Reflecting the fiscal pressures of the times, there has been a great deal of talk recently about improving ‘Value for Money’ in the rail industry. Government ministers and regulators have latched onto the conclusion of the McNulty report, that GB Rail costs are 20% to 40% higher than European peers, and there is an ‘efficiency gap’ of 40%.

They are also unhappy that, despite privatisation, unit costs have not fallen since the 1990s. And they have embraced McNulty’s recommendation ‘the industry should be aiming to achieve a 30% reduction in unit costs (ie costs per passenger km) by 2018/19’.

McNulty’s conclusions suit the Government, which wants to find ways to cut spending, while blaming its predecessors for being profligate. And there is a growing consensus (which I share) supporting McNulty’s main recommendation, that Network Rail should be broken up into independent regional units, potentially semi-integrated with train operators.

However, McNulty’s conclusion that GB Rail costs are higher is not supported by reliable evidence. Moreover, the unit costs measure is not evidence of failure, and indeed rising costs, provided they are matched by rising revenues, could well be a sign of success. There is plenty of room to improve GB Rail, however it is difficult to believe that it is ‘less efficient’ than wholly state-owned French and Swiss rivals. After two (and in some cases three) rounds of competitive franchising, and with the fastest passenger growth of any country in Europe, the opposite seems more likely to be true.

Indeed, the only real evidence that McNulty shows of European railways being more efficient, is the experience of small German train operating companies (TOCs), where competitive tendering by regional authorities has shown savings of about 20% in comparison with Deutsche Bahn. What McNulty doesn’t pick up is that the German TOCs are all much smaller, typically a tenth the size of ours.

A decade after I argued that it was time to break up Railtrack, it is now being accepted as the way forward. Maybe it is also time to end a decade of trying to combine franchises, and accept that it is smaller franchises (think Chiltern, c2c, Merseyrail, Anglia) that are best at delivering improved services and managing costs. This would complement the Government’s ‘localism’ agenda.
Fewer larger franchises may be easier for Whitehall and ministers to understand, but more, smaller franchises (Germany has over 100) will lower the barriers to new bidders, encourage innovation (which McNulty notes is now lacking) and probably also give better value. And, by the way, keep the industry from being consolidated into a handful of foreign, state-owned operators.

Cold numbers
McNulty clearly wanted quantitative evidence to support his recommendations. There is an old saying that ‘if you can’t measure it, you can’t manage it’. And it is convenient when simple numbers prove something you already believe to be true.

But McNulty’s use of benchmarking is deeply flawed. His figures are based on a single benchmarking study by Civty, a German consulting firm. They do not actually prove that GB Rail is inefficient; indeed, on deeper examination, they can equally be used to show the opposite.

McNulty focuses on cost per passenger kilometre as his measure of value for money, ignoring the very different values of travel to different passengers. All other things being equal, it is better to do the same with less, but in the rail industry all other things are never equal. A rail industry that blindly pursues cost reduction is likely to attract fewer passengers, require larger subsidies, and spiral into rapid decline.

Benchmarking is a fancy name used by consultants for doing what managers have always done with numbers. How much does your business produce? How long does it take? And how much does it cost, in comparison with your peers or competitors?

McNulty is new to the rail industry (his background is in aviation). Railways generate masses of statistics, and they all have a story to tell. But each railway is the product of unique geographic and historical circumstances; no two are alike. Benchmarking really only works when you are comparing apples and apples.

Benchmarking railways can be very misleading if one does not look behind the numbers. It is easy to prove anything – or nothing.

Mixed fruit
McNulty’s first problem is that he benchmarked mostly at a national level. This assumes, implicitly, that GB Rail, and the continental systems Civty looked at are each, in aggregate, more or less comparable. In fact, instead of apples, or even oranges, each country’s railway is a basket of mixed fruit.

Benchmarking would be a lot easier if each railway served a single route, for a single type of traffic, with a single type of rolling stock, and with no interaction or overlap with any other railway. Alas this is not the real world.

Civty compared the entire GB Rail system (19 franchised train operating companies [TOCs] or ‘GBR19’), and data subsets for inter-city (IC), London & the Southeast (LSE), and regional TOCs (Civty uses the German name ‘Regio’), against:

- SJ (Swedish Railways), with a mix of higher-speed inter-city services and Stockholm suburban services. Local and regional services are now mostly franchised to competing operators, and not included in the SJ data.
- SNCF (French Railways), which operates mostly long distance high speed trains and intensive commuter services around Paris. Local services are apparently included, although compared with Britain these are few and far between.
- NS (Dutch Railways), which is essentially a large commuter operation. The Randstad is the size of southern England. Longer distance services are mostly operated by Thalys, NS Highspeed, or DB, and so are not included in the NS data.
- SBB (Swiss Railways), like NS, is dominated by suburban and middle distance Intercity services. Most local and regional services are not included in the SBB data.

