

Does peer review predict the performance of research projects in health sciences?

L. ERIK CLAVERÍA,^{1*} ELISEO GUALLAR,^{2*} JORDI CAMÍ,^{3*} JOSÉ CONDE,^{4*}
ROBERTO PASTOR,^{5*} JOSÉ R. RICOY,^{6*} EDUARDO RODRÍGUEZ-FARRÉ,^{7*}
FERNANDO RUIZ-PALOMO,^{8*} EMILIO MUÑOZ^{9*}

¹Neurology Unit, General Hospital, Segovia (Spain)

²Department of Epidemiology and Biostatistics,

National School of Public Health Instituto de Salud Carlos III, Madrid (Spain)

³Institut Municipal d'Investigació Mèdica (IMIM) and Universitat Pompeu Fabra, Barcelona (Spain)

⁴Agency for Health Technology Assessment, Instituto de Salud Carlos III, Madrid (Spain)

⁵Department of Epidemiology and Biostatistics,

National School of Public Health, Instituto de Salud Carlos III, Madrid (Spain)

⁶Neuropathology Unit, Hospital 12 de Octubre, Madrid (Spain)

⁷CSIC Department of Pharmacology and Toxicology, Barcelona (Spain)

⁸Division of Internal Medicine, Hospital Ramón y Cajal, Madrid (Spain)

⁹Institute for Advanced Sociologic Research, Madrid (Spain)

Peer review is a basic component of the scientific process, but its performance has seldom been evaluated systematically. To determine whether pre-approval characteristics of research projects predicted the performance of projects, we conducted a retrospective cohort study of all 2744 single-centre research projects financed by the Spanish Health Research Fund since 1988 and completed before 1996. Peer review scores of grant applications were significant predictors of performance of funded projects, and the likelihood of production was also higher for projects with a basic research component, longer duration, higher budget or a financed research fellow. Funding agencies should monitor their selection process and assess the performance of funded projects to design future strategies in supporting health sciences research.

Introduction

The review and funding of research projects by public agencies has seldom been subject to the scrutiny and accountability of the scientific method.¹⁻⁵ At funding agencies, peer reviewers evaluate the scientific merit of project proposals based on

* On behalf of the Study Group on Policy Research in Health Sciences. Other members of the Study Group were Luis Coma, Victor Díaz, Juan Espinosa, Agustín Gómez de la Cámara, Isabel Gómez-Caridad, Vicente Larraga, Francisco Pozo, María J. Santesmases, Luis Sanz.

usually accepted criteria, such as relevance, methodological quality, and feasibility.⁶⁻⁸ Although some agencies have made important efforts to improve and standardise project peer review and priority scoring,⁶⁻⁹ this process is still based largely on tradition, on common sense, and on the personal opinions of the staff and peer reviewers of funding agencies. Very little is known about the ability of specific aspects of grant applications, including the results of peer review, to predict the performance of research projects.⁵

Objective

The aim of this study was to evaluate whether the pre-approval characteristics of research projects, including the priority scores assigned at the initial peer review, were important predictors of their future performance. To do so, we studied all single-centre projects financed by the Spanish Health Research Fund (Fondo de Investigación Sanitaria) between 1988 and 1994, and completed before 1996, and quantified their performance in terms of scientific production.

Methods

The Spanish Health Research Fund, a public funding agency reporting to the Ministry of Health, is the major source of research funding in the biomedical field in Spain.^{10,11} Competition for these funds is open to all non-for-profit institutions working in health related fields, mainly within the National Health Service. The funds awarded cover disposable materials, equipment (except for large equipment or infrastructure costs, which are requested on a separate call), travel expenses and others. The salaries of project leaders and staff investigators are not covered by the Fund, but some projects can support a research fellow or part-time personnel to perform technical tasks.

