Benchmarking the operations and maintenance of New Zealand's roading sector

January 2017

SB Costello, TFP Henning and H Shivaramu
The University of Auckland

NZ Transport Agency research report 605
Contracted research organisation – Auckland UniServices Limited

ISBN 978-1-98-851204-4 (electronic)
ISSN 1173-3764 (electronic)

NZ Transport Agency
Private Bag 6995, Wellington 6141, New Zealand
Telephone 64 4 894 5400; facsimile 64 4 894 6100
research@nzta.govt.nz
www.nzta.govt.nz

Costello, SB, TFP Henning and H Shivaramu (2017) Benchmarking the operations and maintenance of New Zealand's roading sector. *NZ Transport Agency research report 605.* 59pp.

Auckland UniServices Limited was contracted by the NZ Transport Agency in 2014 to carry out this research.

Keywords: benchmarking, highways, maintenance, performance management.

An important note for the reader

The NZ Transport Agency is a Crown entity established under the Land Transport Management Act 2003. The objective of the Agency is to undertake its functions in a way that contributes to an efficient, effective and safe land transport system in the public interest. Each year, the NZ Transport Agency funds innovative and relevant research that contributes to this objective.

The views expressed in research reports are the outcomes of the independent research, and should not be regarded as being the opinion or responsibility of the NZ Transport Agency. The material contained in the reports should not be construed in any way as policy adopted by the NZ Transport Agency or indeed any agency of the NZ Government. The reports may, however, be used by NZ Government agencies as a reference in the development of policy.

While research reports are believed to be correct at the time of their preparation, the NZ Transport Agency and agents involved in their preparation and publication do not accept any liability for use of the research. People using the research, whether directly or indirectly, should apply and rely on their own skill and judgement. They should not rely on the contents of the research reports in isolation from other sources of advice and information. If necessary, they should seek appropriate legal or other expert advice.

Acknowledgements

The authors would like to thank the Steering Group members: Liane Powell (NZ Transport Agency, chair), Andrew Cooper (ex NZ Transport Agency, chair), David Kelsey (ex NZ Transport Agency), Murray Gimblett (NZ Transport Agency), Philip Blagdon (NZ Transport Agency), Ian Duncan (Ministry of Transport) and Craig Thew (Hastings District Council), as well as the peer reviewers: Professor Susan Tighe (University of Waterloo) and Michael Mason (Auckland Transport).

The authors would also like to thank their colleagues Professor Paul Rouse and Dr Andrea Raith, for their insights into data envelopment analysis, and OPUS International Consultants staff Gregg Morrow, Anna Robak and Elke Beca for their contributions – in particular for collection of the international benchmarking data.

Contents

Executive summary

The first part of this research project aimed to provide an analysis of the suitability of existing benchmarking methodologies for use in the highway operations and maintenance sector in New Zealand. Ideally this should be a translatable best practice benchmarking methodology, rather than creating a bespoke model. In addition, the recommended methodology needed to translate performance, quality and cost into a level of service and value-for-money equation-based measure that could be compared across the New Zealand roading sector. The recommended methodology also needed to normalise for unique network characteristics, outside the control of the supplier/maintainer, that might impact on cost and quality.

The recommended value-for-money equation-based measure is based on a generic framework for designing management control systems in not-for-profit organisations. The report details how it could be adapted for use in the highway maintenance and operations sector. The final framework includes measures of cost or expenditure on highway maintenance and operations, achievement in terms of quantity of maintenance and operations work undertaken and performance or quality of the service provided. It should also be noted that the framework aligns with the NZ Transport Agency's value-for-money objectives and the accountability notions of effectiveness, efficiency and economy therein.

A review of existing benchmarking methodologies used in highway operations and maintenance services, the wider transportation sector and similar industries both in New Zealand and overseas was undertaken to determine candidate benchmarking techniques. These included the partial efficiency measure, total factor efficiency measure, regression analysis and data envelopment analysis (DEA). An analysis of their suitability for use in the roading sector in New Zealand was then undertaken. Each candidate method had advantages and disadvantages to their use in the areas of their ability to handle multiple inputs or outputs, choice of weights, the format of the benchmark produced, their method for dealing with unique network characteristics, the complexity of the technique and usefulness of the outputs. DEA was recommended due to its ability to incorporate multiple inputs or outputs, the optimisation of weights as part of the analysis, the production of an efficient frontier of best performers, its ability to normalise for such unique network characteristics, the usefulness of the outputs and the fact that it has been shown to work in the highway maintenance and operations sector. In such cases, DEA has been supported by other analysis techniques, such as the analytic hierarchy process, regression analysis or peer groups, where appropriate. The only major disadvantage of DEA is its complexity; however, this is offset by the fact that the outputs of the analysis are much easier to explain to decision makers.

If adopted, such a model will be able to compare operations and maintenance costs and performance between networks within New Zealand and will have the potential to do so against similar overseas organisations. It is not envisaged that the resulting model will be ubiquitous in the same way as the Road Assessment and Maintenance Management (RAMM) database. Instead, the model is seen as a high level management tool to help inform NZ Transport Agency (the Transport Agency) policy on issues such as procurement approaches and funding, as well as to help drive efficiency improvements within the regions, territorial local authorities and service providers. However, it is worth noting that the potential efficiency improvements identified by the benchmarking will only be achieved through management's use of this information.

The second part of this research project aimed to collect benchmarking data from two overseas road agencies, to both assess the availability and ease of collection of such data and to enable initial comparisons to be undertaken with the New Zealand roading sector, should the Transport Agency wish to do so. To retain anonymity of the participating road agencies they are referred to simply as international

comparator A and international comparator B, in the report. Contextual data, data on inputs (expenditure), outputs (achievement) and outcome indicators (performance) are provided as per the Transport Agency's value-for-money framework. All expenditure is reported in New Zealand dollars purchasing power parity equivalents. In collecting the benchmarking data from two overseas road controlling authorities, significant challenges were faced including a lack of timely cooperation, composed of delays due to obtaining approvals to release the data followed by delays in interrogating the road and financial databases, as well as differences in performance measurement, definition of maintenance tasks and accounting systems. These challenges are in line with international experiences in this area.

Some basic ratio comparisons between the two overseas road agencies were undertaken, akin to the partial efficiency measure mentioned above. This resulted in a number of individual input-to-output ratios. This is a major drawback of such an approach in that each ratio only tells part of the efficiency story. As a decision maker it is very difficult to draw a definitive conclusion from such partial information, even if all possible partial efficiency ratios are produced. Although all too often the different ratios will provide conflicting information, in this case international comparator B consistently outperformed international comparator A, for the ratio measures reported, by virtue of the fact that they spent less per network length, per lane kilometre, per vehicle kilometre travelled, per head of population and per land area. However, highlighting another major drawback of this approach, such basic ratios do not take into account the service provided or performance. Although it is evident that although international comparator A commits more expenditure to pavement maintenance they also provide better performance than international comparator B – as measured by the international roughness index and wheel track rutting, the only comparable performance measures. In addition, the operating conditions are not taken into consideration. For example, international comparator A has higher rainfall and lower temperatures to contend with. In particular, a sizeable proportion of the routine maintenance budget is dedicated to winter maintenance. Finally, only one normalising factor is considered in each of the basic ratio comparisons. This further demonstrates the need to adopt a best practice benchmarking technique, such as DEA.

If international benchmarking is to be implemented, then the recommendation is to first form a benchmarking club of similarly committed road agencies and then to agree on (or adopt existing) data processing standards and metadata standards. Benchmarking at the national level in New Zealand presents far fewer challenges than international benchmarking, particularly at the state highway level. The relative ease of availability, and potential for comparability, of performance and benchmarking data for New Zealand's state highway networks would suggest that benefits from benchmarking can be realised much earlier than from an international benchmarking exercise. Although not without its own challenges, standardisation of data, if not actually metadata standards, already exists in the form of the RAMM database. Some data processing standards also exist (eg calculation of smooth travel exposure). A top down approach whereby the final performance, expenditure and contextual measures are the starting point and only the data required to calculate these measures is harmonised, is recommended.

Finally, although DEA has been recommended as the technique with which to develop a benchmarking model for the operations and maintenance of New Zealand's road network, considerable further research needs to be undertaken to 'build' the benchmarking model. The suggested next steps and methodology for further development are included in the report.

Abstract

This research project aimed to provide an analysis of the suitability of existing benchmarking methodologies for use in the roading sector in New Zealand. The chosen methodology needed to normalise for unique network characteristics outside the control of the maintainer that might impact on cost and quality. Data envelopment analysis, supported by other analysis techniques, was recommended due to its ability to normalise for such characteristics and the fact that it has been shown to work in the highway maintenance and operations sector.

In addition, this research project aimed to collect benchmarking data from two overseas road agencies, to both assess the availability and ease of collection of such data and to enable initial comparisons to be undertaken with the New Zealand roading sector should the NZ Transport Agency wish to do so. Benchmarking data from two overseas road controlling authorities was collected; however, significant challenges were faced with collection of the data including a lack of timely cooperation, composed of delays due to obtaining approvals to release the data followed by delays in interrogating the road and financial databases, as well as differences in performance measurement, definition of maintenance tasks and accounting systems. These challenges are in line with international experiences in this area.

1 Introduction

1.1 Background to this project

In March 2014, Auckland UniServices Limited was contracted by the NZ Transport Agency ('the Transport Agency') to undertake a research project into benchmarking the operations and maintenance of New Zealand's roading sector.

The research project was split into two parts, with a stage gate review in between. Part 1 involved a review of relevant literature and practices to inform the Transport Agency on current benchmarking state-of-the-practice methodologies and to understand how such methodologies might be applied to the roading sector in New Zealand. Part 2 initially involved the provision of a schema for the design and development of a benchmarking framework for the highway maintenance and operations sector in New Zealand, assuming an existing and appropriate benchmarking framework was discovered in part 1. While an existing and appropriate benchmarking framework was discovered in part 1, the Transport Agency felt that greater benefit would be gained, at this stage, from the collection of performance and benchmarking data from two overseas road agencies to act as international comparators. Consequently, part 2 was revised accordingly.

1.2 Problem statement

The Transport Agency is responsible for maintaining New Zealand's state highway network. It comprises almost 11,000km of carriageway, requiring annual expenditure of approximately $500 million to operate and maintain. While state highways represent only 11% of New Zealand's roading network, they carry half of the country's traffic (NZ Transport Agency 2016).

The local road network in New Zealand is almost 84,000km in length. Local road controlling authorities (RCAs) are responsible for the maintenance of the local road network within their boundary. Central government, through the Transport Agency, funds a portion of each RCA's expenditure on road infrastructure and transportation facilities. Referred to as the funding application rate, this ranges from 43% to 67%, with the national average being 50%. The RCA must provide the balance of the funding from rates or loan-financed (NZ Transport Agency 2016).

Thus, central government through the Transport Agency has a vested interest in ensuring that the funds provided for both state highways and local roads offer value for money, as measured against its goals and priorities for land transport. To this end, the Transport Agency sees performance monitoring and benchmarking as an opportunity to deliver better value for money.

Desirable components of performance monitoring systems in the public sector were defined by Altman (1979) as follows:

- a data component which captures and processes data
- an analysis component which translates data into information
- an action component which refers to management's use of this information.

All three are required to deliver performance improvement. Referring to the first component, many organisations tend to have an abundance of performance data and measures which are captured and processed in a database of some description. Road agencies are no exception. Unfortunately, referring to the second component, while performance management of the state highway and local road networks is well embedded in New Zealand, the ability to benchmark the performance of road agencies and contractors in delivering the performance outcomes is complicated by the variability across state highway regions and RCAs. Such variability includes the annual expenditure, length of the network, traffic loading, geology and the environment. All of the foregoing can influence the efficiency, effectiveness and economy outcomes embedded within the Transport Agency's value-for-money framework. Direct comparison of costs, quantities and quality is, therefore, not a satisfactory way to benchmark performance.

The development of an analysis component that can translate data into useable, in this case comparable, information is therefore required. Such a tool would allow benchmarking across state highway regions or RCAs and, in doing so, identify those exhibiting best practice. The use of such information by management, the third of the above components of performance monitoring, will then be possible. Through the development of networks and the consequent information exchanges between peers, a powerful process of performance improvement can be made to work.

1.3 Objectives of the research

Within the Transport Agency's wider aim of delivering better value for money, the objectives of part 1 of this research project were to:

- document existing benchmarking methodologies for highway operations and maintenance services currently in use in other roading agencies and similar industries both in New Zealand and overseas
- provide analysis on the suitability of existing benchmarking methodologies for use in the roading sector in New Zealand
- discuss the high-level drivers from benchmarking of other jurisdictions and/or sectors that are evident in the best practice performers, opportunities for benchmarking, and barriers and risks to benchmarking.

The objective for the revised part 2 of this research project was to:

- collect benchmarking data from two overseas roading agencies, identified in consultation with the Transport Agency. The purpose of collecting this data was twofold, first to assess the availability and ease of collection of such data and, second, to enable initial comparisons to be undertaken with the New Zealand roading sector should the Transport Agency wish to do so.

1.4 High-level drivers for benchmarking

There are two main high-level drivers for undertaking benchmarking in the infrastructure area, as follows:

1. To better understand the overall performance or efficiency of complex systems with multiple input and output variables such as highway maintenance and operations.
2. Infrastructure owners sometimes have to establish appropriate levels of investment or service for infrastructure having little information from which to establish what an appropriate level might be. Benchmarking with similar agencies, cities or countries often helps to determine such information. In defending the decision, the argument is that the levels are based on appropriate practices elsewhere.

There are a significant number of associated benefits from the benchmarking process that also make it worthwhile to undertake, including:

- The organisational learning through benchmarking processes can be a major benefit. Intuitive and anecdotal views on organisational or system performance are replaced with tangible facts.

- It exposes areas of poor asset management practice, under-investment, neglect or under-performance. It is natural for an organisation to focus on parts of the business that are perceived as urgent or important. However, benchmarking can report on a wider and holistic spectrum of performance aspects and often it directs attention to some less obvious problem areas.
- Benchmarking often requires the collection or acquisition of data from external sources, such as Statistics New Zealand, that could significantly enrich knowledge of a system's performance. Consequently, the asset management team starts considering potentially valuable data that they have not heretofore considered. For example, external data is often used to normalise indices and/or to be used as contextual information. Considering this type of information for a specific region is a significant learning opportunity in itself.

High-level drivers for benchmarking from studies within New Zealand are discussed below:

- In response to the *Government policy statement on land transport funding 2011/12 – 2021/22* (GPS 2012) (MoT 2011), the Road Maintenance Task Force[1] was commissioned with four improvement areas: procurement, asset management planning, collaborations and performance management. The objective of the Waugh Infrastructure Management Ltd (2012) better planning report was to question current practices in asset management. In particular, the primary challenge was 'to consider the hypothesis: "If … we (the sector) ensure that all road network management units are making sound road asset management decisions then the above action will lead to an improvement in efficiency, effectiveness and whole of life value for money in delivery of road maintenance operations and renewals".' The problem definition statement was defined as follows: 'Currently there is a perception that there is an asset management capacity and capability gap within the sector so sub-optimal programmes are being delivered'.
  - Key finding number 3 of the Waugh Infrastructure Management Ltd (2012) report states: 'Encourage and provide leadership to enable study teams and technical working parties to identify and implement more efficient and effective maintenance practices'.
- A key tool for addressing this recommendation would be through a benchmarking process to:
  - identify struggling authorities
  - understand appropriate aspirational targets for achieving road efficiency.
- The main motivation behind the GPS 2012 was the increasing cost of maintenance during the past decade without a noticeable improvement in level of service outcomes. Arguably there are reasons for increasing cost that may include construction cost escalation and increasing traffic loading on roads. Yet, an increase of 50% over the past nine years naturally questions the efficiencies within the road maintenance industry. The only way to explain the trends effectively would be to start benchmarking efficiencies in the sector.

