MOLECULAR ONCOLOGY 4 (2010) 12–18

available at www.sciencedirect.com

www.elsevier.com/locate/molonc

Research Management

Benchmarking biology research organizations using a new, dedicated tool

Willem H. van Harten[a,b,*], Leonard van Bokhorst[a], Henri G.A.M. van Luenen[a]

[a]The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands
[b]School of Management and Governance, University of Twente, Enschede, The Netherlands

ARTICLE INFO

Article history:
Received 2 February 2009
Received in revised form
14 September 2009
Accepted 15 September 2009
Available online 18 September 2009

Keywords:
Organization
Research
Benchmarking
Evaluation

ABSTRACT

International competition forces fundamental research organizations to assess their relative performance. We present a benchmark tool for scientific research organizations where, contrary to existing models, the group leader is placed in a central position within the organization. We used it in a pilot benchmark study involving six research institutions. Our study shows that data collection and data comparison based on this new tool can be achieved. It proved possible to compare relative performance and organizational characteristics and to generate suggestions for improvement for most participants. However, strict definitions of the parameters used for the benchmark and a thorough insight into the organization of each of the benchmark partners is required to produce comparable data and draw firm conclusions.

1. Introduction

International competition in combination with increasing financial restraints and accountability forces academic research organizations to search for means to assess their own performance in a more comprehensive way and to look for ways to improve their internal operations.

There are various definitions of benchmarking, but in short it can be described as ''the process of comparing business processes and performance metrics with good- or best practice organizations''. Systematic benchmarking is used in various business and service domains, including health care, to achieve this purpose. However, we found very little evidence of its application in the domain of fundamental research. Benchmarking of academic biomedical research organizations has usually been limited to a quantitative measurement of a limited number of output parameters. An important system for comparing performance of research organizations is the use of peer review and bibliographic scores such as impact factor and citation index. This, however, usually does not easily provide information on organizational aspects other than the quality and quantity of the scientific publications of the group leader and his/her research group. With the aim to compare research institutes on different levels

* *Corresponding author*. The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. Tel.: +31 20 512 2860; fax: +31 20 6691449.
E-mail address: w.v.harten@nki.nl (W.H. van Harten).
1574-7891/$ – see front matter © 2009 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.molonc.2009.09.002

and at different aspects we embarked on a study with the following objectives:

- To identify an appropriate framework to benchmark similar fundamental research organizations;
- To perform a benchmark with a number of institutions to test data retrieval and comparability;
- To analyze the differences among participants;
- To suggest improvements for participating research organizations.

In this paper we report on a pilot using a benchmark instrument designed to meet the above objectives. We focused our study on molecular/cellular biology research organizations with a basic (fundamental) research program. This research program should not be embedded in a larger research program or organization (for instance at a university) in order to allow accurate identification of general resources to the molecular/cellular research program. These restrictions are necessary because different types of research organizations will pose different requirements on a benchmark model.

The project was initiated by The Netherlands Cancer Institute (Het Nederlands Kanker Instituut, NKI) which is a non-profit, independent research organization. Although the institute is closely associated with the Antoni van Leeuwenhoek Hospital (AVL) we only focused on the research and did not (at this stage) extend the focus to the Comprehensive Cancer Center construct. The research program of the NKI focuses on the etiology and treatment of cancer, including both fundamental and translational research.

1.1. *Literature survey on benchmarking frameworks for academic biomedical research organizations*

To guarantee comparability and to structure data retrieval, a benchmark framework was needed. Three categories of benchmarking frameworks from the literature were considered for our study:

- General benchmarking frameworks.
- Frameworks designed specifically for benchmarking Research & Development units.
- Frameworks designed specifically for benchmarking research institutes.

