

Note on the Construction of CP25/31 Cost Benefit Analysis (CBA)

January 2026

1 Introduction

1. This note sets out more detailed information about the estimates in the CBA for *The framework for a UK equity consolidated tape (CP25/31)*.
2. We begin our explanation of the analysis presented in the CBA and the annex to the CBA (Annex 2 and Annex 3 in CP25/31), as the analysis presented there is used to build the estimates included in the main text.

Cleaning data on demand for the consolidated tape

3. In the April 2025 survey of equity market participants, we asked questions on the demand for four different scenarios for a consolidated tape (CT). We used these questions to build a demand curve for each scenario.
4. We asked respondents the following questions for each scenario:
 - Would you find such data useful?
 - The approximate number of internal users you would buy the tape for under this scenario (assuming only individual user licences are available) if we priced at £50 per month, £200 per month or £500 per month.
5. We undertook cleaning and checking of the data received in response to these questions, to make sure that each respondents' answers were consistent across the survey. Notably, we assumed that, where respondents indicated a level of demand for one scenario at a given price, then they would have at least as much demand for a scenario which offered more data at the same price.
6. We also cleaned the data to ensure the midpoint was used for any ranges reported and to remove any formatting.

2 Estimating demand

7. To estimate the demand for each scenario, we ran the following regression for all respondents that reported non-zero demand for the equity tape for at least one scenario.¹ The regression seeks to identify the relationship between price and demand. The estimated model is:

$$Quantity_{i,p} = \alpha + \beta Price_{i,p} + \delta Price_{i,p}^2 + \varepsilon_{i,p}$$

where Quantity reflects the number of users reported by the respondent, Price is the different price points at which respondents were asked to estimate demand for the tape, i is the index number of the respondent, $p=1,2,3$ for each of the price levels asked about in the survey. In addition, α is the intercept, β is the marginal impact of a 1 unit (£ increase in price), δ is the marginal impact of a 1 unit increase in the square of the price. ε is random variation in the demand for the tape.

8. For example, respondent 50 answered 10, for the question about the number of users they would buy the tape for in scenario 2 at £50 per month. This would mean i is 50, p is 1 for the first price question, Price is £50, $Price^2$ is 2,500 and Quantity is 10 as that is the demand the respondent estimated for this survey response.

9. We use a quadratic function because there is non-linearity in responses. We considered other functional forms but, given for some prices we observe zero demand, we think our choice is appropriate.

10. We also considered whether we should run separate demand regressions for different group types. However, given the size of the sample of groups that provided responses, we did not think this was appropriate and so did not run regressions that considered both group type and size.

11. The estimates generated by these regressions enables us to estimate a demand curve for small and large groups in the survey. The estimates from these regressions are presented in the table below. These are our estimates of the parameters in the regression model described above. Our estimates of α , β , δ are in the columns labelled Intercept, Quadratic and Slope respectively.

Table 1: Slope coefficients ungrouped

Scenario	Group size	Intercept	Slope	Quadratic
Scenario 1	Large	10.62577	-0.06075	0.00008
Scenario 1	Small	3.47593	-0.01831	0.00003
Scenario 2	Large	23.72771	-0.13094	0.00018
Scenario 2	Small	4.54074	-0.02285	0.00003
Scenario 3	Large	54.85940	-0.34423	0.00048
Scenario 3	Small	4.88333	-0.02483	0.00003
Scenario 4	Large	82.83664	-0.47678	0.00064
Scenario 4	Small	5.15185	-0.02630	0.00004

¹ We take into account the respondents with zero demand at a later step.

Additional users of the tape

12. The section above outlines how we estimated total demand for different equity CT scenarios. We then split this observed demand into two categories: additional equity trade data users and existing users. We were able to do this by analysing responses to two questions in the survey:
 - Do you currently purchase display UK equity trade data?
 - Can you approximate how many internal users use the data you purchase?
13. We used responses to these questions to identify the number of current users of equity trade data that each group has. We can compare this to how many users they predict for each equity CT scenario at a given price point, to estimate any additional data users that would be enabled by the introduction of the CT.
14. We calculated the average number of these additional users and then scaled these averages by the population of groups of types of firms (see below for further details) to produce an estimate of the number of additional data users for each CT scenario, assuming it was priced at £50 per month. For Scenario 2, we estimate that there would be 1,819 additional users.

