A BALANCED APPROACH TO NORMALIZING BUS OPERATIONAL DATA FOR PERFORMANCE BENCHMARKING PURPOSES

Paper submitted for presentation at the (2012) Transportation Research Board 91st Annual Meeting

Date submitted: 1st August 2011
Resubmitted: 31 October 2011

Corresponding author: Mark Trompet
Railway and Transport Strategy Centre, Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London
Project Manager of the International Bus Benchmarking Group
Skempton Building, SW7 2AZ London; Tel +(20)75941519; Fax +(20)75946107
E-mail: m.trompet@imperial.ac.uk

Daniel J. Graham
Railway and Transport Strategy Centre, Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London
Skempton Building, SW7 2AZ London; Tel +(20)75946088; Fax +(20)75946107
E-mail: d.j.graham@imperial.ac.uk

Words: 5456
Figures: 4
Tables: 2

ABSTRACT

Peer organizations in a performance benchmarking group are usually carefully selected based on similar characteristics such as the type of services offered, operational characteristics and density of the service area. These similarities enable organizations to compare performance once their operational data are normalized. The most commonly used normalization factors for the demand side output are passenger boardings and passenger kilometres. For the supply side output these are vehicle kilometres and vehicle hours. Through seven years of experience in the International Bus Benchmarking Group (IBBG) a better understanding of differences in service characteristics between 'similar' peers has been achieved. It became clear that relative performance can often not be concluded from a performance indicator normalized in one dimension. Variety in commercial speed, trip length, vehicle capacity, vehicle weight and network efficiency results in the need for a multi dimensional or balanced approach to data normalization. This paper quantifies the variety within these operational characteristics and provides a framework for benchmarking practitioners and policymakers that suggests applicable combinations of denominators for a balanced normalization process. This paper further describes how alternative normalization factors such as revenue service planning capacity kilometres and total tonne kilometres have improved comparability of Key Performance Indicators (KPIs).

KEYWORDS: Normalization, Benchmarking, Urban bus operations, Key performance indicators

1. INTRODUCTION

Based on a definition by Lema and Price (*1*) benchmarking is defined as *a systematic process of continuously measuring, comparing and understanding organizations' performance and change in performance of a diversity of key business processes against comparable peers anywhere else in the world to gain information which will help the participating organizations to take action to improve their performance*.

One of the prerequisites for a successful benchmarking process is the selection of peer organizations based on similar characteristics. In bus benchmarking special focus should be given to the types of bus services offered (e.g. urban, suburban, charter, schoolbus, para-transit, etc), and the service area characteristics, in particular the density of operations and demand. These similarities enable organizations to compare performance once their data are normalized for scale.

As described by Trompet *et al.* (*2*), variability in comparable performance is welcomed as this results in the identification of best practices, which can lead to improved performance. However, the authors also describe that, even within a group of seemingly comparable operators, there is variation in operating characteristics and environment that results in a subset of operators being incomparable. The composition of this 'incomparable subset' differs between key performance indicators, often in relation to the normalization factor chosen.

This paper aims to describe a technique to refine the performance comparisons within an existing peer group. It shows to benchmarking practitioners and policy makers that drawing conclusion from and acting on relative performance differences observed in a single Key Performance Indicator (e.g. only normalized by a single denominator) is often ineffective. Due to variety between peers of operating characteristics such as commercial speed, trip length, vehicle capacity, vehicle weight and network efficiency (e.g. variety in amount of deadheading) it is necessary that performance is at least reviewed from two different dimensions.

To achieve this objective the remainder of this paper is structured as follows. Section 2 reviews normalization factors used or suggested in a number of relevant previous and on-going bus public transport benchmarking initiatives. Section 3 describes the International Bus Benchmarking Group (IBBG) dataset which is used for this study. Section 4 describes the variability of five operational characteristics that can lead to a skewed perspective of relative performance when only one denominator is used for normalization. In section 5 a framework is then presented that schematically suggests which combinations of two denominators could be used to obtain a more realistic, balanced view of relative performance. Section 6 further discusses two alternative normalization factors: revenue service planning capacity kilometres and total tonne kilometres, which within the IBBG have resulted in improved performance comparability. Conclusions are drawn in Section 7.

2. NORMALIZATION IN TRANSPORT BENCHMARKING AND LITERATURE

An overview of public transport benchmarking initiatives has been provided in a variety of reports and papers (e.g. TCRP (*3*), EQUIP (*4*), Geerlings *et al.* (*5*) and Gudmundsson *et al.* (*6*)). With regards to urban bus performance measurement specifically, Mulley (*7*) describes the process and lessons learned from the UK Bus benchmarking Group which was based on the benchmarking handbook developed by EQUIP (*8*). Alongside the references mentioned above, the papers and project reports listed under references (*9-21*) have also been reviewed to understand how practitioners have normalized for scale. The focus in the review was to find recommendations towards improved normalization factors and on discussions in relation to the possible bias of using a single denominator for normalization.

The reports and papers generally underline the importance of peer selection and normalization for successful benchmarking. Complete sets (or examples) of main key performance indicators used are often provided. These confirmed that the most commonly used normalization factors for the demand side output are passenger boardings and passenger kilometres. For the supply side output these are vehicle kilometres and vehicle hours. Useful discussions have been provided on the use of vehicle hours as a preferred denominator over vehicle kilometres (Hencher and Daniels (*9*), Fielding et al. (*10*)). The majority of reviewed papers also mention the use of seat kilometres for supply side normalization. This led to the addition of Section 6 to this paper in which the authors discuss the role and possible bias of seat capacity kilometres as a normalization factor in a benchmarking exercise.