For the purposes of the study, a set of criteria was designed to judge the frameworks:

- The framework should be applicable to non-profit/not-for-profit research organizations with a strong academic focus;
- The framework has to take the central position of the group leader in a research organization into account. The research groups, each under the supervision of a group leader, are the core organizational units in most research organizations. Group leaders have a high degree of autonomy;
- The information for the benchmark should be obtainable;
- The framework should take into account international differences such as culture and legislation.

A number of frameworks were identified in business and service literature (i.e. Kennerley and Neely, 2000; Van Lent and Roijmans, 2005) that belong to one of the categories mentioned above. We went on to analyze the following frameworks to determine which would best serve our purposes:

- Financially oriented models like the DuPont model, which have a strong business focus;
- Input-output models, these are mainly focusing on efficient processes, but provide no insight into the factors that affect the (scientific) performance;
- The Balanced Scorecard (Kaplan and Norton, 1992), which would need to be adapted for external conditions and designed for use in for-profit organizations. Due to its aggregated scores some aspects, such as the position of the group leader, cannot be incorporated into this model. Furthermore, it does not take the external environment into account;
- The European Foundation of Excellence (EFQM) model, which although widely used in non-profit organizations has not been designed for benchmarking and uses mostly qualitative data for the output sections;
- The Performance Prism is mainly directed at the degree to which the interests of stakeholders are managed. It has no output orientation and is not widely used;
- The ''Adapted Balanced Scorecard'' was the only example we identified that was adapted for use in Research and Development departments (Kerssens-van Drongelen and Cooke, 1997). This framework had been designed for use in for-profit organizations and contained indicators that had been verified. The disadvantages were that international differences due to national legislations and culture and the position of the group leader could not be incorporated;
- The NIAB framework had been specifically developed to analyze the performance of an agricultural research organization (Visser et al., 2001). This framework included both the Balanced Score Card and Performance Prism, monitoring systems, which have been widely used in business and service organizations to relate strategy and performance. No mention, however, was made of use of this framework in benchmarking, and it did not take the position of the group leader into account. Furthermore, the unilateral focus on strategy was felt to be inappropriate for use in fundamental research organizations.

While all of the frameworks we reviewed met one or more of our criteria, none fitted absolutely with all our specified criteria, as can be seen in Table 1. None of the frameworks was suitable to compare international research institutes and take into account the cultural and legislative differences that exist between countries. Furthermore, every framework required adjustment.

Based on our analysis of available frameworks, we decided not to use standard frameworks but to design a customized framework combining the positive aspects of the frameworks reviewed. For instance, we included the enablers-results framework in the EFQM Excellence Model and extended it by introducing the Input–Output model focusing on the transfor-

Table 1 – Evaluation of the different benchmark frameworks.

	DuPont pyramid	Input–Output model	Balanced Scorecard	EFQM	Performance Prism	Adapted Balanced Scorecard	NIAB
The framework can be applied on non-profit/not-for-profit research organizations	0	+	0	0	0	0	+
The framework can put emphasis on the position of the group leader	–	–	0	0	0	0	0
The information to be used in the framework can be obtained	+	+	0	0	0	+	+
The framework can take into account international variables like culture and legislation	–	–	–	–	–	–	–

+: Suitable; 0: Can be implemented with modified; –: Not suitable

mation processes. Additionally, we included a feedback loop from performance to enablers aimed at identifying the ways organizations want to improve after evaluating their results. We also found that the focus on stakeholders, found in the Performance Prism, could be incorporated to include the central role of the group leader.

1.2. *Conceptual design of a new benchmark framework for academic biomedical research institutions*

Using essential aspects of existing models, we hypothesized that the group leader is the individual who generates costs and results (performance), while being facilitated by research support facilities and staff departments (enabling factors and functions). Additionally, the external environment influences both the organization and group leaders in their effectiveness.

