VIII. Structured Review of an Article

Juan O. Talavera, Rodolfo Rivas-Ruiz

Several strategies have been attempted to select an article under assumptions of relevance and good quality. They depend largely on the presence or not of a series of features and in other occasions on the judgment of those who classify the article. However, these strategies do not allow for us to know the magnitude of error. Since there is no such thing as a perfect article, it is relevant to identify the magnitude of error and its impact on the final result; hence, it is necessary to develop skills that allow for us to review an article, identify possible errors and generate an idea of their impact on the result. According to the information contained in parts I to VII of this series of articles on clinical research, we have tried to demonstrate its application in a structured review of a causality article, starting with the examination of the baseline state, the maneuver and the result, with the systematic errors (biases) generated in each item, followed by the relevance of the test, the appropriateness of the sample size and, finally, clinical relevance.

Key words
journal article
causality
statistics and numerical data
biases
sample size
association, exposition, risk or outcome measures

This article was originally published in Rev Med Inst Med Seguro Soc 2012; 50 (2): 163-166 and it has been reviewed for this issue.

Introduction

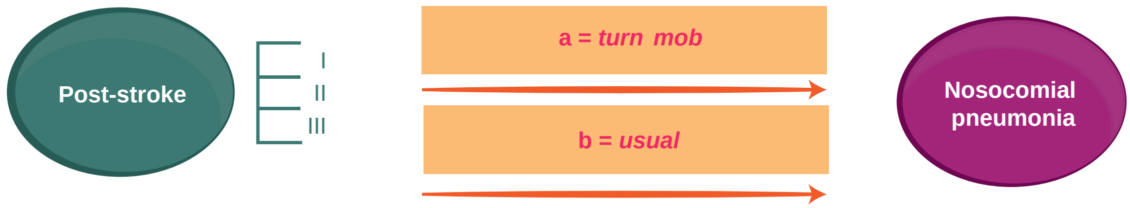
Several strategies have been attempted to select an article under assumptions of relevance and good quality. They depend largely on the presence or not of a series of characteristics and, in other occasions, on the judgment of those who classify the article. This entails a classification of "adequate" or "inadequate", or in best case to a graduation of major to minor quality or relevance. However, these strategies do not really allow for us to know the magnitude of error. And since there is no such thing as a perfect article, it is important to identify the magnitude of error and the impact it may have had on the final result; hence, it becomes necessary to develop skills that allow for us to review an article in a structured way, to identify possible errors and to generate an idea of their impact on the result. That is, we cannot rely on a classification or on the judgment of others to decide what to read and what not to read, or what to consider adequate or inadequate. We will have to learn the minimum basic structure that allows for us to assess ourselves the relevance of each article, its errors and its results.

In parts I and III to VI of this series on clinical research, we have tried to show the characteristics that we consider as being basic to perform a reading and a structured review of an article on causality (risk factor or etiologic agent, prognosis or treatment), once the article has been identified by means of a systematic search (topic addressed in part VII). We started with a model comprising the baseline state, the maneuver and the result (described in article I), with the systematic errors (biases) generated when defining and operating each of these items (article III). And we continued with the appropriateness of the test (part IV), the sample size estimation (part V) and, finally, the clinical relevance (part VI).

Next, we will make an exercise on the use of said information under a structured review proposal; for that, we will use an article of our own authorship: "Reduction in the incidence of post-stroke nosocomial pneumonia by using the 'Turn-Mob' Program", published in the Journal of Stroke and Cerebrovascular Diseases 2010;19:23-28. The purpose of the study was to demonstrate the efficacy of a program of mobilization in bed named "turn-mob" in decreasing the incidence of nosocomial pneumonia in patients with ischemic stroke.

In Figure 1, we can find baseline state characteristics such as the form of test selection and the prognostic demarcation; we can observe that randomization was able to balance the groups' characteristics, with the exception of chronic obstructive pulmonary disease, slightly higher in group *b* (14 %

Population selection method
Patient with acute neurological deficit,
> 12 hours duration referred from
emergency department or internal medicine



Demarcation diagnosis	**Prognostic stratification: group a versus b**	
< 48-hour evolution	*Chronometric*	72 and 74 years of age
No requirement of ventilatory support	*BMI status*	Normal 18 versus 17%; overweight 69.4 versus 70.5%; Obesity 12.6 versus 12.5%
First vascular event	*Clinical*	Motor deficit, hemiparesis 66.7 versus 75.9 % Hemiplegia 33.3 versus 24.1 %; aphasia 50.5 versus 40.2 % Sensory deficit: 56.8 versus 40.2; nauseous reflex 82 versus 79.5 % Glasgow score 15, 40.5 versus 32.1 % NIHSS score 2- 7, 30.6 versus 32.1 % 8-13, 41.4 versus 43.8 % 14-18, 16.2 versus 17.9 % 19-23, 11.7 versus 6.3 %
No clinical evidence of upper/lower RTI		
No psychomotor agitation		
Ischemic stroke tomographic diagnosis	*Morphologic*	Cerebrovascular disease subtype Anterior circulation partial infarction 88.3 versus 90.2 %
Those developing RTI in the first 48 hours were excluded	*Comorbidity*	DM 50.5 versus 42%; HBP 83 versus 84%; COPD 7 versus 14%; CVE 39 versus 40%
	Previous treatment	Corticosteroids; antibiotic
	Socioeconomic, cultural, habits =	Smoking 31 versus 35 %; alcohol 24 versus 24 %