Scaling demand to the population

15. We scaled reported demand for the equity CT by a sample of groups to the full population of equity market participants (excluding those with de minimis trading volumes), taking into account that not all groups will have demand for the data.
16. Considering only data from our demand survey, our demand curve implies that there will be 9,171 users of the tape under Scenario 2, which is very similar to level 1 data, if it is priced at £50. Of these, we estimate that there will be 1,819 additional users according to the responses from the survey. This implies 7,350 users would substitute existing level 1 data for the CT.
17. As noted in the CBA, there are several reasons to suspect that, over the first proposed 5-year contract period for the CT (and thereafter), actual demand will significantly exceed our estimate of initial demand. In paragraphs 12 and 13 of Annex 3, we explain why. We said that:
 - Equity market participants may not immediately realise the value of the CT.
 - They may require time to adjust to using the CT.
 - They may require the CT to become established and trusted to change their demand from existing sources of data to the CT.
18. We also noted that bidders for the bond CT expected significant growth in demand for the bond tape over time.
19. Moreover, a key challenge for the survey was that respondents consistently reported that they would find the tape useful but found it difficult to estimate the level of likely demand for the tape. They found it hard to envisage demand for a hypothetical product that may have uses across a wide array of business areas, especially for large groups. A firm-level survey may have been appropriate, but we have also been told that firms manage data on a group basis. Consequently, either sampling approach has drawbacks. To account for likely underestimate, we compared the

observed demand in the survey with the current observed size of the equity data market and our estimates of the number of current users of display data.

20. We estimate there are currently 50,000 users of level 1 data (which includes both post-trade data and the pre-trade best bid and offer).² We made the following assumptions regarding how many of these current users could switch to using the tape:
 - In our core estimate, labelled *moderate adoption*, we assumed that over the first five years of the CT becoming an established and trusted data source, 50% of existing level 1 direct feed customers switch their demand to the CT.
 - Under *limited adoption*, we assumed that 25% of existing level 1 direct feed customers switch their demand to the CT.
 - Under *substantial adoption*, we assumed that 75% of existing level 1 direct feed customers switch their demand to the CT.
21. We think this is reasonable given that the proposed CT data will be at least as good for most level 1 display data users, compared to existing data sources (and in most cases it would be better). Consequently, we would expect, in most cases, these users would be better off switching to the tape. The range we suggest therefore is likely to contain the true estimate. Greater adoption above the 75% would increase the benefits we estimate and deliver more private benefits.
22. Under the moderate adoption assumption (i.e. if 50% of the current 50,000 users switch to the tape), then the growth in demand would be 3.4 times our observed demand from the survey. The growth rates for the other adoption scenarios are calculated in the same way. Our CBA applies these assumed growth rates to our estimates.

² We base this estimate on the revenue and price information that venues are required to publish under MAR9A.2.7R.

3 Assessing viability

23. To assess viability, we built a revenue function, with respect to price, from the adjusted demand curves described above. That is, we plot the relationship between revenue and prices for each scenario and for each level of adoption of the CT. For example, below is the estimate revenue function for scenario 2 with moderate adoption of the CT.