Grant review in the Spanish Health Research Fund

In 1988 the Fund established a system of external peers to review research proposals.¹⁰ Peers were organised in commissions covering specific research areas (i.e., cancer, neurosciences, cardiovascular disease, etc.). Since 1988, the number of commissions has increased from 16 to 21, with each commission including between 8 and 15 peers. Peers serve in the commissions for four years. After each annual call for proposals, projects received at the Fund were assigned to two of the commissions according to the topic of the project. The first commission was topic-specific, while the

second commission included experts from related fields who reviewed the proposal from a complementary perspective (i.e., methodological review). Both commissions, however, performed a full review of the proposal and both evaluations were equally considered for a final decision.

In each commission, a co-ordinator assigned the project to the reviewer with closest knowledge of the topic, who evaluated the proposal and presented a report to the commission. For each project, the reviewer had to complete a form with 12 open-ended questions judging the capacity of the investigators, the objectives, the methodology, the relevance and the appropriateness of the proposed budget, and a final global report. After panel discussion in the commission, each project was classified as excellent/good (exceptional project or project with good reviews in all aspects of the application; high priority), acceptable (project with some limitations; medium priority), or questionable/rejected (project which should not be funded; low priority). Inter-commission discrepancies were resolved in panel discussion by the Scientific Council of the Fund, composed by the co-ordinators of all the commissions and by external advisors. Funding has been available for projects considered of high priority by both commissions, and for some projects considered of medium priority by both commissions or after review by the Council. Projects were funded for a maximum of three years, with yearly renewals after submission of interim progress reports.

Study design

All single-centre projects funded between 1988-1994 and completed prior to 1996 were assessed in a retrospective cohort design, with prospective and standardised assessment of performance. For each project, the main characteristics of the grant application (year of start, duration, budget awarded and presence of a financed research fellow), and the pre-approval priority scores assigned by the peer reviewers were retrieved from the Fund's registry. Information on financed research fellows and scores given by the complementary reviewers were missing for projects funded during the 1988-1990 period due to administrative reasons.

In April 1996, all documents available at the Fund for each project (including the grant application, the yearly and final reports, and published papers originated by the project) were reviewed individually by 324 reviewers using a form designed to determine the performance and the quality of completed projects. These reviewers were unaware of the objectives of the study. The research level of the project (basic, clinical, epidemiological, or health services research as non-exclusive categories) was recorded. Performance was assessed in two different ways. First, projects were classified as

productive or non-productive (projects which generated no original publications). Second, reviewers evaluated the scientific production of each project in terms of type, number, and quality of the original publications and patents, and scored the project from 1 (minimum) to 10 (maximum) in terms of scientific performance. While the classification of projects as productive vs. non-productive was more objective, the performance score took also into account the quality and relevance of the publications.

To assess the reproducibility of the evaluations, a random sample of 40 projects were each reassessed independently by 3 peer reviewers. The inter-reviewer kappa coefficient for evaluating productivity was 0.68, and the intra-class correlation coefficient for performance scores was 0.77.

Statistical methods

Logistic and linear regression models were used to assess the predictors of productive projects and of performance scores, respectively.¹² Non-parametric logistic regression was used to explore a non-linear relationship of total budget awarded with the likelihood of production.¹³ Since the study included only funded projects, and since the results of the complementary reviewers were available only after 1990, the cross-classification of priority scores (high, medium, or low) assigned by both reviewers resulted in some cells with few projects and unstable results. Thus, for predictive analysis, projects with medium or low priority were classified in a single category. Since more recent projects may still be publishing a their results, year of start was included in multivariate models as a stratification factor, but the association of year of start with performance endpoints is not presented. The budget of the projects was updated to 1995 currency units using the yearly official discount rate in Spain. Statistical analysis was performed using the SAS and S-PLUS packages.^{14,15}

Results

Table 1 shows the characteristics of the 2744 single-centre research projects analysed. Most projects had either a clinical research (53%) or a basic research component (39%), while epidemiological and health services research components were limited to a small number of studies (16 and 10%, respectively). The mean budget was 33,340 US \$, with significant differences among one-year (17,110 US \$), two-year (26,798 US \$) and three-year projects (65,684 US \$; $P < 0.001$). After adjusting for duration, the budget of projects with a basic research component was 44% higher (95%

confidence interval [CI] 37 to 52%). The distributions of the pre-approval priority scores given by the topic-specific and complementary reviewers are also presented in Table 1. The percentages of projects considered of high priority by both, one or neither of the reviewers were 33, 41 and 25%, respectively.