1.5 Benchmarking opportunities

Benchmarking provides a number of opportunities for the Transport Agency and RCAs to help improve their understanding of performance in the highway maintenance and operations sector, as follows:

- The introduction of the network outcome contracts (NOCs) provides a unique opportunity for the Transport Agency to benchmark the performance of service providers over a contract period, thereby

[1] See www.nzta.govt.nz/resources/rmtf-report/ for information on the task force.

helping to inform future decisions on the awarding of contracts. Previous attempts at benchmarking across different procurement methods have made relative comparisons difficult.

- Historically, the state highway network has been managed under a number of different contract forms, for example, performance specified maintenance contracts (PSMCs), alliance and hybrid contracts. Benchmarking provides an opportunity to compare the performance of these historical contracts with the NOCs currently being implemented. Such knowledge would help confirm, or otherwise, the move to the new contract type. Other variables of interest in such a comparison include the performance of service providers under different contract forms.
- There is a noticeable variation in maintenance strategies and regimes across RCAs. Benchmarking the performance of RCAs will help to determine best practice within the local authority sector, thereby providing opportunities for improvement in practices, contract forms, or others such differentiators (particularly between members of the same peer groups)
- There are a number of specific maintenance practices undertaken by authorities that have become accepted practice due to simplicity or cost considerations. However, few authorities have undertaken retrospective analysis to determine the true efficiency outcomes from these practices. Benchmarking will be useful in identifying the specific practices which provide value for money and, indeed, those that do not.

1.6 Barriers and risks to benchmarking

Potential barriers and risks to benchmarking include:

- Differences in task definitions, practices and accounting systems, in particular in international benchmarking exercises, make it difficult to collect information on road maintenance and operation expenses in a way that allows for comparisons.
- Lack of commitment from participating agencies can be a barrier, in particular in international benchmarking exercises where participation relies on good will.
- Appropriate data availability is often held up as a barrier to benchmarking. However, a stage-wise adoption will allow benchmarking to start on the basis of available data, thereby overcoming this barrier. In addition, data required for benchmarking is equally required for the adoption of proper asset management. Consequently, a natural outcome from benchmarking is the opportunity to improve data collection and reporting processes.
- Incorrect data can lead to incorrect conclusions in terms of relative efficiency and effectiveness, consequently care should be exercised in validating any data used for benchmarking.
- An extension of the previous point, but care must also be taken during time-based monitoring to ensure validity of observations. For example, an improvement or decline in performance could in fact be the result of a change in condition data collection contractor or a change in the apparatus used to collect the data.

1.7 Report layout

Chapter 1 provides background to the research, defines the problem statement, objectives of the research, discusses high-level drivers for benchmarking, benchmarking opportunities, and barriers and risks to benchmarking, as well as outlining the layout of the report.

Chapter 2 provides an overview of a generic performance management and benchmarking framework, specifically designed for not-for-profit organisations such as public infrastructure providers, and how it could be adapted to align with the Transport Agency's value-for-money framework.

Chapter 3 provides a review of the relevant benchmarking literature in an effort to identify examples of best practice benchmarking used in industry generally, and more specifically in the roading sector.

Chapter 4 analyses the benchmarking techniques identified in chapter 3 and comments on their suitability for application in the roading sector in New Zealand.

Chapter 5 documents the data collected from the two overseas road agencies. Such data includes expenditure, achievement and performance data, as well as contextual data.

Chapter 6 discusses the high-level drivers for benchmarking that are evident in the best practice performers, identifies benchmarking opportunities for New Zealand roading authorities to implement, details the barriers and risks to benchmarking, and discusses international comparators for benchmarking.

Chapter 7 concludes the research and recommends a benchmarking strategy for the roading sector in New Zealand based on benchmarking best practice.

2 Measuring value for money in highway operations and maintenance

2.1 Introduction

In response to increasing public scrutiny on the performance and public accountability of not-for-profit organisations, Ramanathan (1985) proposed a generic framework for designing management control systems in not-for-profit organisations. While the presented management control system, or performance framework, is general enough to be applicable to a variety of not-for-profit organisations, it is acknowledged that considerable adaptations are required to tailor the framework to any individual organisation. The developed framework has over time become the basis for the design of numerous management control systems in not-for-profit organisations. This section, adapted from Costello et al. (2014), describes the generic framework and how it could be adapted for use in the highway maintenance and operations sector. It should also be noted that the framework aligns with the NZ Transport Agency's value-for-money objectives and the accountability notions of effectiveness, efficiency and economy therein.

2.2 Performance levels

The various levels in the generic framework are summarised in table 2.1. The key challenges in adapting this framework further, to specifically meet the requirements of the highway maintenance and operations sector, are in defining the benefits, outcome indicators, outputs, inputs and costs against which effectiveness, efficiency and economy can be measured. The levels and how they apply to highway maintenance and operations are elaborated on in the following sub-sections.

Table 2.1 Levels in the generic framework

Level	Description
Benefits (B)	Desired high-level outcomes, usually set by government, for the services provided by an agency
Outcome indicators (OC)	Surrogate measures of the outcomes provided by an agency
Outputs (O)	Various measures of activity undertaken by an agency, in order to provide its services
Inputs (I)	Non-financial measures of various types of resources consumed by an agency
Costs (C)	The financial value of all resources consumed by an agency in order to provide its services

2.2.1 Benefits

Outcomes are usually set by the government, and represent the high-level social benefits that an agency provides through the delivery of its services. These benefits are often difficult to quantify and may have several dimensions, resulting in multiple indicators to represent the outcomes. In addition, many parts of an agency contribute to each outcome, thereby making it difficult to quantify the individual contribution from each. External influences can also impact on the outcomes.

The GPS 2012 set out the specific impacts the government expected to be achieved through the use of the National Land Transport Fund. These are included in table 2.2 and are typical of high-level outcomes.

Table 2.2 Impacts the government wished to achieve

Type of impact	Impacts
Impacts that contribute to economic growth and productivity	Improvements in the provision of infrastructure and services that enhance transport efficiency and lower the cost of transportation
	Better access to markets, employment and areas that contribute to economic growth
	A secure and resilient transport network
Other impacts	Reductions in deaths and serious injuries as a result of road crashes
	More transport choices, particularly for those with limited access to a car where appropriate
	Reductions in adverse environmental effects from land transport
	Contributions to positive health outcomes

It should be noted that while all desired impacts are presented in table 2.2 for completeness, not all are directly influenced by highway maintenance and operations. In addition, not all are solely influenced by such funding activities. Taking 'Reductions in deaths and serious injuries as a result of road crashes' as an example, it is clear that highway maintenance and operations directly influences this outcome through, for example, improvements to skid resistance. However, highway maintenance and operations is not the sole contributor to reductions in road crashes as policing, licensing and advertising campaigns all have their part to play. Consequently, defining measures that relate solely to highway maintenance and operations is key if accountability is to be apportioned correctly.

2.2.2 Outcome indicators

Outcome indicators provide a convenient, albeit imperfect, substitute for the outcomes for the reasons discussed above. These allow a review of the performance of an agency, or parts of an agency, to be assessed on its own merits. In short, outcome indicators represent surrogate measures of the outcomes provided by the agency.

Typical outcome indicators in highway maintenance and operations include:

- Surface condition index (SCI) – a combination of the surface faults in the sealed road surface. SCI combines alligator cracking, scabbing, potholes, pothole patches and flushing, into a weighted index.
- Pavement integrity index (PII) – a combination of the pavement faults in sealed roads. PII combines the elements of the SCI with rutting and shoving, into a weighted index.
- Road roughness – a measure of the undulations in the road. Values are obtained using a laser profilometer attached to a high-speed data collection vehicle and are reported using the National Association of Australian Road Authorities counts or international roughness index (IRI).
- Smooth travel exposure (STE) – measures the proportion of vehicle kilometres travelled (VKT) in a year on a 'smooth' road. A smooth road is one that has a road roughness value under a predetermined threshold.
- Rutting – longitudinal depressions in the wheel paths measured in mm. Typically obtained using a transverse laser-based rut bar attached to a high-speed data collection vehicle.
- Texture – macrotexture of the pavement surface, reported as mean texture depth (MTD) in mm. Obtained using a high frequency laser attached to a high-speed data collection vehicle.

- Skid resistance – a measure of the friction of the road surface. It is typically measured using the sideway-force coefficient routine investigation machine and is reported as the sideways force coefficient.

In addition, crash data and the associated social cost of crashes is also available. However, as discussed previously the contributors to this extend beyond the highway maintenance and operations sector.

2.2.3 Outputs

Outputs represent various measures of the volume of activity without regard to whether they lead to successful outcomes. Hence, an increase in output may not necessarily lead to an improvement in outcomes (or outcome indicators), unless they are appropriately targeted. Outputs are generally defined in physical units.

Typical output measures, similar to those available from the Transport Agency achievement reports, are listed below:

- length of road resurfaced
- length road rehabilitated
- length of road reconstructed.

However, due to typical reporting systems for general, or routine, maintenance a true measure of output is often difficult to quantify. Consequently, Rouse et al (1997) proposed the SCI (which is more in the nature of an outcome) and expenditure per network km (which is more in the nature of a cost) as possible surrogates, neither of which they recognised as ideal. This is an area of improvement across the RCA sector in New Zealand which could significantly improve the value of benchmarking.

2.2.4 Inputs

Inputs represent non-financial measures of various types of resources consumed by the agency. In theory, increases/decreases in inputs will result in similar variations in outputs. However, they may not always be strictly proportional, especially where economies of scale are introduced.

Typical measures of various types of resources consumed in delivering highway maintenance and operations services are listed below:

- labour (skilled and unskilled)
- materials
- plant and equipment.

2.2.5 Costs

Costs represent the financial value of all resources consumed by an agency in order to provide its services. In theory, increases/decreases in input will result in similar variations in costs. However, they may not always be strictly proportional especially where economies of scale are introduced. In New Zealand, such costs are recorded within the funding categories set out in the Planning and Investment Knowledge Base (NZ Transport Agency 2011).

2.3 Performance controls

The benefits, outcome indicators, outputs, inputs and costs represent a hierarchy of performance controls which allow an agency to measure its performance, or the performance of parts of its organisation. The resulting control framework integrating the measures and surrogates into a benefit/cost ratio is included in the equation below:

$$\frac{B}{C} = \frac{B}{OC} \times \frac{OC}{O} \times \frac{O}{I} \times \frac{I}{C} \quad \text{(Equation 2.1)}$$

where B represents benefits, OC represents outcome indicators, O represents outputs, I represents inputs and C represents costs.

However, noting that benefits have been replaced by outcome indicators, this yields just three ratios on the right-hand side, as follows:

$$\frac{OC}{C} = \frac{OC}{O} \times \frac{O}{I} \times \frac{I}{C} \quad \text{(Equation 2.2)}$$

where the notation is as previously described.

The various performance controls are included in table 2.3, along with what they are measuring, and the parts within the agency whose performance is being measured.

Table 2.3 Traditional performance controls

Measure	Description
OC/C	Composite measure of the road agency's performance in attaining its outcomes, represented by the surrogate outcome indicator. This is a high-level measure for top management. This is termed 'value for money in the Transport Agency's lexicon.'
OC/O	Measure of the road agency's quality or success in attaining its outcomes. This measures the effectiveness of decision making by professionals, such as asset managers. This is termed 'effectiveness' in the Transport Agency's lexicon.
O/I	Measure of the ability of the road agency's operations department to utilise their physical resources efficiently. This is termed 'efficiency' in the Transport Agency's lexicon.
I/C	Measure of the ability of the road agency's procurement specialists to use their financial resources economically. This is termed 'economy' in the Transport Agency's lexicon.

Note: OC = outcome indicator; O = output; I = input; C = cost.

In New Zealand, most maintenance activities are carried out by contractors whose services are procured through a competitive tendering process, with little or no direct labour being used. Therefore, inputs are not relevant to this application and can be cancelled out, yielding just two ratios on the right-hand side, as follows:

$$\frac{OC}{C} = \frac{OC}{O} \times \frac{O}{C} \quad \text{(Equation 2.3)}$$

where the notation is as previously described.

The revised performance controls are included in table 2.4, along with what they are measuring, and the parts within the agency whose performance is being measured.

Table 2.4 Revised performance controls

Measure	Description
OC/C	As above, a composite measure of the road agency's performance in attaining its outcomes, represented by the surrogate outcome indicator. This is a high-level measure for top management. This is termed 'value for money' in the Transport Agency's lexicon.
OC/O	As above, measure of the road agency's quality or success in attaining its outcomes. This measures the effectiveness of decision making by professionals, such as asset managers. This is termed 'effectiveness' in the Transport Agency's lexicon.
O/C	The ability of the procurement department to acquire economic unit rates, but also of their professional services department to direct use of those resources efficiently. The reciprocal (C/O) provides standard costs per unit of output. This is therefore a combination of 'efficiency' and 'economy' in the Transport Agency's lexicon.

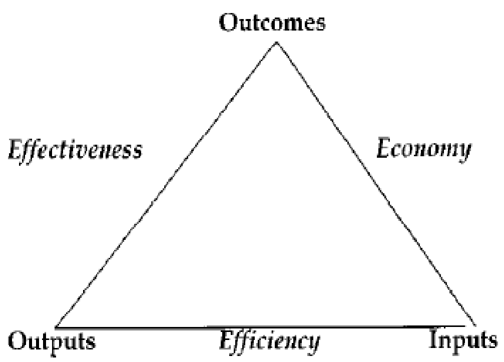
Note: OC = outcome indicator; O = output; C = cost.

The above revised performance controls were also adopted by Rouse et al (1997), albeit with slightly different terminology, as follows:

- Outcomes referred to the state of the network condition prevailing at a point in time.
- Outputs referred to the maintenance activities performed during the year.
- Inputs referred to the resources or cost consumed to provide the outputs.

Referring to figure 2.1 taken directly from Rouse et al (1997), effectiveness can be expressed in terms of the relationship between outputs and outcomes which reflects how well the network condition has been preserved or altered by activities performed. The inputs consumed to provide outputs enable efficiency to be gauged. Finally, although economy refers to the relationship between cost and inputs in the Transport Agency's lexicon, the interpretation adopted by Rouse et al (1997) is the relationship between cost and outcomes. This interpretation of economy, they argue, seems more relevant given that contracting out has become widespread in New Zealand. This interpretation of economy is termed value for money in the Transport Agency's lexicon.

Figure 2.1 The three 'E's' and their relationship measures (Rouse et al 1997)



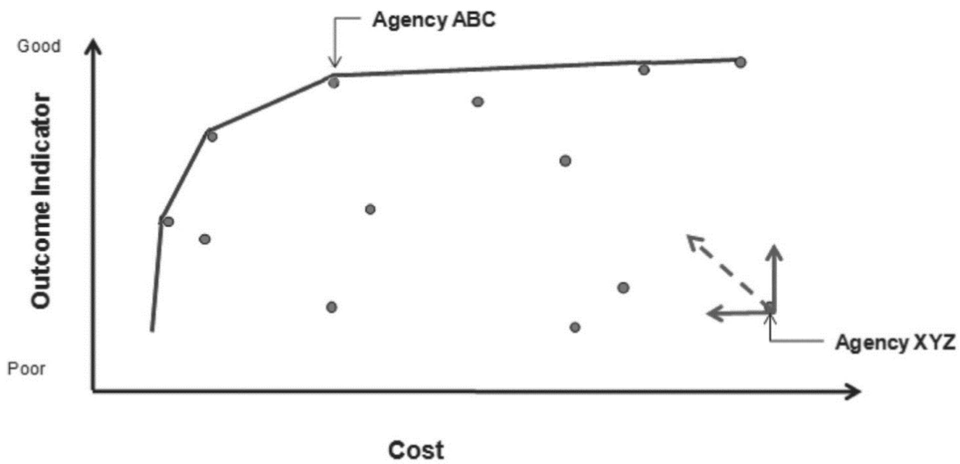
Although the strong emphasis placed on the three 'E's' in the public sector has been criticised (Carter 1991), much of the criticism has revolved around the difficulty of measurement of outcomes. However, in highway maintenance reasonable measures of outcomes pertaining to pavement condition are available, as discussed previously.