The benchmarking projects and literature reviewed did not explain in some level of detail why a certain normalization denominator was chosen (over an alternative one) and what bias could be involved in using that single denominator. This paper aims to contribute to this area.

Fielding *et al.* (*11*) and Stappenhorst (*12*) use cluster analysis to create comparable sub groups to improve comparability within performance indicators. This is feasible in a situation where the total number of participating organizations in a benchmarking group is sufficiently large; the size of each cluster of 'more comparable' peers should also be sufficient for performance comparison. The downside to clustering peers into more comparable subsets before benchmarking analysis is that possibly valid lessons and best practices from 'lesser' comparable peers are discarded. Clustering is done before the performance comparison is executed. This paper discusses a post performance comparison alternative, by understanding relative performance through a multi-dimensional normalization process.

3. THE DATA

The data used for this study have been collected through the International Bus Benchmarking Group (IBBG), which is facilitated by the Railway and Transport Strategy Centre at Imperial College London. The IBBG is now in its eighth year. Its current members are TMB Barcelona, STIB Brussels, Dublin Bus, Los Angeles County Metropolitan Transportation Authority, Carris Lisbon, London Buses, Milan ATM, STM Montreal, NYCT /MTA Bus New York, RATP Paris, STA Sydney Buses, Singapore SMRT and CMBC Vancouver. All members provide normal passenger public bus service operations in large urban areas. Up to ten years of data, from 2001 to 2010, are available for 90 key performance related data items. Some of these items are broken down into further sub-categories such as vehicle type or outsourced versus in-house labour. These performance related data are supported by another 38 background data items to provide context and understanding. Only data from normal service operations is included, filtering out the effect of charter, tourist bus, para-transit and school bus services.

In the IBBG, it took three years of iterative definition development, data collection and analysis before the member operators were sufficiently satisfied with the level of comparability to be able to use the data for performance comparison. As data comparability is key in benchmarking, Trompet *et al. (2)* summarised the necessary benchmarking conditions and decisions in these first years of the IBBG to gradually increase the quality of the dataset. One of the main contributing factors to data quality is a strict confidentiality agreement which in combination with the willingness to help and learn from each other creates an open and honest information sharing environment. This confidentiality agreement also applied to the data used for this study. As a result, the graphs and tables in this paper have been anonymised where necessary. Nevertheless, the IBBG members agreed that the lessons from this study should be shared for the benefit of the wider bus public transport industry and policymakers.

4. VARIABILITY OF SERVICE CHARACTERISTICS

Through seven years of experience in the International Bus Benchmarking Group (IBBG) a better understanding of differences in service characteristics between ‘similar’ peers has been achieved. Using the 2010 data from the thirteen organizations in the IBBG database, the most significant, quantifiable service characteristics variability have been observed in:

- Average passenger trip length, calculated as passenger kilometres divided by passenger boardings,
- Network efficiency, e.g. the proportion of deadheading and dead hours to total vehicle kilometres. Frequently also expressed as the inverse, e.g. the proportion of revenue vehicle kilometres/hours to total vehicle kilometres/hours,
- Average weighted vehicle planning capacity (see section 6.1 for a detailed discussion),
- Average commercial speed, calculated as actual revenue vehicle kilometres divided by actual revenue vehicle hours, and
- Average weighted vehicle weight (in this dataset: unloaded axel weight in tonnes), see section 6.2 for a detailed discussion.

In Table 1, descriptive statistics are used to describe the variation of these five types of service characteristics. These descriptive statistics are: Number of different bus operators in the sample (N), the mean value of the sample (μ), the minimum and maximum value within the sample and the standard deviation (σ).

TABLE 1 Variability in Service Characteristics – 2010 data of the International Bus Benchmarking Group

Type of service characteristic	*N*	*μ*	*Min*	*Max*	*σ*	*CV*
Average passenger trip length - km	13	4.6	2.8	8.0	1.6	0.35
Network efficiency - % of deadheading km	13	10.4	7.3	17.3	3.5	0.34
Weighted average vehicle planning capacity	12	71.2	52.1	94.8	15.7	0.22
Average commercial speed – km/h	11	17.3	12.0	23.3	3.3	0.19
Weighted average vehicle weight - tonne	12	12.5	11.2	14.9	1.0	0.08

N = Number of bus organizations in sample
μ = Sample average
Min = Minimum value
Max = Maximum value
σ = Standard deviation
CV = Coefficient of variation

To be able to directly compare the extent of variability of the different service characteristics, the standard deviation is divided by the mean (σ / μ), to compute the coefficient of variation (CV). The higher this number, the more variability is observed between values of bus organizations for that particular service characteristic.

The results in Table 1 are ranked by level of variability. The most variable service characteristic is average passenger journey length (CV=0.35), closely followed by network efficiency (CV=0.34). The range of values within these two characteristics is considerable. For example, the lowest average trip in the sample (2.8 km) is only 35% of the longest average passenger trip (8 km). As discussed in section 5, this has implications for performance benchmarking when data are only normalized for either passenger kilometres or passenger boardings.

Two medium variable service characteristics are: service planning capacity (CV=0.22) and commercial speed (CV=0.19). Expressed in a coefficient of variation the average weighted vehicle weight is medium/low (CV=0.08), e.g. the majority of bus organizations operate a fleet of similar average weight. However understanding the amplitude of data, e.g. the minimum value is 11.2 tonne and the maximum value is 14.9 tonne, lead to the realisation that some comparability of KPIs could be improved by using weight as a normalization factor. Section 6.2 describes this.