Table 1

Description of single-centre research projects financed by the Spanish Health Research Fund, 1988-1994.

	Number	Percent
Total number of financed projects	2744	
Research level*		
Basic	1068	39
Clinical	1458	53
Epidemiology / public health	440	16
Health services	266	10
Year of onset		
1988	411	15
1989	376	14
1990	536	20
1991	367	13
1992	452	17
1993	396	14
1994	206	8
Duration		
1 year	1045	38
2 years	977	36
3 years	722	26
Total budget awarded**		
< 10,000\$	551	20
10,000\$ - 20,000\$	760	28
20,000\$ - 40,000\$	695	25
40,000\$ - 80,000\$	497	18
> 80,000\$	241	9
Financed research fellow***	189	18
Priority assigned by topic-specific reviewer		
Low	244	10
Medium	878	36
High	1311	54
Priority assigned by complementary reviewer***		
Low	201	15
Medium	468	34
High	696	51

* Non-exclusive categories.

** Funds awarded were updated to 1995 US\$ using the yearly official discount rate in Spain.

*** Information unavailable for projects funded during the 1988-1990 period. Information on the presence of a financed research fellow and on the priority assigned by the complementary reviewer is available for 1069 and 1365 projects, respectively.

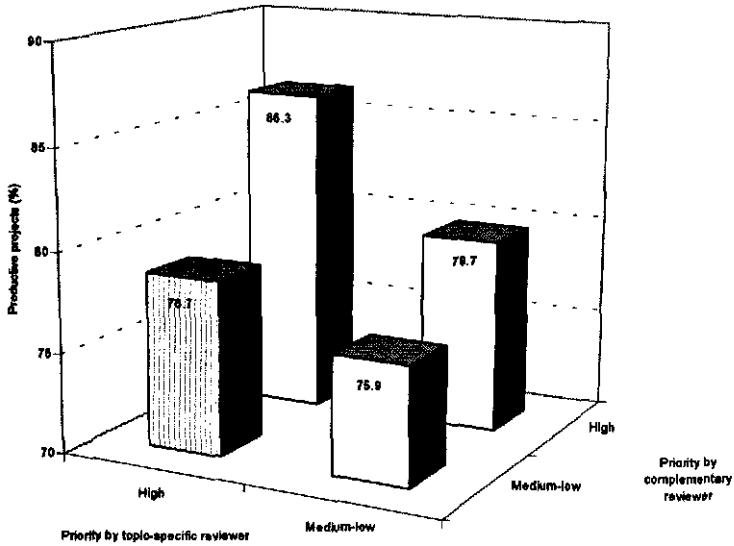


Fig. 1. Percentage of productive projects by priority assigned at initial peer review

The overall proportion of productive projects was 81%. The priority given by the topic-specific and the complementary reviewers were significant predictors of production (Fig. 1 and Table 2). When both, one or neither of the reviewers considered a project of high priority, the percentages of productive projects were 86, 79 and 76% ($P = 0.002$). The odds ratios (OR) of being productive for high-priority vs. medium-low priority projects were 1.68 (95% CI 1.35 – 2.10) and 1.55 (95% CI 1.16 – 2.08) according to the topic-specific and the complementary reviewer, respectively. The predictive ability of peer review results was attenuated, but persisted after adjusting for other project characteristics (Table 2).

A strong positive association between total budget awarded and production was evident from the data (Table 3 and Fig. 2). On average, for each 10,000 US \$ of budget increase, the odds of production increased by 24% (95% CI 17 to 30%). Other significant determinants of production were the presence of a financed research fellow (OR = 1.76; 95% CI 1.12 – 2.77), the presence of a basic research component, and the duration of the project (Table 3).