2.4 Presentation of performance controls

The performance controls developed above are often represented in the form of a simple input-to-output plot. The efficient frontier concept can then be used to determine a benchmark composed of those agencies demonstrating best practice. The efficient frontier concept is used in a number of the benchmarking techniques discussed later in this report; consequently, the concept is described here, albeit using a simple single-input to single-output example.

Referring to the hypothetical example for OC/C in figure 2.2, taken directly from Costello et al (2014) the frontier is determined by first plotting a generic outcome indicator (OC) on the ordinate and some normalised measure of cost (C) on the abscissa, with each agency plotted as an individual point on the graph. Those points on the frontier or boundary are then joined up to indicate the efficient frontier. Depending on whether the scale of measurement is increasing or decreasing for OC, the frontier will be convex or concave, respectively.

Figure 2.2 Efficient frontier (Costello et al 2014)



The agencies on the frontier can be considered as demonstrating best current practice. The proposed efficient frontier is therefore a benchmark of best practice and does not necessarily represent a perfectly efficient organisation. Taking the hypothetical agency ABC as an example, no other agency with similar expenditure has managed to return as high an OC and, conversely, no other agency with a similar OC value has managed to achieve it for less cost. It is equally important, if not more important, to identify agencies that are inefficient and/or ineffective. The further an agency is from the frontier the more inefficient it is, all other things being equal. For example, the hypothetical agency XYZ is the least efficient of those plotted. All other agencies are either returning a higher OC for the same cost or are achieving the same OC but for less cost, or in most cases a combination of both. Clearly, to improve performance, agency XYZ needs to mirror these other agencies and spend less while maintaining standards, increase standards while spending the same, or possibly a combination of both. Theoretically those agencies that exhibit best practice (ie those on the frontier) can be used as exemplars of how to achieve improvement, all other things being equal. While the example above relates to the high-level OC/C performance control that relates to value for money, similar efficient frontiers can be plotted for OC/O and O/C, thereby providing a tool for comparing, with the intention of improving, the effectiveness of asset management decision making and the efficiency/economy of the procurement department, respectively.

3 State of the practice in benchmarking

3.1 Introduction

This section provides a review of the relevant benchmarking case studies from the literature to document state of the practice in this area, and thereby help inform on best practice benchmarking methodologies. The following sections detail case studies in the highway maintenance and operations sector, case studies in highway safety, case studies in the wider transport sector and, finally, case studies in other sectors. The level of detail included in each section decreases in relation to the relevance to this research project.

3.2 Case studies in the highway operations and maintenance sector

This section provides a number of benchmarking case studies in the highway maintenance and operations sector.

3.2.1 New Zealand territorial local authorities

Considerable research has been carried out to date in New Zealand on benchmarking the performance of highway maintenance and operations in local RCAs, previously called territorial local authorities (TLAs). Two techniques in particular have been used, namely ratio analysis and data envelopment analysis (DEA), and the studies are divided along these lines.

3.2.1.1 Using ratio analysis

Ratio analysis involves the calculation of the ratios for an input-output pair or pairs, depending on the complexity of the analysis problem. The ratios can also be presented in the form of a simple input-to-output plot similar to figure 2.2 in the previous section.

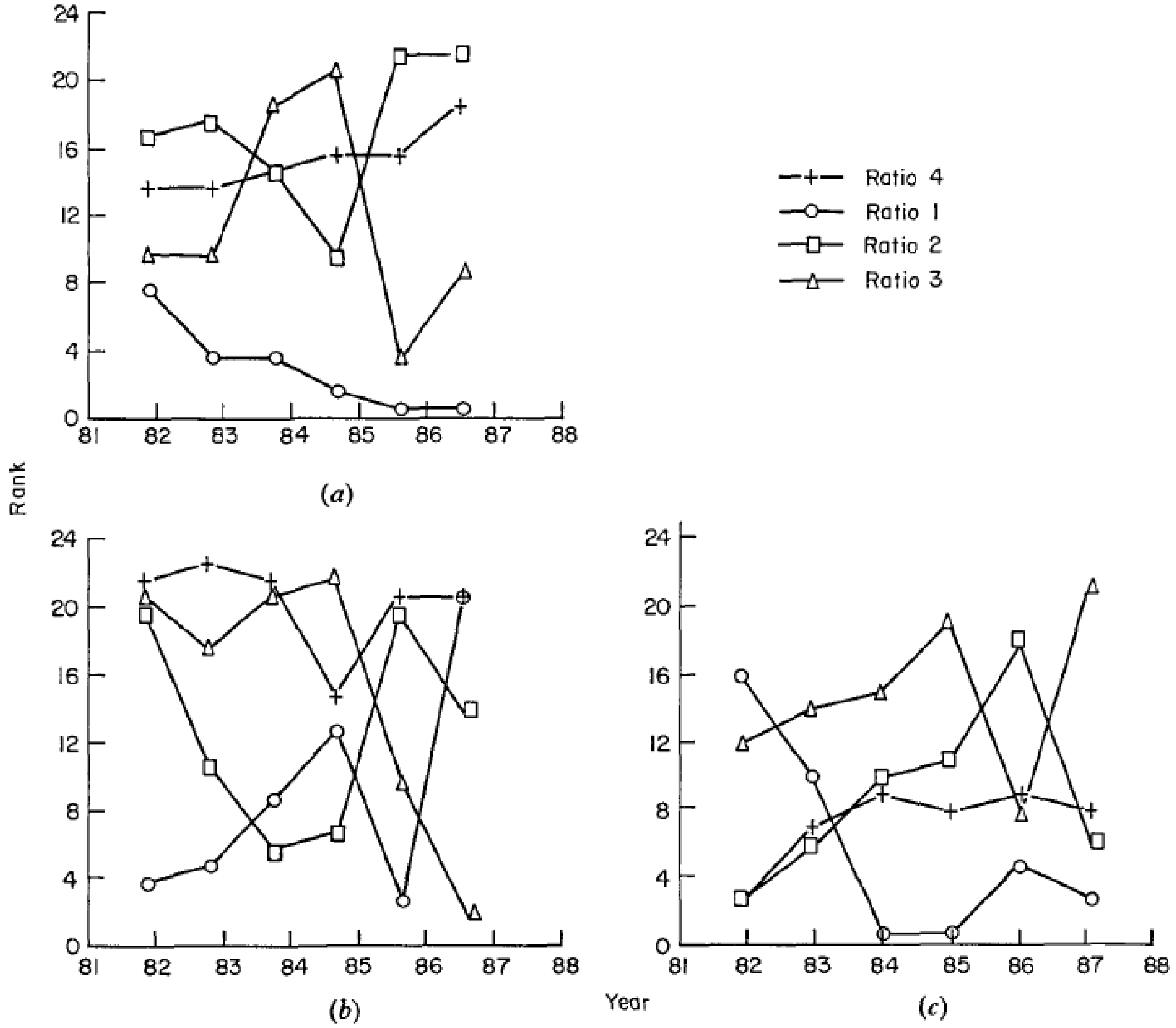
Putterill et al (1990) applied ratio analysis to TLA maintenance management data as a comparison and presentation method. The ratios used are as follows – it should be noted that the condition of the network was not considered in the analysis, possibly due to being unavailable:

- ratio 1: % of general maintenance to the sum of reseals, general maintenance and shape correction
- ratio 2: unit cost of reseals per kilometres of resealing
- ratio 3: % of kilometres resealed to kilometres of sealed length
- ratio 4: unit cost of maintenance per total length of roads.

To account for the variability in operating conditions and facilitate performance comparisons between TLAs, they were compared in peer groups developed in the NRB AD31 road study (Road Research Unit 1982).

The authors then ranked the TLAs within peer groups; this allowed rapid assessment of relative performance and avoided the need to make annual inflation adjustments for monetary measures. They then depicted the ratios as an ordinal time series to help detect performance trends and variations within individual group members, as depicted in figure 3.1. This was termed ranked performance analysis.

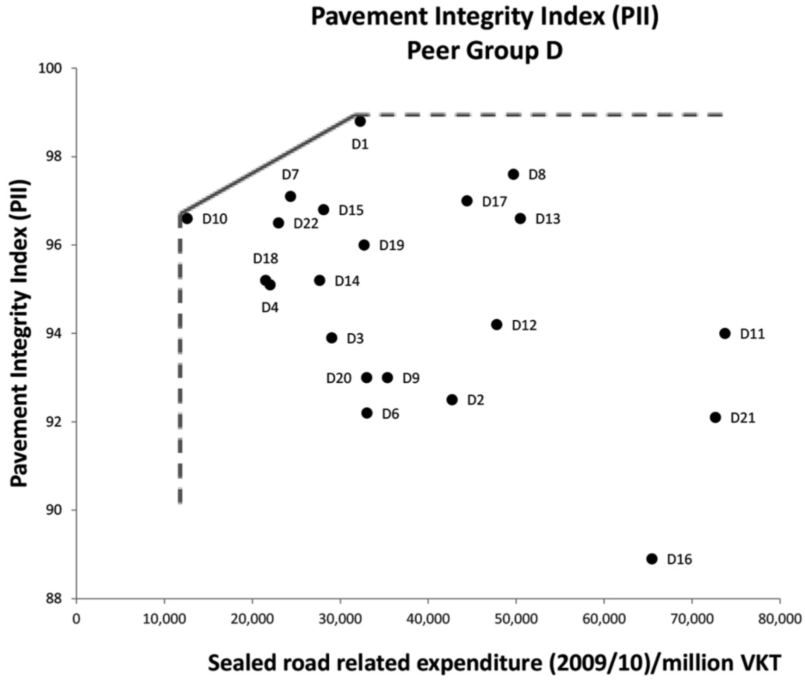
Figure 3.1 Ranked performance ratios for selected local authorities (a) performance of local authority 1; (b) performance of local authority 12; (c) performance of local authority 24 (Putterill et al 1990)



More recently, Costello et al (2014) adapted the Ramanathan (1985) framework previously discussed and applied it with 2010 data to all TLAs using the ratio analysis method. In an attempt to account for some of the variability between TLAs they were compared in their peer groups. An efficient frontier was then incorporated to display those TLAs exhibiting best practice and, conversely, those who were potentially inefficient. An example efficient frontier is shown in the figure 3.2 below for peer group D, displaying PII on the x axis and normalised expenditure on the y axis.

A number of recognised limitations of using ratio analysis were highlighted in the research. These included the necessity to display multiple ratios, given that there are multiple outputs (eg PII, STE) and also multiple normalised inputs (eg expenditure per VKT and expenditure per network km). The per VKT normalisation of the expenditure recognises that roads deteriorate through traffic loading, while the per network km recognises that roads also deteriorate through the damaging effects of the environment; both are therefore required to obtain a clearer picture of performance. All too often, however, the different ratios will provide conflicting information, for example using one ratio TLA 1 may perform better than TLA 2 but using a different ratio the opposite may be true.

Figure 3.2 Efficient frontier for PII – peer group D (Costello et al 2014)



Another limitation of this work is that it relies on peer groups for comparative purposes which, although used for pavement performance comparisons, were originally developed to compare crash rates. The research recommends that either appropriate peer groups are developed for pavement performance comparisons or a more sophisticated technique such as DEA is used to account for the variability between groups.

Finally, there is some difficulty in assigning costs directly to outcomes, given that maintenance expenditure designed to, say, add structural strength to a pavement will also repair the surface roughness and the skid resistance. This can be overcome by defining a composite index for all measures of pavement performance, thereby allowing all costs to be aggregated. However, the challenge in all composite indices is to agree on the weightings in the index, as discussed later in section 4.3.

3.2.1.2 Using DEA

DEA is a non-parametric, optimisation-based benchmarking technique. The underlying idea of DEA is to measure the production efficiency of a so-called decision-making unit (DMU) which consumes inputs to produce outputs. In the context of highway maintenance and operations, a typical DMU is a service provider (maintenance contractor), or maybe a TLA, which consumes funding to produce a safe and comfortable road network, for example. An efficient frontier, similar to that shown in figure 2.2 in the previous chapter, is utilised to set a relative benchmark. However, DEA is not limited to the single-input to single-output frontier shown in figure 2.2 and can instead be multi-dimensional. Outputs from DEA present a relative efficiency score, whereby DMUs on the frontier are given a score of 100%. All others are given a score less than that, the actual value depending on how far they are from the frontier and, therefore, their degree of inefficiency. An example output from DEA is provided in figure 3.3.

Figure 3.3 Output from DEA analysis (Rouse et al 1997)

	EFFIC'Y	EFFECT	ECONOMY
TLA01	100.0%	100.0%	100.0%
TLA02	100.0%	100.0%	100.0%
TLA04	74.6%	100.0%	100.0%
TLA05	100.0%	15.8%	19.1%
TLA06	100.0%	100.0%	100.0%
TLA07	55.2%	55.9%	46.0%
TLA09	100.0%	100.0%	100.0%
TLA10	86.1%	100.0%	100.0%
TLA11	49.5%	77.8%	61.6%
TLA12	100.0%	100.0%	100.0%
TLA13	100.0%	100.0%	100.0%
TLA14	100.0%	59.5%	41.8%
TLA15	65.9%	44.6%	39.7%
TLA16	88.9%	51.7%	33.8%
TLA17	92.7%	45.1%	42.5%
TLA19	75.6%	100.0%	100.0%
TLA20	100.0%	100.0%	100.0%
TLA21	100.0%	100.0%	100.0%
TLA23	78.2%	38.3%	34.5%
TLA24	62.2%	100.0%	100.0%
TLA26	88.7%	100.0%	100.0%
TLA28	91.4%	31.4%	34.1%
TLA29	100.0%	40.5%	40.7%
TLA30	85.8%	100.0%	100.0%
TLA32	100.0%	43.6%	31.7%
TLA33	100.0%	26.7%	22.1%
TLA34	95.9%	60.5%	58.5%

	EFFIC'Y	EFFECT	ECONOMY
TLA35	100.0%	89.9%	78.4%
TLA36	95.5%	30.4%	36.9%
TLA37	64.9%	51.1%	51.3%
TLA39	100.0%	100.0%	34.8%
TLA40	99.1%	40.7%	36.6%
TLA41	100.0%	100.0%	100.0%
TLA42	100.0%	100.0%	100.0%
TLA43	81.9%	66.9%	62.1%
TLA45	76.8%	91.7%	74.5%
TLA46	100.0%	100.0%	100.0%
TLA47	100.0%	100.0%	100.0%
TLA48	100.0%	100.0%	82.2%
TLA49	65.7%	36.5%	28.8%
TLA51	96.9%	10.3%	13.0%
TLA56	71.3%	89.3%	80.9%
TLA57	96.6%	31.9%	26.6%
TLA60	87.5%	20.1%	20.9%
TLA61	100.0%	100.0%	47.8%
TLA64	83.8%	56.9%	51.7%
TLA67	87.9%	25.9%	16.9%
TLA68	100.0%	15.6%	15.8%
TLA69	78.7%	100.0%	100.0%
TLA71	90.1%	39.8%	24.0%
TLA72	63.1%	85.3%	59.3%
TLA73	100.0%	19.5%	18.4%

Referring to figure 3.3, as an example, the efficiency, effectiveness and economy for each DMU, in this case TLA, has been reported. Efficiency, effectiveness and economy are as defined in figure 2.1, whereby the inputs consumed to provide outputs enable efficiency to be gauged, effectiveness can be expressed in terms of the relationship between outputs and outcomes, and economy refers to the relationship between inputs consumed and outcomes. In this example, the inputs consumed refer to total expenditure, the outputs refer to achievement data, such as reseal and rehabilitation kilometres, and the outcomes refer to pavement condition, such as roughness and SCI.

In general, to determine each DMU's efficiency, effectiveness or economy the ratio of output production over input consumption is needed. In the case of one input and one output this is simply output/input. However, when multiple inputs and outputs are concerned, as is the case here, this ratio can only be determined as a weighted sum of inputs and outputs. The weightings, however, are not predetermined but are in fact optimised by the DEA method to ensure each DMU is looked upon as favourably as possible. All other DMUs are then evaluated using these same weights. For example, initially weights are chosen so that DMU 1 looks most favourable. The ratio is then evaluated for DMU 1 and all other DMUs. If, under these weights, another DMU has a better ratio than DMU 1, this implies that DMU 1 is not efficient. Intuitively this is because DMU 1 is not even efficient when looked upon most favourably. Likewise, if no other DMU has a better ratio under the chosen weights, then a DMU is considered efficient. This weight optimisation is an essential aspect of DEA.

