OECD

International Transport Forum

TRANSPORT RESEARCH CENTRE

IMPROVING RELIABILITY ON SURFACE TRANSPORT NETWORKS

International Transport Forum

IMPROVING RELIABILITY ON SURFACE TRANSPORT NETWORKS

TRANSPORT RESEARCH CENTRE

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where the governments of 30 democracies work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The Commission of the European Communities takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

This work is published on the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Organisation or of the governments of its member countries.

ISBN 978-92-82-10241-1 (print)
ISBN 978-92-82-10242-8 (PDF)

Also available in French: *Améliorer la fiabilité des réseaux de transport de surface*

Corrigenda to OECD publications may be found on line at: *www.oecd.org/publishing/corrigenda*.

INTERNATIONAL TRANSPORT FORUM

The International Transport Forum is an inter-governmental body within the OECD family. The Forum is a global platform for transport policy makers and stakeholders. Its objective is to serve political leaders and a larger public in developing a better understanding of the role of transport in economic growth and the role of transport policy in addressing the social and environmental dimensions of sustainable development. The Forum organises a Conference for Ministers and leading figures from civil society each May in Leipzig, Germany.

The International Transport Forum was created under a Declaration issued by the Council of Ministers of the ECMT (European Conference of Ministers of Transport) at its Ministerial Session in May 2006 under the legal authority of the Protocol of the ECMT, signed in Brussels on 17 October 1953, and legal instruments of the OECD. The Forum's Secretariat is located in Paris.

The members of the Forum are: Albania, Armenia, Australia, Austria, Azerbaijan, Belarus, Belgium, Bosnia-Herzegovina, Bulgaria, Canada, Croatia, the Czech Republic, Denmark, Estonia, Finland, France, FYROM, Georgia, Germany, Greece, Hungary, Iceland, India, Ireland, Italy, Japan, Korea, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Mexico, Moldova, Montenegro, Netherlands, New Zealand, Norway, Poland, Portugal, Romania, Russia, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine, the United Kingdom and the United States.

The OECD and the International Transport Forum established a Joint Transport Research Centre in 2004. The Centre conducts co-operative research programmes addressing all modes of transport to support policy making in Member countries and contribute to the Ministerial sessions of the International Transport Forum.

Further information about the International Transport Forum is available on Internet at the following address:
www.internationaltransportforum.org

FOREWORD

This report *Improving Reliability on Surface Transport Networks* examines the extent to which appropriate levels of transport reliability are delivered, examining experience in each of the major ITF regions. With growing prosperity, consumers are increasingly mobile and demand higher quality transport services, for which reliable transport networks are central. Reliable trade connections within economies and between global partners are vital for the growing world economy. It is important to ensure there are efficient levels of reliability – neither too little nor too much.

The report focuses on national and international movements of passengers and goods on roads and railways. Although reliability has long been identified as central to the quality of transport services, research on valuing reliability and how best to reflect it in assessments of transport projects and policies began only recently. The results of this recent research are reviewed and used as a basis to explore a range of reliability performance measures useful to policy makers in identifying strategies to ensure appropriate levels of reliability. Recommendations are made for possible improvements in transport planning and operations that explicitly incorporate values of reliability. Case studies of commercial operations and a range of policy initiatives across OECD and ITF countries provide examples of analytical tools that can be used to deliver more reliable networks in a cost-effective manner.

The report was developed by a group of international experts under the Joint Transport Research Centre of the Organisation for Economic Co-operation and Development (OECD) and the International Transport Forum. It is based on research by a working group of experts from 13 countries, chaired by Mr. Hans Jeekel, Corporate Strategist, Centre for Infrastructure of the Dutch Ministry of Transport, Public Works and Water Management.

In introducing the concept of reliability, there are three factors to consider:

1. **The transport task in our economy and our lifestyles.** Transport networks form the vital conduits of our economies and the arteries that facilitate our patterns of living. Transport enables the trade in our global economy to function and is a vital aspect of the labour market mobility that delivers higher economic productivity and lower inflation. The passenger task enables travellers to commute, to socialise and entertain. Thus, "consuming" transport fulfils a means to an end (in most cases) rather than being an activity in its own right.

2. **Transport unreliability impacts on personal and commercial activities.** Not only do disruptions along transport arteries worsen the "transport experience" of network users but, more importantly, they adversely affect the commercial and personal activities that rely on reliable timetabling. Unreliable transport truncates usable time at the destination, such as missing a critical meeting or seeing half the football game. Similarly, because goods components are increasingly produced by specialist companies, the production of the final product relies on interlocking these separate components; thus, delays in the delivery of individual components can then disrupt the entire production line. Unreliability also encompasses early arrival, which can also adversely impact on personal and commercial activities.

3. **Trends in transport and reliance on reliability.** Declining transport costs – through vehicle and infrastructure improvements – have enlarged the operational sphere of influence for travellers and businesses alike. Personal travel – commuting and leisure activities – have lengthened. Businesses have consolidated into larger, but fewer, physical locations. Complementary activities have been outsourcing of production and just-in-time stockholding. However, while fast, reliable transport shapes industry structure, it also increases the vulnerability of the supply chain to perturbations, especially with lengthening supply lines. The costs of disruptions are likely to be higher than in the past, when, for instance, stocks on hand provided insurance against late delivery.

These factors provide the basic framework and focus for this report. Transport is vital to our personal and commercial well-being. Demands for *reliable* transport are rising; conversely, traffic growth increases the challenge to maintain (let alone enhance) the supply of reliability. The central question of this report is whether appropriate levels of reliability are sought and supplied. For this it is necessary to examine why such levels may not arise and then identify if there are appropriate government policy tools that might then be applied.

The outline of the report is as follows. The first section sets the scene for assessing reliability – notably discussing how it is defined and the primary sources of unreliability. Following this is a review of different indicators of network reliability and how monitoring is already a policy tool. Section three discusses how reliability can be incorporated into cost-benefit analysis, the main tool for delivering optimal reliability levels. Subsequent sections consider a range of government policy tools that can be considered in responding to the assessment of transport reliability. Conclusions are then drawn. The report provides a broad range of case studies of private and public treatments in the supply of reliable transport services.

ABSTRACT

ITRD[1] NUMBER E145017

Supply chains are, more than ever underpinned by global and, often, just-in-time production and distribution systems. This complexity is echoed in passenger movements, both for business and social purposes. These changing patterns have increased the importance of keeping to schedules. This increases the focus on transport reliability.

This report examines the extent to which appropriate levels of transport reliability are delivered in each of the major ITF regions. It provides policy makers with a framework for understanding reliability issues and for designing reliability management policies. The report also reviews policies in ITF/OECD countries showing that few countries explicitly incorporate reliability into transport policy making. There are very few cases where reliability is formally incorporated into the cost-benefit assessment.

Because there generally is no direct market for reliability, cost-benefit assessment needs to be used to determine appropriate levels of reliability and to select cost-effective policies to manage reliability. The report has made significant progress in identifying methodology for incorporating improvements in reliability into project and policy evaluation, while exploring the pitfalls that need to be avoided.

A review of existing reliability indicators suggests that governments have started monitoring and targeting reliability. There is a clear dichotomy in performance indicators: indicators of network quality and indicators of what the user experiences. Therefore, reliability targets need to be applied with caution, distinguishing between the network operator and the user perspective.

A wide range of policy instruments is available to manage reliability. The report presents four principal policy options available to manage reliability: physical expansion of capacity, better management of capacity, pricing mechanisms to deliver a market for reliability and information systems intended to mitigate the adverse consequences of unreliability.

Finally, the report highlights the increased importance of reliability noting that reliability needs policy prominence such as is traditionally given to congestion.

Subjects: Traffic and transport planning (72); Economics and administration (10).

Keywords: Operational research, economics of transport, travel time reliability, passenger transport, freight transport, road transport, rail bound transport, policy, planning, timetable, journey time, evaluation (assessment), capacity (road, footway), capacity (traffic network), railway network, international.

1. The International Transport Research Documentation (ITRD) database of published information on transport and transport research is administered by TRL on behalf of the Joint ITF/OECD Transport Research Centre. ITRD contains over 350 000 bibliographical references, and about 10 000 are added each year. Input to the ITRD database is provided by more than 30 renowned institute and organisations from around the world. For more details about ITRD, please contact itrd@trl.co.uk or see the ITRD website at www.itrd.org.

TABLE OF CONTENTS

KEY MESSAGES

Reliability can be better integrated into transport policy making

The objective of this report is to provide policy makers with a framework for understanding reliability issues and for designing reliability management policies. The report has made significant progress in identifying methodology for incorporating improvements in reliability into project and policy evaluation, while exploring the pitfalls that need to be avoided.

At present, network and service reliability is not systematically incorporated in the transport planning process and thus is not reflected adequately in decision making. Reliability is rarely factored into cost-benefit analysis, the core planning tool for surface transport networks.

Increasingly complex scheduling places more importance on reliability

Technological advances and investments in infrastructure have lowered transport costs and increased average transport speeds. This has facilitated and complemented product specialisation. Supply chains are, more than ever – on a global scale – underpinned by global and, often, just-in-time production and distribution systems. This complexity is echoed in passenger movements, both for business and social purposes. These changing patterns have increased the importance of schedules – and of keeping to those schedules. This increases the focus on transport reliability.

Responses to unreliability

Individuals, companies and infrastructure managers affected by changing reliability can respond in a number of ways; individuals build extra (buffer) time into their journeys to allow for the possibility of delay, companies adapt their pattern and timing of operations, while infrastructure managers often provide traffic flow information to reduce the impact of unreliability.

Reliability improvements can be delivered by both users and network providers. It should not be presumed that the infrastructure (or service) provider/government always has to be the source of reliability enhancements. The low-hanging fruit of cost-effective reliability improvements may come from network users.

Four main instruments are available to optimise reliability on transport networks

A wide range of instruments is available to manage reliability. The policy framework proposed in this study distils these into four principal options:

- Increasing the physical capacity of infrastructure, either through supplying extra capacity or improving the quality of existing capacity. Capacity enhancements are generally costly, time consuming and often politically difficult. Setting appropriate network standards and improving the robustness of infrastructure (for instance, durability of material) also influences reliability.

- Better management of existing capacity can facilitate reliability, just as poor management can increase unreliability. Infrastructure managers can improve reliability through better incident management and appropriate scheduling and publicising of maintenance. The core management skills can be supplemented by pro-active network oversight.

- Where feasible, charging directly for reliability could be used to achieve more efficient levels of reliability. However, it is often difficult to provide different levels of reliability according to the value different users place on reliability, and equally difficult to extract different charges for differential performance.

- Information can be provided to users enabling them to mitigate the adverse effects of poor reliability. This may be a cost-effective way to reduce both unreliability and the impacts of traffic incidents on subsequent business and personal schedules.

Incorporating reliability into cost-benefit assessments encourages proper consideration of options for delivering appropriate levels of reliability

In the absence of a direct market for reliability, cost-benefit assessments can be used to determine appropriate levels of reliability. If a separate market existed for reliability, then prices would encourage an efficient level of reliability and would allocate responsibility for reliability to the party that could bear it at least cost. A cost-benefit analysis attempts to proxy such a market. This study has found that reliability is very rarely embodied in such analyses.

Projects designed to deliver travel time benefits (such as those arising when congestion is reduced) are sometimes credited with generating reliability benefits. However, standard appraisals fail to unbundle improved reliability (reductions in travel time variability) from the benefits due to the reductions in average travel time. This omission removes the factual basis for arguing that a project really does improve reliability.

There are ways to measure and value reliability that can be integrated into cost-benefit analysis. These have been used on a pilot basis in a small number of countries. These approaches provide a foundation for explicitly incorporating reliability benefits into investment appraisals and, consequently, policy frameworks.

Diversity in network user demands for reliability means that no simple mark-up can be applied to incorporate reliability into project assessments

It is difficult to generalise about the value of reliability as it will be project, location, user, and time-specific. For one project studied, the value of improvements in reliability were found to be negligible, whereas for another project they were found to add 25% to the welfare benefits of time savings achieved. It is important to recognise the granularity of reliability – that is that different values are placed on reliability by different network users at different times and for different trip purposes.

Since the demand for reliability varies markedly across users, products, locations and firms, a single monetary value for reliability will be of little, if any, use in project appraisal – a range of values is required that represents the major user groups in each case. Practitioners cannot assume that values used in one study are readily transferable to a project in another situation.

It is also important to avoid potential double-counting when factoring reliability into project assessment. This can arise if the standard values of time used to assess average time savings already have an implicit, crude value for reliability incorporated in them.

Reliability targets need to be applied with caution, distinguishing between the network operator and the user perspective

Reliability targets and performance indicators for services and infrastructure performance can facilitate discussions between users, operators and decision makers regarding the right levels of reliability. But employing fixed targets may be distorting as they can dominate other service characteristics that may be of equal, or greater, importance. Such targets also invariably present an average level of reliability not reflecting diversity in the demand for reliability.

There are also trade-offs to be made. For instance, a rail infrastructure manager may enhance reliability by reducing the number of trains that it operates. The improvements in reliability may then come at the cost of a more limited train schedule and higher overcrowding on the trains. Reliability targets need to be carefully coordinated with other key performance indicators. In those cases where only passenger trains face performance targets, network managers may be inclined to give higher priority to passenger trains over freight trains than is economically justified. Targets should therefore aim at reflecting both the network and the user perspective. For a network provider, the focus is on system vulnerability or operating performance. For a network user, the focus is on the variability of travel times experienced by the user. The incentives that the targets create in relation to other policy goals and the overall efficiency of transport systems need regular review.

EXECUTIVE SUMMARY

Most of us face unreliable travel services in our daily lives, with unexpected delays leading us to miss a train or arrive late for school or work. Whether it be for social or business events or deliveries of goods, reliability is a key quality of movement. A review of policies in ITF/OECD countries shows, however, that few countries explicitly incorporate reliability into transport policy making. This report aims at providing policy makers with a framework for understanding reliability issues, for incorporating reliability into project assessment, and for designing reliability management policies.

The economics of reliability

Reliability is unanimously regarded as a desirable transport network attribute. There is less unanimity in defining reliability. Yet the definition adopted has major implications for policy. Technically, a reliable system is one that performs its required functions under stated conditions for a specified period of time. Under this definition, a road system that becomes choked with traffic during peak hour, reducing speeds to a slow 20 kilometers per hour, could be regarded as unreliable, or 50% reliable, depending on the conditions specified.

An alternative definition of reliability draws on the attribute of predictability. In this context, a congested road system where the speeds at different times of day and different days of the week are consistent, and hence predictable, would be ranked as highly reliable. While both interpretations are valid, the focus of this study will be on the second definition.

Like all desirable features of a transport network, reliability comes at a cost. It is subject to the standard rules of supply and demand: the higher the price, the higher the quantity that will be supplied but the lower the quantity that will be demanded. Conversely, the lower the price, the more consumers will demand it. The challenge for policy makers arises in two areas. The first is in formulating the institutional arrangements that affect the market for reliability. For instance, legal frameworks that prevent discrimination between transport system users can create impediments to differentiating between services on the basis of reliability. The second is in the treatment of reliability when assessing publicly-funded transport infrastructure projects.

In other words, the role of the government is two-fold: encouraging a market for reliability and incorporating reliability into the assessment of transport infrastructure projects. In terms of the first role, it is important to note that, as a service attribute, reliability is often bundled with other attributes such as speed, convenience and cost, making it very difficult to differentiate a separate market for reliability.[1]

An important point that follows from this is that only when, say, two parallel services are provided with reliability being the key differentiating feature is there an explicit market for reliability. Without this, there is a major challenge in developing sound estimates of the value placed on reliability by network users.[2]

Ideally, market incentives would encourage not only an efficient level of reliability but would also allocate responsibility for reliability to the party that could bear it at least cost.[3] This point is also explored in the report.

Factoring reliability into cost-benefit analysis is clearly desirable but also problematic. The values placed on reliability vary from project to project. Using an incorrect value could result in a worse outcome than a failure to incorporate a value for reliability at all. Cost-benefit analysis, as a set of rationalised economic principles, has evolved over more than a century and useful refinements are unlikely to be developed overnight. However, this report has made significant progress in identifying possible methodology for incorporating a value for reliability into project evaluation, as well as exploring the possible pitfalls that need to be avoided.

Demand for reliability is increasing

Changes in commerce and personal travel patterns have increased the importance of a reliable transport system. Reliable transport networks and services are required because of more complex and inter-related supply chains and increasingly complex scheduled activities. The physical way that the economy operates has changed, facilitated by – and demanding – transport system enhancements.

Transport productivity has increased markedly, yielding benefits for business through the specialisation of production on a global scale and the spread of just-in-time production and distribution systems. One aspect of that productivity is the reduction in transit time, which expands the market for the goods and services, and broadens the way in which firms can interact. The increased interaction between businesses puts a premium on reliability. In modern dispersed production systems, time has become the critical factor where timely delivery of components has replaced traditional stock-holding. These developments have facilitated and accompanied the growing operational sphere of influence for businesses. Multinational businesses have consolidated into larger, but fewer, physical locations, growing with the globalised economy. Broadening national and international trade links, with increasing goods movements, have brought greater volumes of goods, moving further and in increasingly complex and – crucially – interdependent ways. That interdependence depends on reliable transport.

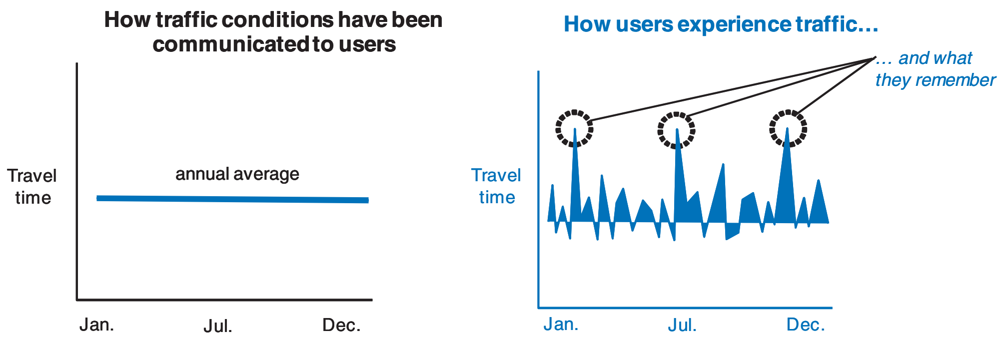
There are also changes in personal lifestyles. Passenger movements, both for business and social purposes, have become more complex with changing patterns of employment, increased disposable income, recreational choices and leisure time. These diverse and geographically-spread activities have led to more intensive use of transport systems, bringing greater dependence on transport to be reliable so that delays do not cascade through the busy calendar of events. The scheduling approach increasingly adopted in private lives echoes the just-in-time deliveries in commerce.

The importance of scheduling in personal and freight activities has grown, so that transport unreliability has an increasingly-marked effect on downstream activities. The expectation from these demand trends is increasingly that transport should provide high levels of reliability.

Unreliability makes trips frustrating

Unreliability makes journeys frustrating and causes stress. The feeling of travelling without control over one's travel time is a disempowering experience, and bad experiences are remembered by travellers. Traffic conditions in the past have often been communicated to travellers only in terms of simple averages (left chart in figure ES1). However, most travellers experience and remember something much different than a simple average of commuting travel time (right chart in figure ES1). Users have deeply negative perceptions of unexpected delays, which colour their attitude to the experience.

Figure ES1. **Travellers' perception of traffic conditions**



Source: FHWA (2006).

Unreliability constitutes a cost

Where performance is inconsistent, network users may simply have to accept the consequences of the delay, albeit that it may have ripple-effects or, worse, snowballing (compounding, or growing) effects, affecting other activities or stages in the personal or logistics chain, constituting a cost to those involved.

The ripple-effect of delays is an important reminder of the inter-connectedness of many individual schedules. A delay at one stage in a person's schedule of activities can mean delays in later related, or unrelated, tasks. Similarly, while logistics chains are built in such a way as to reduce their vulnerability to individual events, any delays in individual consignments can still reverberate through the chain. Indeed, because the transport task is part of a chain, a break in any part of it is a break in the entire chain. An assembled television set with only 99 of its 100 components is an incomplete product that can be neither shipped nor sold.

Costs of unreliability may rival those of congestion. Bearing in mind that the results are not transferable across locations, it is nonetheless significant that there is evidence that the cost of unreliability may cause around half of total underlying delay costs.

Journey-time predictability is the defining feature of reliability

In this report, reliability is defined as:

> *the ability of the transport system to provide the expected level of service quality, upon which users have organised their activities.*

The key word is "expected". According to the definition, reliability can be improved either by supplying a higher level of reliability, or by changing expectations of the level of reliability. In other words, unpredictability (or inconsistency) of network performance is the defining characteristic of unreliability. The more random (less predictable) the performance, the harder it is for the network user to ensure against delays.

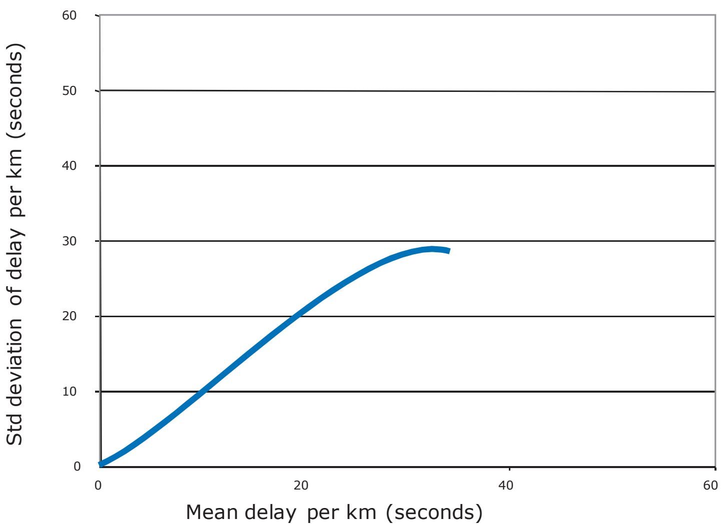
Average travel time between two destinations includes both expected and unexpected delays. It is assumed that network users accommodate expected delays into their travel time through, say, the inclusion of buffer time. However, it is more difficult and costly to incorporate the unpredictable – the unexpected – delays that lead to variation from planned or anticipated travel time.

Disturbances that cause delays can also be classified as recurrent (such as weekday peak-hour congestion) or non-recurrent (such as delays caused by crashes, inclement weather and other events of nature). The essence of the degree of recurrence is that it provides information about the predictability of the event.

The terms unreliability and congestion are often used synonymously. However, as follows from the foregoing discussion, a congested network does not have to be unreliable. Unreliability refers to unanticipated delays, and therefore a congested network is not necessarily unreliable because journey time along a congested road can be fairly predictable.

That said, congestion increases the likelihood of unreliability: as traffic levels increase, the time delays due to slight perturbations tend to increase more than proportionately. This is illustrated by one example, a motorway in the United Kingdom (see Figure ES2 below), where there is a clear correlation between the level of congestion and reliability until high levels of congestion are reached. That said, it is not possible to say whether the variability of travel time was predictable or not.

Figure ES2. **Relationship between reliability (vertical axis) and congestion (horizontal axis) on the M42 motorway in the UK**



Source: Mott MacDonald (2009).

The distinction between reliability and congestion is important because of the different policy implications. However, it is also recognised that remedial actions directed at congestion can improve reliability and, similarly, actions that improve reliability can reduce congestion. For instance, many of the bottlenecks in international supply chains are located in congested urban areas. Reducing congestion at port and hinterland connections may also improve the reliability of the entire logistic chain. There can be overlaps.

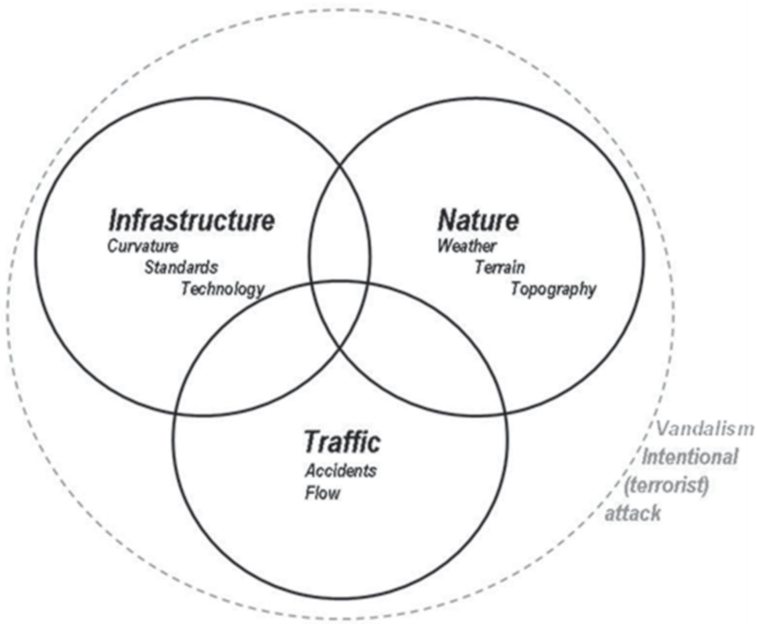
Unreliability arises from multiple sources, each requiring different ways to manage the problem

Unreliability can arise from various activities that are within the control of the network user or network provider. Unreliability of the transport infrastructure network arises from two primary sources:

- Unpredictable demand-related traffic interactions between users (congestion).
- Unanticipated supply-related events:
  - Traffic incidents (crashes and vehicle break-downs).
  - Natural events (*e.g.* floods and earthquakes).
  - Network maintenance (causing temporary reduction in supply).
  - Mismanagement in infrastructure supply, which can also include inappropriate maintenance programs.

Mismanagement of road and railway networks can reinforce other sources of unreliability. It is possible that an uncongested road can be unreliable if the network is poorly-managed; similarly, a congested road with poor management is likely to magnify unreliability. This observation is represented by the intersection of the circles in Figure ES3, showing the primary sources of unreliability.

Figure ES3. **Primary sources of unreliability and inter-relationships**



Source: Derived from Husdal, J. (2004).

Figure ES3 illustrates the interfaces between the various sources of unreliability. For example, low standard infrastructure is likely to be more prone to unreliability arising from events of nature than if the infrastructure is designed to a high standard. This is not to argue for infrastructure to be built to a high standard by default. Given prevailing conditions, such as the likelihood of disruption and levels of traffic, it may be highly appropriate for the infrastructure to be built to a relatively low standard.

Finally, reliability issues are very location- and time-specific – and this affects potential actions to manage the problem, as well as the degree to which costs and benefits from one situation can be inferred to another situation.

Network users develop strategies to deal with unreliability

Individuals and companies affected by deterioration in reliability respond in a number of ways. To reduce the risk of being late at the destination, network users may allow more time for the travel (the so-called safety margin, or buffer). This means, in practice, leaving earlier to ensure arriving on-time. Companies and logistic managers adapt their operations either through changing the way they operate, or by building in buffer stocks of goods. Deliveries can avoid daytime and peak delays, and there has been a growth in evening or night-time deliveries; in some instances companies make greater use of regional depots. Companies also adapt their logistic operations through active traffic management schemes. Increased use of vehicle telematics, routing software and fleet management packages have assisted the adjustment to more congested infrastructure. Minimising the impact of delays on the cost and quality of logistics has become a core skill for freight and logistics managers.

However, each of these options has an associated cost. Leaving earlier to ensure arriving on-time consumes the time available for other, potentially more productive, activities. Holding additional stocks of goods "just in case" involves a capital cost both in terms of the storage facilities and financing the stocks.

Governments have started monitoring and targeting reliability

The first step to recognising the importance of reliability is to monitor it. A number of countries have been exploring ways to monitor reliability. Two distinctive activities are involved here: monitoring service reliability, and setting targets against which the service provider's actual performance is compared.

A review of existing reliability indicators suggests that the purpose of such monitoring is for ascertaining performance and the quality of transport service delivery.

Performance targets are set for three primary reasons:

- Reliability is an important service characteristic in the transport sector.
- The services to which targets are applied often involve monopoly provision underwritten by the taxpayer. Hence governments have an interest in seeking attractive services and efficient provision.
- Reliability targets are important for initiating discussions between politicians, operators, providers and users on the appropriate delivery of service standards.

Most of the existing reliability targets can be found in the rail sector, a transport mode that seeks to run to strict working timetables. Target-setting practice is prevalent in passenger railways. The scheduling of arrival times readily enables these types of targets to be set (setting actual train arrivals against scheduled arrivals) and rail service providers are generally considered a monopoly. To the extent that the provider is perceived to be a monopoly, governments usually oversee supply standards by monitoring and setting performance standards; the target provides a degree of accountability in service quality. Data on service reliability are essential for this oversight. A similar approach is adopted in aviation, where airline service punctuality statistics provide bellwethers for regulatory and policy monitoring.

The actual service performance and the performance against targets are often published as a way for regulators and governments to make service providers accountable, and as implicit encouragement to improve services. The publication of service performance is also relevant for network users to understand the quality of service delivery, enabling users to allow adequate buffer time against delays.

There are several shortcomings in some of the reliability indicators currently available:

- ***Aggregation across users***. Most existing reliability indicators monitor the performance characteristics of the whole system rather than satisfaction of users' needs. That is, whether different users actually receive reliable services.

- ***Aggregation across time***. The indicators normally show overall annual averages only, and therefore mask shorter-term variations in service standards.

- ***Reporting partial data.*** More generally, most of the existing indicators were originally designed to provide feedback to network managers, rather than to measure reliability as perceived by end-users. Thus, the indicators may report operational details such as freight train arrival times, rather than those of primary interest to customers, such as the predictability of collection or delivery times.

As noted earlier, this study has found that, despite its obvious importance, there is generally no explicit view on what travel time reliability *is* precisely; similarly, there is no consensus on how reliability should be monitored. Various definitions for travel time reliability exist, and consequently many different relevant indicators are available. Crucially, there is little recognition of the risks in setting targets (or the difficulty in establishing cost-effective targets); too high might distort desirable management decisions, while low targets might make service provision quality too lax.

There are also trade-offs to be made when influencing reliability through explicit targets. For instance, a railway operator may enhance reliability by reducing the number of passenger trains that it operates. The higher reliability then comes at the cost of less-frequent, higher-loaded trains. Similarly, if performance targets were applied only to passenger trains, the network managers may be inclined to give higher priority to those trains over freight trains than economically justified.

A small number of countries have incorporated reliability into project cost-benefit assessment, but so far fail to reflect diversity in reliability valuations

Case studies reviewed in this study illustrate that some projects are carried out specifically in order to improve reliability. However, there are very few cases where reliability is formally incorporated into the cost-benefit assessment (and hence in the decision making process). Even where decision-making guidelines *do* incorporate reliability, most of the actual project appraisals do not include monetised parameters for reliability.

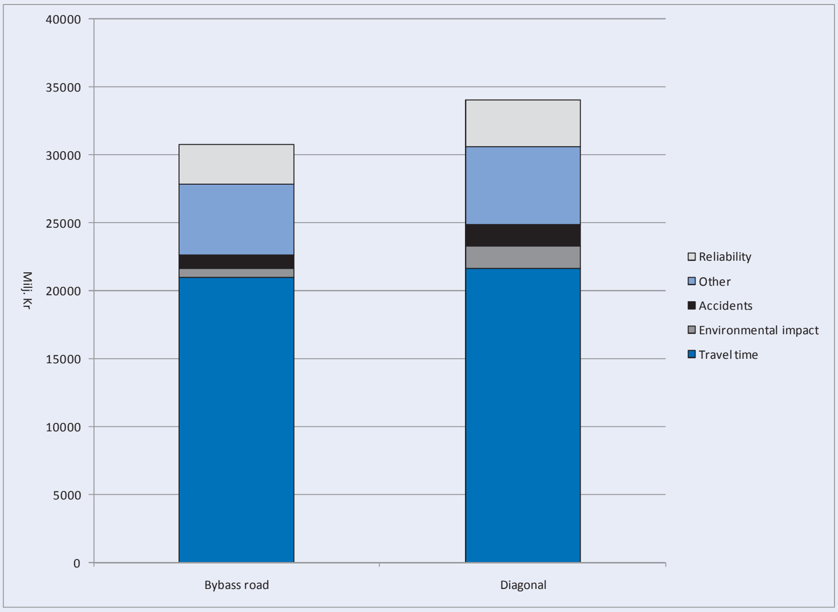
When appraising transport investments, projects are often dominated by improvements in safety and travel time. The time benefits are traditionally measured as the improvement in journey time. In incorporating reliability, those time benefits need to be split into travel time savings and savings in reliability (buffer time). A monetary value is then given to time. Both travel time and buffer time values will vary across users, trip purpose, and location. Differences can be large.

In a small number of countries (the United Kingdom, Netherlands, Denmark, New Zealand, Australia, Norway and Sweden), some project appraisals do incorporate reliability. However, the values used are typically the same for all users. This approach is not fully adequate: the value of reliability is, inevitably, very granular – diverse across users – with a wide spectrum of values. It is important to identify values that differentiate between the transport modes and journey purpose/task. Using a coarse or, worse, a single value for reliability improvement will distort the outcome; even more so if the value is not location-specific.

Box ES1. **North-South road connection in Stockholm, Sweden**

Cost-benefit analysis for proposed North-South road connections in Stockholm, Sweden, includes an estimate of reliability benefits. The following figure summarises results of the cost-benefit assessment for a Stockholm bypass road and for an alternative project, the Ulvsunda Diagonal. The inclusion of reliability into the calculation added around 12-13% to the user benefits.

Benefits of the Stockholm bypass and Ulvsunda Diagonal options



Source: Transek (2006).

Delivering optimal levels of reliability

At present, reliability is generally not taken into account when evaluating a project. If an infrastructure investment, for example, improves travel time reliability rather than average travel time, such project merits will be overlooked.

To appraise reliability effects in cost-benefit analysis it is important to measure both average travel time and travel time variability. If the assessment fails to separate these two measurements, but argues that the project does indeed improving reliability, the assessment lacks a factual basis.

Incorporating reliability requires three sets of data:

1. Existing travel time reliability, defined in minutes.

2. Anticipated reliability level, *e.g.* in minutes, after intervention.
3. Monetary values of reliability, disaggregated at the appropriate level of granularity.

What this report proposes for incorporating reliability into project appraisal is that the temporal journey time improvement should be split into pure journey time improvement and buffer time (or other temporal reliability measure) improvement *for each granulation*. The change in time savings benefit then equals the change in pure journey time multiplied by monetary value of time, plus the change in buffer time multiplied by monetary value of reliability.

Average time savings should be split into travel time reductions and a reduction of travel time variation. Both of these components should be identified. An appreciation of the traveller types using the link would then enable appropriate values for the components to be applied. This unbundling enables planners to gain insight into the relative levels of reliability benefits.

Further, *ex ante*, cost-benefit analysis will require some quantification of the expected reliability effects of policies. This is a poorly documented field and probably requires some improvement of current traffic forecasting tools and models. Ideally, these tools should be able to provide estimates of future changes in the standard deviations of travel times on links, and model the influence of such variables on travel demand and network use.

Above all, because reliability issues are location, user, and time-specific, assessments should avoid applying or repeating the use of a single value for reliability, or applying a value that has been used in one study to a project in another situation. For each project, there are differences in the mix of user groups and time/reliability splits.

Options for achieving reliability should be selected on the basis of cost-effectiveness

A key policy challenge is to create incentive structures that encourage selection of the most cost-effective reliability option – that is, adopting the option that delivers a given level of reliability improvement for the lowest cost. The objective is to ensure that option is chosen ahead of the less-effective options, regardless of whether the responsibility for adopting the option lies with the network provider or the network user. For instance, one project scenario might conclude that the cost-effective way for shippers to achieve greater reliability would be to hold more stocks than for the network provider to incur incremental infrastructure costs.

In order to be able to take into account reliability in policy impact evaluation, only a cost-benefit assessment framework provides consistency in assessing the societal pros and cons of policy interventions in terms of their positive, or negative, effects on reliability.

Network operator perspective and user perspective should be distinguished

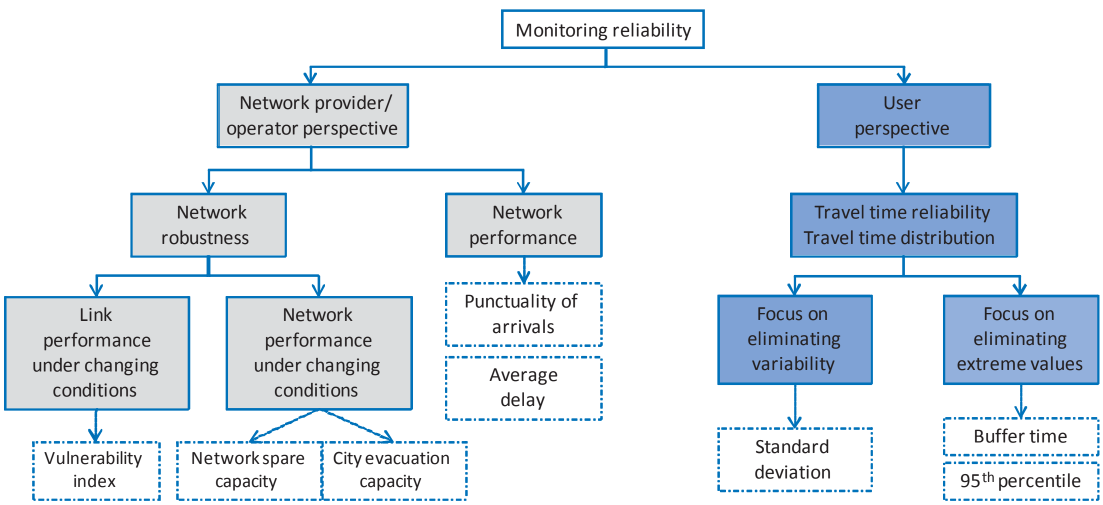
For policy making, it is important to measure and report on both network operator and user perspectives of reliability. The way reliability is measured and communicated provides a policy signal in itself. Also, the better informed regulators are about the appropriate reliability targets, the better the policy.

There is a clear dichotomy in performance indicators: indicators of network quality (what is provided and planned); and indicators of what the user experiences (or how they respond to network experiences). It is recommended that a distinction be made between the system- and user-perspective of reliability.

1. For a network *provider or operator*, the focus is on:
   - System robustness/vulnerability. Here, a further distinction is made between link and network performance indicators, under changing conditions.
   - System operating performance. Here, the focus is on indicators to describe the performance of a system in terms of deviations from expected, or agreed, levels of service.
2. For a network *user*, the focus is on:
   - The variability of travel times experienced by the user. This can provide useful travel-planning information. A further distinction is made between indicators to describe issues regarding general variability of travel times, and issues regarding the elimination of extreme, unexpected, travel times.

Based on the review of existing indicators, the following schematic overview of the different purposes for indicator combinations was derived (Figure ES4). The main conclusion is that it is extremely important to look at both network and user perspectives, as each has different policy implications.

Figure ES4. **Network and user perspective of reliability**



New policy framework

There are many techniques and instruments available that can be used to improve the reliability of the transport network – both individually and in combination with each other. Four principal policy options available to manage reliability can be identified:

- Physical expansion of capacity.
- Better management of capacity.
- Pricing mechanisms to deliver a market for reliability.
- Information systems intended to mitigate the adverse consequences of unreliability (*i.e.* reduce its costs), rather than to reduce the incidence of the unreliability.

These are not necessarily alternative options but, nonetheless, each should be subject to cost-benefit appraisals.

Physical expansion of capacity

On the supply side, infrastructure design and construction can incorporate reliability options. Improving supply-side reliability entails reducing the probability of an unexpected disruption in service. There is a wide array of options to enhance capacity by expanding infrastructure: upgrading and adding line capacity, increasing transport service in corridors and transfer points, construction of new highway lanes and improving alignment, construction of new rail lines and terminals.

Infrastructure can also be built at standards that reduce the need for maintenance or improve the robustness of the capacity. It is notable that these supply side, capital-based, build options are implemented before any incident takes place. Hence, adaptability of the infrastructure is a key issue.

Box ES2. **East Coast Main Line in the UK**

The East Coast Main Line extends 393 miles from London to Edinburgh with a 30 mile branch from Doncaster to Leeds. Investment in the line was undertaken to a low standard compared with other British routes. In particular, economies were applied to the frequency of, and wire tension on, stanchions (electricity poles). The consequence of this standard is that the route is relatively vulnerable to the electrical wires being displaced by winds.

The quality and reliability of the overhead line electrical supply is a cause for concern. This has an impact on performance in two ways. First, as a precautionary measure, electric trains are slowed down in times of strong winds, introducing delays for all journeys on the route. Second, if the wires become detached there can be wholesale delay or cancellation of services until engineers can repair the wiring.

One way to reduce the line's vulnerability to damage from the winds might be to splice in additional catenary support poles (or to respace the poles) at locations where the wiring is most vulnerable to wind damage. However, as there are extended lengths of the northern end of the line that are exposed to the north coast (and the strong north-eastern winds), this remedy could be extremely costly.

Source: East Coast Main Line, *www.absoluteastronomy.com/topics/East_Coast_Main_Line#encyclopedia*.

Supplying new capacity is costly, time consuming, and often politically difficult, while setting higher network standards and improving the robustness of capacity may deliver higher reliability more cost-effectively. It is too often the case that additional supply is considered as the only option, whereas it should be considered as one option among others.

Better management of capacity

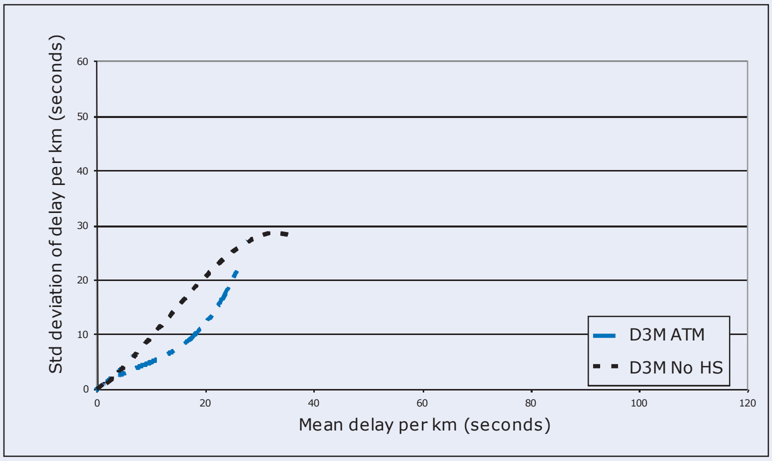
There is a wide range of techniques and instruments available to better manage network capacity to improve reliability. These include pro-active oversight and management of vulnerable parts of the network, and enhanced incident management. For instance, the impact of congestion on reliability can be reduced by the use of variable road speed limits and the temporary addition of road capacity, using emergency hard-shoulder break-down lanes. Similarly, improved management oversight can also be applied on the railway network. Optimised timetabling, a dynamic rescheduling of rail networks in case of an incident, and advanced train management systems can be used.

Box ES3. **Use of motorway hard shoulders in the United Kingdom**

Since late 2005, the UK Highways Agency has allowed use of the hard shoulder of a 16 kilometre section of the M42 motorway at the most congested times. At such times, the speed limit on the section is reduced from 70 miles per hour to 50 miles per hour, with the option for a further reduction to 40mph depending on the operating conditions. The Agency is also investigating the merits of a (higher) 60mph speed limit when the hard shoulder is being used. To provide a safety offset for the loss of the hard shoulder, emergency refuges have been added, providing emergency telephones and lighting, and CCTV monitoring. A variety of monitoring equipment and information systems have been put into place on gantries spaced at 500m intervals. As well as switching between regimes with and without hard shoulder running, the technology is also used to vary the speed limits on the three lanes.

The figure indicates results from active traffic management. The black dashed line (D3M No HS) indicates the variability/average delay plot of the M42. The blue dash (D3M ATM) indicates what happens to this relationship as active traffic management uses the hard-shoulder running regime. The delay penalty due to the speed restriction applied when hard shoulder running is permitted is compensated by reduced variability (improved reliability).

Delay and day-to-day variability for different motorway regimes (M42, UK)



Source: Mott MacDonald (2009).

Enhanced management techniques can assist in reducing the impact of maintenance on network users and reduce the cost of the maintenance activity itself. For example, some contracts in Public-Private Partnership projects have included charges for maintenance works to discourage private network owners from adding too many work zones at the same time.

In summary, an important policy focus for delivering reliability is to better manage capacity through dynamic network management. A focus on interfaces, such as border crossings and ports and hinterland connections where unreliability is likely to occur, might also be appropriate.

Developing mechanisms for charging directly for reliability

Charging for the use of transport networks, or portions thereof, is becoming a more common method of managing traffic demand, and consequently traffic flow and network reliability. It is also possible to charge for information systems, such as GPS guidance systems, which network users can subscribe to in order to mitigate the worst effects of delays. Charges can be applied selectively to segments of the transport network, or more broadly over large sections of the network.

Developments in technology have facilitated an expansion of charging systems and techniques that can be used to manage demand on transport networks. Although most of these techniques are directed at cost recovery and congestion management, they can also be effective in improving reliability.

There are a few situations where access to road networks has been restricted and where charges have been introduced to improve reliability. Dynamic pricing on the Interstate-15 in the USA is one of the few instances. In this case, charges are adjusted up and down to ensure a predictable travel time for the 8 miles of road involved.

Box ES4. **Interstate-15 toll lanes, USA**

Pricing attuned to delivering reliable travel times has been applied to segregated lanes on the I-15 freeway in California, linking San Diego to its northern suburbs. Prices are varied so as to maintain the level of reliability on the tolled lanes. In this case, the speed of traffic is used as a proxy for reliability.

The charges are varied in accordance with the flow of traffic on the road, with charges rising so as to discourage some use of the tolled lanes. Tolls are varied at 6-minute intervals and the level is set so as to attract the level of demand that is consistent with a constant speed. These toll charges also convey information to drivers before entering the free/tolled section of roadway: if the tolls are relatively high traffic is likely to be very heavy on the free lane ways.

Source: Brownstone, K. and Small, K. (2005).

Railways are better placed to use charging as a tool to deliver a consistent level of reliability because full control of access to the network allows network-link charges. There are limited examples (in North America and Australia, for example) where high-reliability freight train services are offered for a premium. In general, however, freight railways' profit-maximising strategy is to move large amounts of freight that does not require very high reliability standards. By contrast, high-speed passenger trains/tracks (ICE, TGV, Pendolino, etc.) have been built to provide near-exclusive rights for services with low transit time and high reliability. Infrastructure charges for these trains are correspondingly high.

In summary, charging directly for reliability by setting differential charges for infrastructure use and service supply, according to the level of reliability, might deliver an appropriate level of reliability. However, it should be noted that it is often difficult (or impossible) to differentiate charges sufficiently to match the level of reliability demanded by different types of user of transport infrastructure. The cost of a charging system that discriminates on the basis of reliability could outweigh its benefits, and must be included in cost-benefit assessments of charging systems.

Mitigating the cost burden associated with unreliability using information

Information systems can reduce the consequences of network incidents. Network demand can be deflected away from the site of congestion or traffic incidents. Information can also reduce stress associated with unreliability, and enable the problems associated with delays to schedules to be managed.

As noted previously, travel time reliability depends, to some extent, on the user's expectation of predictable travel times; this expectation can vary according to the information available. Network providers can facilitate network usage and reduce the impact of incidents by informing users of prevailing conditions. Even if information does not prevent incidents from happening, it can reduce the costs that arise from the incident. For example, the widespread adoption of the mobile phone in recent years has provided the network user with the means to alert interested parties (the warehouse, the family) that arrival will be delayed; the latter parties might then be able to reduce the impact of that delay. Hence, information can mitigate unreliability and reduce the ripple effect that otherwise would be the result of unreliability.

In a specific commercial application, the port of Southampton schedules pick-up times for containers. If a truck is delayed, and will miss its slot, it must phone and reschedule. This can be done up to five minutes before the slot, otherwise a fine is imposed.

Information options may be divided between pre-trip and on-trip measures. Information may be used in different ways to improve reliability depending on whether a traveller has left the origin, whether a traveller can divert to another route, or if the traveller cannot divert but can reduce the ripple effect (consequences). Different tools exist for delivering this information, including variable message signs, car navigators, the internet, and text messages on mobile phones.

Information can be provided to users to mitigate the effects of poor reliability. This is often a cost-effective way to reduce unreliability costs and the cascading impacts of traffic incidents.

Conclusion

A wide range of policy instruments is available to manage reliability. Because there generally is no direct market for reliability, cost-benefit assessment needs to be used to determine appropriate levels of reliability and to select cost-effective policies to manage reliability. Cost-benefit assessment has so far been applied to projects designed to improve reliability in only a small number of countries, with techniques that have been in some important respects unsatisfactory. This report makes significant progress in identifying appropriate methodology for incorporating values for reliability into project and policy evaluations, and it explores the pitfalls that need to be avoided.