Figure 2: Estimated revenue from CT



Source: CP25/31 The framework for a UK equity consolidated tape, p.129

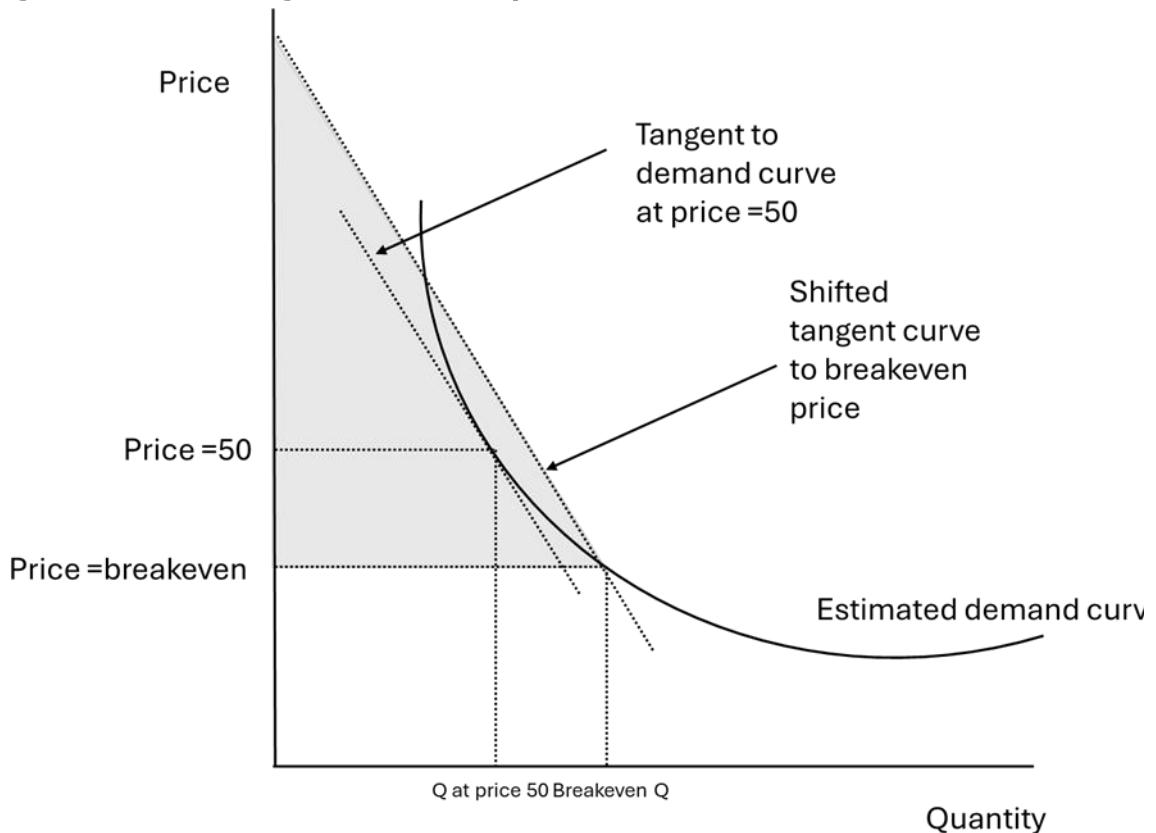
24. We then compare these functions with the range of cost information we gathered on the expected cost of running the tape, plus a 10% margin for the cost of capital and profits. We acknowledge that this margin may be lower than might be expected by bidders for the CT. The price of the CT and its realised profit margin will likely be impacted by the competitiveness of the procurement process for the CT, and we cannot accurately forecast this. We do not think this materially affects the outcomes of our CBA. This is because whilst higher profit margins will reduce viability of the CT and slightly increase the break-even price, the higher profit margin redistributes consumer surplus from data users to the CT provider. For the benefit calculation, the reduced consumer surplus for additional data users will largely be offset by higher profits earned by the CT provider on these data users in the benefit calculation. There would also be a reduction in consumer surplus for existing data users, but this will not affect our benefit calculations.

25. Using the revenue function, we find the lowest monthly price at which these estimated costs would be covered by projected revenue. Where applicable, we use these break-even prices for estimating benefits. However, in some cases there is no price at which the CT appears to break even (for Scenario 3 and 4 the CT will not be viable at the upper end of the cost estimates, in addition for Scenario 3 the CT will not break even at the midpoint of costs).

