

A practical instrument for assessing the quality of pharmacoeconomic studies: how it was developed

Thomas De Rijdt; Professor Ludo Willems; Professor Steven Simoens; Professor Arnold G Vulto

Pharmacoeconomics is playing an increasingly important role in the daily practice of hospital pharmacists. However, a significant proportion of pharmacoeconomic reports have methodological shortcomings, therefore, a quick and easy-to-use scorecard would help to assess their quality. Such a scorecard has been developed by an international expert panel.

Nowadays, economic evaluations play an important role in the good governance of health care and are part of an overall trend towards evidence-based decision-making. Pharmacists are increasingly confronted with economic evaluations in order to decide upon drug inclusion in formularies, to get necessary funding for health technologies or to prove the added value of their services. However, understanding and applying techniques of economic evaluation requires expert training and education. The relatively young age of pharmacoeconomy as a discipline and the complexity of the science have led to a situation whereupon many economic evaluations published in the international literature suffer from shortcomings [1]. A review of the literature revealed that only a few tools for assessing the quality of pharmacoeconomic evaluations have been published. However, most of these suffer from one or more limitations, for instance, they can only discriminate between very high and very low quality papers; the tool itself has not been validated; they do not result in a single quantitative score representing the appraisal, and last, but not least, a rather high level of expertise is required from the assessor. To fulfil the need of a tool for everyday practice, an international panel of experts developed a validated easy-to-use scorecard that can be used by pharmacists with only a basic understanding of pharmacoeconomy. The panel consisted of pharmacists and pharmacoeconomists from all over Europe.

Development of the scorecard

To begin the development process all of the recommendations and checklists concerning methodological quality and transferability of results and conclusions, which were mentioned in articles retrieved from the literature review, were gathered in a questionnaire. This questionnaire would form the basis of the assessment scorecard. These were then discussed by the expert panel with regards to their relevance.



This qualitative checklist focused on two main issues: whether the methodology is appropriate and the validity of the results. The method, developed by Gonzalez-Perez, for assessing the quality of randomly sampled papers by scoring all questions of Drummond's questionnaire as 'yes = 1', 'no = 0', 'don't know = 0.5' or 'not applicable' was used to quantify the answers [2, 3]. After preliminary discussions it was obvious that most of the questions were too complex to be included in the scorecard and further simplification and clarification was necessary so as to make it more accessi-

ble for all pharmacists. Since not all the questions have the same impact on quality outcome a weight factor for each answer was also introduced.

In a further round of discussions the preliminary version of the scorecard's questionnaire was tested on two publications by the multidisciplinary expert panel. They were each asked to assign a weight, on a 5-point Likert scale, to each question (see Table 1). The average values of the individual Likert scores were discussed in a plenary session until a consensus resulting in a factor for each question was reached. The same method of validation was used on the content, phrasing and classification of the questions and subsections of the scorecard. In order to guarantee an easy-to-use scorecard, which makes it possible to score a pharmacoeconomic paper in less than 45 minutes, the questions and categories were revised and the number reduced significantly.

An important observation during the development of the scorecard was the bias introduced by letting the user choose the 'Don't know' option. This allows pharmacists less familiar with pharmacoeconomics to use it as an easy way out, instead of forcing them to think about the question and make a choice/decision. Moreover, in a sensitivity analysis this would lead to a 50% score when blindly applying this answer to each question. Therefore, the option 'Don't know' or a possible replacement by 'Partially met' were deleted from the

Table 1: 5-point Likert scale

The weights are coded on a 5-point Likert scale (weight 5 is the highest quotation).	
5	Indispensable, without this a paper makes no sense.
4	Important to have; if lacking, interpretation of the study will be hampered or invalidated.
3	Nice to have, it will add to the quality of the study.
2	Nice to have, but without the study is still valid, but generalisability may be hampered.
1	Leading to perfection, but will not change the outcome of the study.

scorecard forcing the assessor to decide upon the best fitting answer. Similarly, the option ‘not applicable’ was only added to the questions or sections when obviously relevant.