To the lay person who is not an expert in railways, all five railways may look similar. But the characteristics of each type of service vary widely. French high speed trains run faster, but also generally have further to go, because France is a bigger country.

London commuters mostly like to live in villages beyond the green belt; continentals are more likely to live in tower blocks, perhaps in a 1970s new town.

French, Swiss, and Dutch railways can all use double-deck trains, with 30% more capacity and thus lower costs per passenger. London commuter trains need to be smaller, but they are also usually more frequent, run further out (London uniquely has a statutory green belt) and serve many more small stations. It costs...
more to serve English suburbs, but people also seem to prefer them. However much they may groan about high fares and crowded trains, the high price of houses in places like Woking and Sevenoaks is proof that the price and quality of commuter services is one commuters accept.

Normalisation

Civity normalised the international data for exchange rates, degrees of electrification, multiple tracks, travel speeds and distances between stops. How they did this is not disclosed in any detail, even in the consultancy’s full report. For example:

- How do you distinguish renewals from enhancements? Different railways apply different policies when preparing their accounts.
- How do you apportion shared assets, such as locomotives, which may be used by more than one operator?
- How do you adjust for train frequency, and demand peaking by time of day or direction?
- How do you apportion joint costs between operators with overlapping geographies?
- How do you treat ‘lumpy’ capital charges, for example for rolling stock? How do you treat leasing costs as compared to purchasing?
- How do you distinguish renewals from enhancements? Different railways apply different policies when preparing their accounts.
- How do you match data on Network Rail’s nine Routes to the 25+ TOCs?
- How do you normalise shared assets, for example depots, stations, and even locomotives, which may be used by more than one operator?
- How do you account for differences in gearing, debt-to-equity ratios, tax rates, and interest rates?
- How do you apportion joint costs between operators with overlapping geographies?
- How do you account for differences in operating costs, such as labour, fuel, and maintenance?
- How do you account for differences in capital costs, such as investment in rolling stock and infrastructure?
- How do you account for differences in regulatory frameworks, such as competition, subsidies, and taxes?
- How do you account for differences in economic conditions, such as GDP growth and inflation?
- How do you account for differences in geopolitical considerations, such as war and terrorism?

Figure 1: GB rail train utilisation is significantly lower than comparator countries

This is the full table from Civity’s report. It seems McNulty chose not to include figures for LSE, ‘Regio’ (British Rail’s Regional Railways Sector) and IC in his final report, which actually compare rather favourably with the European comparators.

Figure 2: The GB’s train utilisation is at the lower end of the sample

This actually just shows average train loadings; it comes from McNulty’s summary report.

- Average utilisation of trains is particularly high in France.
- This is very much driven by SNCF’s high speed system which is accounting for a large share of the passenger transport supply. These trains have a large capacity (~500 seats), fairly long sets (~240m), partly use double stack coaches (TGC Duplex) paired with a high demand.
- Utilisation of TGVs was 78% in 2007.
- Is reliable cost data even available, reflecting all costs and subsidies including taxes, pension liabilities etc?
- What value do you place upon operating performance and service quality, both of which come at a price?
- The list could go on and on. Anyone who has ever tried benchmarking railways knows why this is a ‘science’ best done by consenting adults behind closed doors, preferably in the dark.
- Sceptics say that ‘normalising’ is really just a fancy word for ‘we multiplied and divided different numbers together until we got an answer we liked’.

Getting the ‘right’ conclusion

Of course, one can make ‘guesstimates’, informed or otherwise. McNulty clearly wanted Civity to infer some conclusions, and the company obliged. Whether the conclusions are accurate and useful, or misleading and potentially even dangerous, is another matter. Let’s consider just a few.

Let’s start with the contention GB rail train utilisation is significantly lower than comparator countries. Higher utilisation is more efficient, right? As shown in Figure 1, taken from McNulty’s report, GB Rail’s average load of 107 per train, is significantly lower than the other national operators. Looks inefficient, doesn’t it? However, if one digs into the Civity report, available on the Office of Rail Regulation (ORR) website, one finds a slightly different figure (Fig 2), with loadings also for GB ‘Regio’ inter-city and LSE. This tells a somewhat different story.