The outputs in figure 3.3 were developed as part of a general framework for performance measurement in highway maintenance and operations proposed by Rouse et al (1997). They applied the framework to all TLAs in New Zealand and used DEA to benchmark efficiency, effectiveness and economy across the TLAs.

The first trial was undertaken as a baseline and, of particular interest, subsequent trials attempted to account for some of the variability between TLAs. Such factors included:

- vehicle kilometres travelled
- ratio of urban to rural kilometres of highway
- environmental difficulty.

The regression results reported significant positive coefficients for vehicle kilometres and urban/rural mix which is in line with expectations. The lack of significance for environmental difficulty, however, was unexpected as previous studies had shown this to be a highly significant factor at the level of individual roads (Rouse and Putterill 2000). The assessment of environmental difficulty for each TLA on a scale of decreasing difficulty from 1 to 8 was provided by, the then, Transit NZ staff in its Review and Audit department and the counter intuitive result is probably due to using too coarse a measure for this factor.

Rouse and Putterill (2005) went on to demonstrate a further application of their performance framework by evaluating the effects of a major reform of local government in 1989 when over 230 units of local government were restructured into 74 TLAs. The particular focus of their attention was the widely held belief that economies of scale would result in an improvement in service delivery. The objective was therefore to determine if highway maintenance efficiency improved post amalgamation, and to determine if such improvements were the result of amalgamation or the result of other changes to government practices which happened around the same period.

The efficiency of TLAs before and after local authority amalgamation was benchmarked using DEA, and the Herfindahl–Hirschman Index was used to test whether the performance improvement was due to amalgamation. A construction cost index specifically devised for highway activities (which began life as the NZ Ministry of Works Cost Index) was used to convert expenditures to ‘real terms’. While the benchmarking demonstrated that performance improvements were realised, the results (which were confined to highway maintenance activities) showed no evidence that amalgamation was justified in terms of improvements due to economies of scale. Instead, Rouse and Putterill (2005) suggested that the improved performance was almost certainly the result of new governance practices.

Rouse and Chiu (2009) applied the same performance framework to help determine the best practice mix of expenditure on routine maintenance, resealing and rehabilitation from all TLAs. In this case, the TLAs exhibiting best practice, in terms of efficiency, effectiveness and economy provided a best practice indication of the optimal maintenance activity mix to undertake. Measures of quality, quantity and cost were incorporated in DEA to evaluate each TLA’s performance in terms of efficiency, effectiveness and economy. In an attempt to account for the variability between TLAs, a number of ‘environmental factors’ were included in the analysis. In this study, the best practice mix of expenditure was reported as 59% routine maintenance, 27% resealing and 14% rehabilitation.

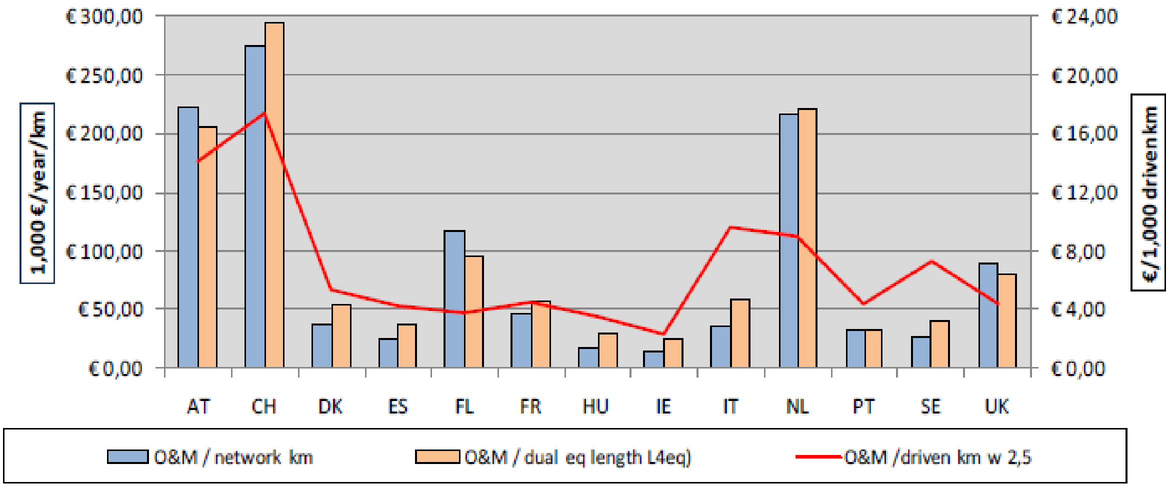
3.2.2 Benchmarking of EXpenditures and PRACtices of maintenance and operation (BEXPRAC), Europe

Road maintenance and operation costs were benchmarked for the first time at a European level by the national road authorities (NRAs) of 13 member countries as reported in CEDR (2010). The resulting Benchmarking of EXpenditures and PRACtices of maintenance and operation (BEXPRAC) survey was initiated in an effort to benchmark the performance of their maintenance and operation policies within the framework of the Conference of European Directors of Roads (CEDR).

Expenditure was compared using some basic normalisation ratios such as expenditure per km, expenditure per VKT and expenditure per weighted VKT, as shown in figure 3.4 taken directly from CEDR

(2010). The weighted VKT assumed that each truck over 3.5 metric tons was equivalent to 2.5 light vehicles; no better equivalent weight being available. The quality of the network was reported, but in a crude manner; defined as the need for maintenance (% of network). As there was no common, clearly defined set of rules for describing the condition of the road networks, the comparisons given leave room for interpretation (CEDR 2010). In any case, the quality of the network, although reported, was not taken into consideration in any of the ratio analyses.

Figure 3.4 Expense ratios (CEDR 2010)



Although the survey gives some indications of the reasons for differences between the countries participating in the survey, it was difficult to collect information on road maintenance and operation expenses in a way that allowed for direct comparisons between countries. Consequently, the differences in the data collected prohibited clear conclusions on efficiency levels (Egger 2012).

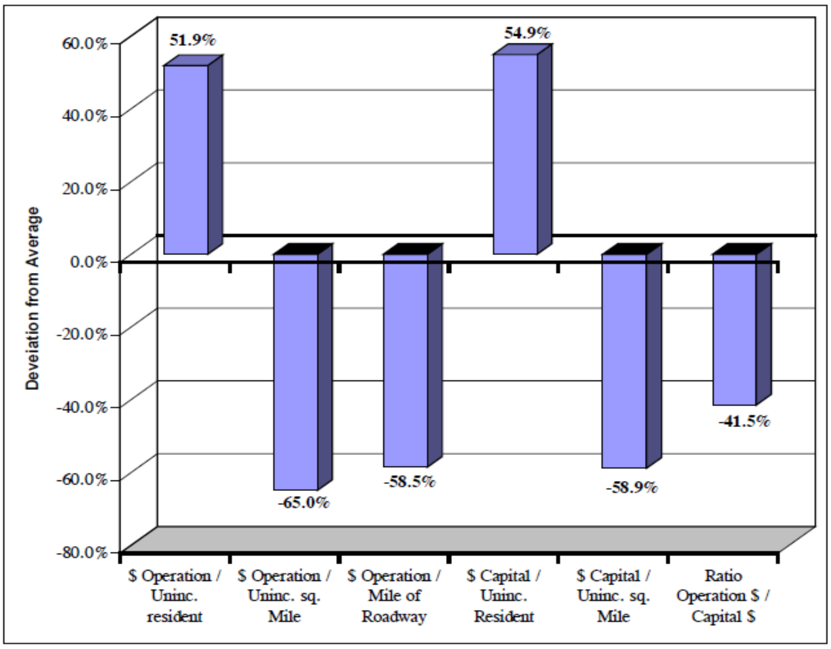
Indeed, the differences in definitions and accounting systems maintained by NRAs appear difficult to overcome; described by Egger (2012) as insurmountable differences in task definitions, practices and accounting methods in the participating countries. This is the main reason why the BEXPRAC members recommend the harmonisation of the definition of tasks and accounting systems before any further efforts are put into a second BEXPRAC study (Egger 2012).

3.2.3 Maricopa County Department of Transportation (MCDoT), USA

Lima and Herz (2003) describe a benchmarking study by the Maricopa County Department of Transportation (MCDoT) in Arizona. The paper describes the process of selection of peer agencies and then the collection, quantification and analysis of the data. The retrieval and comparison of budgetary information was considered the most difficult part of the exercise, given that no standardised reporting mechanism existed for the surveyed agencies and that budget reporting was not always associated with department functions performed.

Figure 3.5 taken directly from Lima and Herz (2003) demonstrates the outputs of the analysis, whereby a ratio analysis was undertaken on expenditure per resident, per square mile and per mile of roadway, as deemed appropriate. These were then displayed as a deviation from the average of the other peer agencies. One of the limitations of such a ratio analysis comparison is highlighted by comparing the performance of MCDoT using the operation expenditure per resident ratio and the operation expenditure per square mile (or mile of roadway) ratio. In the latter they compare very favourably, whereas in the first the opposite is true. Each ratio only tells part of the efficiency story, with no way for the analyst to understand the sum of the parts. As a decision maker it is very difficult to draw a definitive conclusion from such partial information, even if all possible partial efficiency ratios are produced. The risk with such an approach is that incorrect conclusions are drawn by decision makers. It should also be noted that the quality of the network was not considered in the benchmarking analysis.

Figure 3.5 Budget benchmarks for FY 2001 (Lima and Herz 2003)



3.2.4 Virginia Department of Transportation (VDoT), USA

A study by de la Garza et al (2009) applied DEA to the measurement of the relative efficiency of highway maintenance operations of 200 miles of Virginia's Interstate highways maintained by the Virginia Department of Transportation (VDoT) using traditional maintenance practices and 250 miles of Virginia's Interstate highways maintained via a public private partnership using a performance-based maintenance strategy over the fiscal years 2003 to 2007. Regression analysis, the analytic hierarchy process, and an environmental classification method were utilised in conjunction with DEA to account for the impact of uncontrollable factors such as traffic loading, precipitation and temperature differences. These were combined into an 'environmental harshness index/regional effect factor'. The findings show the traditional highway maintenance approach can be as efficient as the performance-based approach.

Ozbek et al (2010a) developed and implemented a comprehensive framework to measure the overall efficiency of the road maintenance operations of VDoT. The framework developed also used DEA to measure the overall efficiency of road maintenance operations while considering the effects of environmental (eg climate, location) and operational (eg traffic, load) factors, both of which are beyond the control of the decision maker, ie the maintenance manager. A companion paper by Ozbek et al (2010b) presents the core implementation stages of the framework in trying to identify 1) the relative efficiency of seven counties of Virginia in performing bridge maintenance; 2) the benchmarks (peers) and targets that pertain to the inefficient counties; and 3) the effects of the environmental and operational factors on the road maintenance efficiency of counties.

3.2.5 Ontario Ministry of Transportation, Canada

Cook et al (1991; 1994) used DEA for the measurement and monitoring of relative efficiency of highway maintenance patrols in the Ontario Ministry of Transportation. Outputs included measures of the area of network served, the traffic level, the improvement in condition and the number of crashes. Inputs into the system included maintenance expenditure, capital expenditure and climate. The relative efficiencies of the highway maintenance patrols were of primary interest; however, an analysis was also undertaken on whether the percentage of privatisation in a district influenced performance and whether the traffic level influenced the performance.

3.2.6 Ontario Roads Coalition, Canada

The Ontario Roads Coalition on performance measurement, benchmarking and best practices was formed in 1999 with the objective of bringing together the various municipal roads associations to form a peer group of road professionals to discuss and recommend appropriate performance measures. Anderson (2003) describes the benchmarking of winter control, their largest operational expenditure item, across municipalities in Ontario. A number of ratios were analysed, including:

- median operating $/lane km by system type
- median operating $/equivalent lane km by system type.

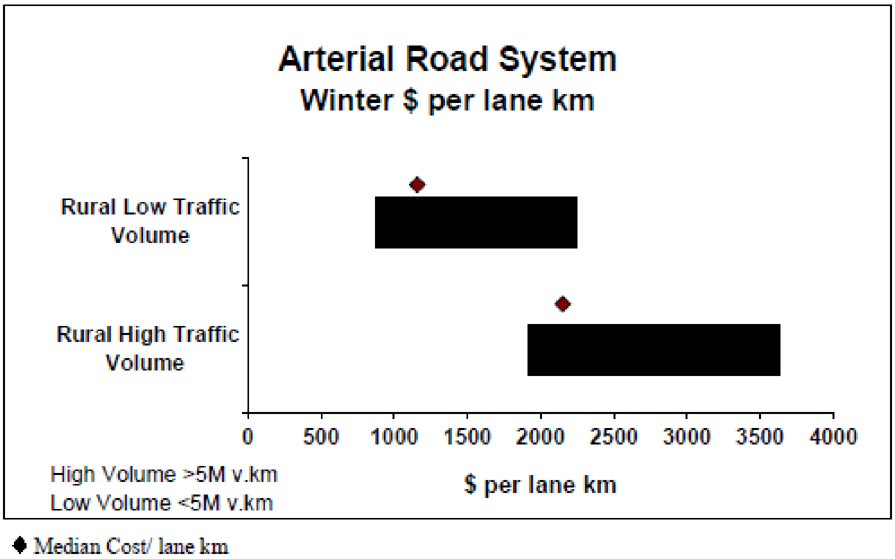
The information was then presented graphically in the form of the median and range per system type. Figure 3.6 displays such information for the ratio median operating $/lane km by system type.

Information which may help explain the differences in performance was also collected, including:

- average cm of snowfall/season
- average days with freezing rain/season
- average tonnes of abrasive/system km
- average tonnes of salt/system km
- average number of lane km/snow plough by road system type

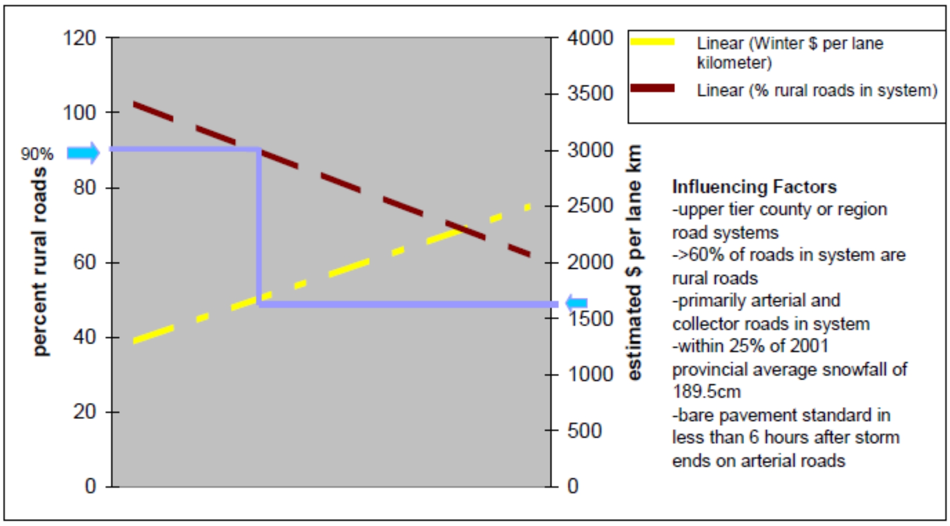
It is apparent from figure 3.6 that there is a significant increase in maintenance costs when the traffic volume is high. This is to be expected given the differing levels of service provided (bare pavement, track bare pavement or snow packed) for high-volume compared with low-volume roads, which were also recorded. The analysis confirmed that level of service (which also varied between municipalities) had a significant impact on cost to deliver.

Figure 3.6 Maintenance costs per lane km (Anderson 2003)



Finally, a simple chart was produced to allow municipalities to compare their results with those provided by the respondents based on a regression analysis of the respondents' results, as shown in figure 3.7. Results that are significantly higher (or lower) than the chart's suggestion, may prompt further investigation to determine reasons for such deviations.