The concept of enablers is taken from the EFQM model, where the tools, assets and structure of an organization, are used to assess its key activities (EFQM brochure, 2003). The indicators used to measure enablers have to be adapted to and focused on research institutes. Performance measures include all results delivered by the group leader (or indirectly through the research group) and the costs generated by research activities. In contrast to frameworks found in literature, we have taken international variables like culture, economic situations and legislation into account, since they are part of the factor ''external conditions''. Such factors are difficult to change by institutions, but nevertheless can have great influence on the strategy of organizations and the operational freedom of group leaders. Identifying such factors helps to explain some of the differences between institutes in different countries. Furthermore, global scientific progress can influence the state-of-art in research and force both group leaders and institutes to continuously evaluate their research. Strategy describes the mission and management principles of an institute. By utilizing such tools the leadership can influence the enablers, for example by changing the reward structure.

The resulting framework with all the related factors can be found in Figure 1. In the appendix some definitions are presented. Using this framework an extensive profile can be made of an institute which can be used for comparison.

2. Methods and material

In the Netherlands there are no other institutes with a similar organizational structure and research focus as the NKI. Furthermore, the NKI is an independent research institute, not part of a university of governmental research organization, which makes the NKI also unique in the Netherlands. This necessitated the need to search for international research institutes resembling the NKI to be used as benchmark partners. In order to allow comparisons of results, the following selection criteria were utilized:

- Active in life sciences. The benchmark partner should be active in the same general field as the target organization, necessitating that they have to be institutes engaged in fundamental research in life sciences. Many organizations qualify for this criterion, but it is still essential for comparison;
- Best-in-class. In his book 'The benchmarking book', Spendolini (1992) describes the process where the first step in the selection of partners is to define the desired state of organizational performance. For this study this would imply that the selected benchmark partners have to be recognized as the internationally, best-in-class research institutes;
- Independent, identifiable institutes. Partners have to be recognized as at least partly independent, identifiable institutes. If such organizations are fully embedded in a university or a hospital, retrieval of data and comparability will be difficult;
- Non-profit/not-for-profit. For-profit institutes have different strategic aims which render comparison with non-profit institutes difficult;
- European. While there are comparable institutes in the United States of America and elsewhere, their external and internal situations differ markedly from European institutes, (in terms of legislation, funding by sponsors, endowments and comprehensive overhead budgets) making comparisons on aspects other than scientific performance difficult;
- Trustworthy. For cooperative benchmarking the sharing of confidential information is essential, making it important for partners to treat such information accordingly.

