
LETTERS TO THE EDITORS

INDEPENDENT OBSERVATION IN THE REVIEW OF SEQUENTIAL DAYS CLUSTERED BY STAY

The Fall 1993 issue included an interesting article on the review of unnecessary hospital use, the aim of which was to show the reliability of a Swiss adaptation of the Appropriateness Evaluation Protocol (4). The author shows a first revision, independently performed by two reviewers, of all of the days (1,847) corresponding to 212 stays, and a second one, independently performed by three reviewers, of all of the days (508) of 59 stays. Results are presented as the percent of agreement between reviewers, percent specific agreement, and κ statistics.

All of these statistical methods used for measuring interobserver agreement assume that the observations are independent of each other (1;2). This assumption is violated in the paper as the review of sequential days of stay does not result in independent judgments by the reviewers, because the judgment on a stay strongly predicts the statement of the successive stays (2;5). This is not a statistical problem but the way of selecting the days to review. The problem could have been avoided by selecting at random one day of each stay or with some other sample formulas that would improve the independence of the observations. This problem does not appear in similar studies cited by the author, because they were performed on a review of only one day of each stay (2;3) or on sample days of each stay (5).

Use of statistical methods designed for independent events in observations clustered by stay biases the results and tends to improve agreement between observers, and subsequently a higher Cohen's κ is obtained. Consequently, results regarding the appropriateness of successive days of stay from Santos-Eggiman's paper do not represent independent judgments by the reviewers, so the conclusions of this study cannot be assessed.

SALVADOR PEIRÓ
SANTIAGO PEREZ
EDUARD PORTELLA
*Valencian Institute for Public Health Studies
Juan de Garay 21
46017 Valencia
Spain*

REFERENCES

1. Fleiss, J. L. *Statistical methods for rates and proportions*. New York: John Wiley & Sons, 1981.
2. Gertman, P. M., & Restuccia, J. D. The Appropriateness Evaluation Protocol: A technique for assessing unnecessary days of hospital care. *Medical Care*, 1981, 19, 855-871.
3. Risphon, S., Lubacsh, S., & Epstein, L. M. Reliability of a method of determining the necessity for hospitalization days in Israel. *Medical Care*, 1986, 24, 279-82.

4. Santos-Eggimann, B. Hospital utilization reviews under field conditions: Potential and improvements. *International Journal of Technology Assessment in Health Care*, 1993, 9, 514–21.
5. Strumwasser, I., Paranjpe, N. V., Ronis, D. L., et al. Reliability and validity of utilization review criteria. Appropriateness Evaluation Protocol, Standardized Medreview Instrument, and Intensity-Severity-Discharge Criteria Set. *Medical Care*, 1990, 28, 95–111.

Dr. Santos-Eggimann replies:

I thank Dr. Pieró and his colleagues for their interest in our paper and add the following comments.

If the appropriateness of one day in the hospital may be related to the appropriateness of the previous day (i.e., delays due to unavailability of support at discharge may last more than one day), there is no reason to believe that judgments made on successive days based on a list of explicit criteria are dependent.

Pieró and colleagues do not provide evidence that judgments made on reviews performed on a single day (1;2), as in the studies to which our results were compared, are more in consonance with the principle of independence of observations than studies based on a longitudinal design. In fact, many delays in the hospital may be related to the availability of given services within and outside the hospital on given days. In addition, the single day, cross-sectional approach is not likely to provide a correct sampling of hospital days from the point of view of patient case-mix.

The third study quoted in the paper (3) was performed on all consecutive days for patients admitted for a stay of 7 days or less and on a sample of 7 days in each stay of longer duration. Since a large proportion of hospital stays are short (i.e., 54% of the stays in Hospital A mentioned in our study were less than 7 days, and an additional 22% were less than 14 days), such a sampling procedure would not have been likely to provide very different estimates of κ values, and comparisons are not unreasonable as far as the sampling procedure is concerned.

There is no doubt that from the purely statistical perspective the random selection of one single day in each stay of a representative sample of stays or the selection of a constant number of observations at each selected date in a representative sample of calendar days would provide better estimates of the true proportion of inappropriate days in the hospital. However, the gain in statistical validity must be weighed against the loss of valuable information when only one day is analyzed in stays reviewed in their whole length. A re-analysis of our data, based on a random selection of one single day in each of the selected stays, as advocated by Peiró et al., produces κ statistics identical to (or slightly higher than) our published estimates (Table 1). This speaks against the strong influence of judgments made on consecutive days of a single stay.

Sample options more in conformity with statistical assumptions, while desirable, are costly, particularly when double or triple reviews are considered for agreement studies. If their advantage is clear for estimating the true proportion of inappropriate days, it is less so when studying agreement concerning the evaluation of appropriateness based on explicit criteria.

REFERENCES

1. Gertmann, P. M., & Restuccia, J. D. The appropriateness evaluation protocol: A technique for assessing unnecessary days of hospital care. *Medical Care*, 1981, 19, 855–71.

Table 1. Kappa Estimates Based on Random Selection of One Day in Each Selected Stay and Based on All Days of Selected Stays

| | Cohen's κ | |
|------------------------|------------------|----------|
| | One day | All days |
| Hospital A (81 stays) | | |
| RHA \times R1 | 0.95 | 0.95 |
| Hospital C* (59 stays) | | |
| RHC \times R1 | 0.90 | 0.80 |
| RHC \times R2 | 0.90 | 0.78 |
| R1 \times R2 | 1.00 | 0.98 |

* Triple review.

RHA, C = Hospital A, C reviewers.

R1, R2 = 1st and 2nd Public Health Department reviewers.

2. Rishpon, S., Lubacsh, S., & Epstein, L. M. Reliability of a method of determining the necessity for hospitalization days in Israel. *Medical Care*, 1986, 24, 279-82.
3. Strumwasser, I., Paranjpe, N. V., Ronis, D. L., et al. Reliability and validity of utilization review criteria. Appropriateness Evaluation Protocol, Standardized Medreview Instrument, and Intensity-Severity-Discharge Criteria Set. *Medical Care*, 1990, 28, 95-111.