

League tables for orthodontists

Frank Dunstan*, Stephen Richmond**, Ceri Phillips*** and Peter Durning**

Departments of *Primary Care and Public Health and **Dental Health and Biological Sciences, Cardiff University and ***School of Health Science, University of Wales, Swansea, UK

SUMMARY The aim of this study was to explore the complexities in constructing league tables purporting to measure orthodontic clinical outcomes. Eighteen orthodontists were invited to participate in a cost-effectiveness study. Each orthodontist was asked to provide information on 100 consecutively treated patients. The Index of Complexity, Outcome, and Need (ICON) was used to assess treatment need, complexity, and outcome prior to, and on completion of, orthodontic treatment. The 18 orthodontists were ranked based on achieving a successful orthodontic outcome (ICON score less than or equal to 30) and the uncertainty in both the success rates and rankings was also quantified using confidence intervals.

Successful outcomes were achieved in 62 per cent of the sample (range 19–94 per cent); four of the 18 orthodontists failed to achieve more than a 50 per cent success rate. In developing league tables, it is imperative that factors such as case mix are identified and accounted for in producing rankings. Bayesian hierarchical modelling was used to achieve this and to quantify uncertainty in the rankings produced. When case mix was taken into account, the four with low success rates were clearly not as good as the top four performing orthodontists.

League tables can be valuable for the individual orthodontist, groups of orthodontists, payment/insurance agencies, and the public to enable informed choice for orthodontic provision but must be correctly constructed so that users can have confidence in them.

Introduction

We live in a society which is increasingly evidence based, in the sense that decisions have to be made on an evaluation of the best possible evidence, rather than merely on professional opinion. In particular, evidence-based medicine and dentistry have assumed major importance in the last decade (Sacker, 2005). While most effort has been expended on evaluating treatments, it is perhaps inevitable that there will be attempts to compare the ways in which these treatments are delivered, whether between countries, institutions, or individual orthodontists (Richmond and Andrews, 1993). It is a short step from there to producing league tables in which these units of comparison are ranked according to specified criteria.

League tables have been used in the United Kingdom (UK) for a number of years to evaluate education, with schools being assessed on the basis of examination results (Department for Education and Skills, 2006). Hospitals have also been graded according to a variety of criteria such as waiting lists and patient throughput (Department of Health, 2006). Recently, tables comparing the mortality associated with individual cardiac surgeons in the UK have been published (Bridgewater, 2005). Also the World Health Organization has produced league tables for healthcare systems (Kmietowicz, 2000). Many of these tables have been heavily criticized for being too simplistic and not accounting for factors which affect the outcome being assessed but which are outside the control of the institution (Poloniecki *et al.*, 1998). For example, a

frequent complaint levied against league tables for schools is that they fail to measure the quality of input, in the form of the ability of pupils, as much as the output, although steps are being taken to adjust for such factors (Goldstein, 2003).

It is important if we are to have league tables highlighting healthcare effectiveness that they are correctly constructed and interpreted as they could affect the careers of individuals and funding of services. It is therefore essential that efforts are made with respect to validation to allow intelligent interpretation of their results.

The aim of this study was to explore three important issues affecting the construction and interpretation of league tables, namely random variation, case mix, and selection bias.

Subjects and methods

Eighteen orthodontic specialist practitioners were randomly selected from the General Dental Service, Hospital Dental Service, and Community Dental Service in Wales (6 in each service; Richmond *et al.* 2005). It was planned to follow 100 patients attending each orthodontist to completion of orthodontic treatment. Patient and orthodontist questionnaires were drafted and completed at the beginning, during, and on completion of treatment. These questionnaires provided information relating to the patient, orthodontist, intervention, management of the orthodontic process, and costs.

The malocclusion was evaluated pre- and post-orthodontic care using the Index of Complexity, Outcome, and Need (ICON; Daniels and Richmond, 2000). An acceptable outcome is defined as a final ICON score of less than or equal to 30 and for the purposes of this research was used as the measure of success.

Statistical analysis

The percentage of subjects achieving an acceptable outcome (less than or equal to 30 ICON points) for each orthodontist was calculated, and the orthodontists were ranked by the percentage of acceptable treatment outcomes, with appropriate confidence limits being calculated. The statistical analysis used hierarchical modelling (Goldstein, 2003). In a hierarchical data model, data are organized into a tree-like structure; here, the orthodontists were a sample from the whole population of orthodontists and nested within each was the set of patients who they treated. The probability of successful outcome for each orthodontist, taking account of case mix, was estimated (initial complexity of a subject as defined by an ICON score of at least 90 was included). The method also allowed estimates to be derived of the probabilities of different ranking positions for each orthodontist. A Bayesian approach (Spiegelhalter *et al.* 2004, Marshall and Spiegelhalter, 2000) and the software Winbugs (Spiegelhalter *et al.*, 1999) were used. This approach offers a flexible method for combining information from different sources to calculate the probabilities of interest, but is not crucial to the arguments advanced in this article; an alternative would be a multilevel modelling package such as MLwiN (Goldstein *et al.*, 1998).

