-69	19	17
70-74	35	32
75-79	5	5
80-84	11	10
85-90	0	0
90 and over	2	2
Types of surgery		
general	78	72
specialty (25 of these heart)	30	28
Sample size (by patients)		
Under 50	10	11
50-99	24	26
100-199	18	19
200-399	22	24
400-599	8	9
600-799	3	1
Over 800	8	8
reported	93	86
not reported	15	14
Sample size (by operations)		
Under 50	3	5
50-99	8	14
100-199	8	14
200-399	18	32
400-599	7	13
600-799	4	7
Over 800	8	13
reported	56	52
not reported	52	48
Method of reporting number of subjects		
number of patients only	50	46
number of operations only	12	11
number of patients and operations	43	40
not reported	3	3
Some Type of Comparison Group	26	24
Method of Computing Mortality		
by operation	22	20
by patient	60	56
both	9	8
method not reported	17	16
Length of Follow-Up		
less than 30 Days (equal for all)	8	7
30 days or more (equal for all)	23	21
to hospital discharge (varied)	31	29
not reported	46	43
Leading causes of death reported	72	67
Reported sample composition by sex	58	54
Reported types of complications	51	48

example, a good design should follow each patient a standard amount of time after operation. Some studies did not provide any information on length of follow-up study. Some used hospital discharge and others reported a standard follow-up, such as 30 days, for everyone. In such situations those not reporting length received a zero, hospital discharge a one, and standard follow-up times a two. As another example, some studies reported mortality by number of operations (which make the mortality rate look low since a patient dies only once), others by patients' operated on (which makes the rate appear higher, unless patients are never reoperated), and others do not specify which was used. Studies should provide both. If they did not record how the rates were computed, they received a zero; by operation a one; by patient a two; and if both were recorded, a three. Other scores were given for complete definition of age, reporting mortality by all appropriate means (elective versus emergency whenever applicable), reasons for deaths, use of age comparison groups, and details about types of surgery wherever applicable. Each study could range from a low of zero to a high of 16 points.

After the rating scale was designed, two persons used the scale independently on the same 20 articles and their item scores were compared to determine reliability by intra class correlations. Items correlated from $r = 0.69$ to $r = 0.96$. To assess validity, a surgeon read 20 of the articles, randomly selected from the 108, and provided a global rating on a 1-5 scale as to the overall quality of study. These 20 global ratings were correlated with the total score from the confidence in results scale for the same 20 articles. The intraclass correlation was $r = 0.84$. Since the scale appeared to have acceptably reliability and validity, it was included in the review.

Data were tabulated first to describe the articles by frequencies of responses and data missing from the articles. Next, mortality rates were stratified by year of study and lower age included for specialty and general surgery separately. Ratings of the confidence in results scale were divided by the mean score into high and low confidence groups and mortality rates reexamined by type of surgery, year, and lower age limit.

Results

Characteristics of Studies

The 108 articles reviewed covered operations in over 50,000 elderly patients. Table 1 describes the studies. Most (80%) were done in the United states, with slightly over half of the authors having university affiliations. Three-fourths described general surgery and of the remaining specialty surgery studies, 83% were of cardiac operations.

In regard to who was considered "elderly," five studies included patients below the age of 60 years and two began with 90 year olds. There was a trend for more recent studies to set a higher lower age limit (age 80 years and above only starting in the early fifties in the articles reviewed); however, the ones including patients below 60 were all done after 1960. The most popular lower age limits were 60 and 70 years. Only about 26% of the articles used some type of comparison group in order to assess differences in surgical risk with age.

Sample size varied from 10% with less than 50 patients to one study of over 7,000 patients. About 14% did not report number of patients in the study, but 79% of those studies reported sample size by operations. In 48% of the studies, number of operations was not reported. Some studies (40%) reported both patients and number of operations.

Although 30 days is frequently thought to be the standard postoperative period, only a fifth of the studies followed all their patients for 30 days or more. Another 29% used discharge from the hospital as end of follow-up, defining mortality as a "hospital death." More surprising is the fact that 43% did not report the amount of time in which mortality was recorded and some, although they used an equal time for all, used as few as eight days following the operations. Less than 70% of the studies reported leading causes of deaths and less than half gave postoperative complications.

Mortality Rates

Table 2 reports overall and elective rates computed by either number of operations or number of patients. Emergency mortality is by patient only. Rates based on number of operations could never be higher, of course, if all studies reported both ways. As seen, rates were generally higher when patients were used as the denominator. The mean overall mortality rate was either 11.7 or 13.6% depending on which way authors chose to compute their rates. A few authors reported both. The numbers not applicable included those who reported by the alternate method (patient or operation) and those who studied only emergency or elective procedures. One study failed to give an overall, emergency, or elective rate, but reported rates for numerous subgroups without providing the number of patients in the subgroups so that an overall rate could be computed.

