Long term prediction intervals for residual growth in English hospitals: a national elicitation exercise providing an outside view based on judgements of experts in support of the New Hospital Programme.

02 December 2024

Prepared by:
Mohammed A Mohammed
Principal Consultant, Strategy Unit on behalf of the Strategy Unit's elicitation team

Contents

Executive Summary

Background

The planning process for a new hospital relies on assumptions about future levels of demand. The demand model for New Hospital Programme (NHP) in England decomposes future total annual growth (G) into shifts in population (P), shifts in population age-specific health status (H), efforts to mitigate (M) demand and residual[1] (R) factors on hospital activity. While national assumptions are available for P, H and M, R had a holding position, pending this report.

The residual (R) is crucial and difficult to estimate, as it has historically driven cost and activity more than demographic changes. R includes unpredictable factors like technology, policy and funding.

The NHP demand model initially used data from 2011-2019 to set assumptions for residual growth (variant 1), but this had limitations. A second variant (variant 2) was developed for non-elective admissions, now the primary assumption, with variant 1 used for sensitivity testing. The use of variant 1 and variant 2 were seen as interim positions[2] pending a national elicitation exercise, which was undertaken in October 2024 and reported here.

Objective

We undertook a national exercise to elicit probabilistic assumptions about annual residual growth rates (R).

Method

An in-person 1 day facilitated elicitation workshop was undertaken over two rounds, using an evidence-based protocol alongside relevant data to elicit prediction intervals for R with two teams - team A (5 senior NHS staff) and team B (4 senior domain experts from academia or policy think tanks) for English hospitals over a 20-year horizon. The primary outcomes are the pooled forecasts representing the % annual growth rate in hospital activity across England based on "surprisingly low" (10th percentile – P10) to "surprisingly high" (90th percentile – P90) prediction intervals across four points of delivery – outpatient attendances, elective inpatients, A&E attendances and non-elective admissions.

[1] The term residual (R) has various synonyms such as non-demographic growth, excess growth, rate of care. In this report we adopt the term residual because it is more precise and more likely to be future proof than its synonyms.

[2] See board paper - Setting 'non-demographic growth' assumptions for further details.

Results

Pooled forecasts for residual growth are shown the table below and the accompanying Figure 1 (overleaf).

Point of delivery	Team A (n=5) Pooled P10 to P90	Team B (n=4) Pooled P10 to P90	All (n=9) Pooled P10 to P90
Outpatient attendances	−0.2% to 4.2%	0.3% to 5.3%	0.0% to 4.9%
Elective admissions	0.2% to 2.8%	−0.2% to 2.6%	0.0% to 2.7%
A&E attendances	−0.7% to 3.0%	−0.3% to 4.0%	−0.5% to 3.5%
Non-elective admissions	−0.7% to 1.9%	0.2% to 3.4%	−0.4% to 2.7%

There was considerable overlap between the pooled intervals for team A and team B, although team B was generally more uncertain than team A. Team B also predicted higher levels of growth for outpatient attendances, A&E attendances and non-elective admissions. The highest pooled P90 values were for outpatient attendances (4.9%) followed by A&E attendances (3.5%). The lowest pooled P10 values were for A&E attendances (-0.5%) and non-elective admissions (-0.4%).

Experts provided rationales for all four points of delivery, with varying levels of detail. Some gave brief key factors, while others described future scenarios to explain their P10/90 predictions. Both team A and team B had similar rationale themes. For P10, experts cited low funding and out-of-hospital care limiting future growth. For P90, they focused on inadequate primary care and higher funding driving demand. Some experts noted that growth in one area could affect another, and technology was seen as influencing growth either way. A key difference was that team B experts referenced workshop analyses, while team A did not. Despite some revisions to P10/P90 predictions in round two, few rationales changed.

The workshop revealed two key challenges in residual (R) forecasting: defining its precise scope and understanding cross-service correlations. While this elicitation process makes previously opaque aspects of R more transparent, these methodological issues require systematic attention in future exercises.

Conclusion

A national elicitation exercise has provided long term pooled forecasts across England that make explicit the variation and uncertainty associated with residual (R) growth from an outside perspective. There was considerable overlap between team A and team B and so we recommend using the pooled prediction intervals. These pooled forecasts may now be incorporated into the NHP demand model, providing a more robust foundation for planning.

