Numbers Are Better Than Words

Verbal Specifications of Frequency Have No Place in Medicine

MICHAEL A. NAKAO, M.D.
Albany, New York
SEYMOUR AXELROD, Ph.D.
Buffalo, New York

We were concerned about the precision (i.e., consensus as to meaning) of adjectives and adverbs used to express frequency in the medical literature. We asked 103 physicians and 106 nonphysicians to assign to each of 22 such modifiers a percentage representing their understanding of the term. As indexed by the standard deviations, the degree of imprecision for 17 terms was so great for both physicians and for laymen as to make their use unacceptable. Consensus was significantly less among laymen than among physicians for 10 of the terms. Greater consensus was shown by native English-speaking physicians than by those with other native languages. Our data suggest further that American graduates of American medical schools show more consensus than American graduates of foreign medical schools. Board-certified physicians did not show greater consensus than physicians who were not board-certified. Verbal expressions of frequency should be eliminated from medical communications; failing that, the author should specify numerically the frequency he intends when he uses any such expression.

The frequency of events is, in principle, quantifiable and expressible as a proportion. For example, the recent revision of the *Diagnostic and Statistical Manual of Mental Disorders* of the American Psychiatric Association says of infantile autism that "the disorder is very rare (2 to 4 cases per 10,000)" [1]. Likewise, the National Center for Health Statistics reports that in their large data set "the overall incidence of low birth weight in 1976 was 54.3 percent for infants in plural deliveries compared with 6.3 percent for infants in single deliveries" [2].

By contrast, examples abound of the use of modifiers expressing frequency, unaccompanied by numerical values: "This indicates that cases of unexplained pulmonary hypertension, although certainly uncommon, are not exceedingly rare" [3], and "Mean pulmonary artery pressures of 40 mm Hg or more are common in this group, and elevations of right atrial pressure and depressions of the cardiac index are the rule rather than the exception. These patients are also more likely to develop acute left ventricular or global heart failures; massive obstruction with survival is extremely uncommon" [4].

The imprecision of such expressions in ordinary discourse has been discussed in the speech and psychology literature for some years [5–8]. Recent letters to editors of medical journals have again raised the problem, asking how well medical personnel communicate their

From the Departments of Anesthesiology and Medicine, Albany Medical College of Union University, Albany, New York (MAN), and the Department of Psychiatry, State University of New York at Buffalo, Buffalo, New York (SA). Requests for reprints should be addressed to Dr. Michael A. Nakao, Room B110, Albany Medical Center, Albany, New York 12208. Manuscript accepted December 5, 1982.