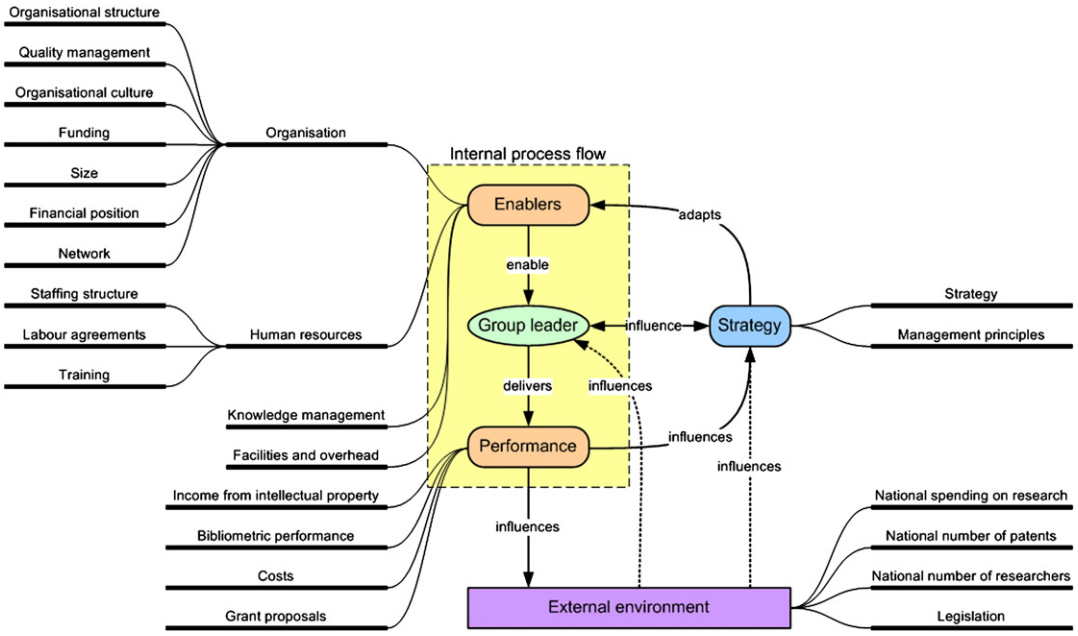

Figure 1 – Benchmarking model for research organizations. Based on existing frameworks a new models has been designed to allow benchmarking of non-profit/not-for-profit research institutions. The parameters are explained in the supplementary data.

For the purposes of our study, 12 organizations that are active in fundamental and preclinical cancer research and generally recognized as best-in-class were selected. This is obviously a subjective element in the benchmark. The institutes were approached by a written request from the Scientific Director of the NKI. Ultimately, six organizations agreed to participate. Reasons for non-participation included lack of interest, time constraints, and other priorities. The collected data was based on annual reports over fiscal years ending in 2006, web sites of the institutes and additional information supplied by each institute. Initially data retrieval was prepared through telephone contact, followed by two of the authors visiting the institutes to assemble and verify information and obtain further information on administrative and cultural issues needed to properly judge the material. Because of the focus on the organization of the institutes and the pilot character of the study, no additional bibliometric studies were performed which would be quite expensive and time consuming. All information was collected in a written profile of each individual institution. The profiles were sent to the institutes for review, factual corrections and approval.

3. Results of a pilot review of six organizations

3.1. *Data retrieval*

Most of the data were obtained within the steps and time frame planned. However, making the data comparable turned out to be difficult. Each institute, often as a result in differences in legislation, used different accounting and reporting principles. Also on other subjects, like the number of employees working in staff departments and research facilities and remuneration, data were difficult to compare or were not available. During site visits face-to-face explanations and review of the data proved extremely useful. Since data collection was the most important step in the benchmark project, sufficient time needs to be allocated to define the data that will be collected and to prepare and perform the data collection.

In our pilot study an extensive dataset was collected, with only a few of the highlights presented in this paper. Since the data are confidential we can only report on results that cannot be traced back to individual institutes.

For purposes of the study we compared the number of researchers (including both faculty and non-faculty) with the overhead costs per researcher. Overhead costs (and other

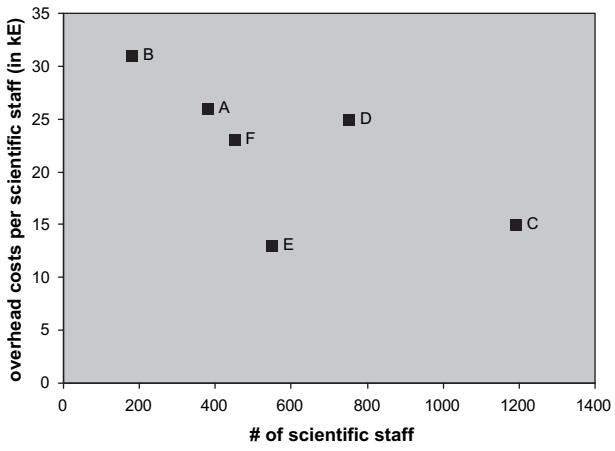

Figure 2 – Number of scientific staff (faculty and non-faculty) plotted against the total overhead costs per scientific staff. The costs are adjusted for purchasing power parity.