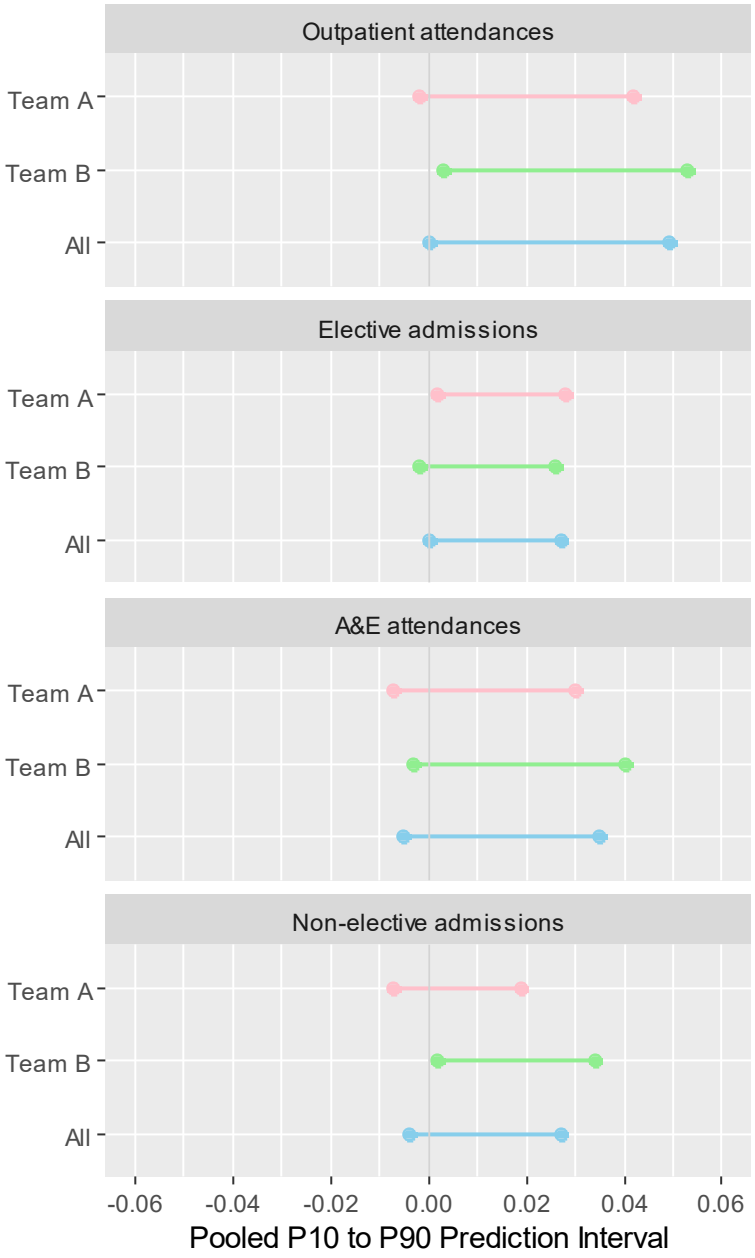


Figure 1 Pooled forecasts for residual growth for each point of delivery.

Introduction

Hospitals are a cornerstone of healthcare systems worldwide. In the English National Health Service (NHS), there are approximately 515 hospitals[1], many of which face structural challenges due to aging infrastructure[2]. This has prompted the government to launch the New Hospital Programme (NHP), which, subject to review, aims to build 40 new hospitals across England. Additionally, seven hospitals compromised by the use of reinforced aerated autoclaved concrete have also been included in the NHP[3].

Building a new hospital is a major investment involving substantial (public) funds. Therefore, accurately forecasting future demand for hospital services is essential. A hospital that is too small risks being overwhelmed and unable to meet the needs of its population, while one that is too large may be underutilised. Either scenario represents poor value for money. Unfortunately, long-term forecasts of hospital activity frequently appear as point estimates, which are prone to pitfalls such as the flaw-of-averages, base-rate neglect. The tendency to underweigh or ignore distributional information is a seen as a major source of error in forecasting.

To support the NHP design process, the Strategy Unit's analytics team has developed a probabilistic demand model, known as the NHP Model. In this model, total annual growth (G) is a function of future population growth (P), shifts in population age-specific health status (H), efforts to mitigate (M) demand and residual (R) factors on hospital activity as per the growth equation[3] below.

G=(1+P)(1+H)(1-M)(1+R) **growth equation (1)**

- G = Total growth in hospital activity
- P = Growth factor due to changes population
- H = Growth factor due to changes in the health status of the population
- M = Growth factor due to mitigation of hospital activity
- R = Residual growth factor

The growth equation considers G across four points of delivery of hospital activity, as shown below:-

Point of delivery (PoD)	Description
Outpatient attendances	Includes outpatient procedures, excludes did not attends (DNAs).
Elective admissions	Includes all admissions and day cases, excludes maternity.
A&E Attendances	Type 1 units only
Non-elective admissions	Includes all admissions including zero length of stay

[3] Note that each term in the equation is assumed to be independent, and that P, H, and R could be +/- but mitigation (M) is theoretically bounded to (0 – no mitigation) to (1 – total mitigation). See Complexity of the residual for further discussion of this limitation.

Assumptions about future population growth (P) are based on ONS population projections[4]. Assumptions about shifts in population-age specific health status (H) are based on a previous unpublished national elicitation exercise. Assumptions about mitigation (M) are based on a previous published elicitation exercise[5]. Assumptions about the residual (R) were set, as an interim position, using base-rates from previous retrospective analyses[6].

The residual (R) is crucial and difficult to estimate, as it has historically driven cost and activity more than demographic changes. R includes unpredictable factors like technology, policy, and funding. The NHP demand model initially used data from 2011-2019 to set assumptions for residual growth (variant 1), but this had limitations. A second variant (variant 2) was developed for non-elective admissions, now the primary assumption, with variant 1 used for sensitivity testing. The use of variant 1 and variant 2 were seen as interim positions pending a national elicitation exercise, which was undertaken in October 2024 and is reported here. The rational for variant 1 and variant 2 are set out in a previous board paper[7].

The term residual (R) has various synonyms such as non-demographic growth, excess growth, rate of care. In this report we adopt the term residual because it is more precise and more likely to be future proof than its synonyms.

In this exercise, we sought to elicit assumptions about the residual (R) from an outside perspective[4-6] for England. P, H and M can, to some extent, be thought of in isolation, but to predict the residual (R) we need to understand what is/is not in the residual term which requires a more detailed and nuanced understanding of the all the terms in the growth equation. This makes predicting R more challenging.