Loads on GB Regional are indeed very low, on average 62 passengers per train. But loads on GB inter-city (138 passengers) and LSE (119 passengers) are pretty similar to SBB at 122 per train, and not a lot lower than NS and SJ with 137 and 140 each.

GB Rail looks ‘bad’ because the data for GB Rail includes regional services, which are (conveniently) mostly excluded from the SJ data (because they are run by other operators), SBB (because they are mostly narrow gauge or other operators), and NS (because the country is so small there hardly are any, and the few there are have been tendered to other operators). McNulty may have been seduced by Civity’s misleading strapline ‘GB train utilisation is at the low end of the sample’.

Read the label

It is easy to be blinded by numbers, and forget to ask whether a chart is even labelled correctly. Normally, ‘utilisation’ is ‘capacity used’ as a proportion of ‘capacity provided’ - right? While the side notes do acknowledge that SNCF’s trains are very large, here there is an implicit assumption that all trains are the same size.

Which, of course, they aren’t. The chart label is incorrect – it should be ‘Average train load’. Why didn’t Civity present data on load factors? This would certainly give a better indicator of ‘utilisation’.

Frequency attracts

But what does average load per train tell us about ‘Value for Money’, anyway? Smart
operators use big trains on busy long distance routes and small trains on low-density regional routes. SNCF Voyages achieves the highest loads, because French cities tend to be further apart – inter-city coaches are banned, and there are tolls on the motorway. So passengers are happy with services running hourly.

UK operators have known since the 1980s that smaller, more frequent trains can attract more passengers, with higher revenues more than offsetting any increase in costs. My own company, GB Railways, did this on the London – Norwich route. We converted an hourly service with electrified 10-car loco-hauled trains to a half-hourly service with a mix including shorter DMUs. Additional revenues more than paid the extra costs.

Virgin increased services to three trains per hour because it makes more money, and requires less subsidy. But it surely also lowers average loads. Reverting to an hourly service wouldn’t increase ‘Value for Money’.

Lower average train loads might just as likely be a good thing, not a bad one. Somebody on McNulty’s advisory panel should have explained this to him.

Use of infrastructure

‘Switzerland and the Netherlands use their infrastructure more intensively than UK. Well, that’s another catchy strapline, but looking at the numbers (Fig 3) the real story is a bit different.

While again GB Rail looks ‘bad’ in aggregate, it seems the London & Southeast routes are used more intensively than the Swiss or Dutch systems. Where traffic density supports it, GB Rail can operate trains as intensively, indeed more intensively, than the Europeans. It’s just that the GB system is much larger, because Britain is a larger country, and also includes lower-density inter-city and regional lines which bring down the average.

SNCF and SJ look even less ‘efficient’, in this respect, because they serve even larger countries.

Exclude Wales, Scotland, and northern England, and what is left of GBR becomes a more intensive user of infrastructure than the comparators. Without actually changing a thing. The caption could have been ‘Network Rail has some of the most intensively used infrastructure

Figure 3: Switzerland and the Netherlands use their infrastructure more intensively than UK

This is extracted from Civity’s report. NR is Network Rail; RFF, ProRail and TRV are the network operators in France, the Netherlands and Sweden. Note that London & Southeast (LSE) actually has the most intensive operation. The Dutch and Swiss networks are also pretty intensively used, but that is because they serve small, densely populated countries and don’t have the long distance routes or remote regional services that bring down the GBR average.

1 Route-km calculated as sum of route lengths of the individual TOCs, normalised to the real total route-km. Approximate values only, therefore.
2 Train-km of all operators would lead to a value of 9 train-km/route-km.
High costs?

'GB's costs for network operations are the second highest.' Well, it depends whether you are one of those people who thinks a cup is half empty or half full. Civity only presents data for four operators (Fig 4), and the comparators are, for some reason, now only identified by code letter.

NR (Network Rail) costs are three times higher than operator 'B', but 'B' doesn't seem to have any stations! And NR's costs are about 50% less than 'A', which seems pretty good. Operator 'D' has slightly lower costs per track-km, but the difference (about 5%) is tiny. And costs per train-km are much higher, so 'D' is hardly a model to emulate.

Civity could equally have said 'GB costs for network operation are below average', but presumably this sort of praise for Network Rail was not what the politicians wanted to hear.

Taxpayer subsidy, the big unknown

'...taxpayer subsidy per passenger-km is substantially higher for GB Rail than in the comparator countries.' Here we have the making of a myth. Civity actually presents data (Fig 5) showing the income sources for GB Rail and four comparator operators, but these are not countries. The comparators are the largest national operators, but they mostly are long distance and commuter operators. It is the regional operators in Sweden and Switzerland, excluded from Civity's data, that get the most subsidy. Bit of a slip here!