Results

Although six orthodontists in each service were approached, two self-employed orthodontists declined to take part; consequently, a further three orthodontists were approached who agreed to participate in the study. Two orthodontists working in the community clinics resigned their posts and one who had originally agreed to take part later withdrew from the study. A further two community orthodontists were recruited. The final sample consisted of seven self-employed, six salaried, and five community orthodontists. The low number of patients treated by some of the orthodontists is a reflection on the timing of enrolling the orthodontists for the study as a result of resignations and subsequent recruitment of newly employed orthodontists.

Fourteen of the 18 orthodontists were male, their average age was 49 years at the start of the study (range 38–59), the median year for obtaining their primary dental degree was 1971 (range 1963–1984) and for their specialist qualification, it was 1977 (range 1965–1995). Twelve of the 18 orthodontists possessed a Fellowship in Dental Surgery from one of the Royal Colleges.

There were 1087 patients with ICON scores for both pre- and post-treatment, with the number of subjects per orthodontist varying between 19 and 94 (Table 1). Not all subjects were in need of treatment as defined by an ICON score of more than 43; this analysis was thus restricted to the 90 per cent who did require treatment. The overall success rate was 62 per cent, but this varied from 19 to 94 per cent between different orthodontists. The rates of achieving a successful outcome (less than or equal to 30 ICON points) for the 18 orthodontists are shown in Figure 1.

The level of random variation is shown in a plot of 95 per cent confidence intervals (CIs) for the success rates arranged

Table 1 Pre- and post-treatment Index of Complexity, Outcome, and Need (ICON) scores for 18 orthodontists.

Orthodontist	<i>n</i>	ICON scores		
		Mean pre-treatment score	Mean post-treatment score	Mean change in score
Self-employed				
A	84	60.0	38.3	21.7
B	84	63.6	20.5	43.1
C	63	65.3	33.0	32.3
D	40	63.0	21.8	41.2
E	88	69.4	31.5	37.9
F	87	61.4	32.5	28.9
G	26	60.4	30.5	29.8
Average	67	63.6	30.3	33.4
Salaried hospital				
H	94	75.0	19.9	55.1
I	53	81.1	33.1	47.9
J	55	68.2	24.7	43.5
K	44	69.9	25.5	44.4
L	54	70.2	29.1	41.1
M	53	73.8	28.7	45.1
Average	59	73.3	26.0	47.3
Salaried community				
N	19	67.5	22.9	44.5
O	78	72.5	29.0	43.5
P	54	70.8	36.8	34.1
Q	55	59.3	24.5	34.8
R	56	70.6	25.2	45.4
Average	52	68.6	28.4	40.2
Overall	1087	68.0	28.4	39.5

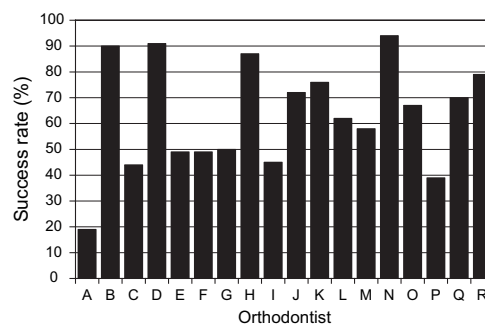


Figure 1 Success rate for the 18 orthodontists in achieving an Index of Complexity Outcome and Need score of less than or equal to 30 on completion of treatment.

in ascending order (Figure 2). The CI shows the range of values plausible for the true success rate in an orthodontist's case load given the observed success rate for the sample of patients treated. It can be seen that orthodontist A provided the poorest orthodontic outcomes of the 18 orthodontists.

Taking into account the case mix, defined here by the percentage of subjects with an initial ICON score of at least 90, the revised CIs for success rates are displayed in Figure 3. Orthodontist A provided significantly poorer outcomes compared with orthodontists L, O, Q, J, K, R, H, N, B, and D.

The distributions of the ranks of the orthodontists are shown in Figure 4. Orthodontist A was ranked 18 of 18 and orthodontist P 17th, but both orthodontists C and I also had a considerable probability of being ranked 17th. Orthodontist D has a 40 per cent chance of being ranked 1st but substantial probabilities of being second and third or even fourth. On the other hand, G could be ranked anywhere between ninth and 17th, although the best estimate from the CIs is around 13th.