There are several approaches to forecasting future healthcare activity including statistical methods, expert judgments, and scenario building. For long-term forecasts, expert judgment is a widely used approach, which we have adopted here. Our objective was to gather outside perspectives on residual growth by eliciting probabilistic forecasts from human Subject Matter Experts (SMEs). The elicitation process is informed by literature on cognitive biases, project planning, and decision analysis as summarised by Hemming et al[8]. "Expert judgement can be remarkably useful when data are absent

[4] Subnational population projections for England Statistical bulletins https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/bulletins/subnationalpopulationprojectionsforengland/previousReleases

[5] Mohammed MA *et al* Long-term probabilistic forecasts of activity mitigation in English hospitals: a national elicitation exercise providing an outside view based on judgements of experts in support of the New Hospital Programme *BMJ Open* 2024;**14:**e084632. doi: 10.1136/bmjopen-2024-084632

[6] Estimating the impact of non-demographic effects on hospital activity growth in England 2011-2019. https://connect.strategyunitwm.nhs.uk/nhp/project_information/modelling_methodology/ndg_analysis.html

[7] Setting 'non-demographic growth' assumptions paper for further details.

[8] Hemming V et al. A practical guide to structured expert elicitation using the IDEA protocol. Methods in Ecology and Evolution. 2018 Jan;9(1):169-80.

or incomplete. However, experts can also make mistakes. This is often due to a range of cognitive biases such as anchoring, availability, and representativeness, groupthink, overconfidence, and difficulties associated with communicating knowledge in numbers and probabilities. Inappropriate and ill-informed methods for elicitation can amplify these biases. Well designed, structured elicitation protocols can enhance the quality of expert judgments. These protocols treat each step of the elicitation as a process of formal data acquisition, and incorporate research from mathematics, psychology and decision theory to help reduce the influence of biases and to enhance the transparency, accuracy, and defensibility of the resulting judgements."

This report describes an elicitation exercise undertaken to obtain probabilistic forecasts about residual growth of hospital activity across England over a 20-year time horizon. We asked experts to provide probabilistic forecast, which were pooled to represent an outside view of residual growth in the form of low to high probabilistic forecasts with an 80% degree of belief, where low and high equate to the 10th and 90th percentiles of an assumed normal distribution. We denote this as the P10 to P90 prediction interval or P10-P90 interval.

Methods

We based our elicitation protocol on the Stanford Research Institute protocol which has five broad steps: motivate, structure, condition, encode and verify, as described in the Handbook of Decision Analysis[9] and the IDEA (Investigate, Discuss, Estimate, Aggregate) protocol as described by Hemming et al[10] adapted to better suit our needs. The elicitation protocol is designed to mitigate well known cognitive biases that usually lead to overconfidence in the judgments of experts.

Participants:

We invited senior managers from NHS England and experts from academia or policy think-tanks to take part in this study who would have some of the following characteristics.

- Domain knowledge or expertise or interest in forecasting future growth of hospital activity
- Consent to participate on a voluntary basis.
- Agree to follow the elicitation protocol.
- Have an appetite for making probabilistic forecasts.

Participation was voluntary, confidential and required informed consent with the option to withdraw at any time. All participants gave informed consent. The list of participants who consented to take part is provided with the acknowledgement at the end of the report. Participants are referred to as Subject Matter Experts (SMEs). Two facilitators made notes during the workshop and participants were given the option to make their own notes and share them with the facilitation team if they wanted to. We did not record the session. SMEs were invited to complete an evaluation form. SMEs were split into two teams. Team A (n=5) included people who worked in the NHS. Team B (n=5, but we had one drop-out) included academics and policy experts, to enable us to examine the extent to which there were systematic differences between the two teams.

To support SMEs in their task we sent SMEs a pre-workshop briefing email which provided an overview of the aims of the elicitation exercise with background reading. We also designed and facilitated a 1 day in-person probabilistic forecasting workshop which incorporated two rounds of data collection using an on-line app. The workshop ran from 10:00 to 16:00 with comfort breaks on 9 October 2024.

Each SME had a supporting information pack which included an overview of the NHP model, descriptions of the four PoD, key table of results from two of the more relevant papers. The first was a retrospective attempt at decomposing growth (G) into P, H and R by the Strategy Unit by PoD[11].

[9] Parnell GS, Bresnick T, Tani SN, Johnson ER. Handbook of decision analysis. John Wiley & Sons; 2013 Jan 24.
[10] Hemming V et al. A practical guide to structured expert elicitation using the IDEA protocol. Methods in Ecology and Evolution. 2018 Jan;9(1):169-80.
[11] Estimating the impact of non-demographic effects on hospital activity growth in England 2011-2019. https://connect.strategyunitwm.nhs.uk/nhp/project_information/modelling_methodology/ndg_analysis.html

The second was a prospective attempt to decompose growth (G) into P, H and R by the REAL centre The Health Foundation[12]). It was noted that the two pieces of evidence had key differences which are highlighted below :-

Feature	Strategy Unit Evidence about growth	Health Foundation Evidence about growth
Perspective	Retrospective	Prospective
Currency	Unweighted activity	Cost-weighted activity
Health status adjustment basis	Single consolidated measure and independent of other drivers of growth	Multi-dimensional and interacting with other drivers of growth
Historical basis	2013-2019	2010/11 – 2018/19
A&E coverage	Type 1 department only	All department types
Residual (R)	Contaminated with Mitigation (M)	Contaminated with Mitigation (M)

The briefing pack included these two key summary tables from the Strategy Unit report and the Health Foundation REAL centre report as shown below.