To understand if similar variety in service characteristics can be observed over time, Table 2 shows the coefficient of variation for the four most variable service characteristics for the period 2006-2010. Backdated details of average weighted vehicle weights were not available. For one operator only 2010 data were available, hence their data have been removed for this trend analysis. The sample size, apart for vehicle planning capacity, is therefore reduced by one. The results in Table 2 clearly show that the variability level and differences observed in Table 1 can also be observed at very similar levels in other years.

TABLE 2 Coefficient of Variability (CV) of Service Characteristics, 2006 – 2010 data of the International Bus Benchmarking Group

Type of service characteristic	*N*	*2006*	*2007*	*2008*	*2009*	*2010*
Average trip length - km	12	0.357	0.356	0.353	0.351	0.347
Network efficiency - % of deadheading km	12	0.316	0.322	0.330	0.327	0.331
Weighted average vehicle planning capacity	12	*NA*	0.224	0.227	0.223	0.220
Average commercial speed – km/h	10	0.204	0.205	0.203	0.203	0.199

N = Number of bus organizations in sample

Detailed description of the reasons for the variability of the service characteristics between ‘similar’ organizations is outside the scope of this study. However, when looking at the ranking, it seems that those service characteristics that are most variable are impacted by external factors, and factors that are harder to change or manage in the short or medium term. Trip length is determined by factors such as city size and density, availability of other transport modes and the complexity of the bus network itself. Network efficiency is strongly affected by depot locations and by the demand and supply profile. However, it is also affected by manageable impacts such as the decision to interline buses and/or the policy to bring buses back to the depot in the inter-peak period. Vehicle planning capacity is a function of the fleet characteristics and the passenger loading policy. Both can be managed on a medium term. Commercial speed is determined by factors such as traffic, dwell times and number of stops per route. Bus priority measures can be implemented on the short and medium term that speed up the organizations in the lower end of the spectrum, resulting in less variety in speed between organizations. Vehicle weights are somewhat different as they are less variable within the sample; however the variability observed is strongly related to regulatory differences in vehicle specification. This is difficult to manage in the short and medium term.

5. A BALANCED NORMALIZATION APPROACH

In the IBBG, common scale denominators used to normalize data are: passenger boardings, passenger kilometres, fleet size, total vehicle kilometres, revenue vehicle kilometres, revenue planning capacity kilometres, total planning capacity kilometres, revenue vehicle hours, total vehicle hours and (categories of) staff hours. For each KPI, the most suitable denominator was chosen. For example, a number of labour productivity KPIs are normalized by total vehicle hours as hours are a more significant labour cost driver. Vehicle kilometres on the other hand are the main cost driver for maintenance activity and hence used as the normalization factor.

Given the extent of variability of service characteristics, as identified in Table 1, within what is considered a group of comparable peers, it is important to realise that the choice of denominator can influence the relative performance (i.e. position in the graph) of individual bus organizations.

If an organization has ‘extreme’ values in any of the five variable service characteristics mentioned in Table 1, the choice of denominator will affect the relative position to peers in the comparison. For example, organisation ‘X’ has a lower average speed of 12 km/h, organization ‘Y’ has a typical average commercial speed of 15 km/h and operator ‘Z’ manages 19 km/h. Therefore operator ‘Z’ produces four more vehicle kilometres in the same time period than ‘Y’ and seven more vehicle kilometres per hour of operation than ‘X’. The relative performance (rank) of organizations ‘X’ and ‘Z’ will significantly differ depending on whether revenue vehicle kilometres or revenue vehicle hours is used as the normalization factor. Using revenue vehicle kilometres will over-normalize the performance of ‘Z’ and under-normalize the performance of ‘X’. For completion, in this example, due to its average ‘operating speed’ within the group of peers, the relative rank (performance) of ‘Y’ will not change if the data is normalized by either vehicle kilometres or vehicle hours.

Figure 3, used in section 6.1 to illustrate the effect of differences in vehicle capacity on relative ranking in a comparison, is another useful example. Operator ‘A’ has the largest average weighted vehicle capacity amongst the IBBG bus operators. In the average bus load KPI: passenger kilometres per revenue vehicle kilometres, operator ‘A’ scores highest; i.e. transporting more passengers per vehicle than other organizations. However, when normalized for the actual weighted revenue vehicle capacity of the fleet operator ‘A’ is ranked 6th. Both statistics are useful to know, but only the combination of the two provides a full picture of capacity utilisation.

The framework in Figure 1 illustrates how different normalization factors, represented with black letters, interact with and complement each other. The five variable service characteristics identified in Table 1 are represented in the dark boxes with white letters. These service characteristics are placed on the border(s) of their relevant normalization factors. Every normalization factor is linked to a 'counterpart' normalization factor, which is needed to counterbalance the bias created by 'extreme' values in a service characteristic. This relation is shown through the arrows.

For example, an operator has a typical passenger trip length, vehicle weight and vehicle capacity. However they also have relatively high levels of dead running as a result of their depot locations and demand profile, and have relatively slow speeds compared to peers. Network efficiency and commercial speed are therefore their key areas that create an unbalanced perspective of their relative performance. Using the framework in Figure 1 this operator should therefore normalize performance data by both revenue vehicle kilometres and revenue vehicle hours to balance the effect of their relatively slow commercial speed. Similarly, they should reproduce any 'revenue' normalized KPI (kilometres, capacity kilometres or hours) also by its 'total' normalized counterpart (for differences in network efficiency).

The two white boxes stating 'vehicle utilization performance' and 'system utilization performance' are not essential parts of the framework itself, but do provide useful additional information on the relationship between passenger based normalization factors and vehicle based normalization factors.