At this stage the scorecard generated a numeric score which represented the assessor’s appraisal of the quality of the paper. In order to introduce both a pass-fail score and reference markers the expert panel arbitrarily defined the following grading rules:

- A paper should score at least 60% on each individual category.
- The first four subtotal scores are calculated as the averages of the weighted individual answers in that section.
- The average of the subtotal scores of each category gives a quantitative appraisal of the methodological quality and the validity of the results, whilst giving an equal weight to all four categories. This results in an absolute number that is not related to a setting or a country.
- A paper is deemed to have an acceptable quality if it obtains a score of at least 60% across all four categories.
- The fifth category, questioning the transferability of economic evaluations to the pharmacist’s own setting, consists of questions that were revised and weighted by the same method as the section on methodology and validity [4]. Scoring in this section may result in different scores for different assessors because answers depend on the applicability of the publication to the pharmacist’s area of practice. To pass the transferability test the report should, on this part also, meet a score of at least 60%.

Table 2: Final score after multiplication of the average score of category 1–4 (quality assessment) with category 5 (applicability)

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.1	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
0.2	0.02	0.04	0.06	0.08	0.1	0.12	0.14	0.16	0.18	0.2
0.3	0.03	0.06	0.09	0.12	0.15	0.18	0.21	0.24	0.27	0.3
0.4	0.04	0.08	0.12	0.16	0.2	0.24	0.28	0.32	0.36	0.4
0.5	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
0.6	0.06	0.12	0.18	0.24	0.3	0.36	0.42	0.48	0.54	0.6
0.7	0.07	0.14	0.21	0.28	0.35	0.42	0.49	0.56	0.63	0.7
0.8	0.08	0.16	0.24	0.32	0.4	0.48	0.56	0.64	0.72	0.8
0.9	0.09	0.18	0.27	0.36	0.45	0.54	0.63	0.72	0.81	0.9
1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

By multiplying the scores achieved on both quality (categories 1 to 4) and transferability (category 5), the resulting score reflects an overall assessment of an economic evaluation. This takes into account the absolute aspect of methodological quality (independent of setting) and the relative aspect of transferability (depending on the area of practice). This calculated value is an appraisal of the overall quality of the article and its usefulness to the appraiser’s setting. Theoretically, this score is between 0 and 1. Upon visualising the possible results in a table and graphic it becomes clear that, due to the multiplication, the result of sufficient subscores will be a rather low value. By arbitrarily setting the limit to 0.3 it is possible to separate the poor from the better reports. Acceptable reports have a score higher than 0.30, whilst very good reports score 0.60 or above (see Table 2).

Subsequent discussions have led to a further version of the scorecard’s questionnaire, which was agreed by the core team of experts after a second review, and is now ready to be tested by a larger panel of users.

Validation of the scorecard

Working together with the members of the expert panel, the adapted version of the scorecard was sent out to be tested by different hospital pharmacists across Europe. They were each asked to read one of two papers [5, 6] and to score them using the scorecard provided to them as a spreadsheet. At the same time a small survey about their experience had to be completed.

A commonly used method for validating a questionnaire such as that in the scorecard is by performing an analysis of the interrater reliability of two raters. A similar technique, called Fleiss’ kappa, suitable for a multirater environment, resulted in a surprisingly low value of 0.5–0.6 for both articles. There are several reasons for this low interrater reliability score. In the survey the assessors stated they easily understood the questions in the scorecard, but that they lacked sufficient knowledge of pharmacoeconomics to fully understand the paper and some of the methods and terminology used. Another common remark was the relatively complex nature of the pharmacoeconomic papers being assessed, again this is related to the assessor’s inexperience of pharmacoeconomics.

To ensure that the low kappa values were associated with an insufficient basic knowledge, two additional simple tests were conducted:

- Two experienced hospital pharmacists familiar with pharmacoeconomics assessed one of the highly rated articles [5] with the unmodified questionnaire.
- Each one of the original articles with a greater focus on clinical practice [7], used in an earlier stage of the development of the questionnaire, was assessed by two experienced hospital pharmacists with no specific interest or education in pharmacoeconomics using the unmodified questionnaire.

A high score on the interrater reliability test is an indication that the scorecard was working well and that the kappa's mentioned previously were low because of the high difficulty level of the subject and terminology used in the paper and due to the absence of sufficient pharmacoeconomic knowledge. Upon completion both additional tests resulted in a high interrater reliability (Cohen's kappa 0.9–1).

Based on the feedback from the assessors the questionnaire was fine-tuned further by adding supplementary explanations to some of the questions. A final validation of the scorecard will be performed by applying the AGREE instrument [8].