Robust and consistent reliability assessments can be developed. Their deployment is important for informing decisions on achieving more optimal levels of reliability on surface transport networks, and for the selection of cost-effective policies and projects.

NOTES

1. This is a common feature of all markets as rarely, if ever, is the array of goods and services so vast that all consumers can select the exact amount of each attribute that they are willing to pay for.

2. For instance, if they cannot be charged directly, they are likely to say they place a much higher value on reliability than otherwise.

3. In this way, reliability is analogous to risk.

1. SETTING THE SCENE

This section provides an overview of the issue of reliability. First, there is a discussion of just what is meant by "reliability". Often unreliability is regarded as being synonymous with congestion, but they are different concepts and this is explained.

In this context, the report then considers the broad sources of transport unreliability, notably, those that can arise from network usage (demand), those arising from network provision (supply) and those arising from factors external to network users and providers – notably, the weather.

Network usage is then examined in more detail. There are two key features. The first is the upward trend in network usage and increasing emphasis on reliability as a key network quality attribute. The second feature of network usage is the wide dispersion across users of the value attached to reliability.

The discussion then considers the signals faced by the players that provide the network and by the network users. Do providers supply appropriate levels of reliability? Do users have appropriate expectations and how do they respond to unreliable services?

The subsequent discussion outlines why optimal reliability solutions may not arise. It addresses the main market failures in the provision of reliability and identifies those market segments with high values of reliability. Case studies reveal there are several key instruments that can improve outcomes. Cost-benefit analysis is then discussed as the key mechanism for assessing and prioritising government strategies and ensuring that the benefits exceed the costs of providing improved reliability.

1.1 Defining transport reliability

Reliability is unanimously regarded as a desirable transport network attribute. However, in a review of literature on transport reliability, it was found that transport reliability was defined in a number of different ways. The choice of definition is important because it has major implications for policy.

This report defines network reliability as being that which provides consistent and, thus, *predictable*, travel times. Thus, the network is reliable as long as travel times are consistent, even if the network underperforms due to regularly-slow speeds caused by traffic congestion.

Applying this reliability attribute to network users, a useful definition recognises that such users then time their actions according to expected network performance. In this report, reliability is defined as:

> *the ability of the transport system to provide the expected level of service quality, upon which users have organised their activities.*

According to this definition, reliability can be improved either by supplying a higher level of reliability, or by changing expectations of the level of reliability. In essence, a *reliable* network has consistent performance and network users are more bothered by reliability as travel time becomes more uncertain. Broadly speaking, network performance can be classified in four main categories:

1. Where the network can be traversed *consistently* in accordance with local speed standards.
2. Where the network *consistently* underperforms. For instance, road congestion on the road link may consistently restrict speeds to 40 km/h rather than the road design speed of 60 km/h.

3. Where the network underperforms *inconsistently*, although the underperformance generally lies within a known band of performance.
4. Where the network underperforms *inconsistently* and so randomly that it is difficult to identify the risk or extent of delay. This can be a particular problem arising with random events of nature.

A consistently slow network has congestion costs, such as the additional driver-wage costs incurred by freight operators, but is nonetheless reliable.

In other words, network unpredictability (inconsistency) – the latter two categories of network performance – is the defining characteristic of unreliability. The more random (less predictable) the performance, the harder it is for the network user to insure against delays. To illustrate the point:

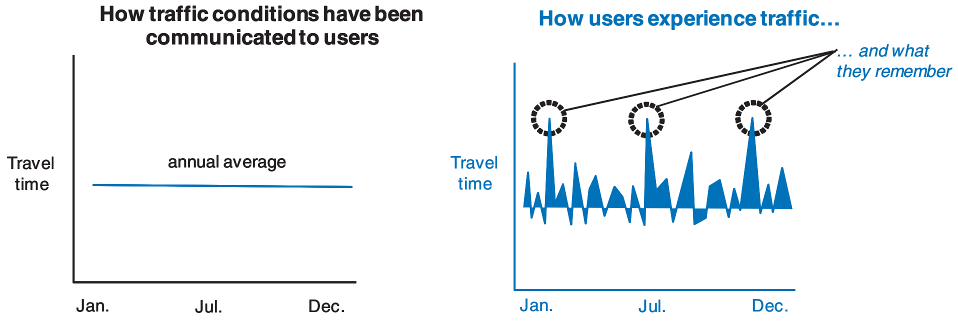
- A network user may know that a journey takes anywhere between 10 minutes and 30 minutes (although there is a small chance of an even longer journey time). This variability tends to occur with recurrent events such as road congestion. If the user adopts a conservative strategy they may allow for a 30-minute trip.
- The incidence or severity of delays is unknown or at least the network user has no information on the network performance; here it becomes difficult for the user to insure against an expected reliability performance. This variability tends to occur with nonrecurring events, such as road accidents.

These areas of unreliability need to be related to the consequences they bring about. Two distinct adverse consequences of journey unreliability arise:

- The first consequence arises during the journey itself: the journey is frustrating and causes stress.
- The second consequence arises from the fact that the travel is usually a derived demand as it enables us to undertake other activities: schedules for our commuting or leisure activities become disrupted as does the scheduled chain of flow of goods. Users may insure against this occurrence by building in extra (buffer) time against expected or unexpected delays.

The feeling of travelling without control over ones' travel time is a frustrating experience and "bad" experiences are often remembered by travellers. Traffic conditions in the past have often been communicated to travellers only in terms of simple averages (left chart in figure 1.1). However, most travellers experience and remember something much different than a simple average of commuting travel time (right chart in figure 1.1). Users may have deeply negative perceptions of unexpected delays, which colour the attitude to the experience.

Figure 1.1. **Travellers' perception of traffic conditions**



Source: FHWA (2006).

The second manifestation that arises from unreliability is arguably the far more significant factor: the unexpected disruption to personal and commercial schedules that cascade from the transport delays. Predictability enables users to plan their journeys to closely match their activities. Inconsistent performance undermines that planning. Where performance is inconsistent, network users may simply have to accept the consequences of the delay, albeit that it may have ripple-effects effects or, worse, snowballing (compounding or growing) effects, affecting other activities or stages in the personal or logistics chain, constituting a cost to those involved.

The ripple-effect of delays is an important reminder of the inter-connectedness of many individual schedules. A delay at one stage in a person's schedule of activities can pass onto delays in later, related or unrelated, tasks. Similarly, while logistics chains are built in a way to reduce their vulnerability to individual events, any delays in individual consignments can reverberate through the chain. Indeed, because the transport task is part of a chain, a break in any part of the chain is a break of the entire chain if the time lost cannot be compensated in other parts of the chain.

1.2 How network users manage unreliability

In contemplating a journey, the network user has not just to consider the expected average travel time but also its variability (including the related uncertainty). If the traveller wants to reduce the risk of being late at the destination, more time will have to be allowed than the mean travel time (the so-called "safety margin" or "buffer").

A comment needs to be made about transport reliability and its costs. It has been recognised that an insurance against unreliable network performance is that the network user modifies their behaviour. Building in buffer time and buffer stocks of goods (or making emergency shipping) are important forms of insurance – and it is assumed that network users taking out such insurance are concluding that such action is less costly than the consequence of late arrivals.

But it is important to recognise that these forms of insurance are, in themselves, not costless. Holding additional stocks of goods "just in case" involves a capital cost both in terms of the storage facilities and financing the stocks. Building in buffer time for goods can be problematic, with goods arriving for onwards delivery or processing before they are required: this might reduce the efficiency of the dispatching or production processes. On the personal front, arriving early for an event might not always be welcome, but might be less costly than arriving late. More importantly, leaving earlier to ensure arriving on-time consumes the time available for other, potentially more productive, activities.

The need to build in these buffers – that is, when the safety margins *can* be perceived – can be costly. Significant variations in travel time therefore reduce the overall efficiency of undertaking the tasks that rely upon transport. Unexpected delays in passenger transport generate costs in the form of prolonged waiting times during travel or costs in the intended activity if the buffer time is inadequate, creating scheduling problems such as missed or delayed connections or appointments. In freight transport the unpredictability may lead to missed connections or to assembly production delays while waiting for delayed components. If conservative approaches are adopted to manage the time variability there will be fewer opportunities gain from just-in-time approaches to distribution, production, and stock management. As a result, the use of vehicle stock becomes less productive, more trucks will be required on the road and higher warehousing costs will be incurred.

Ultimately, of course, the more random the event – the less predictable it is – the less it is possible for the shipper to make informed insurance plans. As is discussed further, below, some sources of network disruption, such as natural events, are less predictable than other events.

Box 1.1. **Warehousing logistics**

McKinnon *et al.* (2008) has noted for UK chemical products such as plastic milk bottles, *current* plastic production/filling/retailing processes are essentially seamless in that there is no warehousing activity at all and no current facilities. Bottles go directly from production to being filled and then to supermarkets (McKinnon *et al.* 2008, p. 40). Thus, if unreliability introduced the need for buffer stocks, this would be a supplementary stage in the logistics chain and it is conceivable that the additional stock handling would increase costs significantly.

1.3 Distinguishing unreliability from congestion

The terms unreliability and congestion are often used synonymously. However, a congested network does not have to be unreliable. A network link that is always congested may still be reliable. For instance, a road may always consistently have "bumper-to-bumper" slow-moving traffic but the consistency in the speed (albeit, slow) enables network users to plan their travel accordingly. Similarly, a generally-uncongested network link can be unreliable – for instance if the road profile is unsafe or the road is subject to flooding, resulting in a high risk of incident-related delay.

The distinction between reliability and congestion is important because of the different policy implications. For instance, a common policy response to a congested road would be to expand capacity. However, if the road is unreliable due to flooding, the response might be to provide road users with sufficient advance warning to enable them to detour when the road is flooded or to elevate the road to reduce flood risk.

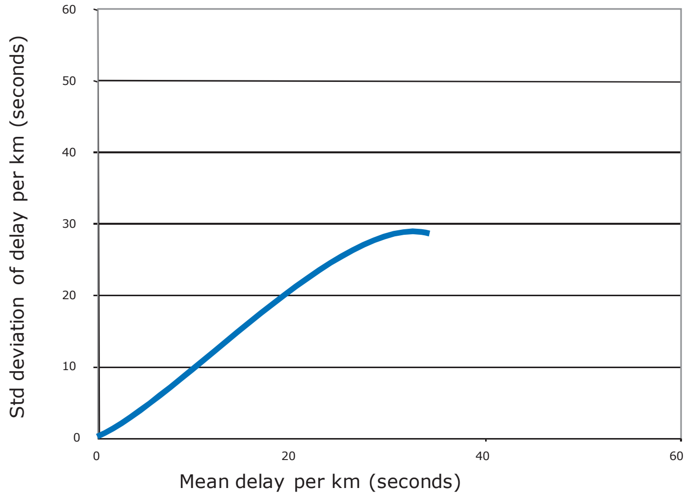
It is recognised, however, that remedial actions directed at congestion can improve reliability and similarly, actions that improve reliability can reduce congestion. That is, there can be overlap between policy initiatives. Nevertheless, this report focuses on reliability and suitable policies directed at responding to reliability issues.

The *unpredictability* of a journey is a defining feature of unreliability. It defines the extent to which the network user can manage the situation. It should be noted that this *can* distinguish "reliability" from "congestion". As long as journey time along a congested road is fairly predictable, the road link is reliable.

That said, congestion increases the likelyhood of unreliability: as traffic levels increase, the time delays due to slight perturbations tend to increase more than proportionately. A consequence of this variability is that network users take out "insurance" on the network being congested: the more the network is congested, the greater the insurance that users take against delays. This is illustrated in the following figure showing the relationship between reliability and congestion a motorway in the United Kingdom. There is a clear correlation between the level of congestion and reliability until high levels of congestion are reached. Because reliability declines with increasing congestion, network users need to increase their buffer time. That said, it is not possible to say whether the variability of travel time was predictable or not.

The distinction between reliability and congestion has important implications for the assessment of congestion costs especially on road links. Put another way, unreliability is a key negative consequence of congestion but measuring congestion should not be used as a proxy for the level of unreliability because there is no automatic direct linkage between congestion and unreliability.

Figure 1.2. **Relationship between congestion and reliability on the M42 motorway in the UK**



Source: Mott MacDonald (2009).

There is a further consequence of the uncertain relationship between reliability and congestion. The failure of road performance assessments to isolate reliability performance from prevailing speed estimates on congested roads has implications for the value of actions to enhance network performance. Indeed, in this context it is notable that studies have found that costs of unreliability can sometimes rival those of congestion. Bearing in mind that the results are *not transferable across locations*, it is nonetheless significant that recent studies found that costs related to unreliability caused around half of underlying costs of delay and that improved reliability might add up to 25% to the welfare benefits compared to time savings (DfT 2006; CEMT/ITF 2007). In other projects these benefits may be closer to zero.

1.4 Sources of transport unreliability

Unreliability can arise from activities that are within the control of the network user or the network provider. Unreliability on the transport infrastructure network arises from two primary sources:

- Unpredictable demand-related traffic interactions between users (congestion).
- Unanticipated supply-related events:
  - Traffic incidents (crashes and vehicle break-downs).
  - Natural events (*e.g.* floods and earthquakes).
  - Network maintenance (causing temporary reduction in supply).
  - Mismanagement in infrastructure supply, which can also include inappropriate maintenance programs.

Reliability literature sometimes classifies the regularity of the disturbances as being "recurrent" (such as weekday peak hour congestion) or "non-recurrent" (such as floods and other events of nature). The essence of the degree of recurrence is that it provides information about the predictability of the event; this report uses mainly that latter terminology.

Deliberate, destructive interventions, such as terrorist attacks and threats, can be a major source of transport unreliability. Analysis of this specific, complex issue lies outside the immediate focus of this report. Nonetheless, the principles of methods and analysis set out in this report have common features that can be applied to that specific issue.

Mismanagement of road and railway networks can reinforce other sources of unreliability. It is possible that an uncongested road can be unreliable if the network is poorly-managed; similarly a congested road with poor management is likely to magnify unreliability. This observation is represented by the intersection of the circles in figure 1.3, showing the primary sources of unreliability.

Figure 1.3 provides a reminder of the interfaces – or inter-linkages – between the sources of unreliability categories. For example, low standard infrastructure is likely to be more prone to unreliability arising from events of nature than if the infrastructure is set to a high standard. This is not to argue for infrastructure to be built to a high standard by default. Given prevailing conditions, such as the likelihood of disruption and low levels of traffic, it may be highly appropriate for the infrastructure to be built to a relatively low standard.

Figure 1.3 also illustrates the essential connection in network reliability between the provision and management of infrastructure (the supply) and its usage (the demand). Sumalee and Watling (2003) present a conceptual framework for the analysis of transport network reliability. They argue that the transport network is a system in which the interaction between demand and supply in the network is the main mechanism defining the service state of the network. The important characteristic of the system is its exposure to various causal sources of variation.

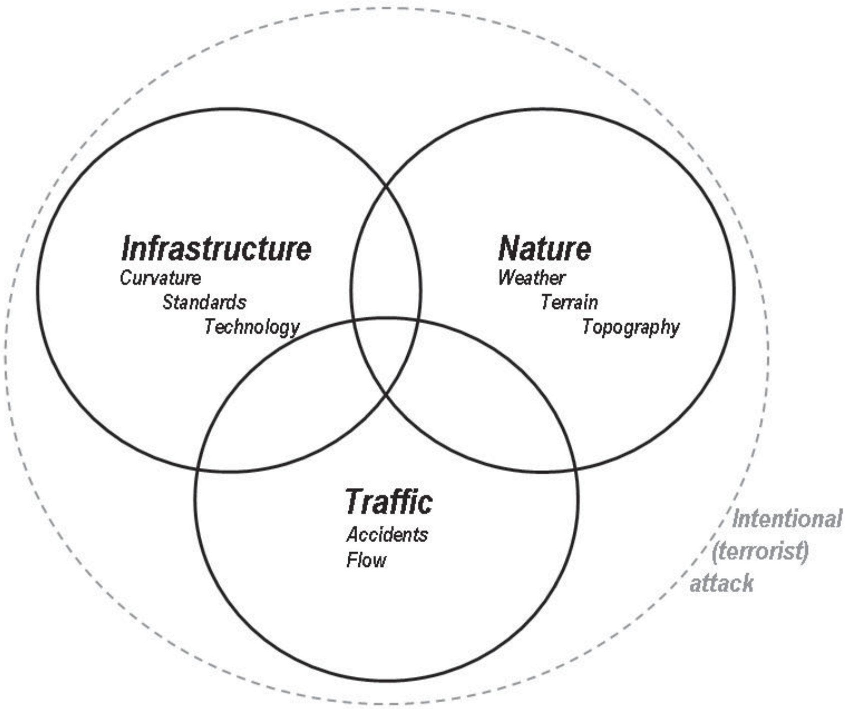
This performance variation can be an element of both demand and supply sides of the network. The variation in performance can arise from an expected or unexpected incident, and the impact can be permanent or temporary. Usually, the network supply interacts with various external factors, such as weather conditions or natural/manmade disasters, which all can cause variation in the link capacities of the network. On the other hand, the demand also fluctuates, both in the recurrent and sporadic (*e.g.* special event) sense, with variation both within the day and between days. All of these causal factors in the system lead to a variable service state of the network.

Reliability issues are very location- and time-specific – and this affects potential actions to manage the problem. For example, in Figure 1.4, it is evident that reliability issues on the Hanshin Expressway are dominated by problems arising from traffic incidents (particularly from accidents) more so than high traffic volumes (albeit those high volumes exacerbate the consequences of an incident). Incidents such as accidents, road works, breakdowns, etc. greatly influence travel time delays. In this example it is noted that the road authority has identified a persistent cause for unreliability; this information in itself can then form an important focus in any management or containment of the reliability problem on the road.

The observation that performance variation arises from both demand and supply elements provides an initial view as to potential policy mechanisms that can be applied to modify reliability levels. One dimension involves focusing on supply (infrastructure provision and network management) and the other dimension focuses on demand (modifying user behaviour).

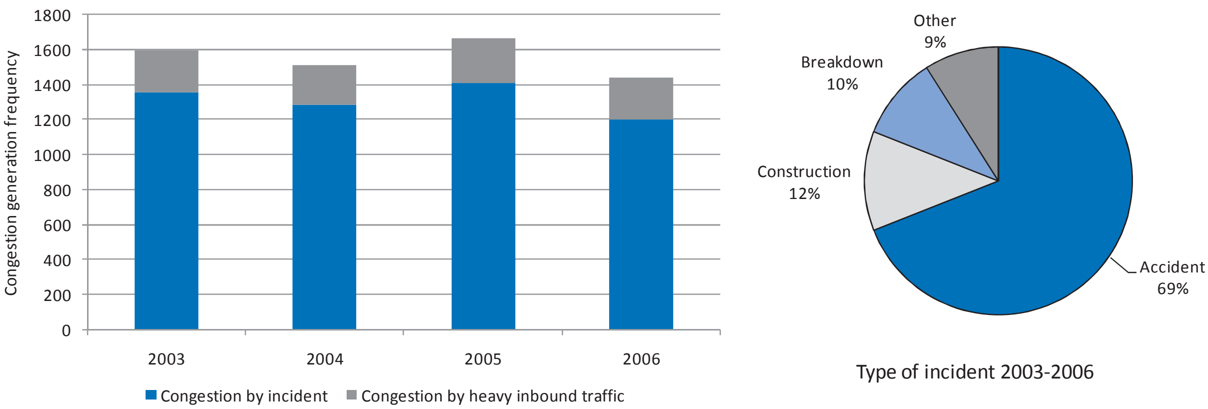
In summary, network reliability performance depends on the occurrence of traffic, nature, and infrastructure events; the impact of incidents is lowered or raised by prevailing infrastructure management and traffic levels. Policymakers and planners face a challenge: to identify the main causes of unreliability on a given link or network and to establish an incentive structure to ensure that the least-cost options are adopted first.

Figure 1.3. **Primary sources of unreliability and inter-relationships**



Source: Derived from Husdal (2004).

Figure 1.4. **Incidents on Hanshin Expressway in Japan, 2003-2006**



Source: Okutani (2008).

The challenge for policy makers grows with rising traffic volumes that provide their own sources of unreliability and exacerbate the impact of other sources of unreliability. The challenge is further heightened because, as the following discussion outlines, there is both a general demand for maintaining levels of reliability and at the same time some network users seek much higher reliability standards.

1.5 Reliability and transport trends

The market for reliability, as expressed by the demand for and the supply of reliability, inevitably changes over time. It is both a product of and a contributing factor to trends in the transport sector.

In fact, changes in commerce and personal travel patterns have increased the importance of reliable transport system. Reliable transport networks and services are required because of more complex and inter-related supply chains and increasingly complex scheduled activities. The physical way that the economy operates has changed, facilitated by – and demanding – transport system enhancements. The changes are within and beyond the sector. A range of economic trends, arising in a large part from the long-term decline in transport costs, have restructured the physical way that the economy operates. The main sources of the declining costs are:

- Significant improvements in infrastructure.
- Productivity enhancements in road and rail vehicles and equipment (such as the development of the container and complementary handling equipment), generating an integrated management system for handling goods from production line to customer.

These developments have facilitated and accompanied the growing operational sphere of influence for businesses. Multinational businesses have consolidated into larger, but fewer, physical locations, growing with the globalised economy. Most marked amongst the changes are:

- Offshore outsourcing of production, bringing lower parts and labour costs.
- Adoption of just-in-time stockholding, enabling inventory costs to be reduced.
- Larger-scale factories, in fewer locations, generating a range of scale economies – the "phenomenon of fragmentation" discussed above.
- Development of fewer, larger regional warehouses, reducing distribution costs.

The geographically-dispersed business activity has therefore been facilitated by, and spurred the need for, the investments in infrastructure and growth in transport productivity.

The importance of scheduling in personal and freight activities has grown so transport unreliability has an increasingly-marked effect on downstream activities. Modern retailing, wholesaling and manufacturing rely on reliability. In modern dispersed production systems, "time" has become the critical factor where *timely* delivery of components has replaced traditional stock-holding. Deardorff (2003) describes the specialisation, or outsourcing, of activities across the global economy as being the "phenomenon of fragmentation". Manufacturing plants are larger but more specialised in componentary and the final product relies on drawing together those "fragments". For the purposes of this report, fragmentation perhaps better describes the production-transport task better than the outsourcing term. Deardorff describes the increasing importance of time and, especially, keeping to schedule. The implications of fragmentation as he describes here are certainly closer to "snowballing" than to "ripple" effects:

> "Time can be especially important here [with fragmentation], since delays in delivery from one stage to the next can shut down the whole chain. Much has been made of the "just-in-time" production methods of the Japanese, introduced in the 1980s, with the usual

interpretation that these methods lessened the costs of holding inventories. This they certainly did, but the more important contribution of these methods may have been the flexibility they offered, especially as production became fragmented across locations, to respond quickly to changes in the need for inputs at various stages of production. ... [I]t appears that the role of trade is becoming increasingly important ... the importance of time seems to loom ever larger in many products." (Deardorff 2003, p. 19).

To the extent that it is representative of broader trends in other industries, in OECD and other countries, the changes in UK brewing industry production and logistics (Box 1.2) provides a strong message for the management and supply of reliability. The example reveals that large productivity gains can be captured from restructuring manufacturing and distribution – notwithstanding the need to embrace complex logistics systems, to operate longer distribution chains and (because of the complexity and lengths of the logistics links) to face the threat of greater unreliability.

Box 1.2. **Logistics restructuring: brewing stockholding and distribution (UK)**

The process of centralisation has been underway for many years in the UK. In the last decade, one large brewer has reduced its number of primary stock locations from 13 to 4, while in the last few years another brewer has reduced its local depots (servicing pubs and restaurants) from 45 to 32. Thus, these depots serve a larger area, despite the fact that worsening congestion during this time have lengthened shipment times and increased shipment time variability. Indeed, companies have not been deterred from pursuing centralisation despite these trends. This is partly because the economic benefits of centralisation – in scale economies, in lower inventories and reduced fixed costs – far exceed the additional reliability and other transport costs.

Source: Paraphrased from McKinnon *et al* (2008).

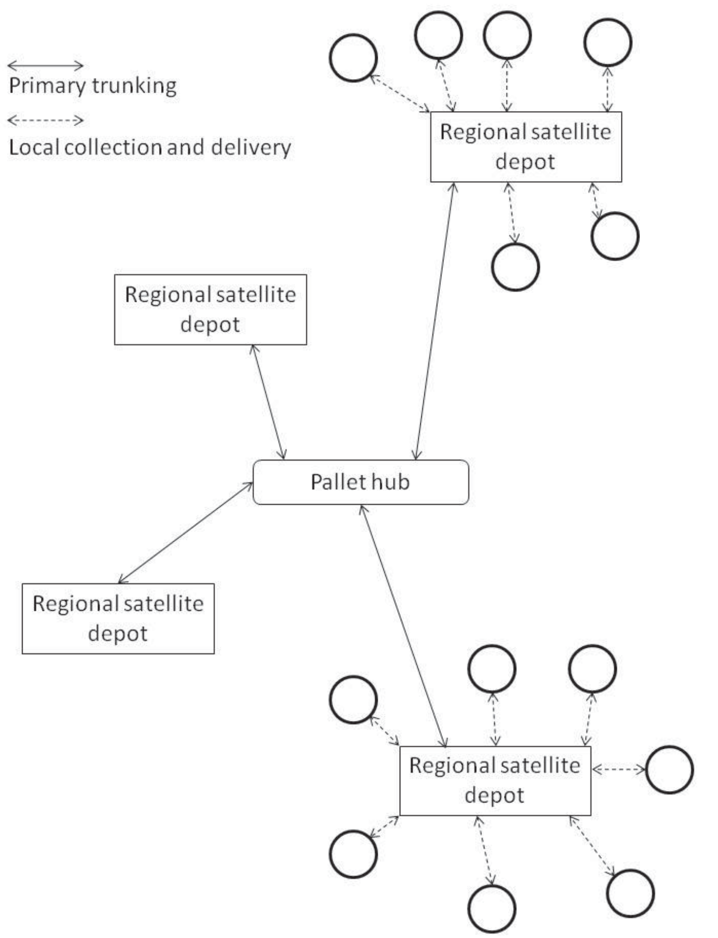
McKinnon *et al.* (2008) present a profile of the growth in hub-and-spoke networks in distribution of goods and the centralisation of production and local inventory. This is illustrated in Figure 1.5. Local inventory holdings and the production locations have been rationalised to centralised locations. For instance, the study notes that the brewing sector in Britain has adopted this approach despite the threat from the worsening congestion, which increases the variance of transit times. This apparently – paradoxical behaviour is explained as being:

"partly because the economic benefits of centralisation, in scale economies, lower inventories and a reduction in fixed costs can far exceed the additional transport costs, even allowing for a congestion cost penalty" (McKinnon *et al*. 2008, p. 11).

Further, the authors note that long-distance movements between the hubs are undertaken overnight, when congestion is relatively low. The congestion arises more at the links radiating from the regional satellite depots.

These transport conduits in the national and international trade links do, however, increase the vulnerability of the supply chain to perturbations. First, there are greater volumes of goods being moved. Secondly, the goods are being moved further. Thirdly, the supply chains are increasingly complex and interdependent, managed by new logistics processes. The longer, complex supply line between businesses and customers increases the likelihood for system and capacity disruptions at some point along the chain. Finally, if disruptions arise, the financial consequences may be more significant than previously because of reduced stock inventories (which provide buffers against unreliability but are costly to keep).

Figure 1.5. **Hub-and-spoke structure of pallet-load and express parcel carriers**



Source: McKinnon et al. (2008).

The changes in commerce provide crucial productivity dividends in the growing national and international economies. While companies inevitably desire greater network reliability, they can reduce exposure to delays by altering their own schedules. In the UK, for instance, the proportion of night-time deliveries (as measured in truck-kilometers) increased from 16% to 20% in the decade to 2005 (McKinnon *et al.* 2008). This shift was facilitated by some easing of restrictions on out-of-hours deliveries, by extending opening hours at commercial premises and by expanding the use of reception boxes for depositing deliveries when the premises are closed.

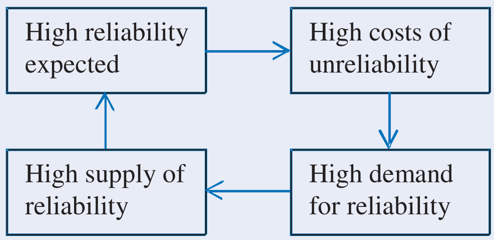
There are also changes in personal lifestyles. Passenger movements, both for business and social purposes, have become more complex with changing patterns of employment, increased disposable income, recreational choices and leisure time. These diverse and geographically-spread activities have led to more intensive use of transport systems, bringing greater dependence on transport to be reliable so that delays do not cascade through the busy calendar of events. The scheduling approach increasingly adopted in private lives echoes the just-in-time deliveries in commerce.

Box 1.3. Expectations and demand for reliability

A greater number of market and non-market activities are dependent on transport and, more specifically, on its reliability. Choice of organisation and location of companies are often based on expectations about transport reliability. For instance, the reconfiguration in the commerce activities of manufacturing and distribution reflects a greater reliance on reliable transport links.

Reliability has become something of a victim of its own success. The more it improves, the more behaviours are based on the assumption of good reliability, leading to choices on organisation and location which are partly irreversible and which make them more vulnerable to unreliability. This, in turn, creates more demand for reliability and – if this demand is reflected in the market – encourages operators to improve the reliability of supply.

Figure. **Expectations and demand for reliability**



Thus, the expectation from these demand trends is increasingly that transport should provide high levels of reliability. Both personal and commercial transport users can be locked into schedules, whether they are chains of commuting, childcare/educational, social and leisure activities or freight logistics chains.

Nonetheless, network users still have diverse needs for reliability and, for many, the cost of delivering such standards (which may be very high) may exceed the value that many players would place on high reliability. For other users, the network providers may not find a practical solution to provide that reliability. These issues are now considered.

1.6 Granulated values of reliability

The demand for reliability varies across users, products, locations and firms. Demand for reliability is differentiated – or "granulated". So the efficient "level" of reliability is *not* one level but, rather, a range of levels. So, strictly speaking, efficient *levels* of reliability is the appropriate term.

Keeping to schedule is often critical, such as for perishable goods *or* where the goods are an integral part of a complex logistics schedule. In such situations, operators are likely to place a high value on reliability. They will be willing to pay a premium on transport costs to ensure the goods arrive at the destination within a tight timescale albeit, taking their own "insurance" actions (such as buffer time and buffer stock). At the other end of the spectrum, where timeliness is not the key factor, firms will not be prepared to pay a premium for a high level of reliability. Table 1.1 below illustrates the range of sensitivities to reliability in a range of product groups – varying from high sensitivity in fresh food to low sensitivity in bulk products.

Table 1.1. **Sensitivity of thirteen product-groups to travel time reliability**

Factor/Product-group	Rapid depreciation product	Rapid depreciation process	Stock-keeping strategy	Stringent customer service requirements	Irrationality	Supply-chain power	Direct influence end-consumer/agility	Time windows/ continuation of disruption in supply chains	Total sensitivity assumed
1. Consumer goods slow/fast		*	*	*	*	*	*	*	++
2. Food (fresh)	*	*	*	*	*	*	*	*	++
3. Clothing	*		*				*	*	++
4. Other durable consumer goods	*		*			*	*	*	++
5. Paper/printed matter	*	*	*				*		++
6. Parts/semi-manufactured products						*		*	+
7. Instruments/tools/equipment/machinery	*		*			*			+
8. Car-parts/trucks/cars etc. (automotive)	*	*	*	*	*	*	*	*	++
9. Waste matter						*		*	0
10. Building material		*						*	+
11. Dangerous goods		*			*	*		*	+
12. Dry/liquid bulk		*				*			0
13. Products sold via internet (b2c)			*	*	*		*	*	++

Source: Kuipers & Rozemeijer (2006).

The spectrum of values that are placed on reliability by different network users is also important for personal travel. For instance, Brownstone and Small (2005) concluded that women value reliability more than twice as highly as men. One reason why this may arise is that women are managing a more complicated schedule of activities than men. Because of this, it seems that women are predisposed to value reliability of travel time more highly.

This range of priorities and complexities in personal and commercial schedules is a reminder that using average levels or values of reliability is highly misleading.

1.7 Why network users' diverse reliability needs are not met by network providers

There can be major difficulties in meeting the diverse needs for reliability because, on a common-user road network, the supply of reliable standard is universal to all users, effectively at a flat charge. All users receive the same level of reliability. Network providers may seek to deliver a high level of reliability to meet a sub-market that desires such a standard. However, low value users will also use the open network and, indeed, such usage is likely to introduce congestion and reliability externalities which act to undermine the high reliability that the providers seek to deliver.

Given that reliability is being undermined in this way, it then becomes costly for network providers to deliver "high" reliability and high-value users may not receive their high-standard of provision. Low reliability value users will also feel aggrieved because they face high charges for use of a network that provides a higher standard than they might otherwise be prepared to pay.

The crucial feature, here, that prevents the optimal supply of reliability – ideally, where users receive the level of reliability that they are prepared to pay – is that *typically* there is open access to the network for a fee that does not vary in accordance with the level of reliability supplied. That is, there is no reliability "market". Achieving such a market requires restricted use of the roadway to maintain a given level of reliability – or, at least, reliability that arises from "low" traffic levels – in return for a fee that reflects that reliability level.

The absence of a market can also arise in railway operations, especially where infrastructure charges are not market-based. For instance, it may not be possible to deliver a specific standard of freight train reliability if government fiat decrees that passenger trains should receive priority movement over the network irrespective of whatever infrastructure charge they pay.

The other principal reason for the failure of a proper-functioning reliability market arises from game-playing between those who demand the service and those who supply it. The absence of a proper-functioning market in road network provision may lead to over-statement of reliability needs by those who most value it, knowing that those who do not value the high reliability will nonetheless pay for it.

The situation is not unique to roads. A shipper that is dependent on railway transport may lobby the regulator to provide a higher standard without bearing the full costs and risks associated with raising that standard. Conversely, where railways are the only transport supplier – shippers or passengers have no practical alternative modes to use – the railway may take advantage of that situation to deliver a lower (less costly) reliability standard.

Such game plans are important considerations in the *efficient* supply of, and demand for, reliability. For instance, some network users may lead network providers to incur high costs in delivering high reliability standard when in fact the network user may be able to deliver that standard far more cheaply by modifying their own behaviour. Box 1.4 provides an illustration where higher transport reliability (with costs and risks borne by the railway) may be more costly than the shipper modifying their behaviour (in this case, increasing their buffer stocks as insurance against lower reliability on the railway).

It is worth noting that an absence or low level of reliability does not necessary imply a market failure. It may simply indicate that the cost of achieving higher levels of reliability exceed the benefit to the users. The critical element for the government is to identify the source of the assumed market failure and, if possible, design instruments that attack the market failure at source. Hence, identifying the source of unreliability is a first step in improving reliability.

As a general guideline, it follows from the discussion above that unreliability is likely to arise in situations where:

- Unrestricted network access makes it impossible to differentiate service level.
- Service is provided by a monopoly, leading to under-provision and over-charging for goods.[1]
- Service is provided by multiple operators or there are intermodal connections or interfaces that results with uncertainty of responsibilities.
- Activities rely on "just-in-time" processes or highly scheduled activities.
- Users of the network have little or no option to rearrange their activities in case of unreliability.

Box 1.4. **Coal transport logistics in USA**

The USA's energy production is heavily reliant on coal-fired power stations. Most of the coal is delivered to those stations by train. In recent years, there have been notable episodes when railways have not provided sufficient haulage capacity. This has arisen when demand for coal has surged unexpectedly, when weather disrupted rail capacity and following railway network reorganisation.

The issue is centred on the Powder River Basin coal fields in the US States of Montana and Wyoming.

In the last decade, the North American railways have enjoyed a surge in freight, particularly in international intermodal traffic. After a long period of relatively minimal investment, the major companies have responded to this growth by adding capacity.

The crucial issue, however, is that the railways are reluctant to build sufficient additional capacity to provide a buffer for service contingencies. Indeed, the railways question whether the shippers would be prepared to pay for that buffer. In terms of the logistics chain, it is pertinent to ask whether railways should supplement their capacity buffer or whether it is more cost-effective for power stations to increase its on-site coal stocks.

Power stations can insure against exhausting their coal stockpiles by maintaining sufficient reserves. However, despite a trend towards longer haulage from coal mine to power stations – with consequently greater potential railway unreliability – the shippers have actually reduced their coal stockpiles. Power stations' stockpile rundown increases the likelihood that a railway perturbation will materially affect power stations.

The consequence of power station stocks being exhausted is that the stations have to use more costly fuel (oil or gas) to fire the stations. In an incident in 2005, when the only primary rail line to move coal suffered major damage, the shortfall in coal deliveries led to electricity cost increases of up to 15% (Consumers United for Rail Equity 2005, p. 1).

The policy consideration is whether there should be government intervention in the supply or usage of railway capacity. Two policy responses have been considered. First, indirect subsidy of additional capacity has been proposed through tax breaks for investment in additional track capacity. The second approach would be to increase regulatory oversight to increasing third-party access to track; the on-track competition that is then assumed to occur might then improve coal services.

The Congressional Research Service notes that railways have become increasingly unwilling to guarantee service quality but notes that this may arise because the companies are unwilling or unable to deliver such service quality when their systems are already capacity-constrained (CRS 2007, p. 2).

This case study provides a useful illustration of the cost of unreliable services. The example also shows that shippers can mitigate against the impact of unreliable deliveries. There is a cost in having buffer capacity (which can reduce unreliability). Given the shippers are largely captive to the railways, however, it means shippers are also captive whatever the quality of service that is delivered. Thus the onus for reliability delivery and controlling the consequences of failures in reliability rests on both the railways and the shipper.

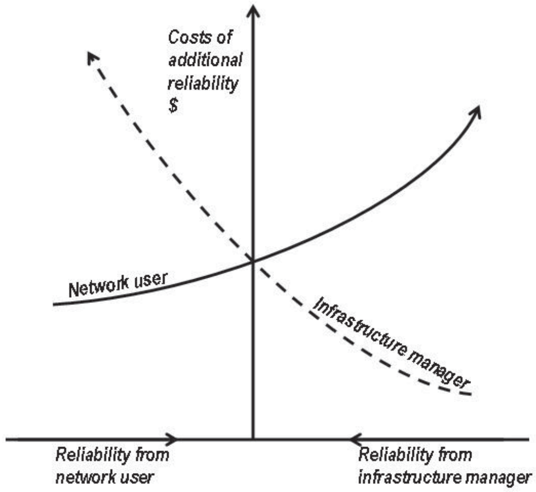
The case study illustrates the reliability trade-off arising from the level of capacity provided. Shippers seek higher capacity (to achieve higher productivity). If railways provided that capacity they would then be taking the risk that the capacity will be used and that the shippers will pay the commensurately higher charges through rail tariffs. With captive markets, railways may be content to supply a lower reliability standard. Shippers may then seek redress with regulators, to enforce higher reliability standards.

Source: Congressional Research Service (CRS) 2007, *CRS report for Congress. Rail transportation of coal to power plants: reliability issues*, Order Code RL34186. Available at *http://ncseonline.org/NLE/CRSreports/07Oct/RL34186.pdf; Consumers United for Rail Equity (CURE) 2005, Rail Report, newsletter, August. Available at www.railcure.org/pdf/newsletter0805.pdf.*

1.8 Determining efficient reliability strategies

Both network providers and network users have a role in managing reliability and, crucially, higher levels of reliability might be more cost-effectively achieved by the network user than the network provider, as illustrated in the following figure.

Figure 1.6. **Infrastructure manager and user costs of enhancing reliability**



The figure illustrates that both network provider and network user can apply strategies for improving reliability. The onus to enhance reliability should not be on one party only – both parties have "low cost" strategies that should be pursued first. For instance, an efficient way for just-in-time shippers to achieve greater reliability may be to hold more stocks than for the network provider to incur incremental infrastructure costs.

Thus both network user and network provider have a role in managing transport reliability.

- Network user actions may be more cost-effective than network provider actions. For instance, shippers' cost-effective options might include transporting goods at less congested or more reliable periods or responding to inherent road reliability problems by holding a higher level of buffer stocks (as noted earlier, UK shippers have shifted some deliveries to quieter, overnight periods). More generally, of course, users need to recognise when their network use is too ambitious for the prevailing network conditions and whether it is therefore more prudent to scale back or reorganise activities.
- Network providers (including governments) also have cost-effective tools, such as managing infrastructure, from which to facilitate the delivery of appropriate levels of reliability.

Whether it is network user or network provider actions, the appropriate government view should be to see the provision of transport reliability as being based on choosing cost-effective strategies. In particular, strategies need to first identify approaches that deliver the best value for money and select projects that perform best under cost-benefit analysis.

Thus, like all desirable features of a transport network, reliability comes at a cost. It is subject to the standard rules of supply and demand: the higher the price, the higher the quantity that will be supplied but the lower the quantity that will be demanded. Conversely, the lower the price, the more consumers will demand it. The challenge for policy makers arises in a number of areas. The first is in formulating the institutional arrangements that impact on the market for reliability. For instance, legal frameworks that prevent discrimination between transport system users can create impediments to differentiating between services on the basis of reliability. The second is in the treatment of reliability when assessing publicly-funded transport infrastructure projects

In other words, the role of the government is two-fold: encouraging a market for reliability and incorporating reliability into the assessment of transport infrastructure projects. In terms of the first role, it is important to note that, as a service attribute, reliability is often bundled with other attributes such as speed, convenience and cost, making it very difficult to differentiate a separate market for reliability.[2]

An important point that follows from this is that only when, say, two parallel services are provided with reliability being the key differentiating feature is there an explicit market for reliability. Without this, there is a major challenge in developing sound estimates of the value placed on reliability by network users.[3]

Ideally, market incentives would encourage not only an efficient level of reliability but would also allocate responsibility for reliability for the party that could bear it at least cost.[4]

Factoring reliability into cost-benefit analysis is clearly desirable but also problematic. The values placed on reliability vary from project to project. Using an incorrect value could result in a worse outcome than a failure to incorporate a value for reliability at all. Cost-benefit analysis, as a set of rationalised economic principles, has evolved over more than a century and useful refinements are unlikely to be developed overnight.[5] However, section three identifies possible methodology for incorporating a value for reliability into project evaluation as well as explores the possible pitfalls that need to be avoided.

1.9 A new policy framework

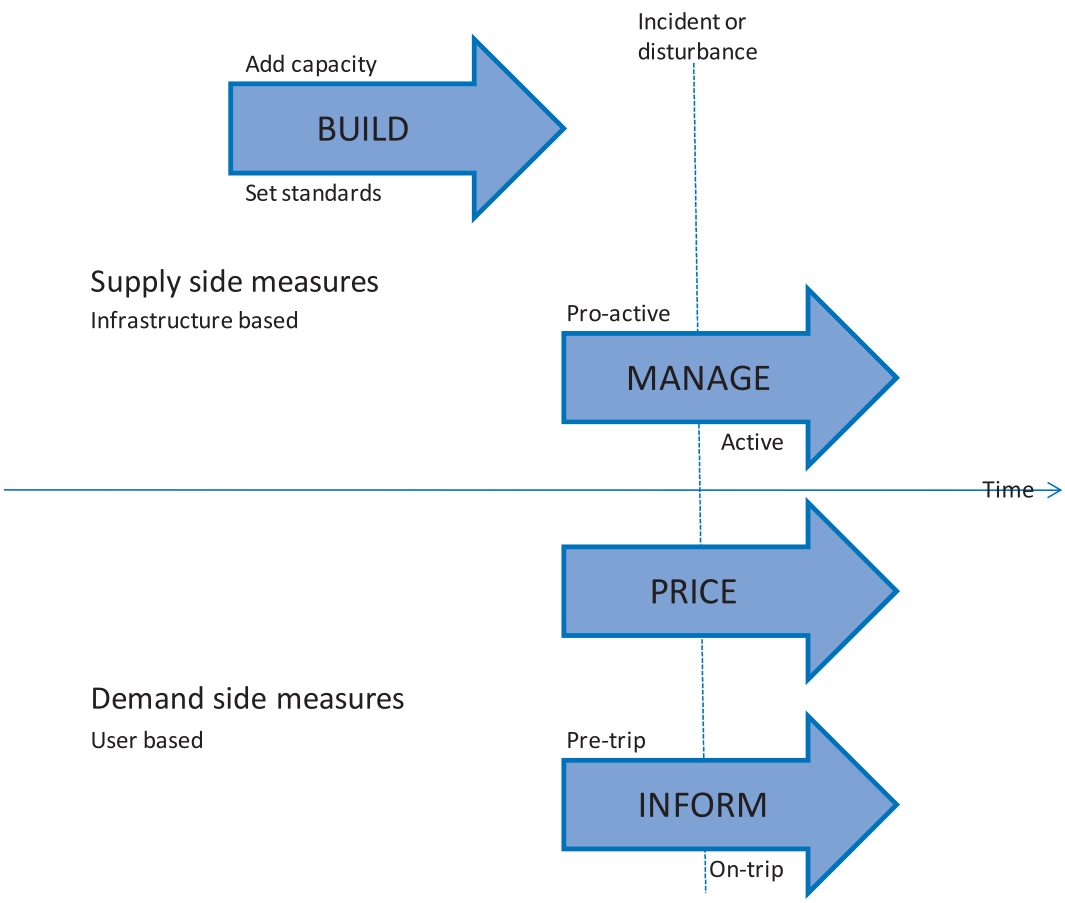
There are many techniques and instruments available that can be used to improve the reliability of the transport network – both individually and in combination with each other. The sources of unreliability identified in Figure 1.3 result in four principal policy options available to manage reliability, illustrated also by the case studies presented in the respective sections of the report:

- Physical expansion and better standard of capacity.
- Better management of capacity.
- Pricing mechanisms to deliver a market for reliability.
- Information systems intended to mitigate the adverse consequences of unreliability (*i.e.* reduce its costs) – these can reduce network demand (to deflect congestion associated with incidents), reduce in-vehicle stress associated with unreliability, and work to ease and manage the problems associated with delays to schedules.

In general, these are not alternative options but, nonetheless, should each be subject to cost-benefit appraisals. Each of these options is discussed in detail in sections 4-7.

Quantifying the value of benefits of greater reliability can be information-demanding. The report considers different options for improving reliability including investments that can mitigate (ease) the impact of an event that causes unreliable travel times. Estimating the benefits from such investments may certainly be difficult to establish.

Figure 1.7. **Policy options to improve reliability**



Monitoring reliability and measuring network users' reliability valuations and options is vital in making a robust assessment of these policy options. Section two therefore considers first reliability monitoring while section three discusses user values of reliability and how those values have been and should be incorporated into cost-benefit analysis.

NOTES

1. Note that a profit maximation strategy under monopoly may also imply better quality and higher differentiation but most likely for a small user group.

2. This is a common feature of all markets, as rarely, if ever, is the array of goods and services so vast that all consumers can select the exact amount of each attribute that they are willing to pay for.

3. For instance, if they cannot be charged directly, they are likely to place a much higher value on reliability than otherwise.

4. In this way, reliability is analogous to risk.

5. For background on the history of benefit-cost analysis in the USA, see *www.chicagoasa.org/downloads/CostBenefitConference2006/benefit%20cost%20history.pdf*.

KEY MESSAGES

- Reliability is the ability of the transport system to provide the expected level of service quality, upon which users have organised their activities.
- Changes in lifestyles and in production, storage and distribution have worked with transport changes to increase the importance of reliability.
- Broadening national and international trade links, with increasing goods movements, have brought greater volumes of goods, moving farther and in increasingly complex and interdependent ways.
- Unreliability is not synonymous with congestion, although congestion increases the likelihood of unreliability; they can have different policy implications.
- Two distinct, adverse consequences of unreliability are that the journey is frustrating – causing stress – and that there can be ripple or snowballing effects disrupting commuting, leisure and business schedules.
- Opportunities to enhance reliability often lie with the network provider and network users – "low cost" strategies that should be pursued first are often available to both parties.
- The absence of a proper-functioning market in road network provision may lead to over-statement of reliability needs by those who most value it, and this is an important consideration in the efficient supply of, and demand for, reliability.
- Reliability issues are very location-specific and this has consequences for potential approaches for managing the problem.
- Network users can insure against unreliability by building in (often costly) buffer time and by holding buffer stocks, although the less predictable the event, the less it is possible to adopt effective strategies.
- Like all desirable features of a transport network, reliability comes at a cost; project cost-benefit analysis should incorporate a value for reliability although using incorrect reliability values could lead to a worse outcome than failing to incorporate reliability.
- Unreliability is likely to arise if there is an open access to the network and, hence, there is no reliability market, service is provided by a monopoly or there are multiple operators with unclear responsibilities, where activities rely on "just-in-time" processes or highly scheduled activities, and where users of the network have little or no options to rearrange their activities in case of unreliability.
- The challenge for policy-makers and planners is to identify the main causes of unreliability on a given link or network and to establish a planning and assessment framework and structure incentives to ensure the least-cost options are adopted first.