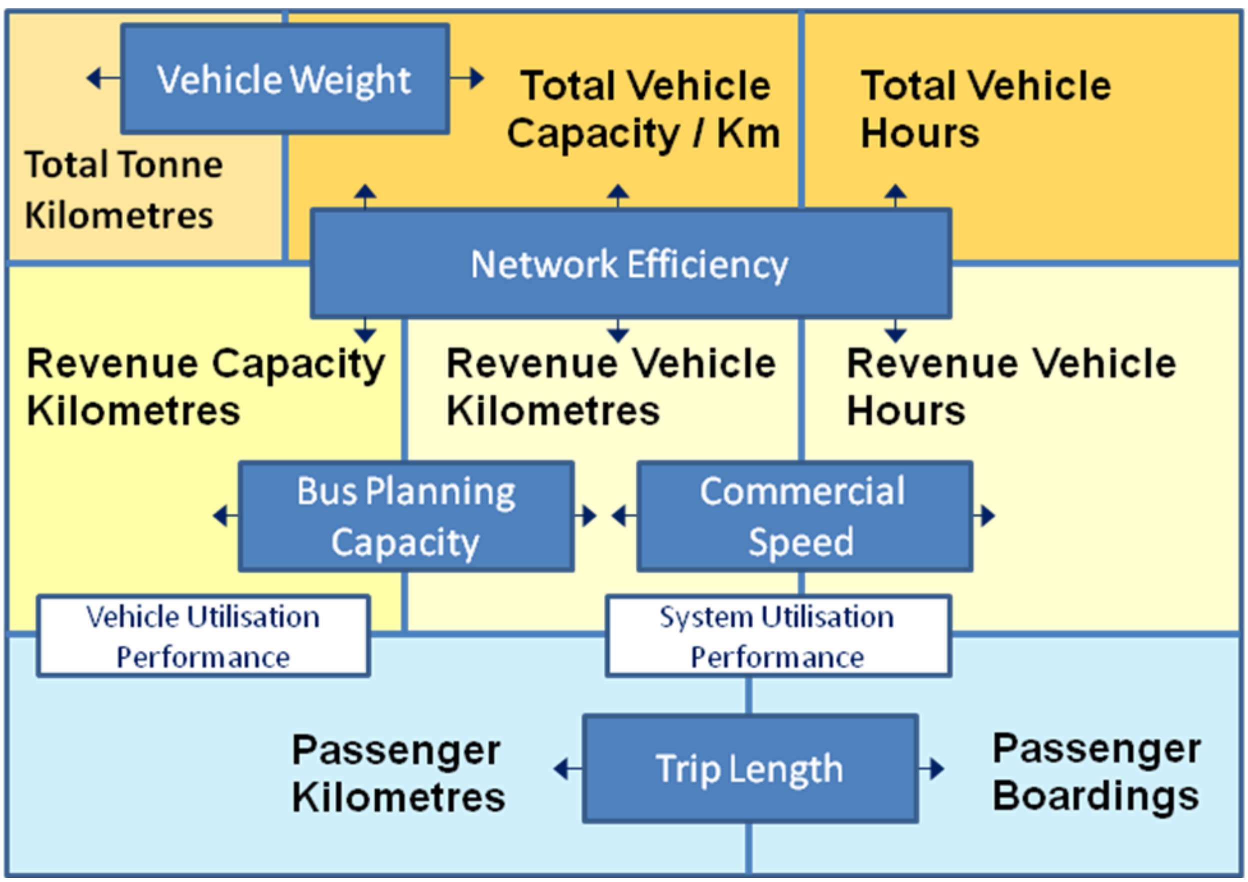

FIGURE 1 Two Dimensional Normalization Framework for Bus Performance Data

6. ALTERNATIVE NORMALIZATION FACTORS FOR IMPROVED PERFORMANCE COMPARABILITY

The International Bus Benchmarking Group has a dynamic approach towards key performance indicator development. Each year, a sub-set of KPIs and their related data items and definitions are reviewed and gradually improved. This section will discuss the improvements made with regards to understanding vehicle capacity and its use as a denominator, and improvements towards the comparability of fuel economy KPIs with the introduction of total tonne kilometres as a normalization factor.

6.1 Normalizing for Vehicle Capacity

The literature review, for example Hensher and Daniels (*9*), TCRP (*14*), MARETOPE (*16*) and Fielding *et al.* (*19*), confirmed that seat capacity kilometres is frequently used or proposed as a denominator to normalize for fleet capacity, especially for passenger loading key performance indicators. Some variations have been observed. Equip (*8*) uses the maximum capacity kilometres (seat and standing) in their load factor. It could not be confirmed if maximum here is defined as the manufacturers 'licensed capacity' or the physical 'crush' capacity. Phillips (*13*) and Fielding *et al.* (*10*) also mention total capacity kilometres, but the exact definition of 'total' is also not clear. The Urban Transport Benchmarking Initiative (*18*) defined capacity kilometres as place km, but only used those in the peak hour for a peak time loading KPI.

Harmonizing average weighted vehicle capacity in a benchmarking exercise is difficult as capacity can be defined differently. In the IBBG four types of vehicle capacity are collected annually: seating capacity, service planning capacity, licensed capacity and crush/max capacity. The difference between service planning and registered capacity is effectively the capacity that according to (service quality) policy should not be used. One can say it should not be considered to be available when scheduling bus services for service quality reasons.

Organizations have different policies on the proportion of registered capacity that can be used for service planning. For example, within the IBBG, one organization plans for their buses to be up to 95% full, but never more as they can be penalised by their authority. The service planning capacity is therefore 95% of the licensed capacity. Another organization, with longer average trip lengths, plans for most of their passengers to have a seat. Service planning capacity is therefore seating capacity + 30%, also called a load factor of 1.3. The Transit Capacity and Quality of Service Manual (*22*) describes that maximum scheduled passenger loads are typically 125 to 150% of seating capacity. Although sometimes inconsistent in policy, service planning capacity is always consistent in concept: e.g. the capacity offered to passengers. After a pilot period, it has therefore been decided within the IBBG that service planning capacity kilometres is the most suitable normalization factor for vehicle size and vehicle capacity.

