

Internal Benchmarking

August 2005

**A report prepared for the London Underground PPP Arbiter by
PricewaterhouseCoopers LLP in association with Bovis Lend Lease**

Contents

IMPORTANT NOTICE	3
1 EXECUTIVE SUMMARY	4
1.1 THE NATURE OF OUR TASK.....	4
1.2 HOW THE BENCHMARKING FRAMEWORK SHOULD OPERATE	6
1.3 COMPLEX ISSUES	9
1.4 NEXT STEPS	10
2 INTRODUCTION	11
2.1 OBJECTIVES	11
2.2 PROCESS	12
2.3 DOCUMENT.....	13
2.4 TEAM.....	14
3 OUTPUTS AND COSTS	15
3.1 OUTPUTS	15
3.1.1 Availability	15
3.1.2 Capability	16
3.1.3 Ambience	17
3.1.4 Other important outputs	17
3.1.5 Prioritisation	18
3.2 COSTS	19
4 BENCHMARKING FRAMEWORK	21
5 FUNDAMENTAL ISSUES WITH REGARD TO COMPARABILITY	25
5.1 OUTPUT MEASURES.....	25
5.2 DIFFERENCES BETWEEN LINES	28
5.3 INTERPRETATION OF BENCHMARK MEASURES	28
6 BENCHMARKS: LEVELS 1 AND 2	30
6.1 LEVEL 1 BENCHMARKS	30
6.1.1 Availability	30
6.1.2 Capability	31
6.1.3 Asset health.....	31
6.1.4 Input specified projects.....	32
6.1.5 Ambience	32
6.1.6 Service points.....	33
6.1.7 Checks.....	33
6.2 LEVEL 2 BENCHMARKS	34
6.2.1 Availability	34
6.2.2 Capability	35
6.2.3 Asset health.....	35
6.2.4 Input specified projects.....	36
6.2.5 Ambience	36
6.2.6 Service points.....	37
6.3 SUMMARY	37

7	BENCHMARKS: LEVEL 3	39
7.1	CAPABILITY	39
7.2	INPUT SPECIFIED PROJECTS	41
7.3	LEVEL 3 ANALYSIS AS PART OF DRILL-DOWN AND DIAGNOSIS	44
7.3.1	Level 1 benchmarking.....	45
7.3.2	Level 2 benchmarking.....	46
7.3.3	Level 3 benchmarking.....	47
7.3.4	Diagnosis.....	48
8	DERIVING COSTS AND NORMALISING FOR SIZE	50
8.1	STEP 1: OBTAIN TOTAL COST DATA (TC)	50
8.2	STEP 2: CLASSIFY COSTS BY ASSET GROUP AND ASSET CLASS.....	50
8.3	STEP 3: EXCLUDE MAJOR PROJECTS FROM THE COST DATA.....	50
8.4	STEP 4: IDENTIFY MOST IMPORTANT ASSET CLASS COSTS	51
8.5	STEP 5: IDENTIFY A SIZE VARIABLE FOR EACH ASSET CLASS COST	51
8.6	STEP 6: ADJUST ASSET CLASS COSTS TO REFLECT DIFFERENCES IN SIZE.....	51
8.7	STEP 7: APPORTIONING COSTS TO OUTPUTS	52
8.8	ALTERNATIVE APPROACHES TO ADJUSTING FOR SIZE.....	58
9	WHOLE LIFE ASSET MANAGEMENT (WLAM)	59
9.1	DISCOUNTED PRESENT VALUES OF COST PER UNIT OF OUTPUT.....	59
9.2	ANALYSIS OF TIME SERIES	60
10	DATA SOURCES & OTHER ISSUES WITH REGARD TO COMPARABILITY	61
11	CONCLUSION & NEXT STEPS	63
	ANNEX 1: EXAMPLE PERFORMANCE REGIME SCORES	65
	ANNEX 2: ASSET GROUPS AND CLASSES	68
	ANNEX 3: THE BENCHMARKING FRAMEWORK - A HYPOTHETICAL EXAMPLE	72
	ANNEX 4: ADJUSTING FOR SIZE (NUMERICAL EXAMPLE)	80
	ANNEX 5: DIRECTORY OF ACRONYMS	83

Important Notice

This report was issued solely to the Office of the PPP Arbiter (OPPPA) for the sole purpose of assisting the OPPPA in developing an internal benchmarking framework which compares the costs and activities of the Infracos with respect to each other. PricewaterhouseCoopers LLP, its partners and staff neither owe nor accept any duty of care to any third party whether in contract or in tort (including without limitation negligence or statutory duty or howsoever otherwise arising) and shall not be liable in respect of any loss, damage or expense of whatsoever nature which is caused by any third party's reliance upon information derived from the report. If any third party wishes to rely upon the report or information derived therefrom, they do so entirely at their own risk.

1 Executive Summary

PricewaterhouseCoopers LLP (PwC) and Bovis Lend Lease (BLL) were appointed in April 2005 to advise the OPPPA on an appropriate approach to the internal benchmarking of Infraco performance. By “internal benchmarking”, we mean seeking to draw inferences about the cost effectiveness achieved by the Infracos in providing the services set out in the London Underground PPP contract, by comparing the costs incurred by the Infracos with each other.

We are about 4 months into an assignment which is likely to continue until towards the end of the year. Our work is divided into two Stages. During Stage 1 the main emphasis of our work has been to develop a benchmarking framework. This framework will then be piloted with real data during Stage 2.

This is the Executive Summary of our Stage 1 Final Report. From the outset, it is important to note that:

- Our objective has been to develop a benchmarking framework which will highlight areas which merit further consideration and investigation. In other words, the framework is not intended to generate accurate results which could be relied upon in isolation from further analysis or other sources of information. For example, we would expect the OPPPA to supplement findings from the internal benchmarking which is the subject of our assignment with the results of external benchmarking (comparing costs incurred by the Infracos with costs in analogous situations on other underground railways and/or infrastructure businesses).
- Our views on the benchmarking framework are evolving over time and should therefore be viewed accordingly. We welcome constructive comments on our approach and would expect such comments to improve the usefulness and relevance of the benchmarking framework.
- We would also expect the piloting which will be carried out during Stage 2 to reveal new insights and to result in the benchmarking framework being further developed.

In the rest of this Executive Summary, we consider:

- The nature of our task.
- How the internal benchmarking framework should operate.
- Some of the more complex issues which the framework will need to address.
- The next steps that are now required in order for us to progress with our work.

1.1 The nature of our task

The PPP regime is, by design, output based. This raises at least five key issues.

First, is whether our approach to comparing the Infracos should consider all of the possible different ways in which the Infracos could deliver these outputs. In other words, we need to consider whether the benchmarking framework should reflect a hypothetical world in which the Infracos are required to deliver the best output at the lowest costs possible and are unconstrained in how they do this.

In our view, such an approach would be inappropriate. What is required instead is an approach which is practical rather than hypothetical in nature, and which strikes an appropriate balance between:

- The unique characteristics of each Infraco. The implications of legacy asset conditions, the standards regime and existing contracts are likely to be important and will vary between the Infracos; and
- The insights that might nevertheless be revealed through an examination of the activities carried out, and the costs incurred, by the other Infracos.

Second, in our view, the output based nature of the regime means that the benchmarking framework should, wherever possible, also focus on outputs rather than inputs. This is because the framework should:

- Be capable of informing judgements that will be made at Periodic Review and which may concern outputs which are different to those which the Infracos are currently required to deliver.
- Use a manageable number of metrics in order not to be administratively burdensome. This could be extremely difficult to achieve if the framework were based on a comprehensive “bottom-up” analysis of costs per unit of intermediate output or costs per unit of input.
- Be anchored in the real world which, for the Infracos, is an environment where they are incentivised to provide, and focus their operations towards the delivery of, outputs.
- Make sense from a conceptual perspective. In our view, making links between costs and final outputs is a crucial part of deciding how to allocate scarce resources.

Third, is to decide what we actually mean by outputs. On the one hand, it would be possible to consider solely those final outputs characterised by the service performance regime (ie capability, availability and ambience). There are, however, a number of other important outputs which should also be considered, such as the requirement to meet specific asset health benchmarks, fault response requirements and the delivery of specific projects.

In our view, the framework should therefore consider:

- Availability.
- Capability.
- Asset health.
- Input Specified Projects.
- Ambience.
- Service points.

It should, however, be noted that it would be possible to extend the definition of outputs to include obligations relating to the standards regime and certain of the Infracos' existing

contracts. Going even further, one might also consider a number of wider issues such as potential incentives to deliver more or less output than would be suggested by the performance regime alone, together with activities that might be described as “good governance” (such as health and safety, investing in people, various corporate functions etc). While these are important issues, we consider them as inputs.

Fourth, is to consider the intricacies of the outputs referred to above, some of which are what we refer to as “negative outputs”. We define negative outputs as those outputs where, when using conventional measures, a higher figure is, all other things being equal, an undesirable outcome (eg Lost Customer Hours and Service Points). Negative outputs present fundamental issues with regard to comparability which, in our view, are best overcome by developing adapted measures of output (which are, however, based directly on existing output measures).

More specifically, we suggest measuring:

- Availability by a metric which we refer to as the Lost Customer Hours Score (LCHS); and
- Service points by a metric we refer to as the Service Points Score (SPS).

Other outputs are particularly difficult to compare between lines on a top-down basis. For example, projects intended to improve capability are often very large, long-term and “all or nothing in nature”. This makes it particularly difficult to link expenditure to output. Moreover, input specified projects are intrinsically line specific. We therefore suggest considering these two outputs (ie capability and input specified projects) on a relatively bespoke, disaggregated basis.

It is also important to note that many of these outputs are inextricably intertwined. Separating costs incurred in the delivery of one output from costs incurred in the delivery of others could be important but this is also likely to be complex.

Fifth, a primary focus of the benchmarking framework will be to analyse cost per unit of output. It is therefore important to be clear about what is meant by the term cost.

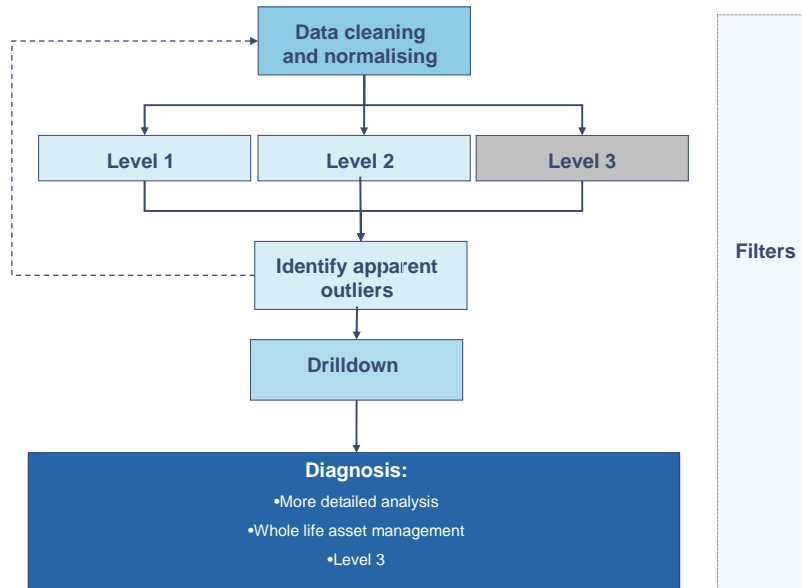
In this regard, we found it useful to:

- Distinguish between total costs and costs excluding major projects. This is important because of how we suggest the benchmarking framework should consider capability and input specified projects.
- Explicitly consider but distinguish between past, current and projected costs. This is important if the framework is to shed light on the Infracos’ approaches to whole life asset management.
- Distinguish between costs which are reported at a number of different levels. The framework should generally analyse costs at the level of line rather than Infraco. But it is also important to look at costs on a more disaggregated basis, and we think the best way of doing this is by also considering costs at the level of what we refer to as asset group and asset class.

1.2 How the benchmarking framework should operate

There are three key points to note about how the framework should operate. A high level summary of the benchmarking framework is provided in figure 1.

Figure 1: The benchmarking framework – A high level summary



First, is to ensure that the costs and outputs of each of the lines are reported and presented on a comparable basis. We refer to this process as data cleaning and data normalising and the challenges presented by this process should not be underestimated. On the one hand, for example, some of the adjustments (eg mapping the Infracos' currently reported costs to different asset groups and classes) are likely to be relatively straightforward. On the other hand, other adjustments (eg precisely how to measure variables such as LCHS and SPS, whether and how to adjust for differences in size and/or asset quality) could be more contentious.

Second, is to make use of a number of different types of benchmark. We draw a distinction between:

- Level 1 benchmarks which consider un-apportioned (ie total) costs per unit of output.
- Level 2 benchmarks which consider apportioned costs (ie costs relevant to the unit of output) per unit of output. There are two types of Level 2 benchmarks – those which refer to asset group costs and those which refer to asset class costs.
- Level 3 benchmarks which consider costs per unit of intermediate output and/or input (ie identifying the costs associated with particular activities rather than measuring costs in relation to the performance achieved in delivering outputs).

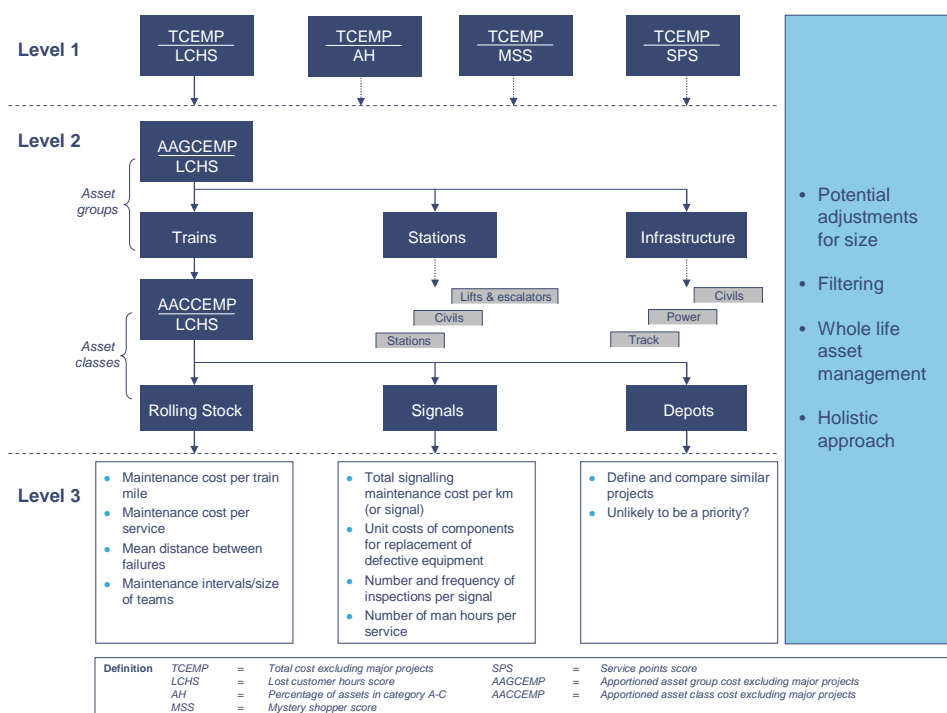
Each of these types of benchmark have strengths and weaknesses but they should, if used intelligently and in conjunction with each other, enable effective internal benchmarking to be carried out and valuable insights into the activities and costs of the Infracos to be generated.

Third, there is a bifurcation in how the framework should consider the outputs referred to above. With regard to availability, asset health, ambience and service points the approach should be to:

- Calculate a full set of Level 1 benchmarks, and a limited but relatively large number of Level 2 benchmarks which should be selected by using what we refer to as filters. In short, as there are so many possible Level 2 benchmarks, the framework should focus on those areas of expenditure which are largest in absolute terms and/or likely to have the greatest impact on costs.
- Ascertain whether there are any clear reasons for anomalous results which it might be reasonable to adjust for. The challenge here will be to work with the PPP Parties to identify and understand these anomalies without carrying out extensive analysis, and then to make only those adjustments which are absolutely necessary.
- Generate a revised set of benchmarks and select a number of areas which it would be particularly useful to analyse in more detail. For example, if our revised set of benchmarks indicate that a specific line has a particularly high rolling stock cost per unit of availability (and that this expenditure is significant and likely to have a high impact on performance), then it might be useful to drill-down into this area in an attempt to diagnose the situation. This process would make use of a range of Level 3 benchmarks which might include maintenance cost per train mile, maintenance cost per service, mean distance between failures, maintenance intervals etc. It could also draw on a range of other diagnostic techniques such as qualitative audits, bottom-up cost modelling and assessments of market based references for price. There are many other possible situations, but the key point to note is that it will be important to drill-down into and attempt to diagnose apparently low, as well as high, costs per unit of output.

A summary of how the benchmarking framework might work in practice for these four output variables, with reference to a specific hypothetical example which is described in the main body of our report, is included as figure 2.

Figure 2: The benchmarking framework in practice – A hypothetical example



With the other outputs, we would carry out a more bespoke and disaggregated analysis which is Level 3 in nature from the outset.

When considering:

- Capability, this is likely to involve some internal benchmarking of the costs involved in signalling projects and a limited amount of external benchmarking of rolling stock projects (even though, strictly speaking, this is beyond the scope of our work).
- Input specified projects, this is likely to involve relatively detailed benchmarking of actual and projected costs for station upgrade projects. Here, the challenge will be to categorise projects, make appropriate comparisons and, if necessary, carry out what we refer to as nature of work analysis.