	Strategy Unit Evidence about growth				Health Foundation Evidence about growth			
POD	**G**	**P**	**H**	**R**	**G**	**P**	**H**	**R**
Outpatient attendance	4.51%	1.07%	-0.02%	3.41%	3.80%	0.90%	0.70%	2.20%
Elective admission	2.29%	1.58%	-0.03%	0.73%	1.80%	0.90%	0.20%	0.70%
A&E attendance	1.29%	0.36%	0.09%	0.84%	3.90%	0.90%	0.20%	2.80%
Non-elective admission	2.69%	0.84%	0.18%	1.65%	2.80%	1.50%	0.50%	0.80%

It was made clear that the growth equation in both the above cases did not explicitly account for mitigation (M) and so the residual (R) is contaminated with M.

[12] Rocks S et al How much funding does the NHS need over the next decade? https://www.health.org.uk/publications/long-reads/how-much-funding-does-the-nhs-need-over-the-next-decade

Nevertheless, these pieces of evidence were our best available attempts at decomposing growth (G). During the workshop Steven Wyatt and Toby Watts were in the room as expert advisors with respect to these two reports and considerable time was used at the start of the workshop to help clarify the challenges and issues associated with understanding and predicting the residual growth (R).

The time horizon for all predictions was 20 years from the baseline year of 2022. It was made clear that SMEs were required to provide annual growth rates (%) not total period growth (which was available to the SMEs as a lookup table in their supporting information pack). SMEs were required to use a bespoke app with a slider and comment boxes (to explain their rationales), to provide P10 and P90 probability intervals forecasts with an 80% degree of belief. The concept of the P10 and P90 interval was explained by using the qualitative terms "surprisingly low" and "surprisingly high" to denote the P10 and P90 respectively. In round one, SMEs were given 15 mins per prediction in silence. In round two, SMEs were presented with their prediction intervals on a dumbbell graph alongside the data from all other SMEs. In round two, SMEs had 30 minutes per question with open discussion and then an opportunity to revise their round 1 responses. In round two, participants could see the rationales from other SMEs. After round two closed, the results were pooled.

Data processing and analysis: All SMEs were deidentified and assigned a unique code which could not be linked back to the SME. We assumed that the P10 and P90 values supplied by each SME described a Normal distribution, from which we derived the mean [P10+P90)/2] and standard distribution [(P90 - mean) / qnorm (0.9, mu=0, sd=1)]. We pooled these SME predictions by creating mixture distributions using the distr package[13], weighting each SME view equally. We present the probability density of these mixture distributions, and extract P10, P90 and mean values, for each team and in total for each point of delivery. The GitHub repository containing the R code is here: https://github.com/The-Strategy-Unit/NDG_prediction_aggregation

[13] https://cran.r-project.org/web/packages/distr/index.html

Results

Overall results

Of the 10 SMEs, one was unable to attend the workshop due to unforeseen circumstances. There were no missing data from round one to round two. Our results are based on round two data. The table below shows the overall pooled prediction intervals for each point of delivery along with a graphical overview.

Point of delivery	Team A (n=5) Pooled P10 to P90	Team B (n=4) Pooled P10 to P90	All (n=9) Pooled P10 to P90
Outpatient attendances	−0.2% to 4.2%	0.3% to 5.3%	0.0% to 4.9%
Elective admissions	0.2% to 2.8%	−0.2% to 2.6%	0.0% to 2.7%
A&E attendances	−0.7% to 3.0%	−0.3% to 4.0%	−0.5% to 3.5%
Non-elective admissions	−0.7% to 1.9%	0.2% to 3.4%	−0.4% to 2.7%

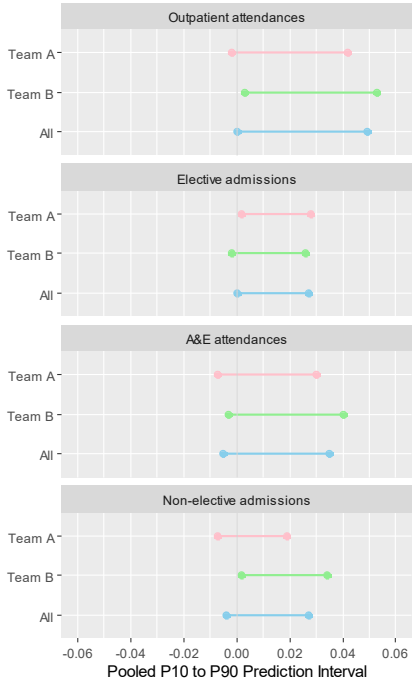


Figure 2 Pooled forecasts for residual growth for team A, team B and pooled.

There was considerable overlap between the pooled intervals for team A and team B, although team B was generally more uncertain than team A. Team B predicted higher levels of growth for outpatient attendances, A&E attendances, and non-elective admissions. The highest pooled P90 values were for outpatient attendances (4.9%) followed by A&E attendances (3.5%). The lowest pooled P10 values were for A&E attendances (-0.5%) and non-elective admissions (-0.4%).

Figure 3 (overleaf) is an overview of individual SME prediction intervals for the four points of delivery pooled by team A, team B and all.