Table 2
Initial peer review and production for projects financed by the Spanish Health Research Fund, 1988-1994

	Unadjusted		Adjusted*	
	OR	(95% CI)	OR	(95% CI)
Priority assigned by topic-specific reviewer**				
Medium-low	1	(reference)	1	(reference)
High	1.68	(1.35 – 2.10)	1.38	(1.09 – 1.74)
Priority assigned by complementary reviewer**				
Medium-low	1	(reference)	1	(reference)
High	1.55	(1.16 – 2.08)	1.33	(0.97 – 1.82)

* Adjusted for research level, duration, total budget awarded and year of onset.

** Odds ratio of being productive with respect to the reference category.

OR = odds ratio; CI = confidence interval.

Table 3
Predictors of productive projects financed by the Spanish Health Research Fund, 1988-1994

Predictor	Unadjusted		Adjusted*	
	OR	(95% CI)	OR	(95% CI)
Research level**				
Basic	2.71	(2.01 – 3.66)	2.29	(1.68 – 3.12)
Clinical	1.03	(0.78 – 1.36)	1.06	(0.80 – 1.41)
Epidemiology / public health	0.93	(0.69 – 1.26)	0.96	(0.70 – 1.31)
Health services	1.07	(0.73 – 1.56)	1.23	(0.83 – 1.83)
Duration***				
1 year	1	(reference)	1	(reference)
2 years	1.04	(0.83 – 1.31)	1.09	(0.83 – 1.42)
3 years	3.03	(2.23 – 4.11)	1.88	(1.26 – 2.80)
Total budget awarded (US \$)***				
<10,000 -	1	(reference)	1	(reference)
10,000 – 20,000	1.05	(0.79 – 1.40)	0.92	(0.68 – 1.24)
20,000 – 40,000	1.54	(1.14 – 2.08)	1.14	(0.82 – 1.58)
40,000 – 80,000	2.54	(1.78 – 3.61)	1.44	(0.95 – 2.19)
> 80,000	10.05	(4.80 – 21.03)	4.75	(2.13 – 10.61)

* Adjusted for the remaining covariates in the table and for year of onset.

** Non-exclusive categories. Odds ratios indicate risk of being productive comparing projects with and without each component.

*** Odds ratio of being productive with respect to the reference category.

OR = odds ratio; CI = confidence interval.

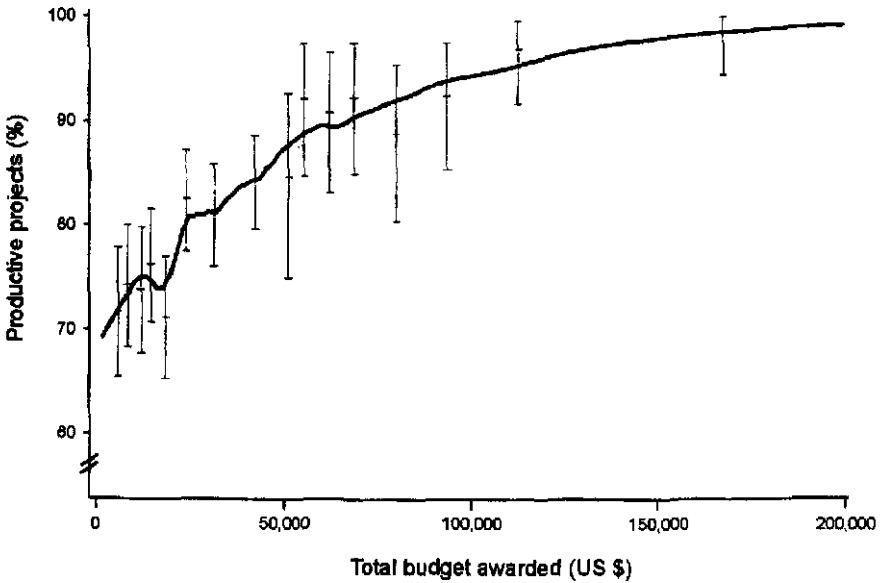


Fig. 2. Percentage of productive projects by total budget awarded. The trend was estimated using non-parametric logistic regression (LOESS with a 30% bandwidth); vertical bars represent the 95% confidence interval for the proportion of productive projects within each decile of budget awarded up to the 80th percentile and for each 2.5% of the projects above the 80th percentile