However, for both of these outputs, the principle of focussed analysis should prevail. In other words, we will calculate an initial set of benchmarks, identify whether there are any obvious anomalies, recalculate the benchmarks if necessary and then drill down into, and attempt to diagnose, particularly interesting results.

Finally, although for the purpose of explanation we have set out in a somewhat mechanistic way the approach we propose should be adopted and the key ratios which we propose should be calculated as part of the internal benchmarking process, we believe that in practice much will be learned simply by compiling and observing the constituent measures (outputs and costs) over time. By the time of the first Periodic Review 7 years' of data will be available, allowing trends as well as levels to be observed. Furthermore, once the OPPPA has undertaken first and second round data cleaning and normalising several times, it will have insights which should make the use of the internal benchmarking framework more effective by reducing the time taken both to calculate measures and to draw inferences.

1.3 Complex issues

While we have managed to address many complex issues which are likely to arise with the benchmarking framework, there are at least five others which require more detailed discussion with the PPP Parties.

First, is precisely how to calculate our suggested measures for availability and service points (ie LCHS and SPS).

Second, is whether and how to make adjustments for inherent differences between lines. As it stands we consider that it might be appropriate to normalise for the size of different lines. We also explain how this normalisation might be carried out. However, we suggest that other potentially important issues, such as differences in asset quality or complexity which we found it particularly difficult to quantify in an unambiguous manner, should be explored during drill-down and diagnosis.

Third, is whether and how it is possible to apportion costs between the different outputs and which areas of cost are likely to have the largest impact on these outputs. We have made provisional assumptions in both of these areas which it will be crucial to discuss with the PPP Parties and, if appropriate, modify. It would be particularly useful for these assumptions to be informed by the various techniques which the Parties use (or have used in the past), such as root cause analysis.

Fourth, is how to capture the Infracos' potentially different approaches to whole life asset management. This is important because, in many cases, the Infracos are incurring costs on large scale investments which are long-term in nature. For example, certain activities are substantial in terms of current costs but are only expected to deliver actual improvements in

performance in the relatively distant future. In this type of situation, an analysis which focuses solely on current costs and performance could indicate that an Infraco is relatively high cost but low performance, despite the fact that over a longer time frame the Infraco might actually be low cost and high performance. There are two ways the benchmarking framework could address this. On the one hand, we could make some potentially complex calculations of the present value of a profile of (past and) projected costs per unit of output. This approach would be conceptually appealing but might result in a certain amount of “spurious precision”. On the other hand, the framework might rely on more straightforward plots and comparisons of these profiles to see what insights this provides. This is less elegant but might actually prove to be more useful.

Fifth, is precisely how to interpret differences between the costs per unit of output of the various lines. Our expectation is that this could provide important insights into the economy and efficiency of the Infracos¹. But eliciting these insights will require judgement, the development of which we would expect to evolve over time.

1.4 Next steps

In order for us to progress our work, the crucial requirements now are to:

- Obtain constructive and timely input from the PPP Parties on the key issues highlighted in this document and in the main body of our report.
- Commence piloting the framework because, to a very large extent, it is only when the framework is populated with real data that we will see what insights it is capable of generating.

DRAFT

31st August, 2005

PricewaterhouseCoopers LLP / Bovis Lend Lease

¹ Although we realise that these are complex concepts and that defining them is beyond the scope of our work.

2 Introduction

2.1 Objectives

PricewaterhouseCoopers LLP (PwC) and Bovis Lend Lease (BLL) were appointed in April 2005 to advise the OPPPA on an appropriate approach to the internal benchmarking of Infraco performance. In giving guidance or directions at a Periodic Review, the Arbiter is required to consider the costs which would be incurred by a Notional Infraco in meeting its obligations. The PPP Contract contains guidance to the Arbiter that benchmarking between Infracos is a relevant factor in his consideration.

There are a number of other reasons why benchmarking between Infracos will be relevant to the Arbiter. More specifically, the Arbiter:

- Has a statutory duty to promote economy and efficiency.
- Has to consider the Metronet Infracos' overall economy and efficiency in the context of the "annual certificate", if this is applied for.
- Has to consider the economy and efficiency of the Infracos in the event of an Extraordinary Review.

The Arbiter has previously commissioned work from Cambridge Economic Policy Associates (CEPA) on how inter-Infraco comparisons might be made². It is intended that the PwC/BLL work should build on this to develop an internal benchmarking framework.

At the outset, it is important to note that:

- The primary purpose of developing an internal benchmarking framework is to inform the Arbiter and assist him in discharging his functions at Periodic Review. While it is possible that some of the information obtained from the internal benchmarking framework will be useful in other contexts (e.g. Metronet Annual Review, Extraordinary Review etc), informing these other contexts is not the primary purpose of developing an internal benchmarking framework.
- Internal benchmarking has a number of inherent limitations. For example, it is difficult to make meaningful comparisons between entities as heterogeneous as underground infrastructure businesses/assets. Moreover, by its very nature, an internal benchmarking framework does not take full account of a range of potential comparators which are external in nature but potentially extremely useful (eg underground systems in other countries, infrastructure businesses in other sectors etc). It will therefore be important for the framework, when made fully operational, to be used with judgement and in combination with other sources of information.
- The framework is intended to develop measures that are sufficiently accurate to highlight areas which merit further consideration and investigation, rather than measures which could be considered to be robust on a stand alone basis.

² "Report to the London Underground PPP Arbiter: Approaches to Benchmarking Infraco Efficiency & Performance", CEPA, July 2003.

- Our work is still at a relatively early stage. We therefore welcome constructive comments on this report, as we would expect this to improve the usefulness and relevance of the benchmarking framework.

It is also important to note that the issues that should be addressed by an internal benchmarking framework are numerous and complex. More specifically, the most important requirements for the framework are that it should:

- Inform the guidance or directions which the Arbiter may be required to give in relation to, in particular, Periodic Review. This means that the framework should be capable of considering not only the current performance regime but also the implications of potential changes to the regime and the outputs which the Infracos may be required to deliver.
- Reflect the impact of the Infracos' obligations and largely output based incentive regimes and make explicit linkages between the costs incurred by, and outputs delivered by, the different Infracos.
- Shed light on differences between Infracos in the ways in which they respond to their obligations and the output based incentive regime. In particular, to the extent appropriate, it should incorporate the different approaches of the Infracos to the concept of "whole life asset management" and to their supply chains.
- Inform the Arbiter's understanding of the Infracos' relative performance, how this is changing over time and how "leading indicators" might be used to understand likely future performance.
- Build on the work already carried out by CEPA.
- Minimise the number of metrics where comparisons are made, in order to reduce administrative burden.
- Strike an appropriate balance between being robust and practical.
- Be flexible and capable of targeting specific differences between the Infracos which merit further investigation and how this investigation might be carried out.

2.2 Process

Our work is divided into two Stages:

- In Stage 1, the main emphasis of our work has been to develop a benchmarking framework.
- In Stage 2, the focus of our work will shift to piloting this benchmarking framework and interpreting what this pilot could indicate about the activities and costs of the Infracos.
- There will, however, be a degree of overlap between these two Stages. The purpose of this overlap is to ensure that the framework developed in Stage 1 is informed by the data that is available and that will be drawn upon during Stage 2, and to ensure that the piloting and interpretation carried out in Stage 2 draws effectively on the

conceptual thinking carried out during Stage 1. There will inevitably be a degree of iteration between design of the framework and data analysis/interpretation.

One of the key aims of our approach has been to work closely with the OPPPA and the PPP Parties. In order to do this, we held:

- An initial meeting with Mike Woods (technical manager, OPPPA) and James Le Couilliard (commercial adviser, OPPPA) on 7th April 2005. The purpose of this meeting was to obtain a better understanding of what information is routinely available to the Arbiter.
- A kick-off meeting with Chris Bolt (the PPP Arbiter), Gaynor Mather (director, OPPPA), Mike Woods and James Le Couilliard on 19th April 2005. The purpose of this meeting was to commence our work formally and to agree a number of broad parameters within which we would operate.

We then submitted an Interim Report to OPPPA on 25th May 2005. The purpose of this report was to outline our emerging views on:

- The key issues that should be addressed within a benchmarking framework.
- The key factors to be considered when addressing the links between cost, work undertaken and performance within the LUL PPP.
- Our proposed work programme to the completion of Stage 1.

We then:

- Participated in the inaugural meeting of the OPPPA's benchmarking Steering Group on 30th June 2005.
- Subsequently held bilateral meetings with each of the Parties in order to explain what we were doing and why, and for us to take account of the Parties' views.

We have also:

- Held a number of conference calls and/or meetings with OPPPA in order to discuss progress with, and issues arising, throughout our work.
- Carried out a high level review of a number of key sources of information which have been provided to us as a result of the above meetings and discussions.
- Drawn on our existing understanding of the London Underground PPP, other industries which are subject to what might loosely be referred to as economic regulation, and relevant engineering practices/issues.
- Carried out our own independent research and analysis.

2.3 Document

This document is our Final Report for Stage 1 of our work:

- In section 3, we provide a high level explanation of why, as much as possible, the benchmarking framework should be based on outputs. We then outline our views on which measures of output the framework should use. Finally, we make a number of relatively high level observations about what we mean by the term “cost”. This section is important because it helps to lay a foundation for understanding whether and how it will be possible to make comparisons of costs per unit of output between the Infracos.
- In section 4, we provide a high level description of how the benchmarking framework should operate. This section introduces a number of key concepts, including: the three different levels of benchmark; data cleaning and normalisation; filters; drill-down and diagnosis.
- In section 5, we describe a number of fundamental issues with regard to comparability. These issues arise because of the nature of the outputs on which the proposed framework is based, inherent differences between the lines, and how difficult it can be to interpret what different levels of cost per unit of output actually imply for benchmarking of relative performance.
- In sections 6 and 7, we define what benchmarks should actually be incorporated into the framework. Section 6 covers benchmarks at those levels where we make links between costs and final output (ie Levels 1 and 2). Section 7 then covers how the framework should use Level 3 analysis, both to consider costs in areas where it is particularly difficult to make clear links between costs and final output, and to understand better the results from Level 1 and Level 2 analysis.
- Sections 8 and 9 are relatively technical in nature. Section 8 describes how it might be appropriate to derive various measures of costs and to adjust (or to “normalise”) Infraco cost data to reflect differences between the size of the various lines. Section 9 describes two possible ways in which the benchmarking framework might address the Infracos’ approaches to whole life asset management.
- Section 10 outlines what we would expect to be the main sources of data for the benchmarking framework and what issues with regard to comparability might arise while using this data.
- In section 11, we summarise our conclusions and outline the most important requirements in order for us to progress with our work during Stage 2.

2.4 Team

The PwC/BLL team responsible for this report are:

- Tim Ogier – Project partner.
- Daniel Hanson – Project leader.
- Wikash Bhagwanbali – Economic adviser.
- Steve Turner – Engineering adviser.
- Julian Williams – Engineering adviser.

3 Outputs and costs

In this section we:

- Provide a high level explanation of why, as much as possible, the benchmarking framework should be based upon outputs. We then outline our views on which outputs the framework should use.
- Provide some high level comments on what we mean by the term “costs”.

Taken together these two areas are of crucial importance in laying a foundation for understanding whether and how it will be possible to make comparisons of costs per unit of output between the Infracos.

3.1 Outputs

In our view, wherever possible the internal benchmarking framework should focus on outputs rather than inputs.

This is because the framework should:

- Be capable of informing judgements that will be made at Periodic Review and which may concern outputs which are different to those which the Infracos are currently required to deliver.
- Use a manageable number of metrics in order not to be administratively burdensome. This could be extremely difficult to achieve if the framework were based on a comprehensive “bottom-up” analysis of costs per unit of intermediate output or costs per unit of input.
- Be anchored in the output based nature of the PPP contract, performance regime, and operations of the Infracos.
- Make sense from a conceptual perspective. In our view, making links between costs and final outputs is a crucial part of deciding how to allocate scarce resources.

In the case of the PPP Contract, the final outputs are defined by the performance regime. The three main elements of service performance are availability, capability and ambience. We now consider each of these outputs in turn.

3.1.1 Availability

Availability is a measure of day-to-day service reliability.

It is measured by recording any disruption to the service lasting two minutes or more (an “incident”) which is attributable to an Infraco. The actual length of each incident is multiplied by a factor which varies by location and time of day to give an estimate of the total amount of delay caused by that incident to customers across the network. Availability is expressed in terms of Lost Customer Hours (LCH).

Key points to note about availability are that:

- The Infracos have performance benchmarks for LCH. These performance benchmarks are based on actual performance during the 17 months prior to Transfer.
- The Infracos receive bonus payments in the event that performance is better than these benchmarks (subject to a “cap” above which further improvements are not incentivised³) and incur abatements if performance is worse than benchmark.
- Unacceptable levels of LCH are defined for each line. If performance is worse than these unacceptable levels, the Infracos are subject to harsher abatements and/or the possibility of Corrective Action Notices.
- The standard measure of availability, LCH, is what we refer to as a “negative output” – a higher figure for LCH is, all other things being equal, a worse outcome. This raises fundamental issues with regard to comparability which we discuss in section 5.
- We have therefore developed an alternative measure of availability which, although based on LCH and therefore the information that is reported as a matter of course in the PPP regime, overcomes these comparability issues. We refer to this alternative measure of availability as Lost Customer Hours Score (LCHS).

3.1.2 Capability

Capability is a measure of the practical capacity of each line.

Infracos can improve their capability score by adding raw capacity (the journey time capability score) and by enabling that capacity to be used more flexibly and reliably (the service control and service consistency scores). Capability scores are measured in minutes.

This measure also raises some fundamental issues with regard to comparability:

- Capability scores are inherently difficult to compare between Infracos or lines. As we understand it, for example, a capability score of 20 minutes on one line cannot be assumed to be superior to a score of 30 minutes on another line.
- Projects undertaken in order to improve capability are large and extremely long-term and “all or nothing” in nature⁴. This makes it particularly difficult to obtain a robust understanding of links between costs and final output.
- Capability is also another “negative output”.

Whilst capability is a crucial part of the PPP Contract and the outputs which the Infracos are required to deliver, and should therefore be an integral part of the benchmarking framework, because of these issues we propose that within the benchmarking framework capability should be considered using what we refer to as Level 3 analysis. Our approach to capability is described in section 7.

³ The cap is zero for LCH but not for the other service performance measures.

⁴ Ie they have to be completed before delivering benefits. A good example would be the introduction of a new fleet of rolling stock which is expensive, can take an extremely long time to procure and deploy, will only improve a line's capability score when it is actually in operation.

3.1.3 Ambience

Ambience is a measure of the quality of the travelling environment.

Ambience is measured through a quarterly Mystery Shopping Survey (MSS). LUL pays for the survey, and is responsible for appointing a suitably qualified survey organisation, after consulting with the Infracos. Ambience scores are measured as a percentage/an index. The Infracos have performance benchmarks, unacceptable performance levels and caps for ambience in much the same way as they do for availability⁵.

A summary of benchmark, unacceptable and cap scores for availability, capability and ambience for each of the lines is provided in Annex 1.

3.1.4 Other important outputs

There are a number of other important outputs which should also be considered to form part of the performance regime.

These include:

- The requirement to meet specific asset health benchmarks. Every asset on the system is classified in terms of its asset condition (i.e. bands A-E)⁶. The Infracos are required to ensure that a specific proportion of its assets are categorised as being in condition A-C by certain time limits.
- The requirement to deliver specific outputs, most notably station and train refurbishments. We refer to these outputs as “input specified projects”.
- Fault response requirements, which are measured by the service points regime. Service points are incurred by Infracos for failure to restore facilities and rectify faults within a given time. Each Infraco is allowed to incur a certain number of service points each period, above which abatements are levied.

Each of these other outputs raises issues with regard to comparability. For example:

- The Infracos have specific and contractually defined targets for asset health. In our view, the benchmarking framework should therefore consider asset health by:
 - Analysing the percentage of all of a line's/an Infraco's assets in categories A-C. This should provide a very high level way of assessing output from the perspective of asset health.
 - Focussing attention on the percentage of a line's/an Infraco's station/civils assets in categories A-C because these are the assets which, if in poor condition, might not otherwise be identified by the framework⁷. We therefore consider this to be the most useful and potentially insightful measure of asset health.

⁵ One key difference is that ambience is measured and incentivised on the basis of Infracos as a whole rather than each line, but we have been advised that data on ambience is available on a line by line basis.

⁶ With the exception of lifts & escalators (and possibly grey assets).

⁷ For example, the health of rolling stock should ultimately affect and therefore be measurable (albeit indirectly) from, in particular, capability. Similarly, we would expect the health of track assets to affect availability, etc.

- Monitoring a number of potentially important “leading indicators”, such as age of rolling stock and track. This could prove to be a useful cross-check to how the framework should be able implicitly to monitor the health of other assets (eg rolling stock through capability, track through availability etc).
- Input specified projects are inherently line specific. In our view, the framework should therefore assess input specified projects by using what we refer to as Level 3 analysis. In other words, input specified projects should be treated in much the same way as capability.
- Service points is another negative output which is also inherently difficult to compare between lines. We therefore consider service points by defining an alternative measure which we refer to as “service points score” (SPS). In other words, service points should be treated in much the same way as availability.

As can be seen from the above discussion, the conceptual issues involved in considering these outputs are significant. But, to the extent that this is possible, we consider that these are the outputs on which the internal benchmarking framework should be based.

3.1.5 Prioritisation

It will also be important for the framework to be focussed. We have therefore attached priorities to the outputs set out above. In other words, we consider that some of them should be analysed in more detail than others.