Discussion

There have been a limited number of studies assessing the treatment need and outcome using ICON in Sweden and Greece (Richmond *et al.*, 2001a,b). The mean ICON scores for seven Swedish orthodontists were 72.5 pre-treatment (range 46.8–78.7) and post-treatment 28 (range 23.5–30.9). For five Greek orthodontists, the mean ICON scores were 69 pre-treatment (range 55–78.9) and 24.5 post-treatment (range 21.9–27.2). These scores compare favourably with the 68 (range 59.3–81.1) and 28.4 (range 19.9–38.3) pre- and post-treatment ICON scores in this study. The overall success rates were 62, 71, and 90 per cent for Welsh, Swedish, and Greek orthodontists, respectively. However, the studies in Greece and Sweden were retrospective and depended on the orthodontists' self-selection of cases ($n = 6-52$). The sample obtained for the present investigation was prospective with patients nominated in advance with less chance of selection bias. These studies have highlighted the variation in pre- and post-treatment ICON scores within, and between, orthodontists and their relative ability to achieve a successful outcome.

However, there are generally two problems with using crude success rates as a measure of the effectiveness of a practitioner of orthodontic treatment. Firstly, it does not take account of case mix. It is possible that some of those orthodontists with apparently low success rates may have had more subjects with complex malocclusions with a smaller chance of delivering an acceptable outcome. This is highlighted by the salaried services (orthodontists H–M) having higher initial ICON scores compared with the other two groups. Secondly, no account is taken of random variation. For example, orthodontist H had a success rate of 87 per cent based on 89 subjects needing treatment. If he were to have another 89 subjects, then almost certainly the

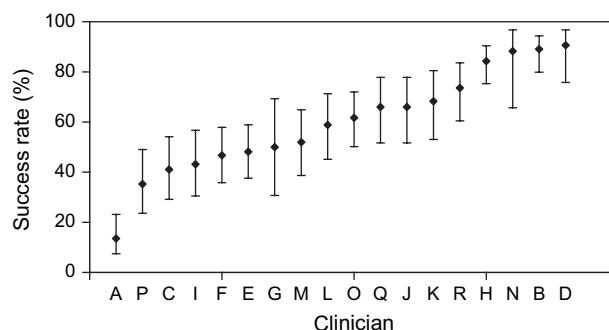


Figure 2 Ordered success rate for the 18 orthodontists (confidence intervals).

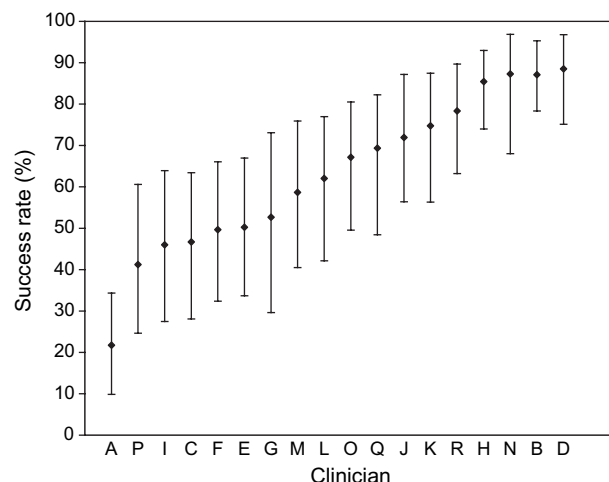


Figure 3 Confidence intervals for the 18 orthodontists arranged in ascending order and taking account of the case mix.

success rate would be different; random factors would make it most unlikely that the rate was identical and it could easily be as large as 92 per cent or as low as 78 per cent. Some of the other estimates were based on smaller numbers of subjects and therefore the resulting uncertainty is even greater.

A plot showing 95 per cent CIs for the success rates is more helpful as it explicitly demonstrates the level of this random variation. While orthodontist M had a success rate of 58 per cent compared with the 76 per cent of orthodontist K, it is quite possible that there is no real difference between them (Figure 2). Indeed these results could easily have occurred by chance if both orthodontists had long-term success rates of 65 per cent. Figure 2 is more useful than Figure 1 but it does not answer all the questions. While it appears likely that A has the lowest rank, it would be useful to be able to quantify this. There are several candidates for being the 'best'—N, B, and D in particular. Identifying an orthodontist who produces the best orthodontic treatment outcomes, taking account of case mix, requires more sophisticated methods.