The mean performance score of the 2744 projects was 5.0 (standard deviation 2.7). The average improvements in performance scores for high-priority vs. medium-low priority projects were 1.03 (95% CI 0.81 – 1.26) and 0.54 (95% CI 0.23 – 0.84) when high-priority was assigned by the topic specific or the complementary reviewer, respectively (both $P < 0.001$). After adjusting for research level, duration, total budget awarded and year of onset, the corresponding average improvements were 0.72 (95% CI 0.50 – 0.94) and 0.26 (95% CI -0.04 – 0.55). Projects with a basic research component, three-year projects and projects with a higher budget had also significantly higher performance scores, even after adjusting for other project characteristics (Table 4). Finally, the funding of a research fellow increased performance significantly (average improvement 1.08; 95% CI 0.63 – 1.52).

Table 4
Predictors of performance score for projects financed by the Spanish Health Research Fund, 1988-1994

Predictor	Unadjusted Improvement	(95% CI)	Adjusted* Improvement	(95% CI)
Research level**				
Basic	0.89	(0.62 - 1.17)	0.57	(0.29 - 0.84)
Clinical	-0.35	(-0.61 - -0.10)	-0.31	(-0.56 - -0.05)
Epidemiology / public health	-0.14	(-0.46 - 0.17)	-0.10	(-0.41 - 0.20)
Health services	-0.25	(-0.64 - 0.14)	-0.15	(-0.53 - 0.23)
Duration***				
1 year	0	(reference)	0	(reference)
2 years	0.12	(-0.13 - 0.37)	0.15	(-0.12 - 0.42)
3 years	1.17	(0.91 - 1.44)	0.51	(0.15 - 0.86)
Total budget awarded (US \$)***				
< 10,000	0	(reference)	0	(reference)
10,000 - 20,000	0.21	(-0.10 - 0.51)	0.05	(-0.27 - 0.37)
20,000 - 40,000	0.65	(0.34 - 0.96)	0.35	(0.02 - 0.69)
40,000 - 80,000	1.43	(1.10 - 1.77)	0.91	(0.52 - 1.31)
> 80,000	2.40	(1.99 - 2.81)	1.69	(1.19 - 2.19)

* Adjusted for the remaining covariates in the table and for year of onset.

** Non-exclusive categories. Improvement in performance comparing projects with and without each component.

*** Improvement in performance with respect to the reference category.

CI = confidence interval.

Discussion

Our study shows the ability of peer review to predict the future performance of research proposals and identifies certain characteristics of successful projects in our context. These results are strengthened by the inclusion of all single-centre projects funded throughout the study period, which minimises the risk of selection bias, and by the prospective and simultaneous evaluation of performance. The consistency of our findings when using different measures of performance and the lack of other empirical data on the long-term predictive ability of pre-approval characteristics of funded projects⁹ also adds to the relevance of our findings.

Although only financed projects were evaluated, there were significant differences in performance depending on the priority assigned by the initial peer review, thus providing empirical support for peer review in assessing projects. In the Spanish Health Research Fund, peer review was based on usually accepted criteria of scientific merit,⁶⁻⁸

evaluating the curriculum vitae of the investigators, the originality and clearness of the objectives, the methodological appropriateness, the socio-scientific relevance and the adequacy of the budget, but specific instructions to standardise reviews were not given to the peers. Similarly, the final decision by the scientific council on projects with discrepant reviews depended on panel discussion and was not standardised. It is likely that if the review process had been more uniform, the final scores would predict even better the outcome of the projects. Research on project peer review has been scarce, and focused on the reliability and interpretation of scoring systems or on the composition of peer review commissions.^{3-5,9,16-20} Our results highlight the need to link the outcome of funded projects to the peer review process at funding agencies.