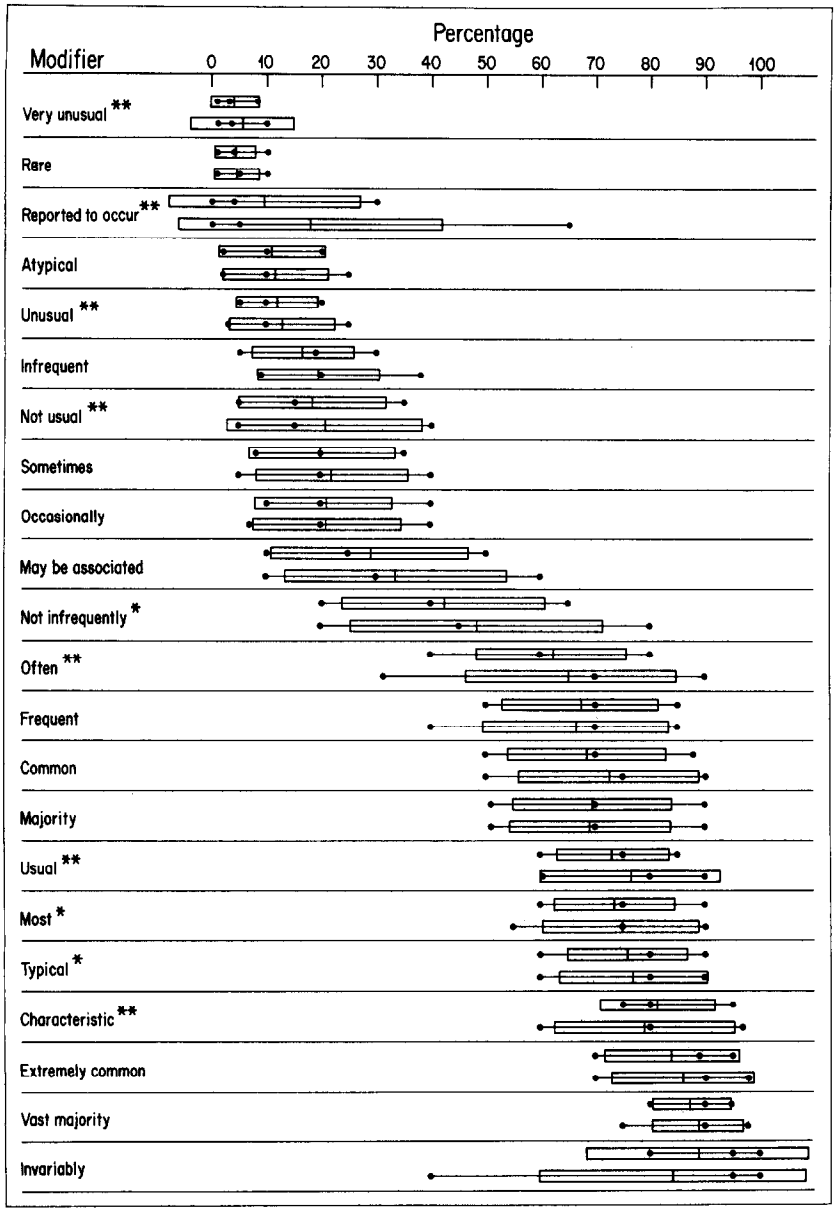

Figure 1. *Central tendency and variability of the modifiers. For each modifier, the* ***upper display*** *represents the data from physicians, and the* ***lower display*** *the data from the nonphysicians.* ***Rectangles*** *mark off one standard deviation (SD) above and below the mean, represented by a* ***vertical line.*** *The* ***dots on the ends*** *of each line segment represent the 10th and 90th percentiles (P10 and P90); the* ***intermediate dot*** *shows the location of the median (P50).* ***Asterisks*** *indicate terms for which the standard deviations were significantly different in the two groups. *$p \leq 0.05$; **$p \leq 0.01$.*

intended meanings when they use such terms. Bryant and Norman [9] generated a list of 30 expressions of probability (e.g., *probable, sometimes, doubtful, always, unlikely*); they asked a small group of physicians (N = 16) to indicate the likelihood implied by each expression, and then repeated the test at least two days later. Although the respondents were consistent over this interval, there was striking interphysician variability. For example, the assigned likelihoods for *sometimes* ranged from 1 percent to 80 percent.

Toogood [10] asked 51 skilled and professional workers, including an unspecified number of physicians, to quantify nine adverbs (e.g., *always, usually, occasionally, never*). As an illustration of the degree of interrespondent variation, for the adverb *often*, two thirds of the respondents assigned percentages ranging from 43 to 75; worse, 5 percent of the respondents assigned values greater than 92 percent or less than 28 percent.

Kenney [11] had 24 members of the Pathology Department at the Massachusetts General Hospital quantify 21 terms, and found ranges of 25 percentage points or more for 15 of the terms.

These reports are based on small numbers of respondents (16 to 51), but they are valuable in raising additional questions. Apart from the imprecision caused by the use of these expressions in interphysician communication, what can we say about their effect

when physicians are communicating to laymen? Are there verbal expressions for which consensus is so high that they can be used without fear of misunderstanding? On the other hand, are there expressions whose ambiguity is so great that they ought to be avoided? Can we begin to specify characteristics of physicians or laymen that influence the understanding that they have of the expressions? We address these questions in the present study.