After consultation with OPPPA, we prioritised the outputs in the following order:

- Availability.
- Capability.
- Asset Health.
- Input Specified Projects.
- Ambience.
- Service Points.

Finally, we also consider that it will be important for the benchmarking framework to include a number of “checks” to ensure that the Infracos are compliant with their contract and/or expectations in a number of key areas (most notably asset health and safety).

These checks should:

- Ensure that an Infraco does not manage to achieve a “superior” level of performance (as measured by cost per unit of output) by taking inappropriate “short-cuts” which might not otherwise be identified by the framework.
- Be of use to the Arbiter in any case in making high level assessments of whether or not Infracos are being economic.

- Be as simple as possible. One possibility would therefore simply be to monitor the number of Corrective Action Notices, but our views on this issue are evolving over time and we would welcome comments from others.

3.2 Costs

Having obtained an understanding of what we mean by the term “output”, it is also important to be clear about what is meant by the term “cost”.

There are five key points to note in this regard.

First, it is useful to distinguish between Infraco total costs (TC) and Infraco total costs excluding costs incurred for major projects⁸ (TCEMP). This is because of our proposal that the benchmarking framework should consider two of the key outputs (ie capability and input specified projects) using what might be described as a bottom-up approach based on Level 3 analysis. We therefore exclude the costs incurred for major projects from our analysis of cost per unit of the other outputs.

Second, we explicitly consider and distinguish between past, current and projected costs. This will be important if the framework is to shed light on the Infracos’ approaches to whole life asset management. We discuss this concept in more detail in section 9.

Third, we distinguish between costs which are reported at a number of different levels. In this regard, wherever possible we consider it appropriate to focus our analysis at the level of lines (ie Bakerloo, Central, Victoria etc), although we realise that it is questionable how far each of the lines can be regarded as independent of each other (because of inherent similarities, their common history under LUL management/ownership and the fact that they are all ultimately run by Tube Lines or Metronet). The main alternative to this approach would be to carry out analysis at the level of Infraco or parent company/consortium but this would provide us with a smaller number of comparators.

We also find it useful to distinguish between asset:

- Group. We consider the main asset groups to be:
 - Train systems (rolling stock, signalling and depots).
 - Station systems (stations, lifts & escalators, station civil).
 - Infrastructure systems (track, power and non-station civil).
- Class. We consider the main asset classes to be:
 - Rolling stock
 - Signalling
 - Stations
 - Track
 - Power

⁸ ie those costs incurred for capability upgrades and input specified projects.

- Civil
 - Depots
 - Lifts & escalators
 - Admin & other / “on-costs”.
- Category - descriptions of which are highly detailed.

We provide a more detailed description of what we mean by asset groups, asset classes and asset categories in Annex 2.

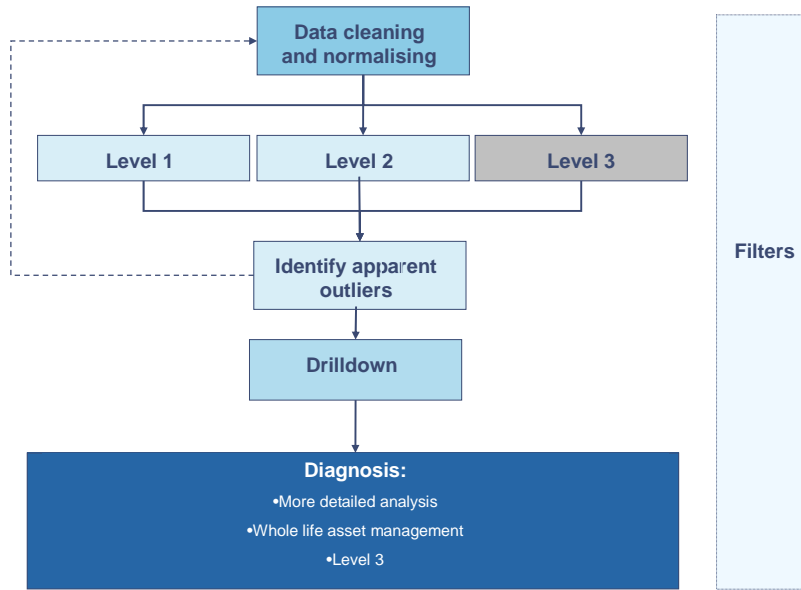
Fourth, we note that the benchmarking framework is intended to consider “above the line costs”. We were advised by OPPPA not to include financing costs or taxation in our analysis.

Fifth, we assume that the framework will be populated with annual data (or information based on 13 reporting periods each of which lasts for 4 weeks). This is necessary to ensure that development and use of the framework is manageable, and to ensure that it will be able to draw on reliable/audited information (if appropriate). We do, however, realise that this approach may have the result of masking potentially important information (if, for example, certain of the Infracos’ activities change radically during the course of one year). This should therefore be considered within the framework, during what we refer to as “drill-down” and/or “diagnosis”. It is also important to note that the amount of data used to populate the framework will increase over time, as more information become available. This will be particularly important to how the framework will consider whole life asset management.

4 Benchmarking framework

In this section we provide a high level description of how the benchmarking framework should operate.

Figure 1: The benchmarking framework - A high level summary



1

There are seven key points to note.

First, as explained previously, our objective has been to develop a benchmarking framework which will highlight areas which merit further consideration and investigation. In other words, the framework is not intended to generate accurate results which could be relied upon in isolation from further analysis or other sources of information.

Second, the framework should make use of benchmarks at three different levels (ie Level 1, Level 2 and Level 3), each of which have different strengths and weaknesses but which if used together should enable effective benchmarking to be carried out, and valuable insights into the activities and costs of the Infracos to be generated.

At the highest level, the framework will consider un-apportioned (ie total) cost per unit of final output. We refer to this as “Level 1”.

The main advantages of Level 1 benchmarking are that:

- It focuses directly on cost per unit of final output.
- Some Level 1 benchmarks⁹ are relatively straightforward to calculate.

⁹ In particular, Level 1 benchmarks using costs which are unadjusted for differences in line size.

- It is administratively manageable because it requires the calculation of a relatively small number of benchmarks.

The main disadvantages of Level 1 benchmarking are that it:

- Is likely to be better at informing assessments of the levels of Infracos costs rather than whether the Infracos are actually carrying out appropriate activities. In other words, Level 1 analysis is more suitable for making high level assessments of efficiency rather than economy¹⁰.
- May generate results which appear to be inconsistent and difficult to understand. For example, it is conceivable that an Infraco/line could have a high cost per unit of one output and a low cost per unit of another output, but it may be unclear what this implies.
- Likely to leave many questions unanswered, in part because of the above.

The framework should therefore also incorporate a more detailed analysis of apportioned costs (ie costs relevant to the unit of output) per unit of output. We refer to this as Level 2 benchmarking. We also define two different types of Level 2 benchmarks – those which consider asset group costs and those which consider asset class costs.

The main advantages of Level 2 benchmarking are that:

- It is also directly focussed on cost per unit of final output.
- It is likely to provide far more insight into the costs and activities of the Infracos than would level 1 analysis.

The main disadvantages of Level 2 analysis are that:

- It could be difficult to calculate these benchmarks in a manner that is completely uncontroversial. We have made some provisional assumptions with regard to how much of expenditure in each asset class/group might be apportioned to each output given a view of the impact that the level of spending in each asset class/group is likely to have on that particular output. One of the key challenges in making the framework operate effectively in practice will be to obtain a better understanding from the PPP Parties of whether our provisional assumptions are reasonable and/or should be modified. This is likely to be an ongoing and evolving process. Our approach to apportioning costs and the likely impact of costs on output is described in section 8.
- There are many potential Level 2 benchmarks. In our view, the framework should therefore make use of what we refer to as “filters” in order to decide which benchmarks should be calculated and analysed. We would do this by calculating benchmarks for those areas of expenditure which are largest in absolute terms and/or which are likely to have the most significant impact on output. This should ensure that the framework and associated analysis is focussed, although it means producing results which are not fully comprehensive.

¹⁰ Although we realise that economy and efficiency are complex concepts and that a definition of them is beyond the scope of our work.

The framework should also use what we refer to as Level 3 benchmarks. Level 3 benchmarks are primarily concerned with costs per unit of intermediate output or input (ie identifying the costs associated with particular activities rather than measuring costs in relation to the performance achieved in delivering outputs).

In our view, the framework should use Level 3 analysis in order to:

- Understand better some of the costs per unit of final output identified by our analysis at Levels 1 and 2. Level 3 analysis would therefore be an integral aspect of what we refer to as “drill-down” and “diagnosis”.
- Analyse the costs incurred by the Infracos in those areas which are likely to be particularly difficult to assess from the perspective of cost per unit of final output (ie capability and input specified projects).

Third, it will be necessary to carry out an initial, or first round, of data cleaning and normalisation:

- Data cleaning refers to adjustments to the costs of the various lines in order to ensure that these costs are reported on a like for like basis. A detailed explanation of these adjustments will be provided in Stage 2, but we make a number of comments about what adjustments may be required in later sections of this report. For example, section 8 explains that it will be necessary to ensure that we have a consistent mapping between the costs in the Infracos’ financial models and/or AAMPS, and the definitions of asset groups and asset classes that we suggest should be incorporated into the framework.
- Normalisation refers to adjustments that may be required to the costs of the various lines in order to ensure that comparisons can be made on a like for basis. As it stands, we consider that it is likely to be appropriate to adjust costs for differences in line size¹¹. We also realise that it could be appropriate to normalise for a range of other factors such as asset quality/complexity and nature of line (ie deep-tube or sub-surface), but we found it difficult to quantify these factors in a relatively unambiguous manner. We are therefore currently of the view that these are issues best explored during what we refer to as drill-down and diagnosis.

Fourth, we would then make some initial calculations of the benchmarks. With regard to Level 1 and Level 2 benchmarks, we would expect this to show that the lines have significantly different costs per unit of output. Through discussions with the PPP Parties, we would then attempt to identify whether or not there are any clear reasons for these differences and whether the data should be adjusted for these reasons. This is likely to be an iterative, but hopefully not unduly protracted, process at the end of which we would produce a revised (and potentially final) set of benchmarks (effectively this would be a second round of data cleaning and normalisation). We would then select a sample of these benchmarks for further analysis at Level 3 when we “drill-down” and attempt to “diagnose” reasons for these levels of, and differences between, costs. The sample of benchmarks which are subject to drill-down and diagnosis would be selected in order to ensure that we examine “outliers” (ie unusually high or low cost per unit of output).

With regard to those areas where we start off with Level 3 analysis, we would carry out a similar process (ie initial calculations, identification of “obvious” reasons for differences, recalculation of benchmarks, further analysis of particularly interesting results etc).

Fifth, it is important to define what the “end product” of the framework will be. With regard to:

¹¹ Although we recognise that whether and how these adjustments should be made could be the subject of debate.

- Level 1 benchmarks, the framework will generate absolute values for unapportioned cost per unit of output for each line and plots of how these absolute levels of cost per unit of output have changed, and are likely to change, over time. The framework may also generate discounted values of projected cost per unit of output.
- Level 2 benchmarks, the framework will generate absolute values for apportioned cost (at the level of asset group and asset class) per unit of output for each line and plots of how these absolute levels of apportioned cost per unit of output have changed, and are likely to change, over time. The framework may also generate discounted values of projected cost per unit of output.
- Level 3 benchmarks, the framework will produce a bespoke analysis of capability and input specified projects, and insight into why some lines have relatively high or relatively low scores for the Level 1 and Level 2 benchmarks.
- All of the above when considered together, should provide OPPPA with a better understanding of the Infracoss' costs and, most importantly, where it would be appropriate to focus attention if the OPPPA needs to find out more.

Sixth, we considered the use of a range of sophisticated quantitative techniques (eg various forms of econometrics etc) but came to the view that these techniques were unlikely to be particularly useful, given the specifics of the PPP regime (not least the number of outputs which the Infracos deliver, the limited number of lines which can be used as comparators and the relatively short space of time since Transfer).

Seventh, is that we considered carefully whether the primary focus of the framework should be to:

- Analyse absolute levels performance (and therefore cost per unit of output) and make comparisons of these absolute levels between different lines; or
- Analyse the rate of change of performance (and therefore cost per unit of output) and make comparisons of how a particular line has performed against itself.

The former might be referred to as focussing on inter-line comparisons and the latter as focussing on intra-line comparisons. This distinction is important because a number of PPP Parties made it clear to us that an intra-line approach is more consistent with the PPP Contract and how they actually operate on a day-to-day basis, and because the lines are so heterogeneous in nature.

However, in our view an inter-line approach is more consistent with the concept of internal benchmarking and with the likely use of the benchmarking framework at Periodic Review. Moreover, an intra-line approach could involve protracted debates over what the appropriate "starting point" for comparisons should be. We have therefore focussed on developing a benchmarking framework which will enable inter-line comparisons to be made. Notwithstanding this, as the framework will produce time series of how absolute levels of cost per unit of output of each line change of time, it should also be able to shed insight into the performance of particular lines over time and key concepts such as the "marginal cost" of changes to performance/output.

5 Fundamental issues with regard to comparability

While developing the framework, we encountered a number of fundamental issues with regard to comparability which arise as a result of:

- The outputs on which the benchmarking framework should be based.
- Inherent differences between the lines.
- Difficulties in interpreting what different values of benchmarks (ie costs per unit of output) might mean.

5.1 Output measures

As discussed in section 3, in our view the benchmarking framework should be based on the following outputs:

- Availability.
- Capability.
- Asset Health.
- Input Specified Projects.
- Ambience.
- Service Points.

With the exception of Input Specified projects, all of the above outputs are measured using metrics which could, at first sight, be used for inter-line comparisons. However, availability, capability and service points are what we refer to as “negative outputs” (ie an output where, on the basis of conventional metrics, a higher score represents a worse outcome).

For example, the metric used to measure availability is LCH. In simple terms, Infracos receive bonuses for reducing their LCH, and incur abatements if their LCH increases. The metric for capability is journey time minutes and, as with availability, Infracos are rewarded for reducing this metric. Service points are allocated for certain failures by Infracos to meet their contractual obligations, according to a tariff set out in the contract. The lower the score for service points the better.

Benchmarking negative outputs introduces a complexity which is best highlighted using a simple example (see Scenario 1 below).

Scenario 1	Line A	Line B
Cost	50	50
LCH	100	200
Cost/LCH	0.5	0.25

The table above shows hypothetical cost and LCH data for two lines. We assume that Line A and Line B are identical with the sole exception that line B incurs twice the number of LCH.

Based on the data in the table above, it might be reasonable to conclude that Line A is the better performer, as it generates less LCH than Line B (and incurs identical costs). When comparing lines by using a benchmark metric (ie cost per LCH), this might lead us to think that, for negative outputs, a high cost per unit of (negative) output represents superior performance.

However, we can see why this would be inappropriate by considering a slightly different example (see Scenario 2 below). In this example, Line A and Line B are again assumed to be identical, with the sole exception this time being that Line B's costs are twice as high as those of Line A.

Scenario 2	Line A	Line B
Cost	50	100
LCH	100	100
Cost/LCH	0.5	1.00

In this example, it might be reasonable to conclude that Line A is the better performer because it is producing the same output as Line B but with half the costs. However, if we were to compare the lines on the basis of a cost per unit of output, we can see that Line A's cost per unit of LCH is half that of Line B. In other words, in this situation a low cost per unit of output represents a superior performance (which is the opposite of our finding in scenario 1).

There are at least two main ways of dealing with this problem. The benchmarking framework could restrict itself to making comparisons between lines which produce relatively similar levels of output. But this would have the effect of severely restricting our number of comparators (which is already very limited anyway). Our preference is therefore to adopt the second solution, which is to convert the negative output measure (ie LCH) into a positive output measure (i.e. an output where more is preferable to less). However, this is far from straightforward.

One possibility would be to estimate the maximum number of LCH that a line could incur in a given period and subtract actual LCH from this figure. Setting aside for a moment the question of how it might be possible to derive a figure for maximum LCH, this would enable us to arrive at a figure that we might refer to as Lost Customer Hours Score (LCHS). Replacing LCH with LCHS would overcome the issues described above. This is explained in the simple examples below.

The tables (Scenario 3 and 4) below are identical to the tables in Scenario 1 and 2, respectively. The only difference is the introduction of an estimate of Maximum LCH and the use of LCHS. In both scenarios, we have a cost per unit of output benchmark (ie cost/LCHS) where a lower figure is "better".

Scenario 3	Line A	Line B
Cost	50	50
Maximum LCH	500	500
LCH	100	200
LCHS	400	300
Cost/LCHS	0.13	0.17

Scenario 4	Line A	Line B
Cost	50	100
Maximum LCH	500	500
LCH	100	100
<i>LCHS</i>	400	400
Cost/LCHS	0.13	0.25

However, the use of LCHS rather than LCH introduces another issue which we refer to as the “gearing effect”. This is best described using the example below (ie Scenario 5).

Scenario 5	Line A	Line B	% Diff (line A rel. to line B)
Cost	50	40	+25%
Maximum LCH	1000	1000	
LCH	100	200	-50%
<i>LCHS</i>	900	800	+12.5%
Cost/LCHS	0.055	0.05	

In Scenario 5, Line A has higher costs but fewer LCH than Line B (ie Line A has a higher LCCHS than Line B). If we were to use a simple comparison of cost per unit of output (cost/LCHS), we would conclude that Line B is a better performer than Line A. But it is unclear whether this would be appropriate. It might, for example, be argued that Line A is better because it delivers half as much negative output (ie LCH) as Line B for only 25% more cost. It is important to note that this would likely become an increasingly important issue if we were to populate the model with real data. This is because the maximum LCH (defined, say, by calculating the LCH that would result from full line and station closures) is likely to be very much greater than the LCH figures delivered by the Infracos.