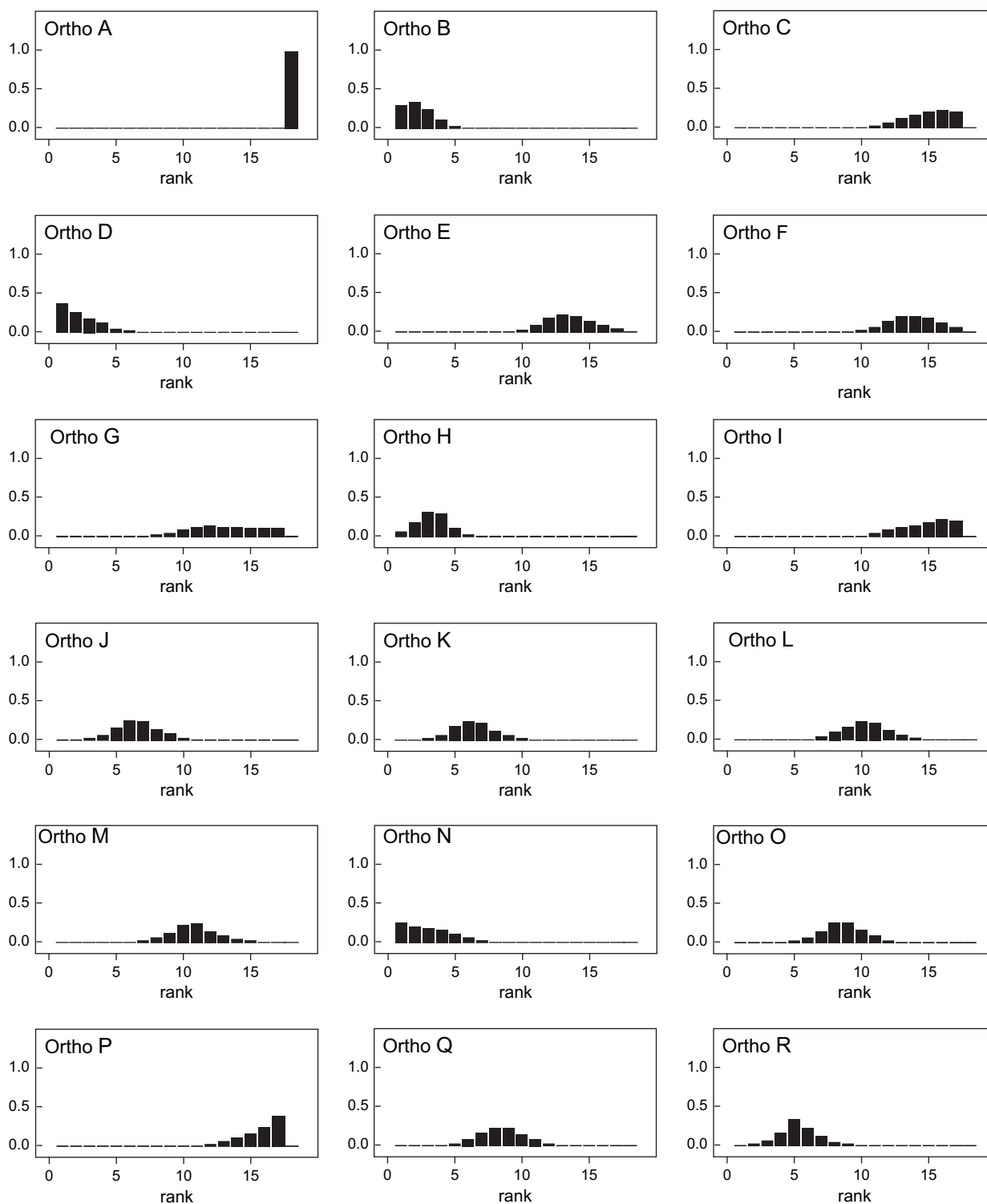


Figure 4 Probability distributions of ranks for the 18 orthodontists.

Figure 3 shows the CIs for the 18 orthodontists, arranged in ascending order and taking account of the case mix. The ranking has not changed significantly even though the case mix varies considerably between the orthodontists from less than 5 per cent of subjects classed

as severe for one practitioner to 33 per cent for another. It appears that initial severity, while important, does not appear to be a strong predictor of a successful outcome. There are some small differences; for example, not only have orthodontists I and C changed places but also the

estimates have changed and the CIs are rather wider, reflecting greater uncertainty.