4 Benefits for users

26. As described in Annex 3 of the CP, we estimate two types of benefits for data users. These are:
 - *Additional data users* who access the CT and without the CT would not be accessing pre- and post-trade equity data, and
 - *Additional data use* by users of the CT with previous access to pre- and post-trade equity data due to improved data availability in the CT relative to these sources.
27. We address each of these calculations in turn.
28. For additional data users, the benefits come in two parts: the private benefits (consumer surplus) to the data customers and the revenue (producer surplus) generated by the CT provider from the new consumers of data.
29. The producer surplus is a benefit for the purposes of the CBA as we have already accounted for the costs of the tape. Additional revenue for the provider is a pure benefit for the CT. This is because we have assumed zero marginal cost for the tape. In reality, there will be some marginal cost to supply data users, but we think this will be small. In most cases, accessing the tape would be automated with only limited human interaction for the CTP that would likely lead to material marginal costs.
30. The consumer surplus is the value users get from the tape over and above the price they pay for the CT.
31. The producer surplus is simple to calculate. It is merely the break-even price multiplied by the number of additional users.
32. To calculate the consumer surplus, for these additional data users, we firstly need to calculate the total consumer surplus for each scenario at each break-even price. The chart below illustrates the calculations. As we have estimated a non-linear demand curve, this makes it less straightforward to estimate the consumer surplus. As the demand curve is non-linear, measuring the area under the demand curve above the price means that we may overestimate the consumer surplus, or the quadratic function means that the consumer surplus is not bounded (i.e. the demand curve does not cross the price axis implying that date users value the CT infinitely much). Also, as we move further away from the reported estimates under the survey, we become more uncertain about the likely demand. We think in all likelihood the value to users of the tape will have some upper limit (as users can spend resources accessing and aggregating existing data sources). Consequently, we use a linear approximation to the demand curve. Some of the break-even prices were above and below the £50. Choosing the tangent to the demand curve at that price may lead to higher estimates of the consumer surplus for higher prices (as the demand curve is steeper there as the second derivative is negative). We therefore used the slope at the £50 level for all our calculations.

Figure 1: Calculating consumer surplus



33. Secondly, once the total consumer surplus is calculated, we then need to allocate this consumer surplus to new users in proportion to the ratio of new users to existing data users.
34. We estimate that of current users, 50% already have access to level 1 data for Cboe venues (25,000 users), 20% already have access to level 1 Turquoise data (10,000 users), and 4% already have access to level 1 data from Aquis (2,000 users). For the users with existing access, the value of the additional data will be £0. These estimates are informed by data published by venues on their revenue and the prices they charge for data. We assume that all users of existing data already have access to LSE data and therefore there are no additional use benefits from existing users for accessing LSE data through the tape. We base these numbers on revenue and price data from the respective venues. We cannot be precise here as there is no specific price per user (prices vary by data accessed, form of access and number of users) and we have not asked venues for their split of revenues from different data sources. Using publicly available information from public sources (rather than commercially sensitive information collected from firms), we are more able to provide transparent calculations. We think that the numbers are broadly reflective of the market situation.
35. For all other users, the maximum value of the additional data will be its current market price, as given this price these users currently do not purchase the data (implying they do value it this highly). These current market prices are approximately:

- £25 per month for Cboe data.
- £10 per month for Turquoise data.
- £25 per month for Aquis data.

36. Given the variation in current customers of data from these venues at prevailing market prices, we should also expect that depending on the venue, there may be a substantial portion of the population that value this additional data at £0. Instead of assuming that all non-customers have some positive value of this additional data, we therefore assume that the population that values a venue's data above £0 but below its current market price is proportional to the current customer base of the data. That is:

- 50% (25,000) of users of the CT do not have access to Cboe data and all these 25,000 value it above £0.
- 80% (40,000) of users of the CT do not currently have access to Turquoise data and of those 25% (10,000) value it above £0.
- 96% (48,000) of users of the CT do not have access to Aquis data and of those 4.2% (2,000) value it above £0.

37. We based the estimates on the reported prices and revenues for equity data that are publicly available. Trading venues are required to publish the cost of their trade data and the revenues they earn from it under MAR9A.2.7R.