One way of dealing with this issue would be to adjust the maximum LCH in such a way that it better represents the maximum LCH that might realistically be delivered by the Infracos. For example, we might:

- Multiply the maximum LCH by an X factor which would have to be selected by the OPPPA in conjunction with the PPP Parties (see Scenario 6); or
- Simply use the unacceptable level of performance as defined by the performance regime.

Scenario 6 shows how the use of an X-factor or an adjusted figure for maximum LCH could address the gearing effect.

Scenario 6	Line A	Line B	% Diff
Cost	50	40	+25%
Maximum LCH	1000	1000	
X Factor	75%	75%	
<i>Adj. Maximum LCH*</i>	250	250	
LCH	100	200	-50%
<i>LCCHS</i>	150	50	+300%
Cost/LCHS	0.33	0.80	

Note: * Assumed to be equal to the unacceptable performance level as defined by the performance regime.

In summary, in our view:

- The adjustments listed above (ie converting a negative output into a positive output, and making appropriate adjustments for the gearing effect) should be the subject of discussion between OPPPA and the PPP Parties, and then applied to availability.
- A similar approach should be adopted for service points where the X factor/maximum service points figure could be proxied by the highest 25th percentile figure that has actually been observed over a specified period of time, or some estimate of what an unacceptable performance level would be.
- No similar adjustments are required for capability or input specified projects (because we suggest that these outputs are analysed using Level 3 benchmarks), or for asset health and ambience (because they are “positive outputs”)¹².

5.2 Differences between lines

When making comparisons between lines, it might be appropriate to account for differences in:

- Size. For example, if we were to find that Line A had the same LCH but a higher cost than Line B, we might initially conclude that Line B is the better performer (because it would have a lower cost per unit of output). However, if Line A were physically very much larger than Line B (eg longer track, more stations, larger fleet), it might be reasonable to expect Line A to have a higher cost base and to make an allowance for this accordingly.

The question of how this might be done, however, is complex which we discuss in section 8.
- Asset quality and/or complexity. For example, it might be reasonable to argue that if two lines had similar levels of output and costs but one line had more complex assets or was operating a lower quality set of assets (either of which might result in higher costs), then this should be taken into account when calculating benchmarks.

However, in our view it would be difficult to make any such adjustments in a relatively uncontroversial manner. For example, our discussions with the PPP Parties have revealed a number of situations where older and potentially lower quality assets actually perform better than newer and potentially higher quality assets. We therefore suggest that the issue of asset quality and/or complexity should be addressed as part of what we refer to as “drill-down” and “diagnosis”.

5.3 Interpretation of benchmark measures

On the assumption that it is possible to calculate insightful benchmarks, it is also important to note that these benchmarks should be interpreted with care.

For example, as suggested in the sections above, a low cost per unit of output would generally be considered to represent a superior performance but:

- In many cases the Infracos are incurring costs on large scale investments which are long-term in nature. Certain activities are substantial in terms of current costs but are

¹² We assume that gearing effects do not apply to these variables but this assumption may need to be revisited when the framework is populated with real data.

only expected to deliver actual improvements in performance in the relatively distant future. In this type of situation, an analysis which focuses solely on current Infracos costs and performance could indicate that an Infraco is relatively high cost but low performance, despite the fact that over a longer time frame the Infraco might actually be low cost and high performance. In our view, this means that the benchmarking framework needs to incorporate some concept of whole life asset management. We describe our views on how this could be done in section 9¹³.

- Differences in cost per unit of output between different lines could reveal important insights into the different strategies being adopted by the Infracos. For example, at a macro-level it might be assumed that some lines will adopt a high cost-high quality approach while others adopt a low cost-low quality approach. It is likely to be difficult to conclude whether either represents a superior approach where both are complying with their contract.

For example, if Line A delivered LCHS of 100 while incurring costs of 10 and Line B delivered a LCHS of 200 while incurring costs of 20, both would have a cost per unit of output of 10, and it would be difficult or impossible to say which would represent a superior performance. Similar arguments will apply at a more detailed level when considering, for example, different asset group or asset class costs per unit of output.

¹³ To a certain extent we would also expect this issue to be addressed by how we suggest the framework should analyse capability (ie by using Level 3 analysis and by excluding the costs of major projects from the rest of our analysis).

6 Benchmarks: Levels 1 and 2

This is the first of two sections which describe what benchmarks should be incorporated into the framework.

In turn, it considers benchmarks at the two levels where we make explicit links between costs and final outputs (ie Level 1 and Level 2).

6.1 Level 1 benchmarks

As discussed previously, Level 1 benchmarks consider un-apportioned costs per unit of output.

The outputs on which the framework should be based are:

- Availability.
- Capability.
- Asset health.
- Input specified projects.
- Ambience.
- Service points.

6.1.1 Availability

The key points to note with regard to availability for Level 1 analysis are that:

- The standard measure of availability (ie LCH) presents a fundamental issue with regard to comparability (ie it is a “negative” output).
- We have therefore defined a variable which we refer to as lost customer hours score (LCHS).
- We define LCHS as:
 - Maximum customer hours (which it might be possible to calculate from the values for Line Suspensions and Station Closures in the performance regime) adjusted by a factor to overcome what we refer to as the gearing effect. In the interest of simplicity, we have assumed that it may be possible to proxy this figure by using the unacceptable level of performance for each line as defined by the performance regime; minus
 - Lost customer hours.
- Cost would be measured as TCEMP for each line, but it might also be appropriate to consider a figure for these costs adjusted by size. Our methodology for adjusting for the different sizes of lines is described in section 8. We refer to this adjusted level of

costs as total costs excluding costs incurred on major projects, adjusted for size (TCEMPadj).

- We therefore have two Level 1 benchmark variables for availability:
 - TCEMP/LCHS.
 - TCEMPadj/LCHS.
- Each of these variables would be calculated on an annual basis for each line.
- This will generate a backward and forward looking profile for each benchmark variable.
- These profiles could also be converted to a single figure, ie a discounted net present value of cost per unit of output. The methodology for this calculation is described in section 9.

6.1.2 Capability

As explained previously, we suggest using Level 3 analysis for capability so there are no Level 1 benchmarks for this output.

6.1.3 Asset health

The key points to note with regard to asset health for Level 1 analysis are that:

- Every asset on the system is classified in terms of its asset condition (ie bands A-E)¹⁴.
- The Infracos are required to ensure that a specific proportion of its assets are categorised as being in condition A-C by certain time limits.
- The framework should therefore:
 - Consider the percentage of all of a line's assets in categories A-C. This should provide a very high level measure of output from the perspective of asset health.
 - Focus on the percentage of a line's stations/civils assets in categories A-C because these are the assets which, if in poor condition, might not be picked up by any of the framework's other main benchmarks. We therefore consider this to be the most important Level 1 asset health benchmark.
- Costs would be measured using TCEMP for each line but it might also be appropriate to consider a figure for these costs adjusted by size, so we would also measure costs using TCEMPadj.
- We therefore have four Level 1 benchmark variables for asset health:
 - TCEMP/% of all assets in categories A-C.

¹⁴ With the exception of lifts & escalators and possibly grey assets.

- TCEMPadj/% of all assets in categories A-C.
- TCEMP/% of stations and civils assets in categories A-C.
- TCEMPadj/% of stations and civils assets in categories A-C.
- Each of these variables would be calculated on an annual basis for each line.
- This would result in a backward and forward looking profile for each benchmark variable.
- These profiles could also be converted to a single figure, ie a discounted net present value of cost per unit of output. The methodology for this calculation is described in section 9.
- We would also recommend monitoring a number of other potentially important “leading indicators” (eg age of rolling stock).

6.1.4 Input specified projects

As explained previously, the framework should analyse input specified projects using Level 3 analysis, so there are no Level 1 benchmarks for this output.

6.1.5 Ambience

The key points to note with regard to ambience for Level 1 analysis are that:

- Ambience is assessed using the MSS score.
- Costs would be measured using TCEMP for each line but it might also be appropriate to consider a figure for these costs adjusted by size, so the framework should also measure costs using TCEMPadj.
- We therefore have two Level 1 benchmark variables for availability:
 - TCEMP/MSS.
 - TCEMPadj/MSS.
- Each of these variables would be calculated on an annual basis for each line.
- This would result in a backward and forward looking profile for each benchmark variable.
- These profiles could also be converted to a single figure, ie a discounted net present value of cost per unit of output. The methodology for this calculation is described in section 9.
- The usefulness of ambience benchmarks is likely to be affected by the lack of a strong direct correlation between inputs and the Mystery Shopping Survey.

6.1.6 Service points

The key points to note with regard to service points for Level 1 analysis are that:

- The standard measure of service points presents a fundamental issue with regard to comparability (ie it is a “negative output”). It is also unclear whether it is inherently possible to make inter-line comparisons of levels of service points.
- We have therefore developed an alternative measure of service points which we refer to as service points score (SPS).
- We define SPS as:
 - An estimate of a realistic maximum for the service points (MSP) that each line might actually incur. We would estimate MSP by plotting service points over a specified period of time and taking the worst 25th percentile¹⁵. Alternatively, it might be reasonable to assume that MSP could be proxied by a performance level which the Parties could agree would be unacceptable; minus
 - Actual service points.
- Costs would be measured as TCEMP for each line but it might also be appropriate to consider a figure for these costs adjusted by size, so we would also use TCEMP adj.
- We therefore have two Level 1 benchmark variables for service points:
 - TCEMP/SPS.
 - TCEMPadj/SPS.
- Each of these variables would be calculated on an annual basis.
- This would result in a backward and forward looking profile for each benchmark variable.
- These profiles could also be converted to a single figure, ie a discounted net present value of cost per unit of output. The methodology for this calculation is described in section 9.

6.1.7 Checks

As explained previously, it will be important for the benchmarking framework to include a number of checks to ensure that the Infracos are compliant with their contract and/or expectations in a number of key areas.

Our current view is that the main checks that would be incorporated into the framework is a straightforward assessment of the number of Corrective Action Notices incurred by each line/Infraco but our views on this issue are evolving and we welcome comments from others.

¹⁵ We chose the 25th percentile in order to minimise the likelihood of selecting an unusually extremely low level of performance, while at the same time minimising the likelihood of the SPS becoming negative.

6.2 Level 2 benchmarks

As discussed previously, the framework should also incorporate Level 2 benchmarks which concern apportioned costs per unit of output.

6.2.1 Availability

The key points to note with regard to availability for Level 2 analysis are that:

- Availability would be measured using LCHS.
- We would consider apportioned costs at the level of asset group and asset class.
- Cost would be measured using:
 - Apportioned asset group cost excluding major projects (AAGCEMP). There will be 3 such measures of cost (ie one each for trains, stations and infrastructure).
 - Apportioned asset class cost excluding major projects (AACCEMP). There will be 9 such measures of cost (ie one each for rolling stock, signalling, stations, track, power, civil, depots, lifts & escalators, and admin & other/“on-costs”).
- It may, however, also be important to adjust these measures of cost to reflect differences in the size of each of the lines. Our approach to making these adjustments is described in section 8.
- We refer to these adjusted levels of cost as:
 - AAGCEMPadj; and
 - AACCEMPadj.
- We therefore potentially have 24 Level 2 benchmark variables for availability:
 - AAGCEMP/LCHS (3)
 - AAGCEMPadj/LCHS (3)
 - AACCEMP/LCHS (9)
 - AACCEMPadj/LCHS (9)
- These benchmarks would be filtered using the approach outlined in section 4.
- The remaining benchmark variables would be calculated on an annual basis.
- This would result in a backward and forward looking profile for each benchmark variable.

- These profiles could also be converted to a single figure, ie a discounted net present value of cost per unit of output. The methodology for this calculation is described in section 9.

6.2.2 Capability

As explained previously, the framework should use Level 3 analysis for capability. There are therefore no Level 2 benchmarks for this output.

6.2.3 Asset health

The key points to note with regard to asset health for Level 2 analysis are that:

- Every asset on the system is classified in terms of its asset condition (ie bands A-E)¹⁶. The Infracos are required to ensure that a specific proportion of its assets are categorised as being in condition A-C by certain time limits.
- The framework should therefore:
 - Consider the percentage of all of a line's assets in categories A-C. This should provide a very high level measure of output from the perspective of asset health.
 - Focus on the percentage of a line's stations/civils assets in categories A-C because these are the assets which, if in poor condition, might not be picked up by any of the other main benchmarks. We therefore consider this to be the most important Level 2 asset health benchmark.
- Cost would be measured using:
 - AAGCEMP. There will be 3 of these measures; and
 - AACCEMP. There will be 9 of these measures.
- It may also be appropriate to adjust these costs to reflect differences between the size of each of the lines.
- We therefore potentially have 48 Level 2 benchmark variables for asset health:
 - AAGCEMP/% of all assets in categories A-C (3)
 - AAGCEMPadj/% of all assets in categories A-C (3)
 - AAGCEMP/% of stations and civils assets in categories A-C (3)
 - AAGCEMPadj/% of stations and civils assets in categories A-C (3)
 - AACCEMP/% of all assets in categories A-C (9)
 - AACCEMPadj/% of all assets in categories A-C (9)

¹⁶ With the exception of lifts & escalators (and possibly grey assets).

- AACCEMP/% of stations and civils assets in categories A-C (9)
- AACCEMPadj/% of stations and civils assets in categories A-C (9)
- These variables would be filtered using the approach described in section 4.
- The remaining variables would be calculated on an annual basis.
- This would result in a backward and forward looking profile for each benchmark variable.
- These profiles could also be converted to a single figure, ie a discounted net present value of cost per unit of output. The methodology for this calculation is described in section 9.

6.2.4 Input specified projects

As explained previously, we would use Level 3 analysis for input specified projects. There are therefore no Level 2 benchmarks for this output.

6.2.5 Ambience

The key points to note with regard to ambience for Level 2 analysis are that:

- Output would be measured using MSS scores.
- Costs would be measured using apportioned asset group and asset class costs (ie AAGCEMP and AACCEMP). It may also be appropriate to reflect for differences in the size of each line so we would also consider AAGCEMPadj and AACCEMP adj.
- There are therefore 24 possible Level 2 ambience benchmarks:
 - AAGCEMP/MSS (3)
 - AAGCEMPadj/MSS (3)
 - AACCEMP/MSS (9)
 - AACCEMP/MSS (9)
- These variables would be filtered using the approach described in section 4.
- The remaining variables would be calculated on an annual basis.
- This would result in a backward and forward looking profile for each benchmark variable.
- These profiles could also be converted in a single figure, ie a discounted net present value of cost per unit of output. The methodology for this calculation is described in section 9.

- The usefulness of ambience benchmarks is likely to be affected by the lack of a strong direct correlation between inputs and the Mystery Shopping Survey.

6.2.6 Service points

The key points to note about service points for Level 2 analysis are that:

- Output would be measured using SPS.
- Costs would be measured using apportioned asset group and asset class costs (ie AAGCEMP and AACCEMP). It may also be appropriate to reflect for differences in the size of each line so we would also consider AAGCEMPadj and AACCEMP adj.
- There are therefore 24 possible Level 2 benchmarks for service points:
 - AAGCEMP/SPS (3)
 - AAGCEMPadj/SPS (3)
 - AACCEMP/SPS (9)
 - AACCEMP/SPS (9)
- These variables would be filtered using the approach described in section 4.
- The remaining variables would be calculated on an annual basis.
- This would result in a backward and forward looking profile for each benchmark variable.
- These profiles could also be converted in a single figure, ie a discounted net present value of cost per unit of output. The methodology for this calculation is described in section 9.

6.3 Summary

The table below summarises all of the potential Level 1 and Level 2 benchmarks.

	Availability	Capability	Asset health	Input specified projects	Ambience	Service Points	Checks
Level 1	TCEMP/LCHS TCEMPadj/LCHS	NA	TCEMP% assets in category A-C TCEMPadj% assets in category A-C TCEMP% stations and civil assets in category A-C TCEMPadj% stations and civil assets in category A-C Leading indicators (age & resid value of rolling stock, track?)	NA	TCEMP/MSS TCEMPadj/MSS	TCEMP/SPS TCEMPadj/SPS	TBC TBC
No. of benchmarks	2	NA	4	NA	2	2	
Profile	Yes (2)	NA	Yes (4)	NA	Yes	Yes	
NPV	Yes? (2)	NA	Yes? (4)	NA	Yes? (2)	Yes? (2)	
Fillers	No	NA	No	NA	No	No	
Total no. of variables	10						
Level 2	AAGCEMP/LCHS (3) AAGCEMPadj/LCHS (3)	NA	AAGCEMP% assets in category A-C (3) AAGCEMPadj% assets in category A-C (3) AAGCEMP% stations and civil assets in category A-C (3) AAGCEMPadj% stations and civil assets in category A-C (3)	NA	AAGCEMP/MSS (3) AAGCEMPadj/MSS (3)	AAGCEMP/SPS (3) AAGCEMPadj/SPS (3)	None (all done as level 1)
No. of benchmarks	6	NA	12	NA	6	6	0
Profile	Yes	NA	Yes	NA	Yes	Yes	
NPV	Yes? (6)	NA	Yes? (12)	NA	Yes? (6)	Yes? (6)	
Fillers	Yes	NA	Yes	NA	Yes	Yes	
Total no. of benchmarks	Up to 30						
Level 2	AACCEMP/LCHS (9) AACCEMP(adj)/LCHS (9)	NA	AACCEMP% assets in category A-C (9) AACCEMPadj% assets in category A-C (9) AACCEMP% stations and civil assets in category A-C (9) AACCEMPadj% stations and civil assets in category A-C (9)	NA	AACCEMP/MSS (9) AACCEMPadj/MSS (9)	AACCEMP/SPS (9) AACCEMPadj/SPS (9)	None (all done as level 1)
No. of benchmarks	18	NA	36	NA	18	18	0
Profile	Yes	NA	Yes	NA	Yes	Yes	
NPV	Yes? (18)	NA	Yes? (36)	NA	Yes? (18)	Yes? (18)	
Fillers	Yes	NA	Yes	NA	Yes	Yes	
Total no. of benchmarks	Up to 90						

7 Benchmarks: Level 3

As explained previously, the framework should use Level 3 benchmarks in order to:

- Assess costs which are incurred in order to improve capability (ie through major line upgrades) and to deliver certain input specified projects.
- To better understand the results of Level 1 and Level 2 analysis. In this respect, the construction and analysis of Level 3 benchmarks will be an integral part of drill-down and diagnosis and is likely to make particular use of, and build on, the CEPA report.