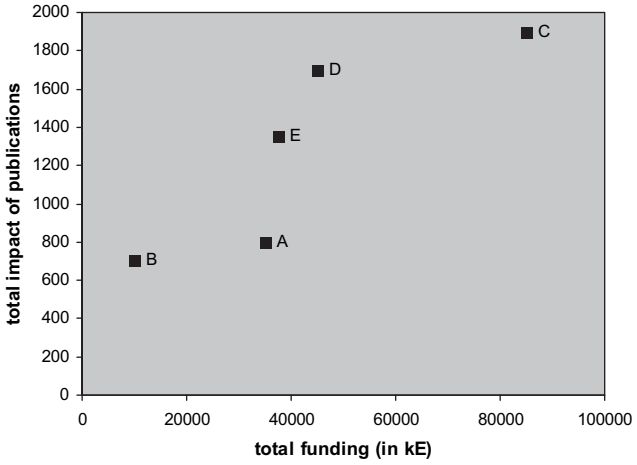

Figure 3 – Total funding for 2004 (adjusted for purchasing power parity) versus the total number of impact point of all publications from 2004.

data) were retrieved from the annual reports or provided additionally on request. Our calculation of the overhead costs was included in the profiles of each partner and feedback on these profiles was incorporated as a means of verifying our interpretation of the data. The costs per country were corrected with the purchasing power parity for international comparison. The purchasing power parity equalizes the purchasing power of different currencies in their home countries (Edison, 1987).

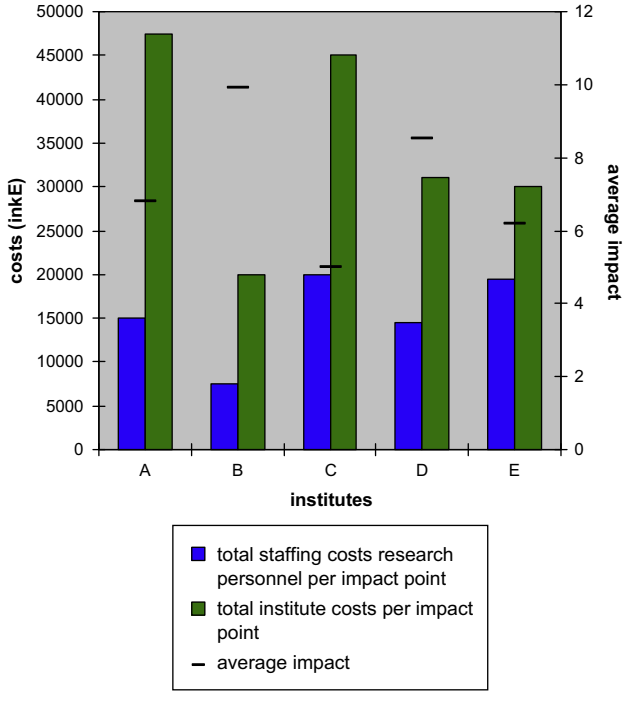

Figure 4 – The relation between costs and impact of publications. The personnel costs of all researchers per institute and the total costs of an institute were divided by the total number of impact points per institute for the year 2004. The costs were corrected for purchasing power parity. The average impact of a publication per institute is also indicated.

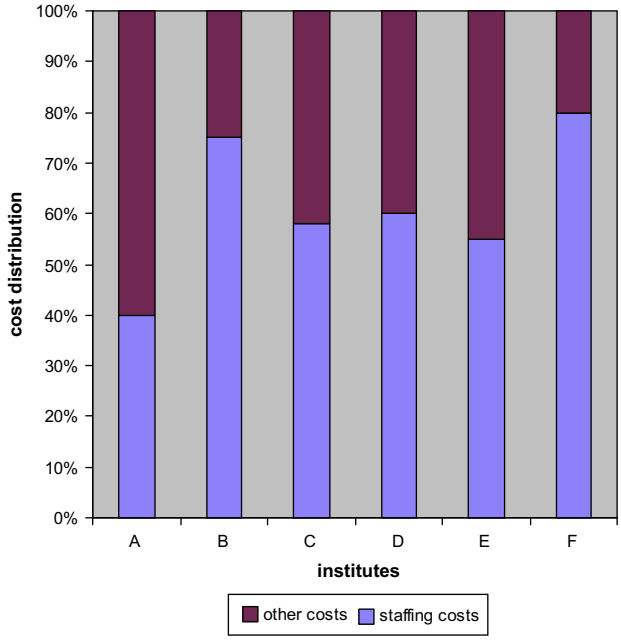

Figure 5 – The costs directly related to research split in personnel costs and other costs (benchfee and investments) for the year 2005 per institute.