RTI = respiratory tract infection; BMI = body mass index; DM = diabetes mellitus; HBP = high blood pressure, COPD = chronic obstructive pulmonary disease

Figure 1 Baseline state characteristics: diagnostic demarcation (selection criteria) and prognostic stratification (demarcation) (variables that impact on the outcome regardless of the maneuver)

versus 7 %, $p = 0.088$), which could have impacted on the final result. Since a stratified analysis was not performed, is not possible to observe the impact of each maneuver according to different risk factors and thus, the result can be attributable mainly to the average characteristics of the population under study.

In Figure 2, the quality of the maneuver application (turn-mob program against usual position changes) has to be considered, verifying that peripheral maneuvers are implemented similarly in both groups.

Although there were no differences in peripheral maneuvers, the application of the turn-mob program was initially standardized and verified day by day; on the other hand, the application of the usual treatment was never standardized or verified and, therefore, there is no guarantee that it was carried out; furthermore, at hospital discharge, the patient did not receive nursing support at home. This could represent more than superiority for the turn-mob program over the usual treatment: the result of application of the turn-mob program against nothing.

Regarding the outcome, there was no possibility of having differentially detected the presence of nosocomial pneumonia, since all patients were submitted to chest X-ray at discharge or upon the slightest clinical suspicion. Similarly, there was no problem due to patient losses (transfer bias), since only two patients were excluded out of a total of 225 due to the presence of pneumonia within the first 48 hours of admission to the hospital (Figure 3).

General Comments

As an overall comment on the methodologic design and development of the project, we could say that the population selection was adequate (adequate assembly), by considering subjects with high probability of developing nosocomial pneumonia and in whom the application of the program turn-mob was feasible. The distribution of different prognostic factors was shown to be similar between groups, which