2. MONITORING RELIABILITY AS A POLICY SIGNAL

2.1 Introduction

In the foregoing discussion, it was concluded that reliability in transport has become increasingly important to commerce and personal lifestyles. As a network provider, Government may be a pivotal player in delivering optimal levels of reliability. It is therefore appropriate to ask whether government have strategies for ensuring that optimality. The key factor in such policy strategies is to identify prevailing reliability levels. That is, monitoring reliability can provide a policy signal. This section considers broad reliability indicators that can guide policymakers and planners.

A number of stakeholders take a potential interest in the reliability of the transport system and the efficiency of their services depends on the reliability of the system. These include:

- The direct users (car and truck drivers, public transport passengers, carriers, shippers, etc.).
- The network manager responsible for the quality of traffic operations (speed, traffic circulation, safety, comfort, reliability, the possibility of finding a seat in public transport, etc).
- Other network authorities, because their system may possibly serve as a back-up in the case of disruptions.
- Public transport operators.
- Emergency services and private services such as breakdown assistance and salvage companies.

A number of countries have been exploring approaches to monitoring reliability. The heart of the task is to find indicators that provide appropriate measures of the *inconsistency* in traversing the network.

A review of existing reliability indicators suggests that an increasing number of countries monitor reliability and a number of reliability indicators are available. However there are also several shortcomings in some of the reliability indicators currently available:

- **Aggregation across users.** Most existing reliability indicators monitor the performance characteristics of the whole system rather than satisfaction of users' needs. That is, whether different users actually receive reliable services.
- **Aggregation across time.** The indicators normally show only overall annual averages and therefore mask shorter-term variations in service standards.
- **Reporting partial data**. More generally, most of the existing indicators were originally designed to provide feedback to network managers, rather than to measure reliability as perceived by end-users. Thus, the indicators may report operational details such as freight train arrival times, rather than those of primary interest to customers, such as predictability of collection or delivery times (BTCE 1996).

It is important to note that "reliability" will mean different things to each of the parties involved. The network provider (or operator) perspective and the user-perspective are clearly distinguishable in these and the differences imply that different parties will want and need to use different indicators representing reliability for their purposes. In the following part of this section recommendations are made on the use of various indicators for reliability for different purposes. In these recommendations the distinction is made between the network-manager and user perspectives.

1. For a network provider or operator, the focus is on:

    - System robustness/vulnerability. Here, a further distinction is made between link and network performance indicators, under changing conditions:

    - System operating performance. Here, the focus is on indicators to describe the performance of a system in terms of deviations from expected, or agreed, levels of service.

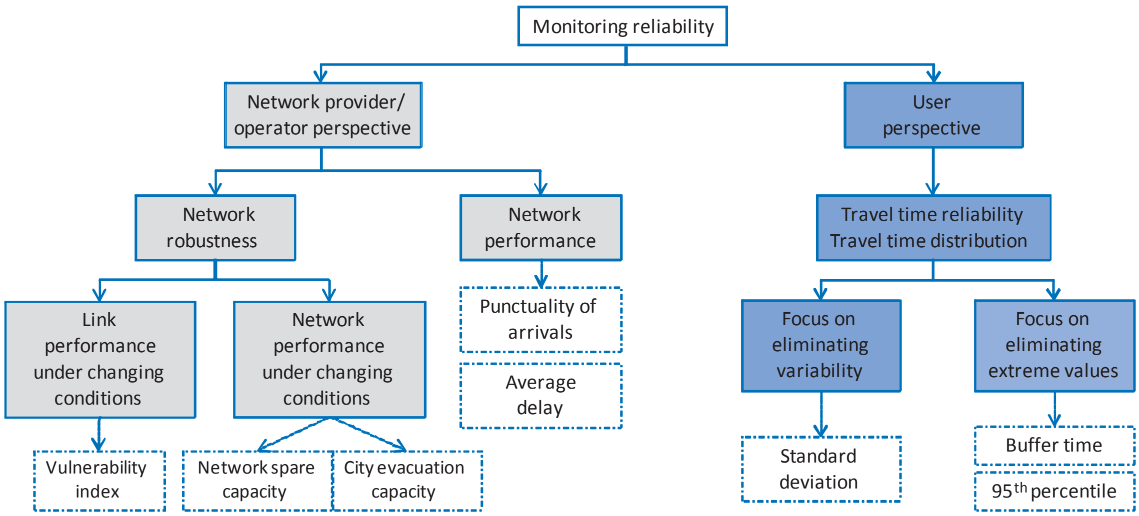
2. For a network user, the focus is on

    - The variability of travel times experienced by the user. This can provide useful travel-planning information. A further distinction is made between indicators to describe issues regarding general variability of travel times, and issues regarding the elimination of extreme, unexpected, travel times.

Thus, there is a clear dichotomy in performance indicators: the indicators of **network** quality (what is provided and planned); and indicators of what the **user** experiences (or how they response to network experiences). For policy making, it is important to monitor and report on both network operator and user perspectives of reliability.

Based on the review of existing indicators, the following schematic overview of different purposes for indicator combinations was derived (Figure 2.1). The main conclusion is that it is extremely important to look at both network and user perspectives, as each has different policy implications. The following part of this section, in which the proposed indicators are presented, is structured along the lines of this figure.

Figure 2.1. **Network and user perspective of reliability**



2.2 Data collection

Reporting on network reliability relies on data of network performance and usage. In principle, it is easier to monitor railway performance because access to the network is restricted, that is that access generally applies to a handful of users (at most) and because real-time usage is monitored. Road access and usage is very different from railways, making data collection a significantly more challenging proposition. There are two main methods to monitor road travel time:

- *Direct travel time measurement* is based on measuring the time interval that a particular vehicle takes to travel from one point to another.

- *Indirect travel time estimation* from traffic flow characteristics – principally, density, flow and speed – is obtained from road-based mechanical equipment such as magnetic loop detectors.

Indirect measurement is relatively cost-effective, in particular, when direct measurement is extremely difficult or costly, or when monitoring equipment for indirect measurement is already available. Direct travel time measurement surveys have been undertaken on specific road links under regional travel time research projects, mainly in the United States and Western Europe.

The lack of data has been a recurrent problem for practitioners in developing their road management schemes. Recognising that travel time is the preferred information for stakeholders (managers and users), filling the gap of data should be a main objective. However, although authorities are aware of the paucity of data, efforts in direct measuring travel times have been very rare.

The following text provides some insights into the challenges in direct and indirect methods of road reliability monitoring.

2.2.1 Direct measurement

There are two main approaches to collecting direct travel time data. One approach is to identify the vehicle in at least two roadside control points. The other approach is to follow specific vehicles. In roadside identification based techniques, the vehicle is identified at the entrance and at the exit of the road segment; this, enables travel time to be estimated. Vehicle identification can be undertaken with, for instance, license plate matching or the Automated Vehicle Identification (AVI) from toll road infrastructure.

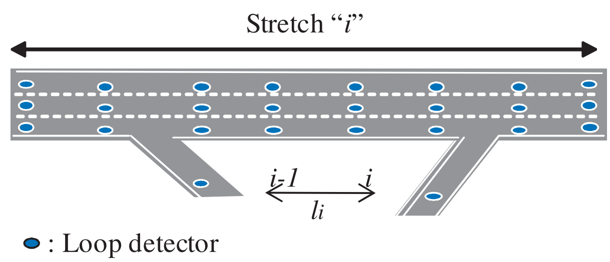
The second group of techniques for the direct travel time measurement involves following the vehicle. It is obviously very costly to collect data with this approach. Historically the monitoring – or "probe" – vehicles have been dedicated cars. The only purpose of these cars is to gather travel time data.

New technologies offer less costly ways of measurement. The development of Intelligent Transport Systems (ITS) and the popularisation of Global Positioning System (GPS) technologies have made it possible that each vehicle which travels in a particular road could be sending data on travel time and its variability. These GPS-equipped vehicles can be regular transport fleets (as buses or parcel companies) which travel regularly over a selected route. Currently, some research is being done to identify the practicality of obtaining travel time data from mobile phone tracking.[1] If these technologies prove to be accurate, there is the potential for every single vehicle to send data on travel time and its variability.

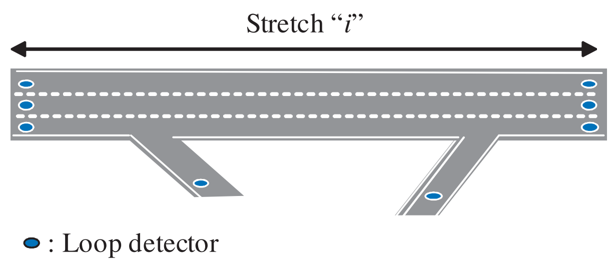
2.2.2 Indirect measurement

Indirect travel time estimation seeks to measure a range of fundamental traffic flow variables (flow, speed and density). The data are likely to be collected in a particular spot of a highway; once collected, the data are extrapolated from these point measurements to a stretch of the highway. In aggregate, these fundamental variables capture the whole physical traffic process, and therefore it should be possible to derive other variables based on them. Loop detectors are widely-used technology that can collect data on traffic flow, speed and density. Simple, single detectors only collect data on traffic flow so speed and traffic density need to be approximated. Dual loop detectors are capable of measuring all three variables.

Figure 2.2. **Required surveillance configurations**



Spot speed algorithm required surveillance configuration



Cumulative flow balance algorithm required surveillance configuration

Source: Soriguera and Robusté (2008).

Travel time estimation from loop detector data arise from two basic methodologies. First and most widely used is the spot speed algorithm. This method is based in the extrapolation of the point speed measurement from the loop to a complete freeway section. The hypothesis considered in the application of this algorithm is that traffic flow characteristics maintain constant in the whole stretch and in the whole time period. This means that for the algorithm to be effective a high surveillance density (loop detectors every 500m) and a frequent actualisation of parameters (every 5 minutes) are needed. Moreover, in highly congested highways with frequent stop and go situations, travel time estimation using this algorithm can be very different from reality. Different approaches differ in how to smooth these variations.

Due to the required high surveillance density and the lack of accuracy of the spot speed algorithm in congested situations, an alternative travel time estimation methodology has been developed. The alternative relies on a cumulative flow balance algorithm, which estimates travel time directly from loop detector flow measurement, without the previous imprecise calculation of speed. The algorithm uses the entrance and exit flows in the highway stretch to calculate the travel time using the conservation of flow equation. Obviously, to apply this algorithm, all the highway ramps must be equipped with loop detector units. The detector drift represents a major issue in the lack of accuracy of this methodology. The required surveillance configurations for both methodologies are presented in Figure 2.2.

2.2.3 Deciding on the level of monitoring

The benefits that come from collecting travel time data for reliability monitoring varies greatly. The benefits vary with the variability of travel time, on characteristics of network and on user requirements. However, because of obvious budget limitations, monitoring schemes need to be prioritised. The level of surveillance will directly affect the accuracy of the resulting travel time information.

The following concept for choosing priorities on network link monitoring has been applied to the Catalan road network in Spain (Soriguera and Robusté 2008). The analysis considered existing data collection, the frequency of congestion and traffic levels. Taking also into account the severity of this congestion, the rush hour durations, and the most suitable technology to measure travel time in each corridor, priorities to implement the information system have been obtained (see Table 2.1).

Table 2.1. **Priority corridors to implement an information system in the Catalan road network, Spain**

Priority	Corridors	Justification
1	Toll highways, primarily those following a SW-NE axis near the coast.	– Severe congestion – Medium frequencies (high unreliability) – High traffic volumes – Low cost (existing tolling infrastructure)
2	Freeways around Barcelona	– Severe daily congestion – High traffic volumes – Already existing intensive surveillance equipment
3	Seasonal corridors – Winter N-S corridors (skiing) – Summer coastal corridors (beach)	– Severe sporadic congestion – Low frequencies (high unreliability) – Low surveillance at the current time
4	Rest of the network	– High cost for low benefits

Source: Soriguera and Robusté (2008).

A very detailed and practical set of guidelines on how to measure travel times, travel time delays, variation, and reliability is presented in the Transportation Research Board/National Cooperative Highway Research Program report "Cost-effective performance measures for travel time delay, variation and reliability" (NCHRP, 2008).

2.3 Monitoring provider or operator reliability

Reliability indicators from a network provider or operator perspective need to be further divided between *robustness* and *performance* indicators. Immers *et al.* (2004) contrasts these two system characteristics: "performance" refers to the performance of a system while robustness characterises the condition of the network.

Robustness indicators should represent whether a system is reliable as a whole in the sense that the system remains its properties and its functionalities under changing conditions or that it is vulnerable to disturbances in functionalities as a result of changing conditions in and outside the system.

Relevant indicators for the *network performance* should represent the reliability performance of a network in terms of deviations from expected or agreed service levels. Different indicators for both types of indicators are presented in the next paragraphs.

2.3.1 Indicators to monitor network robustness

Network robustness (or vulnerability) is a complex concept and several definitions are available. One of the best known is from Ziha (2000), who defines robustness as the capacity of a network to respond to adverse conditions. Network characteristics are important factors to determine the vulnerability and robustness. For instance, in very dense networks, many alternatives will be available for the user in case of the blocking of one or two links, while there will be very few alternatives in case of rural networks or situations with a limited number of links. Initial studies on network robustness focused on few concepts of reliability:

- **Connectivity reliability** considers the probability of the presence (or the lack) of a connection between a given origin-destination pair. Connectivity is useful in representing impacts of catastrophic events such as interruptions due to earthquakes, floods or accidents that completely block a road section and therefore (at least in transition situations) all relative links (Wakabayashi and Iida 1992, Bell and Iida 1997).
- **Travel time reliability** considers the probability of making a trip between two network centres within a specified time interval given daily random travel demand variation (Asakura and Kashiwadani 1991, Clark and Watling 2005).
- Chen *et al.* (1999) also introduced the concept of **capacity reliability**, which is concerned with the probability that a network can accommodate a certain travel demand at a given level of service. This concept focuses on the supply conditions offered by a degraded network.

Recent research has focused on the potential *consequences* of network link failure as well as the *probability* of failure. This research proposed the concept of “vulnerability” or “potential reliability” (Watling *et al.* 2004). D'Este and Taylor (2003) define vulnerability as being the susceptibility of a network to a significant reduction in accessibility in case of loss (or degradation) of a small number of network links. These links should be those whose failure provokes the largest reduction in network performance. An overview of specific indicators for *link* and *network robustness* is presented in the following paragraphs (based on Santos *et al.* 2007).

Indicator for link performance under changing conditions

The failure of some network links can have serious consequences on the performance of the network, particularly when there are no physically close alternative routes. For representing the vulnerability of a network, the vulnerability of a link is defined taking two aspects into account: the number of least-cost routes where the link is included and the traffic flow on the link.

Unlike in electronic networks, for example, taking into account the flow is essential in the transport network. In a transport network, the costs on a link are not constant with the demand. When a threshold is exceeded, costs increase more than linearly. The vulnerability of a link (and the corresponding transportation network) has to take, therefore, into account the cost functions (capacities) of each link in addition to an assignment assumption (Aymerich and Robusté 1990). The number of links and their flows can be weighted according to their perceived importance. A more detailed example of such an analysis is the application of a network assessment tool as described in section five of this report (Robustness scanner).

Indicators for network performance under changing conditions

The interest in the robustness or vulnerability of a specific link is quite limited, as a network operator is seldom responsible for the functioning of only one link in a network. Therefore, the vulnerability of a single link is often less relevant than the vulnerability of a total network. Two suitable types of indicators enabling insight in the robustness/vulnerability of a network are 'network spare capacity' and 'city evacuation capacity'.

Road networks go through important variations of travel demand and infrastructure supply throughout their lifetimes (due to demographic changes, special events, construction works, natural disasters, etc.). These variations require that the network is capable to respond in such a way that a satisfactory level of service is always kept. The lack of this capability can lead not only to serious local disturbances but also to their propagation across the network. In order to represent the aptitude of a network for dealing with these disturbances, *a network spare capacity index* is defined as the sum of the spare capacity of each link, weighted by the total number of kilometres travelled on the link, and dividing the sum with the total number of kilometres travelled on the network. Weights can be applied, for instance, to emphasise the importance of spare capacity in long links with higher traffic flows. An extra parameter can be used to reflect the importance attached to the spare capacity in each link.

Situations of local high travel demand can be critical for the performance of a road network. The capability of the network to allow the fast evacuation of cities can be essential for public safety in case of unexpected events (*e.g.* terrorist attacks, weather calamities) or for the mobility of the visitors of major planned events (*e.g.* music festivals, sport events). This subject has become extremely important in recent years, after terrorist attacks and the 2005 Hurricane Katrina with the flawed evacuation of New Orleans. In order to represent the aptitude of a network for dealing with these circumstances, summing the total capacity of the links with origin in each city, weighted by the population of the city, and dividing the sum with the total population of the region calculate *a city evacuation capacity index*. Weights are applied to emphasise the importance of evacuation capacity in large cities.

2.3.2 Indicators to monitor network performance

In transport planning, reliability performance is generally expressed by the probability of realising trips within a certain travel time. As travel times depend on many factors, the travel times in a given network have some randomness arising mainly from interaction between users and available network capacity as well as variations in road capacity due external factors (see Section 1).

There are numerous indicators used to express the reliability system performance. The reliability of system performance of public transport is often expressed by the punctuality of arrivals and/or departures at stops and stations. Table 2.2 presents a typical public transport system performance indicator.

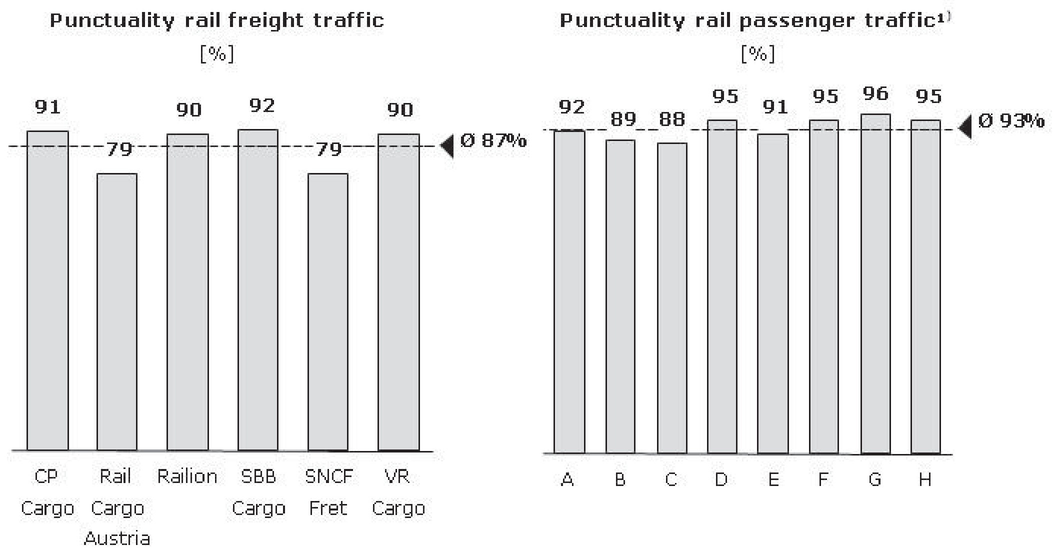
Table 2.2. **Punctuality of intercity rail services in the UK (arrival times only, May 2009**

Long distance route	On time-10 mins (%)
Great Western	94.6
East Coast	92.5
London Midland	90.4
TransPennine	94.7
Cross-Country	92.5
West Coast	84.6

Source: Network Rail (2009).

Reliability in this case is defined as punctuality, or the proportion of trains running 'on time'. The definition of 'on-time performance' however differs between the railways and also between freight and passenger traffic – ranging from 5 minutes to 30 minutes in freight. The following figure illustrates how punctuality can also be rather misleading because the definitions of 'punctual' differ in different countries.

Figure 2.3. **Punctuality of Western European railways**



Source: BSL (2008). Freight train punctuality definitions are not known in detail but thresholds differ between companies from 5 to 15 minutes. Passenger trains are defined as still punctual if they arrive with a delay between 3 minutes and 7 minutes. Due to confidentiality reasons data is made anonymous. Sample consists of Western European countries. Data is normalised.

In comparison with rail, the reliability system performance of road sector is often measured by an average speed level per road category even though a slow speed can still be "reliable". A slightly more complex example from the road sector is presented in Box 2.1.

Box 2.1. Monitoring Network Reliability in the UK

The UK Department for Transport is monitoring the average vehicle delay. It is derived from the differences between observed journey times and a reference journey time, experienced on the slowest 10% of daytime journeys on each of a set of the 103 routes that comprise England's Strategic Road Network.

The monitoring is undertaken, for each 15-minute departure period between 0600 and 2000 hours for each day of the week. Reducing the value of the indicator means faster journey times within the slowest 10% of the distribution, resulting in more reliable journeys overall.

Source: DfT technical note – PSA target 1, available at *www.dft.gov.uk/about/howthedftworks/psa/psatarget1?page=1*.

Thus far we have considered *network* performance indicators. However, as illustrated in Figure 2.1, policy makers need to also monitor *user* experiences and their responses to the prevailing network conditions. This can be illustrated by means of the following example: A goods-train may be six hours late, arriving at 07:00 instead of 01:00. This is a serious network failure. However, from the user perspective, the 6-hour delay will not matter if the shipper can still collect the goods when the depot opens at 08:00. That is, although a train may be late *by railway standards*, the late train may have no bearing on when the goods are available for collection and, thus, end-user reliability is maintained.[2] By contrast, even a modest 15-minute delay in the afternoon may make all the difference between the goods arriving at a factory that day and being delivered at the factory the next day. The focus therefore turns to the user perspective of reliability.

2.4 Monitoring user experience of reliability

The user perspective considers specific characteristics of travel times as experienced by network users and how the users may response to prevailing reliability levels (such as buffer times). Research findings suggest that route travel time characteristics are key reliability indicators on road networks (Lo, 2002, Cassir *et al.*, 2001). In the Dutch national strategic transport policy for example, travel time reliability plays a central role and improving "door-to-door" network reliability is considered a key policy objective (Dutch Ministry of Transport, Public Works and Water Management 2004b).

Despite its obvious importance, there is no explicit opinion on what travel time reliability precisely is or how it should be monitored. Various definitions for travel time reliability exist and many different relevant indicators have been proposed. An overview of such indicators for road traffic can be found in Lomax *et al.* (2003). Van Lint and Van Zuylen (2005) analysed the properties of different indicators and produced a comprehensive overview (see Table 2.3).

There are two primary data sources on the user perspective:

- Data collected by network agencies, replicating typical journeys and journey-time experiences.
- Data collected by network agencies from interviews with users.

The interview approach will be more representative of user experiences and priorities but it will also be considerably more costly approach for monitoring.

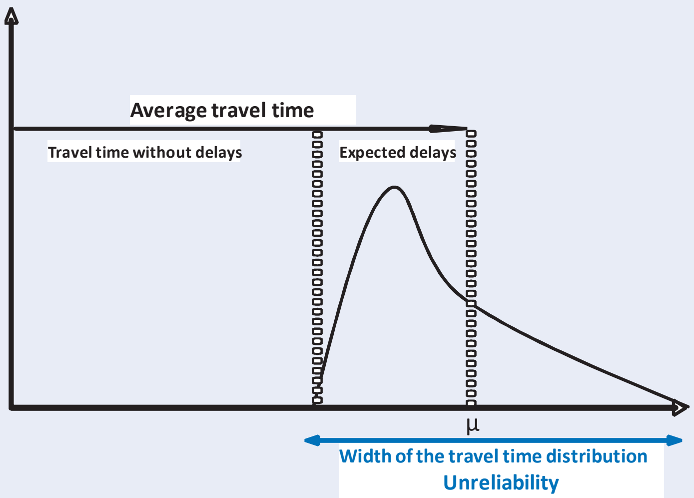
The common feature of all these reliability indicators is that they relate to properties of the (day-to-day) travel time variation – that is, its distribution, and in particular, the shape of the distribution. Day-to-day travel time distribution can be characterised by two characteristics of this distribution; width (journey time dispersion) and skew (the pattern of the distribution of travel time). Intuitively, the wider and/or more skewed the travel time distribution is, on a particular time-of-day and day-of-the-week, the less reliable (predictable) the travel time.

Box 2.2. **Travel time distribution**

Average travel time includes expected and unexpected delays. Unexpected delays lead to variation in travel times. Two forms of unexpected delays can be identified. There is random day-to-day variability, which affects the travel time for journeys undertaken at the same time each day. There are also occasional catastrophic delays as a result of incidents or temporarily unavailability of parts of the network. Thus, in contemplating a journey, the driver has to consider the expected average travel time and its variability. This variability can be quantified by various characteristics of the distribution of travel times over a certain period of time, such as the standard deviation of the travel time distribution[3].

To reduce the risk of being late at the destination, the driver needs to allow rather more time than the average (mean) travel time. This is illustrated in Figure below where μ represents the average travel time. Observing the curve, it can be said that travel time unreliability increases with the widening (the longer the tail) of the travel time distribution. This perspective of reliability as a characteristic related to the travel time distribution forms the basis for the definition of indicators representing reliability.

Figure. **Average-, median- and the travel time distribution**



Source: Dutch Ministry of Transport, Public Works and Water Management (2004b); (Ritsema van Eck *et al.*, 2004).

The user-perspective indicator should be related to the expected travel time value or a variable representing this parameter. Practical examples of the indicators all relate to the reliability of the road system. Comparable information concerning public transport users or individual trips was not found in the literature. Reliability indicators for public transport are generally limited to describing network performance. Table 2.3 presents a range of user-perspective indicators that have been advocated in a range of studies.

Table 2.3. **Several different travel time reliability measures, all for given time of the day or day of the week**

Category	Acronym	Formula	Remarks
Statistical range	STD	$\sqrt{\frac{1}{N-1}\sum_N (TT_i - M)^2}$	Standard deviation or variance of travel time.
	COV	$\frac{STD}{M}$	Coefficient of variation.
Buffer index	BI	$\frac{TT95 - M}{M}$	Buffer index indicates the percentage extra travel time a traveller should leave earlier than on average, to still arrive on time (in 95% in the cases)
	PT	$\frac{TT95}{TTfreeflow}$	Planning Time index, indicates the total travel time that should be planned when an adequate buffer time is included; in this case computed as the 95th percentile travel time divided by the free flow travel time.
Tardy trip	MI	$\frac{M\vert_{TT_i \triangleright TT80} - M}{M}$	Misery index, calculates the relative distance between the mean travel time of the 20% most "unlucky" travellers and the mean travel time of all travellers.
Probabilistic	PR(α)	$P(TT_i \geq \alpha.TT50),$ $\alpha = 1.2$	This indicator calculates the probability that travel times occur larger than α times the median travel time.
Skewness and width	λvar	$\frac{T90 - T10}{T50}$	Robust indicator for width of the travel time distribution.
	λskew	$\frac{T90 - T50}{T50 - T10}$	Robust indicator for the skewness of the travel time distribution.
	UI_r	$\frac{\lambda^{var}\ln(\lambda^{skew})}{L_r}$	Combination of skew and width of the travel time distribution.

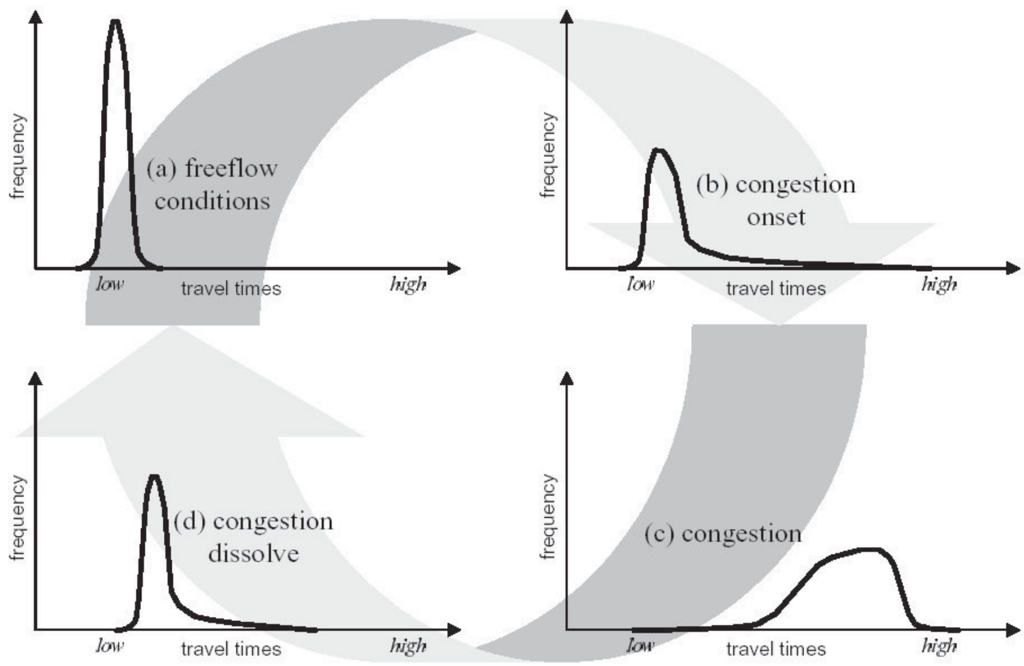
M denotes the mean travel time, *TTi* a travel time observation and *N* the number of travel time observations in a particular time of day or day of week period. *Source*: Van Lint and Van Zuylen (2005).

2.4.1 Properties of different indicators

To illustrate the ability of different indicators to answer the question whether or not travel time should be regarded as (un)reliable on a particular road during specific time-of-day/day-of-week, Van Lint and Van Zuylen analysed travel time data from road segments of the Dutch freeway network. According to the analysis, covering a large number of time periods, the travel time dispersion is very wide (Van Lint 2004). For example, it proved not to be uncommon during weekday peak periods that the 90th percentile value was almost double the median (middle) value. The travel time distribution is also often heavily (left) skewed, particularly in periods when congestion sets in or dissolves.

This is illustrated in Figure 2.4, representing travel time distributions at different periods of the day. Clockwise, the charts show free-flow, the onset and development of congestion and back to free-flow conditions. The charts reflect the pattern of unreliability. The distribution evolves from small and symmetrical (a) to wide and left-skewed (b) to wide and slightly right skewed (c) to wide and left-skewed (d), and back to small and symmetrical (a) again.

Figure 2.4. **Four times-of-day periods and the corresponding shapes of the travel time distribution changes during the day**



Frequency is the proportion of trips that experience a given travel time. *Source*: Van Lint (2004).

The skew in the travel time distributions is very significant since it reveals that a small percentage of travellers experience delays which are much higher than the average delay of the 50% most fortunate travellers. Given the fact that extreme delays also may have extreme consequences, this skewness merits attention – certainly from a traveller's viewpoint, but also arguably from a policy viewpoint.

The application of the different indicators by Van Lint (Table 2.3) showed that different indicators give very different answers to the question whether or not travel time should be regarded (un)reliable on a particular road during specific time-of-day/day-of-week periods. This is partially due to the fact that some measures do not represent well the relatively scarce extreme values in the travel time distribution that are relevant to describe reliability from a users perspective; and partially since not all measures address the skewness of this distribution as an indicator of unreliability.

Nonetheless, none of the measures presented provide undisputable arguments to deduce that travel times are indisputably either reliable or unreliable. That is, there is not a time-based benchmark that enables the policy maker to ascertain whether a given travel time pattern requires remedies. Therefore, in order to base policy decisions on the results for each indicator, a policy maker should translate indicators into economical or societal costs (see section 3 on cost-benefit analysis). A sharply left skewed distribution may well be much more costly (from a economic or societal point of view) than a moderately wide distribution, even if in the first case on average (or in terms of median travel times) travellers are better off than in the second case. However, more complex indicators, accounting in detail with the skewness and/or width of the travel time distribution have the disadvantage of being quite difficult to interpret and communicate and therefore will not appeal very well to the network user.

2.4.2 Indicators to monitor user perspective of reliability

A wide variety of temporal indicators can be used to provide a range of perspectives of the reliability issue. A range of suggested indicators are presented, taking into account some considerations regarding the situation for which they can best be used.

Standard deviation

In situations where there is a need to look at the variability in travel times around an average value and it is expected that this variability is not much influenced by (a limited number of) extreme delays, the travel time distribution will be not very much skewed. In these cases, statistical range indicators can be considered useful.

These indicators generally consider travel time windows in the form of expected travel time plus/minus a factor times the standard deviation (Bates *et al.* 2001, Lomax *et al.* 2003). This "plus or minus" type expression indicates the possible spread of travel time around some expected value, while implicitly assuming that travel times are symmetrically distributed (unskewed) around an average value. The standard deviation (or spread) of travel times can be used to describe the extent of travel-time dispersion. In case of unskewed ("normally-distributed" or bell-shaped) travel times, around 68% (or "one standard deviation") of the travel times are recorded. This can be dissected by days, peak periods or whatever time period is suitable for reporting. A further consideration to use the standard deviation as a reliability indicator is the fact that it is linked to scheduling approach and, because of pragmatic reasons, it is recommended for use in the cost-benefit assessment (see Section 3).

The 95-percentile value

To overcome the eventual problem of not giving much specific attention to possible extreme, the 95-percentile value of the distribution can be used or added to the analyses. This indicator is very appropriate to focus on the width of the travel time distribution and can be very useful to analyse the development of high travel time values. However, as long as this indicator is not combined with information on average expected travel times or delays, the indicator does not directly represent reliability.

Buffer time

The use of so-called "buffer time" related indicators is becoming more and more common. The buffer time can be explained as the extra percentage of travel time due to travel time variability on a trip that a traveller may take into account in order to have a "high" probability of arriving on time. Examples of buffer time related indicators are the Buffer Index and the Planning Time Index, used in the US Federal Highway Administration's Urban Congestion Reports, aimed at monitoring traffic congestion and travel reliability on a national scale.

The *Buffer Index*, as developed in the United States, is the extra time that travellers should add to their average travel time when planning trips relative to the average travel time. In practice, the buffer time varies across the users because of each user's individual experiences with variability and because of each user's individual requirements for arriving at the destination on time. For example, a buffer index of 40% means that a traveller should budget an additional 8-minute buffer for a 20-minute average peak travel time to ensure on-time arrival "most" of the time (where "most" is defined as 95% of the time).

The *Planning Time Index* represents the extra time most travellers should add to a free-flow travel time so as to be fairly confident of arriving at the destination by a certain time. The measure differs from the Buffer Index in that it includes typical delay as well as unexpected delay. For example, a planning time index of 1.60 means that travellers plan for an additional 60% travel time above the free-flow travel time to ensure on-time arrival most (95%) of the time.

Because these indicators use the 95-percentile value of the travel time distribution as a reference of the definitions, they take into account more explicitly the extreme travel time delays. These means that in comparison with using the standard deviation, these indicators better take into account the complete pattern of the travel time distribution and are therefore more appropriate to use in case of expected extreme delay values.

A recent US National Cooperative Highway Research Program report concludes that the Buffer Index appears to relate particularly well to the way in which travellers make their decisions (NCHRP, 2008). The Buffer Index is useful in the user's assessment of how much extra time has to be allowed for uncertainty in the travel conditions. It hence answers simple questions such as "How much time do I need to allow?" "When should I leave?". In addition to the Buffer Time index, the Planning Time Index represents the total travel time that should be planned when an adequate buffer time is included. In the NCHRP report both these indicators are advised as cost effective measures to monitor travel time variation and reliability.

Box 2.3. **Urban congestion report in the United States**

In the USA, the Federal Highway Administration (FHWA) 'Mobility Monitoring Program' aims at monitoring traffic congestion and travel reliability on a national scale. The program objectives are to monitor traffic congestion levels and travel reliability trends using archived traffic detector data, and to provide "proof of concept" and technical assistance to encourage local/regional performance monitoring programs. The archived traffic detector data used were originally collected for traffic operations purposes. Thus the extent of the Program is limited to those cities and roadways where real-time traffic data are collected and archived. The Program started in 2001 in 10 cities. In 2004, the Program has grown to include nearly 30 cities with about 3 000 miles of freeway.

The Program monitors traffic congestion and reliability using the following indicators: Travel time index, percent of congested travel, delay, buffer index, and planning time index. The information in the Mobility Monitoring Program is presented in a very clear and comprehensive way.

Source: FHWA (2005); *http://ops.fhwa.dot.gov/perf_measurement/ucr/index.htm.*

To summarise, buffer time related indicators such as the Buffer Index and the Planning Time Index are appropriate monitors to describe and communicate travel time reliability to planners as well as network users. Other more simple indicators such as travel time percentiles, median travel times and the standard deviation may serve as appropriate indicators, but they should be used with caution, as relevant characteristics of the travel time distributions could be easily overlooked.

2.5 Targets as a policy signal

A number of countries have started targeting reliability. Performance targets are set for three primary reasons:

- Reliability is an important service characteristic in the transport sector.
- The services to which targets are applied often involve monopoly provision underwritten by the taxpayers. Hence governments have an interest in seeking attractive services and efficient provision.
- Reliability targets are important for initiating discussions between politicians, operators, providers and users on the appropriate delivery of service standards.

Most of the existing reliability targets can be found in the rail sector, a transport mode that seeks to run to strict working timetables. Target-setting practice is prevalent in passenger railways. The scheduling of arrival times readily enables these types of targets to be set (setting actual train arrivals against scheduled arrivals) and rail service providers are generally considered a monopoly. To the extent that the provider is perceived to be a monopoly, governments usually oversee supply standards by monitoring and setting performance standards; the target provides a degree of accountability in service quality. Data on service reliability are essential for this oversight. A similar approach is adopted in aviation, where airline service punctuality statistics provide bellwethers for regulatory and policy monitoring.

Thus, with the use of reliability indicators, targets can provide a performance envelope within which operators are expected to work. Such targets may spur transport providers to higher operational standards: providers may perceive that repeatedly failing to achieve a target may lead to adverse government action.

Examples from three countries that have set reliability targets at a broader national level are presented in the Table 2.4 below.

Table 2.4. **Examples of national reliability targets**

Responsible authority	Target	Policy document
Netherlands/Ministry of Transport	Travel time reliability plays a central role and improving “door-to-door” network reliability is considered a key policy objective. 95 per cent of travellers will arrive at their destination on time by 2020.	Mobility Memorandum
United Kingdom/Highways Agency	Average vehicle delay on the 10% slowest journeys is less in 2007-08 than in the baseline period of August 2004 to July 2005.	DfT Public Service Agreement Targets
New Zealand/Ministry of Transport	No overall deterioration in travel times and reliability on critical routes by 2015.	Ministry of Transport – Transport Monitoring Indicator Framework 2008

Governments also use service performance data as bellwethers of individual service health. With such information, governments then often set reliability performance targets for transport operations to encourage service usage, for regulatory and service-provision contract oversight – see the targets set in a range of public transport operations in Australia, in Table 2.5. For example, the government expects most passenger railway franchises to ensure that around 80 to 95% of their train services arrive at their destinations "on-time" (where "on-time" may mean a train arriving within 5 to 15 minutes of scheduled arrival time).

Table 2.5. **Reliability targets in public transport operations in Australia**

Operation		Target (% of services)
Melbourne	Heavy rail	92
	Light rail (all	80
	Buses	95
Victoria	Regional (short)	92
	Regional (long)	92
Sydney	Urban rail	92
	Interurban rail	92
Perth	Urban rail	95
	Urban buses	85
Western Australia long-distance rail		90
Western Australia short-distance rail		90-95

Sources: Economic Regulation Authority, Western Australia, Key Performance Indicators (*www.era.wa.gov.au/3/236/48/key_performance.pm*); CityRail, Customer Charter (*www.cityrail.info/about/customer_charter/*); Department of Transport, Victoria, Track Record (*www.doi.vic.gov.au/DOI/Internet/transport.nsf/AllDocs/8407D0E208BC64A44A256ACE00074FA4?OpenDocument*)

While these targets can provide useful benchmarks of desired performance standards, the target *levels* are often arbitrarily set. Indeed, the cost of achieving such levels may unintentionally exceed the benefit derived. As a related point, targets that provide information regarding averages might lead to delivering level(s) of reliability above or below users' needs, not reflecting the diversity in the demand for reliability.

Thus, targets can, if poorly designed, provide perverse incentives that make the situation worse. The danger is that when the focus is on reliability, providers may, for instance, set less onerous timetables to ensure the reliability targets are met. If achieving the targets is paramount, and other service qualities are not measured, operators may debase other service qualities (such as skipping scheduled stops on a train service to reduce the late-running of a train).[4] Hence, there is a trade-off between reliability and other service levels, such as overall passenger capacity. Reliability targets need to be carefully coordinated with other key performance indicators.

To be of benefit to policy and network provider incentives, then, performance targets need to be realistic and be informed by the costs arising from achieving them – the less "arbitrary" they are, the better. Setting targets has to take into account costs and benefits of these targets, compared to reliability output in the absence of public intervention. Though often difficult to document in practice, this cost-benefit analysis should be at least roughly estimated. It should also be noted that the targets tend to be set in relation to network indicators rather than user perspective indicators and therefore, provide an insufficient picture of reliability performance.

2.6 Conclusions

When monitoring reliability, it is essential to distinguish between network or operator perspective and user perspective. Both perspectives require indicators. Presenting both indicators will also facilitate policy discussion between users, operators and decision makers.

From the network/operator perspective, both robustness and performance should be addressed. There are several indicators available, such as “punctuality of arrivals” and “average delay” to measure deviations from scheduled or expected performance.

Network robustness characterises the condition of the system as a whole. Robustness indicators should show whether the system keeps functioning under changing conditions or is vulnerable to disturbances. Robustness is a complex concept, with many potential indicators. These indicators consider the potential consequences of network link failure and the probability of failure. Some appropriate indicators are the “vulnerability index”, “network spare capacity” and “city evacuation capacity”.

For public transport and freight movement, the focus has hitherto been largely on the network performance perspective. This perspective appears to be more developed for these categories, compared to the use of the road system by private cars, because these indicators play a role in monitoring service level agreements. On the other hand, there is not much research available on the user perspective for public transport and freight transport. On the basis of the examples available, the user perspective appears to be mainly applied in the domain of car traffic.

Despite its obvious importance there is no consensus on single “best” measure for the user perspective. Various statistical measures for travel time reliability exist, and subsequently a range of indicators have been proposed.

Because they relate particularly well to the way in which travellers make their travel decisions, the United States’ Buffer Time Index and the Planning Time Index provide effective indicators to describe and communicate travel time reliability from a user perspective. However, all indicators need to be used with some caution, as relevant characteristics of the travel time distributions, such as width and skewness could be overlooked.

Initial response to unreliability for many countries has been to set targets for reliability. Reliability targets and performance indicators for services and infrastructure performance can facilitate discussions between users, operators and decision makers regarding the “right” levels of reliability. But employing fixed targets may be distorting as they can dominate other service characteristics that may be of equal, or greater, importance. Such targets also invariably present an average level of reliability not reflecting the diversity in the demand for reliability. To be of benefit to policy and create appropriate incentives for network providers, performance targets need to be realistic and be informed by the costs arising from achieving them – the less “arbitrary” they are, the better.

NOTES

1. Transportation researchers from the University of California, Berkeley are, together with Nokia, in a field experiment, testing the feasibility of using GPS-enabled mobile phones to monitor real-time traffic flow while preserving the privacy of the phones' users *www.universityofcalifornia.edu/news/article/17289*.

2. See, for instance, BTCE (1996), *Quality of rail freight service*, page 40. "This [observed late-train running] performance was less inconvenient to customers than it looks, however, because some of the late trains arrived during the night, when Kewdale [the end-terminal] is not open for collection of cargoes".

3. The standard deviation measures the dispersion of a set of values. Where the values are close to the average value (the mean), the standard deviation is small; where a lot of the values are spread away from that average, the standard deviation is large. In the context of reliability, a high standard deviation in travel time would imply a high level of uncertainty in travel times.

4. See, for example, "Why trains miss out stations to save time", *Sunday Mirror [London]*, 28 November 1999. *http://findarticles.com/p/articles/mi_qn4161/is_19991128/ai_n14497878*.

KEY MESSAGES

- In monitoring reliability, policy makers must appreciate the distinction between the network/operator perspective and the user perspective. These perspectives require different indicators.
- From a network provider and operator perspective, both robustness and reliability are relevant parameters to consider.
- Network robustness is a complex concept, characterising the condition of the system as a whole. There are many available indicators, which consider the potential consequences of network link failure and the probability of failure. Appropriate indicators are the "vulnerability index", "network spare capacity" and "city evacuation capacity".
- There are several indicators available to monitor network performance reliability, such as "punctuality of arrivals" and "average delay", describing the deviations from planned or expected performance.
- Despite its obvious importance, there is no consensus on the best measure of travel time reliability. Various definitions for travel time reliability exist, and subsequently also many different relevant indicators are available.
- Indicators such as the USA's Buffer Time Index and the Planning Time Index are useful to describe and communicate travel time reliability from a user perspective, because they relate particularly well to the way in which travellers make their travel decisions.
- The initial response to unreliability for many countries has been to set targets for reliability. Reliability targets and performance indicators for services and infrastructure performance can facilitate discussions between users, operators and decision makers regarding the "right" levels of reliability. However, they can also be distorting as they are usually based on averages.

3. INCORPORATING RELIABILITY INTO COST-BENEFIT ANALYSIS

A number of indicators were identified in the previous section that can be developed as policy signals to flag the need for possible government action on improving reliability. As discussed in Section 1, in the absence of direct pricing mechanisms for reliability, the cost-benefit analysis represents the best option for delivering optimal levels of reliability. Subsequent sections will review four main policy instruments for which cost-benefit analysis needs to be used.

3.1 Cost-benefit analysis as a tool

Cost-benefit analysis (CBA) techniques in transport planning often incorporate estimates of benefits and costs that are often not explicitly priced, such as travel time benefits that accrue to road users. Survey techniques are used to provide estimates of how much the users value infrastructure enhancements. However, cost-benefit analyses typically do not recognise "reliability" as an explicit network quality attribute.

The case for assessing network reliability attributes explicitly in transport policy is stronger than previously thought. Network users are perceived to place greater value in a reliable transport network than has been apparent. However, evaluations of such policies rely on the rigorous application of cost-benefit analyses to assess the merits of different reliability options.

In current cost-benefit approaches, it is possible that analyses do implicitly account for reliability; this can arise when "value of time" estimates are derived from observed network user behaviour. This is because observed user behavioural changes can reflect journey time reductions and reliability improvements even though the benefits are only labelled as "journey-time" benefits. Because stated preference surveys are often calibrated from such observed behavioural changes, it is then possible that stated preference estimates of value of time also incorporate an element of value of reliability.

Unreliability of travel times is increasingly recognised as a major network characteristic that should be separately recognised when considering investment options. A number of policy measures identified in this report illustrate that government interventions are motivated by reliability improvements as well as the headline task of improving average travel time. Some projects are carried out specifically in order to improve reliability. However, there are very few cases where reliability is formally incorporated into the cost-benefit assessment (and hence in the decision making process. In a small number of countries (the United Kingdom, Netherlands, Denmark, New Zealand, Australia, Norway and Sweden), some project appraisals do incorporate reliability.

Difficulties of incorporating reliability as a separate attribute into cost-benefit analysis are acknowledged and it is likely that refinements will not be developed overnight. A number of promising techniques are emerging, however, and these are discussed in the following chapters.

3.2 Applying cost-benefit analysis to reliability policies

Incorporating reliability requires three sets of data parameters:

1. Existing travel time reliability, defined in minutes.
2. Anticipated reliability level, *e.g.* in minutes, after a policy initiative.
3. Monetary values of reliability, disaggregated at the appropriate level of granularity.

Bearing in mind this issue, it is worth summarising how cost-benefit analysis applies the value of time and how reliability would be incorporated. A project is assessed on the basis of the *incremental* benefits that are accrued (or change in the consumer surplus). For a network enhancement, for instance, the current approach identifies the average number of minutes of time saving (typically measured on a per annum basis) and multiplies those minutes by the "value of time" for each identified user group (typically, "business" and "leisure"). Thus the incremental time savings arising from a project can be given an incremental monetary value. This can be illustrated in the following equation, where "Δ" signifies "change in".

Money Value of Time Savings Benefit = Δ Journey Time x Monetary Value of Time

It can be seen that the change in benefit includes the scope for an investment to bring about benefit by reducing the journey time; the journey time is monetised by attaching a value for each temporal change for each granulation (user group).

What is being proposed for incorporating reliability is that the temporal journey time improvement should be split into pure journey time improvement and buffer time (or other temporal reliability measure) improvement *for each granulation*, as such:

Money Value of Time Savings Benefit = Δ Pure Journey Time x Monetary Value of Time + Δ Buffer Time x Monetary Value of Reliability

Thus, to explicitly incorporate changes in reliability, revisions to the CBA approach would involve:

- *Temporal adjustments:* splitting the time (minutes) into time savings for average journey time and time savings from improved reliability (such as the reduction in buffer time).
- *User granulation:* more granulation in (greater differentiation between) network users when considering the "value of time" and the "value of reliability" – for example, differentiating between freight service providers, commuters and leisure travellers.
- *Monetisation adjustments:* having separate measures of the value of time and the value of reliability.