Since 'seat capacity kilometres' is regularly proposed in literature as a normalization factor, it is useful to understand how seating capacity relates to service planning capacity. Figure 2 compares both the average weighted seating capacity per bus with the average weighted service planning capacity per bus across 12 IBBG bus organizations. The organizations are ranked from left to right based on their seating capacity and ranked alphabetically based on their service planning capacity.

The first observation is that there is a substantial difference in the rank of most bus operators in both types of capacity. For example, operator H has the second largest average weighted seating capacity per bus, while 'just' the 8th largest service planning capacity per bus. Or operator D, which has the 4th largest service planning capacity per bus, and the 10th largest seating capacity.

This variety observed in Figure 2 shows that in practice the average weighted seating capacity per bus is not the most effective normalization factor for vehicle size due to the different type of vehicles used by operators. Double-decker and mini/midi buses have a high proportion of seats and a small amount of additional standing capacity, while standard and especially articulated buses have relatively few seats compared to the total capacity offered. Therefore by using seat kilometres as a denominator, a normalization discrepancy occurs due to fleet composition differences between bus operators.

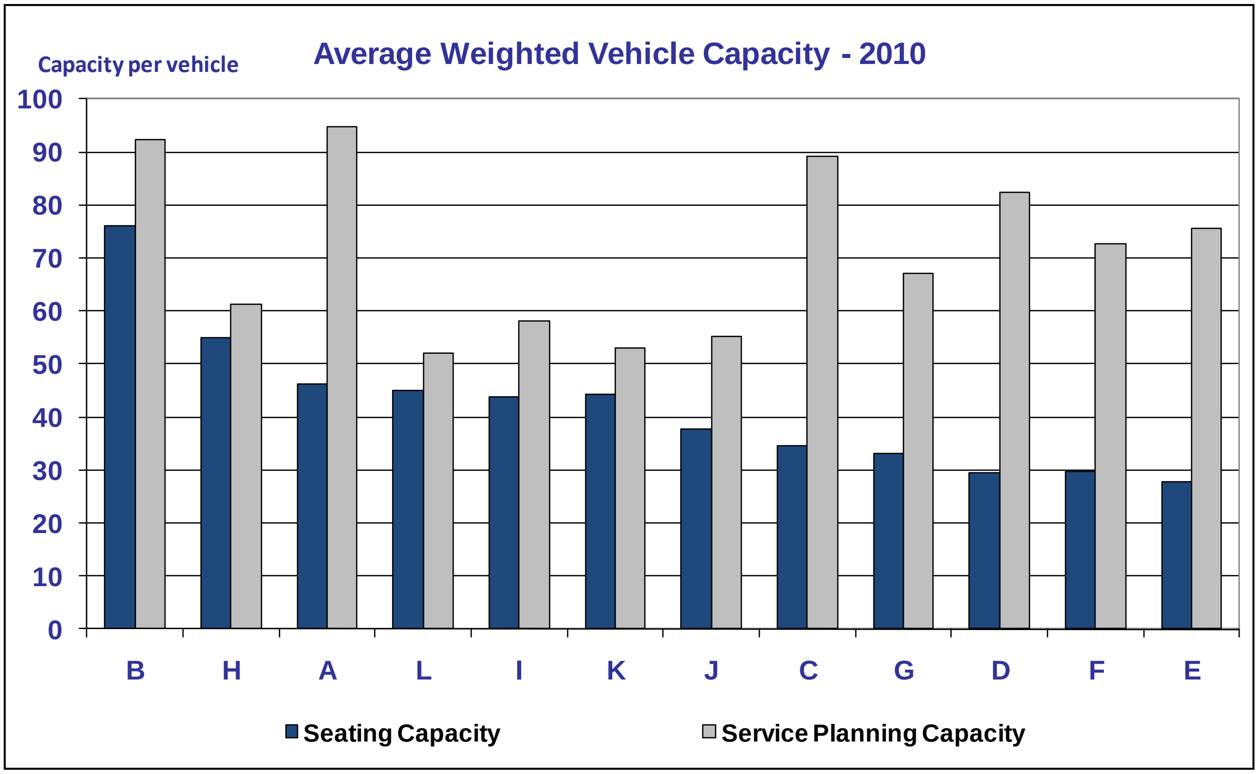

FIGURE 2 Average Weighted Vehicle Seating Capacity versus Average Weighted Vehicle Service Planning Capacity of the IBBG member organizations – ranked by seating capacity

Secondly, it can be concluded that the variety between the IBBG members within either type of capacity is quite significant. The average weighted seating capacity of operator E is just 37% of the seating capacity of operator B. Similarly, but less extreme, the service planning capacity of operator L is 55% of the service planning capacity of operator A. Clearly vehicle size is a factor that needs to be normalized for in performance comparison (as also shown in figure 1).

Figure 3 shows the effect of differences in vehicle size on two capacity utilisation key performance indicators. Passenger kilometres are normalized either by revenue vehicle kilometres (i.e. the average bus load KPI) or by the IBBG preferred service planning capacity kilometres (i.e. a vehicle capacity utilisation KPI). The results are indexed to a group average of 1 to enable comparison. Also here the bus organizations are ranked twice. From left to right operators are ranked based on passenger kilometres per actual revenue vehicle service planning capacity kilometre. Operators are ranked alphabetically based on passenger kilometres per actual revenue vehicle kilometre.

