Short Communication

Benchmarking as a means to improve conservation practice

William J. Sutherland and Michael J. S. Peel

Abstract Benchmarking, the comparison of efficiency measures of an organization against those of other organizations, is widely used in industry, medical practice and agriculture as a means of learning where practice can be improved. This could be used by conservationists for routine repeatable activities, such as the treatment of invasive species or the survival rate of transplanted plants. We give three examples of the benefits of cross-site comparisons: grazing management in South Africa, husbandry of captive penguins and management of lagoons for wading birds. Benchmarking, by comparing effectiveness with others, is the initial stage in identifying weaknesses and leads on to learning how to improve through cross-site comparisons, comparisons with better performers or examination of the published literature. We suggest that the term best practice, which is often used as part of benchmarking, is unsuitable as it implies a comparison of all options, which rarely takes place, and it is subject to change as knowledge and techniques develop. An alternative term is current effective practice.

Keywords Benchmarking, conservation, current effective practice, effectiveness, evidence-based conservation, intervention, monitoring

Organizations and individuals are usually seeking to improve practice so that they can become more effective. One of the most widely used approaches for making management decisions is to learn from previous experience (Sutherland et al., 2004). This is reasonably straightforward if the previous use of the intervention was either a complete failure (e.g. the treatment had no impact on the invasive plant) or an absolute successes (e.g. the cheap and trouble-free treatment efficiently eliminated the invasive plant with no side effects or problems). It is then usually reasonably clear whether to continue with the technique in the same manner.

In reality, most interventions are not complete failures or absolute successes. There is typically some intermediate level of success: the invasive plant population, for example, may have been reduced but not eliminated. In such cases the decision as to whether to continue with the management is harder. Practitioners then have to decide whether their management is the best that could be achieved or whether it is realistic to achieve better outcomes.

The challenge is that it is difficult for an individual or organization to assess how well it is performing, whether it is realistic to be more successful and, if so, how this could be achieved. Benchmarking is a standard method for determining effectiveness in both the industrial (Keehley & Abercrombie, 2008) and public sectors (Saul, 2004). The concept originated with Xerox who compared a range of aspects of their performance with that of other organizations, learnt from those with apparently superior efficiency and identified best practices (Camp, 1998). Although comparisons are sometimes with single organizations, benchmarking usually involves a number of units, such as different companies or different hospitals, which calculate and collate performance measures, such as the cost or success rate of activities. The data are collected across units and, typically, the anonymous results are then provided to all so that the organization can compare its effectiveness in relation to a range of others. This provides the opportunity for reflecting on why performance is higher elsewhere and considering the lessons that could be learnt from others. The practice of benchmarking can then lead to identifying good practice: the changes in policies, procedures, activities or operations that lead to superior performance. Benchmarking has become sophisticated, with means of balancing multiple objectives (ten Raa, 2009).

However, although usually attributed to Xerox, benchmarking has been used in agriculture since the late 19th century, with many countries operating schemes in the early 20th century. This entails farms comparing components of profit such as inputs, yield and labour costs as the initial stage to identifying areas where profitability can be improved (Jack, 2009).

Keehley & Abercrombie (2008) define a benchmark as a point of reference and benchmarking as a methodology used to improve performance by finding high-performing organizations and importing their practices to the home organization. An example from medicine is the use of benchmarking in the treatment of cystic fibrosis in the USA. Treatment takes place in 117 specialized centres and, since 1964, the Cystic Fibrosis Foundation has gathered annual measures of mortality rate and lung function from

William J. Sutherland (Corresponding author) Conservation Science Group, Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, UK. E-mail w.sutherland@zoo.cam.ac.uk

Michael J. S. Peel Agricultural Research Council–Animal Production Institute, Nelspruit, South Africa

Received 9 February 2010. Revision requested 14 May 2010. Accepted 1 July 2010.

each; this was initially devised to examine claims by a centre that it had dramatically lower mortality rates than others, which was substantiated (Gawande, 2007). These results are then presented to all centres, initially anonymously but now openly. Each centre then knows where it lies along the approximately normal distribution of success. This is considered a major force in improving standards by showing what is possible and encouraging improvement. Average survival rates have increased dramatically from an average age of three in 1964 to 18 in 1972 and 33 in 2003. Most interestingly, the centres in the top quartile improved most rapidly, perhaps because they have the capacity to learn and change (Gawande, 2007). Here, we give three examples of how comparisons across sites have shown the capacity for improvement and then consider how benchmarking could operate to improve conservation practice.