METHOD

Twenty-two verbal modifiers expressing frequency were selected as representative of those encountered in medical textbooks and journals. One hundred three physicians and 106 nonphysicians responded to a form listing the expressions; the list was preceded by the following instructions: "What do the following terms mean to you when used in a medical context? For example, if you read that a finding is 'atypical' in a certain disease, what percent of the cases of that disease do you think would have it? Please assign each term a SINGLE percentage. Thank you."

Sex, age, and place of birth were also elicited. Nonphysicians were asked to specify their occupation and the number of years of postsecondary education they had completed. Physicians were asked to specify where they had graduated from medical school; board certification was determined from the *Directory of Medical Specialists* [12].

Descriptive statistics were derived from the computer program BMDP2D, and inferential statistics (*t* tests for mean differences and *F* tests for differences in standard deviation) from BMDP3D [13]. Differences were considered statistically significant when $p \leq 0.05$.

RESULTS AND COMMENTS

Of the 94 physician respondents who gave their sex, eight were female and 86 were male; of the 94 nonphysicians who specified their sex, 47 were female and 47 were male. The physicians' (N = 100) mean age was 33.9 years (SD = 8.41); for the nonphysicians (N = 98), the mean age was 40.9 years (SD = 15.61). Seventy of 99 physicians were born in English-speaking countries, as were 96 of 104 nonphysicians. The mean number of years of postsecondary education in the 106 nonphysicians was 5.4 (SD = 2.11).

Mean values, median values, standard deviations, and 10th and 90th percentiles (P10 and P90) for the 22 expressions are shown in **Figure 1.** The expressions are arranged in order of increasing magnitude of the mean for the physician group. The upper display for each expression represents the data from the physicians; because not all respondents assigned a percentage to all expressions, the numbers of respondents vary from 98 to 103. The lower display shows the measures derived from the nonphysicians; numbers of respondents vary from 98 to 106.

Table I lists the expressions in order of descending size of the difference P90-P10 in the physician group; this index of variability represents the range of responses of the middle 80 percent of that group. We use this index rather than the total range because it eliminates the effects of aberrant values at the extremes.

TABLE I The 22 Modifiers in Decreasing Order of P90-P10 in the Physician Group

Modifier	P90-P10 for Physicians	P90-P10 for Nonphysicians
Not infrequently	45	60
May be associated	40	50
Often	40	58.5
Majority	39	39
Common	38	40
Frequent	35	45
Not unusual	30	35
Occasionally	30	33
Reported to occur	30	65
Typical	30	30
Sometimes	27	35
Extremely common	25	28
Infrequent	25	29
Most	25	35
Usual	25	30
Characteristic	20	37
Invariably	20	60
Atypical	18	23
Unusual	15	22
Vast majority	15	23
Rare	9	9
Very unusual	7	9

Which Terms Are Acceptable? For 15 of the 22 terms (*not infrequently* through *usual* in Table I), the index P90-P10 was at least 25 percentage points in both occupational subsets; for two additional terms, *characteristic* and *invariably*, the index was 20 percent for the physicians and at least 25 percent for nonphysicians. In our opinion, the meanings of these 17 terms are too variable for the writer or speaker to be sure that his intended meaning is being communicated when he uses them.

Kenney [11] reported that those of his expressions whose mean values were near 0 percent or 100 percent were less subject to variation than those whose mean values were more centrally located; the extent of possible variation is naturally more constrained at the extremes than in the central range. Similarly, we found that expressions whose mean values were between 20 percent and 80 percent for both physicians and nonphysicians had values of P90-P10 equal to or greater than 25 percent in both groups. However, of the nine terms whose mean values were lower than 20 percent or higher than 80 percent in both groups, four had values of P90-P10 of 25 percent or greater in one or both.

Differences between Physicians and Nonphysicians. The mean values of only three of our terms

were significantly different for the two groups of respondents; for two of these, *infrequent* and *not infrequently*, the differences were only 2.83 percent and 5.88 percent, respectively. The larger difference for *reported to occur*, 8.41 percent, appears to indicate that this is a jargon term for physicians, who understand it to imply a "low likelihood." Laymen, on the other hand, may tend to take the words *reported* and *occur* in their usual English acceptations, and therefore to understand the combination to imply a "greater likelihood." Supporting this interpretation is the fact that P90 for the lay group for *reported to occur* extended to 65 percent, in contrast with 30 percent for the physicians (Figure 1). Furthermore, we found that the subgroup of physicians whom we expected to be particularly vulnerable to jargonization of the term, namely those born and educated in the United States or Canada (N = 53), had a mean of only 6.2 percent (SD = 11.64) for *reported to occur;* even highly educated, native English-speaking laymen (six or more years of postsecondary education, N = 43) showed a bimodal pattern of responses for this term.