The consequence of these revisions to the CBA framework would be that the calculation of incremental monetary time saving benefits would be split into two distinct measures of incremental time saving benefits for each user group:

1. Time savings x value of time (*revised* estimates for each identified user group – "revised" when the original value estimate implicitly incorporated reliability).
2. Buffer time savings [or other reliability time reductions] x value of reliability.

These estimates will be country-, location-, user- and time-specific. Each project that is analysed must also identify the appropriate degree of granulation – that is, the number of network user groups for whom monetary estimates are required.

What is apparent from the two equations is that it is quite possible for the outcome of any given cost-benefit analysis to be completely changed in level, in situations where:

- The level of granulation changes – if granulation is increased, then diversity in reliability preferences is likely to change the outcome.
- The change in the pure journey time plus the change in buffer time is different from the previous temporal measure (the change in journey time) – the temporal measures are unlikely to remains the same if granulation is increased to reflect diversity in reliability requirements.
- The monetary value of time is different for the "pure journey time" and the "buffer time". However, the challenge for surveying should not be understated. This issue is considered further, later in the report.

3.3 Existing treatment of reliability in cost-benefit analysis

3.3.1 Options for arriving at reliability parameters

To recap the discussion in the previous chapter, incorporating reliability in cost-benefit analysis requires data parameters on existing travel time reliability, anticipated reliability level after a policy initiative, and monetary values of reliability disaggregated at the appropriate level of granularity.

Anticipated reliability level after a policy initiative

To obtain more data on the impact of policy initiatives, such as adding a new lane, on travel time reliability is clearly a challenge. Ideally improvements to traffic forecasting techniques should be able to provide estimates of future changes in the standard deviations of travel times (or other metrics reflecting travel time reliability, *e.g.* changes in buffer time) on links and be able to model the influence of such variables on travel and demand and network use. Specifically for cost-benefit analysis applications, estimates for the future development of the standard deviation of travel times are required. Many of the existing traffic and transport models used for strategic impact assessment have a steady state form. That is they will provide predictions of average demand flows and average (or expected) travel times. These models can also be used to provide predictions of the change in the standard deviation of travel time (see for example Eliasson 2006).

Of the more fundamental approaches, the characteristics of the travel time distribution are treated as endogenous variables in the system, using prospect theory and the development of activity based modelling in combination with multi-agent simulation (see for example Aviniery and Bovy 2007; Rieser *et. al* 2007).

While these approaches are being developed, the focus will be on the more pragmatic approaches, in which the estimation of the effects of specific exogenous factors on reliability is the basic principle. Such approaches require basic research into the probability distribution of the variation in such factors and into their specific influence on demand and capacity.

Three possible pragmatic approaches to solving the issue of inappropriate models have been utilised and deserve further research:

- The reliability analysis can focus exclusively on the impact of incidents on travel time reliability, as these are an important cause for unexpected variations in travel times in situations where the transport network is operating under-capacity. Classic steady state queuing theory can then give predictions of the average additional delay associated with each incident type, from which a travel time distribution can be calculated.

- For situations where incidents are not the principal cause of travel time variability or the network is over-capacity a detailed model of a network or corridor can be developed in a software package that explicitly models travel time variability (*e.g.* microsimulation).

- A relationship between travel time variability and specific traffic demand related variables calculated or used as input by the existing models can be developed for policy measures that are being appraised. By means of further basic research into the relationship between travel time reliability and the demand related variables used in the model, the capability of these simple tools can be further enhanced.

Several existing (micro)simulation route choice models are available to generate information on travel time variability under different assumptions. Such information can be useful to predict travel time on network links. Specific examples of these are developments on real time forecasting of travel times (see, for example Linauer, *et al.* 2006; Hollander & Liu 2005). In his recent dissertation Tu (2008) has formulated the specifications of such a model, specifically designed to evaluate the impacts of a wide array of exogenous influencing factors and traffic flow characteristics on travel time reliability on freeways.

The British reliability parameter measures travel time variability, in common with the approach adopted in a number of other countries. In focusing on the “unexpected” variability, the measure assumes that people have a good understanding of the average travel time (given the time of day and particular patterns associated with day travelling). The unpredictable element then can be divided into day-to-day variation in travel times and unreliability associated with incidents (unforeseen occurrences which lead to reduced capacity, such as accidents or breakdowns).

Monetary values of reliability

A methodology to establish generally accepted values of reliability has yet to be defined. So far, there is no established consensus on how to define and value travel time variability.

The traditional practice is to consider reliability as a qualitative measure, not a quantitative measure, in textbook logistics costs (see ECMT 2005). Nonetheless, significant research has been undertaken to provide insight into perceptions and economic values of travel time reliability in recent years. Recent studies such as HEATCO[1] (2006) and Fosgerau *et al.* (2008) have recommended defining travel time reliability in cost-benefit analysis as the standard deviation of travel time. This places the focus in cost-benefit analysis on reliability arising from both recurrent and non-recurrent congestion.

In general, research shows that reliability is highly valued by travellers and commercial vehicle operators (SACTRA, 1999), reflecting the fact that a reliable transport network is a net benefit for society and that an unreliable (or vulnerable) network represents a net cost to society (Husdal, 2005).

Thus, while planners recognise the benefits of reductions in average travel time, there is an argument for also placing values in benefits of improved reliability. Based on a review of British research, Wardman (2001) also concluded that reliability can be seen as an important quality aspect of a journey.

Box 3.1. **Reliability module of the Dutch National Model System**

The reliability module of the Dutch National Model System (NMS) can be used to assess the impact of policy measures on travel time reliability. It is a specific, temporary module, for use with existing transport models used to calculate future traffic demand and congestion under different policies and scenarios. Enabling the future estimation of the effects of specific exogenous factors on reliability requires basic research into the relationship between such a specific exogenous factor and the base model variables.

In the module four indicators define travel time reliability:

- Probability of a trip being "on time" (defined as less than 10 minutes deviation from the expected travel time for short trips and less than 20% of the expected travel time for long trips).
- Probability of a trip being not "too long" (defined as more than 10 minutes deviation from the expected travel time for short trips and more than 20% of the expected travel time for long trips).
- Percentile-10 of the speed distribution.
- Percentile-90 of the speed distribution.

The model predicts intensity (number of cars per hour) and travel speed on each major road in the Netherlands (characterised by route length and speed limit). Therefore, the underlying empirical model for the reliability module was restricted to use only available variables (*e.g.* vehicle intensity, travel speed, route length, speed limit) as its input.

The modelling seeks to find empirical functions that relate the four reliability indicators to the national and regional model variables (travel speed, route length, and speed limit). Regressions, using the least squares method were used to find the best functional form and the best model coefficients. Data were obtained from road induction loop readers from 212 routes on the highway network.

The best suitable regression function was defined by average travel time, average speed limit and journey length. The function is then used to forecast future journey time reliability. The module can also predict the impact on reliability of a range of exogenous factors, like accidents (minor and major), road works and precipitation. When the observed frequency of occurrence or the impact of an exogenous factor changes, the average speeds (as predicted by the national and regional models) are revised and new reliability values are calculated.

Source: Kouwenhoven *et al.* (2006).

Box 3.2. **Incident related variability – INCA in the United Kingdom**

The availability of high-frequency data from the UK strategic highway network has expanded considerably over the past few years. This has meant that quantifying both the level and changes in day-to-day variation have become possible to a statistically satisfactory level of precision. The data demands are considerable.

To calculate the variation in travel times, the journey times for a particular link have to be observed repeatedly under comparable situations. Due to the Highways Agency's network of road induction traffic loops and a variety of complementary data collection systems (such as that by the private firm Trafficmaster or the ITIS collection where tracking devices have been fitted to a sample of vehicles), data has been available for researchers to provide a picture of variability in day-to-day travel times. Such research requires considerable preliminary work. Separating out those observations that report on incident related variation in travel times is a complex preliminary stage.

The starting point for analysing incident related variability was a detailed study of the likelihood of different types of incidents on different road types. The study determined twelve categories for incidents analysing detailed logs of those occurring on four motorways. Then for each category, the characteristics of the incident were determined, such as the time taken to clear, the average number of lanes closed or remaining available and some measures of diversion potential. One of the important parameters for each incident category is the rate of occurrence. Incident rates are measured as incidents per million vehicle kilometres of flow. Using the data available, category incident rates can be estimated for dual three lane motorways with and without a hard shoulder. The table below indicates those incorporated into the modelling.

Table. **Default incident rates on three lane motorways (per million vehicle km)**

Incident category	Standards hard shoulder	Non-standard hard shoulder
Multi-lane accident	0.0267	0.0439
Single lane accident	0.1173	0.1473
Non-HGV breakdown	0.1047	0.1544
HGV breakdown	0.2412	0.2877
Non-HGV fire	0.0084	0.0146
HGV fire	0.0110	0.0201
Load shedding	0.0025	0.0068
Debris	0.1928	0.4175
SL emergency roadworks	0.0410	0.0833
ML emergency roadworks	0.0118	0.0187
Spillage	0.0022	0.0074
Animal	0.0032	0.0027
Combined fire, spillage, load shedding	0.0140	0.0140

An incident leads both to delay and variability around the estimated average delay. This was analysed providing parameters for each incident category that can be used to assess the expected change in the variability of journey times were a particular incident to occur.

When an incident has occurred, the modelling of the delays imposed on road users analyses the impacts on the vehicle facing maximum delay and then assumes that the delay faced by the average vehicle will be half this. This assumes that demand is constant during the incident. There is a recognition that the flow levels will vary significantly by time of day and different flow groups are constructed to represent this heterogeneity in flow levels. The delay impact of any incident will vary for each group.

For each flow sub-group, the maximum delay is calculated using the parameters estimated in the study. In particular, for each incident category, the impact on road capacity is determined using the number of lanes that will be closed during the duration of the incident as a proportion of the number of lanes normally available. In a particular flow group, the traffic entering the incident area, combined with the information about the impact on capacity can be used to derive delay times.

Source: Mott MacDonald (2009).

De Jong *et al.* (2004) have reviewed literature on monetary values of time in the context of reliability. Most of the studies reviewed used stated preference surveying of road network users. Analysis of revealed preference data is scarce; probably reflecting the difficulties associated with collecting data. Very few research projects assess reliability aspects of public transport and freight transport.

According to De Jong *et al.* (2004), three different approaches have been applied to arrive at values for reliability. These are strongly related to the types of indicators discussed in the previous section. There must be naturally consistency between the valuation of impacts and the measurement of impacts. If reliability is measured as the standard deviation of travel time, unit values must have the same metric.

The approaches are: (a) the mean versus variance approach, (b) percentiles of the travel time distribution, and (c) scheduling models. The first two methods assume that either the variance or percentiles of time distribution reflect demand for reliability – they do not incorporate behavioural models. Scheduling models, in turn, intend to estimate the demand function for reliability without assuming that given indicators correctly reflect demand. They try to reflect the ability and costs of re-arranging activities in case of transport delays or disruptions.

The mean versus variance approach

Unreliability is measured as the standard deviation (or variance) of a travel time distribution. This approach can be defined as an analytical or mathematical method. Data for the valuation of the standard deviation can be obtained through a stated preference survey by including a representation of the variance and the mean travel time as attributes.

A utility function is specified that includes the mean journey duration as well as the variance (or the standard deviation) of the journey duration. Parameters for both variables are estimated, usually on stated preference data. In the stated preference interviews, respondents are not shown the variance of travel time as such, because this is recognised as too difficult a concept for a large number of respondents. Instead, each choice alternative contains as an attribute (in addition to average travel time and travel cost), a set of 5 to 15 possible journey durations (sometimes presented graphically). It is possible to calculate the variance that is consistent with each set (or generate a set of journey durations that matches a target variance). Average journey time and the variation in travel time presented in the stated preference survey can be constructed such that between observations they are not or only lightly correlated. Because both attributes are presented to the respondents in the stated preference survey and vary more or less independently, no double-counting will occur when in a cost-benefit analysis one would include travel time and reliability gains, with values for both coming from the stated preference survey.

From the estimated model, the ratio of the coefficient for the standard deviation to the coefficient for the mean travel time can be calculated. This gives the disutility of a minute standard deviation of travel time in terms of minutes of mean travel time. A monetary value for unreliability can be derived by combining this with a value of travel time (or directly if travel cost is also in the utility function). For the application of these outcomes in practical CBA's of transport projects it is necessary that not only the change that the project causes in expected (mean) travel times is predicted, but also the change in the standard deviation of travel time.

Percentiles approach

With this approach, unreliability is measured as the difference between the 80th or 90th percentile of the travel time distribution and the median (*i.e.* 50th percentile). Again the valuation can be derived from stated preference experiments among travellers. This method is closely related to the mean versus variance approach. Unreliability is measured and valued as the 90th percentile of the travel time distribution minus the median (or the 80th percentile minus the median). The left-hand side of the travel time density (shorter than average travel times) is not used as this is regarded as being of little value to the travellers (assuming they are risk averse). The 80th or 90th percentile indicates a considerable delay, but the most extreme journey durations are not considered but rather seen as outliers. To obtain a value for unreliability measured like this, models need to be estimated on stated preference, revealed preference or combined stated/revealed preference data, in which travel time and the measure of unreliability are separate variables. Again, use of both values in a CBA will not imply double-counting.

Scheduling models approach

With this method, unreliability is measured as the number of minutes that one will depart or arrive earlier or later than preferred (schedule delay). This approach may have the most meaning to the users, undertaking trips. The measure can also be offered as an attribute in a stated preference experiment, together with other attributes such as journey duration and travel cost.

These models are based on work by Vickrey (1969) and Small (1982). The monetary values obtained for being early or late are very difficult to implement in a CBA framework, because the link to travel time period choice is not made in the CBA (there is no reference to clock time, only to journey durations), and the preferred arrival times are unknown.

Summary of different approaches

The standard deviation approach can with relatively little difficulty be applied in the cost-benefit assessment but lacks firm theoretical motivation. The scheduling approach is more fundamental but has been considered, until recently, difficult to apply in practice since it requires knowledge of the preferred arrival times of individual travellers. Recent findings by Fosgerau and Karlström (2007) and Fosgerau *et al.* (2008) have provided valuable contribution to this discussion by showing that the standard deviation, the percentile and the buffer time approaches can all be derived from scheduling preferences. They are found to be equal as long as the shape of the travel time distribution is unchanged. This is an important contribution in reducing the knowledge gap, especially through better understanding of the economic and social impacts of disruptions in activities.

Unpredicted unreliability value is linked to activities and schedules disruption, which may be independent of regular travel times. In this respect, scheduling models approaches are likely to provide more relevant values than mean – variance approaches.

The scheduling approach aims at discriminating between buffer time (which value is, in theory, close to travel time), and "pure" un-reliable time (*i.e.* non anticipated delays or disruptions), which value is likely to be significantly higher, and strongly dependent of the type of users. The mean versus variance approach aggregate both types of time (anticipated and un-anticipated). In a recent study, by Yin-Yen Tseng (2008), two different values for unreliability are identified, both applicable to cost-benefit assessment: scheduling delay reduction (arriving less early or late) and pure reliability (standard deviation).

The scheduling approach may be helpful to better assess the "granularity" of unreliable time values, and, hence, demand for reliability. For example, this approach illustrates how low-income users may have a relatively higher value of unexpected delay, reflecting their lower capacities to manage an unexpected disruption of their professional and individual activities. It might also be useful to better quantify the impact of ex ante information on buffer time (unit value and volume) and impact of on-trip information on unreliable time (unit value and volume).

Stated preference surveying techniques are often put forward as ways of identifying benefit levels of schemes for which proxy values are difficult to ascertain. In behaviour or stated preferences surveys, scheduling approaches may be useful to separate those two different nature of times, and to identify links between the value of the unreliable time, journey purposes (*i.e.* activities) and amount of unreliable time (short delays versus long disruptions).

Finally, the following tables summarise some quantitative results on values of reliability and methods used in different studies. Results reinforce the fact that values of reliability are country-, location, user- and time-specific.

Table 3.1. **Quantitative results on value of reliability in passenger transport (travel time or 2003 Euros)**

Study	Results	Method
Accent and HCG (1996)	Doubling the change of delay equals 13-20 min travel time, halving the change of delay equals 3-5 min travel time.	SP, road transport, UK
AVV (2003)	Reliability is the most important aspect of service.	SP, 3387 bus, tram and metro users in the NL
Brownstone and Small (2002)	Value for 80^{th} minus 50^{th} percentile 11-14 Euros/hr (males) and 28-30 Euros/hr (females)	RP, travel time measurement on State Route 91 with variable tolls
Brownstone and Small (2002)	Value for 80^{th} minus 50^{th} percentile 26 /h	RP (above) and SP
Copley *et al.* (2002)	Value of standard deviation of travel time equals 1.3 times value of travel time.	SP, 167 car drivers in Manchester, mean versus variance method
Copley *et al.* (2002)	One minute late or early valued less than one minute travel time.	SP, scheduling model
De Jong *et al.* (2003)	One minute late or early equals 1-1.5 times one minute travel time (commuting, business, leisure travel).	SP, 1000 car drivers and train users in the peak periods in the NL, scheduling model
MVA (2000)	Value of standard deviation of time equals 24% (when seated) and 48% (when waiting) of the value of travel time.	SP, 309 bus users in France, mean versus variance method
Rietveld *et al.* (2001)	A decrease in the probability of a 15 min delay from 50% to 0% equals 2.35 . One minute delay is 2.4 times as bad as one minute travel time.	SP, 781 public transport users in the NL
Senna (1991)	Disutility of the standard deviation is around 2.5 times the average travel time.	SP, 301 respodents in Porto Alegre, Brasil, mean versus variance approach

SP = stated preference survey, RP=revealed preference survey. *Source*: Adapted from De Jong *et al.* (2004).

3.3.2 Country experiences

Current practices to incorporate reliability into cost-benefit analysis are next reviewed. These examples show that few countries have already taken steps to incorporate reliability and, more importantly, that it is possible to take into account reliability in the policy or cost-benefit assessment. The case studies also illustrate that if the value of the time savings accounts for a very large share of the monetised transport project benefits, then even relatively small changes in temporal levels of travel times may have large consequences on appraisals (that is, whether a project should be embarked upon, and the priority of project).

Table 3.2. **Quantitative results on value of reliability in freight transport (travel time or 2003 Euros)**

Study	Results	Method
Accent and HCG (1995)	1% increase in the probability of delay of 30 min equals 0.45-1.8 Euros per transport.	SP, road transport in the UK
Bruzelius (2001)	1% increase in the frequency of delays equals 4.7-7.0 Euros per wagon (rail), 3.5-32.6 Euros per transport (for road).	SP, shippers in Sweden 1989/1990
Fowkes *et al.* (2001)	The value of difference between the earliest arrival time and the departure time is on average 1.18 Euros per minute per transport. For the deviation from the departure time the value is 1.12 Euros.	SP, 40 shippers and carriers in the UK 1999
HCG (1992a)	10% increase in the percentage on time equals 5-8% additional transport costs.	SP, 119 shippers and carriers in 1991/1992 in the NL
HCG (1992b)	A decrease in the probability of delay from 15% to 5% is worth 0.5-2 cents per ton-km.	SP, 150 interviews in France, Germany and the NL in 1992
RAND Europe *et al.* (2004)	A change of 10% in percentage not on time equals 1.77 Euros per transport (road).	SP/RP, 194 shippers and carriers in road transport
Small *et al.* (1999)	One hour deviation from the scheduled delivery is worth 393 Euros per transport.	SP, hauliers in the USA, scheduling model

Source: Adapted from De Jong *et al.* (2004).

The Netherlands

Previously, in the Dutch cost-benefit assessment practice, most of the consultants involved in the cost-benefit assessment work arbitrarily added an extra benefit of 25% of the direct travel time benefits due to reduced congestion. This "reduced congestion" implies something about "reliability" benefits but it remains a tenuous linkage between benefits of reduced "congestion" (that is, presumably, reduced travel time) and the reliability benefits (that is, reduced travel time variability).

A lot of work has been carried out in the Netherlands improving estimates for the future development of the standard deviation of travel time after a policy initiative (see Box 3.1 above for a discussion on the Dutch National Model System). Because evidence based values for reliability are not available, the Transport Research Centre of the Dutch Ministry of Transport (AVV) organised a meeting with national and international experts to establish a range of provisional values to use in cost-benefit analysis until specific evidence based values are available (Rand Europe 2005). Based on the meeting, the Dutch ministry has adopted reliability ratios (defined as value of one minute of standard deviation/value of one minute of average travel time). These reliability ratios (see more below chapter 3.3.3) differ for passenger transport by car and public transport (0.8 and 1.4 respectively).

This approach is only considered provisional as it can be highly misleading; research is therefore being undertaken to establish empirical-based values for reliability in the Netherlands. The on-going study acknowledges granularity by investigating values for different user groups, modes and trip purposes.

United Kingdom

In the United Kingdom, the Transport Analysis Guidance (TAG) Unit of the Department for Transport provides detailed guidance on the appraisal of transport projects. The guidance should be seen as a requirement for all projects that require governmental approval. The most recent guidance document recommends ways to assess improving journey time reliability for private vehicle travel and public transport (DfT 2009).

For private vehicle travel, the report recommends that reliability should be measured by the standard deviation of travel time or by the coefficient of variation. In order to forecast the change in the level of journey time variability, different methods have been developed for inter urban motorway and dual carriageway roads and urban roads. For motorways and dual carriageway roads incidents are the main source of unpredictable variability and the Unit recommends the use of Incident Assessment Approach using INCA software programme (see also above Box 3.2). INCA reflects how delays caused by incidents vary according to the severity and length of the incident, the number of lanes blocked and the volume of traffic at the time. For urban roads, day to day variability is considered to be the main source of unreliability and the UK Department for Transport has collected data to forecast the standard deviation of journey time from journey time and distance for each origin and destination flow.

To estimate monetised benefit of changes in the variability of journey time, the Unit advocates the use of reliability ratios (defined as the value of variability in journey times in comparison to the value of journey time), derived from the Dutch study mentioned above.

For public transport (rail), delay is measured as a difference between scheduled arrival time and actual arrival time. Disutility of arriving late is then measured using mean delay and valued applying appropriate lateness factor. Estimates of these lateness factors (value of consequences arriving late in relation to in-vehicle time) vary between one and five. The second disutility is related to the unpredictable variation in delay. According to the TAG report, this should be measured by the standard deviation of journey time and valued using the reliability ratio obtained from the Dutch study. However, because there rarely is data on delay distribution, the Unit recommends using lateness factor of three (including an uplift of 20% representing the additional disutility incurred as a result of variability in delay) in a general case. This means that one minute of average lateness is valued by passengers equal to three minutes of scheduled journey time. Using the value of time as basis, allows then the use of monetary value on reliability.

The report acknowledges that there is a limited amount of evidence on the values to be applied to the standard deviation of travel time. The Unit report also notes that reliability is a rapidly developing area where new research will likely bring new insights. The recommendations are thus made with using the existing knowledge base.

Denmark

Although reliability is not yet included in the Danish economic appraisal practice, several authorities have used different measures to evaluate travel time variability (the following is based on Fosgerau *et al*. 2008). The Danish Road Directorate has used delay as a proxy for travel time variability. The Orestad Transport Model (OTM) has been previously used for providing cost and benefit estimates for infrastructure scheme appraisals in the Greater Copenhagen area. OTM evaluates changes in behaviour and the consequences for the total travel time for proposed schemes. Travel time for different road segments are generated on the basis of speed-flow relationships for a number of matrices at different times. The ways in which demand and route choice are affected by congestion are outside the greater Copenhagen area are normally not included. The valuation of the delay is based on the official unit prices. One minute delay is evaluated as 1.5 minute of travel time, based on the "UK value of time" study (Accent and HCG 1996).

In a similar way, the national rail authority uses delay as a proxy for travel time variability. Variability is measured as the total number of passenger delay-minutes relative to timetable. When forecasting variability, two types of sources of delay are considered separately; incident related delays and other delays (timetable and physical capacity related). A unit price of two times the value of travel time is applied to assess the value of reliability.

The most recent Danish study on travel time variability, commissioned by the Danish Ministry of Transport and its road and rail agencies, sets recommendations for incorporating reliability into CBA both on short-term and long-term basis based on new theoretical results (see Fosgerau *et al.* 2008).

The study proposes the use of an approach based on optimal scheduling considerations. In this approach, travel time variability is defined as standard deviation of travel time. In order to arrive at value of reliability, a concept of variability ratio is introduced, defined as value of travel time variability in relation to value of travel time. Based on new theoretical results, the variability ratio can be obtained through a function of the ratio of lateness cost to the value of time, the optimal share of trips arriving late, and the average standardises lateness from the travel time distribution.

Based on literature review, the report recommends to use a value of lateness around three times the value of travel time while average traveller is late optimally on one out of every three trips (the optimal share of trips arriving late) according to findings in the report. The average standardised lateness that is determined by the shape of standardised travel time distribution (assumed fixed for the road and rail sections) was estimated at 0.33 for road and 0.28 for rail. Hence, the study concludes that, until more empirically established values are available, a variability ratio (reliability ratio) of 1.0 (3 x 0.33) should be used for road and 0.84 (3 X 0.28) for rail.

Box 3.4. **Rail link Copenhagen – Ringsted in Denmark**

The railway between Copenhagen and Ringsted is a central part of the Danish rail network. The capacity of the network is currently fully used, which not only affects the number of trains running but also train reliability (punctuality). In order to improve the situation, the Danish government investigated a range of options. The cost benefit assessment of these options included reliability (regularity).

To calculate reliability benefits of the different project options, the total delay minutes for each network link, time period and direction was calculated; these estimates were then divided by the number of trains in order to arrive at average delay per train (or passenger). Improved punctuality was calculated as a fixed coefficient of time savings. Similarly, in case of goods transport, total delay minutes per ton-kilometer was estimated by multiplying the number of total delay minutes per train by average weight of goods per train. According to the calculation, the building option was estimated to reduce delays by 25% compared with the base option while other options showed fewer improvements.

The value of time savings and savings in delay times were based on official values of time. Some recognition of user granularity was built in to the estimates. In particular, values of time for passengers differ by trip purpose (work, commerce, other) and by average travel time, delay time, and frequency. The value of delay time was estimated at double the value of average travel time for all trip purposes. In case of goods transport, granularity was not considered – one value of travel time was used for all goods transported, that is, even though the value of time/reliability depends on the type of good transported.

Applying the personal values of delay (as double the value of time), and the reliability-insensitive goods value of delay and time/time, to cost-benefit analysis, added around 18% of user benefits for passenger transport and 7% for goods transport in the new building option.

Source: Cowi (2005).

Box 3.3. North-South road connections in Stockholm, Sweden

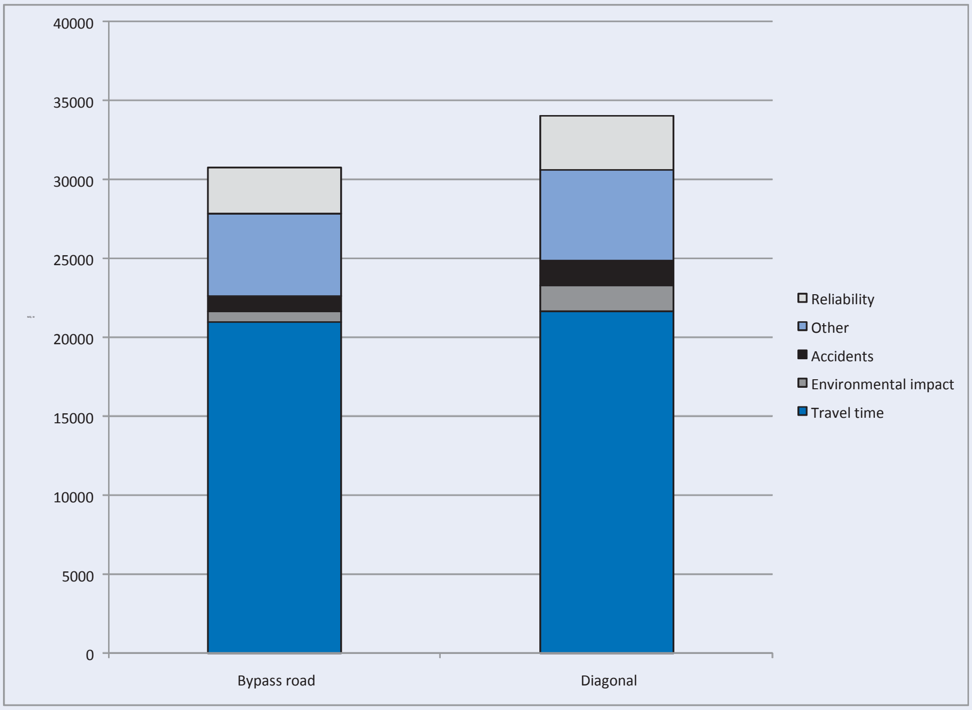
Cost-benefit analysis of the North-South road connections in Stockholm, Sweden, includes an estimate of reliability benefits. Two alternative road investments were considered – so called Stockholm Bypass Road and Diagonal Ulvsunda.

Data on travel time was obtained from the city of Stockholm's traffic cameras that identified each vehicle and its travel time for certain links. Travel times were observed for 46 links both directions during weekdays from September to December 2005. Standard deviation of travel time was then calculated for each link at certain time of the day.

The Stockholm study assumed that the value of unreliability is based on reliability ratio describing how many minutes of travel time a one minute reduction in the standard deviation corresponds to. Granularity was ignored: a ratio of 0.9 times the travel time value was used for all users. Based on observed traffic volumes, a relationship between travel time, congestion and travel time variation was estimated. The value of travel time variation was estimated to be 35kr per hour of standard deviation. Because granularity (diverse user groups) was not taken into account, the value of travel time was implicitly assumed to be same for all users.

The following figure summarises results of the cost benefit assessment for a Stockholm bypass road and for an alternative project, the Ulvsunda Diagonal. The inclusion of reliability into the calculation added around 12-13% to the user benefits.

Figure. **Benefits of the Stockholm bypass and Ulvsunda Diagonal options**



Source: Transek 2006:18

Sweden

The Working Group for Cost Benefit Calculations (so called ASEK-group) has recommended that delay and unreliability should be included into the cost-benefit analysis (SIKA 2008). The report recommends that travel time unreliability for business trips by car and trips by car from home to work should be valued as 0.9 times the average value of time. This is based on the reliability ratios proposed in the Netherlands as well as on Swedish studies on travel time variability (see for example Transek, 2003).

For other travel types, a top-down approach is recommended, in which the aggregated value of delay and congestion time (incorporating both variation in travel time and other factors, such as un-comfort of travel) is defined as 1.5 times the value of average travel time for cars and 2.5 times the average value of travel time for other modes. These values are taken from international studies. In practical applications, however, calculations are done using the standard deviation approach for all travel purposes. In case of goods transport, similar top-down approach has been adopted. The value of delay on goods transport has been considered two times the value of time by product group. This is considered only a preliminary estimate until more research is carried out.

New Zealand

The Economic Evaluation Manual for the New Zealand Transport Agency notes that improved trip reliability should be incorporated in economic efficiency evaluation of land transport projects. The concept used applies both work and non-work trips. However, the procedure for calculating trip reliability is considered suitable only for day-to-day variation in travel time, not for incident related delays. The New Zealand definition of travel time variability is as follows: "trip time reliability is measured by the unpredictable variations in journey times, which are experienced for a journey undertaken at broadly the same time every day. The impact is related to the day-to-day variations in traffic congestion, typically as a result of day-to-day variations in flow. This is distinct from the variations in individual journey times, which occur within a particular period." (see New Zealand Transport Agency 2008, Section A4.5, p. A4-13).

Travel time variability is measured by the standard deviation of travel time reliability. Reduced variability arises from a reduction in congestion on links and at intersections along a route and the manual provides detailed data to calculate and value the standard deviation of travel time in different contexts.

The benefits from improving trip reliability are then calculated as a function of reduction in the network variability, traffic volume, and the value of travel time reliability. The value is arrived at by multiplying respective value of travel time by a factor of 0.9 for typical urban traffic mix (for projects with significantly different vehicle mix, 0.8 for cars and 1.2 for commercial vehicles is used). Finally, a correction factor is applied to adjust calculations for percentage of variance outside of study area.

The New Zealand Transport Agency has recently commissioned a study to find a method of measuring the value placed on public transport reliability. As a part of that study, a stated preference survey was implemented to collect information on passenger's valuation of reliability. The survey focused on two components of unreliable services – departure and in-vehicle variability and produced estimates for these parameters in relation to normal in-vehicle time (Vincent, 2008).

Australia

Similarly to New Zealand, the Australian National Guidelines for Transport System Management (Australian Transport Council 2006) incorporates reliability into cost-benefit analysis. It suggests using the standard deviation of travel time as basis for measuring unreliability for road traffic. The guidelines refer to the New Zealand method for calculation of the standard deviation based on the volume/capacity ratio for the road link, intersection or route segment. For public transport, the frequency of running behind schedule, from which an 'average unexpected waiting time' can be estimated, is recommended as the measure of travel time variability.

There are no monetary values specified for reductions in variability of travel time but research findings are quoted (*e.g.* Bates *et al.* 2001) and it is suggested that a reliability ratio of 1.3 (value of reliability in relation to the value of time) be used for road traffic. This value is for commercial vehicles, because the Australian guidelines are intended for project evaluations for freight transport. If private car travel is included, then a weighting of 0.8 is included for that component of the travel demand. For public transport project evaluation, a weight of 3.0 relative to the value of time is recommended.

Current research is challenging the approach by raising questions on current applications of travel time reliability. On-going research is looking at the distribution of travel time variations and correlation of travel times on route sections (see for example Taylor 2008).

Norway

Norwegian Institute of Transport Economics is currently carrying out several valuation projects, aiming at the development of new unit values for several non-market goods in transport. Valuation of time and reliability in passenger transport is covered in a comprehensive study commissioned by Norwegian transport authorities. Valuation of time and reliability in freight transport is the purpose of a project funded by the Research Council of Norway.

In both studies, valuation of reliability (or travel time variability) is approached in two types of stated choice experiments which are based on the mean-variance approach and scheduling approcah, respectively. The design of the experiments is partly inspired by the design for the Dutch study described by de Jong *et al.* (2007). By applying two alternative approaches, a comparison of the approaches is enabled, and different types of values for reliability can be provided from these studies.

Both studies are to be finalised and reported by 2010. Further work is needed with respect to incorporation of reliability into cost-benefit analysis.

Canada

Transport Canada has recently completed a study on value of time and reliability for local trips in Canada. Although the report focuses mainly on values of travel time, it concludes that the small body of research reviewed on values of reliability suggest that the value of reliability can be related to the value of travel time. Therefore, the report proposes to calculate value of reliability as a linear function of the value of travel time (InterVistas 2008).

Based on the findings from a literature review, the report recommends the use of the difference in percentile approach (difference between the 50th and 90th percentile travel times or alternatively buffer index) to arrive at measure of travel time variability. However, because the value of reliability in relation to value of travel time (in studies reviewed) ranged from 60% to 150%, the report recommends, in the absence of better information, that value of time should be used as a proxy of value of reliability.

3.3.3 Reliability ratio

As the previous examples illustrate, most available country experiences on valuing reliability refer to the use of the so-called reliability ratio. This ratio is defined as the ratio of the value of one minute of standard deviation (*i.e.* value of reliability) to the value of one minute of average travel time. These ratios have been considered rather similar in all studies and they are mainly derived from international case studies, and more specifically from the workshop of international experts convened by AVV, the transport research centre of the Dutch Ministry of Transport. At this meeting, some consensus regarding reasonable reliability ratios for passenger transport was reached – 0.8 for cars and 1.4 for public transpot (Hamer *et al*., 2005, Kouwenhoven *et al*., 2005).

However, the value of reliability is inevitably very granular with a wide spectrum of values, such as across users, journey purposes and times-of-day. Unless empirical values are established, using some standard reliability ratio implies that the value of reliability is related to the value of average travel time for all user groups or trip purposes. As was noted earlier in this study, current classifications for value of travel time may not necessary reflect characteristics of reliability (see for example Small *et al.,* 2002 where it was found that the value of reliability was twice higher for women than for men).

As the tables 3.1 and 3.2 as well as the following table illustrate, reliability ratios are very different and very much location-, time-, and user- specific. Using single values (or ratios) might be highly misleading.

Table 3.3. **Reliability ratios in different studies**

Source	Ratio	Notes
Abdel-Aty *et al.* (1995)	0.35	
Black and Towriss (1993a)	1.27	
Lam and Small (2001)	1.30	
Small *et al.* (2001)	1.30	
Bates *et al*. (2001)	1.10 – 2.20	
Eliasson (2004)	0.95	Private trips in Stockholm in the morning
Eliason (2004)	0.59	Private trips in Stockholm in the afternoon
Brownstone and Small (2002)	0.75	
Hamer *et al*. (2005), Kouwenhoven *et al*. (2005)	0.80	

Source: Adapted from Transek (2006).

3.4 Conclusions

Reliability of travel times is increasingly recognised as a major network characteristic that should be recognised when considering investment options.

The treatment of reliability when assessing publicly-funded transport infrastructure projects is, however, a challenge. Due to the absence of a direct pricing mechanism for reliability, cost-benefit analysis represents the best option for delivering optimal levels of reliability.

Despite its obvious importance, the number of countries that incorporate reliability in cost-benefit analysis is limited. Only a few countries include monetised values of reliability in their appraisal.

Reliability is generally not taken into account when evaluating projects and, as a consequence, CBA will not distinguish between two infrastructure projects where benefits (such as the expected time savings) are identical but where the variability of expected travel time differs. It is obvious that if an infrastructure investment, for example, is aimed at improving travel time reliability rather than average travel time, lack of systematic analysis of the economic benefits is a major drawback.

This section outlined a range of examples of cost-benefit analysis and appraisal guidelines that incorporate reliability benefits. There are ways to measure and value reliability that can be integrated into cost-benefit analysis. These have been used on a pilot basis in a small number of countries. These approaches provide a foundation for explicitly incorporating reliability benefits into investment appraisals and, consequently, policy frameworks.

However, further work is required. Incorporating reliability into cost-benefit analysis requires data on existing travel time reliability, anticipated reliability level after a policy initiative, and monetary values of reliability disaggregated at the appropriate level of granularity. Furthermore, the work on valuation has been mainly focused on passenger transport and the knowledge gained so far cannot be automatically transferred to freight transport, especially for multimodal transport networks and trade corridors.

Even though a number of studies examine the value of reliability, knowledge on how travel time, congestion and quality of network affect reliability is limited. Without exception, almost all the existing traffic and transport models used for strategic impact assessment provide predictions of average demand flows and average (expected) travel times. These models can also be used to provide predictions of change in the standard deviation of travel time; such changes are the building block to identifying the impact on reliability of an investment. It is therefore necessary to improve the traffic forecasting tools currently available.

What is proposed for incorporating reliability into project assessment is that temporal journey time improvement should be split into pure journey time improvement and reliability improvement, for each user group, location, etc. Revisions to the cost-benefit analysis approach would involve temporal adjustments, user granulation, and monetisation adjustments.

It is emphasised that the estimates obtained will not be transferable, but are country, location, user and time-specific. Therefore, using a single reliability ratio (often obtained from other studies) can be highly misleading. Until empirically-based values for reliability are available, using the standard value of time as a proxy for the value of reliability may be as good as any other approach.

NOTE

1. The HEATCO (Harmonised European Approaches for Transport Costing and Project Assessment) research project has made an inventory of the way in which congestion and reliability are included in infrastructure appraisal processes in European countries.

KEY MESSAGES

- The reliability of travel times is increasingly recognised as a major network characteristic that should be recognised when considering investment options.
- Cost-benefit analysis represents the best option for delivering optimal levels of reliability.
- The number of countries that incorporate reliability in cost-benefit analysis is limited. Only a few countries include monetised values of reliability in their appraisal.
- Standard appraisals fail to unbundle benefits from improved reliability (reductions in travel time variability) from the benefits due to the reductions in average travel time. This omission removes the factual basis for arguing that a project really does improve reliability.
- Incorporating reliability into cost-benefit analysis requires data on existing travel time reliability, anticipated reliability level after a policy initiative, and monetary values of reliability disaggregated at the appropriate level of granularity.
- The temporal journey time improvement should be split into pure journey time improvement and reliability improvement for each user group, location, etc.
- Revisions to the cost-benefit analysis approach would involve temporal adjustments, user granulation, and monetisation adjustments.
- Estimates of the value of reliability will not be transferable, but are country-, location-, user- and time-specific.
- Using a single reliability ratio to establish a value for reliability is unlikely to be adequate given the diversity of values the principal user groups place on reliability.

4. INFRASTRUCTURE SUPPLY – PROVISION AS A POLICY TOOL

This section reviews the principles of infrastructure provision to influence the level of network reliability. Options reviewed for improving reliability include supplementing capacity when reliability is undermined by heavy demand and setting cost-effective infrastructure standards. The capacity options include building new rail lines and adding an additional highway; examples of cost-effective capacity include technical standards for the materials to be used and hence affecting maintenance levels, renewal rates or robustness to natural disasters. The section then discusses the practicality of infrastructure provision as a policy tool to improve reliability.

The extent and quality of the transport network is central to the functioning of the economy and society. Building and maintaining heavily used links are, however, a major challenge to network providers.

Network capacity is a crucial element of performance. The capacity of the system is defined by supply side characteristics, such as road and rail spine capacity, road and rail link and node capacities and scheduling. These supply side characteristics are both spatial and temporal. The special characteristics are defined by physical network structures and temporal characteristics by the availability of the network, services on the network over time and the level of disturbance on capacity caused by the use of the system. Given these characteristics, the adaptation of supply to accommodate demand with minimal repercussions on the system performance can take many forms.

Unreliability of system performance of any kind has repercussions on travel time. As discussed in Section 1, it is the interrelationship between supply, demand and external factors that collectively generate unreliability. For example, a heavily congested capacity is more vulnerable to disturbances than capacity with less traffic. An extreme example of the performance failure is a full blockage of the capacity of the system, reducing travel speeds through the system to zero.

In case of a performance failure, it then becomes important how redundant the capacity is. Redundancy here is defined as the existence of more than one means to accomplish the given function for the network. A lack of redundancy may have catastrophic consequences. Redundancy of the capacity can be improved either through provision of additional (alternative) capacity or by increasing the flexibility of the existing capacity through different kinds of standards and hence reducing the vulnerability of the network.

In this section, the two primary supply-based strategies are considered. First, the capacity expansion is reviewed. Secondly, the issue of the quality of that capacity is reviewed – high physical standards are likely to make the system more robust to perturbations. Clearly, such options for capacity expansion and high quality need to be considered against formal cost-benefit analyses that incorporate the net benefits of reliability.

4.1 Supplying capacity

There is a wide array of options available to expand infrastructure capacity. The most common options include construction of new infrastructure (such as new highways or new rail lines) or expanding existing infrastructure (supplementing extra road lanes or rail tracks, adding new stations or increasing port capacity).

The options listed above should, however, be carefully considered by analysing their cost-effectiveness against other, often less costly, alternatives. That is, cost-benefit analyses are required. It is too often the case that additional supply is considered as an option of first resort whereas, in many cases, it should be considered only when other options have fallen out.

Even though this report focuses on reliability, many examples in this section are remedial actions directed at congestion. Most of the supply side measures have been related to congestion relief, not reliability. This is mainly due to the fact that reliability is relatively new issue while congestion has been on the transport policy agenda for many years.

It is still important to recognise that congestion does not necessarily mean unreliable. It should be remembered that congestion and reliability may have different policy implications, as discussed in Section 1. Still, while this report focuses on reliability and suitable policies directed at responding to reliability issues, it is recognised that remedial actions directed at congestion can improve reliability and similarly, actions that improve reliability can reduce congestion. Thus, many so-called reliability policies are actually policies to reduce congestion.

This is intuitively understandable. If the demand on any link increases faster than the available capacity, congestion (and unreliability) is likely to increase. A study that compared this relationship in the United States found that greatest increase in congestion was witnessed in areas where demand grew more than 30% faster than supply (Schrank and Lomax 2005, p. 82).

The case of M6 Toll Motorway in the United Kingdom is a classical example of a congestion policy that most likely has had a positive impact on reliability as well. The reduction in travel time arising from the Tollway indicates a better traffic flow. On the other hand, increased traffic arising from the new link may increase a risk of incidents which again may reduce reliability. As no data is available quantifying impacts from reliability point of view, it is, at the end, difficult to be conclusive on the reliability benefits of the M6 Motorway.

Box 4.1. M6 Toll Motorway in the United Kingdom

Britain's first new major toll road in modern times, the M6 Toll in the Midlands, was opened in 2003. The Tollway is a 27mile long motorway bypass, aimed at relieving congestion on the parallel M6 motorway. Previously, the busiest section of the public motorway was carrying up to 180 000 vehicles per day – more than double its design capacity. The toll is charged by time-of-day, the toll station used (as a proxy for distance travelled) and according to vehicle size. The new motorway is reported to have achieved its aim of relieving congestion, and travel time has halved during peak periods between Birmingham and Manchester.

Source: M6 Toll After Study (2005).

However, there seem to be few cases where capacity-enhancement investments have been made to specifically improve reliability although they lack any impact assessment from the reliability point of view.

Countries adopt different approaches to network planning. That said, countries usually have strategic plans that focus on network links that are vital to the national economy. Identification of freight network corridors is one such strategic focus, seeking to reduce the vulnerability of the freight supply chains to perturbations. Identifying vulnerable parts of the network is therefore an important task in understanding reliability strategies (see Section 5 on pro-active management of the network).

Recognising the twin problems of reliability and congestion, many countries have developed Strategic Plans which include time-maps of investment options. These proposals are not only focused on selected infrastructure investments but also include investments in traffic management technology and other management-related investments. It is essential, however, that such strategic plans be grounded in cost-benefit appraisals that incorporate reliability issues.

Box 4.2. **Corridors of the Future**

On September 10, 2007, the U.S. Department of Transportation announced six interstate routes that will be the first to participate in a new federal initiative to develop coherency in the management of multi-state corridors. The "Corridors of the Future" program is aimed at developing innovative national and regional approaches to reduce congestion and improve the efficiency of freight delivery. The concepts include building new roads and adding lanes to existing roads, building truck-only lanes and bypasses, and integrating real time traffic technology like lane management that can match available capacity on roads to changing traffic demands.

Particularly, it includes projects which are said to improve reliability through additional capacity, including:

Interstate 95 (I-95) – This project would reconstruct and expand a 1 054 mile stretch of I-95 from Florida to Washington, D.C. that will accommodate future demand, safety, and reliability.

Interstate 70 (I-70): Dedicated Truck Lanes - Missouri to Ohio – This project proposes dedicated and segregated truck lanes along I-70 from the Interstate 435 beltway on the eastern part of Kansas City, Missouri to the Ohio/West Virginia border near Bridgeport, Ohio/Wheeling, West Virginia.

Interstate 69 (I-69) - Texas to Michigan – This 2 680-mile international and interstate trade corridor extends from Mexico to Canada. From the Mexican border to Indianapolis, Indiana, the proposed corridor would be built on a new location for about 1 660 miles.

Although these projects are likely to have reliability benefits, no data is available on existing variability of travel time or the anticipated reliability levels after the improvements. It is, therefore, difficult to be conclusive on the reliability benefits of these investments.

Source: *www.corridors.dot.gov/*.

Strategic investment plans are often large-scale, multimodal, projects. In international trade, the carriage of goods by a combination of two or more modes of transport is common. The long distances and the varying geographical features make it necessary to move international cargo by more than one mode of transport. Such movement might involve any combination of rail, road, air, inland waterway and sea transport. Capacity expansions should therefore also take into account the interoperability of these modes and interfaces (see also Chapter 5.3 on managing interfaces).

Another example of a strategic approach is the Sydney Rail Clearway Lines programme. The review process identified the locations with greatest operational unreliability and these areas received funds for capacity improvements, including by removing operational interfaces between otherwise-independent services. Currently, delays on one part of the system can affect trains in a completely different section of the network, due to the heavily interwoven system in which infrastructure, carriages and services are shared between lines. The Clearways project aims to divide the network's fourteen metropolitan rail lines into five independent "clearways". This is intended to isolate incidents to one part of the network so other lines will still be reliable.

The plan has two primary benefits. First, it reduces the area of impact of an incident on the network. Secondly, by enhancing reliability, it reduces the (unused) capacity that must be set aside in order to maintain a given level of reliability. Sydney's clearways plan thus illustrates how increased unreliability can choke up the otherwise reliable capacity. Improving reliability frees underutilised capacity. In the presence of core bottlenecks, delivering a relatively reliable rail system requires that the entire network capacity has to be constrained at a relatively low level.