According to the graph, total income (and thus by implication total cost) on the other four railways is about half the level of GB Rail. Is this really true? And if it is, why is McNulty only complaining of a 40% efficiency gap? This looks more like 100%!

Without seeing Civity's underlying data, it's impossible to tell exactly what is missing. Only the GB data is broken down between state funding of infrastructure, state funding of operations, and passengers. For the other operators, there is identification of 'infrastructure state funding' and in one case 'rolling stock funding'.

We know operating subsidy is substantial on all continental railways, but practically nothing is shown. Some operators include it with 'train operations revenue', even though it is paid by government, not passengers. Civity actually states that GB is the only country in the sample where train operating companies' income includes a large share of state funding which seems misleading in the extreme! SNCF and SJ long-distance services may indeed be profitable, or at least cover their operating costs, but French, Swiss, and Swedish local services receive large operating subsidies that don't seem to show up in Civity's graph.

Elsewhere, McNulty seems to recognise that nobody really even knows how much the railways are subsidised. Subsidies are now paid through local and regional governments. There are unfunded pension liabilities and periodic debt writeoffs. In some countries rolling stock, once purchased by the taxpayer, is treated as 'free', while in the UK and some other countries it is leased. It is accepted that average fares are higher in the UK, but there is no good factual evidence to support the claim that taxpayer support is higher in Britain: very likely it is not.

Staff utilisation

'Compared to country B staff numbers per train-km are higher in GB' says the heading to Fig 6. It is like the graph on track utilisation. All other things being equal, it is better to do the same with fewer staff, but maybe the additional staff are actually doing more?

With all its 'normalising', why is Civity presenting data per train-km? Wouldn't it be more useful to show it per train-hour?

In an efficient railway, each driver can work about 1,800 hours per year. How far they go depends on the line speed. Are SNCF TGV drivers working a leisurely 900 hours a year, but...
at 300km/h, really as efficient as British drivers working twice as many hours but at half the average speed?

Read the strapline and look at the graph carefully. It shows that GBR 19 staffing (the franchised TOCs), per train km is only about 10% above comparator ‘B’. If the average speed of GBR 19 trains is just 11% slower than ‘B’, then GBR staff are actually being more productive.

Note also the breakdown between operation and customer management staff. Presumably the first is drivers, the second station and onboard service (conductors, catering). The caption could have read ‘GBR operational staff are significantly more productive than comparator operators’. After privatisation, we did agree new terms with our drivers, increasing productivity 30% or more, so this seems perfectly believable.

That we have more ‘customer management’ staff, on trains and stations serving passengers, does not sound like such a bad thing either. Maybe they are even earning their keep, selling tickets, helping passengers on and off trains, and serving refreshments?

Actually, this is very odd ‘benchmarking’ because the conclusion is drawn against only one, anonymous comparator. Was there no data for operators A, B and D? Surely their total staff numbers are presented in their annual accounts, as is their traffic. Or did they make GB Rail look too good?

Costs
McNulty’s favoured measure of ‘Value for Money’, and the one he puts front and centre in his report, is cost per passenger km. Given that operators tailor train size and frequency to match local market conditions and infrastructure constraints, it would have been more useful to compare cost per seat-km (which is what airlines usually benchmark) or per car-km (which would nicely ‘normalise’ for different train lengths). These are what train operators actually ‘produce’.

‘Turning these into passenger-km depends on the pattern of market demand, especially peaking by time of day and direction, as well as on government regulation about fares, loading and service standards. Ryanair can get 80% load factors by cherry picking its routes, and so can offer very low fares. Network carriers like British Airways serve different, and wider markets, often flying more frequently (hourly between major cities) but accepting lower load factors. BA still makes a profit, because some travellers will pay for the frequency.

Geography counts
Even very similar UK train companies have very different load factors: South West Trains (SWT) averages 135 passengers per train while Southern only gets 106, 22% less. It’s not that Southern is less ‘efficient’: both companies run very similar trains packed full into London every morning.

But glance at a map and you will see that SWT serves Basingstoke, Southampton and Bournemouth, business centres that generate good contra-peak flows, filling some seats going in the other direction. Southern trains are more likely to return empty.

SWT has a single efficient four-track line into Waterloo, while Southern has to run trains to several different London terminals. According to recent data from the Office of Rail Regulation, total cost per passenger km is £0.19 per km in Sussex, but only £0.15 in Wessex. On this basis, would McNulty conclude that Wessex (SWT) gives better ‘Value for Money’? And if so, can anything be done about it? It’s a bit hard to change geography. (By the way, London Underground’s overall
load factor is reportedly about 15%. Until Morgan Stanley can be persuaded to move its offices to High Barnet, the Northern Line will always run pretty empty in the contra-peak direction. Good thing McNulty wasn’t asked to look at it too!