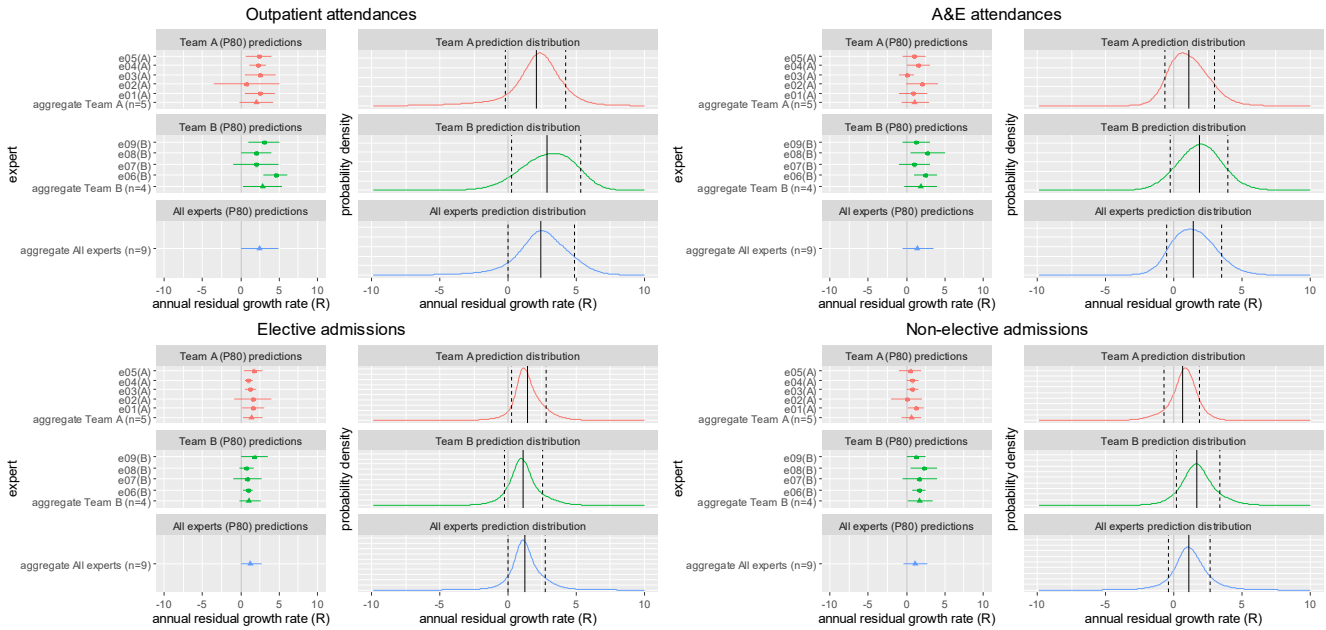


Figure 3 Showing round two individual and pooled prediction intervals by team (team A, team B and all) across each point of delivery. Left panels shows line plots from each expert and right panels are pooled density plots with P10 and P90 indicated by dashed vertical lines and mean indicated by a solid vertical line.

There were no apparent outliers in round two results for individual SMEs. However, in team A, e02(A) had the widest prediction intervals indicating greater uncertainty. In team B, e07(B) had the widest intervals. In general, the means of the pooled distributions for team B were more likely to be higher than those of team A. Distributions for elective and non-elective admissions were more peaked than other distributions reflecting lower variance across the individual SMEs.

In the next section we report the detailed round two results for each point of delivery in table 1 to table 4.

Detailed results for each point of delivery

Individual results for experts for each point of delivery.

Note that, given the time constraints of the live workshop and the need for experts to provide rationales within short time intervals, the verbatim free-text comments are not fully refined or "camera-ready." As a result, such comments may appear to lack the usual level of precision as there was no opportunity during the workshop for SMEs to craft their responses.

Outpatient attendances

The table overleaf shows the raw results for each SME for outpatient attendances.

P10 rationales

A key theme present in the rationales of three team A and three team B experts was the rationing of care due to low levels of future funding. Some experts made explicit that this would be a result of limited economic growth and/or high input prices. Two Team B experts described an increase in patients seeking private care due to a lack of access to this point of delivery.

The only experts (one team A and one team B) to predict negative average annual growth rates in their P10 forecasts included within their rationales the role of technology in reducing demand for outpatient care.

While one team A expert reasoned that there could be “successful movement of some follow-up activity to primary care”, an expert from team B stated that “limited access to primary care leads to lower referrals”.

P90 rationales

Most experts described a significant shift in point of delivery with activity moving from inpatient to outpatient care. This was a common theme across all team B rationales and 3 team A rationales. Some experts suggested that technology would be a key factor in the residual growth of outpatient activity they predicted. Other factors included a “growth in patient expectation and willingness of clinicians to refer/treat” (team A expert) and similarly, a team B expert stated that “private funding of diagnostics leads more people to come back into NHS for advice and treatment.” Another team B expert included within their rationale a description of a scenario in which high levels of funding results in “greater demand/supply.”

Table 1 Round two prediction intervals experts in each team (A) or (B) for outpatient attendance.

Expert	P10	P90	P10 comments (verbatim)	P90 comments (verbatim)
e01(A)	0.5	4.5	There is less opportunity to move activity from daycases as in the past. Lower growth in diagnostics. Successful movement of some followup activity to primary care	Greater push for proactive elective care. More movement from DC to OP
e02(A)	-3.49	4.99	Tech changes reduce need for OP e.g. asynchronous	Tech change permits many more elderly patients to be dealt with on a OP basis
e03(A)	0.51	4.51	Rationing of healthcare continues to restrict provision.	Continued left shift and new technologies increase demand. Technology allows more patients to be treated with fewer contacts
e04(A)	1.13	3.33	Low funding growth for the NHS as a result of low per capita economic growth	Significant shift in activity from inpatient PODs to Outpatients
e05(A)	0.68	4.02	I have assumed that funding for outpatients decreases (one pay and other price increases factored in)	Growth in patient expectation and willingness of clinicians to refer/treat
e06(B)	2.98	6.05	I chose P10 to be close to the point estimate of the Strategy Unit study. The rationale is that I would expect a large proportion of growth in OP attendances explained by shifting of activity from IP to OP settings due to technological advances. I expect that this shifting will be remarkably higher than what can be captured by historical data.	Similar rationale as for the P10 prompted me to provide a P90 that is significantly higher than the Strategic Unit and HF point estimates.
e07(B)	-1.01	4.96	Major rationing due to low GDP growth, tight funding/high input prices. This would still have outpatient appts growing (due to population growth etc) but conditional on age/health, there would be about 20% less. May also be technological	Large shifts to outpatient care relative to inpatient care (told this doesn't go into M). High GDP and funding growth, mean greater demand/supply. Estimate was 3.4% for last decade, very plausible it could be substantially higher if higher funding growth. Technological developments