Operator A is the same organization in both Figure 2 and Figure 3. As shown in Figure 2, operator A has the largest average weighted vehicle capacity amongst the IBBG bus operators. In the average bus load KPI operator A scores highest; i.e. transporting more passengers per vehicle than other organizations. However, when normalized for the actual weighted revenue vehicle capacity of the fleet operator A is ranked 6th. In other words, five organizations see more of their offered service planning capacity used by passengers than operator A.

Overall the conclusion is that service planning capacity kilometres normalizes well for vehicle size, not only from an operational point of view, but also from a customer service/policy point of view.

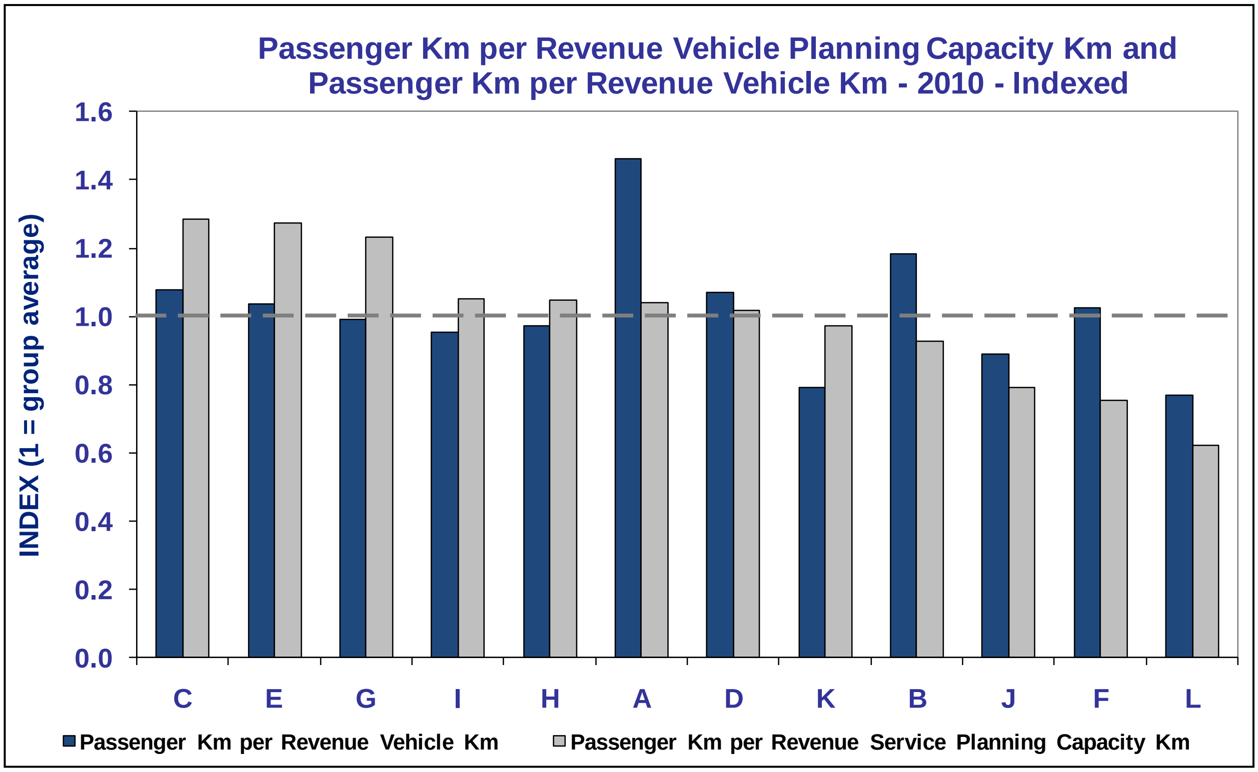

FIGURE 3 Average Bus Load versus Average Bus Planning Capacity Used

6.2 Normalizing for Vehicle Weight

Due to regional differences in vehicle specifications some bus organizations operate fleets that are considerably heavier than fleets used in other cities. From Table 1 the range of average weighted vehicle weight data, could be seen, i.e. a minimum value of 11.2 tonne and a maximum value of 14.9 tonne. These numbers relate to unloaded vehicle axle weights.

The IBBG operators were interested to understand the effect of these weight differences on the relative performance of bus operators in the different fuel economy KPIs, the hypothesis being that heavier buses would be less fuel efficient. The original fuel economy KPI used is: fuel type (e.g. diesel) used per total fuel type (e.g. diesel) vehicle planning capacity kilometres. As discussed in 6.1, vehicle planning capacity kilometres is the preferred IBBG denominator to normalize for vehicle size. The pitfall here is that size differences do not necessarily correlate to weight differences. To help understand the impact of weight the same fuel use is therefore also normalized by total weighted tonne kilometres. Figure 4 combines the results of these two KPIs, indexed to a group average of 1 to enable comparison.

The bus organizations are ranked in two ways. From left to right operators are ranked based on diesel fuel used per diesel vehicle tonne kilometre. Operators are ranked alphabetically based on diesel fuel used per diesel vehicle service planning capacity kilometre.

Operators A and C have the highest average weighted vehicle weight in the sample, operator K has the lowest average weighted vehicle weight in the sample. When normalized for capacity kilometres operator A is very fuel inefficient compared to peers and operator C also performs significantly worse than the group average. However, when normalized for tonne kilometres, operator C moves from the 3rd least fuel efficient operator, to the 3rd most fuel efficient operator. The fuel economy of operator A, although still the highest, is much more in line with other peers when normalized by total tonne kilometres.