The most frequent problem in reporting mortality was the failure to give elective and emergency rates when applied. About 72% of the studies included multiple types of operations, where both elective and emergency procedures could have been computed. Often the authors cited rates for the different kinds of surgery such as biliary tract, abdominal, etc., but each of these could vary in terms of the composition by emergency and elective procedures. For example, the not applicable for elective rates refers to those who studied only emergency surgery or those who reported by the alternate method (patient or operation). Therefore, it can be seen that over half of the studies could have reported an elective rate but did not do so. Likewise, about the same proportion did not report the emergency rates when it was possible to do so. Even with crude rates, such as those in Table 2, it is obvious that emergency rates varied widely but averaged three-fold greater risk than the risk of elective surgery.

Table 3 breaks the mortality rates down further by type of surgery, year reported, and lower age included. For this table, the patient mortality figures were used except in cases where only the operative mortality rates were known. It can be seen that overall rates for general surgery, especially the elective surgery rates, have increased over the years. The increase in overall rates are

TABLE 2. *Comparison of Mortality Rates as Reported by Patient and Operation*

Variables	Per Cents	
	By Operation	By Patient
Overall rates		
0-5%	15	12
6-10%	39	28
11-15%	22	28
16-20%	11 Mean = 11.7	20 Mean = 13.6
21-25%	11	4
26-30%	1	4
over 30%	0	5
reported	43	70
not reported	1	1
not applicable	56	29
Elective rates		
0-5%	29	42
6-10%	33	32
11-15%	24 Mean = 8.8	16 Mean = 7.9
16-20%	14	0
21-25%	0	11
reported	19	17
not reported	58	57
not applicable	23	26
Emergency rates		
0-5%		4
6-10%		2
11-15%		7
16-20%		17
21-25%		20 Mean = 29.0
26-30%		6
31-35%		11
36-40%		17
41-45%		4
over 45%		11
reported		43
not reported		55
not applicable		2

TABLE 3. *Comparison of Mortality Rates by Years Published, Lower Age Included, and General or Specialty Surgery*

	Mortality Rates					
	Overall		Elective		Emergency	
Variables	General	Specialty	General	Specialty	General	Specialty
Years						
1931-40	10.3	13.0	11.0	4.2	31.0	34.5
1941-50	10.8	—	5.0	—	32.1	62.0
1951-60	11.1	—	7.3	—	21.3	—
1961-70	14.7	16.8	9.2	—	29.3	—
1971-80	12.4	13.7	9.5	9.1	25.1	41.6
Lower age						
Under 60	8.4	—	1.3	—	—	—
60-70	11.3	13.9	7.5	8.1	25.7	50.9
70-79	11.4	17.1	8.2	14.0	26.7	49.0
80-89	15.1	4.7	11.3	2.3	26.6	10.5
90 and Over	22.1	—	—	—	28.0	—
Number reporting	(66)	(20)	(31)	(5)	(41)	(6)

probably a reflection of increase in the elective rates, since emergency mortality rates in general surgery tended to go down. Specialty surgery mortality rates, on the other hand, showed no definite trends over the years. When emergency rates for general and specialty were compared, the overall emergency rate for general was 27.6% and for specialty was 42.8%. This was statistically significant ($F = 5.27$, $p < 0.05$) for a difference between the two groups. Considering the lower ages included in the studies, it is obvious that the mortality rate increased as the lower age included became higher for overall and elective general surgery, but does not increase as much for emergency general surgery with ages studied. Again, no definite pattern seems to exist with specialty surgery and lower age included.

Confidence in Results Scale

The average rating of the 108 studies was 7.15 (SD = 1.99), with only one study scoring under three and only four over 10. The highest possible score was 16. Table 4 shows distribution of scores and the average scores for different types of journals. The studies published in specialty surgical journals had the highest average rating. Ratings for geriatric journals were a close second and ratings for general medical and surgical journals were lower.

TABLE 4. *Classification of Articles by Confidence in Results Scale, Journals, and Types of Surgery*

Variable	Number	Per Cent	Mean Score
Rating on scale (total score)			7.15
Under 3	1	1	
3-6	39	36	
7-10	64	59	
Over 10	4	4	
Type of Journal			
general medical	32	30	6.72
general surgical	41	38	6.92
geriatric	20	19	7.50
surgical specialty	15	14	7.93

Studies were divided into high and low scoring articles, with those scoring at or above the mean (seven points) on the scale assigned to the high score group. Table 5 shows how the studies compare on several variables when confidence in results was considered high or low. Two variables discriminated between high and low confidence studies. Those published more recently were more often found in the high confidence group ($p < 0.05$). Furthermore, those reporting specialty surgery were considered significantly better by the scale ($p < 0.01$). Items which did not discriminate between studies were university affiliation, being done in the United States, number of patients included, lower age studied, or mortality rates reported. In regard to mortality rates, however, the trend is consistently toward higher rates of mortality being associated with studies having higher confidence ratings. In fact, the overall rate discriminates at $p < 0.09$, which might be considered a strong trend.

Discussion

What conclusion can be drawn from the studies reviewed? First, a rather large number have been done over the past 40 years, and the 108 reviewed are by no means an exhaustive list, although all efforts to be as

complete as possible were exerted. There is agreement that the risk of emergency surgery in the elderly is much greater than elective surgery. The average from these studies would support about three times as great a risk.