Figure 3.7 Factors influencing the need for benchmarking (Anderson 2003)



Interestingly, feedback from one municipality stated that 'The dialogue and networking opportunities which take place throughout the benchmarking process are as valuable as the qualitative and quantitative data collected, and go further to improve processes and develop best practices than any other tool to date' (Anderson 2003).

3.3 Case studies in highway safety

This section provides a number of benchmarking case studies in the highway safety sector. The road safety measures incorporated in the case studies are not focused on road maintenance, with some of the measures directed towards other factors affecting the road users. Furthermore, the concept of road safety is not limited to engineering studies and measures, but also relies heavily on variables related to policy making.

In 25 countries of the European Union, about 40,000 people lost their lives in road crashes in 2006, which is a reduction of almost 30% when compared with the number of fatalities in 1997 (Khorasani et al 2013). Some of this success can be credited to initiatives like the EUROpean Road Assessment Programme (EuroRAP), which was undertaken to address the unnecessary and preventable toll of death and serious injury on Europe's roads. Initially, this programme incorporated the road safety datasets from Britain, the Netherlands and Sweden, along with some additional data from Catalonia, because of their superior track record in terms of safety performance and data collection. Roads of similar type and similar flow function were compared, both within and between countries. This study demonstrated that for most road networks, apart from Sweden, the assessments of risk rate remained constant over time on most routes (Lynam et al 2003). In a similar study, the star rating of Australia's Network of Highways was undertaken by the Australian Automobile Association which was performed as a part of the Australian Road Assessment Program (AusRAP). This analysis covered 21,921km of national highway, with a speed limit of 90km per hour or above, awarding star ratings based on their level of safety (Australian Automobile Association 2013). EURORAP and AusRAP paved the way for addressing several important issues related to risk assessment, crash savings, investment costs and safety improvement programmes. More importantly, these programmes help in understanding best practices between the countries or organisations within the country.

Hermans et al (2008), in their development of a composite road safety performance indicator, focused on the assignment of weights to the individual indicators. A number of weighting methods were considered: factor analysis, analytic hierarchy process, budget allocation, DEA and equal weighting. Advantages and disadvantages of each are highlighted. As expected, the position of a country in the ranking differs by the method used. Of the five techniques, the weights based on DEA resulted in the highest correlation with the road safety ranking of 21 European countries based on the number of traffic fatalities per million inhabitants. Hermans et al (2009) followed up this work with a DEA based benchmarking model using the same dataset. Conclusions were drawn regarding the influence of alcohol and drugs, speed, protective systems, vehicle, infrastructure and trauma management. For each country that performed relatively poorly, a particular country was assigned as a useful benchmark. It was, however, acknowledged that results obtained from the DEA analysis were highly sensitive to measurement error, input and output specification and sample size.

Despite reports suggesting recent success in the road safety sector in the US (Sivak and Schoettle 2011), Egilmez and McAvoy (2013) make the case for a state-by-state analysis and comparison considering other characteristics of the holistic national road safety assessment problem. They developed a DEA-based index model to evaluate the relative efficiency of 50 states in reducing the number of fatalities caused by road crashes. DEA is known to produce highly satisfactory results when evaluating the productive efficiency of homogenous DMUs with comparable inputs and outputs (Kuah and Wong 2011; Ozbek et al 2009; Sherman and Gold 1985). Egilmez and McAvoy (2013) incorporated five pre-determined input parameters into the DEA model along with a single output – fatal crashes. They concluded the declining trend in fatal crashes is largely attributed to technological improvements, since the analysis returned a negative efficiency growth. Consequently, the utilisation of societal and economical resources by states towards the goal of zero fatality is not still efficient.

Wegman and Oppe (2010) proposed a conceptual framework based on the SUNflower approach. The first SUNflower report (Koornstra et al 2002), compared road safety in Sweden, the United Kingdom and the Netherlands (SUN). The SUNflower approach was developed to analyse various performance indicators like safety measures, policies, casualties, injuries and social costs. In addition, the structure and culture of the country serve as background variables. Wegman and Oppe (2010) describe their framework for the development of a comprehensive set of indicators to benchmark the road safety performance of countries or regions. They also discuss the advantages and disadvantages of combining such indicators and, if combined, how to aggregate the different indicators into one composite performance index. Finally, they propose grouping countries into different classes, where all countries in a particular class are more or less comparable, and use different procedures for this grouping.

Knowles et al (2010) carried out a benchmarking exercise of road safety in Northern Ireland (NI). An examination of the road safety exposure factors in NI compared with the rest of the UK, concluded that no single country or region could be used as a comparator for NI. Consequently, in order to determine an appropriate comparator, a hypothetical region was developed from the data for the rest of the UK, adjusting for variables such as traffic flow and road type. The resulting model indicated that road safety on motorways and urban roads in NI is better than the hypothetical benchmark. However, safety on rural roads was worse. The evidence from this study suggests that young drivers and undivided rural highways in NI are a major cause of concern. The study was limited to killed or seriously injured data, probably because of the discrepancies involved in reporting non-fatal crashes and lack of uniformity in crash definitions.

Akaateba (2012) compared the road safety performance of 20 European and African countries using two different techniques. In the first, a road safety index using simple averaging techniques for all variables and the assumption of equal weightings was developed. In the second, multiple regression analysis was used with time-series data over a number of years to develop a composite index. He argues that the composite index derived using the multiple regression analysis produces more meaningful results when compared with the simple averaging method because it integrates seven independent safety variables into a simple and aggregate index. Not surprisingly, the results of the study revealed different country rankings of the composite road safety index from those produced from the traditional rankings based on fatality rates.

Oluwole et al (2013) compared the road safety performance of 10 African countries using two different techniques: a simple average technique and a composite index approach. The weightings for the composite road safety index were determined based on a literature review and experience. This limitation is acknowledged and the use of regression analyses is seen as the next step in the development of the index. They also highlight a number of limitations of the simple average technique: no insight is provided into indicator importance; no added value is provided for policy makers; and finally, there is a risk of double weighting.

Pešić et al (2013) proposed a technique for evaluating the traffic safety level of territories in Serbia, called the benchmarked traffic safety level. This gives a solitary numerical value to assess the safety level for the region. The main steps involved in deriving the benchmarked traffic safety level include selection of the indicator, calculating the transformed value of the indicator, assigning the weights and, finally, aggregation of data. Allocating the weight coefficients plays a very important role in determining the appropriate safety indicator. Much of the available literature on road safety benchmarking deals with approaches like DEA, analytic hierarchy process (AHP), factor analysis, budget allocation and equal weighting to assign the suitable weight factor to each indicator (Hermans et al 2008). However, commenting on these approaches, Pešić et al (2013) argue that budget allocation is the most effective technique, mainly because of its ease of understanding for decision makers and practitioners who may not be familiar with the previously mentioned statistical modelling techniques.

Al-Haji (2007) developed a composite quantitative index, called the road safety development index (RSDI) to examine the road safety condition in different countries. Various weighting combinations were analysed in order to determine the 'ideal' performance indicators. Al-Haji (2007) investigated the relevance and efficiency of RSDI through empirical comparisons between countries with different levels of motorisation. The research provides a list of key performance indicators which can be applied to evaluate the road safety condition of most countries around the world. Overall, the study not only suggests the need to have a universally acceptable composite index (RSDI), but also outlines the benefits of using such an index.

3.4 Case studies in the wider transportation and other sectors

Various benchmarking studies in the wider transportation sector, and a diverse range of other sectors, have been found in the literature. Examples of these are provided purely to demonstrate the diversity of application of benchmarking and the techniques used therein.

Benchmarking of airport efficiency has been undertaken using a variety of techniques, for example regression analysis (Oum et al 2003; Oum and Yu 2004), DEA (Sarkis 2000; Adler et al 2013; Schaar and Sherry 2008; Lin and Hong 2006) and total factor productivity, a form of ratio analysis (Hooper and Hensher 1997; Oum et al 2003).

DEA, in particular, is prominent in the wider transportation sector, for example benchmarking of train operations (Merkert et al 2010; George and Rangaraj 2008), bus operations (Sexton et al 1994; Kumar 2011) and freight operations (Bhanot and Singh 2014).

DEA has also been widely applied in benchmarking a diverse range of other sectors, such as measuring operating efficiency of US airforce units (Charnes et al 1985), project planning and management (Trappey and Chiang 2008), bank branch efficiency (Paradi and Zhu 2013) and benchmarking of health care providers (Ozcan 2014).

3.5 Summary

Through a number of case studies in the highway maintenance and operations sector, the highway safety sector, the wider transport sector and other sectors, it is evident that a number of benchmarking methodologies have been employed as state of the practice. These include:

- the partial efficiency measure approach (or ratio analysis) which requires a single input to single output ratio to be produced for each relevant input and output
- the total factor efficiency measure (or total factor productivity) approach which produces only one input-to-output ratio incorporating all the inputs in a single measure and all the outputs in a single measure
- regression analysis (or multivariate analysis) which allows the linking of inputs and outputs through a parametric equation modelling the process under investigation
- DEA, a non-parametric, optimisation-based benchmarking technique, based on a frontier production model.

Each has their own advantages, disadvantages and limitations, which are analysed in the following section with a view to recommending one for adoption by the Transport Agency.

4 Analysis of benchmarking techniques

4.1 Introduction

This section provides an analysis of the benchmarking techniques from the literature through highlighting the advantages, disadvantages, and limitations of each. Based on the foregoing, their suitability for use in the Transport Agency's context is discussed and a recommended benchmarking technique/methodology is proposed for adoption.

A common measure of efficiency is the simple input-to-output ratio in the equation below:

$$Efficiency = Output/Input \quad \text{(Equation 4.1)}$$

However, due to the existence of multiple inputs and/or outputs in a complex process such as highway maintenance and operations this simplistic model is inadequate. Consequently, the following techniques, for use with multiple inputs and/or outputs, are discussed based on Rouse and Swales (2003), Ozbek et al (2010a) and others as referenced.

4.2 Partial efficiency measure

The partial efficiency measure approach (or ratio analysis) requires the above single input to single output ratio to be produced for each relevant input and output. Multiple outputs exist in the highway maintenance and operations sector, for example roughness, rutting, SCI, PII, skid resistance, and texture. Inputs, in this case costs, need to be matched to each output. However, accurate classification of the costs to each output is difficult as expenditure that improves one output can also improve one or more additional outputs.

This results in a number of individual input-to-output ratios. The major drawback of such an approach is that each ratio only tells part of the efficiency story, with no way for the analyst to understand the sum of the parts. All too often the different ratios will provide conflicting information, for example using one ratio DMU 1 may perform better than DMU 2 but using a different ratio the opposite may be true. As a decision maker it is very difficult to draw a definitive conclusion from such partial information, even if all possible partial efficiency ratios are produced. The risk with such an approach is that incorrect conclusions are drawn by decision makers. However, the use of ratio analysis has been shown to be useful in combination with other more sophisticated techniques (eg Thanassoulis et al 1996). In such cases, the simple ratio analysis can be used to explore further the details of an inefficient DMU identified by a more sophisticated approach.

Normalisation for the unique network characteristics, outside the control of the supplier/maintainer, that may impact on cost and quality is undertaken through the definition of 'peer groups'. DMUs are grouped into 'peer groups' of organisations that operate under similar conditions (Putterill et al 1990). Similar conditions may include for example, traffic loading, network complexity, urban to rural split, geology or climate. DMUs are then compared within peer groups and not as a whole cohort, limiting the comparisons that can be made. In addition, variability still exists within peer groups, as the variables used to distinguish one group from another are often continuous variables. Despite these disadvantages, the use of 'peer groups' in benchmarking is popular and, if appropriately defined, they are a very useful tool.

4.3 Total factor efficiency measure

The total factor efficiency measure (or total factor productivity) approach produces only one input-to-output ratio which incorporates all the inputs in a single measure and all the outputs in a single measure.

Typically, the inputs are combined into a composite input which requires the assignment of weights to each input. However, in highway and maintenance operations the total expenditure (ie the input) can be used making this a significant advantage over the previous partial efficiency measure. Similarly, the outputs are combined into a composite output which requires the assignment of weights to each output.

While this overcomes the inherent problem with the partial efficiency measure, in that the decision maker only has one ratio to draw conclusions from, it does have one major disadvantage. It requires the subjective assignment of weights by decision makers to each input and output. Although techniques such as the AHP can be used to assist in determining the subjective weightings, this presupposes sufficient technical understanding of the relative importance of each input and output in determining the efficiency. Given the subjective nature of such weights, differing opinions will exist on the relative importance of the inputs and outputs, and depending on who is asked to participate in the determination of the weightings a bias may be introduced. Regardless of what weights are eventually decided upon, the ability to get 'buy in' from the different participants is compromised and they may therefore choose not to take heed of the results, citing a bias in the weightings as their defence. Despite that, the use of AHP in particular is popular in the benchmarking literature where weightings do need to be assigned. Nevertheless, getting agreement on weightings would be a valuable process in its own right.

Normalisation for the unique network characteristics, outside the control of the supplier/maintainer, that may impact on cost and quality is undertaken through the definition of 'peer groups', as previously described.

4.4 Regression analysis

Regression analysis (or multivariate analysis) allows the linking of inputs and outputs through a parametric equation which models the process under investigation. An independent variable, such as performance or output, is predicted using one or a number of independent variables such as expenditure and traffic. The average relationship represents the average output level expected for a given input level. Each DMU is compared with the average relationship and DMUs returning larger outputs than the average are labelled efficient and, conversely, DMUs returning smaller outputs are labelled inefficient. The major disadvantage of this approach is that it compares performance against a hypothetical average and not against the best performers, ie those exhibiting best practice.

However, regression analysis has been shown to help in determining the strength of various unique network characteristics on the efficiency scores determined by other techniques.

4.5 Data envelopment analysis

DEA, a frontier-based production model, is by far the most popular benchmarking technique in the literature. It is a non-parametric, optimisation-based benchmarking technique which originated from Charnes et al (1978) and was later extended by Banker et al (1984). The technique is sophisticated and can be difficult to explain to decision makers (as mentioned in the limitations below); however, readers who wish to understand the technique further are referred to a paper by Ozbek et al (2009) entitled *Data envelopment analysis as a decision-making tool for transportation professionals.*

DEA is considered the best practice benchmarking technique due to its ability to overcome the limitations of the aforementioned techniques, as follows:

- The problem of producing separate single input-to-output ratios for each input-output pair in the partial efficiency measure is addressed by allowing for multiple inputs and outputs in the analysis.

- The weights for the inputs and outputs do not need to be identified by the decision maker, as per the total factor efficiency measure, and instead they are optimised in the analysis to show each DMU in the best possible light (Thanassoulis 2001). Therefore, if a particular DMU is identified as being inefficient relative to others when shown in its best possible light then regardless of the weightings that a decision maker from the DMU may wish to assign they will not be able to improve its efficiency.
- The issue with comparisons to the average in the regression analysis is overcome by creating an efficient frontier of the best performers as the benchmark.

Additional advantages of DEA include the capability of incorporating non-economic factors as inputs or outputs, and the absence of the necessity to specify a fully functional relationship between inputs and outputs, as is the case for regression analysis. The latter advantage is particularly important in highway maintenance given the complexity and number of possible influencing variables (Rouse et al 1997).

DEA does, however, present its own challenges, although some are common to all the techniques mentioned. In particular, Ozbek (2009) highlights the following challenges, taken from Golany and Roll (1989), Ramanathan (2003) and Rouse et al (1997):

- Since DEA is an extreme point technique, errors in measurement or recording of data for input or output variables may result in significant problems. Thus, utmost care should be given to ensure that data is accurate. Such care should, obviously, be taken with all the techniques discussed.
- As efficiency scores in DEA are optimised by running a series of linear program formulations, it becomes intuitively difficult to explain the process of DEA to the non-technical audience and/or decision makers for the cases in which there are more than two inputs and outputs. Thankfully, the outputs of the DEA analysis are much easier to explain to decision makers.
- As mentioned previously, the weights for the inputs and outputs do not need to be identified by the decision maker and instead they are optimised in the analysis to show each DMU in the best possible light. However, difficulties remain with allowing total flexibility in assigning weights (Cook et al 1994) resulting in a large number of DMUs returning high efficiency scores. This can be overcome by assigning upper and lower limits to the weightings based on engineering knowledge. It is easier to agree upper and lower limits to a set of weights than to define a definitive, inflexible, set of weights.