38. Finally, if we assume that the value of the data for those that assign it a positive value is uniformly distributed, the average value of the data will be halfway between the lower bound (£0) and the upper bound (£25 or £10). However, the distribution may be more left-skewed than a uniform distribution if valuations are more concentrated around £0. We therefore assume that the average value of the data for those with demand is 50% of the value assuming a uniform distribution (or 25%) of the upper bound. This would be a highly skewed distribution implying bifurcation between users to suggest a lower percentage. We therefore get valuations of:

- £6.25 per additional user of Cboe data.
- £2.5 per additional user of Turquoise data.
- £6.25 per additional user of Aquis data, all per month.

39. We think these assumptions are reasonable estimates of average valuation. We have taken the prices levels from the publicly available prices from trading venues to inform these assumptions. For example, the benefits of accessing Cboe data is £6.25 * 0.5 * demand for the CT at the break-even price level.

40. We also take in to account additional benefits of access to a greater depth of pre-trade data than is offered by level 1 data for Scenarios 3 and 4. The added value comes from the value users have for additional pre-trade data of further depth.

41. We estimate this based on the cost of level 2 data (full depth, c. £190 per month per user) relative to level 1 data (c. £50 per month) on the main market. The value for level 2 data will therefore be distributed between £0 and £140. If a level 1 customer valued this data above £140, they would purchase level 2 data instead. We assume the following:

- Scenario 4 provides 50% of the value of level 2 data (5 levels vs full depth), and
- Scenario 3 provides 25% of the value of level 2 data (3 levels vs full depth).

42. These estimates reflect that pre-trade data at levels closer to the top of the book should be more relevant for most participants than pre-trade data at higher levels.
43. If we assume that level 1 customers' value of level 2 data is uniformly distributed between £0 and £140, and that this scales proportionally with the above percentages, then:
 - The average value of the additional pre-trade data per level 1 user is £35 under Scenario 4 (£70/2 is the mean of a uniform distribution between 0 and £70).
 - The average value of the additional pre-trade data is £17.5 under Scenario 3 per user.
44. A uniform distribution, however, may overestimate the value of the additional data if the true distribution is skewed left towards lower values. We therefore use £17.5 (50% of £35) and £8.75 (50% of £17.5) as our estimates for the average additional monthly value of pre-trade data to account for a possible left skew in the distribution. Again, we use a conservative estimate of the benefits of this additional data (50% of the mid-point) to account for a potential left-skew in the distribution of valuations.

Calculating Net Benefits

45. To calculate the net benefits, we compared the estimates described in the previous sections with the overall costs of building the tape. We also adjusted these costs by £100k to reflect the familiarisations costs attributable to the tape (our estimated familiarisation costs are lower than this but difference makes no material difference to the findings of the CBA). This adjustment explains the difference between our estimates of the direct costs to the CT provider and the overall costs of the proposals.
46. We did not include the costs of participating in a procurement process within the calculations as we think they may inversely move with the overall costs of the tape. The more providers compete to provide the tape and so incur these costs, the more likely we are to move down the distribution of the likely costs of the tape. We also note that these costs are relatively small proportion of the overall costs of the tape.

5 Annex: additional tables

Table 1: Slope coefficients ungrouped

Scenario	Group size	Intercept	Slope	Quadratic
Scenario 1	Large	10.62577	-0.06075	0.00008
Scenario 1	Small	3.47593	-0.01831	0.00003
Scenario 2	Large	23.72771	-0.13094	0.00018
Scenario 2	Small	4.54074	-0.02285	0.00003
Scenario 3	Large	54.85940	-0.34423	0.00048
Scenario 3	Small	4.88333	-0.02483	0.00003
Scenario 4	Large	82.83664	-0.47678	0.00064
Scenario 4	Small	5.15185	-0.02630	0.00004