Second, when studies were separated out by year, there was a trend toward increasing mortality rates for elective, but not emergency, operations in general surgery since 1941. Two explanations are possible. One is that surgeons are taking patients with greater risks. The other is that operating techniques and/or care are deteriorating. A frequent conclusion concerning operating on the elderly has been that elective surgery may need to be considered more often, since a reluctance to operate could result in the need for emergency surgery, which is known to carry a higher risk of death. Another advantage cited for doing elective surgery is that more attention can be devoted to preoperative preparation of the elderly patient so that the best possible results can be obtained. One can think of other factors associated with increased risk in elective surgery such as trends toward doing more operations among the elderly, more elderly patients in general who are living longer with more chances of having coexisting diseases, older ages being considered for surgery, and more extensive types of surgery being offered than in the past. However, to our knowledge, what seems to be an increase in elective surgery mortality rates over the past few decades has not been reported, and requires a more careful review by surgeons in order to determine the reasons for its occurrence.

What cannot be told from these studies? Probably the one thing that most of these studies hoped to answer (i.e., the relative risk of surgery with age) cannot be determined. Without controls such as comparable patients of younger ages matched for type of operation and co-existing disease with the same surgeon or surgeons of comparable skill, it is almost impossible to sort out the differences attributed to age alone. To properly do this, an operative risk scale is needed which incorporates degree of co-existing diseases, obesity, nutritional status, physiological decrements associated with normal aging, physiologic outlook on life, and other factors which make up a person's physiologic age. This would permit clearer assessment of the impact of chronologic age in and of itself upon operability, morbidity, and mortality.

The confidence in results scale used with these studies points out some of the basic information and design issues that should routinely be provided in any article in order to permit a comparison of mortality rates with other articles. In so many instances, such essential information was missing. For example, many studies failed to provide mean age of the patient population studied, no less a breakdown of age subgroups in terms of outcome. Such information as number of patients and operations was often missing, and failure to compute mortality by elective and emergency categories was found in over half of the studies where this could have been done. Since studies varied in proportion of emergency and elective procedures, the overall rate alone tells very little. The fact that a comparison group of mortality rates for younger patients undergoing comparable surgery in the same institution was seldom used also leaves no way of comparing how an institution does with older versus younger patients. Another serious deficiency was the lack of an equal amount of postoperative follow-up within a single study, much less between studies. It was even more distressing when authors did not provide any information concerning the period of time over which the mortality rate was computed. A few studies separated deaths due to cancer from other mortality rates. There is a need to define this clearly and perhaps report both with and without malignancies included until some standardization of death related to surgery is achieved. The need to report complication rates and causes of death seems obvious. More than half of the studies failed to do so. Our scale dealt only with some of the information needed to compute and assess mortality rates. Other important issues such as reliability of data, criteria for sample selection, control of variance, and proper use of statistical methods were not addressed.

One encouraging note is that there was a definite and statistically significant trend toward better confidence scores appearing in more recent years. This is probably

TABLE 5. *Classification of Studies and Their Results by High and Low Ratings on Confidence in Results Scale*

Variables	Means		F-Ratio
	Low Group	High Group	
Average # of years since publication	19.76	14.32	4.12*
Study conducted in U.S.	.84	.75	.37
Author university affiliated	.52	.48	.09
Number of patients studied	224.72	544.81	.28
Lowest age included	65.40	68.35	.13
Specialty surgery	.12	.40	6.73†
Mortality			
overall	11.25	14.01	2.88
elective	7.41	8.37	.28
emergency	25.94	30.60	1.13

* $P < 0.05$.
† $P < 0.01$.

to be expected. The fact that specialty surgical studies scored higher and that mortality tended to be higher in the better rated studies could reflect the significant difference found between emergency mortality rates in general and specialty surgery and/or that lower confidence studies underestimated mortality rates.

The existence of low confidence scores also suggests a need for more rigorous standards regarding acceptance for publication of surgical studies in the elderly. Again, specialty surgery journals, which reported more of the specialty studies, had the highest mean scores on the scale. Surprisingly, the general surgical journals, which published over a third of articles dealing with surgery in the elderly, ranked among the lowest. The country where the study was conducted and whether or not the authors were university affiliated did not discriminate between high and low scoring studies.

If the focus of this review were simply to be critical in regard to what was missing from studies, this would have been a sterile exercise. Its purpose has been to illustrate some of the sources of variance that need to be taken into account when comparisons are made between studies. Furthermore, in doing this, it is hoped that some control of these sources of variance would be exercised in conducting studies and standards set for reporting and examining mortality rates in the elderly.

References

1. Anderson B, Genster M, Langberg K. Geriatric surgery in a community. Acta Chir Scan 1965; 354(Suppl):1-103.
2. Cole WH. Operability in the young and aged. Ann Surg 1953; 138:145-157.
3. Miller R, Marlar K, Silvary G. Anesthesia for patients aged over ninety years. NY State J Med 1977; 77:1421-1425.
4. Rankin FW, Johnson CC. Major operations in elderly patients. Surgery 1939; 5:763-774.