The duration of a project, its total budget, and the presence of a financed research fellow were related to one another and to the performance of the project. Higher budget projects were usually granted to established investigators, with a record of successful completion of previous projects and a high quality scientific production (a manifestation of the Matthew effect in science^{21,22}). It is also likely that peer reviewers were more critical when reviewing high budget projects. In the Spanish Health Research Fund, fellows were financed as part of the study grant only if the project offered training capability, mainly in three-year projects. The funding of full-time research personnel, however, was a strong predictor of the performance of a project.

Some limitations of our study have to be considered. The validity of our results depends on the stability of the evaluation process throughout the study period. During this period, the fund has maintained the same criteria, key staff personnel, and review forms. The main changes were the increase in the number of expert commissions, and the call for 1-year 'seeding' projects during 1988-1990 intended to promote new research groups, but these changes did not modify the evaluation process. Also, this study was restricted to funded projects, which limited the power to study the discriminative ability of peer review. Had the full range of projects been included, it is likely that peer review would appear as an even more important predictor of performance.

The generalisation of these results to the review process of other national or international agencies also deserves comment. These results are based on single-centre projects of up to three years duration. The predictors of the quality and outcomes of larger, multicenter or multinational projects, may be different.^{7,23} In addition, our results can only be generalised to selection processes based on an approximate balance of the review of the project's objectives, methodology, relevance, and feasibility.

Review processes including or based on other criteria, such as Community added value and subsidiarity for European Union's BIOMED Program,⁷ may result in different predictors of performance.

These results have important implications for research policy in health sciences. Firstly, peer review has been criticised as unreliable, arbitrary, and biased towards mainstream science,^{24,25} but so far it is the only selection system that has shown any predictive value.^{2,5,18,26} Secondly, these results indicate that it may be better to promote projects of longer duration, presented by established groups, and to fund them appropriately. Short-term projects or projects with low budgets may underestimate the effort needed to complete the research, and may indicate some inexperience by the research team. Under current budgetary constraints, these results favour the funding of established units,²⁷ but specific policies will be needed to allow for the creation of independent groups, to promote multidisciplinary groups, and to correct current inequalities in the geographical distribution of research funds. Thirdly, our results show the importance of full-time research personnel to complete research projects. This may be particularly important in systems such as the Spanish National Health Service, which lacks an established professional research career, and where professionals are dedicated mainly to clinical practice.

Conclusions

This analysis of the Spanish Health Research Fund data support the use of peer review in proposal selection, but more research is needed to further improve the selection process. Funding agencies should apply the scientific method to their own practice, including the evaluation and monitoring of the reproducibility of reviews, the determination of the characteristics of successful projects, the cost-effectiveness of masking the review process, and the organisational characteristics of the agency that influence the selection and funding of projects.

*

This project was funded in part by the Spanish Health Research Fund (Fondo de Investigación Sanitaria, FIS 96/1803). The authors acknowledge the assistance of M. F. *Guasch*, J. *Medina*, A. *Jiménez*, D. *Pérez*, and the collaboration of the support personnel, the peer reviewers and the Scientific Council of the Spanish Health Research Fund (FIS). Dr. Javier *Damián* gave helpful comments on an earlier version of the manuscript.