Grazing management in South Africa Reserve managers in South Africa are faced with the problem that the vegetation state fluctuates widely, largely in response to variation in annual rainfall, grazing intensity and the impact of megaherbivores such as elephants *Loxodonta africana*. It is a challenge to make decisions in response to such variability. The Savanna Ecosystem Dynamics Project covers an area of c. 45,000 ha of the eastern Lowveld of South Africa and includes 520 vegetation sampling sites that are monitored annually and an annual mammal survey conducted by helicopter (Peel et al., 2005). The aim is to help practitioners in 21 different areas make rational management decisions based on objectively researched results. Data on woody plants are collected within 100 × 2-m square quadrats. The measures include tree density, height class and stems per rootstock, canopy cover and elephant impact (both percentage impact and impact on marked, undamaged trees). Along 100-m transects the annual and perennial plant species' diversity, abundance and cover is determined. This is used to calculate forage and fuel load. Mammal density is determined in each area from helicopter surveys within predetermined strip widths at set altitude and speed. In addition to two back seat observers, a further individual in the front seat uses a computer linked to a global positioning system to locate each sighting and ensure the transect line is followed. The results are presented to each manager as histograms showing changes in annual vegetation (grasses, trees and elephant impact), animal stocking rate and diversity, energy requirements and fuel load, along with annual rainfall. This allows managers to compare their site with environmentally similar reserves facing similar rainfall patterns. Individuals can then see how their area has responded to both weather and management and whether their site is responding in a typical or atypical manner. These results can also be used to model the system and so also aid decision making (Peel, 2009).

Penguins in zoos Zoo keepers typically have responsibilities for single enclosures of a given species. It is then hard to assess how well their animals are performing and what can be done to improve this. Blay & Côtè (2001) examined the performance (per capita egg productivity, chick productivity and hatching success, measured as the proportion of eggs laid that hatched) of 16 Humboldt penguin *Spheniscus humboldti* enclosures in UK zoos and bird gardens. There was considerable difference in performance between establishments. Breeding output per pair was higher if the substrate used within the nesting boxes was sand or gravel rather than twigs or vegetation. Hatching success increased with larger pool size and was highest in enclosures with a concrete floor. This leads to the production of best practice advice.

Lagoon management in the UK Managers of saline lagoons have the problem that breeding success of birds can vary between years. Robertson (1993) compared a series of coastal lagoons in England that showed considerable variation in the breeding success of avocets *Recurvirostra avosetta*. By collecting environmental and prey density data across sites it was straightforward to show that breeding success was strongly correlated with food items. Salinity and fish abundance were the main factors determining the density of prey. Reserve wardens could then compare their avocet breeding success, lagoon salinity and fish density with that of other reserves and consider if there were ways of improving success, e.g. by adjusting salinity or periodically draining to reduce fish density.

In each of these case studies the application of standard monitoring methods across sites showed considerable variation in performance. This enabled practitioners to identify whether or not there were potential management problems. In each of the case studies this led to greater understanding of the causes of variation.

Our message is that the capacity to compare across sites greatly helps management. It has three advantages. Firstly, it ensures that the assessment of effectiveness is collated and stored so that changes in performance over time can be examined. Secondly, participants learn how their effectiveness compares with others (of course there may be reasons for such variability that are outside the control of the practitioner). Thirdly, it encourages an examination of the reasons for variability in effectiveness by determining how practices differ and the consequences of such differences.

There can be a range of ways in which discovering variation in performance leads to improved conservation. These include comparing approaches with those with greater effectiveness, carrying out cross-site comparisons or collating the evidence as to what is effective (Dicks et al., 2010). In some cases the variation in performance will have causes that are difficult for practitioners to influence, such as climate or soil type. It is still useful to account for this variability, for example by concentrating effort in areas or on issues where the benefits are likely to be highest.

It is not difficult to imagine comparisons within large conservation organizations, i.e. internal benchmarking. It is more challenging to benchmark across a range of organizations, i.e. external benchmarking. One option would be for this to be coordinated through a national regulating organization or through an international conservation organization. In agriculture benchmarking has been particularly effective when provided by an advisor network or run alongside farm accounts, thus reducing the need for double data entry (Jack, 2009). An example of collaboration across sites is the case of the wild dog *Lycaon pictus* in Africa, where the information on the fate of individuals in different sites was collated by a wide range of researchers (Gusset et al., 2008).

Thus, our vision is that comparisons will not only come from within organizations within a region but ideally across multiple organizations in different regions. The challenges are ensuring methods for assessing outcomes that are consistent across sites, ensuring there is an incentive to present results accurately and without exaggeration, and providing an incentive for making comparative information available.

Benchmarking is only appropriate for those activities that are repeatable and have measurable outcomes in terms of a set of objectives. However, there are a range of suitable repeatable activities. Collecting data in a uniform manner and using these for comparison with others would be a fruitful activity. Potential subjects include densities of key species, breeding success of key species, survival of grazers in systems where the grazing resource is limiting, use of artificial nesting sites, responses of visitors to interventions, control of invasive species, and loss of individuals to specific sources such as a particular disease. Benchmarking would be made easier by developing more consistent standards for data collection, e.g. by agreed standardized techniques and means of documenting results (Sutherland et al., 2010). In the South African example, the National Rangeland Monitoring programme aims to achieve this.