Another important issue is selection bias. If the orthodontists were ranked and those with apparently the lowest and highest ranks compared, the mechanism of selection means that they are highly likely to be significantly different. This is different from comparing orthodontists having selected them by chance. To demonstrate this, suppose that 18 orthodontists had 60 subjects (the average number in this study) and each had a probability of success of 62 per cent (the overall average rate here, for each subject treated). If the highest and lowest are compared using a nominal significance level of 5 per cent, there is actually a 35 per cent chance of deciding that two orthodontists are different if no allowance is made for the fact that the extremes are being compared. The significance level needs to be set at 1.7 per cent to achieve the required 5 per cent rejection rate. Care must therefore be taken in deciding what comparisons should be made and how they should be evaluated.

Conclusions

League tables can be useful in making comparisons, whether between orthodontists or between different healthcare systems in which the orthodontic care is delivered. They can add to the evidence concerning particular treatments or treatment modalities. For example, if it appears that two units are performing differently, then an investigation may highlight some important differences from which lessons can be learned. As has been seen in education, however, it is important, if they are to be accepted, that all relevant factors are taken into account to avoid the accusation that a table is measuring input rather than output. The methods illustrated here can adjust for relevant factors; the complexity/severity of the subject's malocclusion did not impact greatly on the outcome. The league table could have been adjusted for additional factors; this example was merely an illustration.

Ranking individuals or institutions is an emotive issue and it is vital that any ranks produced are accompanied by measures of uncertainty and that comparisons are made fairly. As has been shown, comparing extremes can be misleading unless the method of selection is recognized. Comparing orthodontists A and B, having chosen them at random, is quite different from comparing them because they seem to be the worst and best, respectively.

League tables have considerable potential for informing orthodontist, patients, and third party payment agencies; however, they will be quickly discredited unless they are constructed and interpreted correctly.

Address for correspondence

Professor Frank Dunstan
Department of Primary Care and Public Health
Cardiff University
Cardiff CF14 4YS
UK
E-mail: dunstanfd@cardiff.ac.uk

Funding

Wales Office for Research and Development in Health and Social Care (R96/01/094).

References

- Bridgewater B 2005 Mortality data in adult cardiac surgery for named surgeons: retrospective examination of prospectively collected data on coronary artery surgery and aortic valve replacement. *British Medical Journal* 330: 506–510
- Daniels C, Richmond S 2000 The development of the index of complexity, outcome and need (ICON). *Journal of Orthodontics* 27: 149–162
- Department for Education and Skills 2006 School and college achievement and attainment tables (formerly performance tables). <http://www.dfes.gov.uk/performance/tables/>
- Department of Health 2006 Hospital waiting times. <http://www.performance.doh.gov.uk/waitingtimes/>
- Goldstein H 2003 Multilevel statistical models, 3rd edn. Kendall Library of Statistics, Oxford.
- Goldstein H et al. 1998 A user's guide to MLwiN, Version 1.0. Institute of Education, London
- Kmietowicz Z 2000 France heads WHO's league table of health care systems. *British Medical Journal* 320: 1687
- Marshall E C, Spiegelhalter D J 2000 Institutional performance. In: Leyland A H, Goldstein H (eds). *Multilevel modelling of health statistics*. Wiley, Chichester, pp. 127–142.
- Poloniecki J, Valencia O, Littlejohns P 1998 Cumulative risk adjusted mortality chart for detecting changes in death rate: observational study of heart surgery. *British Medical Journal* 316: 1697–1700
- Richmond S, Andrews M 1993 Orthodontic treatment standards in Norway. *European Journal of Orthodontics* 15: 7–15
- Richmond S, Ikonou C, Williams B, Rolfe B 2001a Orthodontic treatment standards in Greece. *Hellenic Orthodontic Review* 4: 9–20
- Richmond S, Dunstan F, Phillips C, Daniels C, Durning P, Leahy F 2005 Measuring the cost, effectiveness and cost-effectiveness of orthodontic care. *World Journal of Orthodontics* 6: 161–170
- Richmond S, Ikonou C, Williams B, Ramel S, Rolfe B, Kurol J 2001b Orthodontic standards in a public group practice in Sweden. *Swedish Dental Journal* 25: 137–144
- Sacker D L 2005 Evidence based medicine. Churchill Livingstone, New York
- Spiegelhalter D J, Abrams K R, Myles J P 2004 Bayesian approaches to clinical trials and health care evaluation. Wiley, Chichester
- Spiegelhalter, D J, Thomas A, Best, N G 1999 WinBUGS version 1.2 user manual. Biostatistics Unit, Medical Research Council, Cambridge

Copyright of European Journal of Orthodontics is the property of Oxford University Press / UK and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.