Finally, as mentioned earlier, one of the strengths of the DEA approach is its capability to measure the efficiency of processes with multiple variables. However, running the DEA model using a very large number of variables would shift the compared DMUs toward the efficient frontier. This can result in a large number of DMUs returning high efficiency scores (labelled in the literature as the curse of dimensionality). This can be a limitation of the DEA approach, particularly when a small number of DMUs are being assessed. As DEA allows flexibility in the choice of input-output variables' weights, the greater the number of variables included in the analysis, the lower the level of discrimination (Ozbek 2009).

This has to be managed as part of the process and only variables which have the most influence on the efficiency scores would be assessed, similar to that undertaken by Rouse et al (1997). Preferably, only variables that have a significant influence on the efficiency scores should be included. An alternative approach is to attempt to group the variables into a composite measure of some description.

Clearly, the aforementioned issues with assigning weights arise and a process such as AHP or factor analysis should be adopted where composite measures are deemed necessary.

4.6 Summary comparison of techniques

To assist in comparison of the techniques table 4.1 has been prepared.

Table 4.1 Summary comparison of benchmarking techniques

	Partial efficiency	Total factor efficiency	Regression	DEA
Ability to handle multiple inputs or outputs	Produces multiple ratios	A single composite measure is calculated	Incorporated in the analysis to a degree	Incorporated in the analysis
Choice of weights	N/A	Subjective assignment of weights	Incorporated in the analysis	Weights are optimised as part of the analysis
Benchmark produced	Best performers for each ratio	Best performers for a single ratio	Hypothetical average performer	Frontier of best performers
Method for dealing with unique network characteristics	Peer groups	Peer groups	Incorporated in the analysis, although peer groups could also be incorporated	Incorporated in the analysis, although peer groups could also be incorporated
Complexity of the technique	Relatively simple	Some complexity added through definition of weights	Complex, given that a full functional relationship has to be defined.	Complex linear programming technique
Usefulness of outputs	Difficult to draw a definitive conclusion from multiple ratios	Single ratio, hence conclusions easily drawn	Benchmarking against average not as useful as comparisons to best performers	Comparisons to best performers, efficiency score and peer DMUs identified

In recommending a suitable technique for adoption by the Transport Agency, the above advantages, disadvantages, and limitations have been considered. In addition, the relative popularity of the techniques should be taken into consideration. Despite the complexity of the DEA technique, it is by far the most popular of all the techniques in the literature. Its complexity, however, is offset by its many advantages and the fact that the outputs from the analysis are relatively easy to explain to decision makers. For the above reasons, DEA is considered the most suitable benchmarking methodology for adoption by the Transport Agency.

5 International benchmarking data

5.1 Introduction

This section details the key benchmarking data from the two overseas road agencies selected for comparison. The networks were chosen in consultation with the Transport Agency but were restricted to overseas networks that OPUS International Consultants (sub-consultants to UniServices) had a degree of access to. The purpose of collecting this data was twofold, first to assess the availability and ease of collection of such data and, second, to enable initial comparisons to be undertaken with the New Zealand roading sector should the Transport Agency wish to do so.

To retain anonymity of the participating road agencies they are referred to simply as international comparator A and international comparator B, in the following sections. For each in turn, contextual data is provided in the first instance, followed by data on inputs (expenditure), outputs (achievement) and outcome indicators (performance) as per the Transport Agency's' value-for-money framework. All expenditure is reported in New Zealand dollars purchasing power parity (PPP) equivalents. The PPP values were sourced from the Organisation for Economic Co-operation and Development using the analytical category – Construction: Capital expenditure on the construction of new structures and renovation of existing structures, where structures include residential buildings, non-residential buildings, and civil engineering works.

A full tabulation of the data collected from international comparator A in included in appendix A and from international comparator B in appendix B.

5.2 International comparator A

Expenditure and achievement data below are for the 2014/15 fiscal year. Although performance data is nominally for the 2014 calendar year, data is not collected on the whole network annually.

5.2.1 Contextual data

International comparator A manages a total network length of 45,803km, equating to 89,819 lane km. The network classification includes primary and secondary highways, as well as side roads. All primary and secondary highways are sealed, totalling 28,803 lane km, and of the remaining 61,016 lane km of side roads, 36,974km are unsealed. The surface type of the remaining roads is predominantly hot mix asphalt, totalling 40,253 lane km, with 8,254 lane km of surface treated, 4,191 lane km of cold mix asphalt and 147 lane km of Portland cement concrete rigid pavements. No information is available on average surface or pavement age.

Vehicle kilometres travelled (VKT) on the primary and secondary network was over 21 million, with over 18 million on the primary highways and over 3 million on the secondary highways. No VKT data was available on the side roads. The percentage of heavy commercial vehicles (HCV) ranges from 4% in urban areas to 65% in heavy logging areas, with 19% being the average. No data is available on the urban/rural split of the network.

Geographic and climatic data was also provided. The geographic area over which the network extends is 944,735 km^2 with a population of over 4.4 million. Mean annual rainfall, depending on location, can range from a high of 3,505mm/yr to a low of 278mm/yr. Average annual temperatures, depending on location, can range from a high of 16^{o}C to a low of -6^{o}C. This is a significant factor.

5.2.2 Expenditure data

Periodic pavement expenditure was NZ$ PPP 258M, split NZ$ PPP 189M for primary and secondary and NZ$ PPP 68M for side roads. Routine maintenance expenditure was NZ$ PPP 568M. The routine maintenance was a lump sum and no split is available between summer and winter maintenance or between primary, secondary and side roads.

5.2.3 Achievement data

Achievement data has been split into planned/periodic maintenance, such as rehabilitation and renewals, and reseals as per availability of data. International comparator A rehabilitated or renewed 700km of the network and resealed 150km of the network.

5.2.4 Performance data

Performance data provided included pavement condition data, safety statistics and customer survey feedback. The pavement condition data provided included the IRI average (1.65) and 85%ile (2.65), wheel track rutting average (3.8mm) and 85%ile (6.2mm), a pavement distress composite index average (7.5) and 85%ile (5.2) and numerous measures of cracking, among others. Interestingly, neither MTD nor skid resistance data was collected on the network. Safety statistics included 109 reported fatalities, 3978 reported injuries and 3696 reported damage only incidents. Finally, customer satisfaction was reported as 4.15 out of 5.

5.3 International comparator B

Expenditure and achievement data below are for the 2014/15 fiscal year. Although performance data is nominally for the 2014/15 fiscal year, data is not collected on the whole network annually.

5.3.1 Contextual data

International comparator B manages a total network length of 18,601km, equating to 39,292 lane km. The network classification includes national and state roads. All national roads are sealed, totalling 10,952 lane km, and all but 981km of the state roads are sealed, totalling 28,340 lane km. Surface type is unavailable, however, anecdotally the majority of the network is chipseal. The average surface age on the network is 10 years and the average pavement age is 37 years.

Vkt on the network was over 15 million, with over 4.8 million on the national network and over 10.2 million on the state network. The average percentage of HCV is 15%. The network is 95.5% rural and 4.5% urban.

Geographic and climatic data was also provided. The geographic area of over which the network extends is 2,526,573km^2 with a population of over 2.5 million. Mean annual rainfall ranges from a high of 1,350mm/yr to a low of 190mm/yr, with an average across the state of 486mm/h. Average annual temperatures, depending on location, can range from a high of 36°C to a low of 8°C.

5.3.2 Input/expenditure data

Periodic pavement expenditure was NZ$ PPP 209M, split NZ$ PPP 47M for national roads and NZ$ PPP 162M for state roads. Routine maintenance expenditure was NZ$ PPP 69M, split NZ$ PPP 14M for national roads and NZ$ PPP 55M for state roads. Little or no winter maintenance is performed.

5.3.3 Output/achievement data

Achievement data has been split into reseals and potholes repaired as per availability of data. International comparator B resealed 807km of the network, 305km on the national network and 502km on the state network. They also repaired 62,415 potholes, 24,640 on the national network and 37,775 on the state network. No achievement data was available for rehabilitation/renewals.

5.3.4 Outcome/performance data

Performance data provided included pavement condition data, safety statistics and customer survey feedback. The pavement condition data provided included the IRI average (2.37) and 85%ile (3.24), wheel track rutting average (6mm) and 85%ile (8.4mm), STE (97.3%) and MTD average (1.54mm) and 15%ile (0.9mm). Interestingly, skid resistance data was not collected on the network. Safety statistics included 184 reported fatalities and 298 reported injuries. No data was available for damage only incidents. Finally, customer satisfaction was reported at 96%.

5.4 Comparison of international data

The international contextual, expenditure, achievement and performance data discussed above is presented in table 5.1 for comparative purposes. As mentioned above, full tabulation of the data collected is presented in the appendices.

Table 5.1 Benchmarking data

Data measure	International comparator A	International comparator B
Contextual		
Total network length	45,803km	18,601km
Total lane km	89,819km	39,292km
Vehicle kilometres travelled	21,638,075,526	15,136,000,000
% HCV – average	19%	15%
Population	4,400,057	2,519,321
Geographic area	944,735	2,526,573
Mean annual rainfall – high/low	3,505mm/278mm	1,350mm/190mm
Mean annual air temperature – high/low	16ºC/-6ºC	36ºC/8ºC
Expenditure		
Planned pavement maintenance	NZ$ PPP 258M	NZ$ PPP 209M
Routine maintenance	NZ$ PPP 568M	NZ$ PPP 69M
Achievement		
Periodic maintenance length	700km	N/A
Reseals	150km	807km
No. potholes repaired	N/A	62,415
Performance		
IRI – average (85%ile)	1.65 (2.65)	2.37 (3.24)
Wheel track rutting – average (85%ile)	3.8mm (6.2mm)	6mm (8.4mm)
STE	N/A	97.3%
MTD – average (15%ile)	N/A	1.54mm (0.9mm)

Data measure	International comparator A	International comparator B
Pavement distress composite index	7.5 (5.2)	N/A
Fatalities	109	184
Injuries	3,978	298
Customer satisfaction	4.15/5 (83%)	96%

Some basic ratio comparisons, akin to the partial efficiency measure mentioned in section 4.2, are included in table 5.2. This resulted in a number of individual input-to-output ratios, which is a major drawback of such an approach as each ratio only tells part of the efficiency story. As a decision maker it is very difficult to draw a definitive conclusion from such partial information, even if all possible partial efficiency ratios are produced.

Although all too often the different ratios will provide conflicting information, in this case international comparator B consistently outperformed international comparator A, for the ratio measures reported, by virtue of the fact that they spent less per network length, per lane kilometre, per vehicle kilometre travelled, per head of population and per land area.

However, highlighting another major drawback of this approach, such basic ratios do not take into account the service provided or performance. Although it is evident that although international comparator A commits more expenditure to pavement maintenance, referring to table 5.1 above, they also provide better performance than international comparator B – as measured by the IRI and wheel track rutting, the only comparable performance measures. In addition, the operating conditions are not taken into consideration. For example, international comparator A has higher rainfall and lower temperatures to contend with. In particular, a sizeable proportion of the routine maintenance budget is dedicated to winter maintenance. Finally, only one normalising factor is considered in each of the basic ratio comparisons in table 5.2. This further demonstrates the need to adopt a best practice benchmarking technique, such as DEA.

Table 5.2 Basic ratio comparisons

Data measure	International comparator A	International comparator B
Pavement expenditure (NZ$ PPP)/ network length (km)	$17,956	$14,908
Pavement expenditure (NZ$ PPP)/lane kilometre	$9,157	$7,058
Pavement expenditure (NZ$ PPP)/VKT	$0.038	$0.018
Pavement expenditure (NZ$ PPP)/ population	$187	$110
Pavement expenditure (NZ$ PPP)/land area	$871	$110

Note: The international comparator A VKT data does not include side roads. Consequently, the pavement expenditure in NZ$ PPP/VKT will err on the high side.

5.5 Ease of data collection

Considerable delays were experienced in data collection due to a lack of timely cooperation, composed of delays due to obtaining approvals to release the data followed by delays in interrogating the road and financial databases. In addition, differences in performance measures collected, definition of maintenance tasks, and accounting systems further compounded the data collection challenges. Similarly, the preceding challenges reduce the potential usability of the data for comparative purposes. These challenges are in line with international experiences in this area, in particular those reported in the BEXPRAC survey in Europe (CEDR 2010). Potential ways to overcome these challenges are discussed in section 6.2.

6 Conclusions and recommendations for further development

6.1 Conclusions

The objectives for part 1 of this research project were to:

- Document existing benchmarking methodologies for highway operations and maintenance services currently in use in other roading agencies and similar industries both in New Zealand and overseas.
- Provide an analysis on the suitability of existing benchmarking methodologies for use in the roading sector in New Zealand.

The brief further specified that the research 'should initially focus on finding a translatable best practice benchmarking methodology rather than creating a bespoke model' and that the recommended methodology needs to:

- translate performance, quality and cost into a level of service and value-for-money equation based measure that can be compared across the New Zealand sector
- normalise for unique network characteristics that may impact on cost and quality outside the control of the supplier/maintainer.

A literature review of existing benchmarking methodologies for highway operations and maintenance services is included in chapter 3. An analysis of their suitability for use in the roading sector in New Zealand is included in chapter 4, resulting in a translatable best practice benchmarking methodology that meets the above requirements, as follows:

- The value-for-money equation based measure can be encapsulated in the efficiency, effectiveness and economy benchmarking framework presented in chapter 2 and the performance controls therein. The framework was specifically developed for not-for-profit organisations and has been shown to work for the highway maintenance and operations sector both overseas and in New Zealand.
- Normalisation for unique characteristics that are outside the control of the supplier/maintainer can be undertaken through the use of DEA. The use of DEA to normalise for such characteristics has also been shown to work for the highway maintenance and operations sector, both overseas and in New Zealand. In such cases, DEA has been supported by other analysis techniques, such as the AHP, regression analysis or peer groups, where appropriate.

Such a model will be able to 'compare operations and maintenance costs and performance between networks within New Zealand' and will have the potential to do so 'against similar overseas organisations'. It is not envisaged that the resulting model will be ubiquitous in the same way as the RAMM database. Instead, the model is seen as a high-level management tool to help inform Transport Agency policy on issues such as procurement approaches and funding, as well as to help drive efficiency improvements within the regions, RCAs and service providers. However, referring back to Altman's (1979) components of performance monitoring systems in the public sector, it is worth noting that the potential efficiency improvements identified by the benchmarking will only be achieved through management's use of this information.

The revised brief for part 2 of this research project was to collect benchmarking data from two overseas roading agencies, identified in consultation with the Transport Agency. The purpose of collecting this data was seen as twofold, first to assess the availability and ease of collection of such data and, second, to enable initial comparisons to be undertaken with the New Zealand roading sector should the Transport Agency wish to do so.

The performance and benchmarking data from two overseas road controlling authorities are included in chapter 5, with supplementary data included in the appendices. The challenges faced with collection of the international benchmarking data from the two overseas networks in this project are in line with international experiences in this area. These are elaborated on further in the following section.

6.2 Recommendations for further development

Recommendations for further development of the concepts identified within this report, so that potential benefits of the research can be realised by the Transport Agency, are suggested below. The recommendations have been split into two categories, international benchmarking and national benchmarking:

6.2.1 International benchmarking

As discussed previously, the challenges faced with collection of international benchmarking data include a lack of timely cooperation, composed of delays due to obtaining approvals to release the data followed by delays in interrogating the road and financial databases, as well as differences in performance measurement, definition of maintenance tasks and accounting systems.

A solution to delays due to obtaining approvals to release the data is the formation of an international benchmarking club through, for example, Austroads in the first instance. With buy in from all agencies the collection of benchmarking data can be agreed in advance, therefore making it part of business as usual.