In the absence of cost-benefit analysis it is not possible to identify exactly the reliability benefits of the project. However, it is more likely that benefits will be higher when reliability is taken into account. As discussed in Section 3, a cost-benefit analysis incorporating reliability benefits and related costs would provide a valuable tool in assessing the priorities and values in such schemes (see also case studies in Chapter 3.3.2 of country experiences incorporating reliability into cost-benefit analysis).

A basic level of connection between social and economic players in an economy is essential for a functioning economy; reliability is one of the quality attributes of that network. The success of land transport routes of developing economies, for example, remain sensitive to numerous factors including cost, reliability and time (United Nations, 2008).

The example of the Ukraine's transport infrastructure below illustrates a situation where underinvestment has made the network highly unreliable, with serious repercussions for the country's export potential and the internal workings of the economy.

Box 4.3. **Underinvestment in infrastructure in Ukraine**

For past decades on network of roads of general use of Ukraine, as a result of insufficiency or absence of financing, repairs and maintenance works (asphalt concrete pavements and surface treatment) were not undertaken. There has been a reduction in the volume of road works of over 95%. As a result, in some areas less than one fifth of roads are in normal operating condition.

This has led to a situation of fundamental disintegration of the road formation structure and destructions such as break of edges, pot-holes, overall cracking, settlings, rutting and others (mostly on regional and district roads). The physical integrity of the roads has, in some situations, led users to use the road shoulders, because the carriageway itself is ruined. The network itself becomes extremely unreliable to even the most basic weather conditions such as modest falls of rain. That is, the network is not just slow to move along, but also lacks reliability.

Under conditions of inadequate funding the network is regressing in physical structure and, thus, performance at a time when national economic growth and growing global links require higher performance of speed and robustness.

Source: Gamelyak *et al.* (2008).

To summarise, in principle, supplying more capacity can improve reliability, particularly when unreliability arises from high traffic levels. This may also mean less susceptibility of network if multiple links are provided on a corridor. Providing additional capacity with parallel links can make it easier to undertake maintenance on the network. Carrying out extensive maintenance work on an already-busy link will lead to extended traffic delays.

However, aside from the typically-high capital cost, there are other aspects of this policy option that do not resolve underlying reliability issues. In particular, adding capacity does nothing to deliver differential reliability as required by granular values of reliability unless it is reserved for specific user groups (see Chapter 6, Box 6.4 on truck-only lanes). Further, capacity itself is not a source of non-recurrent congestion or unreliability. The various sources of non-recurring congestion can still interact in the presence of inadequate base capacity and lead to non-recurrent congestion.

4.2 Setting network standards and improving robustness of capacity

If network reliability is to be considered as an explicit transport policy option then the second aspect of infrastructure provision that needs to be considered is that appropriate levels of infrastructure quality need to be assessed. Investment in transport network that makes infrastructure more robust to weather and other incidents will enhance reliability. Network providers can set quality standards or increase the size or volume of capacity affecting its ability to accommodate external impacts, such as natural disasters or weather variations. Investments can also be directed towards the life-long durability of the facilities and equipment and therefore affecting the levels or frequency of maintenance and hence reducing the adverse impacts of work zones on travel time reliability.

It is possible to put the case this way: where policy is focused on reliability, the case for inexpensive but short-lived, maintenance-intensive infrastructure has to be considered as much as the case for long-lived, low-maintenance infrastructure. The presumption should not be that "gilt-edged" infrastructure is preferred; the case for the appropriate infrastructure standard needs to be considered, however. The policy challenge is to ensure that cost-benefit analysis (with reliability appropriately factored in) provides the criteria for deciding the quality, rather than the decision being made entirely on the basis of government budgetary constraints.

Network maintenance can be costly. It also generates disruptive delays to users, such as putting up work zones. The duration of their installation affects supply. Road users, suddenly confronted by an unexpected work zone, will normally encounter a queue and experience unexpected delay. It follows from this that strategies to reduce the extent and duration of work zones that are part of capacity improvements, will improve travel time reliability. The extent to which users of the network are exposed to the disturbance can be reduced by minimising the need for additional maintenance. The disruptions can also be reduced by managing impacts of the work zone in a way that minimises the negative impacts for users. This section describes infrastructure capacity provision-related policy options while infrastructure management issues are dealt with in Section 5. Mitigating user impacts with information is discussed in Section 7. Identifying an optimal solution would include assessing all policy measures.

Two primary characteristics of infrastructure quality that impact on network reliability are, first, the frequency and intensity of maintenance and, secondly, the robustness of the infrastructure to external perturbations.

4.2.1 Low maintenance – durable – infrastructure

The adaptation capability of supply to accommodate demand with minimal repercussions on the system performance can have a significant impact on network reliability. These supply-side measures include standards and performance metrics that reduce the need for maintenance or reduce maintenance frequencies for the network. These then include measures that are directly related to pavement or rolling stock, they can relate to censors or traffic remote censing and can relate to design for bridges and tunnels.

Capacity improvements that involve long-lived materials requiring little maintenance over long periods will sharply reduce non-recurring congestion over the facility's lifecycle (and hence improve reliability). Long-life pavements may be considered desirable especially in heavily-congested parts of network where (other things being equal) the aggregate reliability costs of maintenance works will be much higher than on less-congested links.

Installing higher-durability infrastructure is more costly. On-going research on long-life pavements indicates there are high initial construction costs. However, the research suggests that these costs may be outweighed by the benefits through lower maintenance costs and lower user (delay) costs (OECD 2005). Thus, decisions on infrastructure provision will incorporate higher capital costs and lower maintenance costs, but should also identify the reliability enhancement arising from less disruption when the maintenance is undertaken (see Table 4.1.).

Table 4.1. **Economic evaluation of long life pavements: one km of 3-lane motorway surfacing (present value $000s)**

Contributing factors	Traditional surface	Advanced (low maintenance) surface
Initial works costs	480	1440
Maintenance works	1080	280
User costs (delays)	1280	520
Traffic management costs	260	170
Residual value	-40	-90
Net Present Value	3060	2320
Difference in NPV of costs		-740

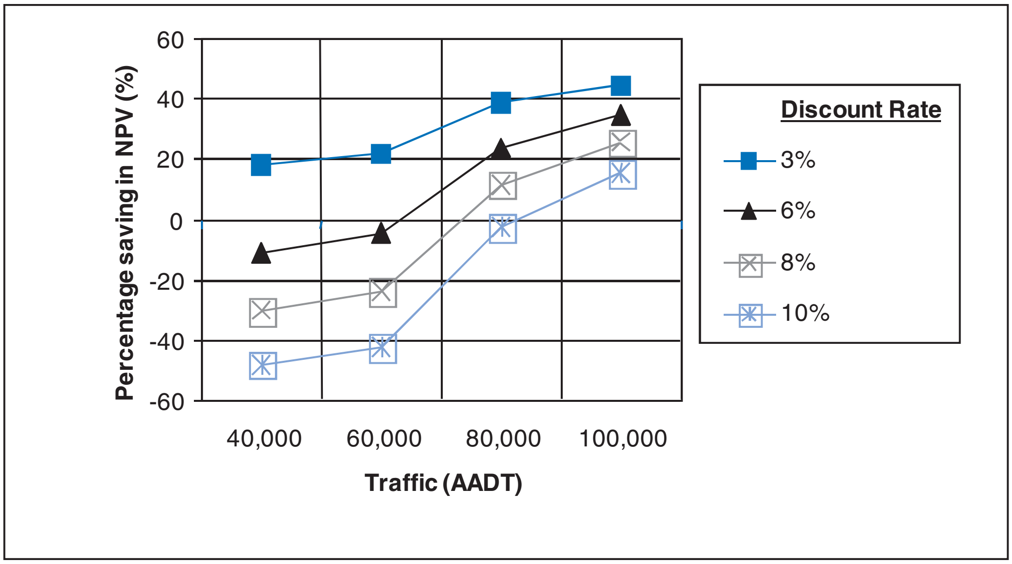
Indicative economic evaluation results. *Source*: OECD (2005).

The OECD study reveals that, in one evaluation exercise, once user-delay costs (a proxy for reliability costs) are considered, the optimal choice of road surface is changed. The study on long-life pavements has identified at least two surfacing material warranting further investigation: epoxy asphalt and high performance cementious materials. What can be achieved from the administration's perspective is better value for roads, more effective use of existing budgets, and more importantly, indicative costs of advanced surfacing materials. From the user perspective the most important benefit arises from improved service levels and greater network reliability because the need for road maintenance over an extended service life (above 30 years) is avoided. This all means better value roads with lower present value of life-cycle costs.

It is intuitive that the identified preference for more-durable road surfaces is greater when road links are highly utilised. As noted earlier, a network that is operating near its capacity is more vulnerable to disturbances, and the magnitude of the disturbances is greater, than in situations where traffic flows are small. Again, intuitively, highly congested network are more likely to benefit from improved standards of networks design. It is up to policy, however, to ensure that such costs and benefits are incorporated into appraisals.

Thus, a key message here is that incorporating travel time and unreliability in project appraisal impacts on the final solution, *i.e.* selection of the pavement material in this case. The OECD's preliminary work carried out on the economic evaluation of long life pavements in specific situations found that at high discount rates, "long-life" pavement was cost-effective on roads where annual average daily traffic was above 80 000 (see Figure 4.1). Obviously, such findings are specific to location and pavement materials.

Figure 4.1. **Economic evaluation of long life pavements: one km of pavement surfacing – potential benefits**



Source: OECD (2005).

4.2.2 Robustness of capacity

Network reliability can be influenced by infrastructure quality; infrastructure quality affects the frequency and intensity of maintenance, and the robustness of the infrastructure to external perturbation. The focus on infrastructure robustness is to set the quality of existing capacity to optimise its performance relative to perturbations.

Concerns about climate change heighten the importance of awareness of infrastructure robustness. There is a range of weather-related conditions that may affect transport infrastructure but greater weather variation and increased weather extremes require greater focus on network robustness. For instance, the likelihood of railway tracks buckling rises with rising maximum temperatures (TRB 2008) while increasingly low temperatures (with snow and ice) increase the incidence of points failures on railway turnouts. In such circumstances, stronger track sleepers and additional shoulder ballast can reduce the incidence of buckling; point heaters and greater human oversight can ensure that turnouts continue to function.

Extreme weather conditions have an impact on network manager and on users. Poor road conditions lead to reduced travel speed and delays. Severe weather conditions increase the risk of accidents and hence increase the possibility of disruption of traffic flow and unreliability. Hence, increasing number of extreme weather conditions will make it more difficult to deliver safe and predictable travel times.

The East Coast main line in the United Kingdom provides an example of the importance of improving the robustness of the existing network. In this case, the quality of electrical supply lines has both direct and indirect impacts on the reliability of train operations. A remedial plan has been set to improve infrastructure robustness.

Box 4.4. **East Coast main line in the UK**

The East Coast Main Line extends 393 miles from London to Edinburgh with a 30 mile branch from Doncaster to Leeds, and a project to electrify the route was approved by Government on 27th July, 1984. The investment was undertaken to a low standard compared with other British routes. In particular, economies were applied to the frequency of, and wire tension on, stanchions (electricity poles). The consequence of this standard is that the route is relatively vulnerable to the electrical wires being displaced by winds.

The quality and reliability of the overhead line electrical supply is a cause for concern. This has an impact on performance in two ways. First, as a precautionary measure, electric trains are slowed down in times of strong winds, introducing delays into all journeys on the route. Second, if the wires become detached there can be wholesale delay or cancellation of services until engineers can repair the wiring.

One way to reduce the line's vulnerability to damage from the winds might be to splice in additional catenary support poles (or to re-space the poles) at locations where the wiring is most vulnerable to wind damage. However, as there are extended lengths of the northern end of the line that are exposed to the north coast (and the strong north-eastern winds), this remedy could be extremely costly.

Note: This problem is well known within the industry. See, for instance, *East Coast Main Line www.absoluteastronomy.com/topics/East_Coast_Main_Line#encyclopedia*.

Infrastructure can be designed and constructed to standards that mitigate the effects on reliability of external factors such as weather. This can include embedding sensors in the infrastructure or along the right of way to detect critical conditions such as freezing conditions, accumulation of snow and ice. This assists in prevention of accidents and provides the means to monitor ambient conditions. Furthermore it is possible to deploy resources, such as snow and ice control equipment, in a more optimal manner. The provision of robust infrastructure hence also enables better management of that infrastructure. This issue is discussed in the next section in more detail.

Given the disruptive nature of weather on reliability, infrastructure should be planned, designed and maintained to in a way that recognises weather extremes, and the possibility of more frequent and greater weather extremes. Those extremes require construction focus on the infrastructure materials and focus on the maintenance regime. In more detail, the measures would include reviewing the case for less frost-susceptible materials, retrofitting side drains on the road structure, more extensive and frequent maintenance and using higher strength materials or insulating road sub-surface. Similarly, in rail tracks, measures might include strengthening earthworks and improving drainage to increase resilience into the renewals work.

4.3 Conclusions

In principle, supplying more capacity can improve reliability, particularly when unreliability arises from high traffic levels. Additional capacity also means less vulnerability of network in case of incidents if alternative links are provided. Basically, however, adding capacity does not deliver differential levels of reliability as required by users and their different valuation of reliability. Adding new capacity is likely to lead, at some point, to more traffic that will, in turn, increase unreliability. Unless the improvements address the root-causes of non-recurring congestion, for example, it is difficult to asset there will be an improvement in travel time reliability.

In any case, it is difficult to be conclusive on the reliability benefits of infrastructure investment projects because there is little – or no – *ex post* analysis done regarding the actual reliability benefits of investment projects.

As outlined elsewhere in this report, capacity expansion is not necessarily the most cost-effective approach to optimising reliability and there are clear limits to the build-only option. Many countries are facing limitations of the build-only option either because there is little additional space available for new infrastructure or because of the large investment requirements. There are more and more people on the network but less and less manoeuvrability because network is often constrained by available space which again has an impact on the cost of new capacity. Therefore, not all ambitious will be achieved with investments alone.

Careful choice of the robustness of capacity is a fundamental instrument for managing reliability. For instance, capacity provision that involves low-maintenance, long-lived materials, can reduce non-recurring congestion that is associated with infrastructure degradation and repair. Considering the merits of such a strategy is particularly important on heavily congested parts of network

The prospects of greater weather extremes have implications for infrastructure provision and maintenance. Infrastructure should be planned, designed and maintained in a way that ensures the appropriate level of resilience.

KEY MESSAGES

- In principle, supplying more capacity can improve reliability, particularly when unreliability arises from high traffic levels.
- Additional capacity will also mean less vulnerability of network if alternative links are provided.
- Undertaking maintenance on other parts of the network may be easier where such additional capacity is added. Nonetheless, if such reserve capacity is provided, it needs to be subject to cost-benefit analysis.
- It is difficult to be conclusive on the reliability benefits of infrastructure investment projects because there is little – or no – ex post analysis done regarding the actual reliability benefits of investment projects.
- However, adding capacity does nothing to deliver differential reliability as required by users and their granulated values of reliability unless it is reserved for specific user-groups.
- Many countries are facing limitations of the build-only option due to little additional space available, the cost of new capacity and long lead time needed for implementation of investment proposals.
- Careful choice of the robustness of capacity is a fundamental instrument for managing reliability. For instance, capacity provision that involves low-maintenance, long-lived materials can reduce non-recurring congestion that is associated with infrastructure degradation and repair. Considering the merits of such a strategy is particularly important on heavily congested parts of network.
- The prospects of greater weather extremes have implications for infrastructure provision and maintenance – and, hence, reliability. Infrastructure should be planned, designed and maintained in a way that ensures the appropriate level of resilience.

5. INFLUENCING SUPPLY – MANAGEMENT AS A POLICY TOOL

This section reviews the principles of managing the network to influence the level of its reliability. Management options reviewed can be divided into two categories: *pro-active and active*. Pro-active management of infrastructure mainly includes identification of network vulnerability to recurrent and non-recurrent unreliability. Dynamic processes, in turn, focus on active management of network to intensify oversight of network use once a network incident arises; such management systems include traffic control, accident clearing teams and rerouting strategies. Further, this section also discusses delays faced in different types of interfaces, such as those between ports and hinterland connections, borders of organisational interfaces.

The section first considers some pro-active management measures adopted by network providers. It then discusses dynamic processes through series of examples. Finally, conclusions are drawn about supply-side management as a reliability policy tool.

5.1 Pro-active identification of network vulnerability

Pro-active identification of network vulnerability consists of measures to identify those parts of the network that are more vulnerable to external disturbances. The probability and the consequences associated with the failure or ill-functioning of some network elements became an important topic of research in transportation planning in the 1990's. However, in spite of the increasing interest in developing methodologies to assess the robustness (or reliability) of transportation networks, only a few studies have focused on the integration of robustness objectives into network planning models.

The use of models to assist in the vulnerability assessment of any network may have significant reliability benefits. This type of modelling gives insights regarding the most vulnerable segments in a network. This information can then be used in future planning of network improvements.

The degradation of the quality of service provided by the network that may occur in case of fluctuations in travel demand or disruptions in infrastructure supply is typically not taken into account in models designed to represent those problems. Yet, this type of occurrences can have a severe impact both on the welfare of individual network users and on the performance of economic system as a whole. Most research efforts made so far with regard to the robustness of transportation networks have concentrated on the development of reliability indexes to be applied for the analysis of networks. Some authors have also made use of simulation tools to evaluate the vulnerabilities of networks. For instance, dynamic traffic assignment programs are used to assess network vulnerability and to show the importance of considering congestion spillback to evaluate the performance of a congested network when a link is blocked (Knoop *et al.* 2007).

The Netherlands has developed a strategic modelling tool to assist the vulnerability assessment and planning of the road network. The Dutch Network Robustness Scanner is used by the Dutch Road Directorate to assess vulnerability of network links. The Directorate uses the Scanner in its long-term traffic forecasts (see Box 5.1).

Box 5.1. Network Robustness Scanner in the Netherlands

The Network Robustness Scanner is used by the Dutch Road Directorate in long-term traffic forecasts in addition to its New Regional Model. The Scanner calculates traffic flows on the network for different user classes and for three different time periods (morning peak, evening peak, off-peak).

The Scanner is used to test the impact of perturbations on the network. Model runs are undertaken where the capacity of a link is reduced. By assigning traffic under these circumstances and comparing the effects with the 'normal' situation the effects of incidents can be evaluated. Ideally, all road segments should be taken into account. However, given the size of the model network (some 80 000 links) such an approach would take far too much time. Therefore the number of relevant links is limited by aggregation of closely related links to segments and by selecting the top 500 of a segment ranking list, based on the chance on accidents and the traffic volume. After the traffic is reassigned from restrained network links, new link volumes which may affect travel times are obtained.

The output of the model enables the identification and ranking of vulnerability network links segments. The model gives thus an insight on the most vulnerable segments in a network. Interpretation and insight into the results is facilitated by generation of maps.

Source: Kouwenhoven *et al.* (2006).

The kind of modelling approach presented above can facilitate infrastructure managers to identify reliability weaknesses on the network and, thus, where remedial actions might be required. The modelling and the visual map output may help in identifying suitable alternative routes in case of, for example, an incident. The network operator can use them to analyse which alternative routes are acceptable or what measures should be taken to improve these routes.

Another Dutch example of a system performance assessment tool regarding service reliability in public transport is the SIMONE model. SIMONE is a simulation model to check vulnerability of the railway network. The approach to this model of network vulnerability is similar to that for the road Scanner model.

The starting point for a SIMONE analysis is the given timetable and the available infrastructure. The network operator can then implement a perturbation. After simulation of a given timetable, the model generates easy accessible information on, for example, the delays by cause, the use of the buffer times included in the timetable, the number of realised connections and the way in which an initial delay was built up or reduced in network.

The key elements of planning rely on assigning adequate priorities to different links of the network. The main factors influencing these priorities are the level and frequency of congestion and incidents in addition to the implementation costs and traffic demand. Ultimately, such parameters can be incorporated into cost-benefit analyses of strategies to manage such vulnerabilities.

In the United States, non-recurring congestion (such as road accidents) accounts for roughly half of all congestion. There has been relatively little research undertaken into the effect of incident management on travel time reliability; what information is available is largely anecdotal.

However, recent research undertaken as part of the USA's second Strategic Highway Research Program (SHRP 2) has produced an experimental analysis and data collection plan for determining the extent to which various actions mitigate non-recurring congestion. Results of this work should yield equations that describe the incremental change in travel time reliability due to an action to improve reliability, given the characteristics of the road, the presence of other sources of delay known to affect travel time reliability, and other factors. This type of information is highly valuable in determining the best possible managing strategies.

Computing software programs can be used to estimate reliability levels and to provide real-time information to network users. A travel time forecast estimates the time a particular vehicle will take to travel between two places. This forecasting aims to provide a useful trip-planning tool and to provide real time information for long journeys. Travel time forecasting allows the operator/manager to bypass incidents and operational problems, while real time information allows monitoring the evolution of the incidents. In the context of Section 7 (where the use of mitigating actions is discussed), the modelling results can be also used to evaluate how users of the network can be informed about the incident and thus reducing unreliability.

Finally, identifying vulnerable links in the network may also provide valuable information on identifying and prioritising future infrastructure investment needs. The most vulnerable elements in the network can be determined by looking at the performance of the network under the demand predicted for future years and under different scenarios for regional development. The studies present a valuable insight in the specific road links that are crucial to the vulnerability of the network. Given this insight, the research highlights the importance of alternative (diversionary) routes, thereby pointing to useful supply strategies for alternative routes. In the Netherlands, for example, the Robustness Scanner has been used to compile a "vulnerability score" of several alternative schemes to provide new infrastructure. This score is defined as the total time lost in congestion at the moment the capacity on the potentially new link was disrupted. This analysis provides extra information on which construction alternatives could be compared. It should be noted, however, that a cost-benefit analysis incorporating unreliability would be the optimal tool for this purpose.

5.2 Active management of infrastructure

Active management of road and railway networks involve measures that intensify oversight of network use and involve management actions to influence traffic flow, including improving reliability. Active management systems enable transport service providers to react faster and effectively to service disruptions. Such systems reduce the vulnerability of the transport system, and increase the reliability in general, because incidents have less impact on usual travel times. The dynamic detection of all kinds of incidents and consecutive traffic jams is essential to provide an adequate response. It is to note that in this regard the term "incident" covers both unforeseeable accidents and those regularly occurring such as peak hour demand during holiday travel times.

An application of "active management" is the use of safety hard shoulders on motorways for conventional traffic movement. Improvements in traffic management technology have meant more countries are able to actively manage the hard shoulder. The use of the hard shoulder as a running lane at times of heavy traffic was introduced in the Netherlands and Germany in 1996 and, more recently, in the UK, the US and France. The management is applied in a range of ways but they generally involve monitoring of traffic flows and/or speeds so that as congestion rises above a level, road users are directed to use the hard shoulder as well as the main carriageway. This use of hard shoulder might be undertaken in conjunction with a reduction in speed limits.

At times, heavy traffic can cause congestion and a reduced traffic flow, creating a stop/start style of driving on the motorway. However, research has shown that use of variable speed limits and temporary supplementation of capacity with hard-shoulder running can significantly reduce flow breakdown. This improved flow of traffic reduces braking and subsequent acceleration by drivers and thus reduces speed variability.

In the United Kingdom the use of hard shoulders is applied with lowering speed limits; that is, enhanced reliability is delivered but with longer travel times. In the following Box 5.2, an application of this active management suggests that the reliability benefits outweigh increase in travel time. In other words, the management intervention has traded a longer travel time due to the reduced speed with improved reliability (reduced variability). Reliability benefits are gained without too much additional delay. However, because the hard shoulder is in use, improved reliability may come with a cost of increased risk of accidents that needs further studying. Hence, they may be a trade-off between increased reliability and safety.

Box 5.2. **Use of motorway hard shoulders in the United Kingdom**

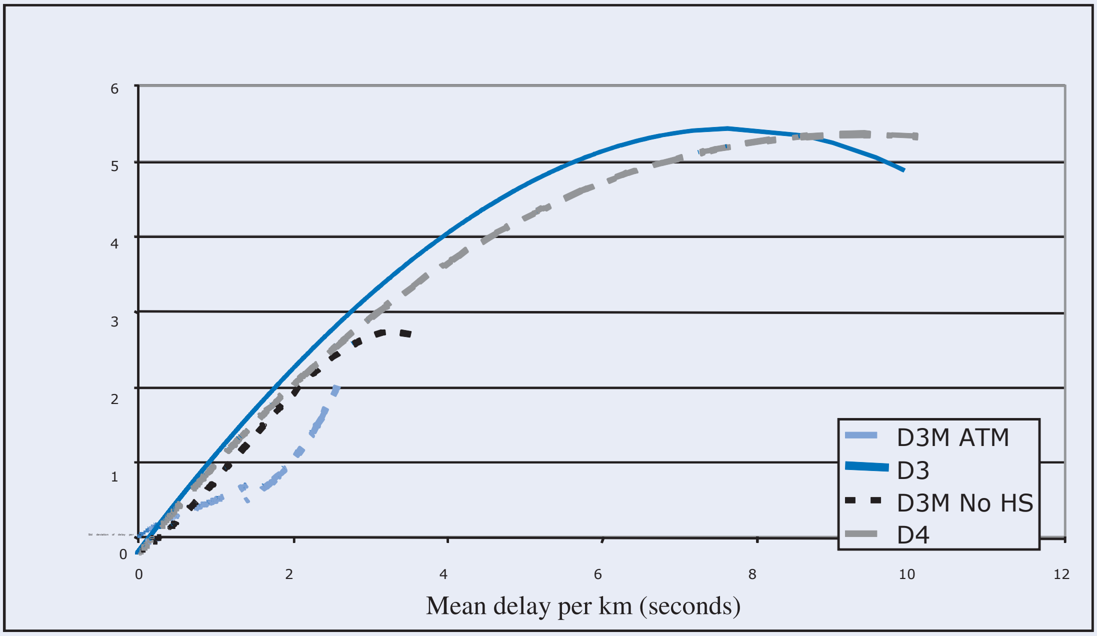
Since late 2005, the UK Highways Agency has allowed use of the hard shoulder of a 16 kilometre section of the M42 motorway at the most congested times. At such times, the speed limit on the section is reduced from 70 miles per hour to 50 miles per hour, with the option for a further reduction to 40mph depending on the operating conditions. The Agency is also investigating the merits of a (higher) 60mph speed limit when the hard shoulder is being used. To provide a safety offset for the loss of the hard shoulder, emergency refuges have been added, providing emergency telephones and lighting, and CCTV monitoring. A variety of monitoring equipment and information systems have been put into place on gantries spaced at 500m intervals.

As well as switching between regimes with and without hard shoulder running, the technology is also used to vary the speed limits on the three lanes.

The following figure presents delay and day-to-day reliability for four types of roads in the United Kingdom. The standard UK motorway, D3M, is the blue line. The vast majority of observations fall in the left hand part of the line, so this part of the curve is most confidently estimated. It indicates that as delay rises, the variability in delays rises too but the rate of increase declines. The D3M curve is similar to the grey dashed line for four lane motorways. The data indicates that the fourth lane does lessen variability for the same level of delay.

The black dashed line (D3M No HS) indicates the variability/average delay plot of the section of the network under study, when operating as a three lane motorway. It looks very similar to D3M. The blue dash (D3M ATM) indicates what happens to this relationship as active traffic management uses the hard-shoulder running regime. The delay penalty due to the speed restriction applied when hard shoulder running is permitted is compensated by reduced variability (improved reliability). However, because the hard shoulder is in use, improved reliability may come with a cost of increased risk of accidents.

Figure. **Delay and day-to-day reliability for different motorway regimes (M42, UK)**



Source: Mott MacDonald (2009).

Preliminary results from French experience (see Box 5.3.) suggests that the use of hard shoulders in active traffic management can increase capacity by at least ten percent but the consequent impact on reliability has not been researched. In France, hard shoulders have been used on a section of the motorway to the east of Paris; the hard shoulder is open to cars when traffic becomes saturated. It has been deduced that the increased capacity provided by the hard shoulders, leading to reduced congestion, has increased travel time reliability. Further investigation from the reliability perspective is still needed to confirm this.

Box 5.3. **Dynamic use of motorway hard shoulders in France**

There has been a hard shoulder running scheme in France since summer 2005 on the joint section of the motorways A4 and A86, which is renowned as the greatest traffic bottleneck in Europe. In the Val de Marne *department,* the joint section of motorway A4-A86 passes through the town of Joinville-le-Pont on a flyover more than 2200m long. Until summer 2005, 280 000 vehicles using this stretch of road every day faced a severe bottleneck, with over 10 hours' congestion a day and tailbacks regularly averaging 10 km.

To ease the recurrent congestion, drivers are given peak-time access to the hard shoulder. The size of the traffic lanes has been adjusted. From the standard width of 3.50 m, they were reduced to 3.2 m. In the event of an incident or accident occurring when the special lane is open, the system detects any stationary vehicles in that lane, which is then closed.

The hard shoulder lane is opened depending on dynamic changes in real-time speed and lane occupancy. Based on preliminary results, there is a significant impact on capacity, showing evidence of an increase in capacity of around 10% for traffic out of Paris, and 7.5% for traffic into Paris.

Source: Cohen and Zhang (2007).

Capacity can also be increased on railways through "optimised timetabling". A traditional rigid timetable specifies arrival and departure times for each train and station. When a train is delayed, the effects can ripple through the network. A dynamic rescheduling of rail networks can help train controllers to keep the network running smoothly despite an incident.

An example of where reliability policies can be applied to active and pro-active management in maintenance is with seasonal weather patterns. For instance, in some regions, snow and ice can have an important impact on reliability. Travel time predictability is often more difficult in winter. Traffic and transport are expected to function well in both summer and winter.

A Finnish study indicates that wintertime and summertime safety risk levels are at nearly the same level, on average. Nevertheless, wintertime risks are still higher on the busy main roads. Here, the uniformity and predictability of driving conditions becomes important. For this reason, the Finnish Road Administration has set winter maintenance principles and guidelines according to which winter maintenance is to be implemented. The administration sets quality standards for different road classes and for special events such as holiday seasons. Bearing this in mind, the administration sets standards for snow removal (see Table 5.1).

Table 5.1. **Quality standards for snow removal in Finland**

Winter maintenance class	Is	I	Ib and TIb	II	III	K1	K2
Maximum snow depth when snowing	4 cm	4 cm	4 cm (8 cm – night)	8 cm (10 cm – night)	10 cm	3 cm (8 cm – night)	
Cycle time, clean after snowing stops	2.5 h (slush 2 h)	3 h (slush 2.5 h)	3 h	4 h	6 h	3 h	4 h
If snowing stops after 22 at night	Plowed clean within cycle time		05	06	06	05	06

Source: Finnish Road Administration (2001).

Better management of incidents may also deliver significant reliability improvements. The use of modern incident management systems enable transport service providers to react faster and more efficient on disruptions of their services thus reducing the vulnerability of the transport system and increasing the reliability in general as incidents will have less impact on usual travel times (see Box 5.4). The dynamic detection of all kinds of incidents and consecutive traffic jams is essential to provide an adequate response. Establishing services to monitor conditions on the network helps responding quickly when incidents do occur. Improved cooperation with different actors, such as emergency services, can further improve response times and reduce any negative impacts.

Box 5.4. **Use of GPS equipment at accident sites to reduce post-incident disruptions**

GPS equipment is currently being used to survey and record the details of accident sites in the United Kingdom. The average time saved by the eleven police forces trialling the equipment is 40 minutes per incident. This time saved in getting the road re-opened reduces snowballing effects of the initial accident. At one incident, police was able to survey and collect enough data from an accident scene in ten minutes, where previously the road would have been closed for two hours.

Source: DfT (2008).

Investments in new technology to help incident management may reduce significantly negative impacts after serious collisions or incidents. The implementation of Traffic Control Units on the motorway network in Austria provides a further example of positive results obtained through active traffic management (see Box 5.5). While detailed analysis regarding the extent of impacts on reliability is missing, it is clear that reducing the number of incidents will directly have an impact on network reliability.

It should not be presumed that the infrastructure manager will be sufficiently focused on ensuring that reliability is optimised. Incentive systems can be put in place to facilitate such focus. Shadow tolls are a special context where pricing could be used to improve reliability through a contract between government (as network owner) and service provider (whether public or private). In traditional shadow toll schemes, the concessionaire charges a toll (to the Government) that accounts for the number of vehicles which have used the infrastructure. In this situation a Bonus/Malus (reward/penalise) incentive could be applied to this amount accounting for the provided reliability. This is already being applied to public transport concessions (contracts, franchises), where there are incentives for punctuality. This will not affect the demand but gives an incentive to the service provider to deliver a specified service quality. It encourages the service provider to deliver a certain "high" level of reliability; the shadow tolling does little to ascertain that network users face the appropriate pricing signals themselves.

Box. 5.5. Pro-active road management: Traffic Control Unit in Tyrol, Austria

Austria has applied a pro-active road management system on a strategic part of its road network; the systems that have been installed have facilitated the reduction of accidents and improved recovery time from non-recurrent events. The outcome has been to improve reliability, together with other safety and network capacity benefits.

The traffic control unit in Tyrol in the Alps covers a network of 125 kilometres. The roads A12 and the A13 were chosen for two reasons: First, efficient traffic management in these sections has absolute priority due to the high rate of international traffic. Second, for reasons of environmental and residential protection, it is priority to monitor traffic flow variability, not only dependent on the traffic volume, but also dependent on weather conditions.

After almost two years of operation traffic accidents were reduced by 35%, hours of congestion was reduced by 20%, and the number of accidents with personal injury has been reduced substantially by 31%.

Calculating the benefit in relation to the cost of the Traffic Control Unit, a quite positive result is observed. On an annual basis, there are total costs of 72 million Euros. The operator's benefit can be mainly seen as saved financing costs. More importantly, planned extensions of infrastructure can be postponed due to the increase of capacity through the use of the Traffic Management System. This benefit has been calculated to amount 33 million Euros per year. The user's benefit consists of a time cost reduction amounting to 74 million Euros per year. The macro-economic benefits account for 141 million Euros per year due to avoided accidents. Therefore a CBA factor of 3.35 can be calculated.

Source: Kummer and Nagl (2007).

Finally, also speed management can be considered as a way to improve reliability. Limitations of the speed limits to 80 km/h on a freeway to Barcelona has been motivated, in addition to environmental and safety aspects, by reliability improvements. It should be noted, however, that no assessment of impacts has been made so far.

5.3 Managing interfaces

Network users often face delays at different types of interfaces. These interfaces can be either organisational or modal or between network providers or between political entities. They can be a source of delay that can be reduced by good management. There are four main types of interfaces that are liable to delays:

- Those between ports and hinterland connections.
- Borders.
- Different network providers.
- Organisational interfaces.

The contention is that better management of these interfaces can deliver improvements in reliability.

It should be noted that the number of interfaces does not necessarily have to be *minimised.* It is true that network users face delays at organisational/modal interfaces. However, logistics chains – which can actually increase the number of interfaces – illustrate that interface issues (such as unreliability) can be sufficiently addressed and managed so that the risks/costs of the interfaces are less than the benefits from this supply chain approach.

These interfaces are now considered, with a focus on how operations can be improved through better management.

5.3.1 Ports and hinterland connections

The main challenge in the port-hinterland interface is to maintain reliable land services. The interface between land and maritime networks is crucial for most of the goods other than heavy bulks. Very large ships can be used to carry large quantities of goods by sea, but moving such large flows overland is much more difficult: funnelling those large volumes of goods into and out of the port in a short berthing time is a logistics challenge that has important reliability consequences. A virtuous circle has emerged in which demand for transport is stimulated by the reliability and service quality offered by intermodal transport chains, which in turn generates further supply. The growth of traffic may be in the process of breaking this virtuous circle due to increased congestion and the impossibility of ensuring the same conditions of reliability and service quality for all traffic in port hinterlands.

Hinterland congestion is not only a port problem. It is suggested that from the port and supply chain perspective, reliability – which is correlated with but different from congestion – may matter more than congestion itself (JTRC 2008).

Road is the predominant transport mode for carriage to and from ports. Blumenhagen (1981) argues it offers many advantages over alternatives. The density of the road network offers comprehensive coverage of the territory. The haulage industry is often fragmented, with a handful of large firms and a very large number of self-employed drivers, so that availability is almost immediate. It is convenient, both for the document chain (a limited number of parties involved) and for terminal operations (loading of the container and immediate departure of the vehicle). The journey from the port to the customer's warehouse is direct, with no unloading/reloading. It thus offers a high level of reliability and service quality.

Port authorities, shipping lines, freight forwarders and shippers are all, for different reasons, looking for alternatives to road transport. Even assuming very proactively policy-driven scenarios in which combined transport would play a much greater role, road transport is likely to continue growing strongly in absolute value in coming years. But volumes have increased to a point where decreasing capacity of access roads undermine the efficient operation of the port.

Combined waterway/road or rail/road transport may help to ensure the long-term reliability of end-to-end transport by increasing the massification of land transport in response to increased volumes. However, it has to offer a level of reliability and service quality at least equal to if not greater than that of road transport (Vellenga *et al.*, 1999).

Service integrators offering shippers an integrated end-to-end service through a vertically integrated transport chain are essential for the development of combined transport (ECMT, 2006b). Further increases in traffic – under conditions of reliability and service quality comparable to those that already exist – are bound to require new forms of organisation in ports and at maritime and inland terminals.

All transport modes are likely to come under very strong pressure. Even if there is a relatively sharp rise in combined transport, growth in road transport in absolute value terms will remain significant. Given that motorway networks are already extensive and that environmental considerations are hardly favourable to further expansion,[1] that is likely in the long run to mean higher prices for road transport and more congestion in and around major port cities.

The world's biggest ports are all situated in conurbations with several million inhabitants; this aggravates the situation, especially since there is virtually no competition for trucks when it comes to carrying goods over relatively short distances to and from terminals. To take Hamburg as an example, trucks account for over 80% of journeys of less than 150 km to and from terminals, whereas road transport accounts on average for less than 70% of all traffic. In fact, port traffic and urban traffic mingle on the same motorway. The bigger the metropolis and the more varied its activities, the less the population and, thus, local politicians, like the port. They are seen more as a nuisance than as creating added value, a syndrome that affects virtually all of the world's major ports.

Port authorities respond to congestion by lobbying for new roads and by seeking to extend terminal opening times so that trucks can travel outside peak times. That has been the case with Los Angeles/Long Beach, which launched the PierPass OffPeak Program in July 2005. Another solution is to develop relatively close inland centres linked to terminal containers by mass transport modes like rail or barge (Slack, 1999, Notteboom and al., 1999). Where there is an existing waterway, it is relatively inexpensive to develop inland centres, as has been the case with Antwerp and Rotterdam.

In contrast, rail corridors require massive investment (Notteboom *et al.*, 2007). The best-known example is probably the Alameda Corridor, a 32 km dedicated freight line linking the terminals of Los Angeles and Long Beach ports with the inland Burlington Northern -Santa Fe Intermodal Container Transfer Facility (ICTF) where double-stack trains can be assembled. The Corridor greatly improved the reliability and travel times of train traffic. The length and range of travel time was reduced from range of 2 to 6 hours to about 45 minutes (Giuliano, 2004). The same rationale applies to the Betuweroute, opened in June 2007, linking the Maasvlakte terminals at the port of Rotterdam to Zevenaar on the German border.

The interface between maritime networks and land transport networks may also be affected by poorly functioning port terminals.

Long waits at terminals have serious consequences for truckers because they limit the number of their daily rotations in urban areas. The consequences of such delays are equally negative for barge operators, as can be seen from the situation at Antwerp or Rotterdam. Barges currently have to wait between 24 and 72 hours at the terminal even if they are announced and arrive on time. The delays are all the more harmful for short-distance waterway lines, since the wait represents a substantial proportion of the total journey time. Whereas it used to be possible, without congestion, for barges to make the trip between Rotterdam and Duisburg twice a week, it is now no longer realistic to envisage more than one trip a week.[2]

5.3.2 Border management

By their nature, transit operations are extended in time and space involving several countries. Transit trade requires more oversight than domestic trade over similar distances. Such trade depends on measures taken by countries to regulate vehicle movement (drivers), and trade in services. Crossing borders means waiting time for document controls. Border delays can be a great concern if institutional arrangements are inadequate. For example, US trade with Canada and Mexico has grown about 90% since NAFTA took effect in 1994. The US highway and rail networks are now strained, especially at border crossings. Predictable travel times are especially important in a global economy where many goods are needed in tightly scheduled manufacturing and distribution systems (Public Roads November/December, 2004).

Government administrative procedures can be a major source of unreliability. As illustrated in Table 5.2, there can be a very wide range of time that road freight operators spend in waiting to cross the USA land borders.

Table 5.2. **Waiting time at the USA land borders (minutes)**

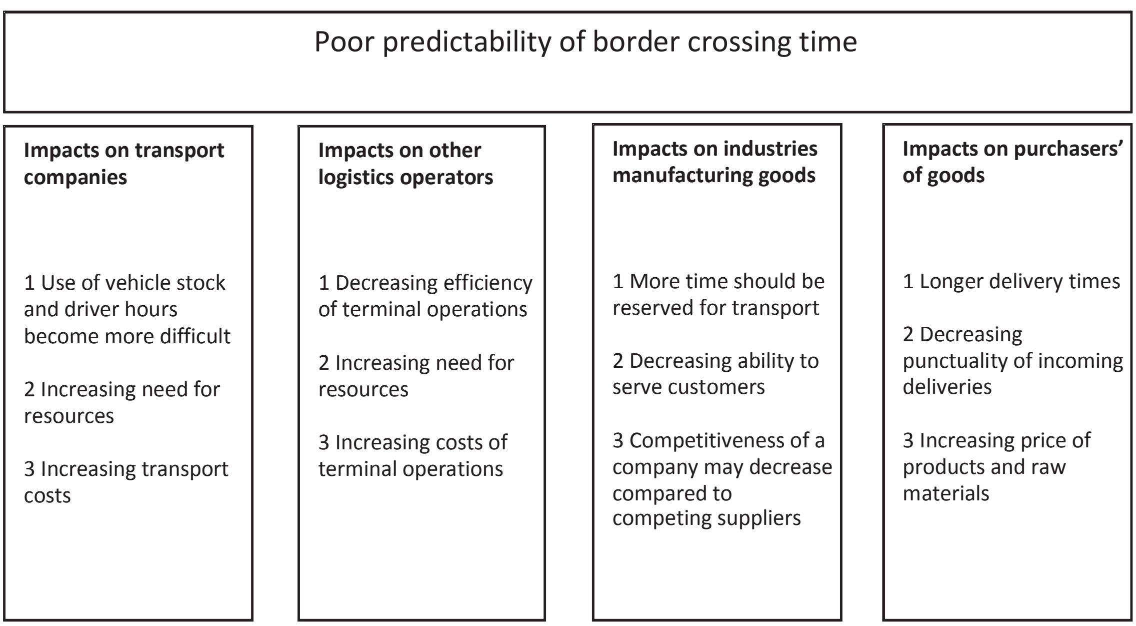
Border Post	State of Entry	Baseline	Average	95th Percentile
Ambassador Bridge	Michigan	5.7	8.8	13.9
Blaine Crossing	Washington	8.1	17.3	35.6
Blue Water	Michigan	11.1	34.2	80.3
Peace Bridge	New York	8.3	21.5	83.4
World Trade	Texas	12.2	31.2	54.9

Source: Arvis *et al.* (2007).

A long time spent at the border and poor predictability of border crossing operations may constitute a significant problem for trade.

Unreliability of the border crossing causes inefficiency in the operation of the vehicle stock, generating additional costs in terms of salaries, capital cost of the vehicle stock, capital cost of the cargo, fuel, and emissions. In addition, sales revenues and even some customers may be lost if products are not delivered in-time. These impacts are illustrated in Figure 5.1.

Figure 5.1. **Impacts of poor transit-time predictability at border crossings**



A further example of border-crossing unreliability is at the primary Finnish-Russian border crossing. Border crossing issues have created a bottleneck for freight transport directed to Russia. Road E18 in Finland is the main road leading from Turku along the southern coast through Helsinki to the Russian border. There is an investment plan to upgrade the whole road section into a motorway. This should have a significant positive effect on the national competitiveness of Finland and the operating conditions of companies along the entire transport corridor. However, the benefits of this major infrastructure investment will not be fully utilised if the reliability of cross-border transport operations is not improved. The severity of border delays is described in Box 5.6.

Box 5.6. **Unreliability problems at the Finnish-Russian border**

Freight transport between Finland and Russia has increased by about 20-30 %/year since 2002. There have been record-breaking, even 50-kilometre long queues at the border stations between Finland and Russia in Southeast Finland in recent years. In addition to serious safety, environmental and trade concerns, this undermines freight transport reliability.

A recent study on the Finnish-Russian border crossing operations concluded that it took 6 working days for a lorry to make a round-trip from Finland to Moscow. Four days of this was the driving time and two days was spent for various delays mainly in border crossing. According to a follow-up study, the total time spent in border crossing operations from Finland to Russia typically varies between 1.5 and 16 hours.

In addition to the high volume of traffic crossing at the borders, delays also arise from the different types of goods being transported. Weather conditions, too, play a part, as well as customs and other border crossing operations on the Russian side.

It would seem that the appropriate solution would be for border authorities to seek, in the first instance, bureaucratic solutions to the border problems. One of the proposed solutions has been to increase border-crossing capacity in cooperation with the Russian authorities. Another solution proposed has been to set up large waiting areas for lorries. In order to mitigate problems, measures include several temporary waiting areas, guidelines for measures to control the lorry queues and directing traffic more evenly over the various border crossings.

In the absence of a resolution, some shippers have found longer routing around the border. The mobile-phone company, Nokia, is located in southern Finland along Road E18. The company has selected a totally different route for transporting consumer goods to Moscow. As punctuality, reliability and security are the ultimate demands for an efficient transport chain, Nokia has selected more expensive air transport from Finland through Dusseldorf to Moscow. This decision has been made due to unreliability in crossing the Finnish-Russian border along Road E18.

Source: MINTC (2006).

5.3.3 *Managing organisational and modal interfaces*

Interfaces between organisations may be a source of significant unreliability that can be improved through enhanced coordination.

A multi-firm logistics team can be used to reduce co-ordination costs in freight movement and focus attention on bottlenecks in the logistics chain that both constrain capacity and reduce service reliability. The case study on Hunter Valley coal supply chain illustrates how the overhaul of the institutional framework can enhance capacity delivery and utilisation and, therefore, reduce the occurrence of incidents that lead to unreliability (Box 5.7).

This also provides a case study on how government may have an underlying “facilitation” role. This includes protection from government regulations that might otherwise see the co-operation as collusion and/or anti-competitive. In this case, improvements in network reliability are being achieved through the co-operation of entities – particularly train operators and coal mining companies – that would otherwise compete directly with each other. Indeed, mandated access regulation encourages that competition.

Box 5.7. **Improving reliability and capacity in Australia's Hunter Valley coal supply chain by improving inter-organisational interfaces**

Coal trains operate between the Hunter Valley and Port Waratah, where most of the coal is transferred to ships for export. In 2009, there is expected to be a demand for shifting around 100 million tonnes from the Valley to the coast. Rail capacity on the network is uneven, with bottlenecks on the network causing reliability problems. The consequence of these perennial problems is that, ultimately, the amount of coal available for export is constrained. That is, the railway haulage capacity is less than the mining or the port capacity. Resolving these bottlenecks has been prolonged due to the fragmented ownership/management nature of the coal supply chain.

Two approaches have enhanced reliability and capacity. The first has involved targeted investment to expand capacity. The government, through its rail infrastructure corporation (Australian Rail Track Corporation), has invested in removing key physical bottlenecks that have constrained capacity and led to service delays.

The second policy has been to establish a logistics team consisting of employees of the coal mining companies, train operators, track managers and the port operator. The Hunter Valley Supply Chain Logistics Team co-ordinates operational activities (train operation and track maintenance especially) with coal and maritime requirements

Government has an underlying "facilitation" role. This includes protection from government regulations that might otherwise see the co-operation as collusion and/or anti-competitive. In this case, improvements in network reliability are being achieved through the co-operation of entities – particularly train operators and coal mining companies – that would otherwise compete directly with each other (See Affleck 2005, p. 18). Although participants in the logistics team co-operate in order to maximise coal throughput, tensions inevitably remain between the members. In particular, network users (shippers and train operators) would prefer additional capacity to be provided, and to advance the delivery date of any agreed capacity expansion. Inevitably, network providers prefer a higher degree of certainty regarding demand as the investment costs can be recovered only when users take up the extra capacity

This case study focuses on strategic capacity expansion more than reliability, per se. Nonetheless, it illustrates how the multi-firm logistics team can be used to reduce co-ordination costs in freight movement and to focus attention on bottlenecks in the logistics chain that both constrain capacity and reduce service reliability. The case study illustrates how the overhaul of the institutional framework can enhance capacity delivery and utilisation and, therefore, reduce the occurrence of incidents that lead to unreliability.