Analysis of the survey

Analysis of the survey which accompanied the scorecard showed that the average time spent completing the list of the scorecard's questions for the first time is about 43–53 minutes (range from 10–90 minutes). However, once an assessor is familiar with the scorecard it is believed that this time frame will reduce significantly. The assessors were from Belgium (26%), France (6%), The Netherlands (41%), Italy (9%) and Spain (18%). Responses from the UK were not received. All of the assessors were hospital pharmacists, most with little experience of pharmacoeconomics in daily practice; 80% were connected to an academic facility, 10% to a large hospital and the remaining 10% to a community hospital. The most common remarks noted in the survey were regarding the lack of basic knowledge on pharmacoeconomics, the high level of difficulty of the assessed papers (possibly not the kind of papers a hospital pharmacist is confronted with regularly) and the request to explain terms such as cost-utility, willingness-to-pay, valuation of costs, etc. About 17% would use the scorecard without adaption, 57% requested revision, mainly a clarification, before using it in daily practice.

Conclusion

While the number and the impact of economic evaluations in health care are increasing, the methodological quality

and transferability of scientific papers is lagging behind. A quick, simple and useful scorecard can help a less experienced practitioner or decision maker in selecting high quality, transferable economic evaluations, and thus improve decision-making and implementation in practice. To support pharmacists in applying the outcome of pharmacoeconomy studies, a standardised evaluation is mandatory and the proposed scorecard is a possible way to fulfil this need.

Our experience has also highlighted the fact that many hospital pharmacists are not familiar with the principles of pharmacoeconomics. This is not surprising when looking at the current curricula and training for hospital pharmacy. Since the modern hospital pharmacist is a major stakeholder in the evaluation and purchasing of medicines for the hospital it would be strongly advisable to train them in pharmacoeconomics during undergraduate education.

Acknowledgements

The authors wish to acknowledge the work and support of the members of the expert panel: Dr Isabelle Borget, Dr Ana Ortega Eslava, Dr Paul McNamee, Dr Andrea Messori and Professor Frans Rutten, who were instrumental in testing the scorecard and providing the necessary feedback. We would like to thank Mr Jan-Dietert Brugma for his invaluable help during the process.

We are grateful to Amgen Europe GmbH for their trust in supporting this project with an unrestricted educational grant.

Authors

Thomas De Rijdt
Hospital Pharmacy
University Hospitals Leuven
49 Herestraat
B-3000 Leuven, Belgium
thomas.derijdt@uz.kuleuven.ac.be

Professor Ludo Willems
Hospital Pharmacy
University Hospitals Leuven
49 Herestraat
B-3000 Leuven, Belgium
ludo.willems@uz.kuleuven.ac.be

Professor Steven Simoens
Research Centre for Pharmaceutical
Care and Pharmaco-economics
Katholieke Universiteit Leuven
O&N 2 PO Box 521
49 Herestraat
B-3000 Leuven, Belgium
steven.simoens@pharm.kuleuven.be

Professor Arnold G Vulto
Hospital Pharmacy
Erasmus University Medical Center
PO Box 2040
NL 3000 CA Rotterdam, The Netherlands
a.vulto@erasmusmc.nl

References

1. De Rijdt T, Willems L, Simoens S. Economic effects of clinical pharmacy interventions: a literature review. *Am J Health-Syst Pharm.* 2008;65:1161-72.
2. Gonzalez-Perez JG. Developing a scoring system to quality assess economic evaluations. *Eur J Health Econom.* 2002;3:131-6.
3. Drummond M, Sculpher M, Torrance G, et al. *Methods for the economic evaluation of health care programmes.* 3rd edition. 2005 Oxford University Press. ISBN 019-852945-7.
4. Sculpher M, Pang F, Manca A, et al. Generalisability in economic evaluation studies in healthcare: a review and case studies. *Health Technology Assessment.* 2004;8(49) ISSN 1366-5278.
5. Mernagh P, Campbell S, Dietlein M, et al. Cost-effectiveness of using recombinant human TSH prior to radioiodine ablation for thyroid cancer, compared with treating patients in a hypothyroid state: the German perspective. *Eur J Endocrinology.* 2006;155:405-14.
6. Machnicki G, Ricci JF, Brennan DC, et al. Economic impact and long-term graft outcomes of mycophenolate mofetil dosage modifications following gastrointestinal complications in renal transplant recipients. *Pharmacoeconomics.* 2008;26(11):951-67.
7. Ramsey SD, Shimei L, Boer R, et al. Cost-effectiveness of primary versus secondary prophylaxis with pegfilgrastim in women with early-stage breast cancer receiving chemotherapy. *Value in Health.* 2009;12(2):217-25.
8. Appraisal of guidelines for research & evaluations (AGREE) instrument. The AGREE collaboration September 2001. ISBN 1 8981 8321 X.