However, the challenges faced in benchmarking such data go far beyond a lack of timely cooperation. The BEXPRAC survey in Europe was in essence a benchmarking club. The aim was to benchmark road maintenance and operation costs, for the first time at a European level, from the NRAs of 13 member countries (CEDR 2010).

While the BEXPRAC survey provided some indications of the reasons for differences between the countries participating in the survey, it was difficult to collect information on road maintenance and operation expenses in a way that allowed for direct comparisons between countries. Consequently, the differences in the data collected prohibited clear conclusions on efficiency levels (Egger 2012). Indeed, the differences in definitions and accounting systems maintained by NRAs appear difficult to overcome; described by Egger (2012) as insurmountable differences in task definitions, practices and accounting methods in the participating countries. This is the main reason why the BEXPRAC members recommend the harmonisation of the definition of tasks and accounting systems before any further efforts are put into a second BEXPRAC study (Egger 2012).

Similar challenges were faced with the collection of international benchmarking data from the two overseas networks in this project. Consequently, the first task of any benchmarking club should be the harmonisation of the definition of tasks and accounting systems. This should be undertaken before any further attempts are made to collect international benchmarking data. A term in current use for such an exercise is the definition of metadata standards.

Interestingly, Austroads is currently undertaking an ambitious project to establish a harmonised road asset data standard for use in Australia and New Zealand (Austroads 2016a). ‘The Road Metadata Standard Project has been initiated in response to requests from stakeholders who increasingly need to share data with other road management agencies but are frustrated by the lack of common data standards… The lack of harmonised road asset data standards means that each road manager collects similar, yet slightly different, information. The minor but consequential differences in road asset data limit the comparability of asset information between road networks, and increases the costs of working across different road networks. The types of data to be considered in the project include descriptions and locations of assets, maintenance activities and cost metrics, asset condition and performance and road classification.' (Austroads 2016b)

If a benchmarking club is initiated within Austroads, then it is recommended that this metadata standard, when completed, informs the next step of the process: the standardisation of data processing to calculate the performance, expenditure and contextual measures.

The risk with such a self-stated ambitious project is that it may be too ambitious, resulting in its outputs never being implemented or even the project never being completed. To mitigate such a risk, it is recommended that any benchmarking club take a different approach to the development of metadata standards. Such an approach should not be 'bottom-up' like the Austroads metadata standards project, but instead 'top-down' whereby the final performance, expenditure and contextual measures are the starting point and only the data required to calculate these measures are harmonised.

Such a top down approach is also likely to be transferrable internationally, outside of Austroads, where the metadata standards project will not be implemented. Invitations to join the benchmarking club would then be accompanied with the processed data standards and metadata standards.

A key outcome is that the production of such measures becomes business as usual and not seen as a 'request for data' requiring many hours of interrogating the road database and financial management systems. Instead, the data processing standards should be encapsulated in a query which can be run at the touch of a button. This will assist with timely cooperation by minimising delays due to interrogating the road and financial databases.

In summary, if international benchmarking is to be implemented then the recommendation is to first form a benchmarking club of similarly committed road agencies and then to agree on (or adopt existing) data processing standards and metadata standards. When all of the above has been achieved then the steps detailed below for the national benchmarking model can be replicated in an international benchmarking exercise.

6.2.2 National benchmarking

Benchmarking at the national level in New Zealand presents far fewer challenges than international benchmarking, particularly at the state highway level. The relative ease of availability, and potential for comparability, of performance and benchmarking data for New Zealand's state highway networks would suggest that benefits from benchmarking can be realised much earlier than from an international benchmarking exercise.

Although not without its own challenges, standardisation of data, if not actually metadata standards, already exists in the form of the RAMM database. Some data processing standards also exist (eg calculation of STE). A top down approach whereby the final performance, expenditure and contextual measures are the starting point and only the data required to calculate these measures is harmonised is also required here.

The next step is the further development of the recommended benchmarking methodology from the literature review and subsequent analysis. The recommended technique is DEA, supported by other analysis methods as required.

Although DEA has been recommended as the technique with which to develop a benchmarking model for the operations and maintenance of New Zealand's road network, considerable further research needs to be undertaken to 'build' the benchmarking model. It is therefore recommended that a combined Transport Agency and RCA working group is set up to champion and facilitate the development of the benchmarking model. In particular, to:

- Agree on the appropriate outcome indicators (roughness, rutting, etc) from which to measure performance of the DMUs or RCAs in the roading context.
- Agree on the relative weightings between the outcome indicators. Although absolute weightings are desirable, in reality these are difficult to reach consensus on. Consequently, appropriate ranges for the weightings can be determined if required. One of the features of the DEA analysis is that the software is capable of handling such information, showing each DMU in the best possible light (within the defined ranges for the weightings).
- Agree on a list of candidate unique network characteristics which influence the performance of road controlling authorities in New Zealand (eg VKT, urban/rural split, climate, geology).
- Make Transport Agency resources available to collect and subsequently process the relevant benchmarking data.

This will then allow the research team to run the combination of DEA analysis and other analysis methods as required to:

- Identify the unique network characteristics that have the greatest impact on the reported efficiencies and therefore should be included in further refinements to the DEA analysis.
- Report on the relative efficiencies of the RCAs, having accounted for the main unique network characteristics. This will allow for the identification of an 'efficient frontier' made up of the best performers. Similarly, it will highlight those whose relative performance is less than ideal, having allowed (normalised) for the unique network characteristics which have the greatest impact.

The above benchmarking model should then be incorporated into the Transport Agency's performance management framework so annual benchmarking can be undertaken. Please note it is likely that two benchmarking models will be implemented, one for the state highway network and one for the local authority network, due to differences in availability of outcome indicators.

As part of further implementation, the Transport Agency might consider the formation of 'national or regional benchmarking clubs' based on the DEA analysis but also based on network characteristics, ie the challenges faced by city councils can be very different from the challenges faced by rural councils.

While all possible unique network characteristics cannot be accounted for (eg some are unavailable), the refined DEA analysis will allow discussions to take place between agencies knowing that the most important unique network characteristics (from those available) have been accounted for. The focus of discussions will then be on determining what the 'best performers' are doing that the others can learn from and adopt.

7 References

Adler, N, V Liebert and E Yazhemsky (2013) Benchmarking airports from a managerial perspective. *Omega 41*: 442–458.

Akaateba, MA (2012) Comparing road safety performance of selected EU and African countries using a composite road safety performance index. *Journal of Natural Sciences Research 2*, no.8: 31–45.

Al-Haji, G (2007) Road safety development index: theory, philosophy and practice. Dissertation. Sweden: Linköping University.

Altman, S (1979) Performance monitoring systems for public managers. *Public Administration Review 39*, no.1: 31–35.

Anderson, B (2003) Benchmarking municipal roads – the Ontario experience. *2003 Annual Conference of the Transportation Association of Canada*, St. John's, Newfoundland and Labrador, Canada.

Australian Automobile Association (2013) *Star rating Australia's national network of highways*. Australian Automobile Association, Canberra.

Austroads (2016a) *Austroads draft standard: data standard for road management and investment in Australia and New Zealand. Classification and inventory/condition and performance/access/demand/ work and costs*. Sydney: Austroads.

Austroads (2016b) *Road data harmonisation project*. Accessed 27 April 2016. www.austroads.com.au/road-operations/asset-management/road-data-harmonisation-project

Banker, R, A Charnes and W Cooper (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science 30*: 1078–1092.

Bhanot, N and H Singh (2014) Benchmarking the performance indicators of Indian railway container business using data envelopment analysis. *Benchmarking: An International Journal 21*, no.1: 101–120.

Carter, N (1991) Learning to measure performance: the use of indicators in organizations. Public Administration 69: 85–101.

CEDR (2010) Report on BEXPRAC. *Conference of European Directors of Roads*, Paris.

Charnes, A, C Clark, W Cooper and B Golany (1985) A developmental study of data envelopment analysis in measuring the efficiency of maintenance units in the U.S. air forces. *Annals of Operations Research 2*: 95–112.

Charnes, A, W Cooper and E Rhodes (1978) Measuring the efficiency of decision making units. *European Journal of Operational Research 2*: 429–444

Cook, WD, A Kazakov and Y Roll (1994) On the measurement of relative efficiency of highway maintenance patrols. Pp 195–210 in *Data envelopment analysis: theory, methodology, and applications.* AW Charnes, W Cooper, AY Lewin and LM Seiford (Eds). Boston: Kluwer Academic.

Cook, WD, A Kazakov, Y Roll and LM Seiford (1991) A data envelopment approach to measuring efficiency: case analysis of highway maintenance patrols. *The Journal of Socio-Economics 20*, no.1: 83–103.

Costello, SB, NE Smith, TFP Henning and M Hendry (2014) Towards measuring road maintenance efficiency and effectiveness in local authorities. *Road & Transport Research 23*, no.2: 3–14.

de la Garza, JM, K Triantis and S Fallah-Fini (2009) Efficiency measurement of highway maintenance strategies using data envelopment analysis. In *Proceedings of the 2009 NSF Engineering Research and Innovation Conference*, NSF CMMI, Honolulu.

Egger, M (2012) Benchmarking of expenditures and practices of maintenance and operation (BEXPRAC). *Procedia – Social and Behavioral Sciences 48*, no.0: 1733–1742.

Egilmez, G and D McAvoy (2013) Benchmarking road safety of US states: a DEA-based Malmquist productivity index approach. *Accident Analysis & Prevention 53*: 55–64.

George, SA and N Rangaraj (2008) A performance benchmarking study of Indian railway zones. *Benchmarking: An International Journal 15*, no.5: 599–617.

Golany, B and Y Roll (1989) An application procedure for DEA. *Omega, International Journal of Management Science 17*, no.3: 237–250.

Hermans, E, T Brijs, G Wets and K Vanhoof (2009) Benchmarking road safety: lessons to learn from a data envelopment analysis. *Accident Analysis & Prevention 41*, no.1: 174–182.

Hermans, E, F Van Den Bossche and G Wets (2008) Combining road safety information in a performance index. *Accident Analysis & Prevention 40*, no.4: 1337–1344.

Hooper, PG and DA Hensher (1997) *Measuring total factor productivity of airports: an index number approach. Transportation Research-E 33,* no.4: 249–259.

Khorasani, G, MRM Verki, M Fereidoon, H Motamed, A Yadollahi, A Tatari and M Khorasani (2013) Evaluation of road safety performance based on analytic hierarchy process. *International Journal of Innovative Technology and Exploring Engineering (IJITEE) 3*, no.5: 168–172.

Knowles, J, B Sexton, B Lawton and S Charman (2010) Benchmarking road safety in Northern Ireland. *Published project report PPR478*. Transport Research Laboratory, Wokingham.

Koornstra, M, D Lynam, G Nilsson, P Noordzij, H-E Pettersson, F Wegman, and P Wouters (2002). *SUNflower; A comparative study of the development of road safety in Sweden, the United Kingdom, and the Netherlands*. SWOV Institute for Road Safety Research, Leidschendam, the Netherlands.

Kuah, CT and KY Wong (2011) Efficiency assessment of universities through data envelopment analysis. *Procedia Computer Science 3*, no.0: 499–506.

Kumar, S (2011) *State road transport undertakings in India: technical efficiency and its determinants. Benchmarking*: *An International Journal 18*, no.5: 616–643.

Lima, P and T Herz (2003) Benchmarking as a tool for assessing a transportation organization's performance. *Institute of Transportation Engineers (ITE) 2003 Annual Meeting and Exhibit*, Seattle, Washington, 24–27 August 2003.

Lin, LC and CH Hong (2006) Operational performance evaluation of international major airports: an application of data envelopment analysis. *Journal of Air Transport Management 12*: 342–351.

Lynam, D, T Sutch, J Broughton and SD Lawson (2003) *European road assessment programme pilot phase technical report*. Farnborough, UK: European Road Assessment Programme (EuroRAP) and AA Foundation for Road Safety and Research.

Merkert, R, ASJ Smith and CA Nash (2010) Benchmarking of train operating firms – a transaction cost efficiency analysis. *Transportation Planning and Technology 33*, no.1: 35–53.

Ministry of Transport (MoT) (2011) *Government policy statement on land transport funding 2012/13-2021/22*. Wellington: MoT.

NZ Transport Agency (2011) *Planning and investment knowledge base.* Accessed 16 October 2016. www.nzta.govt.nz/resources/planning-and-investment-knowledge-base/

NZ Transport Agency (2016) *Transport data.* Accessed 27 April 2016. www.nzta.govt.nz/planning-and-investment/planning/transport-data/

Oluwole, AM, MR Bin Abdul Rani and JM Rohani (2013) Integrating road safety indicators into performance road safety index. *ARPN Journal of Engineering and Applied Sciences 8*, no.9: 757–762.

Oum, TH and C Yu (2004) Measuring airports' operating efficiency: a summary of the 2003 ATRS global airport benchmarking report. *Transportation Research Part E 40*: 515–532.

Oum, TH, C Yu and X Fu (2003) A comparative analysis of productivity performance of the world's major airports: summary report of the ATRS global airport benchmarking research report – 2002. *Journal of Air Transport Management 9*: 285–297.

Ozbek, ME, JM de la Garza and K Triantis (2009) Data envelopment analysis as a decision-making tool for transportation professionals. *Journal of Transportation Engineering* 135, no.11: 822–831.

Ozbek, ME, JM de la Garza and K Triantis (2010a) Data and modeling issues faced during the efficiency measurement of road maintenance using data envelopment analysis. *Journal of Infrastructure Systems 16*, no.1: 21–30.

Ozbek, ME, JM de la Garza and K Triantis (2010b) Efficiency measurement of bridge maintenance using data envelopment analysis. *Journal of Infrastructure Systems 16*, no.1: 31–39.

Ozcan, YA (2014) *Health care benchmarking and performance evaluation: an assessment using data envelopment analysis (DEA).* 2nd ed. Newton, MA: Springer.

Paradi, C and H Zhu (2013) A survey on bank branch efficiency and performance research with data envelopment analysis. *Omega 41*: 61–79.

Pešić, D, M Vujani, M, K Lipovac and B Anti (2013) New method for benchmarking traffic safety level for the territory. *Transport 28*, no.1: 69–80.

Putterill, MS, KE Maani and DG Sluti (1990) Performance ranking methodology for roading operations management. *Transport Reviews 10*, no.4: 339–352.

Ramanathan, KV (1985) A proposed framework for designing management control systems in not-for-profit organizations. *Financial Accountability and Management 1*, no.1: 75–91.

Ramanathan, R (2003) *An introduction to data envelopment analysis: a tool for performance measurement*. New Delhi: Sage Publications.

Road Research Unit (1982) *NRB RRU project AD31 – management of road maintenance*. Wellington: National Roads Board, Road Research Unit.

Rouse, P and RW Swales (2003) *Methodology for the NZ health sector national reference pricing framework – a review*. Auckland: Meridian Consulting.

Rouse, P and T Chiu (2009) Towards optimal life cycle management in a road maintenance setting using DEA. *European Journal of Operational Research 196*, no.2: 672–681.

Rouse, P and M Putterill (2000) Incorporating environmental factors into a highway maintenance cost model. *Management Accounting Research 11*, no.4: 363–384.

Rouse, P and M Putterill (2005) Local government amalgamation policy: a highway maintenance evaluation. *Management Accounting Research 16*: 438–463.

Rouse, P, M Putterill and D Ryan (1997) Towards a general managerial framework for performance measurement: a comprehensive highway maintenance application. *Journal of Productivity Analysis 8*, no.2: 127–149.

Sarkis, J (2000) An analysis of the operational efficiency of major airports in the United States. *Journal of Operations Management 18*: 335–351.

Schaar, D and L Sherry (2008) Comparison of data envelopment analysis methods used in airport benchmarking. In *Proceedings International Conference on Research in Air Transportation (ICRAT-2008)*, Fairfax, VA, Feb 2008.

Sexton, T; S Sleeper and R Taggart (1994) Improving pupil transportation in North Carolina. *Interfaces 24*: 87–103.

Sherman, HD and F Gold (1985) Bank branch operating efficiency: evaluation with data envelopment analysis. *Journal of Banking & Finance 9*, no.2: 297–315.