Source: Affleck (2005).

A similar co-ordinated approach to infrastructure oversight is evident in road corridors that cross jurisdictional boundaries. The following example from the Interstate-95 road coalition in the United States illustrates the benefits flowing to reliability and to incident mitigation from adopting a co-ordinated approach to managing a transport corridor (Box 5.8). Roads form a national network but management of some corridors (including incident management) can be administered at a local level. Levels of reliability may be impeded by the dispersed management system. Similarly, incident management at this disaggregate level may result in inconsistent responses. Reliability is enhanced indirectly because of better, more coherent corridor management of infrastructure provision and management. The organisations formed on a range of corridors provide compatible infrastructure to facilitate seamless traffic movements across jurisdictions. The organisations can also ensure there are coherent information systems that disseminate traveller information on a corridor basis rather than a jurisdictional basis.

While the formation and operation of such organisation is not costless, their impact on supplying reliable infrastructure, and information that can mitigate the impact of incidents, may be relatively cost-effective.

Box 5.8. **Reliability enhancement through cross-jurisdictional network management – the Interstate -95 coalition**

The example considered here is part of a US National Highway System coordination programmes, known as the Multi-state Transportation Operations Programs (MSTOPs); the Interstate-95 Coalition is an early application. (TranSystems Corporation, undated)

The Interstate 95 Coalition is a co-ordinating partnership of around 60 organisations, including relevant State and local government transport and law enforcement agencies and transport providers. The geographic coverage of the coalition is between Maine and Florida, with relevant Canadian affiliate members. Essentially, the coalition seeks to manage the corridor as a seamless network. By providing co-ordinated management of major transport incidents, the impact of that incident can be reduced through co-ordinated mitigating actions. During major incidents, the partnership coordinates traffic management, law enforcement, fire, safety and other emergency activities.

The links between the agencies are strengthened and this facilitates the coherent transmission of information to network users about the network performance. The coalition enhances interaction between the agencies. This interaction can facilitate the uptake of compatible technologies (where appropriate) across the different agencies, such as electronic tolling systems. Such compatibility can reduce the likelihood of delays. Similarly, the coalition facilitates the development of information across the agencies to ensure consistency in multi-jurisdictional operational and capacity-enhancing activities.

The coalition therefore facilitates reliability because of better management and decision-making in recurring and non-recurring traffic incidents. The Transportation Research Board concludes that the MSTOPs: "have proven vital to the reliability and security of key interstate corridors. Weather information, emergency operations, goods movement, homeland security, and traveller information have been substantially enhanced through interstate relationships and partnerships that typically engage the transportation and public safety communities".

Source: Baniak (2002); TRB (2007); TranSystems Corporation (undated).

So far, the management of infrastructure in delivering reliability has been discussed; the management of the supply of services can be equally important in delivering reliability. For instance, a range of complementary actions by a logistics provider and train operators can deliver greater reliability for any products that reliy on on-time delivery.

The actions seek to streamline train operations so as to reduce the potential activities that can be sources of delay. The following example illustrates, for instance, how the siting of the terminals facilitate the reduction of reliance on roads (see Box 5.9). Having left the terminal, the financial involvement of the railways in the terminal provides incentives to the railway to ensure that the train is given high priority across the network. The operation of a unit train keeps the train service operationally simple so as to reduce activities that might cause delays; routeing the train so as to bypass railway yards reduces activities and delays that arise within the railway yard. Operating the train across multiple railways but treating it as if it is on a single system reduces the likelihood of delays that might arise when the train is interchanged across the systems. Finally, reliable delivery is backed by a GPS tracking system that assures shipper, logistics company and partner railways alike as to the progress of the goods. The GPS also beams monitored statistics on the environment (temperature) within the refrigerated vans.

Box 5.9. **Managing service reliability: fresh produce trains**

Unreliable and slow freight across USA have prevented north-western farmers from supplying their produce to the primary population centres on the eastern side of the country. Traditionally, railways, in particular, offered unreliable services, which was unsuitable for perishable products. Two key sources of delay arise when trains are at terminals (including intermediate yards between the origin and destination) and when trains are interchanged at railway operator "borders".

This case study provides a holistic management solution that focuses on each source of delay and find ways of reducing them. For instance, the produce trains are permanently coupled as unit trains; this obviates the need for the trains to be routed via intermediate terminals and so removes one potential source of delay. The potential for delays at the interchange between the two railway companies is minimised by planning, managing, marketing and operating the train as if it is operating over just one railway operator. Finally, the two railways and the logistics shipper introduced a tracking system – a single tracking system – for the perishables wagons. The tracking provides vital information that can be used by the relevant parties in managing/responding – mitigating actions – to the progress of the train.

The project was undertaken in conjunction with local farmers and authorities and with the railways. The aim was to streamline processes required to convey fresh farm produce from Wallula (Washington State) to Rotterdam (New York). Crated goods are conveyed to the Wallula terminal by road for onwards movement to the east coast by a unit train (the "Produce Railexpress" or "Produce Unit Train") of louvered rail wagons.

Railex committed to at least one train in each direction per week. In return, the railways (who part-funded the terminal infrastructure investment) agreed to prioritise the shipments across the network. The railways' investment in the service provides them with an incentive to work with the logistics company to maintain reliability; this includes ensuring that trains are given sufficient priority over other trains. The trains are stopped only for crew changes and a single refuelling. Close co-operation between the two railways – Union Pacific and CSX – ensures that the physical interchange between the railways does not become a source of unreliability. The priority given to the trains ensures that they can (and do) traversed the distance in 72 hours – well within the 124 hour guarantee window (including 8 hours of "cushion" time) (Philp 2007). The railways can deliver this degree of reliability, and the guarantee time is still considerably faster than the previous transit times.

The location of the terminals facilitates maintaining reliability on the intermodal aspects of the logistics chain. The reliability extends beyond the railway part of the logistics chain because, at either end of the corridor, the trains are loaded and unloaded at locations that are within a day's drive of the source or destination of the goods. The proponents of the service argue that such short-haul has relatively fewer reliability issues than when operations from terminals are beyond one-day's travel.

The value of this service to shippers relies entirely on its reliability. The ability to provide a reliable service has been achieved by applying a range of measures that collectively reduce the likelihood of delays. Various aspects of this holistic approach have been applied to a number of other long-distance trains in USA, including the BNSF/CSX "Cold Express", the Union Pacific/Norfolk Southern "Blue Streak" service and the California Northern/Union Pacific/CSX/Hub Group (logistics company) "Wine Connection service.

Source: Hopkin (2006); Natarajan *et al.* (2005) ; Philp (2007) ; Washington State Transportation Committee 2006.

5.4 Conclusions

This section focused on how the management of infrastructure and services can be used to deliver better transport reliability. A range of examples illustrate how management systems can improve reliability. Improving management can provide higher reliability, reducing capacity needs. Thus, the gains from better management of networks and services can be far more cost effective in delivering higher reliability – especially relative to capacity expansion.

Pro-active management of network through integration of robustness objectives into the network planning models enables assessment of network vulnerability. The Dutch network planning system illustrates how reliability is being incorporated into network planning models; this enables pro-active assessment of network vulnerability. Such planning may have significant reliability benefits. The essence of the strategy is that active network management, involving intensified oversight of network use, allows faster reactions to disruptions, hence, increasing reliability.

Better active management of incidents can provide significant reliability benefits. An example of active management is the use of motorway hard shoulder capacity. Preliminary analysis on their use during congested periods has shown reliability benefits. Sometimes maximum road speeds are reduced when hard shoulders are in use. The longer transit time arising from the reduced traffic speed is traded with the reduced travel time variability. The use of incident management systems, in turn, enables transport service providers to react faster to disruptions thus reducing the duration and severity of the events.

Network users often face delays at organisational, modal, or other interfaces. Poor active management at border interfaces can generate unreliability. A long time spent at the border and poor predictability of border crossing operations constitutes a significant problem for trade. Unreliability of border crossings causes inefficiency in the operation of the vehicle stock, generating additional costs in terms of salaries, capital cost of the vehicle stock, capital cost of the cargo, fuel, and emissions.

The use of single-point management teams (such as used in the Hunter Valley coal logistics and the Interstate-95 Coalition), and targeted strategies at country borders, can ease unreliability in the network by identifying bottlenecks and co-ordinating responses to the bottlenecks. The single-point teams facilitate breaking-down problems arising at organisational interfaces.

In case of organisational interfaces, governments may have an important facilitation role and can improve reliability through enhanced interface coordination or coherent corridor management.

NOTES

1. For example, in October 2007 in the context of the "Grenelle de l'Environnement" environmental forum, the French Minister for Ecology and Sustainable Planning and Development announced that, with the exception of bypasses, no new motorways would be built in France.

2. Werner Kühlkamp, transport expert for the Lower Rhine and Duisburg Chamber of Commerce and Industry, quoted in *Journal de la Marine Marchande*, 24 August 2007, p.26.

KEY MESSAGES

- Reliability can be influenced through pro-active network management. Improving network management reduces capacity needs and improves reliability.

- Pro-active management of the network through integration of robustness objectives into network planning models enables assessment of network vulnerability. This type of model can identify vulnerable segments in a network, helping network operator to analyse acceptable alternative routes, evaluate how users can be inform about incidents and provide valuable information on future infrastructure needs.

- Active network management involves measures to intensify oversight of network use, allowing faster responses to disruptions thus reducing vulnerability and increasing reliability.

- An example of active management is the use of motorway hard shoulder capacity; preliminary analysis on their use during congested periods has shown reliability benefits. Sometimes maximum road speeds are reduced when hard shoulders are in use; the longer transit time arising from the reduced traffic speed is traded with the reduced travel time variability (improved reliability).

- Better incident management may provide significant reliability impacts. Efficient incident management systems enable infrastructure managers to react faster to disruptions thus reducing the vulnerability of the transport system and increasing the reliability.

- Network users often face delays at organisational, modal, or other interfaces. Organisational delays can be minimised through management interfaces such as the Hunter Valley coal logistics team and the Interstate-95 Coalition.

- Poor active management at border interfaces can also generate unreliability. Time spent at borders and poor predictability for border crossing operations constitutes a significant problem for trade. Unreliability causes inefficiency in the operation of vehicle stocks, generating additional costs in terms of salaries, capital costs, fuel, and emissions.

- In the case of organisational interfaces, governments may have a facilitation role and can improve reliability through enhanced interface coordination or coherent corridor management.

6. PRICING NETWORKS TO OPTIMISE RELIABILITY

This section reviews the principles of pricing for network use to influence the *level* of network reliability through demand management; and to supply *differential* levels of reliability reflecting granularity in user values of reliability. It has been noted, earlier, that reliability is often bundled with other attributes such as speed, convenience and cost, from which it can be very difficult to differentiate a separate market for reliability. Following on from this issue, this section considers how practical it is for pricing to deliver the levels of reliability consistent with diverse values placed on reliability.

This section, first, considers the prevailing strategies adopted by network providers in delivering reliable networks; and different users' needs for reliability. The discussion then outlines the difficulties in influencing reliability and providing a market with differential levels of service. Next, practical examples are given that illustrate the potential and limitations of pricing to influence the level of reliability. Conclusions are then drawn about pricing as a reliability policy tool.

6.1 Prevailing pricing strategies

Insight into the benefits of pricing reliability can be obtained by considering briefly the application of pricing to electricity provision. It can be shown that using pricing to isolate power users who seek high levels of reliability can actually benefit all power consumers. The example also warns of the dangers inherent in using an aggregated reliability values to reflect all users.

Box 6.1 provides a brief overview of pricing for reliability in energy supply in USA. It should be noted that the provision of differential reliability in energy provision is inhibited; this is because government regulations specify a given (high) level of reliability – that is, minimal power outages. In spite of this, a market in reliability has developed that reflects the needs of some network users for higher levels of energy-supply reliability than the mandated level.

Some electricity generating companies have introduced pricing differentials in services offered to companies that require high levels of supply security. The generators offer such users a discounted electricity price for those able to opt out of the grid at times of high electricity usage. These users can opt out because they own back-up generators as insurance against outages. Generation costs in general are then lower (and the benefits are shared between generating companies and users) because peak generation levels and, therefore, costs can be kept lower than they might otherwise be.

An important aspect of this example is the value of understanding diversity in needs for reliability. In general, surface transport network providers do not explicitly identify user "reliability" as a specific level of service attribute. It follows that there is neither a price applied to reliability nor, more generally, a market for reliability.

The absence of a pricing system removes crucial guidance as to the appropriate network standards. Where analyses do incorporate reliability benefits, the approach usually assumes only one level of reliability. That is, the analysis assumes that all road users are prepared to pay the same price for a given level of reliability. However, as discussed earlier, road users will not be homogenous in their demand for reliability.

Box 6.1. Lessons from pricing energy supply reliability in the US

Electricity supply provides a useful case study on how reliability is priced in other sectors. The focus on optimal reliability in the electricity industry is relatively mature because there is considerable regulatory oversight of the security of power provision. The major consequence of unreliable electricity delivery is the cost of power outage. What is worth noting is that there is a wide disparity in the value that different users place on reliability, as expressed by their willingness to pay: the cost of a power failure (or outage) varies significantly across electricity users and also between locations and time of day.

Regulatory intervention in electricity supply has generally led to a requirement that specific levels of reliability be maintained. An important tool to achieve these target levels have been prices that vary directly with the level of reliability. The lower the level of reliability, the lower the price. The lowest prices are available to those customers willing to be disconnected from the grid during periods of high power demand.

At the other end of the scale, a ceiling on the price that users will pay for high levels of reliability is provided by the cost of alternate back-up systems. Where an extremely high level of reliability is required, the lowest-cost option may be for the users to have their own emergency power generators. Alternatively, since they have access to their own system, they may also opt for the lower unit rate that involves being disconnected from the grid during periods of peak demand. In this way, the system has the potential to encourage an outcome that reflects both an efficient *level* of reliability and an efficient *sharing* of the responsibility for achieving this level. Despite pricing regulatory rigidities, prices can still be utilised to ration power between customers according to the value they place on reliability.

It is clear from a review of the electricity sector that the industry, its customers and regulators have different opinions on delivering an efficient level of reliability. Diverse transport reliability needs – and the need to have an open market that optimises the supply of, and demand for, reliability – has resonance in the electricity sector. Diverse power outage costs reflect diverse users; it warns of the dangers inherent in using an aggregated reliability value for all users. Woychik (2006) stresses the need – very akin to reliability in transport – for understanding electricity user needs. He notes the marginal cost of incremental capacity, the price cap on electricity prices and the diverse needs of users, may dictate different responses – and price caps certainly may discourage the provision of additional generation.

In the context of different user needs, Burns (2004) argues that it is because of the very diversity of electricity users/outage costs, that aggregated averages of costs of outage do not provide policy-makers with useful information as to what should be done. He cites a range of costs of an hour of outage on a summer's day: $2.90 for residential customers; $1 200 for small commercial and industrial customers; and $8 200 for large commercial and industrial customers (Burns 2004). One group of customers will be far more likely than another customer to pay to achieve higher reliability while another group will be content to accept lower reliability in return for lower overall electricity charges.

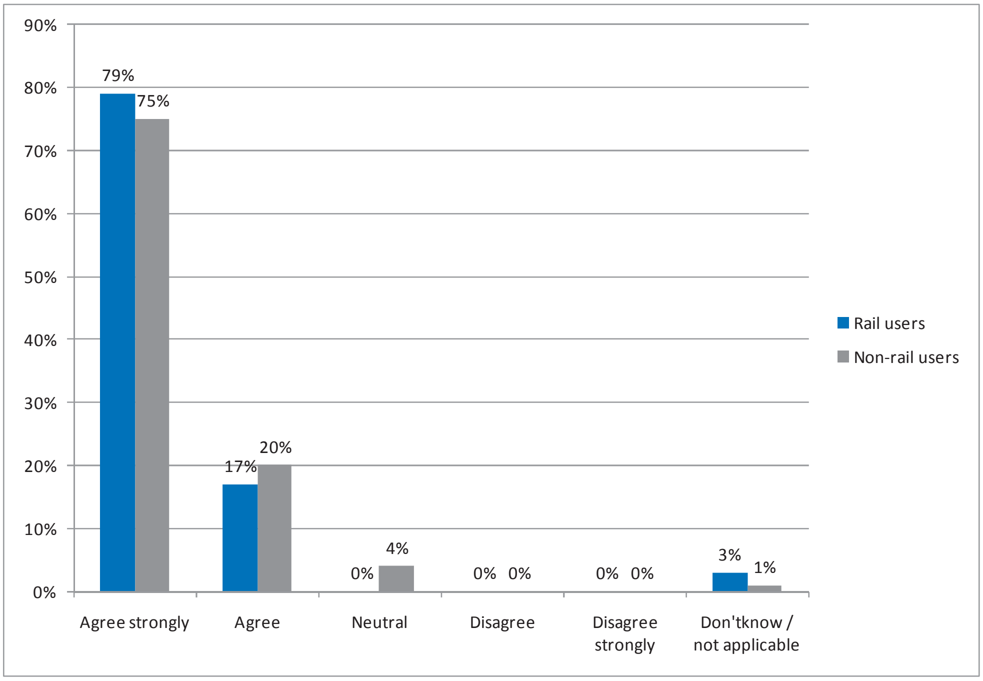
Burns argues that given the rich diversity, it is necessary to take a more granular approach to reviewing the cost of outage. He notes that following this "granularity" in the individual needs for reliability, a range of policy responses can emerge. Such responses may be to restructure the pattern of electricity generators, or to enhance or expand grid transmission systems between generators and customers or, finally, to undertake demand-side management of electricity consumption.

Thus, experience from electricity warns us of the different policy responses that come from diverse users and, thus, the need to have a good, disaggregate knowledge of users. Charging mechanisms are one of the systems used to tailor electricity supply to appropriate levels of reliability.

Source: Burns (2004); Woychik (2006).

Because the link is very weak between pricing access to the network and the standard of network that different users require, road infrastructure managers must estimate a level (generally only one level) of reliability to be provided. Stated-preference surveys are often used but results always need to be treated with care. Surveys (such as Figure 6.1. below) are of limited, if any, value to the infrastructure operator as the preferences stated do not necessarily represent a willingness to pay[1]. This provides no guidance as to what they would be willing to pay for higher levels of reliability.

Figure 6.1. **Illustration of demand for reliability**



Source: City of Chicago Department of Transportation (2003).

More generally, network provision does not directly recognise reliability as an attribute nor recognise diverse values in reliability. Use of the road network is commonly charged with two-part pricing: a fixed annual charge for a given vehicle type and a variable charge (typically applied as a tax on the fuel that is used). That price does not vary if there is a poor level of service or, in particular, a poor level of reliability. On the flipside, the road charges generally do not vary if the provider does enhance reliability.

As discussed in preceding sections, network providers can influence the level of reliability, albeit that the financial or political costs may be unpalatable. This is particularly the case with capacity expansion so it can be vexing to planners that users tend to see unreliability as arising from an insufficient level of capacity. Users then place pressure on the network provider – that is, the government or its agency – to expand capacity.

In facing the challenge of delivering passenger and freight transport reliability, the core policy issue is determining, from a community perspective, an "optimal" (or "efficient") level. As the foregoing discussions have indicated, "optimising" cannot mean maximising reliability. This can be readily understood in the context of road congestion. The "optimal" level of road congestion does not mean zero congestion, as noted by the UK's Department for Transport:

> "… it does not imply that free flo*w speeds are the aim.* There is no case for providing free flow conditions for all road users at all times. The cost of doing so would greatly

> outweigh the benefits. It would be extremely inefficient to build a wide road wide enough to provide free flow *speeds* during the peak because for most of the day most of the lanes would be empty." (DfT 2006)

From a similar perspective, even if it were possible to achieve 100% reliability, the costs of doing so would generally outweigh the benefits. Efficient levels of reliability are determined by the same principles of supply and demand as other goods and services. In a market economy, the market price and quantity are determined by the interaction of demand and supply. For the purposes of this report, it is sufficient to note that the economists' standard downward-sloping demand curve and upward-sloping supply curve reflect declining marginal utility and the increasing marginal cost of supply of a given level of reliability.

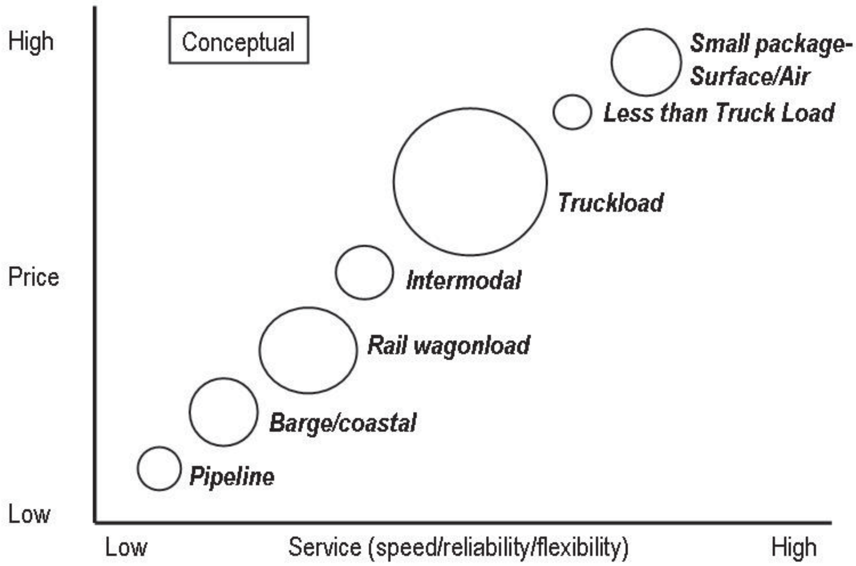
Crucially, as discussed in Section 1, the downward-sloping demand curve for reliability embodies the diverse values that users place on reliability, which varies across products, locations and firms. Demand for reliability is differentiated – or "granulated".

In principle, then, network providers need to be responsive to the need for reliability and, in particular, to differential levels of reliability. It is possible to outline a price-reliability pattern across freight movements; the policy challenge for network providers is to be responsive to that. That said, it is also the case that if the transaction costs of achieving that pattern is excessive, then in the worst-case scenario it may mean retaining current strategies, with little or no differentiation.

It needs to be recognised that, *in aggregate*, the various transport modes provide services with different reliability levels. Figure 6.2 illustrates the size of different transport markets in USA and where the markets tend to lie in terms of the price/reliability trade-off. This means that a shipper *may* be able to choose a price-service quality (including reliability) combination.[2] Some modes, such as barges, have low service characteristics (that include reliability standards) but freight is conveyed at a relatively modest price; other modes, such as air freight, have high service quality products for a relatively high price (such as premium-priced small packet shipments). There are also options within modes, such as truckload and less-than-truckload, each with different service standard-price characteristics.

However, in practice, shippers' choice of reliability levels will be limited, in the first instance, by the choice of modes that is available between given origin and destination pairs and by the extent to which a given service provider can control that reliability. There are a number of factors that influence reliability, some of which are within the control of the service provider.

Figure 6.2. **A typical price-reliability spectrum, with circle size illustrating traffic volume (USA)**



Source: Derived from The Tioga Group (2003).

Even if the network provider can reasonably control reliability, the issue is then whether there is sufficient profit in delivering that reliability or in offering different levels of reliability. For example, a railway network provider could expand capacity to reduce the likelihood of congestion-based delays to a given commodity of freight but it might not be profitable and the network provider may conclude that the shipper is not prepared to bear the incremental cost (see, for instance, Box 1.4.).

The common outcome, in particular, is that the shipper faces only one mode and one reliability standard (subject to temporal time variation). The shipper faces an average charge an "average" level of reliability. This leaves dissatisfied both those that are willing to pay more for a higher level of reliability and those that would prefer to pay less for a lower level of reliability. This is illustrated in Figure 6.3., where the price for reliability stays constant on the public road (with no entry restrictions) while the level of reliability varies, rising to but never exceeding R_a. All users face a single price, P_a, with a supply line S_0, for reliability; and for any point in time the level of reliability will be between 0 and $R_{a.}$

Even crude reliability level and pricing differentiation can be important to shippers. The addition of a tolled lane provides the opportunity for specific shippers to use and pay for higher levels of reliability, for which they are prepared to pay; there is no congestion on the lane because many network users are not prepared to pay the supplementary price (toll).

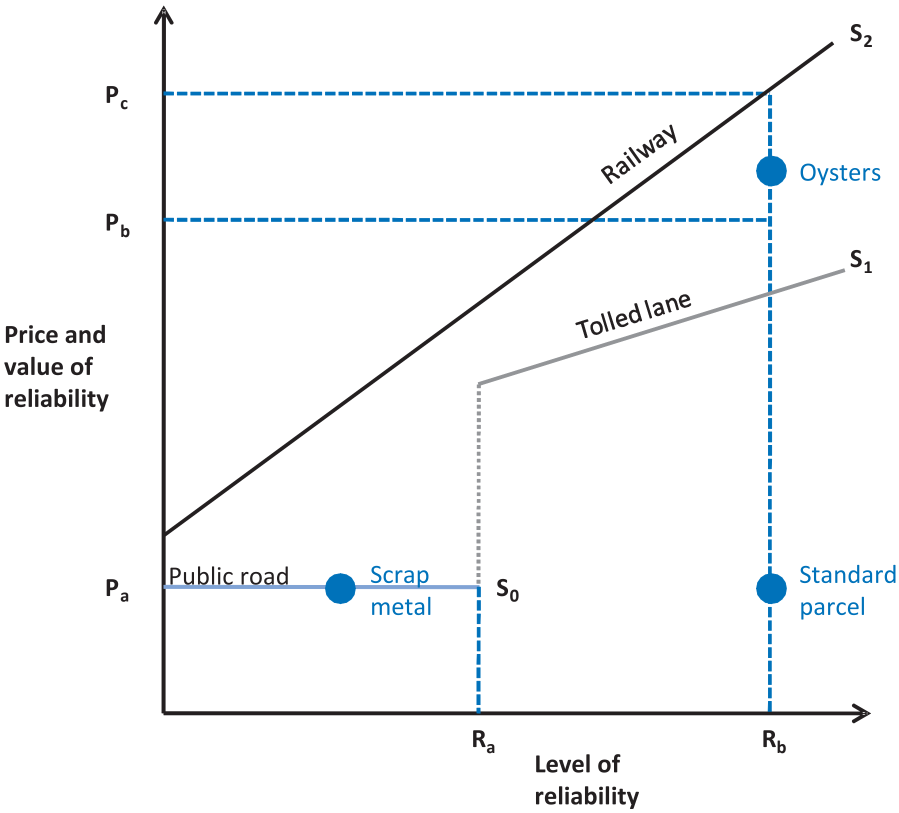
What does this price-reliability level combination mean for different network users? In figure 6.3, three commodity groups are presented – a scrap metal dealer, a standard parcel operator and an oyster merchant – showing different willingness to pay for reliability.

The scrap metal shipper is not dealing with a time-sensitive product and hence places a low monetary value on reliability – in this illustration, a value equating to the cost of using the road.

Some shippers are willing to pay for a higher level of reliability than the maximum that is provided by the market (at R_a), while others might like a high level of reliability, they are not prepared to pay more than the going fixed price, P_a. (as illustrated by the "standard parcel" shipper).

The fresh-oyster shipper requires a high level of reliability to meet the market requirements (with the food being highly-perishable). The shipper is prepared to pay a high price for this reliability – in fact, higher than is necessary if the tolled road is available. In this case, reliability level R_b can be achieved at price P_b.

Figure 6.3. **Illustrative markets for reliability**



In the example shown, the railway can also provide variable levels of reliability, S_2. In this case, the price for a given reliability on the railway exceeds that which can be achieved on the road or toll lane (S_1) and exceeds the price the oyster shipper is prepared to pay.

On most road networks, however, there is neither the alternative toll road nor a railway option as a close substitute. In these situations, network users have no alternative to facing whatever is the prevailing reliability. That said, as discussed earlier, users can improve reliability by travelling at less congested times.

Thus, in the absence of direct pricing signals – which provide information about how much users value reliability – network providers (which are typically government agencies) must then decide on a level of reliability that they should provide (facilitated by planning tools such as cost benefit analysis). And, temporal issues aside, the network providers will supply only one level of reliability – users will not be able to choose different levels of reliability.

6.2 Obstacles to achieving differentiated reliability levels

Given that different market players require different levels of reliability, the core question that arises is: what factors determine whether it is technically *and* economically feasible to segment the market? That is, what issues arise in delivering given standards of reliability and applying differential pricing to users who value those standards.

While market segmentation costs vary between modes, the cost for segmenting road performance and pricing is generally high. The fundamental difficulty with attempting to control road reliability is that access to the network is usually "open"; it then becomes difficult and costly to influence the level of reliability that can be achieved by influencing demand levels. There remain very few examples of where demand has been restricted and supplementary prices applied and the degree of reliability differentiation is low.

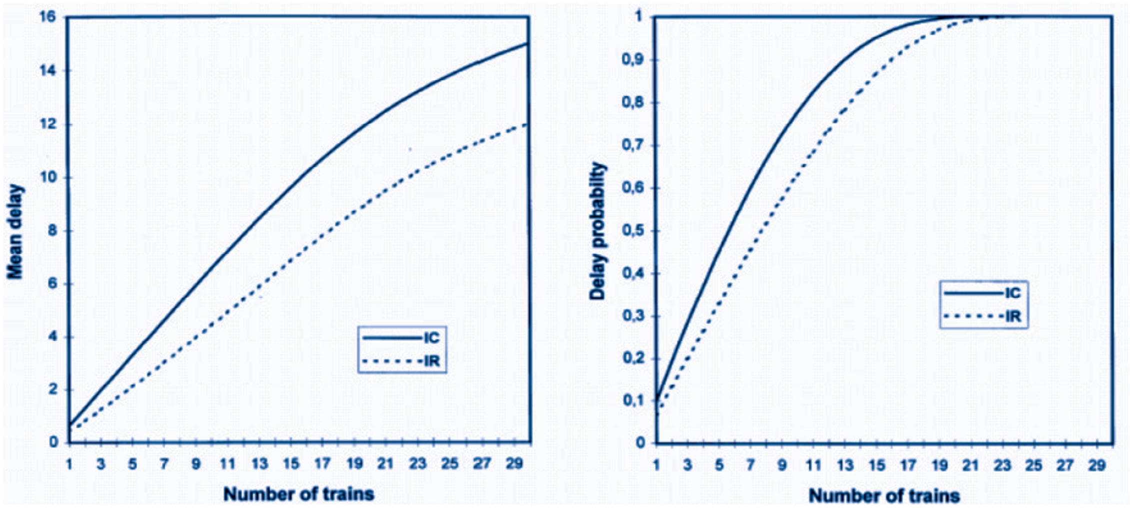
The rail network has different characteristics and, from one perspective, more readily lends itself to providing a segmented market for reliability than the road network. The physical nature of rail operations precludes both the "open" and "random" network use as seen in road use. Trains do not simply appear on the network out of a private driveway. By necessity, there is a "Train Controller" who allocates timed "paths" across the network. The entire network is run according to aggregations of these paths, which are mapped out as timetables. This scheduling is facilitated by the fact that there are generally very few parties seeking access to the railway network – historically only one party. While timetabling is not costless, it is essential for rail operations (including safety) and hence can be assumed into the fixed network running costs.

Enforced timing of the activity is not unique to railways; it can be seen in many areas of activity, such as dental, solicitor and medical appointments. It also can appear in the freight logistics chain, such as with the scheduling of container movements at ports. One such example is Pier*PASS*, a container movement scheduling/reservation system used at the Ports of Los Angeles and Long Beach[3]. An important difference between such systems and railway scheduling is that the administrative (transaction) costs of such systems may exceed the reliability benefits that the scheme delivers, although if the reservation systems are not obligatory then the container movement will not be reserved. In the case of the railways, reserved (scheduled) train pathing is essential for safety purposes: the benefits of scheduling (which bring reliability benefits) invariably outweigh the transaction costs.

Railway reliability can thus be finessed to facilitate the operation of trains to follow the rigid – but predictable – timetables. Further, the "real time" control of individual train movements across the network gives the infrastructure manager the power to respond to unexpected events (on both the supply and demand side) to make "real time" changes to the running order of different trains.[4] This means that in the event of an incident that disrupts traffic flow, a train with time-sensitive goods can be given priority over other movements, allowing it to overtake other trains.

Further, railways can influence the level of reliability they achieve by controlling the number of trains on the network. Like roads, there is a strong trade-off between the number of vehicles permitted on the network and the impact of the interaction of vehicles on the network on the ability to maintain on-time performance. Figure 6.4 illustrates this trade-off. Other things being equal, as the number of trains permitted on a railway increases, the probability of delay and the average length of delay also increases. Thus, while financial considerations encourage high network utilisation, the downside of the greater network utilisation is reduced reliability.

Figure 6.4. **Relationship between railway reliability and capacity utilisation**



IC =InterCity trains, IR=InterRegional trains. Source: Huisman & Boucherie (2001).

This apparent advantage of railways over roads can break down, however, either because railways face other significant reliability issues or because the cost of delivering high reliability is too "high". For instance, on railways, high levels of reliability might be achieved by introducing more sophisticated systems to manage trains or by paring back the number of trains using the network[5]. To justify these actions, the supplementary reliability costs need to be less than the value that shippers place on the reliability achieved by them. The challenge for railway management is to determine the mix of the service characteristics required by their customers, including reliability, which maximises rail's net profits. Thus, as illustrated in Figure 6.3, for reliability level R_b, even the highly time-sensitive oyster shipper finds that the reliability on rail comes at too high a price. The oyster shipper would not be prepared to pay P_c. Given the underlying economics – particularly, the absence of shippers prepared to pay very high rates for the reliability – railways may not seek to cater for the highest levels of reliability.

In brief, while the preference for higher levels of reliability can be readily understood, only when this translates into a willingness to pay would the infrastructure operators endeavour to supply that level of reliability. Whether higher or even different levels of reliability can be commercially provided depends critically on the cost of, and returns to, doing so. This limit to the available options is common to all markets. For example, while there is a wide range of motor vehicles and household goods available for purchase, there will still be some consumers who will not be able to buy their preferred goods and are forced to compromise.

We can draw some conclusions about the obstacles to achieving given reliability levels and offering differential reliability standards. The following are observed from the market treatment of reliability and the extent to which prices play a role:

- Reliability needs to be considered as a normal service characteristic and thus subject to standard rules of demand and supply.

- Because the demand for reliability varies between and within products and users and over time, prices can potentially provide an important avenue for communicating these preferences.

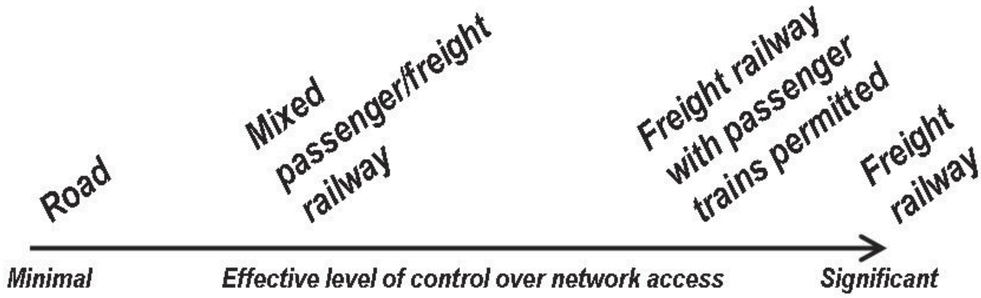
- If market segmentation costs are extremely high, then it may not be possible to provide transport services with different levels of reliability.

- In principle it is harder to segment the road market than the railway market, where access control costs are already incurred for operational reasons and where trains can be prioritised.

- Nonetheless, the costs to railways in supplying a high degree of reliability (in terms of foregone capacity utilisation) may well make it unprofitable to market segment at this level.

Figure 6.5 presents a continuum of control over access to the network. Road appears at the lower end, reflecting typical open access with no control of flows. Rail can appear at the other end of the continuum, where a network manager has complete control of access on dedicated lines for prioritising freight.

Figure 6.5. **Spectrum of effective network controls over reliability level**



6.3 Application of pricing to network reliability

The foregoing discussion indicates the difficulties in applying reliability pricing to transport networks. This is particularly the case with the road network, which is generally open-access to all users who pay the registration/licence fees and fuel excise. Examples are now provided of situations where reliability has been improved by using pricing to temper demand; and where differential standards of reliability are being provided.

6.3.1 Managing reliability through tempering demand

Some road networks use prices and regulations to control access to roads and this can moderate traffic levels generally. The Stockholm and London congestion-charging schemes are illustrations of where price can be used to moderate the prevailing demand generally.

However, only for a few roads around the world are prices dynamic enough to adjust for different levels of traffic. In these cases, the essential pre-condition needed for pricing is that access to the roadway is not open – the entry to the road is conditional on supplementary monetary payment. Where this occurs, the network provider can set access charges that can influence traffic flows and hence the level of reliability. Important examples of this principle are the Interstate-15 outside of San Diego and the MnPASS lanes in Minneapolis. A third illustration of this principle will be laneways on the Interstate-680 in California’s Alameda County, currently under construction.

Box 6.2. **Interstate-15 toll lanes and MnPASS lanes, USA**

In this case, the growth in road traffic has reduced network reliability due to increasing congestion.

Pricing that is attuned to delivering reliable travel times has been applied to segregated lanes on a roadway. The road prices are varied so as to maintain the level of reliability on the lanes. In this case, the speed of traffic is used as a proxy for "reliability". The principle has been applied to laneways on the Interstate 15 (I-15) road that links San Diego and its northern satellite suburbs; and on MnPASS lanes in Minneapolis, Minnesota. A third example, on laneways on the Interstate 680 in California's Alameda County, is under construction, with operation due to commence in 2010.

Road tolls that aim to deliver a given level of reliability are termed "dynamic pricing". The charges are varied in accordance with the flow of traffic on the road, with charges rising so as to discourage some use of the tolled lanes. In the case of the I-15, tolls are varied in 6-minute intervals and the level is set so as to attract the level of demand that is consistent with a constant speed. Indirectly these toll charges also convey information to drivers before entering the free/tolled section of roadway: in particular, if the tolls are relatively high, the level is likely to indicate that traffic is very heavy on the free lane ways (Brownstone and Small 2005, p. 281).

The purpose of Minnesota's MnPASS is to "maintain traffic flow and alleviate congestion. If you drive alone and want a more convenient and predictable trip, open a MnPASS account…" (MnPASS, n.d.). In this case it is evident that the pricing seeks to deliver a consistent travel time and it does this by dynamically changing the charges in response to prevailing traffic levels. It is notable that the charge seeks to alleviate congestion – the congestion may not be eliminated but it may be interpreted that the traffic volumes will not be as such as to lead to significantly longer travel times.

The three examples cited here – the I-15, the MnPASS lanes and the I-680 – relate to lanes on roads. However, the "dynamic pricing" that is used in these cases can also be applied to separate (toll) roadways. The case study is an example of the close link between pricing for reliability and pricing to restrain congestion. Inevitably, a charge that is applied to a specific road or road segment has greater potential to deliver a given level of reliability than a zonal (cordon or area) charge, which apply to a grid of roads.

The essence of this case study, then, is the segregation of part of the road space for those who are prepared to pay for access to that road space; and that the charge that is applied is varied so as to maintain a certain travel speed/time across the road link.

Source: 680 Smart Lane (n.d.), "Frequently asked questions about the southbound I-680 Express Lane", *I-680 Express Lane* web site, *www.680smartlane.org/faq.html*; Brownstone and Small (2005); MnPASS (n.d.), "MnPASS. What is it? How does it work", *MnPASS* web site, *www.mnpass.org*.

Priced roads or road networks (such as in Central London) are not uncommon. However, such charges are either fixed throughout a period or are subject to pre-determined pricing levels.

In the case of the toll lanes allocated within the Interstate-15 road corridor, the tolled lanes are subject to "dynamic pricing". Prices are set within a lower and upper price envelope. However the pricing level that is applied at any given time is determined by the prevailing traffic volumes. Tolls are varied in 6-minute intervals and the toll level is set so as to attract the level of demand that is consistent with a constant speed. The project's primary benefit was the reliability of on-time arrival for users.

These examples illustrate that pricing can be used to moderate demand to achieve a given level of reliability – or, at least, removing unreliability arising from "excess" traffic. It still needs to be asked whether it is feasible to apply this pricing approach more generally across the network. The likely answer for the road network, with current technology, is "no". The costs of technology and other transaction costs involved in applying the system are likely to outweigh the benefits.[6] Nonetheless, the examples indicate that there can be situations where such pricing can be applied, albeit not on the wider network.

Box 6.3. Segragated rail freight lines (the Netherlands and Australia)

Rail freight movement has been impeded in accessing terminals. The interface between freight and passenger trains, in particular, leads to difficulties in planned train paths and actual difficulties in operating those trains, and this is manifested in train delays.

In The Netherlands and Australia, the governments have added railway capacity. However, in the specific examples here, the capacity has been allocated exclusively for freight traffic. As with the road freight "truck-only lanes" examples noted elsewhere, the segregation of freight trains from passenger trains is intended to enhance the operational performance of the network by consolidating trains with similar characteristics.

In 2007, the double-track 160 km Betuwe Route freight railway opened. The line provides a short route between the Port of Rotterdam and Germany and additional capacity. A key attribute of the investment is that the government-funded route is "…exclusively for freight trains, which for a logistics service provider or shipper means short waiting times, fewer stops and high availability, punctuality, and therefore reliability." (Rail Cargo n.d.).

There is a similar philosophy behind the 36 km publicly-funded Southern Sydney Freight Line, which allows freight trains to operate along the southern intercity corridor into Sydney, independently of passenger trains. Currently, freight trains are held outside of this area during commuter peak periods; this causes a bottleneck in freight flows. Freight train reliability can be maintained along the 1 000 km link between Sydney and Melbourne but service quality can be impeded due to passenger train disruptions in the Sydney area. The benefits arising from the removal of the curfew on freight trains along this corridor include improving freight train reliability and reducing delays to passenger services that arise from the conflict between the two different types of train (Australian Rail Track Corporation, n.d., p. 1).

The two infrastructure investment examples differ from other capacity enhancements in that they involve segregation of freight operations from passenger operations. In essence, the separation delivers a degree of robustness in the specific passenger and freight train services with passenger train performance impacting only on passenger trains and freight only on freight. As with the truck-only road laneways, the premise is that the physical separation of freight from passenger will deliver reliability benefits. The operational characteristics (train length, speed, acceleration and stopping patterns) of the two train types differ markedly and segregation is intended to deliver performance benefits by reducing the potential conflict in train speeds that trigger train delays.

Unlike the Truck-Only Toll (TOT) lanes, however, there is no explicit premium charge for users of the freight-only lines. That is, greater reliability is provided to all freight operators. As noted by Badcock (2007), at 1.15 per train kilometre, the access charge on the new route is also identical to the charge on the rest of the Dutch network.

As with truck-only lanes, freight-only railways are not a cure-all for congestion. There are potential reliability benefits arising from the additional capacity. Again, as with truck-only lanes. The additional capacity should still to be subject to standard financial and economic evaluations.

Source: Australian Rail Track Corporation; Badcock (2007); Rail Cargo Netherland.

While the principles of pricing the road network apply equally to the railway network, the nature of use of the railway network differs fundamentally from that of road networks. The explicit pricing of railway infrastructure has only been adopted in recent years and generally *infrastructure usage* pricing is often not applied to temper demand. It is true that railway network demand is constrained but typically this is done by actions of fiat and never done by price. The network provider's fiat decisions usually involve preferential access for passenger trains over freight trains. [7]

High standards of passenger train reliability often lead to curfews for freight train operations or very restricted numbers of train paths. The following Box 6.3 illustrates that authorities in the Netherlands and Australia have elected, instead, to construct separate freight railways (the Betuwe Route and the Southern Sydney Freight Line, respectively) rather than continue with what is often unreliable access over a railway where passenger trains receive (uncosted) priority and higher reliability.

Existing examples of truck-only facilities are available from the Netherlands, where two truck/bus-only lanes exist on 5.3 kilometre and 3.4 kilometre sections of the A16/A20 Rotterdam Ring Road. Evaluations results show that, under congested conditions, trucks and busses have an average time gain of 3.5 minutes (Rijkswaterstaat 2004). A similar approach is under consideration in USA, where “Truck-only Toll Lanes” have been proposed; the lanes are (upon payment of a toll) for the exclusive use of trucks.

In railways, the control over access and prioritisation of traffic can provide additional monetary levers on reliability. In particular, that control can be used to ensure that network users operate according to the time slot (train path) allocated to them. Options that are open to rail network providers include:

- **Performance regimes.** Train operators can be financially penalised for late operation that they are responsible for (as applied with the British “performance regimes”). On crowded railways, in particular, late running affects the reliability of other train operations. Thus, imposing a penalty on the operator provides an incentive for the operator to operate within its time window and to address the sources of unreliability[8]. Performance regimes may also be applied to the network provider to ensure they give adequate priority to third-party train operators. For instance, there are performance regimes between Amtrak (the national long-distance passenger operator in USA) and freight railway owner-managers of the track; freight railway infrastructure managers receive incentives and penalties for keeping Amtrak trains running to published timetables.[9]

- **Premium pathing.** In Australia, the national rail network provider sells “super premium” train paths to train operators – scheduled priority train paths. The intimation of this scheduled priority is that such trains will also receive priority over other trains in the event that delays occur. That is, faster train paths may also bundle in higher levels of reliability. As illustrated in Figure 6.4, to achieve the high reliability involves setting generous train path priorities to the premium services – which can involve building in a degree of (essentially) excess capacity – and the Train Controller actually giving real-time priority to the services. Unlike road networks, a railway can drop or discourage or lower the service quality on low-margin traffic, pricing off business if their demands on the network prove to be unprofitable.

There are examples of similar incentive schemes, albeit only indirectly monetary in nature. Operators can be operationally penalised, by rescheduling late running trains. For instance, the new Lötschberg Tunnel in Switzerland will be operating at up to 97% of capacity. As a consequence, a late-running train will have an immediate and rippling effect on the reliability of immediate *and* subsequent trains. To deal with this, late-running trains will be rescheduled: if a train is more than five to seven minutes late, the path will be lost and the train will have to wait for the next spare train path before it proceeds through the tunnel (International Railway Journal 2007).

6.3.2 Delivering differential standards of reliability

Network reliability can be improved by using infrastructure pricing to restrict demand, such as used with the Interstate-15 road, which delivers a given reliability level. Road network providers can also use pricing to deliver different levels of service although the pricing is not usually attuned directly to reliability. Toll roads paralleling public roads (for instance, the M6 Tollway shadowing the public M6 Motorway in Britain’s Midlands) provide opportunities for passenger and freight users to pay to (probably) receive higher reliability. This is a crude form of where pricing can be used to deliver differential reliability standards. Of course, such “crudeness” may nonetheless be optimal in that there may be very high transaction costs involved in catering for an additional (particularly higher) reliability standard.

Box 6.4. **Truck-only lanes/truck-only toll lanes (USA)**

Truck-Only Toll lanes have been, or are, being considered in several US States. A more extensive proposal is the development of truck-only lanes along the Interstate-70 as it passes east-west through a number of States. The truck-*only* laneway policy contrasts with the counter-policy that *restricts* truck use to (usually) one lane only of a roadway – this lane, itself, remaining a general user lane. See, for example where Florida has applied a "truck lane restriction" on parts of the Interstate-95 (Florida Department of Transportation Research 2008, n.p.). This practice has also been adopted on some roads in Alabama, Delaware, Georgia and New Jersey. A primary objective with truck-lane restrictions is safety, particularly the safety of passenger vehicles, which is sought through enforced allocation of trucks to specific lanes. A further objective is to improve operational performance of the network, particularly for *passenger* vehicles. It is not surprising, then, that it is acknowledged that a consequence of such restrictions is the *slowing* of truck journeys (see, in particular, Moses 2007, passim.).

Two significant options are currently under consideration across the USA. These are: Truck Only Lanes (TOL); and Truck Only Toll lanes (TOT)**.** By implication, in the case of the TOL, there would not be a direct user charge for the facility. This would make the facility more difficult to finance.

The objective of TOT lanes is to facilitate the financing of the lanes and to influence usage patterns, thereby influencing the level of reliability. Despite the research, a 2005 study of TOT in the US state of Georgia concluded that "Trip reliability is an important potential benefit of TOT lanes that cannot be readily measured" (SRTA 2005, Steering Committee Meeting of 25 March 2005.). A later, separate, Georgia Department of Transportation study identified reliable goods delivery times as a major performance objective of truck-only lanes. The study recommended constructing lanes but did not specify whether the lanes were TOL or TOT. Lanes were recommended for the I-75 north and south of Atlanta, the I-20 west of Atlanta and part of the I-285 to the west of Atlanta. The study warned that such lanes would require "major" public funding and that "alternative" funding could be considered but "could come with 'strings attached'". The study summary does not disclose reliability benefits (Georgia Department of Transportation 2008, passim.).