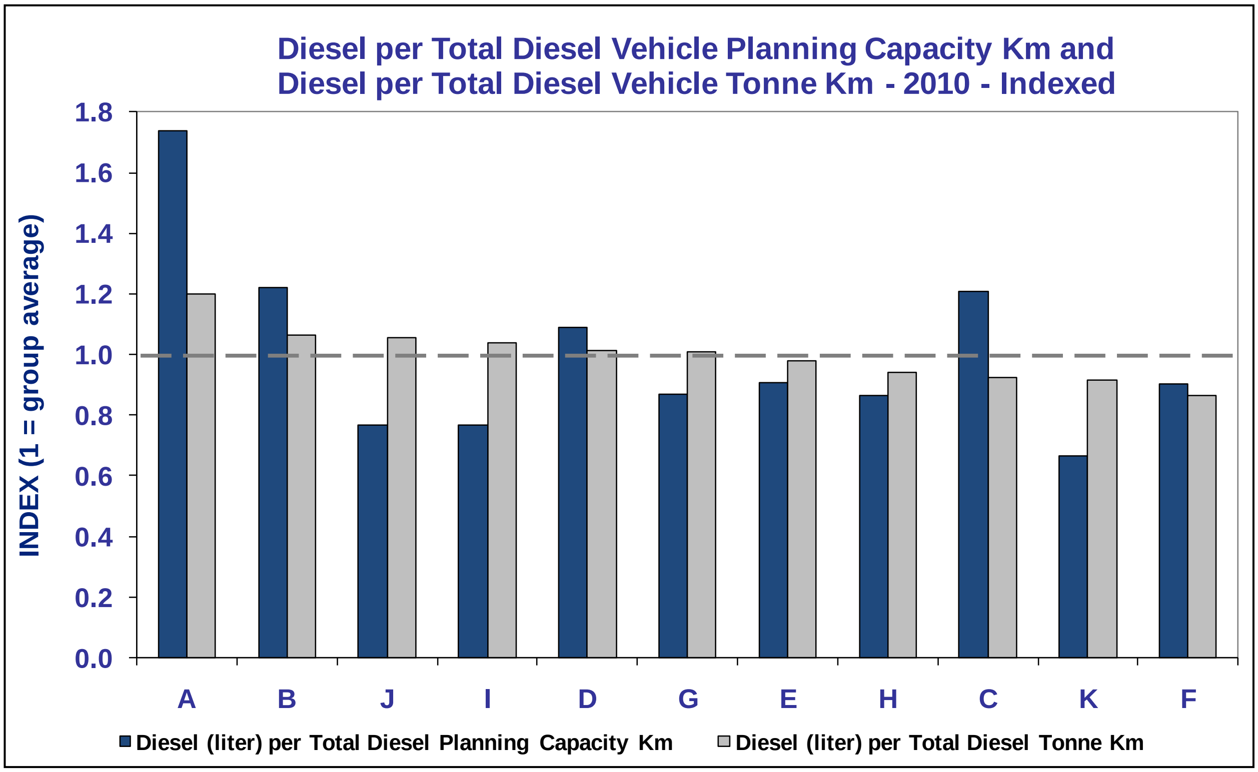

FIGURE 4 Impact of Using Total Weighted Actual Tonne Kilometres as a Normalization Factor instead of Total Weighted Actual Revenue Capacity Kilometres

Overall the variability of the sample in both KPIs has changed. The coefficient of variation for the fuel efficiency per capacity kilometres is 0.30, while the fuel efficiency per tonne kilometre has a CV of 0.13. When normalising for vehicle weights, the fuel efficiency performance between operators is more similar.

From Figure 4 it can therefore be concluded that that vehicle weight has a significant impact on total fuel used. Total vehicle planning capacity kilometres, a proxy for vehicle size, only partly normalizes for weight. The use of total tonne km as a denominator has improved the comparability of different fuel economy key performance indicators by fully normalizing for weight.

It is important to reiterate that, for a ‘full understanding’ it remains necessary to also review fuel economy in terms of capacity kilometres supplied or even passenger kilometres produced. Those KPIs give a different, but certainly not less valid view of relative fuel efficiency performance, especially if one keeps in mind that vehicle weight specifications can be changed in the medium to long term.

7. CONCLUSIONS

The thirteen members of the International Bus benchmarking Group (IBBG) are all urban bus operators in large cities. They consider themselves comparable peers and successfully learn from each other through the exchange of best practices. Still, significant variety of operating characteristics has been observed, especially with respect to (in order of variability): passenger trip length, % of dead running, commercial speed, and vehicle capacity. This research shows that a considerable variability in key service characteristics has implications for normalized performance comparison, i.e. benchmarking. For example, passenger kilometres, a frequently used denominator, is defined by the average trip length and the number of passenger boardings. The analysis showed that the shortest average trip length observed is just 34% of the longest average trip length in the sample. Therefore using only either passenger kilometres or passenger boardings as a normalization factor, the relative performance of a bus operator observed in that single key performance indicator can be skewed.

This paper aims to illustrate to benchmarking practitioners and policy makers that one key performance indicator often only answers part of a question. If your organization has average values in all four variable service characteristics mentioned above, the choice of denominator (for example vehicle kilometres or vehicle hours) will not affect the relative position to peers in the comparison, just which organizations are ranked on either side of you. However, if an organization has extreme values in any of the four service characteristics identified, then the relative rank of that organization will significantly differ depending on the normalization factor used. To understand the ‘full picture’, two KPIs using complementary normalization factors (such as vehicle kilometres and vehicle hours) need to be produced and simultaneously analyzed.