Expert	P10	P90	P10 comments (verbatim)	P90 comments (verbatim)
			developments that reduce demand.	increase demand for outpatients - e.g. lower costs from online
e08(B)	0	4	shift to private care or emergency care because of undersupply	New technologies are actually shifting care away from emergency care. Obesity keeps increasing in the population.
e09(B)	1	5	aging population with comorbidity and frailty makes it harder to shift care to outpatient settings. National push to reduce follow-ups. Rationing care and limited access to primary care leads to lower referrals. More people go private due to lack of access and concerns about quality.	technology and innovation likely to lead to lots more treatments being managed as inpatients. Investment in diagnostics leads to increase in referrals. New drugs require specialist prescribing. Private funding of diagnostics leads more people to come back into NHS for advice and treatment.

Elective admissions

The table overleaf shows the raw results for each SME for elective admissions.

P10 rationales

Some of the experts who had described a shift in activity from inpatient to outpatient care in their response to the previous question, reiterated this view in their P10 rationales for elective admissions. The rationing or 'undersupply' of care in this point of delivery was also referenced by multiple experts across both Teams. While some experts suggested this could be a result of low economic growth constraining future funding levels, other experts included different potential drivers: one team A expert outlined a future in which the "Government no longer seeks to prioritise waiting list" while another stated that "resources continue to be soaked up by emergency care". The impact of the "left shift" or "treatment in community settings" was also cited by experts (one team A and one team B).

P90 rationales

Multiple experts from both teams focussed on technology and increased productivity as key drivers of future residual growth in this type of activity. Some experts also suggested that patient expectations would be a contributing factor. Two experts from team A and one expert from team B stated that the increase in activity would be caused, at least in part, by a shift from emergency care to elective care. One team B expert described a scenario in which amongst other factors there would

be “high political focus on elective waiting times” and “high funding”, but also noted that “elective activity seems more responsive to income than others”.

Table 2 Round two prediction intervals experts in each team (A) or (B) for elective admissions.

Expert	P10	P90	P10 comments (verbatim)	P90 comments (verbatim)
e01(A)	0.2	3	Significant movement to OP for some daycases. Resources continue to be soaked up by emergency care	Reinvigoration of elective care rather than emergency.
e02(A)	-0.9	4	I would expect EL to follow P&H&M and I'd be surprised if it were much lower than that	Plausibly we are able to predict better and much more care becomes amenable to DC
e03(A)	0.5	1.99	Rationing of healthcare coupled with left shift may be expected to result in a low rate of residual growth of electives	Technology and patient expectations my drive non-demographic pressures, but these will be limited by spending constraints. Revised upward to recognised shift from non-electives.
e04(A)	0.52	1.5	Government no longer seeks to prioritise waiting list and therefore we see a shift of activity to the independent sector	Increase in the adoption of new technology
e05(A)	0.39	2.84	Funding constraint - essentially funding stays close to GDP growth (say 1-1.5%) and this all used by D, H and price increases, with little productivity improvement	Relatively unconstrained growth made possible by significant increases in productivity e.g. future acceleration of day/outpatient cases, robotic surgery costs fall significantly, skill mix
e06(B)	0.31	1.6	P10 based on the point estimates of the two studies	Ditto for P90
e07(B)	-1	2.72	Major rationing of elective care, perhaps due to low GDP and funding growth/high input prices. Technological improvements reduces the need for inpatient treatment. Shift to private demand for treatment.	High political focus on elective waiting times, driving large increases in activity. High funding/GDP growth. Means you get growth much faster than last decade. Greater demand because of improvements in performance. Elective activity seems more responsive to income than others for me.
e08(B)	-0.2	1.7	Inadequate supply shifts population care to private and emergency care.	Increased efficiency that shifts emergency care towards ellective

Expert	P10	P90	P10 comments (verbatim)	P90 comments (verbatim)
e09(B)	0	3.48	low rate is above previous decade as that was a decade in which standards fell substantially and waiting lists rose. shift to outpatient procedures and more medical interventions in place of surgery and treatment in community settings even for comparatively complex conditions.	Baby boomer generation reaching early old age which significant expectations for healthy aging. Economic need to ensure people between 50 and 68 are able to work so need more access to deal with lower level work inhibiting health problems. growing recognising of inequality and unmet need. More people having access through tech and cheap private investigations to earlier diagnosis of conditions and expecting / seeking interventions. technology making it safer to do procedures on people where previously would have been concerns.

A&E attendances

The table overleaf shows the raw results for each SME for A&E attendances.

P10 rationales

Although not mutually exclusive, here the rationales largely fell into two categories: those which describe the provision of care delivered outside of hospital (or emergency settings) being sufficient in the future and therefore resulting in low or negative growth rates in ED attendances (three team A and two team B) and those which suggest that the key factor constraining growth is the rationing of care (two team A and one team B).