In 2007, the US Department of Transportation named six interstate routes as "Corridors of the Future" that would be used to develop "innovative" ways to deal with congestion, including construction of truck-only lanes. The studies of the I-15 and I-95 specifically address the issue of reliability enhancement. The segregation can influence freight reliability but this is further influenced by the application of variable tolls. In the case of Georgia's State Road & Tollway Authority (SRTA), the pricing strategy for the freight lanes would be "a cost per mile that will keep the TOT lanes performing at a level of service that provides more reliable travel" (SRTA 2005, p. v).

An issue arising from the analysis is whether or not freight hauliers would be forced to use the tolled laneways. The 2005 Atlanta study assumed that operators would not be obliged to use the lanes (SRTA 2005, p. 3). As noted by Saporta, there was disagreement over this. The truckers preferred the voluntary approach but potential infrastructure funders insist that trucks be obliged to use the lanes, in order to ensure that there was enough revenue to pay for the lanes (Saporta 2008). At present there are no operational examples of truck-only toll lanes. It is likely that retrofitting truck lanes to existing road will be very costly. This has been acknowledged by planners in Georgia. In any case, the assessment of the reliability merits of such initiatives needs to separate the benefits arising from supplementing capacity from the benefits arising from segregating freight movement from passenger movement.

The merits of truck-only lanes in optimising reliability need to be judged against the terms for accessing that capacity. Reliability cannot be influenced if the capacity is provided to freight users without direct user charges. If these lanes become heavily used, the absence of the pricing mechanism will prevent reliability standards being maintained. Truck-only lanes are not a cure-all for congestion. They provide reliability benefits because they provide additional capacity and provide freight users with the option of paying for a higher level of reliability.

Source: Florida Department of Transportation Research (2008); Georgia Department of Transportation (2008); Moses (2007); Saporta (2008); State Road & Tollway Authority (2005); US Department of Transportation (2007).

The technical efficiency of pricing roads or road networks depends on whether pricing can be undertaken in a way that satisfies network users' demand for reliability – and whether the costs that arise from applying the system are very modest relative to the benefits that are captured.

Box 6.5. **Offering reliable train services – UP/CSX express lane (USA)**

This case study illustrates a service product with differential reliability standards. With the "Express Lane" product, railways manage the train so that there is a high probability that freight arrives within a given delivery time window. The service product is priced and marketed as a premium service, backed by delivery time guarantees and refunds.

The Union Pacific/CSX "Express Lane" trains shift perishables goods across USA in conventional merchandise trains. Wagons are added to, and detached from, the trains. Implicitly, the additional handling of goods increases the potential for delays relative to unit trains.

The transit times of Express Lane relative to the Railex case study (Box 5.9) reflect this additional handling. However, the commonality of the services is that the railways provide a guaranteed maximum delivery time. The Railex trains guaranteed to take under 5 days and the Express Lane trains guaranteed to take between 5 and 10 days (depending on the eastern seaboard destination). On-board tracking and sensing equipment provides railways and shippers with information that facilitates the progress of the train and the logistics planning when the goods arrive at their destinations.

The central feature of this service product is that a premium is levied for the guarantee of service reliability; for their part, the railways are encouraged to honour their guarantees through monetary compensation for failing to provide a given service standard. Thus, shippers face a market in service standard reliability and the compensation mechanism provides railways with an incentive to provide that standard.

For their "Express Lane" product, Union Pacific and CSX guarantee the service reliability and they reimburse the shipper if the goods do not arrive at the destination on time. BNSF Railway offers a similar money-back guarantee on a range of its services and commodity groups. By way of example, the specific products "BNSF Premium" and "BNSF Expedited" are offered to shippers with a 15% premium on its standard freight rates.

The case study illustrates that a market in reliability is possible. This relies, however, on the ability to provide differential levels of reliability. It is possible to provide a degree of differential reliability. This arises where train operators can set train priorities (when they are infrastructure managers as well) or can pay for train prioritisation (when train operations are vertically-separated from infrastructure management).

The new wagon and intermodal priority service products are relatively new in USA. The development and expansion of these services indicates that there *is* a market for these higher standards but it has been suggested that the market is fairly limited and, indeed, that what is required is on-time delivery rather than money-back guarantees:

"The demand for such services, although improving, is not necessarily extensive. One rail user suggested that shippers do not want money-back guarantees, they want their freight where it's supposed to be, when its supposed to be there. Another indicated that the market for such express services is not yet well developed." (Progressive Railroading 2003).

However, while the money-back guarantee may thus be seen to be poor compensation for lack of reliability, the value of the guarantee lies in that it can provide a strong incentive for train operators to provide a reliable service – particularly because the shippers are paying for that quality.

Source: Entrepreneur.com (2007); Progressive Railroading (2003).

Because access to, and use of, railways is restricted, delivering (and pricing) differential reliability is more achievable. That said, it is worth remembering that it is unlikely to be economically possible for railways to deliver the *highest* standard of service (transit time and reliability). Railways may see their business as capturing the "middle ground" in reliability standards, offering a service range between fairly reliable/slow and usually reliable/fast. By way of example, while US railways capture significant volumes of business for the time-sensitive United Parcel Service business, they do not win it all. The company uses USA railways extensively but in 2007, for instance, it shifted some of its traffic onto roads for its "fast lane" service because the railway network could not provide a fast enough service.[10] It is conceivable that the railways could have met the standards required for this service, by prioritising the UPS trains. However this might have been achieved only at the expense of a considerable amount of track capacity, that is, by not running many other trains or by building a lot of additional capacity.

In this context, it is notable that railways do not seek to deliver the highest reliability standards. Bryan, *et al.*, conclude that USA railways do not seek to meet the highest standard of service quality:

> "…rail intermodal capacity seems to have gravitated toward the international container market, where service demands generally are less stringent [than high-priority domestic container movements]" (Bryan, et. al 2007, p. 7).

Such a stratification of business is common on railways:

> "Except for premium intermodal and some other operations, rail shipments may be slower, require longer lead times, and perform less reliably than trucks." (*Ibid.,* p. 208)

The fact that railways offer slower, less reliable services as well as faster, highly reliable services may, however, reflect shipper needs. Shippers will prefer more reliable services but they may not be prepared to pay for that reliability.

On heavily-trafficked railways it is likely that it would not be profitable to offer high-reliability services. Indeed, in the second half of the twentieth century, USA railways abandoned scheduled train services completely in order to improve their financial performance; trains would depart only when after reaching a certain train length. However, since the turn of this century, shippers have been increasingly willing to pay for greater certainty in deliveries (which has been demanded in the strongly growing intermodal rail services). This had led North American railways to embark on a widespread re-adoption of scheduled services for certain commodities.[11]

USA railways have also adapted their reliability standards in other ways to meet their customers' needs. In particular, they have tailored coal movements from mines to reflect power stations' needs. The railways undertake to deliver a given volume of coal within a given time window, rather than guarantee the arrival time of a given train.

Some integrated North American railway products offer examples where pricing has been applied directly to delivering differential reliability:

- The joint Union Pacific-CSX train product, the "Express Lane" is a fast freight train service for shippers wishing to move perishable goods between California/Pacific North West and the North East/Florida. The railways incur a penalty if the goods are delivered late. (See Box 6.5.)

- As discussed in the last section, Union Pacific and CSX, in conjunction with Railex, operate fresh produce trains between Washington State and New York State, guaranteeing delivery in just over 5 days. The service compares with the previous only option – a non-guaranteed time of 12-14 days. The guarantee provides farmers with a guarantee that the goods will arrive at the market without spoilage and still in marketable condition (see Box 6.5.).

- BNSF Railway offers a 100% on-time guarantee on its BNSF Premium and Expedited service products in return for a 15% premium on its rates.[12]

- Norfolk Southern and Union Pacific offer a Los Angeles-Atlanta premium service, Blue Streak, offering a range of reliability standards. The "Standard" product offers access to Blue Streak trains, subject to availability; the "Premium" product offers shippers priority train space, freight monitoring and improved cut-off and availability times. The "SuperFlyer" product offers guaranteed equipment availability and on-time delivery. If a SuperFlyer load is not available for pick-up at the scheduled time, the shipper is not charged for the freight movement.

A related area where differential standards of reliability can be provided is in mitigation against unreliability. As is discussed in the next section, those who place a high value on reliability (that is, are prepared to pay for greater reliability), may often have access to premium information services that are available on subscription. For instance, in finding their way across a network, users can use a simple paper road map or can pay to use GPS-based tracking systems to direct them. Similarly, it may be possible for network users to pay for "live" information systems that provide more useful up-to-date information than free-to-air systems. Thus, there may be limited opportunities for choosing differential levels of reliability but there may be a range of opportunities for choosing differential levels of reliability *mitigation* systems.

6.4 Conclusions

The principles and applications of pricing for transport reliability have been reviewed. When transport unreliability arises from traffic flows (largely from congested links) then pricing, in principle, can be used to moderate demand – necessarily depending on being able to restrict access. There are few real-world situations where access has been restricted and where this reliability-sensitive pricing has been undertaken. "Dynamic pricing" on the Interstate-15 is one of the few instances. Through the pricing system, motorists on that roadway should have a good chance of a reliable journey (to the extent that reliability can be delivered from capping traffic volumes). The system relies on pricing having an adequate "bite" on travel behaviour. If an upper limit is placed on prices then, ultimately, large increases in traffic could reduce reliability.

In principle, because of their control of access and pricing, railways are better placed to use pricing to deliver a consistent level of reliability. In practice, it is their control over access to the network – rather than price – that is used to constrain demand. This access fiat ensures that the network timetabling is not so ambitiously constructed as to severely undermine reliability. Nonetheless, there will always be a balance between adding extra trains onto the network and the effect that extra trains have on system reliability.

An example where pricing is used as a mechanism in railways to deliver a reliability standard arises from performance regimes and operational incentives. Performance regimes provide network users and providers with commercial incentives to manage the unreliability that they, themselves, have within their own control; such sources of unreliability impact on other users.

There are examples where differential levels of reliability on road network are being offered. A person paying for use of a toll road tends to expect lower journey times *and* higher reliability than on the parallel free public road. But opportunities for supplying toll roads are limited; offering differential reliability standards means providing both the high reliability and the low reliability networks and there will always be relatively limited opportunities to undertake this.

The need to recognise the heterogeneity in network users is a key issue in delivering optimality in reliability. As is illustrated by the energy supply sector, policy solutions need to recognise diversity in network users' desire for reliability. For electricity users, the cost of outages (and thus the value of reliability) varies very widely across users. A pricing solution using an average product and an average price will leave important users dissatisfied.

Transport networks become increasingly less reliable as capacity is taken up. Unless network users can be segregated in a cost-effective way, it is unlikely that diverse user needs will be met. That said, policy solutions to optimising reliability using charges are realistic only when it is cost-effective to separate these network user products. If there are significant costs involved in separating network users, then charging may not offer a viable policy solution.

NOTES

1. A similar example is the opening hours for receival and delivery at the port terminals in Brisbane, Australia. Importers/exporters complained about lengthy queues at the terminal. The stevedores responded by lengthening operating hours overnight but "there has been little take up of overnight pick up/pre-delivery with cargo owners still preferring to only use normal business hours". Australasian Freight Logistics (2009), pp. 10-11.

2. It is highly likely that a shipper may face only one mode. The figure shows the aggregate market spectrum, not the options open to each individual shipper.

3. See, for example, the Pier*PASS* web site, at www.pierpass.org/about_pierpass.

4. Or, at least, the infrastructure manager can offer differential reliability where that reliability is within the manager's control. Some sources of unreliability lie outside of the manager's complete control – such as the effect of bad weather, albeit that the manager has some ability to influence the level of vulnerability to such events.

5. As noted in Figure 6.4, the level of capacity utilisation influences the level of train punctuality.

6. However, that clearly does not preclude the use of other cost-effective methods to improve reliability.

7. This priority is often not reflected in the access charge. Note, also, that not all railways give priority to passenger trains, although passenger train operators may provide ancillary financial incentives to freight railway infrastructure managers in order to allow passenger trains to take priority. See, for instance, the arrangements between the USA long-distance passenger train operator, Amtrak, and freight railways, p. 15. See, also, the case study on proposed truck-only lanes in the USA, where, again, government fiat is intended to restrict trucks to specific road lanes, in order to improve passenger operational performance.

8. Where a railway is integrated and has no third-party operators on its network, the impact of a train delay is internalised within the company – there is no external impact of the delay and so there is no merit in applying any penalty system on the train operation.

9. BTRE (2003) observes "A punctuality-based performance regime exists between the USA passenger train operator, Amtrak, and freight infrastructure owners to the extent that freight owners are rewarded and penalised for variances from an agreed standard of time keeping." (p. 142)

10. Committee on Transportation and Infrastructure (2006).

11. For instance, in 1998, Canadian National adopted scheduled services; in 2003, Norfolk Southern progressively introduced its "Thoroughbred Operating Plan", which included adoption of scheduled services and real-time tracking of goods.

12. See Thuermer (2003).

KEY MESSAGES

- Demand for reliability varies between and within products and users and over time so prices can potentially provide an important avenue for communicating those preferences.
- Surface transport network providers do not generally identify reliability as a specific service attribute; there is neither a price applied to reliability nor, more generally, a market for reliability.
- The absence of a pricing system removes crucial guidance as to the appropriate network standards.
- Optimal reliability does not mean maximising reliability – even where it possible to achieve 100% reliability, the costs of doing so would outweigh the benefits.
- High transaction costs involved in differentiating reliability across users can also exceed the benefits – it follows that not all diverse user demands for reliability will be met.
- For road networks, the typical outcome is that the user faces only one reliability standard; a parallel toll road may provide the only alternative reliability standard
- Railways restrict users to the network and control movement on the network and so can provide a relatively consistent reliability standard to customers seeking a given service standard, although providing services with the highest reliability level is unlikely to be optimal.
- In practice, then, there are limited opportunities to use prices for supplying differential reliability standards that meet network user preferences.

7. INFORMING – MITIGATING THE IMPACTS OF UNRELIABILITY

This section focuses on the role for information in mitigating the negative impacts of unreliability. Travel time reliability depends on the user's expectation of travel time. This expectation can vary according to the information available. Developments in technology may make it increasingly cost-effective to supply up-to-the-minute information to network users. Traffic information can be used by network users to adapt their behaviour at short notice.

Infrastructure managers can facilitate network usage and reduce the impact of unreliability by informing users of prevailing conditions. The information does not stop an incident happening but rather reduces the costs that arise from the incident. Hence, information can mitigate the unreliability and reduce the "ripple effect" or "snowballing" consequences of unreliability.

7.1 The role of information

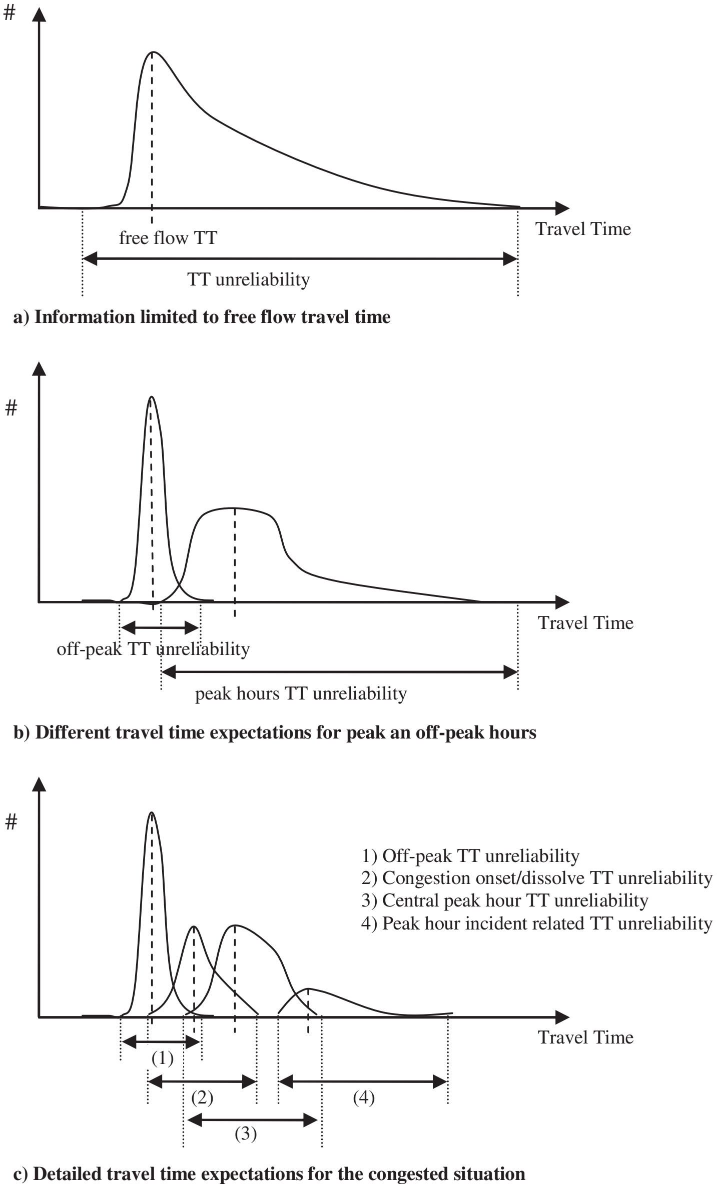
Information plays an important role in managing reliability. A crucial issue in travel time is whether the network user faces recurrent, or non-recurrent, delays. If delays are recurrent and predictable, then network users can alleviate the worst effects of those delays by adjusting their travel plans to accommodate the delays. An irregular network user would not be familiar with these recurrent delays *unless* informed of them.

The provision of information enables network user to manage their travel behaviour (such as when to travel and which routes to travel on) and to mitigate against the consequences of late arrival at the destination. For instance, a delay on a route may be unavoidable if there are no alternatives to the route or if it is too late to alter routeing around the delay. In this instance, the role of information may at the least ease the user's mind about the extent of the delay; the information also enables the user to advise parties at the destination that a late arrival should be expected. Such information can therefore facilitate rescheduling at the destination.

Figure 7.1 illustrates the role of information. The more information provided to the user (in this case a car driver), the closer is the *expected* travel time to the real travel time and, as a result, the higher the reliability of the trip travel time. If infrastructure managers can communicate the travel-time variability at different times of the day, unreliability can be reduced.

Where the infrastructure manager has not given the driver any travel time information, it is probable that the user's knowledge is limited to the free flow travel time, which would be inferred from the *expected average speed* for the type of road and the distance. In this situation travel time reliability is very low, if travelling in an area subject to congestion. The poor reliability results from a very wide travel time frequency distribution with a high range of possible travel times. If the driver is informed about travel times in peak and non-peak hours, and the associated expectations of delay (*i.e.* knowledge of recurrent delays), then the travel time unreliability is reduced. Finally, drivers with good knowledge of the traffic conditions likely to be encountered on a trip (a daily commuter or a user of very accurate travel time information for a freeway), can predict travel time fairly reliably.

Figure 7.1. **Variation of travel time unreliability in relation to the drivers' informationin a multilane freeway**



Source: Soriguera and Robusté (2008).

In this context the relationship between travel time variability and reliability is not direct, as a road with heavy peak hour traffic with large travel time variations according to time of day, could also be reliable if accurate information is provided to the driver and there is an efficient incident management system. In contrast, a road subject to less travel time variability could be very unreliable for a driver provided with no information.

Box 7.1. **Value of information**

There is abundant evidence that travellers place a high value on travel time information. Airline pilots generally provide information to passengers on predicted arrival time and any expected delay. In a study of motorists' preferences, Harder *et al.* (2005) found that travellers are willing to pay up to $1.00 per trip for convenient and accurate travel-time predictions, such as when traffic is delayed and alternative routes would be faster. The value of this information is higher for commuting, special event trips, and when there is heavy congestion.

Studies of the welfare impacts of providing different types of information to road users have found that (under certain assumptions) provision of both perfect and imperfect information leads to a strict Pareto improvement (see for example Emmerink *et al*, 1998). The results of these studies suggest that it will always be beneficial (from the welfare economic consideration) to provide any information, even if incomplete, rather than no information at all.

Thus, travel time information is a key element in managing reliability. The fundamental variable to provide network users with is traffic time information. Several surveys have confirmed that travel time is the single most valuable piece of information from the user point of view, allowing the user to decide in advance the best time to start a trip and the best routing option, or to modify this initial planning once on route.

7.2 Travel time information dissemination

7.2.1 Dissemination techniques

Dissemination techniques can be divided into two main types: pre-trip information and on-trip information. Both forms of information can reduce the ripple/snowballing cost consequences of likely delays. However, pre-trip information allows trip planning (to possibly avoid the unreliability entirely) while on-trip information enables network users to (possibly) modify the initial planning according to current traffic conditions. Users will be able to optimise their routing behaviour and their travel time management on the spot or just-in-time while information on rerouting will enable users to keep to their travel schedule more easily.

To be effective, traffic information must be short, concise, quantified and specifically addressed to the network user. Travel time information itself fulfils the three first conditions, and the dissemination technology must fulfil the last one. That is, the travel time information to be conveyed must be that of specific interest to the driver.

There are several different techniques for travel time information dissemination, each one related to a particular technology. Their detailed characteristics are shown in Table 7.1.

Funding information infrastructure is a challenge, whether state-funded or recovered from network users. One difficulty for managers is how to integrate new communication technologies as they emerge while retaining sufficient homogeneity to allow them to be maintained (Janin 2003). Traffic information is often organised by the state for national roads, by motorway concessionaires on toll motorways, and counties and towns for their local road networks. Information equipment is generally funded by the various authorities responsible for their own networks. The question of funding becomes more complex for systems seeking to provide integrated transport information for both road and rail services which concern the overall infrastructure in a geographical area. There is then no "natural" funder and specific organisations have to be put in place for these new services.

Table 7.1. **Travel time dissemination techniques**

Techniques	Characteristics
Press	• Only pre-trip information • No user discrimination • Discrete information times
Internet	• Pre-trip information (on-trip using mobile phones) • It is a service on demand. User must log-in and ask for a specific itinerary
TV Broadcasts	• Only pre-trip information • No user discrimination • Discrete information times
Information Points	• Capable of disseminating all types of information and discriminate between users. However its accessibility is very low because the driver must stop at the service area to obtain the information.
Radio Broadcasts	• Traffic information bulletins • Capable of disseminating pre-trip and on-trip information • No user discrimination. Each driver must carefully listen to the whole bulletin and select his own information of interest. • Discrete information times, subject to the scheduled bulletins. • In case of short range dedicated radio signal, these last two limitations can be overcome.
Traffic Call Center	• Capable of disseminating pre-trip and on-trip information, with the limitation of on-trip telephone calls • It is a service on demand. Driver must ask for it and usually pay a price for it. This implies a limitation of access to the information.
Variable Message Signs	• Capable of disseminating pre-trip and on-trip information • Specifically addressed to the driver, as only inform the drivers who travel below them. • Continuous and very accessible information
Car Navigator (RDS-TMC radio signal)	• Capable of disseminating pre-trip and on-trip information • Specifically addressed to the driver (GPS/GSM/UMTS) • Continuous and very accessible information
Cellular Phone (text service)	• Equivalent to a call centre with the improvement that you can subscribe to a particular corridor and receive information without asking every time for it.

Source: Soriguera and Robusté (2008). Note, that the GPS (global positioning system) is used to provide information on positioning and routing, RDS (radio data services) is used to transmit real-time traffic information, while cell phone networks is used to transmit real-time traffic information and to calculate traffic condition while providing real time traffic data.

Network users, for their part, have grown accustomed to considering that information should be provided free of direct charge. The need for traffic information is considerable, but users often do not see why they should have to pay for access to information. Nonetheless, for those who place a high value on reliability (that is, are prepared to pay for greater reliability), there are often premium services available on subscription. For instance, in finding their way across a network, users can use a simple paper road map or can pay to use GPS-based tracking systems to direct them. That is, there can be a price for reliability mitigation.

The range of policy options quoted above require tools to monitor reliability and to inform operators and users in real time with the precise location of events. With funding and space for large-scale infrastructure construction becoming increasingly scarce, governments, infrastructure operators and public authorities *may* find that new technology such as Intelligent Transport Systems and Services (ITS) offer cost-effective systems. Such technologies include ramp metering systems, traffic and incident detection and variable message sign systems. Such investments should, of course, be subject to cost-benefit analysis.

It is important to note that ITS technologies may be applied to a range of objectives: easing congestion, providing data to support management and pricing systems, improving safety and reliability. Some technologies are more dedicated to congestion (*e.g.* ramp metering), others to reliability (*e.g.* incident detection), others are clearly multi-objective (*e.g.* variable message sign systems, or video-surveillance). Benefits may also be numerous. Road users may benefit from higher transport safety, optimised traffic flow and shorter and more reliable travel times. The national economy may benefit from less accident and congestion related costs. Surface network administrations may benefit from improved maintenance planning, more efficient network use and consecutively from postponed costs related to construction and reconstruction.

7.2.2 Information before departure

Pre-trip travel-time information allows the user to decide the time of travel, the mode of travel, or even cancel the trip all together. Pre-trip information can reduce the risk of delivering goods late or arriving late at a destination.

There are many options available to disseminate pre-trip information. Traditionally, newspaper or radio has been a source of traffic information, especially in case of an already-scheduled event such as a sporting event or a festival. These types of information basically warn network users of the possible delays and provide information on the extent of the disturbance on the network. Even though these types of traditional measures might be considered as too static, they can provide valuable information to the users of the network and mitigate possible unreliability impacts, if correctly targeted (see, for example, Box 7.2).

Traffic management and information services are an important strategy in many countries. There are many examples available of this type of service, especially related to pre-trip information. Most of them are currently internet-based services while some also provide up-to-date information to mobile phones. Some service companies offer the calculation of journey times or provide travel information with added value. The first websites grew up in the mid 1990s. Most of them originally provided traffic information for a certain region but have since been extended to cover whole networks. These websites initially targeted the general public but have subsequently offered tailored systems for targeted network users with specific reliability needs.

Box 7.2. **Travel warning for the Grand National race**

Racegoers are being urged to use public transport to get to Aintree racecourse as the 161st Grand National festival gets under way in Liverpool.

About 150 000 people are expected to attend the three-day event and police are warning the area will become heavily congested.

People are being urged to use trains, taxis or local park-and-ride services to avoid traffic delays.

The Grand National steeplechase takes place on Saturday, on the final day.

Chief Superintendant Mark Matthews said: "Merseyside Police would encourage race goers to travel to the event by public transport where possible to avoid congestion and parking problems.

"I would like to thank the public in anticipation of their cooperation for what promises to be a memorable and enjoyable event."

Source: BBC News.

Bison Futé, for example, is a French website directly managed by the ministry of transport. Bison Futé was founded in the 1970s during a period of severe traffic jams on trunk holiday route. Originally it operated through more traditional media (television, radio and newspapers) but now has a very well used website.

Box 7.3. **Bison Futé (France)**

Bison Futé is operated by the ministry of transport in France. It is an interactive website that gives information about current traffic conditions in France, provides safety advice, information about road works and weather. Information about traffic conditions for the week-end are given each Thursday and Friday evening after the news on the TV at 8h40 PM, when the audience is very large.

Source: *www.bison-fute.equipement.gouv.fr.*

Access to lineside and roadside cameras can provide would-be network users with information on network performance, including reliability. Many road administrations use cameras to facilitate traffic management. Network operators are often able to swivel these cameras, to monitor and identify incidents or congestion. In recent years there has been a growing demand to view images from traffic cameras by different parties, such as travel information businesses and more importantly general public. The cameras may also provide information regarding weather conditions. Being able to consult weather conditions on a certain part of network may be very useful in order to assure on-time arrival at the destination. Box 7.4 provides an example from Finland, where weather can have a strong influence on estimated travel times (especially during the winter).

Box 7.4. **Weather camera provided by the Finnish Transport Agency**

The following web image illustrates how the output from a roadside camera is used by the Finnish Transport Agency to inform road users on the impact of current weather on road driving conditions. The website also provides information on the air and road temperature and other important information regarding road conditions.

Photo. **Webcam capture**

Source: *www.liikennevirasto.fi.*

In the British example in Box 7.5, the Highway Agency website uses an interactive map to display planned events and real-time traffic conditions. Icons pinpoint incidents and road works. Information posted on the variable message signs can be also viewed from the website (see more on next chapter on on-trip mitigation). By ticking the "Future Conditions" box, the web site shows all known *planned* events. Constantly updated, real-time traffic conditions are shown through color-codes, with green or blue showing when traffic is flowing freely on roads and motorways, yellow for delays of between 15 and 30 minutes and purple for delays of 30 minutes or more.

Box 7.5. **Traffic England**

Traffic England is a new service from the Highways Agency that provides real-time traffic information from the National Traffic Control Centre. This real-time traffic service is also available as a downloadable desktop application and as traffic radio news service delivered live to compatible mobile phone or PDA.

Source: *www.highways.gov.uk/traffic/traffic.aspx.*

In 1999, the U.S. Department of Transportation (USDOT) petitioned the Federal Communications Commission (FCC) to designate a nationwide three-digit telephone number for traveller information. This petition was formally supported by 17 State DOTs, 32 transit operators, and 23 Metropolitan Planning Organisations and local agencies. In 2000, the Federal Communications Commission designated “511” as the single traffic information telephone number to be made available to states and local jurisdictions across the country. An interesting point here is that the number is national but information is local.

Box 7.6. **US Call 511 traffic information**

The US Call 511 phone-based traffic information system uses one easy-to-remember number, regardless of the traveller’s location. The system gives travellers’ choices – choice of time, choice of mode of transportation, choice of route.

There are no Federal requirements and no mandated way to pay for 511; however, USDOT and FCC expect to see nationwide deployment. In 2005, the FCC reviewed progress in implementing 511.

While the flexibility provided in the FCC ruling is highly desirable, it also presents a challenge. If not thoughtfully planned, 511 services could devolve into an inconsistent set of services widely varying in type, quality and cost. There is a great deal of interest in using 511 throughout the U.S. It is expected that there will be multiple requests for 511, at least in some parts of the U.S., from State DOTs, transit agencies, regional and local transportation agencies, as well as private service providers who will offer to implement 511 services for some sort of compensation

Source: www.fhwa.dot.gov/trafficinfo/511.htm.

Information tools are also available for train passengers. Live train arrival and departure updates by station are helpful in planning a journey. The VR Passenger Services in Finland provides a website where passengers can view departures and arrivals at or from selected stations. User can also view train punctuality data by train, comparing train’s arrival and departure time at the stations to that of timetable. The site also includes info on the time which the train left or arrived at the station and an estimated arrival time on next station. Additional information is provided in case of a cancellation or delay of the train. The same data is also accessible by mobile phone. A similar example is provided in Melbourne, where the passenger train operator sends automatic text messages to a patron when services on a specified line are delayed (or are subject to disruption) by more than 15 minutes[1].

Box 7.7. **Live train updates by the Finnish Rail**

The Finnish rail company VR offers up-to-date information for each train regarding their scheduled departure and arrival times, actual observed times as well as causes for possible delays. For passengers, this information may be useful when planning the trip or for those going to pick up a passenger from the railway station, reducing extra buffer time.

Source: *www.vr.fi*.

A specific internet tool to tackle unreliability is provided by the Washington Department of Transportation. It uses travel time data to provide a reasonable approximation of the “worst case” travel time scenario. That is, the tool provides users with the likely buffer time required if the user wants a high degree of confidence that they will arrive at the destination “on time”. The web calculator shows how long it will take to make the trip and when the user needs to leave the origin to arrive at destination on time in 19 cases out of 20 (that is, with 95 per cent probability). The calculator does not use real-time data; it uses travel time data from year 2006 and cannot be considered as fully accurate in terms of reliability measures. However, it is an interesting example of information provided for the public.

Box. 7.8. **Washington Department of Transportation webpage calculator**

The Washington Department of Transportation has a webpage that calculates the 95th percentile travel time for selected routes mainly in urban areas, and determines what time a person must leave on a trip from some origin to some destination in order to arrive on time 19 out of 20 times, or with 95% reliability. The user needs to enter information on journey origin, destination, and preferred arrival time.

The calculator presents the journey time and the time of departure needed to arrive at the destination with a reasonable approximation of the worst case travel time scenario.

Source: *www.wsdot.wa.gov/Traffic/Seattle/TravelTimes/reliability/.*

To sum, pre-trip information enables network users to improve their reliability; it also enables them to mitigate against the adverse effects of late arrival at the destination. Information on real-time performance of the network, incidents, weather and possible delays help in planning the trip. It should be acknowledged that quite often the same pre-trip information is now available on-trip through, for example, mobile phones. Thus, the distinction between pre-trip and on-trip information has blurred in recent years.

7.2.3 Information en route

Using on-trip information may mitigate undesired impacts of network incidents. Depending on the information, users may decide to change their route, if an alternative is available. Users may also reduce the impact of arriving late by rescheduling their deliveries or planned destination activities and hence reduce the ripple or snowball effect on that activity and subsequent activities. Even where the user cannot take alternative en-route or destination actions, just passing on the information of being late can reduce the stress related to not knowing how long the possible delay may last.

Roadside information signs are increasingly used to provide real-time information to road network users. Electronic information signs are now a familiar sight across the world on motorways and trunk road networks. These signs provide advance warning to drivers of emergencies, incidents and road works. Variable Message Signs (VMS) is a term often used to describe these signs.

The main purpose of VMS is to communicate information and advice to drivers about emergencies, incidents and network management, aimed at improving safety and minimising the impact of congestion. Messages displayed on VMS are often limited to those that help drivers complete their journey safely and efficiently. There are a number of types of VMS in use around the world and they provide the capability to display a wide range of warnings, messages and other traffic information.

Box 7.9. **The UK variable messages signs**

The UK Highway Agency provides travel and delay times on roadside Variable Message Signs. The service is available on the majority of motorways and some major A-roads across England. The messages advise drivers about the road conditions ahead, therefore allowing them to make informed decisions during their journeys.

Customer research on the trial of VMS ahead of general introduction found that 89% of respondents thought travel and delay time messages on motorways across England were a good idea. Over half of respondents said they would consider taking action, such as changing route, if they saw a delay time message. The service compares historic data for a route with current traffic conditions to set estimated travel time messages. These travel times are recalculated every five minutes and the messages updated. If a traffic event, such as a collision, causes delays on the network above an agreed threshold the message will automatically switch from the travel time message to an estimated delay message. The service is operated by the National Traffic Control Centre.

Source: *www.highways.gov.uk/knowledge/knowledge.aspx.*

Some applications currently are provided on a commercial basis. Many in-car navigator models, for example, already provide real-time information on incidents, weather, and traffic to car navigators. They calculate estimated travel times and take into account incidents in order to improve the estimated travel time. However, they are often limited in their capability to take into account changes in the traffic conditions due to new information in real-time. Cell phone networks can be used to collect data about current traffic conditions. A GPS receiver is the basis for all personal navigation systems. Each unit equipped with GPS-receiver and GSM/UMTS can be used as a sensor providing real time traffic data. Systems exploiting this opportunity are being introduced by various navigation service providers.

Box 7.10. **"Dash" – a car navigation system incorporating real time information**

"Dash" is an in-car navigation system that presents up to three different routes to a destination, and uses traffic information to calculate Estimated Time of Arrival (ETA) for each route. The traffic-based arrival times are expected to be more accurate, helping users decide which route is best. Even after a route is selected, Dash automatically alerts users when traffic conditions change significantly. Each Dash in-vehicle monitor anonymously and automatically sends its position and speed back to servers at the Dash Network Operations Center. The Dash servers then update all other Dash devices in the area with current road speeds. The system utilises real-time route information sent automatically back to Dash's central servers by each Dash user's equipment. Then the central system sends specific route and traffic information back to individual users so that they can benefit from the experience of fellow Dash users ahead of them. Similar systems are being introduced by other providers such as TomTom and Navigon.

Of course, in this case, the benefit offered by the aggregation of traffic data is only as good as the total amount of information being sent back to Dash. Thus, it will clearly depend on a critical mass of users in order to work as advertised.

Source: *www.dash.net*.

Many public transport systems now offer real time information. See, for example, Box 7.11, which describes two of the systems in use in London. This information provides many benefits (Turnbull and Pratt, 2003). It reduces waiting stress and allows passengers to better use their time and coordinate activities. For example, if a passenger knows when the next bus will arrive they can decide whether there is sufficient time to stop at a nearby store to make a quick purchase, when they are likely to arrive at their destination, and whether they should use an alternative mode, such as calling a taxi. In situations with multi-route options, passengers use the information for enroute travel decisions. Customer response to this innovation has been positive.

Box 7.11. **Information reduces stress – train and bus travel in London**

In 1984, signs providing real-time information on the status of London Underground service were tested at several platforms on the Northern Line (Turnbull and Pratt, 2003). Passenger surveys indicated a small, but significant, stress reduction in response to the information system. Passengers both with and without access to the information tended to overestimate actual wait times for trains. However, when passengers are given the information, the over-estimation was reduced by 0.68 minutes on average. The platform signs gave order of arrival information for the next three trains, route and terminal destination as needed, and the number of minutes before expected arrival. Passenger response to the system was very favourable: 95% of respondents indicated it was useful and 65% reported it helped reduce uncertainty in waiting for a train.

In more recent years, Transport for London has extended the information service to buses – traditionally one of the most difficult areas for public transport arrival time uncertainty and delays. The "Countdown" system is a real-time bus arrival estimation system, using GPS technology to detect each bus's location relative to each bus stop. The electronic display at the bus stop indicates the likely number of minutes away a bus is from the stop. The Countdown signs list the order in which buses will reach the stop, their destinations, and the number of minutes to arrival. Information on traffic and safety conditions can also be displayed. Visual observations indicate that 90% of passengers at the equipped stops looked at the sign at least once during their wait time. Average perceived wait time declined from 11.9 minutes before the trial to 8.6 minutes with the Countdown system, although there was no actual change in bus frequencies. 83% of respondents agreed that "if you know the bus is coming, time seems to pass more quickly" and 89% agreed that the signs made the wait time more acceptable. Respondents expressed a slight willingness to pay higher fares for the system.

Source: Turnbull and Bratt (2003);
www.tfl.gov.uk/corporate/projectsandschemes/technologyandequipment/7204.aspx.

7.3 Conclusions

This section presented a few case studies on information technology used in mitigating the negative impact of unreliability. It is argued that providing pre-trip and on-trip information can be a cost-effective way of improving reliability and reducing unreliability-related costs. In particular, information helps in rescheduling tasks and reduces the snowballing disruptions of schedules that otherwise might result from unreliability.

However, funding information infrastructure is a challenge and needs to be subject to cost-benefit analysis. In the meantime, numerous privately-available systems are being developed and those road users who find high value in information and reducing unreliability can purchase information.

Even though network users may be able to mitigate negative impacts of unreliability, the role of network operators is an important one. Travel time information is not only useful to the driver, but also to the road system operator as it is basic operational management and planning of the network. Travel time forecasting allows the operator to plan for responding to incidents and operational problems while real time information provides effective monitoring of the evolution of incidents.

Network operators need to consider the cost-effectiveness of providing pre-trip and on-trip information on travel time variation. The diffusion of expected travel times (from travel time patterns) is the first step, but in addition to mean value, variance in travel time should also be provided. This allows drivers to accommodate safety margin times and reduces anxiety levels. Even though some information is not possible to be provided in advance (*e.g.* accidents), information regarding the incident once it happens still reduces the cost of unreliability.

NOTE

1. See *www.metlinkmelbourne.com.au/using-public-transport/sms-services#1*.

KEY MESSAGES

- Informing users of travel times and variability is a core policy option for managing reliability. The information may enable users to set appropriate buffer times for travel and can facilitate mitigation of the adverse consequences of delays.

- The more the information provided to network users, the closer is the expected travel time to the real travel time and the higher the reliability of service provided.

- The network user can use pre-trip information to assist in deciding the time of travel, the mode of travel, or even to cancel the trip all together. Pre-trip information reduces the risk of delivering goods late or arriving late at the destination in general.

- On-trip information mitigates undesired impacts. Depending on the information provided, network users may decide to change their route, or reduce the impact of arriving late by rescheduling their deliveries or planned activities. Even where no mitigating action can be taken, information can reduce the stress of d to not knowing how long the delay may last and can be used to reduce the downstream impact of delay.

8. CONCLUSIONS

Why focus on reliability?

Transport trends suggest there is a need for greater policy focus on reliability:

- Transport unreliability impacts on personal and commercial activities – the way personal and business activities are organised is increasingly dependent on robust schedules.
- In recent years, economies have captured significant gains from activities based on centralised and specialised production, centralised storage and longer distribution lines – but these gains are underpinned by resilient schedules.

Changes in disposable income, in leisure time and in the geographical pattern of commercial activities have led to greater expectations for reliable transport networks. There have been great productivity gains from commercial restructuring but it leaves businesses generally more vulnerable to disruptions. The cost of these disruptions tends to be higher for a system built on the premise of a high level of reliability.

It is important to be able to distinguish between congestion and reliability. Congestion tends to increases the likelihood of unreliable service but it can affect travel time in predictable patterns and on permanently congested routes journey times can be lengthened but show little variability. Addressing congestion can therefore fail to improve reliability and generally a wider range of policy tools exists for addressing reliability.

Thus, reliability needs policy prominence such as is traditionally given to congestion.

The policy challenge

A key policy challenge is to create incentive structures that encourage selection of the most cost-effective reliability option – that is, adopting the option that delivers a given level of reliability improvement for the lowest cost. The objective is to ensure that option is chosen ahead of the less-effective options, regardless of whether the responsibility for adopting the option lies with the network provider or the network user.

The challenge for policy makers arises in two areas. The first is in formulating the institutional arrangements that affect the market for reliability. For instance, legal frameworks that prevent discrimination between transport system users can create impediments to differentiating between services on the basis of reliability. The second is in the treatment of reliability when assessing publicly-funded transport infrastructure projects. The role of the government is two-fold: encouraging a market for reliability and incorporating reliability into the assessment of transport infrastructure projects.

Determining *appropriate* levels of reliability is a major issue. This is because there is generally no priced market for network users to choose levels of reliability according to their needs. It is usually difficult to provide differential network quality to users who would be prepared to pay more for a higher reliability standard. Network users generally take some action to cushion themselves against system unreliability, and it may be more cost-effective for the users to reduce their exposure to unreliable situations, or to recast their activities, than to invest in expanding capacity or otherwise making the network more resilient. This is especially the case when users do not pay directly for reliability.

Government is the predominant road and rail network provider and setting appropriate reliability standards is therefore implicitly or explicitly a policy issue – indeed, a policy challenge. Governments face a formidable task in ascertaining appropriate network standards. Where users do not pay directly for those standard, users will be tempted to call upon government to deliver the highest network quality even when the most effective/lowest cost option to enhance reliability might be for users to modify their own behaviour.

For road networks, in particular, it is common for only one reliability standard to be provided. There are only rare examples of road sections where network users are given the option of a higher reliability standard through variable tolling or where a public road is adjacent to a tolled road that provides drivers with different reliability levels (to the extent that the travel-time variability arises from traffic volumes rather than non-recurrent events).

Because these approaches are rare, road infrastructure providers in particular need to use cost-benefit analysis techniques to estimate a prevailing reliability standard. In the absence of direct pricing information, network providers rely upon cost-benefit analysis to put together the case for given levels of infrastructure quality.

Historically, cost-benefit analysis has not explicitly measured reliability. Given trends in commerce and personal travel, there is a case for greater focus on teasing out the value of reliability options within cost-benefit analyses.

Valuing reliability needs to recognise diversity among users

The value different network users place on reliability varies greatly. Thus incorporating diversity, or "granularity", in user valuations of reliability is important when cost-benefit analysis is to be used assess reliability. Put another way, using a cost-benefit analysis with only one defined user group (that is, applying a single value of time) will ignore the diversity in network user valuations can introduce large errors in calculating user benefits.

The value of reliable networks is location, time and user-specific and is not generally transferable across situations. The large variation in the value users attach to reliability make it unwise to try and factor reliability into cost benefit assessments by applying a general uplift factor to average time savings. This can seriously over or under estimate benefits. Values for reliability should only be transferred across situations that are sufficiently similar in terms of users, patterns of use, levels of congestion, etc.

For incorporation in cost-benefit reliability needs to be expressed in suitable units (in more technical terms, estimates of the travel time-equivalence for unreliability are required). Most are based on the standard deviation of journey time around the mean travel time. One example is buffer time, the time needed to be added to the journey to be sure of arriving on time to some degree of probability.

Monitoring reliability

The monitoring of reliability is an essential tool in guiding policy. In the first instance, of course, there is a need to define reliability. This report defines reliability as the ability of the transport system to provide the expected level of service quality on which users have organised their activities.

There is a clear dichotomy in performance indicators: indicators that measure quality for network *providers* (what level of service is planned and delivered, how robust the system is to disruption); and indicators that reveal what the *user* experiences (or how they respond to network experiences). It is extremely important to look at both network and user perspectives, as each has different policy implications. Presenting both indicators will also facilitate policy discussion between users, operators and decision makers.

The monitoring of reliability is useful in two important ways. First, it is relevant for network users to help them plan travel and help mitigate the worst effects of unreliable performance. It is also useful for policy guidance. In recent years, for instance, public transport authorities have introduced a number of performance statistics, outlining how well services perform. Governments can use "target" quality indicators to measure service providers' performance and to "encourage" improvements. However, it must be stressed that such targets need to be set carefully, to ensure that network providers do not distort their behaviour simply in order to meet targets (for example reducing the number of trains operated on a route might improve reliability at the cost of a much poorer service overall).

Four main policy instruments

Unreliability arises from multiple sources each requiring different ways to manage the problem. Unreliability of the transport infrastructure network arises from two primary sources:

- Unpredictable demand-related traffic interactions between users (congestion).
- Unanticipated supply-related events:
  - Traffic incidents (crashes and vehicle break-downs).
  - Natural events (*e.g*. floods and earthquakes).
  - Network maintenance (causing temporary reduction in supply).
  - Mismanagement in infrastructure supply, which can also include inappropriate maintenance programs.

There are many techniques and instruments available that can be used to improve the reliability of the transport network – both individually and in combination with each other. Four principal policy options available to manage reliability can be identified:

- Physical expansion of capacity.
- Better management of capacity.
- Pricing mechanisms to deliver a market for reliability.
- Information systems intended to mitigate the adverse consequences of unreliability (*i.e.* reduce its costs), rather than to reduce the incidence of the unreliability.

In principle, supplying more capacity can improve reliability, particularly when unreliability arises from high traffic levels. Additional capacity also means less vulnerability of network in case of incidents if alternative links are provided. However, capacity expansion is not necessarily the most cost-effective approach to optimising reliability and there are clear limits to the build-only option.

Infrastructure can also be built at standards that reduce the need for maintenance or improve the robustness of the capacity. For instance, capacity provision that involves low-maintenance, long-lived materials, can reduce non-recurring congestion that is associated with infrastructure degradation and repair. The prospects of greater weather extremes have implications for infrastructure provision and maintenance. Infrastructure should be planned, designed and maintained in a way that ensures the appropriate level of resilience.

The gains from better management of networks and services can be far more cost effective in delivering higher reliability – especially relative to capacity expansion. Pro-active management of network through integration of robustness objectives into the network planning models enables assessment of network vulnerability.

Active network management, involving intensified oversight of network use, allows faster reactions to disruptions, hence, increasing reliability. Better active management of incidents can provide significant reliability benefits. An example of active management is the use of motorway hard shoulder capacity. The use of incident management systems, in turn, enables transport service providers to react faster to disruptions thus reducing the duration and severity of the events. A focus on interfaces, such as border crossings and ports and hinterland connections where unreliability is likely to occur, might also be appropriate.

Charging directly for reliability by setting differential charges for infrastructure use and service supply, according to the level of reliability, might deliver an appropriate level of reliability. In recent times, dynamically-priced roads in the USA have been introduced; variable pricing of roads can be used to change traffic levels which, in turn, impacts on reliability (to the extent that changing the traffic levels influences reliability).

However, because of generally open-access roads and low financial returns from offering multiple reliability levels, there are few situations where price can be used to deliver more than the most basic differential reliability levels. Railways are better placed to use charging as a tool to deliver a consistent level of reliability because full control of access to the network allows network-link charges.

Providing pre-trip and on-trip information can be a cost-effective way of improving reliability and reducing unreliability-related costs. In particular, information helps in rescheduling tasks and reduces the snowballing disruptions of schedules that otherwise might result from unreliability.

Information may be used in different ways to improve reliability depending on whether a traveller has left the origin, whether a traveller can divert to another route, or if the traveller cannot divert but can reduce the ripple effect (consequences). Different tools exist for delivering this information, including variable message signs, car navigators, the internet, and text messages on mobile phones.

Travel time reliability depends, to some extent, on the user’s expectation of predictable travel times; this expectation can vary according to the information available. Network providers can facilitate network usage and reduce the impact of incidents by informing users of prevailing conditions. Even if information does not prevent incidents from happening, it can reduce the costs that arise from the incident.