Table 2: Estimated additional data users by Scenario

Scenario 1	Scenario 2	Scenario 3	Scenario 4
1,362	1,819	1,979	1,979

Table 3: Estimated break-even price of the tape by Scenario

Scenario	Cost scenario	Cost estimate	Cost estimate including 10% profit margin	Break even price
Scenario 1	Lower end	£ 5.9m	£ 6.5m	£23.1 per month
Scenario 2	Lower end	£ 6.7m	£ 7.4m	£16.4 per month
Scenario 3	Lower end	£ 7.4m	£ 8.1m	£11.6 per month
Scenario 4	Lower end	£ 7.4m	£ 8.1m	£8.7 per month
Scenario 1	Midpoint	£ 9.6m	£10.6m	£42 per month
Scenario 2	Midpoint	£11.0m	£12.1m	£28.8 per month
Scenario 3	Midpoint	£27.8m	£30.6m	NA
Scenario 4	Midpoint	£27.8m	£30.6m	£39.2 per month
Scenario 1	Upper end	£13.4m	£14.7m	£69.9 per month

Scenario	Cost scenario	Cost estimate	Cost estimate including 10% profit margin	Break even price
Scenario 2	Upper end	£15.3m	£16.8m	£43.5 per month
Scenario 3	Upper end	£48.1m	£52.9m	NA
Scenario 4	Upper end	£48.1m	£52.9m	NA

Table 4: Estimated benefits using break even price by Scenario

Cost scenario	Scenario	Type	Break even price	Estimated initial benefits	Estimated benefits (limited adoption)	Estimated benefits (moderate adoption)	Estimated benefits (substantial adoption)	Cost estimate	Net benefits
Lower end	Scenario 1	Total benefits	£23.1	£2.6m	£ 4.4m	£ 8.8m	£13.2m	£ 6.0m	£ 2.8m
Lower end	Scenario 2	Total benefits	£16.4	£3.6m	£ 6.2m	£12.4m	£18.5m	£ 6.8m	£ 5.6m
Lower end	Scenario 3	Total benefits	£11.6	£5.8m	£ 9.9m	£19.8m	£29.6m	£ 7.5m	£12.3m
Lower end	Scenario 4	Total benefits	£8.7	£9.2m	£15.6m	£31.2m	£46.8m	£ 7.5m	£23.7m
Midpoint	Scenario 1	Total benefits	£42	£2.5m	£ 4.2m	£ 8.4m	£12.6m	£ 9.7m	£-1.3m
Midpoint	Scenario 2	Total benefits	£28.8	£3.5m	£ 6.0m	£12.0m	£18.0m	£11.1m	£ 0.9m
Midpoint	Scenario 3	Total benefits	Not viable	Not viable	Not viable	Not viable	Not viable	£27.9m	Not viable
Midpoint	Scenario 4	Total benefits	£39.2	£8.0m	£13.6m	£27.2m	£40.9m	£27.9m	£-0.7m
Upper end	Scenario 1	Total benefits	£69.9	£2.3m	£3.9m	£ 7.7m	£11.6m	£13.5m	£-5.8m
Upper end	Scenario 2	Total benefits	£43.5	£3.4m	£5.8m	£11.6m	£17.4m	£15.4m	£-3.8m
Upper end	Scenario 3	Total benefits	Not viable	Not viable	Not viable	Not viable	Not viable	£48.2m	Not viable
Upper end	Scenario 4	Total benefits	Not viable	Not viable	Not viable	Not viable	Not viable	£48.2m	Not viable