Figure 2 shows that, as expected, the overhead costs per researcher drop with an increase in the total number of researchers, indicating an efficiency of scale effects. Of all the organizations considered, institute E showed relatively low overhead costs for the size of the research population; however no specific measurable characteristics could be identified that provided an explanation for the cost efficiency.

Although, there seems to be a relationship between budget size and the total impact of all publications (Figure 3), a more interesting comparison proved to be the cost per impact point (see Figure 4). The personnel costs for all researchers (blue bar) and the total institute costs (green bar) were divided by the total number of impact points of all publications per institute for the year 2004. From this it is apparent that some institutes spent considerably more per impact point than others. However, this does not seem to correlate with the average impact score per publication. In other words, a larger budget does not seem to result in higher impact publications.

These results should obviously be interpreted with care, even though the data have been verified by the partners, because sometimes elements of overhead costs are not always reported as such. The degree to which research activities are outsourced was not compared, but such practices may provide some explanation of the findings that in some organizations total costs are high despite staffing costs being relatively low.

In Figure 5, the relative distribution of expenditures on personnel, consumables and equipment that are directly associated with the performance of research (i.e. overhead costs are not included) produces interesting findings. Analysis of the expenditure shows that different choices are being made by research executives in budget allocation. While some institutes have chosen to invest in state-of-the-art equipment and to supply the group leader with an excellent infrastructure

(like institute A); others choose to invest in personnel (such as institutes B and F). Such choices, however, are not reflected in output: neither the average impact score per publication per institute or the total impact score of all publications per institute (data not shown).

4. Conclusions and discussion

As the benchmarking organizations are active in the same field, the organizations look similar in their organizational structures and cultures independent of their funding structure. The research groups are the central units in the organizational structure and group leaders have a large degree of independence. From our data we could not differentiate the degree of independence of group leaders between the institutes. It was difficult to distinguish clear strategic choices within organizations. Five institutes seemed to have no defined overall strategic policy.

Although some small differences were found in the structure and culture, no firm relationship between organizational structure, culture and performance were identified. While some institutes had extensive centralized research facilities (such as facilities for microscopy and microarray analysis) and others had only a few centralized facilities, no relationship between performance and cost structure could be found.

Some of the institutes outsourced a large fraction of overhead activities such as IT or salary administration. However, outsourcing does not appear to lead to lower total overhead costs, and from this sample it even seems to be less cost effective to out source.

As might have been predicted, large institutes spent relatively less on overheads. This, however, is the only instance where economies of scale seem to apply. The size of a research institute, for example, could not be related to its performance. In one organization with exceptionally low overheads no other explanation could be found other than a consistent management focus on cost reduction.

One of the key items of this benchmark study was to find the drivers of bibliographic performance. It became clear that identifying these drivers was more difficult than expected, and might relate to two issues. First, the small sample size makes statistical analysis impossible. Second bibliographic performance data supplied by the institutes were based on their own annual reports, using different definitions, so that sometimes data were not comparable. Such factors could be overcome by having a bibliometric analysis performed by an independent organization on multiple institutes. In our pilot benchmark, we found no relation between the degree of structural funding (distinct from project funding) and performance. Although these data have to be handled with care due to possible differences in administrative procedures, higher budgets per individual research group do not seem related to either the quality or quantity of output performance.