References

1. N. J. BIRKETT, The review process for applied-research grant proposals: suggestions for revision, *Canadian Medical Association Journal*, 150 (1994) 1227-1229.
2. R. SMITH, Problems with peer review and alternatives, *British Medical Journal*, 296 (1988) 774-777.
3. S. A. GLANTZ, L. A. BERO, Inappropriate and appropriate selection of peers' in grant review, *The Journal of the American Medical Association*, 272 (1994) 114-116.
4. A. T. EVANS, R. A. MCNUTT, S. W. FLETCHER, R. H. FLETCHER, The characteristics of peer reviewers who produce good-quality reviews, *Journal of General Internal Medicine*, 8 (1993) 422-428.
5. S. WESSELY, Peer review of grant applications: what do we know?, *Lancet*, 352 (1998) 301-305.
6. NATIONAL INSTITUTES OF HEALTH, *NIH peer review of research grant applications*, US Department of Health and Human Services, Public Health Service, National Institutes of Health, Bethesda, 1996.
7. BIOMEDICAL AND HEALTH RESEARCH PROGRAMME, BIOMED 2 (1994-1998), *Guide to the evaluation and the selection of proposals*, European Commission, Directorate General XII for Science, Research and Development, Brussels, 1996, 1-22.
8. E. GUALLAR, J. CONDE, M. A. DE LA CAL, J. M. MARTÍN-MORENO, en nombre del Grupo de Evaluación de la Actividad del Fondo de Investigaciones Sanitarias entre 1988 y 1995, Guía para la evaluación de proyectos de investigación en ciencias de la salud, *Medicina Clínica*, 108 (1997) 460-471.
9. Committee on rating grant applications, National Institutes of Health. Report of the committee on rating grant applications, World Wide Web <http://www.nih.gov/grants/peer/rga.htm>, 1996, 1-32.
10. J. R. RICOY, M. F. GUASCH, L. E. CLAVERÍA, Fondo de Investigación Sanitaria (1988-1995). Una aproximación al análisis de un Programa I+D, Instituto Nacional de la Salud, Madrid, 1996.
11. J. R. RICOY, La financiación de la investigación médica, *Medicina Clínica*, 100 (1993) Suppl. 1, 6-8.
12. P. MCCULLAGH, J. A. NELDER, *Generalized linear models*, Chapman and Hall, London, 1989.
13. T. J. HASTIE, R. J. TIBSHIRANI, *Generalized additive models*, Chapman and Hall, London, 1990.
14. SAS INSTITUTE INC, *SAS language: reference, version 6*, SAS Institute Inc, Cary, NC, 1990.
15. STATISTICAL SCIENCE INC, *S-PLUS reference manual, version 3.2*, MathSoft Inc, Seattle, 1993.
16. R. SMITH, Promoting research into peer review [editorial] (see comments), *British Medical Journal*, 309 (1994) 143-144.
17. C. A. OLSSON, W. A. KENNEDY, Urology peer review at the National Institutes of Health, *Journal of Urology*, 154 (1995) 1866-1869.
18. H. J. SUTHERLAND, E. M. MESLIN, R. DA CUNHA, J. E. TILL, Judging clinical research questions: what criteria are used?, *Social Science and Medicine*, 37 (1993) 1427-1430.
19. K. J. VENER, E. J. FEUER, L. GORELIC, A statistical model validating triage for the peer review process: keeping the competitive applications in the review pipeline, *FASEB Journal*, 7 (1993) 1312-1319.
20. F. J. MEDMAN, R. A. DE MELKER, The extent of inter- and intrareviewer agreement on the classification and assessment of designs of single-practice research, *Family Practice*, 12 (1995) 93-97.
21. R. K. MERTON, The Matthew effect in science, *Science*, 159 (1968) 36-63.
22. R. K. MERTON, The Matthew effect in science. II. Cumulative advantages and the symbolism of intellectual property, *Isis*, 79 (1988) 606-623.

23. NATIONAL INSTITUTES OF HEALTH, *PHS grants policy statement*, US Department of Health and Human Services, Public Health Service, Washington, 1987.
24. D. F. HORROBIN, *Referees and research administrators: barriers to scientific research?*, *British Medical Journal*, 2 (1974) 216-218.
25. D. F. HORROBIN, *Peer review of grant applications: a harbinger of mediocrity in clinical research*, *Lancet*, 348 (1996) 1293-1295.
26. J. P. KASSIRER, E. W. CAMPION, *Peer review: crude and understudied, but indispensable*, *The Journal of the American Medical Association*, 272 (1994) 96-97.
27. A. ABBOTT, *Funding cuts put pressure on peer review*, *Nature*, 383 (1996) 567.

Received September 8, 1999.

Address for correspondence:

L. ERIK CLAVERÍA
Sección de Neurología, Hospital General de Segovia,
Carretera de Avila s/n, 40002 Segovia (Spain)
E-mail: lclaverias@medynet.com