7.1 Capability

By way of background, there are 4 key points to note with regard to Level 3 analysis of capability.

First, the PPP Contracts make it clear that the majority of capability improvements on the lines come from line upgrades.

These line upgrades consist principally of rolling stock and signalling renewal projects. Examples of these projects are shown in the table below.

SSL

New trains - Circle and Met and Hammersmith & City : 2013, District : 2015

New Signalling - Circle and Met and Hammersmith & City : 2013, District : 2016

BCV

New trains - Victoria : 2009, Bakerloo : 2019

New Signalling – Victoria : 2009

JNP

New trains Picc : 2014

New Signalling Jubilee : 2009, Northern : 2011, Picc : 2014

Contracts have already been let by the Infracos for a number of these projects.

Second, we would recommend Level 3 analysis of a sub-set of the projects listed above. In our view, this sub-set should comprise:

- Victoria line rolling stock.
- Victoria line signalling.
- Jubilee line signalling.

These projects have been chosen as they are the first major renewals projects under the PPP contracts.

Third, we would expect Level 3 analysis of these projects to provide a useful insight into the relative economy/efficiency of these projects themselves, but also into the relative economy/efficiency of other rolling stock and signalling renewal projects in the future.

Fourth, from the projects selected above it can be seen that there are two signalling projects and one rolling stock project. The signalling projects can therefore be benchmarked with each other but the rolling stock project will need other references for comparison.

Considering signalling projects first, in our view the approach of benchmarking these projects against each other is appealing as the design and installation of new signalling systems on the underground presents some quite specific issues which would make robust external comparators particularly difficult to find. Moreover, the technology employed in the design of new signalling systems is likely to have a significant impact on the initial and long-term costs of these projects. Thus comparing two projects internally will provide a useful insight over time as to the relative merits of the approaches taken by the different Infracos.

The drawback to this approach is that both these projects are at relatively early stages and therefore much of the data used in the development of the metrics for the benchmarks will be based on the Infracos' forecasts rather than actual outturn costs. However, over time more actual cost data should become available.

The exact level of disaggregation of unit costs to be used in this analysis would be finalised during Stage 2 but would be likely to be based on at least the following breakdown:

- Fixed costs (design).
- Variable costs (cost per km of track or signal section).
- Installation times / costs (per km of track or signal section).
- Maintenance times / costs (per km of track or signal section).

Moving on to rolling stock, we note that the CEPA report suggests that other London Underground projects could be used for this comparison such as Jubilee Line, Northern Line and Central Line fleets. However, as the newest of these fleets is now approaching 10 years of operation, we are unsure whether indexing costs over such an extended period of time will provide robust results. While, strictly speaking, this is beyond the scope of our work, we have therefore identified some more recent external comparators. Examples of suitable external projects include Singapore NEL and Guangzhou Metro Line 2 in China. Both these projects have very comparable performance levels and configurations to the rolling stock which is to be provided for the Victoria Line. Whilst these two example projects were derivatives of existing designs (whereas the Victoria Line trains will be a new design), external comparison with these projects should provide a useful insight into both the current costs and timescales for this project and the expected costs for the other Metronet rolling stock fleet renewals.

We would again expect exact details of the level of disaggregation and comparators to be finalised in stage 2, but we would expect to start with high level comparisons of:

- Cost per train set/car.
- Timescales from contract award to first delivery.

- Timescales from first delivery to first revenue service operation.
- Rate of delivery of trains.
- Rate of introduction into service of new trains.
- Reliability growth of new trains (ie time taken to achieve contractual reliability targets).
- Maintenance costs.

7.2 Input specified projects

By way of background, there are five key points to note about Level 3 analysis of input specified projects.

First, for the purposes of the internal benchmarking framework, we define input specified projects as covering:

- The Station Modernisation & Refurbishment Programme; and
- The Station Accessibility Programme (where this is delivered as an integral part of a Station Modernisation & Refurbishment Programme).

We therefore define input specified projects as excluding a range of other projects which might be referred to as “specified projects”, ie:

- Line Upgrades (which is considered in Level 3 analysis of Capability).
- Enhancements to Trains and Station Facilities (capex measured within Rolling Stock & Stations Asset Classes).
- Train Refurbishments and Replacements (capex measured within Rolling Stock Asset Class).
- Station Accessibility Projects (capex measured within Stations Asset Class).
- Station Maintenance (opex measured within Stations Asset Class).
- Lifts & Escalators (separate Programme within Lifts & Escalators Asset Class).
- Other Station Capex – ie any works not listed above such as Major Enhancements Services & Works (measured within Stations Asset Class).

Second, it will be important to classify station projects by line. In this regard, it is worth noting that:

- The Underground system comprises a total of 250 stations.
- Stations are generally allocated to specific lines in their entirety (ie stations supporting multiple lines are not apportioned between respective lines or indeed Infracos).

- There are some instances where there is a split in interconnecting stations which are known by the same name due to a clear separation in infrastructure i.e. two stations in close proximity. In such cases each individual station is treated separately and referred to by adding the line abbreviation as an extension eg “Edgware Road (Bak)” and “Edgware Road (Cir)”.
- 19 stations are allocated as National Rail Stations and would therefore be excluded from the analysis, as would the 5 stations which are allocated to the East London line.

Third, there are a number of different ways of categorising stations. For example by:

- Size / Quality. This is measured A-C, indicating the categories of finish and quality of the materials used in Station Modernisations and Refurbishments (“A” being a station with the highest level of finish, fabric quality and durability required due to passenger traffic characteristics). The precise allocation of station categories is defined by the nature, location, size and passenger flow requirements of a particular station¹⁷.
- Depth from Surface. Deep stations are categorised as Section 12 stations and require enhanced specification particularly with regard to fire protection etc. Uncategorised stations are closer to the surface and often in suburban locations, resulting in a reduced specification.

Fourth, there is also a number of different ways of categorising the station projects themselves. More specifically, there are 3 recognised categories of works within the Modernisation & Refurbishment Programme (ie Station Modernisation, Enhanced Station Refurbishments and Station Refurbishments)¹⁸.

As stated previously, there is also an associated programme of works that is often integrated with station projects (ie Accessibility Projects). Where these works are not integrated¹⁹ they will be commissioned separately as Major Enhancements and not assessed as part of the Level 3 analysis. In our view, however, it would be worthwhile flagging those projects which incorporate Accessibility works as a potential explanation for outlying results.

Fifth, it should be possible to drill down beyond the work classifications described above into standard work types. Such distinctions in work types can be extracted from the descriptions used within the definition of work content for each classification²⁰. However, we recommend that they be aligned to headings used in reports from the Master Projects Database (MPD) so that appropriate costs can be extracted.

With all of the above in mind, the approach would be to:

- Group together similar stations that currently have active or planned works, along the lines of the table below.

¹⁷ See Schedule 2.1, Appendix 15, Item 2.2 of the PPP Contracts for a definition of Station Categorisations.

¹⁸ See Schedule 2.1, Appendix 15, Item 6 of the PPP Contract for a list of all station Modernisations, Enhanced Refurbishments and Refurbishments allocated by Line. This source also states the due date for completion.

¹⁹ Integrated Accessibility works are listed in the guidance document “Summary of the PPP Contract Documents, Annex: Station Projects by line & due date”.

²⁰ See Schedule 2.1, Appendix 15, Items 2-4 of the PPP Contracts

Group by:	Examples:
Work Type	Modernisation Enhanced Refurbishment Refurbishment.
Station Depth	Section 12 Yes / No
Station Size/Quality	Category A-C
Line Allocation	Bakerloo etc

- Obtain actual and/or forecast costs from the MPD.
- Make cost comparisons of comparable station projects by line.
- If appropriate, explore outliers in more detail by considering costs on what we refer to as a “work type” basis. Examples of possible work types include:
 - Station Facilities including Ticket Halls, Ancillary Areas and Staff Accommodation (Renewal, Enhancement or New)
 - Floor, Wall, Ceiling or Staircase Finishes
 - Renewal of Electrical & Mechanical, Fire Protection and Communications assets
 - Renewal of Doors, Ironmongery, Signage, Seating, Toilets, Waiting Rooms and Enclosed Waiting Areas
 - Measures to Prevent Water Ingress
 - Platform Repairs
 - Renewal or Repair of Structures including Facades, Canopies, Glazing, Roof and Platform Canopies

Finally, it is important to note that:

- The work type headings cited above should not be confused with similar cost headings within AAMP reports that sit outside the Stations Modernisation and Refurbishment Programme (and which are not part of the Level 3 analysis of input specified projects). The exact headings used to analyse work type costs should be tied into those used in the MPD to ensure data is readily available and in a consistent manner.

- Analysis by work type could be used where the sample of grouped stations is too small or where individual stations are felt to be too heterogeneous and therefore not directly comparable with other stations in their classification.

7.3 Level 3 analysis as part of drill-down and diagnosis

As explained previously, the framework should use Level 3 benchmarks as part of the drill-down into, and diagnosis of, analysis carried out at Levels 1 and 2.

It is currently impossible to define precisely how this would work or which benchmarks should be analysed because so much will depend on what results are generated at Levels 1 and 2, but we would expect the Level 3 benchmarks to build on the recommendations of the CEPA report²¹.

However, it is possible to obtain a better understanding of how the framework would actually work in practice and what some of the Level 3 benchmarks might be by constructing a hypothetical example.

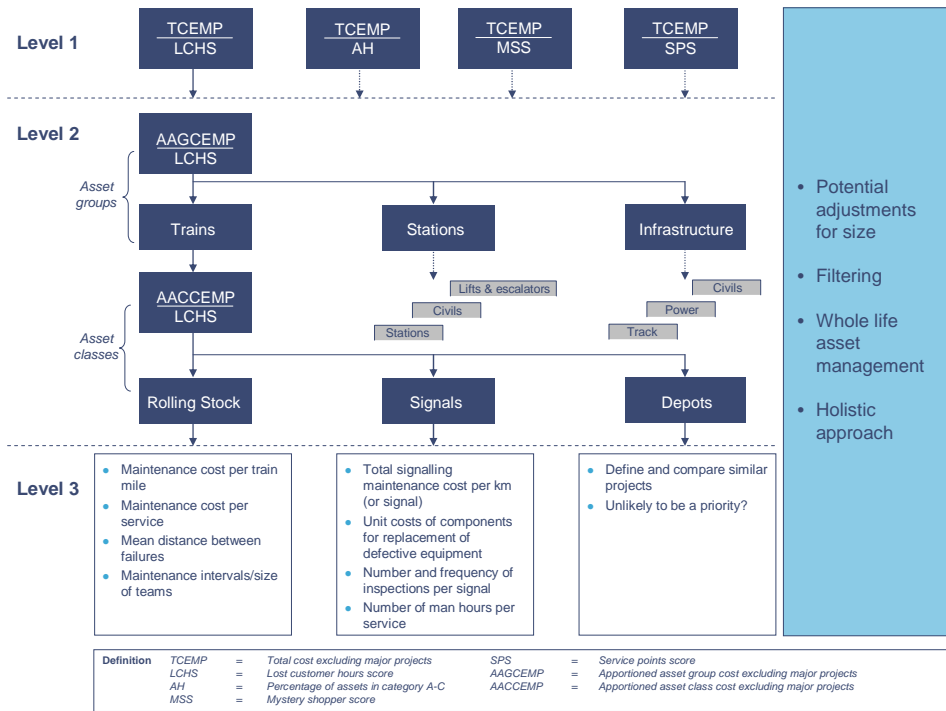
The rest of this section therefore does this by considering, in turn:

- Level 1 benchmarking.
- Level 2 benchmarking.
- Level 3 benchmarking.
- Diagnosis.

A schematic summary of the hypothetical example is provided in figure 2 and corresponding data is provided in Annex 3. This section should therefore be read in conjunction with Annex 3.

²¹ See, in particular, Annex 4 – Appendix 1.

Figure 2: The internal benchmarking framework in practice – a hypothetical example



7.3.1 Level 1 benchmarking

First, let us assume that we have carried out an initial round of data cleaning and that this provides us with the information in table 1.

Converting this raw data into Level 1 benchmarks (see table 2) indicates that 1 of the 11 lines has a particularly high cost per unit of availability (ie 1.5 as compared with 1.0-1.1 for the other lines).

For the avoidance of doubt, with regard to:

- Cost, we mean unapportioned (ie total) line costs excluding those costs incurred on major projects. In other words, we have carved out those costs involved with major line upgrades and station projects.
- Unit of availability, we mean lost customer hours score (LCHS). In other words, we have subtracted actual lost customer hours for each line from a maximum lost customer hours figure for each line.

We also see from table 2 that all other measures of cost per unit of output (ie cost per unit of asset health, ambience and service points supplied) are relatively high for line 1 but by less than for availability.

This is important because if, when considering line 1 as compared with the other lines:

- Only cost per unit of availability was high (ie cost per unit of the other outputs were similar for each of the lines), then this could indicate that the situation is solely the result of low performance on availability for line 1.

- If costs per unit of output were just as high for the other outputs as for availability (ie all of line 1's Level 1 benchmarks are high), then this might suggest that line 1's cost are unusually high.

On the face of it, line 1 might therefore be characterised by high costs and/or poor performance on availability. But how might we better understand this situation?

One possible explanation for the situation might be inherent differences between the lines, such as line size. In order to deal with this, we:

- Calculate Level 1 benchmarks using costs adjusted for size (see table 4).
- See that adjusting costs for size does not explain all of the difference. More specifically, cost per unit of availability for line 1 continues to look relatively high (ie 1.4 compared with 1.0-1.1 for the other lines).

We therefore need to find out more by considering benchmarks at Level 2. And we would want to do this anyway because of the inherent limitations to Level 1 benchmarks.

7.3.2 Level 2 benchmarking

With regard to Level 2 benchmarks, it is important to remember that:

- Level 2 benchmarks consider apportioned costs (ie costs relevant to the unit of output).
- We effectively have two sets of benchmarks, ie those which consider costs at the level of asset group (3) and asset category (9)²².

Asset group benchmarks

We first consider Level 2 asset group benchmarks (see tables 10 and 11) and find that:

- Station system and infrastructure system costs per unit of availability are broadly similar between all of the lines; but
- Train system cost per unit of availability is high for line 1, regardless of whether or not we adjust these costs for size.
- More specifically, line 1 has an apportioned train system cost per unit of availability of:
 - 0.3 compared with 0.13-0.14 for the other lines, on an unadjusted basis; and
 - 0.286 compared with 0.13-0.14 for the other lines, on an adjusted basis.

On the face of it, this would lead us to believe that the situation is the result of high train system costs.

²² With regard to "admin & other", we currently assume that this would be considered as an asset class in its own right but rolled into one of the main asset groups but our thinking on this is evolving.

Asset class benchmarks

We therefore now examine Level 2 asset class benchmarks (see tables 17 and 18). Those asset classes which are most important to the train systems asset group are rolling stock, signals and depots.

We now see that line 1 has:

- High rolling stock and signalling costs per unit of availability.
- An average level of depot cost per unit of availability. This, together with the relative unimportance of expenditure on depots, leads us to filter and therefore discontinue our analysis into this area.

Let us also assume that:

- Initial discussion with the PPP parties does not reveal any clear explanation for the situation with regard to rolling stock and signalling asset class costs per unit of availability which also clearly warrants a further round of normalisation. (Possible such explanations are included in Box 1); and
- Availability is primarily driven by opex rather than capex.

The next logical step is therefore to consider Level 3 benchmarks for rolling stock and signals maintenance expenditure.

Box 1: High rolling stock and signal costs per unit of availability

One possible explanation for high train system costs per unit of availability is an endemic failure of a system on a particular asset such as fatigue failures of gearboxes on rolling stock. Failures like this are likely to occur based on mileage and therefore would occur at roughly the same time on each train. This would have the effect of requiring an enhanced maintenance regime to both check and repair the gearboxes and could also result in disruptions to service.

Other possible causes of high cost per unit of availability could be indirect, for example the installation of a new signalling system. The costs of the new signalling system would likely already have already been removed from the benchmark metric as this would be classified as a major project cost. However the process of installing the new signalling system, because it has to be installed over the existing system which will have to continue to operate until the new system is fully installed and commissioned, might cause faults to occur on the old system as it is disturbed through the installation of the new system. As the installation and commissioning of new signalling systems can take years, this could cause a prolonged performance impact on the old system.

7.3.3 Level 3 benchmarking

Drawing on the CEPA report, Level 3 benchmarks for rolling stock maintenance expenditure might include:

- Maintenance cost per train mile.
- Maintenance costs per service.
- Mean distance between failures.
- Maintenance intervals / size of maintenance teams.