P90 rationales

A common theme was the delivery of care outside of hospitals. Three team A and three team B experts depicted future scenarios in which the provision of primary, community and/or social care would be inadequate to meet the needs of the population leading to higher growth in ED attendances. In addition to "inadequate primary and planned care", the expert predicting the highest level of growth also described "increasing levels of deprivation" in their rationale. Some experts also referenced patient expectations as a driver of activity growth: one team A expert described "rising public expectations of rapid treatment" and one team B expert noted the "risk aversion of families".

The only reference to the role of technology was from a team B expert who explained that there would be "limited scope for technology to reduce need."

Table 3 Round two prediction intervals experts in each team (A) or (B) for A&E attendances.

Expert	P10	P90	P10 comments (verbatim)	P90 comments (verbatim)
e01(A)	-1	2.7	Large investment in primary and community care moves care outside hospitals and to type 3 depts	Failure to invest in primary care. Realisation that some care is more cost effectively provided in hospitals.
e02(A)	-0.03	4.11	Capacity constraints push more UEC first-contact to something like the app or alternative	Could abolish type 3 A&E!
e03(A)	-1	1	Improved provision of out of hospital care and increased use of direct admission pathways may be expected to divert patients away from A&E settings	Deterioration in out of hospital care may direct more care towards A&E settings. I might consider increasing this estimate, but it depends to a degree on on pessimistic I am about the future.
e04(A)	0	3	Policy shifts, patients seek care or are directed to care that are alternatives to A&E departments.	Primary care model becomes unsustainable, and thus A&E picks up the demand.
e05(A)	-0.53	2.5	Thresholds for A&E raised (NB: not sure if this is a mitigator or not)	Rising public expectations of rapid treatment, growth in clinical practice (e.g. more primary care and type 3 send for advanced diagnostics)
e06(B)	1.01	4.01	P10 more in line with the HF study that predicts higher overall growth	Ditto for P90
e07(B)	-1.01	3	Major rationing because of low funding growth/GDP/high input prices. Poor performance reduces demand. Greater use of technology means much more care can be provided in other settings.	High GDP/funding growth. But seems less responsive to that than elective activity. Limited scope for technology to reduce need. Problems in social care.
e08(B)	0.5	5	There is a mild increase in material deprivation and the supply of primary and planned care almost covers the demand	Increasing levels of deprivation and inadequate primary and planned care supply push the population towards ED

Expert	P10	P90	P10 comments (verbatim)	P90 comments (verbatim)
e09(B)	-0.51	3.03	tech and innovation allowing many more options to support people at home, Ambulance services have many more options for 'do not convey'. Integrated neighbourhood teams in place and delivering full model of proactive support. Dementia drugs being used, slowed progression and fewer people with advanced dementia in the community.	Increasing numbers of frail elderly with complex comorbidities living at home without support networks. Risk aversion of families and out of hospital services. mental health crisis increasing due to lack of access to services.

Non-elective admissions

The table overleaf shows the raw results for each SME for outpatient attendances.

P10 rationales

Three team A experts made explicit in their rationales that growth in non-elective admissions would be constrained by future levels of capacity. Similarly, one team B expert stated that "tight funding growth" would result in "more rationing". The P10 values of these 4 experts were negative or zero with none of the other experts predicting lower growth rates. Some of the other rationales cited the role of primary, community or social care (one team A, two team B) in constraining the growth of non-elective admissions. Technology was also referred to with one team A and one team B expert suggesting that it could have a downward pressure on growth while another team A expert described a future where "technological advances net off against one another (e.g. new treatments that prevent admissions balance those [which] generate admissions)."

P90 rationales

Here three experts from team A made specific reference to future growth in either same day emergency care admissions (SDEC) or zero-day admissions which was not described in the rationales of team B experts. Some of the drivers of residual growth described by team A experts were a failure to move care out of hospital, risk aversion by clinicians, increased appetite for diagnostics and the advance of technology resulting in treatments which can only be delivered in hospital. Two experts from team A also suggested that "an expansion SDEC" or "real terms increase in funding" would increase supply and therefore demand in non-elective admissions.

Two team B rationales could be interpreted as descriptions of failure demand fuelling non-elective admissions: one expert commented on the inadequacy of the social care budget and the supply of primary and planned care while the other detailed "wider cuts to public services" and a lack of progress in "end of life care" resulting in high rates of admissions for "older people and the vulnerable". Another team B expert described "high funding growth" and "limited technological improvements."

Table 4 Round two prediction intervals experts in each team (A) or (B) for unplanned admissions.