The literature review confirmed that seat capacity kilometres is frequently used or proposed as a denominator to normalize for fleet capacity. The data from the IBBG shows that in practice the average weighted seating capacity per bus is not the most effective normalization factor for vehicle capacity or vehicle size due to the different types of vehicles operated by operators. Double-decker and mini/midi buses have a high proportion of seats and a small amount of additional standing capacity, while standard and especially articulated buses have relatively few seats compared to the total capacity offered. When using (only) seat kilometres, a normalization discrepancy occurs due to the fact that the fleet composition differs between bus operators. To overcome this, the IBBG now successfully uses vehicle service planning capacity. This denominator is calculated as the weighted average number of seats per bus plus the additional weighted average number of standees that each operator plans for.

It can also be concluded from this study that due to regional differences in vehicle specification, some bus organizations operate fleets that are considerably heavier than fleets used in other cities. This paper showed that the weight of vehicles has a significant impact on total fuel used. Total vehicle planning capacity kilometres, a proxy for vehicle size, only partly normalizes for weight. The use of total tonne km as a denominator, however, has improved the comparability of different fuel economy key performance indicators by fully normalizing for weight.

ACKNOWLEDGEMENTS

The authors thank the members of the International Bus Benchmarking Group for their willingness to share the normalization improvements learned though the benchmarking group, with the wider transport industry. We also acknowledge the work of Richard Anderson and Eric Randall who played important roles in the development of the IBBG and thank Nicholas Findlay, Alexander Barron, Thomas Lyscom and Judith Cohen for their useful comments and their diligence in maintaining such a high quality data set.

REFERENCES

(*1*) Lema, N. and A. Price. Benchmarking: performance improvement towards competitive advantage. *Journal of Management in Engineering*, Vol 11, 1995, no 1, pp. 28-37.
(2) Trompet, M., R.J. Anderson and D.J. Graham. Variability in Comparable Performance of Urban Bus Operations. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2111, Transportation Research Board of the National Academies, D.C., 2009, pp.177-184.
(3) The Transit Cooperative Research Program (TCRP). *Report 141: A Methodology for Performance Measurement and Peer Comparison in the Public Transportation Industry*, Transportation Research Board of the National Academies, Washington, D.C., 2010.
(4) EQUIP (Extending the Quality of Public Transport) consortium. *State-of-the-art in Benchmarking of Quality in Public Transport.* Report delivered to the DGTREN, Brussels, March 2000.
(5) Geerlings, H., R. Klementschitz and C. Mulley. Development of a methodology for benchmarking public transportation organisations: a practical tool based on an industry sound methodology. *Journal of Cleaner Production*, 14, 2006, pp.113-123.
(6) Gudmundsson, H., A. Wyatt and L. Gordon. Benchmarking and Sustainable Transport Policy: Learning from the BEST Network. *Transport Reviews*, Vol 25, No. 6, November 2005, pp. 669-690.
(7) Mulley, C. Improving efficiency as a means to improving transport quality. *Proceedings of the Institution of Civil Engineers: Municipal Engineer*, 157, 2004, pp. 17-24.
(8) EQUIP consortium. *The Benchmarking Handbook*. Transport Programme Project No. UR-98-RS.3076, Contract Report 4, Report delivered to the DGTREN, Brussels, August 2000 .
(9) Hensher, D. and R. Daniels, Productivity Measurement in the Urban Bus Sector. *Transport Policy*, Vol. 2., 1995. pp. 179 -195.
(10) Fielding, G., T. Babitsky and M. Brenner. Performance evaluation for Bus Transit. *Transportation Research Part A*, Vol. 19A, No. 1, February 1985, pp73-82.
(11) Fielding, G., M.E. Brenner and K. Faust. Typology for Bus Transit. *Transportation Research*, Vol. 19A, No. 3, 1985, pp. 269-278.
(12) Stappenhorst, T. *The Benchmarking Book: A How-to-Guide to Best Practice for Managers and Practitioners.* Butterworth-Heinemann, Oxford, 2009.
(13) Phillips, J. An Application of the Balanced Scorecard to Public Transit System Performance Assessment. *Transportation Journal*, Vol. 43, No. 1, 2004, pp. 26-55.
(14) The Transit Cooperative Research Program (TCRP). *Report 88: A Guidebook for Developing a Transit Performance-Measurement System*. Transportation Research Board of the National Academies, Washington, D.C., 2003.
(15) ISOTOPE (Improved structure and organisation for transportation operations of passengers in Europe). *Final report by ISOTOPE consortium on behalf of European Commission DGVII.* Luxembourg, Office for Official Publications of the European Communities, 1998.
(16) MARETOPE (Managing and Assessing Regulatory Evolution in local public Transport Operations in Europe). *MARETOPE Handbook, Transport Programme Project No. 2000 – RD.10668.* Report delivered to the DG TREN, Brussels, 2003.
(17) MVA Limited, *London Buses Strategic Review Comparable Cities Study: Final Report.* August 2003
(18) Urban Transport Benchmarking Initiative, *Annex A1: Review of the Common Indicators from the Urban Transport Benchmarking Initiative.* Report delivered to the DG TREN, Brussels, July 2004.
(19) Fielding, G., R. Glauthier and C. Lave. Performance Indicators for Transit Management. *Transportation*, Vol. 7, 1978, pp. 365-379.
(20) Badami, M. and M. Haider. An Analysis of Public Transit Performance in Indian Cities. *Transportation Research Part A*, 41, 2007, pp. 961-981.
(21) Mackie, P. and C. Nash. Efficiency and Performance Indicators: the case of the bus industry. *Public Money*, December 1982, pp. 41-44.
(22) The Transit Cooperative Research Program (TCRP). *Report 100: Transit Capacity and Quality of Service Manual: 2nd Edition – PART 4 Bus Transit Capacity*, Transportation Research Board, National Academy Press, Washington, D.C, 2003.