The differences in variability were more systematic than the differences we obtained for the mean values. To permit inferential statistical comparisons, we shifted to standard deviation as the measure of scatter. As shown in Figure 1, the standard deviations for 10 of the terms were significantly smaller for the physician group than for the lay group; the only difference in the opposite direction (for *atypical*) was not significant. For our total samples, then, the physicians showed greater interrespondent agreement than the laymen.

Sex: Because only 9 percent of the physician group were female, we turned to the nonphysician group (47 female, 47 male) for information on possible effects of sex. For three modifiers (*invariably, rare, very unusual*), the standard deviation was significantly higher for the females than for the males; for *not usually*, the difference was significant in the opposite direction.

Amount of education: Our nonphysician group had significantly fewer years of formal education than our physician group, and we wondered whether amount of education could account for the physician/layman differences. We compared male, native English-speaking nonphysicians with six or more years of postsecondary education (N = 33) with male, native English-speaking physicians (N = 58). For seven of the expressions, the standard deviation for the physicians was significantly smaller than that for the nonphysicians (*characteristic, invariably, may be associated, not usual, reported to occur, sometimes, usual*); for *very unusual*, the gradient was significant in the reverse direction. This finding suggests that differences in amount of education were not the major determinant of the physician/layman differences in variability. However, when we divided our native English-speaking lay group into those with five or fewer years of postsecondary education (N = 48, mean = 3.8 years, range 2 to 5 years) and those with six or more years (N = 48, mean = 6.8 years, range 6 to 12 years), we did find significantly larger standard deviations in the former group for four modifiers (*invariably, most, occasionally, very unusual*), and a reverse significant gradient for one (*not usually*). We therefore cannot rule out amount of education as a restricter of variability. If in fact it is, then the differences between the physician and lay groups that we observed surely underestimate the potential for faulty communication between physicians and their patients. Our lay sample was certainly not representative of the lay population with respect to education. None of our lay respondents had less than one year of postsecondary education (median = 6.0); the median for persons 25 years old or over in the United States in 1979 is estimated to be 0.5 years [14]. It is worth noting that the curtailed range of education in our lay sample would by itself have the effect of underestimating the absolute magnitude of the effect of education on variability.

Does Scientific Training Promote Consensus? All physicians have received extensive education in science; by contrast, half of our nonphysicians were not employed in scientific fields. We compared our group (N = 18) of native English-speaking nonphysicians with six or more years of postsecondary education and scientific occupations (e.g., computer programmers, science professors, medical students) with our group (N = 30) of native English-speaking nonphysicians who also had six or more years of education but who were in nonscientific occupations (e.g., librarians, professors of humanities, administrators). Only for *characteristic* was there a significant difference in standard deviation, with variability greater in the nonscience group. Scientific training and occupation, then, do not appear to make for systematically decreased variability.

Are There Subgroups of Physicians Who Are More Consistent Than Others? Board certification: We expected that the filtering process involved in specialty-board certification would result in systematically more consistent use of these expressions. However, when we compared, among the native English-speaking physicians in our sample, the board-certified subgroup (N = 27) with the subgroup without board certification (N = 43), we found that while for two expressions the former subgroup had significantly smaller standard deviations, for three they had significantly larger standard deviations.

Native language: Expressions in any language should have greater commonality of meaning among native speakers than among those who acquired the language later in life. Our group of native English-speaking physicians (N = 70) did indeed have signifi-

cantly smaller standard deviations for seven terms (*atypical, characteristic, infrequently, invariably, rare, typical, usual*) than the physicians whose native language was not English (N = 29), with a significantly larger standard deviation for only one term (*very unusual*).