Moreover, Level 3 benchmarks for signal maintenance expenditure might include:

- Total signal maintenance cost per km (or signal).
- Unit costs of components for replacement of defective equipment.
- Number and frequency of inspections per signal.
- Number of man hours required per service.

Let us now assume that the benchmarks referred to above suggest that the underlying situation appears to be the result of a combination of both of these possible causes (ie the line has rolling stock and signalling which requires a lot of expensive servicing, as measured by these benchmarks).

On the face of it, this would suggest that the underlying situation is predominantly cost rather than output based but we might want to confirm this through more detailed analysis of the output variable. For example, it might be useful to ascertain whether or not line 1 has had a large number of unusually severe incidents which have impaired availability (but which might not be expected to recur). Let us assume that they have not.

7.3.4 Diagnosis

It may now be possible to draw meaningful conclusions, but only if we consider our benchmarks:

- In a holistic fashion; and
- In a manner which encompasses the concept of whole life asset management.

Let us therefore assume that:

- Line 1 is in the process of carrying out an important rolling stock refit. We might identify this from our Level 3 analysis of capability (ie major line upgrades).
- Rolling stock asset class cost per unit of availability for line 1 declines considerably during future periods to the extent that, over the period as a whole²³, it is at a similar level to the other lines. This decline in rolling stock asset class costs is, however, not sufficient to take train system costs per unit of availability anywhere close to being the same as the other lines over the whole period (see figure 2).
- Line 1's signal maintenance cost per unit of availability remains above average throughout the period (see figure 3).

It might therefore be reasonable to conclude that signal maintenance is an outstanding issue which needs to be explored further.

There are a number of ways that this could be done, including but not limited to:

- Qualitative audits.
- Market based references for prices.

²³ Which, for simplicity, we consider to be until 2011.

- Bottom up cost modelling.

This analysis might shed more light on the specific potential cause outlined earlier (ie installation of a new signalling system) and, at least as importantly, it might also uncover completely new potential explanations.

8 Deriving costs and normalising for size

As explained previously, it may be appropriate to normalise for differences in line size. This section outlines our views on how this might be done and how to derive various measures of cost.

We would suggest using 7 steps, each of which is described below.

8.1 Step 1: Obtain total cost data (TC)

Step 1 is to identify what total costs are incurred by each line. Generally, this should be relatively straightforward as the Infracos' financial models and AAMPs report data on a line specific basis.

The main exceptions to this appear to be in the areas of:

- Rolling stock – where, on occasion, fleet may be shared by the different lines.
- Stations – which are/will have to be allocated between different lines.
- Overheads and information technology systems – which may be reported only at the Infraco level and may therefore have to be allocated between an Infraco's lines.

Step 1 provides us with total costs per line (TC).

8.2 Step 2: Classify costs by asset group and asset class

Step 2 is to classify line-specific costs by asset group and asset class. We consider there to be three asset groups²⁴ and nine asset classes²⁵.

This should also be relatively straightforward, as there is a relatively clear relationship between the reporting headings used by the Infracos in their financial models and their AAMPs and our definitions of asset groups and asset classes. It will, however, be important to ensure that a consistent approach is adopted between different Infracos and this may require some adjustments to be made. For example, as we understand it, Tube Lines adopts a different reporting approach to Metronet which results in as much as 25% of Tube Lines' costs being reported as admin/on-costs.

In short, Step 2 provides us with Asset Group Costs (AGC)²⁶ and Asset Class Costs (ACC)²⁷ for each line.

8.3 Step 3: Exclude major projects from the cost data

Step 3 is to exclude from the above those costs which are incurred as a result of major projects (ie line upgrades intended to improve capability and input specified projects).

Step 3 will therefore provide us with figures for:

²⁴ Trains, Stations, Infrastructure

²⁵ Rolling stock, Signalling, Stations, Track, Power, Civil, Depots, Lifts & escalators, Admin & other / "on-costs"

²⁶ Of which there will be three.

²⁷ Of which there will be nine.

- Asset Group Cost Excluding Major Projects (AGCEMP).
- Asset Class Cost Excluding Major Projects (ACCEMP).

It is therefore also important to note that aggregating either AGCEMP or ACCEMP by line should equal total costs excluding cost for major projects (TCEMP) per line.

8.4 Step 4: Identify most important asset class costs

Step 4 is to identify the, say, four most important/largest items of asset class cost.

While it is impossible to say which these asset classes are without analysing the data in detail (or indeed whether each of the lines has a similar breakdown of asset class costs), it might be reasonable to assume that these will be:

- Rolling stock.
- Signalling.
- Stations.
- Track.

8.5 Step 5: Identify a size variable for each asset class cost

Step 5 is to identify an appropriate size variable for each of the asset class costs referred to above.

For example, it might be reasonable to assume that:

- Rolling stock costs vary according to fleet size (ie number of cars).
- Signalling costs vary according to number of signals.
- Station costs vary according to the number of stations on the line.
- Track costs vary according to the length of the line (ie km of track).

For the avoidance of doubt, we realise that each of these items could be modelled in much more detail (eg by weighting stations according to their various classifications and/or track according to whether it is open section, sub-surface or deep tube etc) but it is currently unclear to us whether this would be a fruitful exercise.

8.6 Step 6: Adjust asset class costs to reflect differences in size

There are very many different ways in which the asset class costs referred to above might vary with size.

In the interest of simplicity, it might be reasonable to:

- Assume that there is a linear relationship between the size parameters and asset class costs referred to above and that asset class costs are completely “variable” in nature (ie there are no fixed costs which have to be considered).
- Adjust a line’s asset class costs by a factor which depends on the extent to which its line size parameter is greater or lower than the average level of this parameter for all of the lines. In other words, if a line has a larger line size parameter than average, then its costs for this asset class should be adjusted downwards. A worked example showing how these adjustments would work is included in Annex 4.

Step 6 therefore provides us with adjusted asset class costs excluding major projects (ACCEMPadj).

These figures could then be aggregated to provide us with:

- Asset Group Cost Excluding Major Projects Adjusted for size (AGCEMPadj) for each line; and/or
- Total Cost Excluding Major Projects Adjusted for size (TCEMPadj) for each line.

It would, of course, be possible to adjust all asset class costs for size but we considered that this would result in unduly complex calculations.

8.7 Step 7: Apportioning costs to outputs

Step 7 is to apportion asset class costs to the various outputs on which the benchmarking framework is to be based.

This would be done by:

- Multiplying ACCEMP or ACCEMPadj by the relevant percentage selected from the table below; in order to calculate
- Apportioned Asset Class Cost Excluding Major Projects (AACCEMP) or Apportioned Asset Class Costs Excluding Major Projects Adjusted for size (AACCEMPadj)

It is important to note that it should be possible to sum AACCEMP or AACCEMPadj to arrive at²⁸:

- Apportioned Asset Group Costs Excluding Major Projects (AAGCEMP) or Apportioned Asset Group Costs Excluding Major Projects Adjusted for size (AAGCEMPadj); and/or
- Apportioned Total Costs Excluding Major Projects (ATCEMP) or Apportioned Total Costs Excluding Major Projects Adjusted for size (ATCEMPadj).

Provisional assumptions - Apportioning costs to outputs & Impact of costs on outputs

We have made provisional assumptions with regard to how certain costs should be apportioned to the four main outputs relevant to Level 2 analysis. These assumptions are summarised in tables 1 and 2.

²⁸ Although there are situations where this might not strictly be accurate.

It should be noted that:

- We have excluded costs incurred on major projects because of how the framework should consider capability and input specified projects.
- For asset class costs we have differentiated between capex and opex, although this might not strictly be necessary or appropriate given the benchmarks we propose should be calculated by the framework. We have, however, included this differentiation for the sake of completeness.
- In the interest of simplicity, we have assumed that these apportionments would be valid over the 30 year period of the PPP Contract.
- Our intention would generally be to apply the same apportionments to each of the lines.
- We realise that it will be important to consider interrelationships between each of the outputs and that this could be complex.
- These assumptions are very provisional and subjective in nature. We would welcome comment from the PPP Parties.

We have also made provisional assumptions with regard to the impact of spending in each area on the outputs. This could be important because it will be important to understand the extent to which certain expenditure is more important/generates a “bigger bang per buck” than other expenditure. Our provisional assumptions are summarised in table 3. Similar comments to the above apply.

Table 1 - Provisional apportionment of asset group costs to outputs

	Train Systems	Station Systems	Infrastructure
	Total Costs	Total Costs	Total Costs
Availability	40%	45%	30%
Ambience	10%	10%	5%
Asset Health	45%	15%	55%
Service Points	5%	30%	10%
Description	<p>Train systems (rolling stock, signalling & control systems and depots) expenditure mainly affects the availability and asset health output scores. Through the rolling stock expenditure there is also a contribution to ambience through the opex spend affecting the train MSS score. Asset health and service point contributions are to be verified.</p>	<p>The majority of expenditure on station systems (stations + lifts and escalators) will affect the availability and service point scores. Ambience will be affected through the opex expenditure on stations improving station MSS scores and asset health will also be influenced by the capex expenditure on the stations and lifts and escalators.</p>	<p>The infrastructure expenditure on track, power and non station civil structures mainly affects the availability and asset health scores. Capability score are also improved through track renewals allowing faster line speeds and ambience is indirectly affected through improved ride quality affecting train MSS scores.</p>

Table 2 – Provisional apportionment of asset class costs to outputs

	Rolling Stock		Signalling		Stations		Track	
	Opex	Capex*	Opex	Capex*	Opex	Capex*	Opex	Capex
Availability	50%	10%	80%	10%	35%	5%	35%	20%
Ambience	20%	10%	0%	0%	35%	20%	5%	10%
Asset Health	10%	70%	10%	80%	10%	70%	30%	60%
Service Points	20%	10%	10%	10%	20%	5%	30%	10%
Description	<p>The majority of the Opex spend will be on routine and non routine maintenance therefore this will contribute mainly to the availability, ambience and service point scores. Asset health will also be maintained as part of this expenditure. SP allocation is improved through improved reliability of the assets.</p>		<p>Capex spend on rolling stock will be on renewals, refurbishments and upgrades. Therefore the main improvement area will be asset health. Ambience, availability and service point scores will also see some improvement from the introduction of new equipment.</p>		<p>The majority of the Opex spend will be on routine and non routine maintenance therefore this will contribute mainly to the availability and service point scores. Asset health will also be maintained as part of this expenditure.</p>		<p>Capex spend on signalling will be on renewals and upgrades. Therefore the main improvement area will be asset health. Availability and service point scores will also see some improvement from the greater reliability of the new equipment.</p>	
					<p>The opex expenditure on stations will mainly be on the maintenance of the station systems and cleaning. The availability, ambience and service point scores will therefore be the main beneficiaries of this spend. Service points will be improved through lower failure rates.</p>		<p>Capex expenditure on stations will give improved asset health and better ambience. A small contribution to availability and service points improvements are expected through improved reliability of new station systems.</p>	
							<p>Track opex expenditure on maintenance, planned and non planned, will provide improved availability and service point scores. Asset health will be maintained.</p>	
							<p>Capex expenditure on track will mainly show an improved asset health score, greater availability, and reduced service points. Ambience will also improve indirectly through better ride quality.</p>	

Table 2 (cont.) – Provisional apportionment of asset class costs to outputs

	Power		Civil		Depots		Lifts & Escalators		On Costs	
	Opex	Capex	Opex	Capex	Opex	Capex	Opex	Capex	Opex	Capex
Availability	40%	25%	40%	40%	60%	40%	40%	10%	tbd	tbd
Ambience	0%	0%	0%	0%	0%	0%	20%	30%	tbd	tbd
Asset Health	20%	65%	30%	50%	20%	40%	20%	50%	tbd	tbd
Service Points	40%	10%	30%	10%	20%	20%	20%	10%	tbd	tbd
Description	The opex expenditure on the power system will principally benefit the availability and service points score. Asset health will also be maintained from this expenditure.		Capex expenditure on the power system will principally improve the asset health score. Improvements will also be seen in availability through the improved reliability of new equipment.		The split in benefits from opex expenditure on Civil works is approximately even between availability, asset health and service points.		Capital expenditure on civil works will have similar spread of benefits to opex spend with the addition of a small contribution to capability through upgrades to some bridges and other structures allowing higher line speeds and or greater passenger access.		The depots exist principally to service the rolling stock and infrastructure systems. Therefore the main contribution from opex expenditure at depots will be improved availability, asset health and service point scores.	
	Capex expenditure at depots will give a similar range of benefits to opex with the addition of a contribution to capability as capex on depots can allow the depot to handle larger fleet sizes for example.		The opex spend on lifts and escalators will mainly affect the availability, service points, ambience and asset health. The predominant factor being availability. The ambience contribution is principally through cleaning activities.		Capex expenditure on lifts will have a similar range of benefits to opex but with greater emphasis on asset health.					

Table 3: Provisional assumptions with regard to impact of costs on outputs

	Rolling Stock	Signalling	Stations	Track	Power	Civil	Depots	Lifts & Escalators
Availability	H	H	M	M	L	L	M	H
Ambience	H	L	H	L	L	L	L	H
Asset Health	H	H	M	H	M	M	M	H
Service Points	M	M	H	L	L	L	L	M

H= High

M= Medium

L=Low

8.8 Alternative approaches to adjusting for size

We recognise that it would be possible to adopt a more “top down” approach to adjusting costs for size, but in our view this would create a number of conceptual problems which would be more difficult to overcome than adjusting for costs using the relatively “bottom up” approach described above.

9 Whole Life Asset Management (WLAM)

Whole Life Asset Management (WLAM) plays a central role in the PPP contract, which states that:

*“Infracos shall adopt and demonstrate an efficient and economic whole-life cost approach to decisions regarding the balance between maintenance, renewal and enhancement of Assets, regardless of when in the Contract Period such decisions fall to be made which approach shall be established by reference to Good Industry Practice”.*²⁹

As explained previously, it will also be important for the benchmarking framework to consider the Infracos’ different approaches to WLAM. In our view, this could be achieved by:

- Calculating discounted present values of cost per unit of output.
- Analysing time series of costs per unit of output³⁰.

9.1 Discounted present values of cost per unit of output

This approach would require us to:

- Consider what time period should be considered. On the one hand, use of a relatively short time period should mean that forecast costs and outputs are likely to be relatively accurate. However, short time periods may not fully take into account the longer term focus of WLAM. On the other hand, whilst a relatively long time period would be more consistent with the concept of WLAM, average asset lives in the context of the PPP, and the contract itself, the main drawback is that forecasts further into the future tend to be less accurate.
- Obtain forecasts of costs. These forecasts would be sourced from the Infracos’ AAMPS and/or their financial models and could be Level 1 or Level 2 in nature. Moreover, these forecast costs could be unadjusted or adjusted for size.
- Obtain forecasts of output (eg in the case of availability this would be LCH, from which we would have to calculate projected levels of LCHS).
- Consider whether the value which we attach to output should vary over time. For example, it might be appropriate to index forecasts of output to expected increases in real wages or real personal disposable incomes (on the grounds that as consumers become better off they would attach more importance to a better service).
- Calculate cost per unit of output for each period under consideration.
- Select an appropriate cost of capital. This could be the subject of considerable debate, particularly for Level 2 asset group or asset class benchmarks.
- Calculate discounted present values of these forecasts of cost per unit of output by using the opportunity cost of capital.

²⁹ SCHEDULE 3.1 - Asset Management Regime of the PPP Contract.

³⁰ These approaches are, however, not mutually exclusive.

- Aggregate these discounted present values of forecast cost per unit of output.
- Compare these aggregated figures between lines.

In summary, this approach is conceptually appealing but:

- The calculations could be complex and might result in a certain amount of “spurious precision”.
- Asset groups may, as a result of their different characteristics, require analysis over different time periods (for example, certain asset classes might be considered to have longer lives than others).
- It places a great deal, perhaps too much, emphasis on the reliability of the Infracos’ forecasts.

9.2 Analysis of time series

The alternative would be to plot the various benchmarks (and underlying cost and output data) over time and see, by inspection, what this tells us. These plots would cover past, current and projected data.

Whilst this approach is less elegant from a theoretical perspective, in our view it has a number of advantages. For example:

- It would make it easier to understand what is actually happening with the Infracos’ costs and outputs.
- It would retain some of the granularity which might be lost if we were to rely solely on NPV calculations.

For example, at the moment it is difficult to predict precisely which information generated by the framework will be particularly interesting. In our view, “seeing” the data (along the lines of how it is depicted in Annex 3, figures 2 and 3), rather than relying on relatively abstract NPV calculations should help to overcome this and shed insights into crucial issues such as:

- Whether all of the lines can genuinely be compared against each other. On the one hand, it might be the case that some lines are so different to the others that comparisons simply cannot be made. On the other hand, if such differences do exist, then it might be possible to overcome this through some type of normalisation process. But, in order to make this judgement, it will be important to understand the nature and magnitude of whatever differences actually do exist.
- What it might be reasonable to expect from a particular line, given what is being/has been delivered by others. For example, it might be easier to make such decisions if the plots of benchmarks for the various lines are relatively similar/not too variable. But it will again be important to understand the nature and magnitude of whatever differences actually do exist in order to make this.
- What the “marginal cost” of possible changes to future outputs might be by, say inspecting how various lines’ costs and outputs have changed over time. This would be particularly difficult to capture from NPV calculations.

10 Data sources & other issues with regard to comparability

This section summarises our initial views on what data sources we suggest should be used during piloting of the framework.