A key policy issue that links these four policy options is ensuring the application of the most cost-effective options first, regardless of whether responsibility lies with the network provider or the network user. Cost-benefit analysis is the central process to achieving this objective. This report has suggested ways in which such analyses can be modified to incorporate reliability.

Assessing Reliability

Cost-benefit assessment has so far been applied to projects designed to improve reliability in only a small number of countries, with techniques that have been in some important respects unsatisfactory.

This report makes significant progress in identifying appropriate methodology for incorporating values for reliability into project and policy evaluations, and the pitfalls to be avoided.

Incorporating reliability into project assessment requires splitting time savings into pure journey time improvement (average time savings) and reliability improvements (reduction in travel time variability). For this, the transport and traffic models currently used for transport planning need refining. Reliability benefits need to be monetised separately from average time savings, and a number of statistical techniques have been developed to produce credible values of reliability. Assessing reliability also requires a good level of disaggregation in the calculation of benefits by different category of user as the values attached to reliability tend to vary greatly by user group.

Because of this wide variation in values, the main pitfall to be avoided in assessments is transferring values for reliability from one location or situation to another without demonstrating that the cases are sufficiently similar. Applying a uniform uplift factor to average time savings to attempt to reflect the increasing importance of reliability to transport network users can produce highly misleading results.

Robust and consistent reliability assessments of reliability can be developed. Their deployment is important for informing decisions on achieving more optimal levels of reliability on surface transport networks and for the selection of cost-effective policies and projects for improving transport services. Cost-benefit assessments are the only way of determining which options are most cost-effective in improving reliability and choosing between public investments in capacity, network management and information systems and user responses through changes to travel patterns and the logistical organisation of businesses.

REFERENCES

Abdel-Aty, M., Kitamura, R. and Jovanis, P. (1995), Investigating effects of travel time variability on route choice using repeated-measurement stated preference data, *Transportation Research,* 1493, pp. 39-45.

Accent and HCG (1996), The value of travel time in UK roads – 1994; Final report, Report for the Department for Transport, Accent and Hague Consulting Group.

Affleck, F. (2005), Coal railway infrastructure in Queensland and NSW Hunter Valley, ACCC 2005 Competition Conference, Gold Coast, Australia, July. *www.accc.gov.au/content/item.phtml?itemId=658141&nodeId=def2448c97ba44180fa0746a3b57c396&fn=Rail%20session%20-%20Fred%20Affleck.pdf.*

Ahuja, S., Van Vuren T., Porter S. and Fearon, J. (2002), Assessing measures which reduce incident related delays and travel time variability, in *Proceedings European Transport Conference*, 9-11 September 2002, Homerton College Cambridge, UK.

Ansari, A. and Heckel, J. (1987), *JIT Purchasing: Impact on Freight and Inventory Costs*, Journal of Purchasing and Materials Management, vol. 23, no. 2.

Arvis, J.-.F, Raballand, G., Marteau J.-F. (2007), *The Cost of Being Landlocked: Logistic Costs and Supply Chain Reliability,* World Bank Policy Research Working Paper No. 4258.

Asakura, Y. (1996), Reliability Measures of an Origin and Destination Pair in a Deteriorated Road Network with Variable Flows, in *Proceedings of the 4th Meeting of the EURO Working Group in Transportation*.

Asakura, Y. and Kashiwadani M. (1991), Road network reliability caused by daily fluctuation of traffic flow, *Proceedings of the 19th PTRC Summer Annual Meeting (Seminar G)*, Brighton, pp.73-84.

Asensio, J. and Matas A. (2007), Value of time and reliability on C-32 highway near Barcelona, *Studies on public works and transportation*, n. 107, pp. 105-124.

Australasian Freight Logistics (2009), *The network of knowledge*, 16, March/April.

Australian Rail Track Corporation n.d., *Southern Sydney Freight Line project. Macarthur to Sefton. www.ssfl.artc.com.au/docs/SSFL_180505.pdf.*

Avineri, E., and Bovy, P. (2007), Practical aspects in applying prospect theory approach in modelling travellers'responses to network uncertainties, *Proceedings of the Third international Symposium on Transportation Network Reliability*, 19-20 July, 2007, The Netherlands.

AVV (2003), *Klantenbarometer Openbaar Vervoer: Het meten van belang*, AVV, Rotterdam.

Aymerich, O. and Robusté F. (1990), Network reliability under exceptional conditions, *Journal of the Spanish Ministry of Transportation, Tourism and Communications,* n. 42, pp. 25-37.

Badcock, P. (2007), Betuwe Route opens for business, *Railway Gazette International*. *www.railwaygazette.com/features_view/article/2007/07/7581/betuwe_route_opens_for_business.html*.

BAH and ITS (2004), Freight User Benefits Study, 2004. Final Report for Strategic Rail Authority, Booz Allen Hamilton and Institute for Transport Studies, University of Leeds, *www.dft.gov.uk/stellent/groups/dft_railways/documents/page/dft_railways_611117.pdf*.

Baniak, J. (2002), *A decade of partnership, evolution and growth: the I-95 Corridor Coalition*, 9th Workd Congress on Intelligent Transport Systems, Chicago, 14-17 October.

Bartley, R. (2007), Marginal valuations of travel time and scheduling, and the travel time premium, *Transportation Research E*, 43 (4). pp. 387-408

Bates, J., Polak, J., Jones, P., and Cook A. (2001), The valuation of reliability for personal travel, in *Transportation Research Part E* (Logistics and Transportation Review), 37-2/3, pp. 191-229.

Beuthe, M. and Bouffioux, C. (2008), Analysing qualitative attributes of freight transport from stated orders of reference experiment, *Journal of Transport Economics and Policy,* Volume 42, Part 1, pp. 105-128.

Becket, H. (2007), Competition drives rail freight renaissance, *Railway Gazette International*, December 2007.

Beddow, M. (2007), Private practice, in *Containerisation International*, August 2007, pp. 49.

Bell, M. (1999), Measuring Network Reliability: a Game Theoretic Approach, *Journal of Advanced Transportation*, 33 (2), p.125-134.

Bell, M. and Iida Y. (1997), *Transportation Network Analysis*, J. Wiley, Chichester, New York, USA.

Bertini, R., Lyman, K. (2007), Developing improved travel time reliability measures for real-time and archived ITS data applications, Paper presented at the ITS Europe, June 20 2007, Aalborg, Denmark.

Bhouri, N. Haj-Selam, H. (2009), Improving Travel Time Reliability Using Ramp Metering: Field Assessment Results on the A6W Motorway in Paris, 12th IFAC Symposium on Control in Transportation Systems, September 2-4, Renodo Beach, CA, USA.

Bhouri, N. Haj-Selam, H. (2009), Motorway ramp metering as a tool for improving the travel time reliability, EWGT2009 – 13th Euro Working group on Transportation Meeting, September 23-25, Padua, Italy.

Black, I. and Towriss, J. (1993), *Demand effects of travel time reliability*, Centre for Logistics and Transportation, Cranfield Institute of Technology.

Blanchard Company, undated, *How Union Pacific measures train performance*. *www.rblanchard.com/resources/texts/unionpacificperf.html*.

Blumenhagen, D. (1981), Containerisation and Hinterland Traffic, *Maritime Policy and Management*, 8 (3), 197-206.

Bogers, E. and van Lint, H. (2007), Traveler's perception of reliability: How to measure and how to influence? *Proceedings of the Third international Symposium on Transportation Network Reliability*, 19-20 July, 2007, The Netherlands

Bogers, E. and Van Zuylen, H. (2004), The Importance of Reliability in Route Choice in Freight Transport for Various Actors on Various Levels, *Proceedings European Transport Conference*, October 2004, Strasbourg, France.

Bogers, E., Viti, F. and Hoogendoorn S. (2005), Joint modeling of ATIS, habit and learning impacts on route choice by laboratory simulator experiments, *Proceedings of the 84th Annual Meeting of the Transportation Research Board*, Washington DC, USA 9-13 January 2005.

Brooks, M. (2000), *Sea Change in Liner Shipping*, Oxford, Pergamon.

Brooks, M. and Cullinane, K. (2007), Introduction, in Brooks, M.R., Cullinane, K. (eds), *Devolution, Port Governance and Port Performance, Research in Transportation Economics*, vol. 17, Elsevier, pp. 3-28.

Brownstone, K. and Small, K. (2002), Valuing time and reliability: assessing evidence from road pricing demonstrations, University of California, Irvine.

Brownstone, K. and Small, K. (2005), Valuing time and reliability: assessing the evidence from road pricing demonstrations, *Transportation Research Part A*, 39, pp. 279-293.

Bruzelius, N. (2001), *The valuation of logistics improvements in CBA of transport investments – a survey*, SAMPLAN, SIKA, Sweden.

Bryan, J., Weisbrod, G., Martland, C. (2007), *Rail freight solutions to roadway congestion – final report and guidebook*, National Cooperative Highway Research Program Report 586, Transportation Research Board, Washington, D.C.

BSL (2008), *Survey on price and demand elasticity in terms of reliability in freight railway services*, BSL Management Consultants GmbH & Co. *www.internationaltransportforum.org/jtrc/infrastructure/networks/documents.html*.

BTCE (1996), *Quality of rail freight service: the customer perspective*, Bureau of Transport and Communication Economics, Working Paper 96, Australian Government Publishing Service.

BTRE (2003), *Rail infrastructure pricing: principles and practice*, Bureau of Transport and Regional Economics, Report 109, Canberra, Australia. *www.btre.gov.au/publications/61/Files/btre_r109.pdf*.

Burns, R. (2004), Electricity reliability: how much, by what means, at what cost?, *International Association for Energy Economics Conference*, Washington DC, July, *www.iaee.org/documents/washington/Robert_Burns.pdf*.

Cambridge Systematics (2006), *Traffic congestion and reliability: Trends and Advanced Strategies for Congestion Mitigation*. Final report prepared for Federal Highway Administration. *http://ops.fhwa.dot.gov/congestion_report/congestion_report_05.pdf*.

Cassir, C., Yang, H., Ka Kan, L., Tang W., and Bell, M. (2001), Travel time versus capacity reliability of a road network, *Reliability of transport networks*, Research Studies Press Ltd., United Kingdom.

CEMT/ITF (2007), Congestion, a global challenge, OECD, Paris.

Chen, A., Yang, H., Lo, H. and Tang, W. (1999), A capacity related reliability for transportation networks, *Journal of Advanced Transportation*, 2 (vol. 33) pp. 183-200.

City of Chicago Department of Transportation (2003), Freight rail futures report (Appendices and bibliography). *http://egov.cityofchicago.org/webportal/COCWebPortal/COC_EDITORIAL/3appendices.pdf*.

Clark, S. and Watling D. (2005), Modelling network travel time reliability under stochastic demand, *Transportation Research Part B: Methodological*, 2 (vol. 39) pp. 119-140.

Cohen, S. and Zhang, M.Y. (2007), Tronc commun A4-A86. Utilisation dynamique de la voie auxiliaire: quel impact sur le trafic ?, *Revue générale des routes*, Octobre 2007, pp. 44-48

Committee on Transportation and Infrastructure (2006), *U.S. rail capacity crunch*, 109th Congress House Hearings, Hearing before the subcommittee on railroads of the Committee on Transportation and Infrastructure, House of Representatives. *http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=109_house_hearings&docid=f:28281.wais*.

Copley, G., Murphy, P. and Pearce, D. (2002), Understanding and valuing journey time variability, European Transport Conference 2002, Cambridge.

COWI (2005), Strategieanalyse København-Ringsted, Samfundsokonomisk analyse, Baggrundrapport.

CRS (2007), *CRS report for Congress, Rail transportation of coal to power plants: reliability issues*, Congressional Research Services. *http://ncseonline.org/NLE/CRSreports/07Oct/RL34186.pdf*.

Cullinane K., Song D-W. (Ed.) (2007), Asian Container Ports. Development, Competition and Co-operation, Palgrave Macmillan.

CURE (2005), Rail Report, Consumers United for Rail Equity, newsletter, August. *www.railcure.org/pdf/newsletter0805.pdf*.

Danielis R., Marcucci, E. and Rotaris, L. (2005), Logistics managers' stated preference for freight service attributes, *Transportation Research Part E 41*, pp. 201-215.

Deardorff, A. (2003), The importance of the cost and time of transport for international trade, Paper presented to "Time and transport" Round Table 127, ECMT, Paris, 4th-5th December.

De Jong, G., Daly, A., Pieters M., Vellay, C., Bradley, M., and Hofman, F. (2003), A model for time of day and mode choice using error components logit, *Transportation Research E*, Logistics and Transportation Review, 39, pp. 245-268.

De Jong, G., Kroes, E., Plasmeijer, R., Sanders, P., Warffemius, P. (2004), The value of reliability, *Proceedings of the European Transport Conference 2004*, 4-6 October, Strasbourg, France.

De Langen, P. (2007), Stakeholders, Conflicting Interests and Governance in Port Clusters, in Brooks, M. and Cullinane, K. (eds), *Devolution, Port Governance and Port Performance*, Research in Transportation Economics, vol. 17, Elsevier, pp. 457-477.

D'Este, G. and Taylor, M. (2003), Network vulnerability: an approach to reliability analysis at the level of national strategic transport networks, in *The Network Reliability of Transport*, Bell, M. and Iida, Y. (Eds.), Pergamon, Oxford, UK, pp. 23 - 44.

DfT (1998), A new deal for trunk roads in England: guidance on the new approach to appraisal, United Kingdom Department for Transport. *www.dft.gov.uk/pgr/economics/rdg/multimodal/anewdealfortrunkroadsinengla5491*.

DfT (2002), Multi-modal transport appraisal investment , United Kingdom Department for Transport. *www.dft.gov.uk/162259/165223/multimodalappraisal*.

DfT (2006), Journey time reliability on motorways and trunk roads: measure for PSA target published, Press Notice, 14 February, United Kingdom Department for Transport. *www.gnn.gov.uk/environment/fullDetail.asp?ReleaseID=187497&NewsAreaID=2&NavigatedFromDepartment=True*.

DfT (2007), *Delivering a Sustainable Railway*, White Paper CM 7176, United Kingdom Department for Transport.

DfT (2008), Roads – Delivering Choice and Reliability, Department for Transport, July 2008. *www.dft.gov.uk/pgr/roads/introtoroads/roadcongestion/roadscommandpaper1.pdf*.

DfT (2009), The Reliability Sub-Objective, TAG Unit 3.5.7., United Kingdom Department for Transport. *www.dft.gov.uk/webtag/webdocuments/3_Expert/5_Economy_Objective/3.5.7.htm*.

Dobias, G. (2003), Vers la voiture automate, circulation et sécurité. Odile Jacob (Ed.), Mai 2003.

Douglas Economics (2004), Value of Rail Travel Time, RailCorp. *www.railcorp.info*.

Douglas Economics (2006), Value and Demand Effect of Rail Service Attributes, RailCorp. *www.railcorp.info*.

Dutch Ministry of Transport, Public Works and Water Management (2004a), Aanvulling leidraad OEI-directe effecten, (Appendix to Evaluation of infrastructural projects; Guide for cost-benefit analysis), Den Haag.

Dutch Ministry of Transport, Public Works and Water Management (2004b), Nota Mobiliteit, The Netherlands. *www.notamobiliteit.nl*.

Dziekan, K. and Vermeulen, A. (2006), Psychological Effects of and Design Preferences for Real-Time Information Displays, *Journal of Public Transportation*, Volume 9, No. 1 (*www.nctr.usf.edu/jpt/jptv9n1.htm*), pp. 71-89.

ECMT (2006a), Inland Waterways and Environmental Protection, OECD, Paris.

ECMT (2006b), *Strengthening Inland Waterway Transport. Pan-European Co-operation for Progress*, OECD, Paris.

Eliasson, J. (2004), Car drivers' valuations of travel time variability, unexpected delays and queue driving, *Proceedings of the European Transport Conference*, 2004.

Eliasson, J. (2006), *Forecasting Travel Time Variability*, Working Paper.

Emmerink, R., Verhoef, E., Nijkamp, P. and Rietveld, P. (1998), Information policy in road transport with elastic demand: Some welfare economic considerations, *European Economic Review* 42 (1998), pp. 71-95.

Entrepreneur.com (2007), Railroads rebuilding trust in perishables services, *Frozen Food Digest,* February-March. *www.entrepreneur.com/tradejournals/article/162361506.html*.

FHWA (2004), *Traffic Congestion and Reliability: Linking Solutions to Problems*, United States Federal Highway Administration, Washington, D.C.

FHWA (2005), *Traffic Congestion and Reliability, Trends and Advanced Strategies for Congestion Mitigation*, United States Federal Highway Administration, Washington, D.C.

FHWA (2006), *Travel time reliability: Making it there on time, all the time*, United States Federal Highway Administration, Washington, D.C.

Finnish Road Administration (2000), *Tie- ja liikennetekniikka. Tietyömaiden liikennehaittojen arviointi*. Tielaitoksen selvityksiä 14/2000. Helsinki.

Finnish Road Administration (2001), Winter road maintenance, policy 2001, Road Administration. *http://alk.tiehallinto.fi/thohje/pdf/winter_road_mainten_policy_2001.pdf*.

Florida Department of Transportation Research (2008), Evaluating the effectiveness of various truck lane restriction practices in Florida: Phase II. *www.dot.state.fl.us/research-center/Completed_Proj/Summary_TE/FDOT_BD543_10.pdf*.

Fosgerau, M., Hjorth, K, Brems, C. and Fukuda D. (2008), *Travel time variability – Definition and valuation*. Danish Department of Transport 1:2008.

Fosgerau, M. and A. Karlström (2007), The Value of Reliability, MPRA.

Fowkes, A., Firmin P., Whiteing, A. and Tweddle, G. (2001), Freight users valuation of three different aspects of delay, Proceedings of the European Transport Conference 2001, Cambridge.

Fowkes, A., Firmin, P., Tweddle, G. and Whiteing, A. (2004), How Highly Does the Freight Transport Industry Value Journey Time Reliability – and for What Reasons?, *International Journal of Logistics*, Volume 7, Number 1, March 2004, pp. 33-43(11)

Freight Transport Association (2005), Memorandum by the Freight Transport Association, House of Commons Select Committee on Transport, Transport - Seventh Report, *www.publications.parliament.uk/pa/cm200405/cmselect/cmtran/218/218we19.htm*.

Frémont A., Soppé M. (2007), Northern European Range: Shipping Line Concentration and Port Hierarchy, in Wang, J., Olivier, D., Notteboom, T. and Slack, B. (Eds.), *Ports, Cities and Global Supply Chains*, Ashgate, 105-120.

Frémont, A. and Ducruet C. (2005), The Emergence of a Mega-Port – From the Global to the Local, the Case of Busan, *Tijdschrift voor Economische en Sociale Geografie*, vol. 96, no. 4, pp. 421-432.

Frémont, A., Franc P. and Slack, B. (2007), Inland waterways and port competition: a new challenge for the French ports, Paper presented at the IAME (International Association of Maritime Economists) conference, Greece.

Gamelyak, I., Vyrozhemsky, V., Krayushkina, K., Kishchynskiy, S. (2008), *Materials on reliability of motor roads*, Working Group background document, available at *www.internationaltransportforum.org/jtrc/infrastructure/networks/documents.html.*

Georgia Department of Transportation (2008), Statewide truck lane needs identification study. *www.atlantaregional.com/documents/Statewide_Truck_Lane_Study_TCC_Presentation_Jan_18.ppt*.

Giuliano, G. (2004), Alameda Corridor: A Blueprint for the Future?, Summary and Observations, February 10, 2004 Davidson Conference Center, University of Southern California. *www.usc.edu/schools/sppd/keston/pdf/alameda-corridor.pdf*.

Gouvernal, E., Debrie, J. and Slack, B. (2005), Dynamics of change in the port system of the Western Mediterranean, *Maritime Policy Management*, vol.32, no. 2, pp. 107-121.

Hamer, R., De Jong, G. and Kroes, E. (2005), *The value of reliability in Transport – Provisional values for the Netherlands based on expert opinion*, RAND Technical Report Series, TR-240-AVV, Netherlands.

Harder, K., Bloomfield, J., Levinson, D. and Zhang, L. (2005), *Route Preferences and the Value of Travel-Time Information for Motorists Driving Real-World Routes*, University of Minnesota Department of Civil Engineering. *www.lrrb.org/pdf/200520.pdf*.

HCG (1992a), *De reistidjswaardering in het goederenvervoer, Rapport Hoofdonderzoek*, Report 142-1 for Rijkswaterstaat, Dienst Verkeerskunde, Hague Consulting Group, The Hague.

HCG (1992b), Study into the social benefits of goods transport by road; the conjoint analysis, Report HCG 170-2 for the International Road Transportation union, Hague Consulting Group, The Hague.

HEATCO (2005), *Developing Harmonised European Approaches for Transport Costing and Project Assessment*, Deliverable 1, Current practice in project appraisal in Europe, January 2005. *http//:heatco.ier.uni-stuttgart.de/*

HEATCO (2006), *Developing Harmonised European Approaches for Transport Costing and Project Assessment*, Deliverable 5, Proposal for harmonized guidelines, February 2006. *http//:heatco.ier.uni-stuttgart.de/*

Heaver, T. (2002), The Evolving Roles of Shipping Lines in International Logistics, *International Journal of Maritime Economics*, (4): 210-230.

Heaver, T. (2005), Responding to shippers'supply chain requirements, in Leggate, H., McConville, J., Morvillo, A. (eds) *International Maritime Transport Perspectives*, Routledge, London and New York, pp. 202-214.

Heaver, T., Meersman, H. and Van de Voorde, E. (2001), Co-operation and competition in international container transport: strategies for ports, *Maritime Policy and Management* 28 (3), pp. 293-305.

Hollander, Y. (2005), *The attitudes of bus users to travel time variability*, Association for European Transport and contributors 2005.

Hopkin, M. (2006), “Hitting the rails”, 20 October, *RedOrbit online knowledge network*. *www.redorbit.com/news/business/700561/hitting_the_rails/index.html*.

Huisman, T. and Boucherie, R. (2001), Running times on railway sections with heterogeneous train traffic, *Transportation Research Part B 35*, pp. 271-292.

Husdal, J. (2004), Reliability and vulnerability versus costs and benefits, Paper presented at the European Transport Conference 2004, Strasbourg, 4-6 October. *http://husdal.typepad.com/blog/docs/etc2004_epr08ii_husdal_revised.pdf*.

Husdal, J. (2005), The vulnerability of road networks in a cost-benefit perspective, *Proceedings of the 84th Annual Meeting of the Transportation Research Board*, Washington DC, USA 9-13 January 2005.

Immers, L., Stada, J., Yperman, I. and Bleukx A. (2004), Robustness and resilience of transportation networks, *Proceedings of the 9th International Scientific Conference MOBILITA*, Bratislava, Slovenia, 2004.

Infopolis 2 Consortium (2000), Inventory: At-stop Display, Passenger Information on Public Transport. *www.ul.ie/~infopolis and www.ul.ie/~infopolis/existing/stopdisp.html*.

International Railway Journal (2007), Lötschberg tunnel set for June opening, *International Railway Journal*, June.

InterVistas (2008), *Value of Time and Reliability for Local Trips in Canada, Final Report,* March 2008.

Janin, J.-F. (2003), *Des transports intelligents ? Comment y parvenir*, Les dossiers du CERTU, N°148, CERTU.

Jones, C. and Sedor, J. (2006), Improving the Reliability of Freight Travel, *Public Roads*, July/August.

JTRC (2008), Port Competition and Hinterland Connections, *Joint Transport Research Centre, Discussion Papers,* No. 2008-19.

Keisling, L. (2004), Is network reliability a public good?, International Association for Energy Economics Conference, Washington DC, July. *www.iaee.org/documents/washington/Lynne_Kiesling.pdf.*

Knoop, V., Hoogendoorn, S. and van Zuylen H. (2007), Quantification of the impact of spillback modeling in assessing network reliability, *Proceedings of the 86th Annual Meeting of the Transportation Research Board*, Washington DC, USA, CD-Rom version.

Konings, R. (2006), Hub and spoke networks in container on barge transport, Paper presented at the 85th Annual Meeting of the Transportation Research Board, Washington DC, USA TRB meeting, 2006.

Kouwenhoven, M., de Jong, G. and Rietveld, P. (2005), *Reliability Ratio's voor het Goerderenveroer*, Final report to AVV, RAND Technical Report Series, WR-274-AVV, Netherlands.

Kouwenhoven, M., van Grol, R., Kroes, E., Schoemakers A. (2006), Development of a tool to assess the reliability of Dutch road Networks, *Proceedings of the European Transport Conference 2006*, October, Strasbourg, France.

Kuipers, B. and Rozemeijer, S. (2006), *Strategies by Shippers and Transport Companies in Response of Decreasing Reliability of Travel Times,* TNO, Delft.

Kummer, S., Nagl, P. (2007), *Economic Benefits of Traffic Control Units*, Zentrum Transport Logistik GmbH, January, Wien.

Linauer M., Schneider M. and Hainitz N. (2006), Travel time estimation based on real-time toll data, *13th World Congress on ITS*, 08.-12.10.2006, London, UK

Litman, T. (2007), *Travel Time, Transportation Costs and Benefit Analysis*, Victoria Transport Policy Institute. *www.vtpi.org*.

Lo, H. (2002), Trip travel time reliability in degradable transport networks, *Proceedings of the 15th International Symposium on Transportation and Traffic Theory (ISTTT)*. Adelaide, South Australia.

Lomax, T., Schrank, D., Turner, S. and Margiotta, R., (2003), *Selecting travel reliability measures*, Texas Transportation Institute Cambridge Systematics, Inc.

M6 Toll After Study (2005), *Post Opening Project Evaluation, M6 Toll After Study, Traffic and Safety Summary*. *www.highways.gov.uk/roads/documents/one_year_after_study.pdf*

Mackie, P., Nellthorp, J., Laird J. And Ahmed F. (2003), *Toolkit for the Evaluation of World Bank Transport Projects*, The World Bank, Washington, DC.

Margiotta, R. and Taylor, R. (2006), Traffic Congestion and Reliability: Making the Connection with Operations, Part 1: Measuring and Tracking Reliability. *ITE Journal*, February 2006.

Mattsson, L.-G. (2004), Train service reliability. A survey of methods for deriving relationships for train delays. *http://users.du.se/~jen/Seminarieuppsatser/Forsening-tag-Mattsson.pdf*

McKinnon, A.C. (2004), Benchmarking the Efficiency of Retail Deliveries in the UK, BRC Solutions, *British Retail Consortium magazine*, Issue 5, 2004.

McKinnon, A. (2006), Road Transport Optimisation., in Waters, D. (Ed.), *Global Logistics and Distribution Planning* – 5th edition, Kogan Page, London.

McKinnon, A., Palmer, A., Edwards, J. and Piecyk, M. (2008), *Reliability of Road Transport from the Perspective of Logistics Managers and Freight Operators*, Logistics Research Centre, Heriot-Watt University. *www.internationaltransportforum.org/jtrc/infrastructure/networks/documents.html*.

McKinnon, A., Yongli G. and Duncan L. (2003), *Analysis of Transport Efficiency in the UK Food Supply Chain*, Logistics Research Centre, Heriot-Watt University, UK, April 2003.

MDS Transmodal (2007), *Supply and Demand in the Container Trades. Global and Regional Trends and Forecasts*.

MergeGlobal (2006), Coping with uncertainty, How might market forces create new ocean transport products for service-sensitive container shipments?

MINTC (2006), Measures for improving the smoothness of heavy goods traffic flows at the Finnish-Russian border, Publications of the Finnish Ministry of Transport and Communications 55/2006.

Moses, B. (2007), Evaluating the effectiveness of various truck lane restriction practices in Florida – Phase II, Volume 3 – Evaluation of truck lane restriction in non-limited access urban areas. Final Report, *www.dot.state.fl.us/research%252Dcenter/Completed_Proj/Summary_TE/FDOT_BD543_10_rpt_v3.pdf*.

Mott MacDonald (2009), Development of INCA to incorporate single carriageways and managed motorways, UK Department for Transport, *www.dft.gov.uk/pgr/economics/rdg/ttv/incaresearch/inca.pdf*.

MTC (2006), MTC Transit Connectivity Plan, Metropolitan Transportation Commission. *www.mtc.ca.gov/planning/connectivity/Final_Connectivity_Study/finalsummary.pdf*.

MVA (2000), Etude de límpact des phénomenès dírregularité des autobus – Analyse des resultants, MVA, Paris.

Natarajan, B., Duncan, C. and Simpson, D (2005), Strategic innovations in North American railroad management, Transportation Research Record: Journal of the Transportation Research Board, No. 1924, pp. 146-152.

NCHRP (2008), Cost-effective performance measures for travel time delay, variation and reliability, National Cooperative Highway Research Report 618, Transportation Research Board, Washington, D.C.

Nes, R., Marchau, V., van Wee, G. and Hansem, I. (2007), Reliability and robustness or multimodal transport network analysis and planning: towards a new research agenda. Paper presented at the 3rd INSTR conference, Rotterdam.

Network Rail (2009), Train Performance Report, period 2, 2009/10. *www.networkrailmediacentre.co.uk/imagelibrary/downloadMedia.asp?MediaDetailsID=2361*.

New Zealand Transport Agency (2008), Economic Evaluation Manual, Volume 1, Amendment No 2, September 2008, New Zealand Transport Agency. *www.nzta.govt.nz*.

Nicholson, A. (2007), Road network unreliability: impact assessment and mitigation, International Journal of Critical Infrastructures, Vol 3, Nos 3-4.

Nicholson, A., Dalziell, E. (2001), Assessing and Mitigating the Impacts of Road Network Unreliability. Paper presented at the first INSTR Symposium.

Notteboom, T., Konings, R. (2004), Network dynamics in container transport by barge, Belgéo, 4, pp. 461-478.

Notteboom, T., Rodrigue, J-P. (2007), Re-assessing Port-Hinterland Relationships in the Context of Global Commodity Chains, in J. Wang, J., Olivier, D., Notteboom, T. and Slack, B. (Eds.), Ports, Cities and Global Supply Chains, Ashgate.

Notteboom, T., Winkelmans, W. (1999), Spatial (de)concentration of container flows: the development of load centre ports and inland hubs in Europe, in *Transport Modes and Systems, 8th WCTR Proceedings*, Pergamon, pp. 57- 71.

OECD (2005), *Economic Evaluation of Long-Life Pavements: Phase 1*, 2005(3), OECD, Paris.

OECD/ECMT (2007), *Managing Urban Traffic Congestion*, OECD, Paris.

Okutani, T. (2008), *Road Network Reliability in Japan*, Working Group background document, available at *www.internationaltransportforum.org/jtrc/infrastructure/networks/documents.html.*

Palen, J. (1997), The Need for Surveillance in Intelligent Transportation Systems, *Intellimotion*, 6:1, pp. 1-10, University of California PATH, Berkeley, CA.

PDFC (2002), *Passenger Demand Forecasting Handbook*, Passenger Demand Forecasting Council. *www.atoc.org/RPDFS/index.asp*.

Péan, S. (2004), *Le déploiement des ITS en France. Réalisations et stratégies*, Thèse professionnelle. Mastère SIT, Ecole Nationale des Ponts et Chaussées. Décembre 2004.

Philp, J. (2007), Shipper success story. Railex Produce Unit Train, SEARS Conference, 2 October. *www.railshippers.com/regional/southeast/philp_presentation.pdf*.

Progressive Railroading (2003), Container volume is growing, but railroads' intermodal profit margins are marginal at best, *Progressive Railroading*, 6 June. *www.progressiverailroading.com/commentary/article.asp?id=13104*.

Providing a Highway System with Reliable Times, F-SHRP Web Document. *http://onlinepubs.trb.org/onlinepubs/f-shrp/f-shrp_webdoc_3.pdf*.

Rail Cargo Netherland, n.d., Betuwe Route: Freight only. *www.railcargo.nl/index.cfm/menuid/6/Product/418*.

RAND Europe, SEO and Veldkamp/NIPO (2004), *Hoofdonderzoek naar de reistidwaardering in het goederenvervoer*, Report for the Adviesdienst Verkeer en Vervoer, RAND Europe, Leiden.

Rieser, M., Nagel, K., Beuck, U., Balmer, M. and Rümenapp, J. (2007), Truly agent coupling of an activity-based demand generation with a multi agent traffic simulation, in *Proceedings of WCTR'07*, World Conference on Transport Research, Berkeley, CA, 2007

Rietveld, P., Bruinsma, F. and van Vuuren, D. (2001), Coping with unreliability in public transport chains: A case study for the Netherlands, *Transportation Research A*, 35, pp. 539-559.

Rijkswaterstaat (2004), DZH, Evaluatie doelgroepvoorzieningen A16 en A20, november 2004, DHV Ruimte en Mobiliteit.

Ritsema van Eck, J., Snellen, D. and Hilbers, H. (2004), Travel time reliability: methodology and some results for the Netherlands, Paper presented at WCTR 2004, Istanbul

Robinson, R. (2002), Ports as elements in value-driven chain systems: the new paradigm, *Maritime Policy and Management* 29 (3), pp. 241-255.

SACTRA (1999), *Transport and the economy: full report*, The Standing Advisory Committee on Trunk Road Assessment, Department for Transport, London.

Santos, B., Antunes, A. and Miller, E. (2007), Interurban road network planning model with accessibility and robustness, *Proceedings of the Third international Symposium on Transportation Network Reliability*, 19-20 July, 2007, The Netherlands

Saporta, M. (2008), Freight forecast shows shift to trains makes sense, AJC.COM, 3 March. *www.ajc.com/business/content/business/saporta/stories/2008/03/02/saporta_0303.html*.

Schrank, D. and Lomax, T. (2005), *The 2005 Urban Mobility Report*, Texas Transportation Institute, Texas.

Schreuder, M., Molenkamp, L., Tamminga, G. and Kraan, M. (2007), Vulnerability of a national road network, Policy, research Quick Scan & Model, *Proceedings of the Third international Symposium on Transportation Network Reliability*, 19-20 July, 2007, The Netherlands

Schrijver J., Meeuwissen, H., Hilbers, H. (2003), *Het voorspellen van betrouwbaarheid van reistijden met een vervoerprognosemodel: SMARA*, In Colloquium Vervoersplanologisch Speurwerk 2003, CVS, Rotterdam.

Senna, L. (1991), Risk of delays, uncertainty, and travellers' valuation of travel time variability, 19th PTRC Summer Annual Meeting, Brighton.

SIKA (2000), En hög transportkvalitet för näringslivet – Underlag om mal mätt och metoder. Rapport från arbetsgrupp inom SIKAs måluppdrag, 2000-01-31. *www.sika-institute.se/Doclib/Import/105/sr_2000_1u7.pdf*.

SIKA (2008), *Samhällsekonomiska principer ock kalkylvärden för transportsektorn: ASEK 4*, SIKA PM 2008:3.

Slack, B. (1993), *Pawns in the Game: Ports in a Global Transportation System*, Growth and Change, Vol.24, pp.579-588.

Slack, B. (1999), Satellite terminals: a local solution to hub congestion?, *Journal of Transport Geography*, pp. 7, 241-246.

Slack, B. (2007), The Terminalisation of Seaports, in Wang, J., Olivier, D., Notteboom, T., and Slack, B. (Eds.), *Ports, Cities and Global Supply Chains*, Ashgate.

Slack, B., Comtois, C. and McCalla, R. (2002), Strategic alliances in the container shipping industry: a global perspective, *Maritime Policy and Management* 29 (1), pp. 65-76.

Small, K (1982), The Scheduling of Consumer Activities: Work Trips, in *American Economic Review*, 72, June 1982, 467-479.

Small, K, Noland, R., Chu, X. and Lewis, D. (1999), *Valuation of travel time savings and predictability in congested conditions for highway user-cost estimation*, Report 431, National Cooperative Highway Research Program, Washington, D.C.

Small, K., Winston, C. and Yan, J. (2005a), Uncovering the distribution of motorists' preferences for travel time and reliability, *Econometrica*, vol. 73, issue 4, pp. 1367-1382.

Small, K., Winston, C., Yan, J. (2006), Differentiated road pricing, express lanes, and carpools: exploiting heterogenous preferences in policy design. *Brookings-Wharton Papers on Urban Affairs*.

Soriguera, F. and Robusté, F. (2008), *Travel Time Reliability and Information Systems*, Working Group background document, available at *www.internationaltransportforum.org/jtrc/infrastructure/networks/documents.html*

Soriguera, F., Thorson, L. and Robusté, F. (2007), Travel Time Measurement Using Toll Infrastructure, *Transportation Research Record: Journal of the Transportation Research Board*, n. 2027, National Research Council, Washington, D.C., pp. 99-107.

Soriguera, F., Thorson, L., Grau, J. and Robusté, F. (2006), Fundamentals for a travel time surveillance system in Catalonia, *Proceedings of the VII Congress on Transportation Engineering*, Ciudad Real.

Steer Davies Gleave (2004), Effects of road congestion on rail demand: technical report, A report to the Association of Train Operating Companies (ATOC).

Stopford, M. (1997), *Maritime Economics*, Routledge, London.

Sumalee A. and Watling D. (2003), Travel time reliability in a network with dependent link modes and partial driver response, *Journal of the Eastern Asia Society for Transportation Studies*,5, pp. 1687-1701.

Swenseth, S. and Buffa, F. (1990), Just-in-Time: Some Effects on the Logistics Function, *International Journal of Logistics Management*, vol. 1, no. 2, pp. 25-34.

Takagi (2004), High-speed railways: the last 10 years, *Japan Railway & Transport Review 40*, pp. 4-7. *www.jrtr.net/jrtr40/f04_tak.html*.

Taylor, M. (2007), Using accessibility metrics for road transport vulnerability analysis and the identification of critical infrastructure locations, *Proceedings of the Third international Symposium on Transportation Network Reliability*, 19-20 July, 2007, The Netherlands.

Taylor, M. (2008), *A longitudinal study of travel time variability on a commuter route*, Working Paper, August 2008, Institute for Sustainable Systems and technologies, University of South Australia, Adelaide.

Thuermer, K. (2003), Laying down track. Intermodal gains favor with shippers, *World Trade Magazine*, 1 November. *www.worldtrademag.com/CDA/Articles/Rail/1aba8e96f8af7010VgnVCM100000f932a8c0____*.

Tioga Group (2003), Project 99-130: Goods movement truck and rail study. Final Report. Report prepared for Southern California Association of Governments *www.scag.ca.gov/goodsmove/pdf/truckrail/ch5.pdf*.

Transek (2003b), *Förseningar, restidsosäkerhet och trängsel i samhällsekonomiska kalkyler*, Underlag till ASEK arbetet 2002, Transek 3 maj 2002.

Transek (2006), *Samhällsekonomiska kalkyler för Nord-sydlika förbindelser i Stockholm*, 18:2006.

TransFund (1998), *Project Evaluation Manual*, TransFund New Zealand *www.transfund.govt.nz*.

TranSystems Corporation (undated), *Multistate Transportation Operations Programs. Critical issues and opportunities. www.uppermidwestfreight.org/files/04-09-13_Summary_Report.pdf*.

TRB (1997), *Highway Capacity Manual*, Transportation Research Board, Washington D.C., USA.

TRB (2008), *The Potential Impacts of Climate Change on U.S. Transportation*, National Research Council of the National Academies, Transportation Research Board Special Report 290, Washington, D.C.

Tseng, Y.-Y., (2008), *Valuation of Travel Time Reliability in Passenger Transport*, PhD theses, Tinberger Institute.

Turnbull K. and Pratt R. (2003), Transit Information and Promotion: Traveler Response to Transport System Changes, Chapter 11, TCRP 95, Transportation Research Board, *http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_rpt_95c11.pdf*.

United Nations (2008), *Joint Study on Developing Euro-Asia Transport Linkages*, Prepared under the United Nations Development Account Project on "Capacity Building through Cooperation in Developing Land and Land-Sea Interregional Transport Linkages", United Nations, New York and Geneva.

Van de Kaa, E. (2008), Extended prospect theory: findings on choice behaviour from economics and the behavioural sciences and their relevance for travel behaviour, Trail Research School, Netherlands. *www.narcis.info/dare/RecordID/oaitudelftnl381534/Language/nl/repository_id/tuddare/*.

Van Lint, J. (2004), Freeway Travel Time Reliability Maps; Using the shape of the day-to-day travel time distribution, Paper presented at the 8th TRAIL congress, 2004 November 23, Rotterdam, The Netherlands.

Van Lint, J. and Van Zuylen, H. (2005), Monitoring and predicting freeway travel time reliability, *Transportation Research Record* 1917 (Data initiatives), pp. 54-62.

Van Lint, J., Zuylen, H. and Tu, H. (2008), Travel Time Reliability on Freeways: Different Why measures based on variance tell only half the story, *Transportation Research Part A: Policy and Practice*, Volume 42, Issue 1, January 2008, Pages 258-277

Van Slobbe, R. (2002), Larger volumes, increasing scale, environmental aspects and road congestion require intermodal solutions, unpublished presentation, ITMMAPS Conference, Antwerp, April 18-20, 2002.

Vickrey, W. (1969), Congestion theory and transport investment, in *American Economic Review* (Papers and proceedings), 59, pp. 251-261.

Vincent, M. (2008), *Measurement Valuation of Public Transport Reliability*, Land Transport New Zealand Research Report 339. *www.landtransport.govt.nz*

Wakabayashi, H. and Iida, Y. (1992), Upper and lower bounds of terminal reliability of road networks: an efficient method with boolean algebra, *Journal of Natural Disaster Science* (vol. 14) pp. 29-44.

Wardman, M. (2001), A review of British evidence on time and service quality valuations, in *Transportation Research Part E* 37, pp. 107-128, 285-316.

Wardman, M. (2004), Public transport values of time, in *Transport Policy* 11, pp. 363-377.

Warffemius, P. (2005), *Using the Standard Deviation of the Travel Time Distribution as an Indicator for Valuing the Reliability of Travel Time*, AVV Research Program Report. Dutch Ministry of Transport, Public Works and Water Management.

Washington State Transportation Committee (2006), *Statewide Rail Capacity and System, Needs Study,* Task 10.2 – Washington State Rail Investment Plan, Technical Memorandum. *www.wstc.wa.gov/Rail/TM10_2_WashStateRailInvestPlan.pdf*.

Watling, D., Sumalee, A., Connors, R. and Balijepalli, C. (2004), *Advancing Methods for Evaluating Network Reliability*, Institute for Transport Studies, University of Leeds, Leeds.

Woychick, E. (2006), Capacity for the energy-only market: demand response to educate customers, hedge market power, and ensure revenue adequacy. *http://speakuptexas.com/electric/projects/32853/Eric_Woychik_WP.doc*.

Yang, H., Bell, M. and Meng, Q. (2000), Modeling the Capacity and Level of Service of Urban Transportation Networks, *Transportation Research Part B*, 34, p.255-275.

Ynace, J.-L. and De Banville, E. (1999), Les systèmes de Transports Intelligent : un enjeu Stratégique mondial, *La documentation française*. Décembre 1999.

Yuen-wah L. (2003), Evaluating the Urban Commute Experience: A Time Perception Approach, *Journal of Public Transportation*, Vol. 6, No. 4, pp. 41-67. *www.nctr.usf.edu/jpt/pdf/JPT%206-4%20Li.pdf*.

Ziha, K. (2000), Event oriented system analysis. *Probabilistic engineering mechanics*, vol. 15, No. 3, Elsevier, pp. 261-75.

Ziha, K. (2000), Redundancy and robustness of systems of events. *Probabilistic Engineering Mechanics*, 4 (vol. 15) pp. 347-357.

CONTRIBUTORS TO THE REPORT

Chair

Mr. Hans JEEKEL
Centre for Infrastructure of the Dutch Ministry of Transport,
Public Works and Water Management (Netherlands)

Members of the Working Group

Australia	Mr. Peter KAIN Bureau of Transport Economics
	Ms. Lyn MARTIN Bureau of Transport Economics
Austria	Mr. Florian MATIASEK Federal Ministry of Transport, Innovation and Technology
Canada	Mr. Louis-Paul TARDIFF Transport Canada
Denmark	Mr. Flemming CLAUSEN Danish Road Directorate
Finland	Mr. Juha PARANTAINEN Ministry of Transport and Communications
France	Ms. Neila BHOURI INRETS
	Mr. Xavier DELACHE Ministry for Ecology, Energy, Sustainable Development and Sea
	Mr. Antoine FRÉMONT INRETS
Greece	Mr. Matthew KARLAFTIS National Technical University
Japan	Mr. Tadashi OKUTANI National Institute for Land and Infrastructure Management
Netherlands	Mr. Jan VAN DER WAARD Centre for Transport and Navigation (DVS)
	Mr. Pim WARFFEMIUS

	Kennisinstituut voor Mobiliteitsbeleid (KiM)
Spain	Mr. Francesc ROBUSTÉ Technical University of Catalonia (UPC)
	Mr. Francesc SORIGUERA Technical University of Catalonia (UPC)
Ukraine	Mr. Sergiy KISHCHYNSKY State Road Scientific Research Institute
	Ms. Kateryna KRAYUSHKINA State Road Scientific Research Institute
	Mr. Valery VYROZHEMSKY State Road Scientific Research Institute
United Kingdom	Mr. Prabhat VAZE UK Department for Transport
United States	Mr. William HYMAN Transportation Research Board/SHRP 2
	Mr. Ed WEINER US Department of Transportation
Joint Transport Research Centre of the ITF and the OECD	Mr. Jari KAUPPILA
	Mr. Stephen PERKINS

The principal editors of the report were: Peter Kain and Jari Kauppila. Members of the Editorial Committee were: Xavier Delache (France), Antoine Frémont (France), William Hyman (United States/SHRP 2), Lyn Martin (Australia), Stephen Perkins (Secretariat), Prabhat Vaze (United Kingdom), Jan van der Waard (Netherlands), Pim Warffemius (Netherlands), and Ed Weiner (United States)

The following persons reviewed a final draft of the report and their insightful comments allowed for key improvements to the text and substance:

External reviewers

Professor Dr. Gerard de Jong, Significance, ITS Leeds and NEA

Mr. Bruno Jacques, Director, Economic and Environmental Analysis & Research, Transport Canada

Other commentators

Ms. Rose Anne Amourdon, Transport Canada

Mr. Malcolm Blair, Vanderbilt Center for Transportation Research, United States

Mr. Serge Le Cunff, Sétra/DREX, France

Professor Jonas Eliasson, Royal Institute of Technology, Sweden

Mr. Mogens Fosgerau, Institute for Transport, Denmark

Mr. Joel Franklin, Royal Institute of Technology, Sweden

Mr. Paul Koster, Department of Spatial Economics, VU University Amsterdam, Netherlands

Ms. Joanne Leung, Ministry of Transport, New Zealand

Ms. Alina Mustra, World Bank

Professor Lauri Ojala, Turku School of Economics, Finland

Ms. Stefanie Peer, Department of Spatial Economics, VU University Amsterdam, Netherlands

Ms. Hanne Samstad, TØI, Norway

Professor Michael Taylor, University of South Australia

Professor Erik Verhoef, Department of Spatial Economics, VU University Amsterdam, Netherlands

Mr. Matthew White, Chief Scientific Adviser's Unit, Department of Transport, United Kingdom

The Working Group acknowledges the insightful comments provided also by the members of the SHRP2 Reliability Technical Coordinating Committee and the Joint Transport Research Centre Committee.

The following consulting reports were developed as inputs to this project, both of which are available in their entirety on the Transport Research Centre's website, at

www.internationaltransportforum.org/jtrc/infrastructure/networks/documents.html

McKinnon, A., Palmer, A., Edwards, J., Piecyk, M. (2008), *Reliability of Road Transport from the Perspective of Logistics Managers and Freight Operators*, Logistics Research Centre, Heriot-Watt University.

BSL (2008), *Survey on price and demand elasticity in terms of reliability in freight railway services*, BSL Management Consultants GmbH & Co.

OECD PUBLISHING, 2, rue André-Pascal, 75775 PARIS CEDEX 16
PRINTED IN FRANCE
(77 2009 01 1 P) ISBN 978-92-821-0241-1– No. 57339 2010

IMPROVING RELIABILITY ON SURFACE TRANSPORT NETWORKS

Passengers and freight shippers alike want reliable transport services. Surprisingly, little research has been undertaken in incorporating reliability into the assessment of transport projects despite the increasing importance of scheduling in economic activities.

This report provides policy makers with a framework to understand reliability issues, to incorporate reliability into project assessment and to design reliability management policies. It also explores a range of reliability performance measures. Case studies across OECD and ITF countries provide examples of several core policy tools that can be used to deliver more reliable networks in a cost-effective manner.

The report makes significant progress in identifying appropriate methodology for incorporating reliability into policy and project evaluation, as well as exploring the pitfalls that need to be avoided.

www.internationaltransportforum.org

OECDpublishing
www.oecd.org/publishing

(77 2009 01 1 P1)
ISBN 978-92-821-0241-1

9 789282 102411