American versus foreign medical degree: Finally, we wondered whether American foreign medical graduates would show more consensus on the meaning of the modifiers than Americans who received their medical degrees from American schools. Our sample of United States and Canadian foreign medical graduates (N = 14) did have significantly higher standard deviations than the United States and Canadian American medical graduates (N = 53) for eight of the 22 expressions (*common, extremely common, invariably, not infrequently, occasionally, rare, reported to occur, sometimes*); there were no significant differences in the opposite direction. Any interpretation of this finding needs, of course, to be tempered by consideration of the small size of the foreign medical graduate group.

In conclusion, physicians should take no consolation from the fact that they show greater consensus on the quantitative meanings of verbal expressions of frequency than laymen do; the results shown in Figure 1 and Table I highlight the folly of assuming that any two randomly chosen physicians are likely to have similar percentages in mind when they use any such term; and the likelihood of misunderstanding is even greater in physician/layman communication. During the writing of the text of the present paper, we felt very uncomfortable in doing without expressions of this sort. Nevertheless, we recommend that these terms be carefully avoided in order to eliminate fuzziness in communication; the physician-author should determine from the literature the reported frequency of events, and should use those numbers. When this cannot be done, we propose with Kenney [11] that the author at least provide a numerical estimate of the frequency he has in mind, identifying it as an estimate.

What about expressions like the five that were not classified as unacceptable by our arbitrary criteria (the last five in Table I); might not they be used? We would answer that no helpful function (other than hedging) is served by their use. Minimal ambiguity is still ambiguity. An author who means "5 percent" when he writes "atypical" can be mistaken to mean "15 percent" or even "20 percent." No such mistake can occur if he writes "5 percent" in the first place.

ACKNOWLEDGMENT

We are grateful to Bernard DeLeo, MD, for helpful comments on an earlier version of this paper.

REFERENCES

1. Task Force on Nomenclature and Statistics: Diagnostic and statistical manual of mental disorders. 3rd ed. Washington, DC: American Psychiatric Association, 1980: 89.
2. Taffel S: Factors associated with low birth weight: United States—1976 (DHEW publication no. [PHS] 80-1915). Washington, DC: Government Printing Office, 1980: 10–11.
3. Wagenvoort CA, Wagenvoort N: Primary pulmonary hypertension: a pathologic study of the lung vessels in 156 clinically diagnosed cases. Circulation 1970; 42: 1163–1184.
4. McFadden ER, Braunwald E: Cor pulmonale and pulmonary thromboembolism. In: Braunwald E, ed. Heart disease: a textbook of cardiovascular medicine. Philadelphia: WB Saunders, 1980: 1643–1680.
5. Simpson RH: The specific meanings of certain terms indicating differing degrees of frequency. Q J Speech 1944; 30: 328–330.
6. Hakel MD: How often is often? Am Psychol 1968; 23: 523–534.
7. Pepper S, Prytulak LS: Sometimes frequently means seldom: context effects in the interpretation of quantitative expressions. J Res Personality 1974; 8: 95–101.
8. Bradburn NM, Miles C: Problems in using imprecise quantifying words. In: Bradburn NM, Sudman S, eds. Improving interview method and questionnaire design. San Francisco: Jossey-Bass, 1979: 152–162.
9. Bryant GD, Norman GR: Expressions of probability: words and numbers [letter]. N Engl J Med 1980; 302: 411.
10. Toogood JH: What do we mean by "usually" [letter]? Lancet 1980; I: 1094.
11. Kenney RM: Between never and always [letter]. N Engl J Med 1981; 305: 1097–1098.
12. American Board of Medical Specialists: Directory of medical specialists. 20th ed. Chicago: Marquis' Who's Who, 1981.
13. Dixon WJ, Brown MB, eds: BMD-79: biomedical computer programs, P-series. Berkeley: University of California Press, 1979: 146–154, 170–184.
14. U.S. Bureau of the Census: Statistical abstract of the United States: 1980. 101st ed. Washington, DC: Government Printing Office, 1980: 149.