With regard to outputs, the framework should:

- Use data from CUPID (for availability, ambience³¹ and service points).
- Use the single source truth of assets (SSTA) database for asset health.
- Use MPD for capability and input specified projects.

It will, however, be necessary to discuss with the Parties our suggested approach to calculating LCHS and SPS.

With regard to costs:

- The framework should generally use data from the AAMPS (mainly the pricing schedules³²) and the Infracos' financial models³³.
- It will be important to validate that there is a clear mapping between the costs reported in the above sources and the asset classifications we suggest are used by the framework (ie asset groups, asset classes and asset categories).
- With regard to Level 3 benchmarks:
 - Capability – The framework will require relevant data from the MPD and discussions with the Infracos. Our approach to roiling stock will also require us to obtain information from external sources.
 - Stations – The framework will require relevant data from the MPD and from discussions with the PPP Parties.
 - Other Level 3 analysis – The information required here will depend on the results of Level 1 and Level 2 analysis, so cannot currently be specified.

The main issues with regard to comparability are likely to include:

- PPP Parties' views on whether and how to adjust for size (and other parameters such as asset quality and/or complexity).
- PPP Parties' views on how to apportion costs, and how to assess the likely impact of this expenditure on each of the outputs.

³¹ It will, however, be necessary to ascertain to whether disaggregated information on ambience is available from CUPID or whether it will have to be obtained from an alternative source.

³² If these are available. So far, we have only managed to obtain a relatively dated one for Tube Lines (JNP).

³³ On the assumption that we will be able to obtain up-to-date versions of these models.

-
- Different approaches to reporting admin/on costs. For example, we understand that as much as 25% of Tube Lines' costs are reported as "admin/on costs", which is very different to the approach adopted by Metronet.
 - Potentially different approaches to data reported in the MPD and/or elsewhere.
 - Potentially different approaches to recording capex/opex. This could be particularly important to calculations of TCEMP.
 - Different approaches to contingencies.
 - Station costs. More specifically, it may be necessary to establish whether these costs are appropriately allocated to different lines and, if not, how this might be done.
 - Rolling stock costs. It may again be necessary to establish whether these costs are appropriately allocated to different lines and, if not, how this might be done.
 - Potentially different approaches to and reliability of forecasts. This could be particularly important to the analysis of WLAM.

11 Conclusion & Next Steps

This report concludes Stage 1 of our work. Table 4 provides a brief summary of how the benchmarking framework meets the OPPPA's high level requirements.

In order for us to progress our work, the crucial requirements now are to:

- Obtain constructive and timely input from the PPP Parties on the key issues highlighted in this document and in the main body of our report.
- Commence piloting the framework because, to a very large extent, it is only when the framework is populated with real data that we will see what insights it is capable of generating.

DRAFT

PricewaterhouseCoopers LLP / Bovis Lend Lease

30th August 2005

Table 4: Mapping high level requirements for the framework to proposed approach

High level requirement	Proposed solution
1. Inform the guidance or directions which the Arbiter may be required to give in relation to, in particular, Periodic Review. Framework should therefore be capable of considering not only the current performance regime but also the implications of potential changes to the regime and the outputs which the Infracos may be required to deliver	<ul style="list-style-type: none"> ● Focus on outputs rather than inputs ● Focus on inter- rather than intra-line comparison ● Use profiles of costs per unit of output to obtain an understanding of marginal costs of changes to outputs ● Anchor framework in the Infracos' own information and planning systems
2. Reflect the impact of the Infracos' obligations and largely output based incentive regimes and make explicit linkages between the costs incurred by, and outputs delivered by, the different Infracos	<ul style="list-style-type: none"> ● Focus on outputs rather than inputs ● Use a range of benchmarking techniques but emphasise cost per unit of outputs (rather than costs per unit of intermediate output or input) ● Anchor framework in the Infracos' own information and planning systems
3. Shed light on differences between Infracos in the ways in which they respond to their obligations and the output based incentive regime. In particular, to the extent appropriate, it should incorporate the different approaches of the Infracos to the concept of whole life asset management and to their supply chains	<ul style="list-style-type: none"> ● Focus on inter- rather than intra-line comparisons ● Carve out capability and input specified projects ● Consider two possible approaches to assessing whole life asset management ● Use of data cleaning and normalisation to capture approach to supply chains
4. Inform the Arbiter's understanding of the Infracos' relative performance, how this is changing over time and how "leading indicators" might be used to understand likely future performance	<ul style="list-style-type: none"> ● Focus on inter- rather than intra-line comparisons ● Generate profiles of costs per unit of output to understand key relationships
5. Build on the work already carried out CEPA	<ul style="list-style-type: none"> ● Use a range of benchmarking techniques but shift focus to outputs (and away from intermediate outputs and inputs to outputs) ● Build on CEPA work for Level 3 analysis where appropriate
6. Minimise the number of metrics where comparisons are made in order to reduce administrative burden	<ul style="list-style-type: none"> ● Focus on outputs rather than inputs ● Adopt top-down approach and use filters ● Anchor framework in Infracos' own information and planning systems
7. Strike an appropriate balance between being robust and practical	<ul style="list-style-type: none"> ● Important throughout
8. Be flexible and capable of targeting specific differences between the Infracos which merit further investigation and how this investigation might be carried out	<ul style="list-style-type: none"> ● Use filter based drill-down together with basket of techniques for diagnosis ● Anchor framework in Infracos own information and planning systems

Annex 1: Example performance regime scores

This Annex provides example data on benchmark, unacceptable performance levels and cap scores for availability, capability and ambience.

Availability

Availability Benchmark Table (For Payment Period 1 to 3 2003/04)			
	Cap (hours)	Benchmark (hours)	Unacceptable (hours)
BCV			
Bakerloo Line	0	58,600	83,800
Central Line	0	214,100	305,900
Victoria Line	0	112,900	161,200
Waterloo & City Line	0	4,600	6,600
SSL			
Metropolitan Line, Circle Line and Hammersmith & City Line	0	330,000	471,500
East London Line	0	2,800	4,000
District Line	0	388,300	554,700

Note: for District Line the benchmarks are for periods 1 and 2 only

Source: Schedule 4.1 of Appendix 2 - PPP contract (for BCV, SSL, JNP)

Availability Benchmark Table (For Payment Period 11 2002/3)			
	Cap (hours)	Benchmark (hours)	Unacceptable (hours)
Jubilee Line	0	78,867	112,700
Northern Line	0	112,929	161,300
Piccadilly Line	0	183,000	261,400

Note: the benchmarks for the Jubilee Line change, from the earliest of either the Implementation date for the Jubilee Line Specified Upgrade (as set out in Appendix 5A of Schedule 2.1 of the PPP Contract), or the 14th November 2009

Source: Schedule 4.1 of Appendix 2 - PPP contract (for BCV, SSL, JNP)

Capability

Aggregate Capability Benchmark Table			
	Cap	Benchmark	Unacceptable
BCV			
Bakerloo Line Grouping	12.72 mins	16.37 mins	19.64 mins
Central Line Grouping	20.81 mins	24.00 mins	28.80 mins
Victoria Line Grouping	12.90 mins	16.88 mins	20.26 mins
Waterloo & City Line Grouping	11.43 mins	14.52 mins	17.42 mins
SSL			
Southern SSL Line Grouping	23.91 mins	29.29 mins	35.15 mins
Northern SSL Line Grouping	20.97 mins	27.18 mins	32.62 mins
East London Line Grouping	21.44 mins	22.51 mins	27.01 mins
JNP			
Jubilee Line Grouping	14.99 mins	20.91 mins	25.07 mins
Jubilee Line Grouping	15.63 mins	22.68 mins	27.22 mins
Northern Line Grouping	16.41 mins	21.99 mins	26.39 mins
Piccadilly Line Grouping	19.74 mins	26.63 mins	31.96 mins

Note: the first Jubilee line scores are until 6th January 2012, the second are from 7th January 2012

Source: Schedule 4.1 of Appendix 2 - PPP contract (for BCV, SSL, JNP)

Ambience

Ambience Benchmark Table			
	Cap	Benchmark	Unacceptable
All BCV Lines	82 MSS score	68 MSS score	61 MSS score
All SSL Lines	82 MSS score	70 MSS score	63 MSS score
All JNP Lines	82 MSS score	71 MSS score	64 MSS score

Source: Schedule 4.1 of Appendix 2 - PPP contract (for BCV, SSL, JNP)

Annex 2: Asset groups and classes

Introduction

This section provides more detail on some of the definitions in the report.

We consider there to be:

- 4 “Asset Groups”. These are:
 - Train Systems – Rolling stock, Signalling & Control, and Depots
 - Stations – Stations including station civil structures and M&E systems, and Lift and Escalators
 - Infrastructure – Track, power supply, and all non station civil structures
 - Administration – Overhead costs, finance dept, HR, IT systems
- 9 “Asset Classes”. These are defined in the table and text below.
- A very large number of ‘asset categories’. For example, these include the specific assets within an asset class for example the Jubilee Line rolling stock fleet. These are defined in the text below.

Asset classes

We consider there to be the following nine asset classes.

No.	Asset Name	Description
1	Rolling Stock	Rolling stock fleets
2	Signalling	Signalling and control systems
3	Stations	Stations and associated Civil and M&E systems
4	Track	Track system
5	Power	Power supply systems
6	Civil	All non station civil structures
7	Depots	Depots and depot equipment
8	Lifts & Escalators	Lifts and escalators
9	On Costs	Overheads, HR, Finance dept, IT systems

Rolling Stock

Rolling stock passenger train fleets. Costs included cover maintenance, improvements, refurbishment, upgrades, PFI costs and renewals ie purchase of new stock. On track plant such as cleaning, measuring and other station or infrastructure maintenance equipment is not included (this is covered under depot).

Signalling

This class covers the mainline and depot signalling systems plus the control centre and vital comms systems. The costs included cover maintenance, improvements, refurbishment, upgrades, PFI costs and renewals ie purchase of systems.

Stations

The Station asset class covers the station infrastructure (all station classes), the E&M works at stations, non vital comms, and fire protection systems. The costs included are maintenance, improvements, refurbishment, upgrades, and renewals i.e purchase of systems. Lifts and escalators are not included as these have their own asset class. Note: maintenance costs also includes cleaning.

Track

The scope of the track asset class covers the running rails the track base and its drainage. It does not cover the 3rd and 4th rail power conductors which are covered under the power asset class. The costs covered include inspection, maintenance, renewals and refurbishment.

Power

The power asset class covers the maintenance, renewal and refurbishment of the 3rd and 4th rail conductor rails, the substation feeder cables, intermediate switch, isolation and cross feed equipment and earth structures. The substations are not included as these are part of the Power PFI contract managed by London Underground.

Civil

The civils asset class covers all civils work not covered under the stations asset class e.g. tunnels, bridges retaining walls and includes the maintenance, refurbishment and renewal costs. Note: maintenance also includes cleaning costs.

Depot

The opex (maintenance) costs and capex (renewals, refurbishments and improvements / upgrades) costs for the depot and depot equipment are included in this asset class. Where the depots are operated by others the PFI costs will be used.

Lifts & Escalators

This asset class covers the costs for the maintenance, renewals, refurbishment and improvements of the lifts and escalators used to access the stations.

On Costs

This asset class is intended to cover the overhead administrative costs of the Infraco e.g. HR, Finance dept. IT systems etc. It is not intended to cover the costs of special or transition

projects. These projects should be allocated to or divided between the other asset groups which best fits the scope of the particular project.

Activity Definitions and Cost Allocations

The meaning of the following terms is defined as follows;

Renewal: The replacement of an asset with new like for like equipment. This is a capex cost.

Maintenance: Routine and non routine (i.e. fault repair) of assets to maintain its asset condition assessment score. This is an opex cost.

Refurbishment: Maintenance of an asset to enhance its asset condition score (i.e. greater than routine maintenance less than renewal). This is normally treated as a capex cost.

Improvement: Improvements to assets would be part of refurbishments but would improve asset performance (this would normally improve its availability and or service points score but would not affect its capability score). This is normally treated as a capex cost.

Upgrades: Upgrades can be part of renewals or refurbishments and may affect availability and or service points but must also improve the assets capability score. This is normally treated as a capex cost.

Cyclical based work: Regular work which needs to be repeated on a cyclical basis. This is a opex cost.

Assessment based work: Assessment of assets to determine condition. This is an opex cost.

Inspection based work: Inspection of assets to determine work that may be required to maintain or improve it. This is an opex cost.

Lines and their basic data

We also considered it useful to include some basic data for each line. This information will be updated over time.

Line	No. Stations	No. of Depots	Km of track	Km tunnel	Km surface	Fleet Size Trains	Fleet Size Cars
Jubilee	24	1	16	12.4	3.5	59	59x6
Northern	40	2	58	27.8	30.2	106	106x6
Piccadilly	36	2	71			76	
Bakerloo	10	3	23.2				
Circle	43	1	22.5				x6
Victoria	10	1	21	21	0	43	43x

Central	49	3	74			85	85x4
District	41	2	54				x6
H&C	Inc. in Cir	1	14.5				x6
Met	Inc. in Cir	1	66.7	9.7	57		x8
W&C	Inc. in Cen	1	2.37	2.37	0	5	5x4

Note: East London Line is not included in the benchmarking analysis as this is to be transferred to London Rail and as such will fall outside the scope of the OPPPA.

Annex 3: The benchmarking framework - a hypothetical example

**Note: In this example, costs are calculated on a "top down" basis. In reality, a "bottom up" approach would be used.*

Step 1: Obtain and inspect raw data

Table 1: Raw data

Line	TCEMP (£mn)	LCHS ('000)	Size Index	AH (%)	Amb (MSS)	SPS
1	30.0	20.0	1.05	50.0	50.0	250.0
2	40.0	41.0	1.00	80.0	81.0	410.0
3	20.0	19.0	1.00	39.0	39.0	190.0
4	25.0	26.0	1.00	51.0	51.0	260.0
5	30.0	29.0	1.00	59.0	59.0	290.0
6	35.0	36.0	1.00	71.0	71.0	360.0
7	40.0	40.0	1.00	80.0	80.0	400.0
8	45.0	45.0	1.00	90.0	90.0	450.0
9	50.0	49.0	1.00	99.0	99.0	490.0
10	25.0	26.0	1.00	51.0	51.0	260.0
11	45.0	45.0	0.95	90.0	90.0	450.0

Convert raw data into Level 1 benchmarks

Table 2: TCEMP per unit of output

Line	TCEMP/LCHS	TCEMP/AH	TCEMP/Amb	TCEMP/SPS
1	1.5	0.6	0.6	0.120
2	1.0	0.5	0.5	0.098
3	1.1	0.5	0.5	0.105
4	1.0	0.5	0.5	0.096
5	1.0	0.5	0.5	0.103
6	1.0	0.5	0.5	0.097
7	1.0	0.5	0.5	0.100
8	1.0	0.5	0.5	0.100
9	1.0	0.5	0.5	0.102
10	1.0	0.5	0.5	0.096
11	1.0	0.5	0.5	0.100

Level 1 Benchmarks

Conclusions

1. Cost per unit of availability is high for line 1
2. Result is mirrored in, but less marked for, other outputs
3. Important to understand cost per unit of availability in more detail

Step 2: Adjust costs for size

Table 3: TCEMPad divided by size index

Line	TCEMPad(£mn)
1	28.6
2	40.0
3	20.0
4	25.0
5	30.0
6	35.0
7	40.0
8	45.0
9	50.0
10	25.0
11	47.4

Estimate the same ratios as in table 2, using size adjusted cost (table 3)

Table 4: TCEMPadj per unit of output

Line	TCEMPadj/LCHS	TCEMPadj/AH	TCEMPadj/Amb	TCEMPadj/SPS
1	1.4	0.57	0.57	0.11
2	1.0	0.50	0.49	0.10
3	1.1	0.51	0.51	0.11
4	1.0	0.49	0.49	0.10
5	1.0	0.51	0.51	0.10
6	1.0	0.49	0.49	0.10
7	1.0	0.50	0.50	0.10
8	1.0	0.50	0.50	0.10
9	1.0	0.51	0.51	0.10
10	1.0	0.49	0.49	0.10
11	1.1	0.53	0.53	0.11

Level 1 Benchmarks

Conclusions

4. Some but only a small part of the difference highlighted above is explained by line size

Step 3: Obtain costs per asset group on an unadjusted and an adjusted basis

Allocate actual costs to the asset groups

Table 5: TCEMP allocated by asset group

Line	TCEMP	AGCEMP(trains)	AGCEMP(stations)	AGCEMP(infra)
1	30.00	15.00	7.50	7.50
2	40.00	13.33	13.33	13.33
3	20.00	6.67	6.67	6.67
4	25.00	8.33	8.33	8.33
5	30.00	10.00	10.00	10.00
6	35.00	11.67	11.67	11.67
7	40.00	13.33	13.33	13.33
8	45.00	15.00	15.00	15.00
9	50.00	16.67	16.67	16.67
10	25.00	8.33	8.33	8.33
11	45.00	15.00	15.00	15.00

Table 6: TCEMPadj allocated by asset group

Line	TCEMPadj	AGCEMPadj(trains)	AGCEMPadj(stations)	AGCEMPadj(infra)
1	28.57	14.29	7.14	7.14
2	40.00	13.33	13.33	13.33
3	20.00	6.67	6.67	6.67
4	25.00	8.33	8.33	8.33
5	30.00	10.00	10.00	10.00
6	35.00	11.67	11.67	11.67
7	40.00	13.33	13.33	13.33
8	45.00	15.00	15.00	15.00
9	50.00	16.67	16.67	16.67
10	25.00	8.33	8.33	8.33
11	47.37	15.79	15.79	15.79