Devolution
Scotland’s average load per train is even lower, only 66 passengers, and the cost per passenger-km even higher, at £0.24. Following McNulty’s approach to its (il)logical conclusion, the best way to improve value for money in GB rail is to support Scottish devolution. If the SNP can be persuaded to take Wales and Northern, so much the better!

Aspiration
Although he speaks of a 40% efficiency gap, McNulty never actually says costs per passenger-km are 40% higher in GB Rail than in the comparators. He knows average fares are higher, and guesses that subsidy also must be higher, but as we have seen, he never really came to grips with the latter.

Civity only compared GBR 19 in aggregate against smaller subsets of the European operators, excluding the parts of the French, Swedish and Swiss railways that are most heavily subsidised. Ultimately, the basis for his expectation for reducing costs is really just an aspiration, although he does have some ideas how it might be achieved (and some of his ideas are very commendable).

The costs trap
Probably McNulty’s biggest (and most dangerous) mistake, is to conclude that in an efficient industry, average costs should always come down over time. In some industries this is indeed the case, as new technology is applied to find ways to produce more with less. We have become used to declining real prices for many products, especially those that can be produced in a country with lower wages.

In aviation, McNulty became used to costs falling, because low-cost airlines found ways to pack more passengers into each plane, to fly them more, and to pay crews lower wages. But average fares on long haul flights haven’t fallen in the same way, because fuel and capital charges are such a high proportion of costs, and travellers will still pay for frequency and comfort.

Service standards
Civity does present evidence that franchised German operators, mostly small regional concessions, have achieved 20% cost reductions over the last decade. In comparison, average costs in the UK do not seem to have come down at all, even though traffic has increased about 50%. McNulty seems to be implying that if GB Rail costs have not fallen after franchising, like they did in Germany, then this ‘20%’ is still there to be had. Maybe it is. Another possibility is that underlying unit costs really have been reduced, but this has been masked (and offset) by quality and service improvements.
McNulty implicitly assumes that quality has (and should) remain constant – a seat km in 1990 was the same as one in 2011. Clearly, much of the cost ‘saving’ in aviation was really due to a decline in service standards (and hopefully Ryanair has found the bottom!). But a railway that did this would be doomed to steady decline. To compete with cars and planes (and staying home and using the internet), trains need to run faster and more frequently, more reliably, with air-conditioning, power doors and even wi-fi.

Regulators have also decided that society can afford retention toilets, wheelchair access, delay compensation, and customer service centres that actually answer the telephone. All of this costs money.

Impressive achievement
As incomes rise, people are willing to pay more for better quality (and staff who can’t be outsourced to India are likely to want to be paid more too). It is, in fact, quite remarkable that GB 'average' costs have remained constant, and not actually risen over the past decade. What McNulty considers evidence of stagnation is, in fact, an impressive achievement.

McNulty sees high fares as a bad thing, but they might actually reflect consumer choice. There seems no doubt that average yields in the UK are 30% more than on the Continent. What is not clear, and would be interesting to know, is whether the average passenger (the ‘median’ passenger) pays more or less. Anecdotal evidence is that more than half of UK long distance passengers pay discount fares that are a lot lower than fares in the continental comparators. Maybe this is an opportunity for some useful benchmarking?

What's wrong with a few people (mostly business travellers) paying very high fares, if this pays for new investment and reduces reliance on government support?

It is certainly true that the UK rail industry costs a lot of money. And there are few industries where managers cannot see many ‘problems’ affecting efficiency. If only the raw materials were cheaper, the workers would accept lower wages, and the customers willingly pay high prices!

There is nothing wrong with setting aspirational targets. But rail service is not a homogeneous product, like tonnes of pig iron or barrels of petroleum. Costs depend on the specific route and service, and the quality that is offered.

There's no point in producing cheaper apples if what customers really want is better quality fruit, and are willing to pay more for it. Starbucks didn’t set out to sell cheaper coffee. Finding ways to produce more with less is fine, but blindly cutting costs to pursue misguided benchmarks will increase the burden on taxpayers and may well actually reduce ‘Value for Money’.

The author was a Director of GB Railways (1996-2003) and is co-founder of Hamburg Köln Express, a new train company that will commence services in Germany during 2012. The author thanks Civity for permission to reproduce the company’s graphs.