An area where considerable differences were found between institutes was the composition of the scientific staff, i.e. the percentage of post docs and graduate students. The relative composition, however, could not be related to differences in bibliometric output.

Although the limited number of participating institutes makes it difficult to draw firm conclusions on correlations, the study has provided some interesting comparative insights into research institutes. In future studies, it will be important to identify institutes with comparable research areas while including sufficient subjects.

For our study, a custom-made benchmark tool was preferred over existing tools, primarily because we hypothesized that the central position of the group leader was essential for the type of research institute we aimed to benchmark and none of the existing tools incorporated this aspect. Indeed, the institutes participating in our study did indeed show a central position for their group leaders. Furthermore, none of the existing tools were capable of integrating the external environment into the study.

Since no comparable benchmark studies of research institutes were found in our search, it is impossible to judge whether the framework described in our study performs better than others.

We think that the new tool can be used effectively when comparing international research institutes. As national regulations on workforce, conditions, financing and accounting, differ considerably (especially when US institutions are included), thorough preparation of the data is needed to make them comparable. Most essential to performing a benchmark exercise are clear definitions of the metrics and high quality data, which are also the most difficult to obtain. On-site review by the study authors of the data provided by the participating institutes was essential to improve the interpretation and quality of the data. Further improvement is undoubtedly possible, but this would require one person to spend at least two to three days at each institute. In total, we estimate that it takes a trained analyst with some knowledge of research institutes two weeks to collect and review the data for each institute and make them suitable for comparison within a benchmark study. Such levels of input represent a considerable investment of time for both the investigators and participating institutes.

Molecular biology research organizations can have a different focus, for instance focusing more on mouse models for diseases, high throughput analysis, or chemistry and pharmacology. They can also emphasis more translational research or basic research. These choices should be considered during a benchmark study. For instance, it is known that translational research achieves in general lower impact scores than other areas of study. Such knowledge should be taken into account when considering organizations showing increasing levels of activity in this field, such as comprehensive cancer centers.

Our newly developed benchmark tool seems to incorporate all relevant aspects of the organizations under consideration. However, two aspects, legislation and culture, proved difficult to assess. Legislation was mainly included to find explanations for the observed differences between institutions for different countries. Identifying differences in legislation requires extensive research. Culture can be assessed using existing measurement tools and is likely to be related to performance. The introduction of such an approach would, however, require a representative survey among employees using a validated tool for research organizations, which although available was not feasible within the context of this pilot.

In conclusion, benchmarking biology research organizations proved feasible with the instrument that was specially created for this purpose. Further application in larger series, adding a culture assessment tool, focusing on the position of the group leader and including a management focus, may provide further information on validity and underlying mechanisms related to performance.

Conflict of interest

W.H. van Harten and H.G.A.M. van Luenen are both employed by the Netherlands Cancer Institute.

Acknowledgements

We thank the participating institutes in this benchmark study for their support and open exchange of data.

REFERENCES

Edison, H.J., 1987. Purchasing power parity in the long run: a test of the dollar-pound exchange rate (1890–1978). Journal of Money, Credit and Banking 19, 376–387.

Kaplan, R.S., Norton, D.P., 1992. The balanced scorecard: measures that drive performance. Harvard Business Review, 71–79.

Kennerley, M.P., Neely, A.D., 2000. Performance measurements frameworks – a review. Proceedings of the Second International Conference on Performance Measurement, Cambridge. 291–298.

Kerssens-van Drongelen, I.C., Cooke, A., 1997. Design principles for the development of measurement systems for research and development processes. R&D Management 27, 345–357.

Spendolini, M.J., 1992. The Benchmarking Book. Amacon, New York.

Van Lent, W., Roijmans, D., 2005. The way to benchmarking in oncological centers.

Visser, S., Kressens-van Drongelen, I.C., De Weerd-Nederhof, P., Reeves, J., 2001. Design of a research performance measurement system: the case of NIAB. Creativity and Innovation Management 10, 259–268.