There is a fine line between benchmarking, where the objective is to learn from comparisons with other sites, and performance measures, which are used to judge individuals or organizations and may be associated with rewards. When comparisons are used as a performance measure, these results run the risk of becoming distorted if collected by the individual or programme that is being assessed. If this is a problem, the solutions include the data being collected by a third party (as in our three ecological examples), for the data to be submitted anonymously (as initially in the cystic fibrosis example) or for the data to be coordinated by one group but made available for others to compare in confidence. An example of the latter approach is the Farm Business Survey (Hadley & Irz, 2008), which is an annual government-funded survey of a range of measures of accounting information so that individual farmers can, in confidence, compare their profitability with that of comparable farms.

In the conservation context we suggest that the term best practice is not used. There are two problems with this term. Firstly, it is often used to describe a practice that appears reasonable without any comparison with other practices. Where this is not the case, the terms standard practice or suggested practice are more honest. If there is some evidence that the practice is effective then effective practice is a suitable term. Promising practice is the term recommended for a practice that has not undergone comparison (Keehley & Abercrombie, 2008). Secondly, it implies that it will always be the best practice, yet further research may well change the assessment as to which technique is best while new techniques may be developed that are more effective. Current effective practice may be a better term.

Benchmarking is most likely to be adopted if it obviously leads to useful comparisons from which individuals benefit by improving practice. The alternative would be to make it mandatory. In some countries, such as the USA, some medical funding is provided on the condition that benchmarking occurs. We suggest that, in some circumstances, benchmarking using standard comparisons could be a useful means for individuals and programmes to evaluate their efficiency and thus lead to improvements in the success of conservation actions.

Acknowledgements

Markus Gusset and an anonymous referee made useful comments. Dan Salzer told WJS about cystic fibrosis. WJS is funded by Arcadia.

References

Blay, N. & Côté, I.M. (2001) Optimal conditions for breeding of captive Humboldt penguins (*Spheniscus humboldti*): a survey of British zoos. *Zoo Biology*, 20, 545–555.

Camp, R.C. (1998) *Global Cases in Benchmarking*. ASQ Quality Press, Milwaukee, USA.

Dicks, L.J., Showler, D.A. & Sutherland, W.J. (2010) *Bee Conservation: Evidence for the Effectiveness of Interventions*. Pelagic Publishing, Exeter, UK.

Gawande, A. (2007) *Better. A Surgeon's Notes on Performance*. Profile Books, London, UK.

Gusset, M., Ryan, S.J., Hofmeyr, M., Van Dyk, G., Davies-Mostert, H.T., Graf, J.A. et al. (2008) Efforts going to the dogs? Evaluating attempts to re-introduce endangered wild dogs in South Africa. *Journal of Applied Ecology*, 45, 100–108.

Hadley, D. & Irz, X. (2008) Productivity and farm profit—a microeconomic analysis of the cereal sector in England and Wales. *Applied Economics*, 40, 613–624.

Jack, L. (2009) *Benchmarking in Food and Farming*. Gower Publishing, Farnham, UK.

Keehley, P. & Abercrombie, N.N. (2008) *Benchmarking in the Public and Nonprofit Sectors*, 2nd edition. Wiley, San Francisco, USA.

Peel, M.J.S. (2009) *Ecological Monitoring: Sabi Sand Wildtuin*. Sixteenth Unpublished Annual Report. Agricultural Research Council of South Africa, Nelspruit, South Africa.

Peel, M.J.S., Kruger, J.M. & Zacharias, P.J.K. (2005) Environmental and management determinants of vegetation state on protected areas in the eastern Lowveld of South Africa. *African Journal of Ecology*, 43, 352–361.

ten Raa, T. (2009) *The Economics of Benchmarking*. Palgrave Macmillan, Basingstoke, UK.

Robertson, P.A. (1993) *The management of artificial coastal lagoons in relation to invertebrates and avocets* Recurvirostra avosetta *(L.)*. PhD thesis. University of East Anglia, Norwich, UK.

Saul, J. (2004) *Benchmarking for Nonprofits*. Fieldstone Alliance, Minnesota, USA.

Sutherland, W.J., Pullin, A.S., Dolman, P.M. & Knight, T.M. (2004) The need for evidence-based conservation. *Trends in Ecology & Evolution*, 19, 305–308.

Sutherland, W.J., Armstrong, D., Butchart, S.H.M., Earnhardt, J., Ewen, J., Jamieson, I. et al. (2010) Standards for documenting and monitoring bird reintroduction projects. *Conservation Letters*, 3, 229–235.

Biographical sketches

William Sutherland has wide interests in conservation biology and is committed to making global conservation practice more rigorous. He is particularly interested in predicting the impacts of environmental change, especially on bird populations. Another major theme of his work is using evidence-based conservation to collate experience of the effectiveness of interventions (see http://www.conservationevidence.com) and then use this evidence to advise practice. Michael Peel co-ordinates rangeland ecology research teams across various Provinces of South Africa. He initiated the Savanna Ecosystem Dynamics Project in the eastern Lowveld of South Africa in 1989, and is investigating the potential of the natural resources of the savanna biome to contribute to the economy and development of southern Africa in harmony with social and environmental needs.

https://doi.org/10.1017/S0030605310000852 Published online by Cambridge University Press