Table 5: Estimated benefits for each Scenario by type of benefit

Cost scenario	Scenario	Type	Estimated initial benefits	Estimated benefits (limited adoption)	Estimated benefits (moderate adoption)	Estimated benefits (substantial adoption)
Lower end	Scenario 1	Benefits from additional users	£2.3m	£ 3.8m	£ 7.7m	£11.5m
Lower end	Scenario 2	Benefits from additional users	£3.1m	£ 5.3m	£10.6m	£15.9m
Lower end	Scenario 3	Benefits from additional users	£3.2m	£ 5.5m	£10.9m	£16.4m
Lower end	Scenario 4	Benefits from additional users	£3.3m	£ 5.6m	£11.2m	£16.8m
Lower end	Scenario 1	Benefits from additional use	£0.3m	£0.5m	£1.1m	£1.6m
Lower end	Scenario 2	Benefits from additional use	£0.5m	£ 0.9m	£ 1.7m	£ 2.6m
Lower end	Scenario 3	Benefits from additional use	£2.6m	£ 4.4m	£ 8.9m	£13.3m
Lower end	Scenario 4	Benefits from additional use	£5.9m	£10.0m	£20.0m	£30.0m
Lower end	Scenario 1	Total benefits	£2.6m	£ 4.4m	£ 8.8m	£13.2m
Lower end	Scenario 2	Total benefits	£3.6m	£ 6.2m	£12.4m	£18.5m
Lower end	Scenario 3	Total benefits	£5.8m	£ 9.9m	£19.8m	£29.6m
Lower end	Scenario 4	Total benefits	£9.2m	£15.6m	£31.2m	£46.8m
Midpoint	Scenario 1	Benefits from additional users	£2.2m	£3.7m	£7.4m	£11.1m
Midpoint	Scenario 2	Benefits from additional users	£3.1m	£5.2m	£10.4m	£15.6m

Cost scenario	Scenario	Type	Estimated initial benefits	Estimated benefits (limited adoption)	Estimated benefits (moderate adoption)	Estimated benefits (substantial adoption)
Midpoint	Scenario 3	Benefits from additional users	£NAm	£NAm	£NAm	£NAm
Midpoint	Scenario 4	Benefits from additional users	£3.1m	£5.3m	£10.6m	£15.9m
Midpoint	Scenario 1	Benefits from additional use	£0.3m	£0.5m	£1.0m	£1.5m
Midpoint	Scenario 2	Benefits from additional use	£0.5m	£0.8m	£1.6m	£2.4m
Midpoint	Scenario 3	Benefits from additional use	£NAm	£NAm	£NAm	£NAm
Midpoint	Scenario 4	Benefits from additional use	£4.9m	£8.3m	£16.7m	£25.0m
Midpoint	Scenario 1	Total benefits	£2.5m	£4.2m	£8.4m	£12.6m
Midpoint	Scenario 2	Total benefits	£3.5m	£6.0m	£12.0m	£18.0m
Midpoint	Scenario 3	Total benefits	£0.0m	£0.0m	£0.0m	£0.0m
Midpoint	Scenario 4	Total benefits	£8.0m	£13.6m	£27.2m	£40.9m
Upper end	Scenario 1	Benefits from additional users	£2.0m	£3.5m	£6.9m	£10.4m
Upper end	Scenario 2	Benefits from additional users	£3.0m	£5.0m	£10.1m	£15.1m
Upper end	Scenario 3	Benefits from additional users	£ NAm	£NAm	£NAm	£NAm
Upper end	Scenario 4	Benefits from additional users	£NAm	£NAm	£NAm	£NAm

Cost scenario	Scenario	Type	Estimated initial benefits	Estimated benefits (limited adoption)	Estimated benefits (moderate adoption)	Estimated benefits (substantial adoption)
Upper end	Scenario 1	Benefits from additional use	£0.2m	£0.4m	£0.8m	£1.2m
Upper end	Scenario 2	Benefits from additional use	£0.4m	£0.7m	£1.5m	£2.2m
Upper end	Scenario 3	Benefits from additional use	£NAm	£NAm	£NAm	£NAm
Upper end	Scenario 4	Benefits from additional use	£NAm	£NAm	£NAm	£NAm
Upper end	Scenario 1	Total benefits	£2.3m	£3.9m	£7.7m	£11.6m
Upper end	Scenario 2	Total benefits	£3.4m	£5.8m	£11.6m	£17.4m
Upper end	Scenario 3	Total benefits	£0.0m	£0.0m	£0.0m	£0.0m
Upper end	Scenario 4	Total benefits	£0.0m	£0.0m	£0.0m	£0.0m