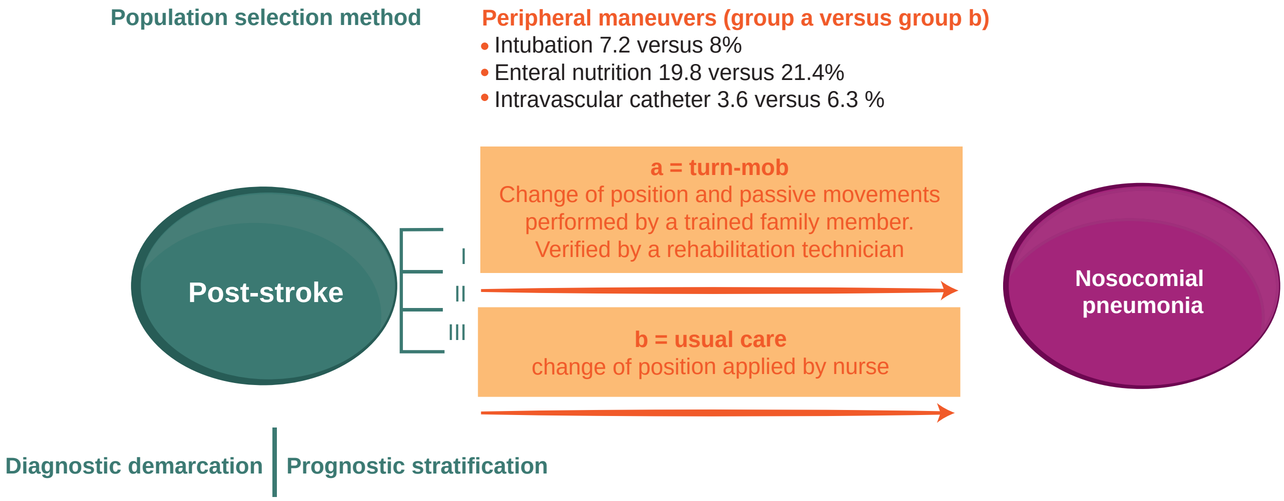


Figure 2 Characteristics to be considered during the application of the maneuver

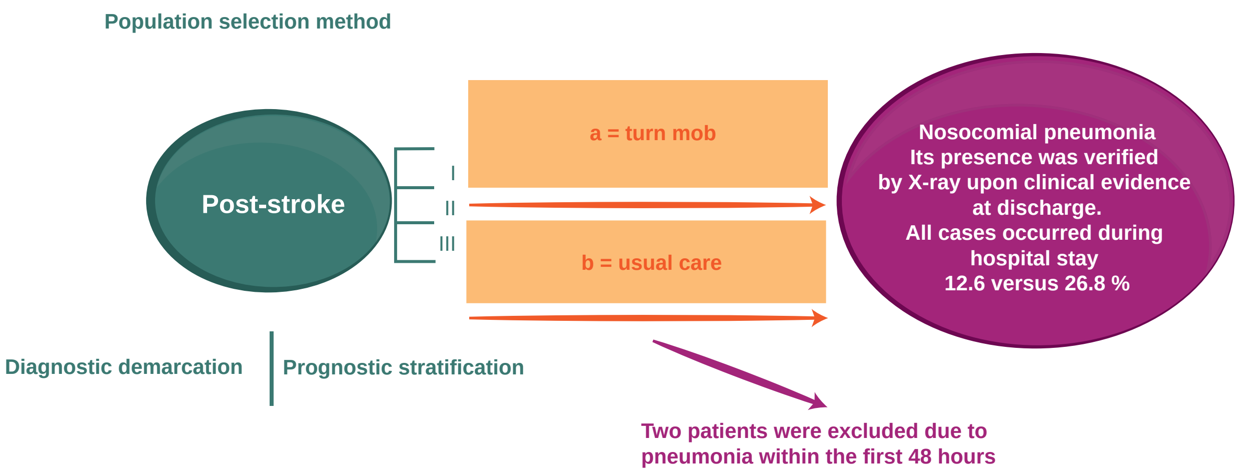


Figure 3 Characteristics to be considered in the outcome

partially prevented susceptibility bias, since no stratified analysis was performed that would allow for the maneuver to be assessed in different risk groups (prognostic susceptibility). As for the maneuver, the adequate execution of the usual maneuver was not properly supervised and, therefore, we cannot guarantee that there was no performance bias. The outcome measure was the same in both groups, which prevented detection bias. Finally, we did not observe losses that could have reversed the observed difference in the outcome between groups (there was no transfer bias).

Regarding the test used (topic developed in Part IV of this series on clinical research), the chi-square shows the comparison of a nominal outcome variable between two groups, such as the presence or not of nosocomial pneumonia.

On the other hand, although the absence of a difference between the presence of diverse characteristics and the treatment group was demonstrated (chi-square test), a multivariate adjustment for the effect of the turn-mob program would have been attractive, due to the multiple characteristics of the baseline state and the co-maneuvers that could have impacted on the outcome. In this case, the multiple logistic regression test would have been appropriate, since the outcome nominal.

As for the sample size (addressed in part V), the method used for its calculation is not mentioned; however, we should remember that this calculation is performed in order to obtain the required number of patients to demonstrate that an expected difference between two groups is real and not by chance. In this case the observed clinical difference of 12.6 % versus 26.8 % was statistically significant and thus, we can assume that it is real, since the probability of it being due by chance is lower than 5 % ($p < 0.05$). And even when the calculations are not described, with the incidence of 2 to 23 % mentioned in the introduction, we can estimate that the highest value was used and a direct reduction of about 15 % was considered, which yields a sample size between 90 and 103 subjects per group (Fleiss-Kelsey formula) and if we add 20 % to this, we obtain a value around the 225 subjects included in the study (sample size estimation for proportions difference).

Finally, in general, direct differences greater than 10 % or an NNT≈10 (CI-VI) were considered clinically relevant. In this case, the difference was 14.2 and the NNT consisted of 7.04 patients (which rounded is equivalent to 8) to see the benefit in one. With these results, we can clearly conclude that it is clinically relevant.

Conclusions

We cannot rule out the presence of a performance bias where the usual treatment would had not been carried out, in which case the conclusion would not be that the turn-mob program is better than usual mobilization performed by nursing staff, but rather it would have to be concluded that the program turn-mob in a post-ischemic stroke patient is better than no rotation or mobilization. On the other hand, we cannot identify whether the turn-mob program retains its benefit in different severity strata, since no stratified analysis was performed and no adjustment was made through a multivariate analysis; probably, these analyses were not performed due to the sample size, since 44 nosocomial pneumonia cases are insufficient when stratifying or adjusting. As we can see, every study has errors and yet, there is valuable information; however, to weigh it, is essential to have some notion on clinical research.

Bibliography

Talavera JO. Clinical research I. The importance of the research design. Rev Med Inst Mex Seguro Soc. 2011;49(1): 53-8.

Talavera JO, Wacher-Rodarte NH, Rivas-Ruiz R. Clinical research II. Studying the process (the diagnosis test). Rev Med Inst Mex Seguro Soc. 2011;49(2):163-70.

Talavera JO, Wacher-Rodarte NH, Rivas-Ruiz R. Clinical research III. The causality studies. Rev Med Inst Mex Seguro Soc. 2011;49(3):289-94.

Talavera JO, Rivas-Ruiz R. Clinical research IV. Relevancy of the statistical test chosen. Rev Med Inst Mex Seguro Soc. 2011;49(4):401-5.

Talavera JO, Rivas-Ruiz R. Clinical research V. Sample size. Rev Med Inst Mex Seguro Soc. 2011;49(5):517-22.

Talavera JO, Rivas-Ruiz R. Clinical research VI. Clinical relevance. Rev Med Inst Mex Seguro Soc. 2011;49(6):631-5.