Step 4: Analyse apportioned asset group costs per unit of availability

Table 7: Costs apportioned to outputs

	Train system Total Costs	Station system Total Costs	Infrastructure Total Costs
Availab	40%	45%	30%
Ambien	10%	10%	5%
Asset H	45%	15%	55%
Service	5%	30%	10%
Total	100%	100%	100%

Apportion unadjusted and adjusted asset group cost to outputs using table 7

Table 8: Costs per asset group apportioned to Availability

Line	AAGCEMP (trains)	AAGCEMP (stations)	AAGCEMP (infra)
1	6.00	3.38	2.25
2	5.33	6.00	4.00
3	2.67	3.00	2.00
4	3.33	3.75	2.50
5	4.00	4.50	3.00
6	4.67	5.25	3.50
7	5.33	6.00	4.00
8	6.00	6.75	4.50
9	6.67	7.50	5.00
10	3.33	3.75	2.50
11	6.00	6.75	4.50

Table 9: Adjusted costs per asset group apportioned to Availability

Line	AAGCEMPadj (trains)	AAGCEMPadj (stations)	AAGCEMPadj (infra)
1	5.71	3.21	2.14
2	5.33	6.00	4.00
3	2.67	3.00	2.00
4	3.33	3.75	2.50
5	4.00	4.50	3.00
6	4.67	5.25	3.50
7	5.33	6.00	4.00
8	6.00	6.75	4.50
9	6.67	7.50	5.00
10	3.33	3.75	2.50
11	6.32	7.11	4.74

Step 4: Analyse apportioned asset group costs per unit of availability (contd)

Calculate unadjusted and adjusted apportioned asset group cost per unit of output

Table 10: Apportioned costs per asset group per unit of output (Availability)

Line	AAGCEMP (trains)/LCHS	AAGCEMP (stations)/LCHS	AAGCEMP (infra)/LCHS
1	0.30	0.17	0.11
2	0.13	0.15	0.10
3	0.14	0.16	0.11
4	0.13	0.14	0.10
5	0.14	0.16	0.10
6	0.13	0.15	0.10
7	0.13	0.15	0.10
8	0.13	0.15	0.10
9	0.14	0.15	0.10
10	0.13	0.14	0.10
11	0.13	0.15	0.10

**Level 2 Benchmarks
Asset Group**

Table 11: Apportioned adjusted costs per asset group per unit of output (Availability)

Line	AAGCEMPadj (trains)/LCHS	AAGCEMPadj (stations)/LCHS	AAGCEMPadj (infra)/LCHS
1	0.286	0.161	0.107
2	0.130	0.146	0.098
3	0.140	0.158	0.105
4	0.128	0.144	0.096
5	0.138	0.155	0.103
6	0.130	0.146	0.097
7	0.133	0.150	0.100
8	0.133	0.150	0.100
9	0.136	0.153	0.102
10	0.128	0.144	0.096
11	0.140	0.158	0.105

**Level 2 Benchmarks
Asset Group**

Conclusions

- 5. Train system asset group costs per unit of availability are high for line 1
- 6. This is not true for stations and infrastructure asset group costs
- 7. Important to understand train system costs in more detail
- 8. Proceed to analysis of train system asset class costs

Step 5: Analyse train system asset group

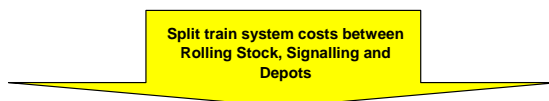


Table 12: Train system costs allocated by asset class

Line	AGCEMP (Trains)	ACCEMP (Rolling Stock)	ACCEMP (Signalling)	ACCEMP (Depots)
1	15.00	9.15	4.50	1.35
2	13.33	6.67	4.00	2.67
3	6.67	3.33	2.00	1.33
4	8.33	4.17	2.50	1.67
5	10.00	5.00	3.00	2.00
6	11.67	5.83	3.50	2.33
7	13.33	6.67	4.00	2.67
8	15.00	7.50	4.50	3.00
9	16.67	8.33	5.00	3.33
10	8.33	4.17	2.50	1.67
11	15.00	7.50	4.50	3.00

Table 13: TCEMPadj allocated by asset group

Line	AGCEMPadj (Trains)	ACCEMPadj (Rolling Stock)	ACCEMPadj (Signalling)	ACCEMPadj (Depots)
1	14.29	8.71	4.29	1.29
2	13.33	6.67	4.00	2.67
3	6.67	3.33	2.00	1.33
4	8.33	4.17	2.50	1.67
5	10.00	5.00	3.00	2.00
6	11.67	5.83	3.50	2.33
7	13.33	6.67	4.00	2.67
8	15.00	7.50	4.50	3.00
9	16.67	8.33	5.00	3.33
10	8.33	4.17	2.50	1.67
11	15.79	7.89	4.74	3.16

Step 6: Analyse apportioned asset class costs per unit of availability

Table 14: Costs apportioned to outputs

	Rolling Stock	Signalling	Depots
Availab	40%	40%	40%
Ambien	20%	0%	0%
Asset H	20%	30%	30%
Service	20%	30%	30%
Total	100%	100%	100%

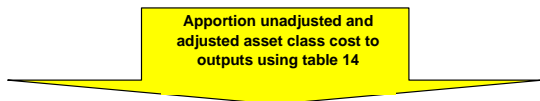


Table 15: Costs per asset class apportioned to Availability

Line	AACCEMP (Rolling Stock)	AACCEMP (Signalling)	AACCEMP (Depots)
1	3.66	1.80	0.54
2	2.67	1.60	1.07
3	1.33	0.80	0.53
4	1.67	1.00	0.67
5	2.00	1.20	0.80
6	2.33	1.40	0.93
7	2.67	1.60	1.07
8	3.00	1.80	1.20
9	3.33	2.00	1.33
10	1.67	1.00	0.67
11	3.00	1.80	1.20

Table 16: Adjusted costs per asset class apportioned to Availability

Line	AACCEMPadj (Rolling Stock)	AACCEMPadj (Signalling)	AACCEMPadj (Depots)
1	3.49	1.71	0.51
2	2.67	1.60	1.07
3	1.33	0.80	0.53
4	1.67	1.00	0.67
5	2.00	1.20	0.80
6	2.33	1.40	0.93
7	2.67	1.60	1.07
8	3.00	1.80	1.20
9	3.33	2.00	1.33
10	1.67	1.00	0.67
11	3.16	1.89	1.26

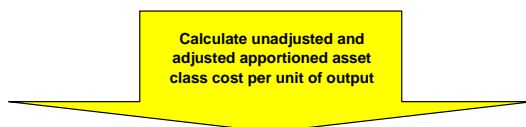


Table 17: Apportioned costs per asset class per unit of output (Availability)

Line	AACCEMP (Rolling Stock)/LCHS	AACCEMP (Signalling)/LCHS	AACCEMP (Depots)/LCHS
1	0.183	0.090	0.027
2	0.065	0.039	0.026
3	0.070	0.042	0.028
4	0.064	0.038	0.026
5	0.069	0.041	0.028
6	0.065	0.039	0.026
7	0.067	0.040	0.027
8	0.067	0.040	0.027
9	0.068	0.041	0.027
10	0.064	0.038	0.026
11	0.067	0.040	0.027

Level 2 Benchmarks
Asset Class

Table 18: Apportioned adjusted costs per asset class per unit of output (Availability)

Line	AACCEMPadj (Rolling Stock)/LCHS	AACCEMPadj (Signalling)/LCHS	AACCEMPadj (Depots)/LCHS
1	0.174	0.086	0.026
2	0.065	0.039	0.026
3	0.070	0.042	0.028
4	0.064	0.038	0.026
5	0.069	0.041	0.028
6	0.065	0.039	0.026
7	0.067	0.040	0.027
8	0.067	0.040	0.027
9	0.068	0.041	0.027
10	0.064	0.038	0.026
11	0.070	0.042	0.028

Level 2 Benchmarks
Asset Class

Conclusions

- 9. Line 1's depot costs per unit of availability not above average and relatively small
- 10. Line 1's rolling stock and signalling cost per unit of availability are above average
- 11. Drill down into rolling stock and signal maintenance costs using Level 3 benchmarks

Figure 2: Train system asset group and rolling stock asset class cost over time

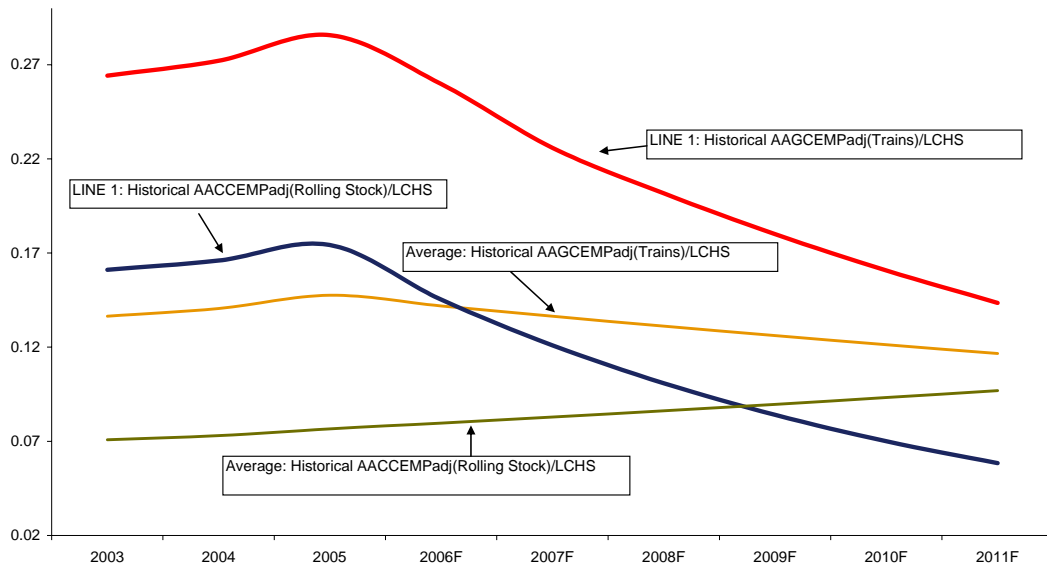
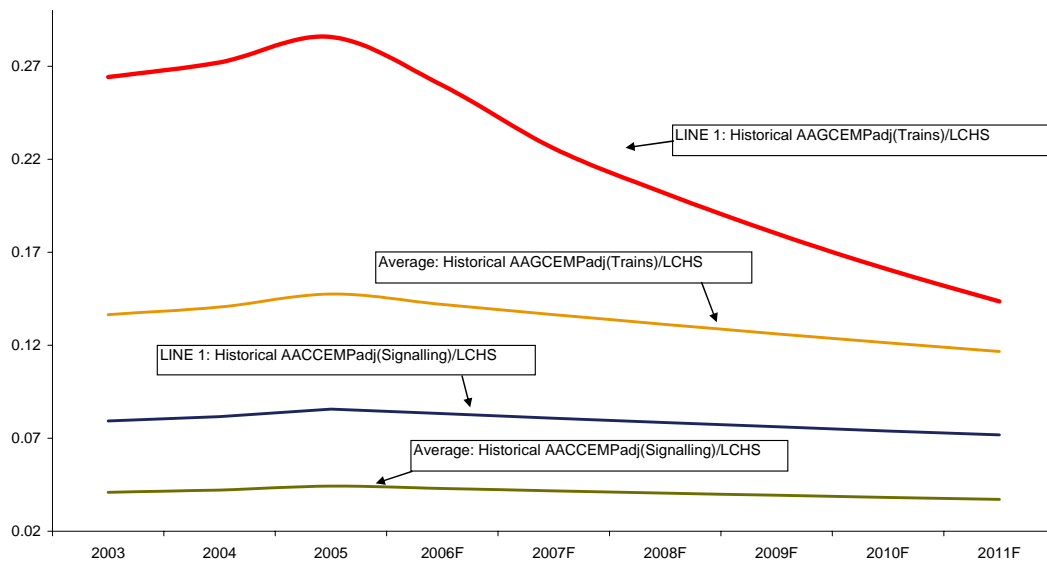


Figure 3: Train system asset group and signalling asset class cost over time



Annex 4: Adjusting for size (numerical example)

As discussed in Section 8, this annex sets out a worked example of how it might be appropriate to adjust costs for size. The example follows the steps that we propose to follow to adjust for size as set out in that section. In this example we assume 3 separate train lines (line 1, line 2, line 3).

Step 1, 2, 3

As per steps 1 - 3 the table below sets out all costs by asset class, excluding major projects. As discussed above, these costs should exclude costs attributable to major projects.

Asset Class Cost Excluding Major Projects (ACCEMP)	Line 1	Line 2	Line 3
Rolling Stock	£24	£15	£13
Signalling	£13	£25	£22
Stations	£23	£55	£28
Track	£13	£32	£18
Power	£1	£3	£2
Civil	£5	£9	£6
Depots	£3	£9	£5
Lifts & escalators	£4	£16	£9
Admin & other / "on-costs"	£15	£42	£29
TCEMP	£101	£206	£132

Step 4

As per step 4, the asset classes identified as being most important, and which will therefore be adjusted for size, are highlighted in yellow.

Asset Class Cost Excluding Major Projects (ACCEMP)	Line 1	Line 2	Line 3
Rolling Stock	£24	£15	£13
Signalling	£13	£25	£22
Stations	£23	£55	£28
Track	£13	£32	£18
Power	£1	£3	£2
Civil	£5	£9	£6
Depots	£3	£9	£5
Lifts & escalators	£4	£16	£9
Admin & other / "on-costs"	£15	£42	£29
TCEMP	£101	£206	£132

Step 5

The highlighted asset class costs in the previous table will be adjusted by using a size parameter for each asset class. The size proxies that will be used are listed in the table below.

Asset class	Proxies for size	Line 1	Line 2	Line 3
Rolling stock	Fleet Size (cars)	420	636	275
Signalling	No. of Signals	155	250	240
Stations	No. of Stations	20	50	30
Track	Track Length Km	36	57	60

Step 6

The three lines are adjusted for size using the formula presented below.

$$SizeAdjustedCost_{ij} = ACCEMP_{ij} - \left(ACCEMP_{ij} * \left[\frac{ACS_{ij}}{\left(\frac{\sum_i ACS_{ij}}{n} \right)} - 1 \right] \right)$$

Where:

i = Line (i = 1, 2, ...,n)

j = Asset Class (j = 1, 2, 3, 4)

ACC = Asset Class Cost Excluding Major Projects

ACS = Asset Class Size Proxy

As per the above formula, the average size for each asset class is calculated (see table below).

Average size by Asset Class (ACS)	
Fleet Size (cars)	444
No. of Signals	215
No. of Stations	33
Track Length Km	51

Then, for each line the percentage deviation from the average size is calculated by line.

Deviation from Average	Line 1	Line 2	Line 3
Fleet Size (cars)	-24	192	-169
%	-5%	43%	-38%
No. of Signals	-60	35	25
%	-28%	16%	12%
No. of Stations	-13	17	-3
%	-40%	50%	-10%
Track Length Km	-15	6	9
%	-29%	12%	18%

The percentage deviation from the average size is then used to adjust ACCEMP data in the first table. A size estimate that is less than the average effectively adjusts the ACCEMP upwards, and a size estimate that is more than the average effectively adjusts the ACCEMP downwards. Note that the costs below are not yet apportioned to the different output measures.

Asset Class Cost Excluding Major Projects adjusted for size (ACCEMPadj)	Line 1	Line 2	Line 3
Rolling Stock	£25	£8	£18
Signalling	£17	£21	£19
Stations	£32	£28	£31
Track	£17	£28	£15
Power	£1	£3	£2
Civil	£5	£9	£6
Depots	£3	£9	£5
Lifts & escalators	£4	£16	£9
Admin & other / "on-costs	£15	£42	£29
TCEMPadj	£119	£164	£134

The ACCEMPadj are then aggregated to arrive at TCEMPadj.

Step 7

The ACCEMPadj can then be apportioned to the different outputs as set out in table 2.

Annex 5: Directory of Acronyms

AACCEMP	Apportioned Asset Class Cost Excluding Major Projects
AACCEMPadj	Apportioned Asset Class Cost Excluding Major Projects, adjusted for size
AAGCEMP	Apportioned Asset Group Cost Excluding Major Projects
AAGCEMPadj	Apportioned Asset Group Cost Excluding Major Projects, adjusted for size
ACC	Asset Class Costs
ACCEMP	Asset Class Cost Excluding Major Projects
ACCEMPadj	Asset class costs excluding major projects, adjusted for size
AGC	Asset Group Costs
AGCEMP	Asset Group Cost Excluding Major Projects
AGCEMPadj	Asset Group Cost Excluding Major Projects, adjusted for size
ATCEMP	Apportioned Total Costs Excluding Major Projects
ATCEMPadj.	Apportioned Total Costs Excluding Major Projects, adjusted for size
BLL	Bovis Lend Lease
CEPA	Cambridge Economic Policy Associates
LCH	Lost Customer Hours
LCHS	Lost Customer Hours Score
MSP	Maximum Service Points
MSS	Mystery Shopping Survey
OPPPA	Office of the PPP Arbiter
PwC	PricewaterhouseCoopers LLP
SPS	Service Points Score
SSTA	Single Source Truth of Assets
TC	Total Costs
TCEMP	Total Costs Excluding Major Projects
TCEMPadj	Total Costs Excluding Major Projects, adjusted for size
TCEMPadj	Total Cost Excluding Major Projects, adjusted for size
WLAM	Whole Life Asset Management