Sivak, M and B Schoettle (2011) Recent major improvements in road safety in the US: changes in the frequency of crashes or the severity of the outcome of crashes? *The University of Michigan Transportation Research Institute report UMTRI-2011-46*. Ann Arbor, Michigan.

Thanassoulis, E (2001) *Introduction to the theory and application of data envelopment analysis*. Boston: Kluwer Academic.

Thanassoulis, E, A Boussofiane and RG Dyson (1996) A comparison of data envelopment analysis and ratio analysis as tools for performance assessment. *Omega, International Journal of Management Science 24*, no.3: 229–244.

Trappey, A and T Chiang (2008) A DEA benchmarking methodology for project planning and management of new product development under decentralized profit-center business model. *Advanced Engineering Informatics 22*: 438–444.

Waugh Infrastructure Management Ltd (2012) Road maintenance task force: better asset management, planning and delivery. Accessed 16 October 2016. *www.nzta.govt.nz/resources/rmtf-report/docs-interim/better-am-planning-and-delivery.pdf*

Wegman, F and S Oppe (2010) Benchmarking road safety performances of countries. *Safety Science 48*, no.9: 1203–1211.

Appendix A: Data collected from international comparator A

Table A.1 Pavement network data for international comparator A

Category	Measure	Network value	Sub- network	Sub- network value
Network length	Total network length	45,803km	Primary	N/A
			Secondary	N/A
			Side roads	N/A
	Total lane kilometres	89,819km	Primary	28,803km (split N/A)
			Secondary	
			Side roads	61,016km
Network usage	Vehicle kilometres travelled (VKT)	21,683,075,526	Primary	18,311,897,214
			Secondary	3,371,178,312
			Side roads	N/A
Structures	Number of bridges	2,859		
	Bridge deck area	1,614,726 m^2		
	Number of drainage structures (span>=3m)	N/A		
	Sign gantry structures	N/A		
Other data	Urban/rural split	N/A		
	Annual growth (in network length)	1%		
	AADT (length weighted average)	N/A		
	% HCV	19%		
	Annual ESAs (length weighted average)	N/A		
	Average seal age	N/A		
	Average pavement age	N/A		
	Average bridge age	32yrs		

Notes: AADT = annual average daily traffic; HCV = heavy commercial vehicle; ESA = equivalent standard axles

1 Annual growth in network length – averaged from all traffic counters on primary/secondary highways:
   - a 1-year average growth rate = 1%
   - b 3-year average growth rate = 1%.

2 Percent heavy vehicles – averaged from all permanent traffic counters on primary/secondary highways:
   - a average = 19%
   - b range = 4% (urban) to 65% (heavy logging).

3 % sealed vs %unsealed:
   - a 52,845 lane-km sealed (59%)
   - b 36,974 lane-km unsealed (41%)
   - c 100% of primary and secondary sealed.

4 Chipseal vs hot mix asphalt vs rigid (Portland Cement Concrete) split:
   - a hot mix = 40,253km
   - b cold mix = 4,191km
   - c concrete = 147km
   - d surface treated = 8,254km
   - e total = 52,845km sealed.

Table A.2 Expenditure data for international comparator A

Category	Measure	Network value	Sub- network	Sub- network value
Pavement expenditure	Planned/periodic maintenance	NZ$ PPP 258M	Primary	NZ$ PPP 189M (split N/A)
			Secondary	
			Side roads	NZ$ PPP 68M
	Routine maintenance	NZ$ PPP 568M (incl summer and winter maintenance)	Primary	N/A
			Secondary	N/A
			Side roads	N/A
Structures expenditure	Bridges and culverts (>3m diameter)	NZ$ PPP 91M	Primary	N/A
			Secondary	N/A
			Side roads	N/A

Notes: PPP = purchasing power parity

1 Expenditure data is for 2014/15.

2 Planned/periodic maintenance includes planned pavement rehabilitation treatments (ie asphalt overlays, milling/overlays) and also includes graded aggregate sealcoats.

3 Routine maintenance (or operational works) includes crack sealing, small patches, filling potholes, etc – we call that maintenance.

4 Component of routine that is winter maintenance unavailable, as bids are lump sum and vary considerably as some parts of the network get no snow, while other parts receive significant amounts.

Table A.3 Achievement data for international comparator A

Category	Measure	Network value	Sub- network	Sub- network value
Pavement achievement	Length of rehabilitation, renewals and reseal (planned/periodic maintenance minus reseals)	700km	Primary	N/A
			Secondary	N/A
			Side roads	N/A
	Length of reseals/seal coating	150km	Primary	N/A
			Secondary	N/A
			Side roads	N/A
	No. of potholes repaired	N/A	Primary	N/A
			Secondary	N/A
			Side roads	N/A

Notes:

1 The 700km would include treatments like asphalt overlays, milling/overlays, full depth pavement reclamation and in-place recycling.

2 A sealcoat is like a chipseal, but the aggregates are varying gradations rather than a single chip size. This treatment is used on existing hard surfaced roads (ie either asphalt or sealcoat), that have lower traffic volumes (ie < 5,000 vehicles per day).

Table A.4 Performance data for international comparator A

Category	Measure	Network value	Sub- network	Sub- network value
Pavement condition	International roughness index – average	1.65m/km	Primary	1.26m/km
			Secondary	1.65m/km
			Side roads	2.55m/km
	International roughness index – 85%ile	2.65m/km	Primary	1.66m/km
			Secondary	2.35m/km
			Side roads	4.33m/km
	Wheel track rutting – average	3.8mm	Primary	4.4mm
			Secondary	3.2mm
			Side roads	4.9mm
	Wheel track rutting – 85%ile	6.2mm	Primary	6.7mm
			Secondary	4.9mm
			Side roads	6.4mm
	Smooth travel exposure	Not collected	Primary	Not collected
			Secondary	Not collected
			Side roads	Not collected
	Mean texture depth – average	Not collected	Primary	Not collected
			Secondary	Not collected
			Side roads	Not collected
	Mean texture depth – 15%ile	Not collected	Primary	Not collected
			Secondary	Not collected
			Side roads	Not collected
Pavement condition	Pavement distress (composite index) – average	7.5	Primary	7.7
			Secondary	7.6
			Side roads	6.5
	Pavement distress (composite index) – 85%ile	5.2	Primary	6.3
			Secondary	6.0
			Side roads	4.4
	Long wheel track cracking (%)	8	Primary	6
			Secondary	8
			Side roads	12
	Long joint cracking (%)	23	Primary	27
			Secondary	21
			Side roads	19
	Edge cracking (%)	15	Primary	15
			Secondary	14
			Side roads	18
	Transverse cracking (%)	11	Primary	8
			Secondary	12
			Side roads	15

Category	Measure	Network value	Sub- network	Sub- network value
	Meandering cracking (%)	4	Primary	2
			Secondary	5
			Side roads	8
	Alligator cracking (%)	1	Primary	0
			Secondary	0
			Side roads	3
	Potholes (%)	0	Primary	0
			Secondary	0
			Side roads	0
	Rutting (%)	53	Primary	59
			Secondary	39
			Side roads	60
	Bleeding (%)	0	Primary	0
			Secondary	0
			Side roads	0
Safety statistics	Fatalities	109	Primary	N/A
			Secondary	N/A
			Side roads	N/A
	Injuries	3978	Primary	N/A
			Secondary	N/A
			Side roads	N/A
	Damage only	3696	Primary	N/A
			Secondary	N/A
			Side roads	N/A
Customer satisfaction	Survey	4.15/5.0 rating	Primary	N/A
			Secondary	N/A
			Side roads	N/A

Notes:

1 Although nominally 2014 data, data was collected for primary and secondary highways in the years 2012 and 2013. Data on side roads included surveys from 2009 to 2014.

2 Sampling is used, and collected on a rotation. Represents 96% + of primary and secondary highway sand about 40% of side roads.

3 Percentages only were given for distresses, while composite index values are averages.

4 Values of ‘0’ are due to rounding.

5 Safety statistics are from 2012.

6 Customer satisfaction rating is based on the 2014 customer satisfaction survey: overall target score: 4.10/5.0.

Table A.5 Contextual data for international comparator A

Category	Measure	Description
Climate	Mean annual rainfall (mm/yr)	Ranges from a high of 3,505mm to a low of 278mm mean annual rainfall depending on location.
	Mean annual air temperature	Ranges from a high of 16°C to a low of -6°C mean annual air temperature depending on location.
Geographic	Population	4,400,057
	Geographic land area	944,735km2
Contract type	Descriptive	Contractors carry out routine maintenance on roads and structures. Separate contracts are awarded for resurfacing and rehabilitation work.

Appendix B: Data collected from international comparator B

Table B.1 Pavement network data for international comparator B

Category	Measure	Network value	Sub- network	Sub- network value
Network length	Total network length	18,601km	National	5,111km
			State	13,490km
	Total lane kilometres	39,292km	National	10,952km
			State	28,340km
Network usage	Vehicle kilometres travelled (VKT)	15,136,000,000	National	4,889,000,000
			State	10,247,000,000
Structures	Number of bridges	1151		
	Bridge deck area	1,080,000m2		
	Number of drainage structures (span>=3m)	13		
	Sign gantry structures	58		
Other data	Urban/rural split	4.5% urban 95.5% rural		
	Annual growth (in network length)	0.3%		
	AADT (length weighted average)	2,193		
	% HCV	15% (average)		
	Annual ESAs (length weighted average)	349,601		
	Average seal age	10 yrs		
	Average pavement age	37 yrs		
	Average bridge age	37 yrs		

Notes: AADT = annual average daily traffic; HCV = heavy commercial vehicle; ESA = equivalent standard axle

1 The national road network is a defined network of nationally important road infrastructure links, as determined by the government.

2 The state road network includes declared highways and main roads (excluding those included in the national road network) within the state.

3 The state road network includes 981km of unsealed pavements – the remainder of the network is sealed.

Table B.2 Expenditure data for international comparator B

Category	Measure	Network value	Sub- network	Sub- network value
Pavement expenditure	Planned/periodic maintenance	NZ$ PPP 99M (road rehabilitation)/ NZ$ PPP 110M (other planned maintenance)	National	NZ$ PPP 23M/24M
			State	NZ$ PPP 76M/86M
	Routine maintenance	NZ$ PPP 69M	National	NZ$ PPP 14M
			State	NZ$ PPP 55M
Structures expenditure	Bridges and culverts	NZ$ PPP 20M	National	NZ$ PPP 3.8M
			State	NZ$ PPP 16M

Notes: PPP = purchasing power parity

1 Expenditure data is for 2014/15.

2 Routine maintenance includes all routine costs incurred in maintaining the roadway and shoulders, excluding periodic costs incurred on sealed roads at a frequency of more than one year, for example:
   a pothole repairs/minor patching less than 500 square metres
   b crack sealing
   c edge repairs
   d shoulder grading
   e re-sheeting of unsealed roads and shoulders
   f administrative and supervision costs associated with above types of works.

3 Periodic surface maintenance of sealed roads includes all periodic costs associated with maintaining sealed roadways and shoulders incurred at a frequency of more than one year, for example:
   a maintenance reseals/enrichments
   b thin asphalt overlays (less than 25mm)
   c asphalt retreatment and regulation
   d administrative and supervision costs associated with above types of works.

   Any costs associated with the provision of materials and preparation for the above work should be included in this category.

4 Bridge maintenance and rehabilitation includes all costs associated with the maintenance and rehabilitation of bridges and culverts, for example:
   a bridge maintenance, including painting
   b bridge repairs, including replacement of bridge railings and decking
   c administrative and supervision costs associated with above types of works.

5 Road rehabilitation includes costs associated with reinstating failed pavements to existing standards to improve ride quality and/or correct pavement shape, including the provision of a wearing course. These costs will normally improve the riding quality of pavements without improving the design standard, for example:
   a major patching in excess of 500 square metres
   b re-sheeting of sealed roads
   c reconstruction of failed pavements
   d asphalt overlays over 25mm
   e administrative and supervision costs associated with above types of works.

   Where an improvement in the design standard was made in conjunction with rehabilitating a pavement, eg pavement widening and reconstructing an existing pavement, only the cost associated with rehabilitating the pavement to the existing standard should be included in this category.

Table B.3 Achievement data for international comparator B

Category	Measure	Network value	Sub- network	Sub- network value
Pavement achievement	Length of rehabilitation, renewals and reseal (planned/periodic maintenance minus reseals)	N/A	National	N/A
			State	N/A
	Length of reseals/seal coating	807km	National	305km
			State	502km
	No. of potholes repaired	62,415	National	24,640
			State	37,775

Table B.4 Performance data for international comparator B

Category	Measure	Network value	Sub- network	Sub- network value
Pavement condition	International roughness index – average	2.37m/km	National	N/A
			State	N/A
	International roughness index – 85%ile	3.24m/km	National	N/A
			State	N/A
	Wheel track rutting – average	6mm	National	N/A
			State	N/A
	Wheel track rutting – 85%ile	8.4mm	National	N/A
			State	N/A
	Smooth travel exposure	97.3%	National	N/A
			State	N/A
	Mean texture depth – average	1.54mm	National	N/A
			State	N/A
	Mean texture depth – 15%ile	0.9mm	National	N/A
			State	N/A
Safety statistics	Fatalities	184	National	N/A
			State	N/A
	Critical injuries	298	National	N/A
			State	N/A
	Damage only	N/A	National	N/A
			State	N/A
Customer satisfaction	Survey	96% satisfied	National	N/A
			State	N/A

Notes:

1 Although nominally 2014 data, the IRI, rutting and texture data are from a 2012 survey.

2 Smooth travel exposure is from 2013/14.

3 Safety statistics are 2014 data.

4 The Customer Perceptions Survey is from 2015.

Table B.5 Contextual data for international comparator B

Category	Measure	Description
Climate	Mean annual rainfall (mm/yr)	Mean annual rainfall is 486mm/yr, but ranges from a high of 1,350mm/yr to a low of 190mm/yr, depending on location.
	Mean annual air temperature	Mean annual air temperature ranges from a high of 36°C to a low of 8°C, depending on location.
Geographic	Population	2,519,321
	Geographic land area	2,526,573km^2
Topography	-	The topography is mainly flat.
Geometry	-	The majority of the road network is straight or with gentle curves
Availability of natural resources	-	In general, natural construction materials are readily available within the regions. However, in the majority of remote rural areas, material such as sealing aggregates and asphalt may not be available locally.
Geology	-	Approximately two thirds of the land area is covered by crystalline rocks, and only a limited area discloses the sedimentary series. In the south, where the prevailing formations are crystalline schists, they are fringed by deposits containing marine shells.
Contract type	-	All maintenance work is delivered through a relationship-based contracting arrangement, including road network operations, operational asset management, maintenance delivery, minor capital works delivery, and project and contract management services.

Appendix C: Glossary

AAA	Australian Automobile Association
AHP	analytic hierarchy process
AusRAP	Australian Road Assessment Program
Austroads	Association of Australian and New Zealand Road Transport and Traffic Authorities
BEXPRAC	Benchmarking of EXpenditures and PRACtices of maintenance and operation (BEXPRAC) (survey)
CEDR	Conference of European Directors of Roads
DEA	data envelopment analysis
DMU	decision-making unit
EURORAP	EUROpean Road Assessment Programme
GPS	government policy statement
HCV	heavy commercial vehicles
IRI	International Roughness Index
MCDoT	Maricopa County Department of Transportation
MTD	mean texture depth
NI	Northern Ireland
NOC	network outcome contract
NRA	national road authority
OECD	Organisation for Economic Co-operation and Development
ORC	Ontario Roads Coalition
PII	pavement integrity index
PPP	purchasing power parity
PSMC	performance specified maintenance contract
RAMM	Road Assessment and Maintenance Management
RCA	road controlling authority
RDSI	road safety development index
RFP	request for proposal
SCI	surface condition index
STE	smooth travel exposure
SUN	Sweden, United Kingdom and Netherlands
TLA	territorial local authority
VDOT	Virginia Department of Transportation
VKT	vehicle kilometres travelled