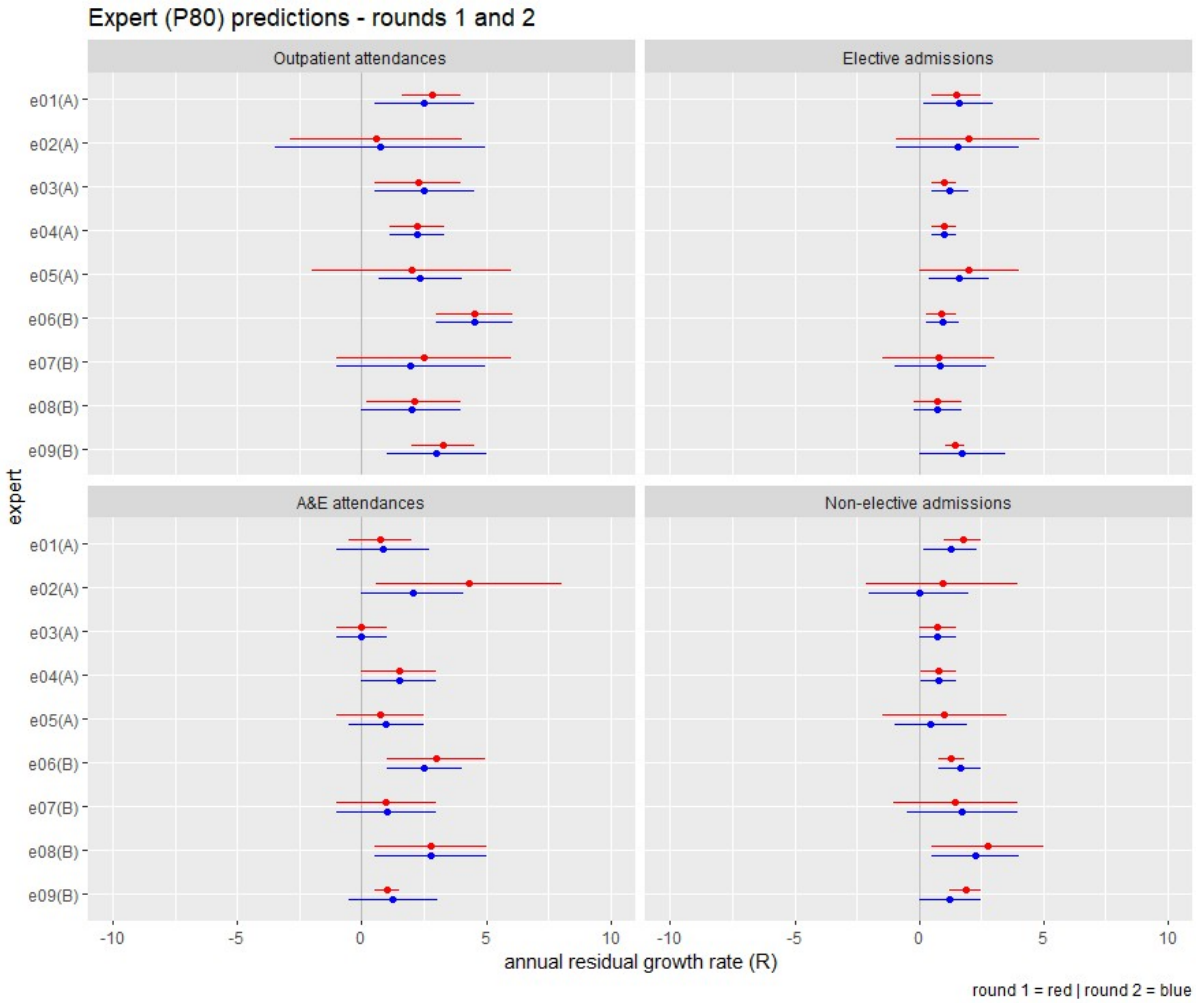
Expert	P10	P90	P10 comments (verbatim)	P90 comments (verbatim)
e01(A)	0.2	2.3	Increase in prevention and primary care reduce avoidable admissions and SDEC falls as patient able to see GP.	NHS struggles to move care out of hospitals. Very high SDEC growth continues
e02(A)	-1.99	2	Capacity constraints lead to NEL basically being held to low growth (assuming what goes on with P&H&M)	We could fail to shift care to preventative or community basis and continue strong growth
e03(A)	0	1.51	Net use of non-elective care is constrained by provision and growth in line with need. Technological advances net off against one another (e.g. new treatments that prevent admissions balance those generate admissions).	Risk aversion by clinicians increases demand for NEL care. Continued focussed on expansion of SDEC provision leads to supply-induced demand for short-stay admissions
e04(A)	0.08	1.52	Technology shift allows more patients to treated in SDEC and thus avoid admissions	Increased appetite for diagnostics (e.g. more diagnostics per patient) which adds to time and thus requires more admissions
e05(A)	-1	1.94	Bed base stays flat and so 1.5% reduction needed in R order to compensate for 1.5% growth from D+H	Enabled by growth in zero day admissions, stimulated by technological frontier on possible treatments which need to be delivered in hospital and some real terms increase in funding
e06(B)	0.8	2.47	P10 roughly the point estimate of the HF study	P90 roughly the point estimate of the Strategic Unit study
e07(B)	-0.5	3.95	Tight funding growth/GDP/high input prices means severity	High funding growth/GDP. Limited technological

Expert	P10	P90	P10 comments (verbatim)	P90 comments (verbatim)
			thresholds are higher, more rationing	improvements. Historic estimate higher than elective/A&E.
e08(B)	0.5	4	Only mild increase in deprivation. The supply side of primary and planned care is almost adequate to cover demand.	Increasing levels of deprivation and inadequate primary and planned care supply push the population towards ED. The social care budget is inadequate, and there are many delayed discharges.
e09(B)	0	2.49	Technology means more patients don't need to be admitted, more effective community based options, good social care with integrated neighbourhood teams managing many more conditions at home. Home based end of life care models in place, most people have advanced care plans. Still above past rate as services quality and standard declined during reference period.	Continued low economic growth and cuts to wider public services leave older people and the vulnerable without support and networks to manage them in the community, admission rates are high and LOS longer. Technology means more people surviving but end of life period prolonged with frequent admissions are others services (community, primary care etc) can't cope and don't have capacity. No significant progress on end of life planning.

Comparison of round one versus round two results

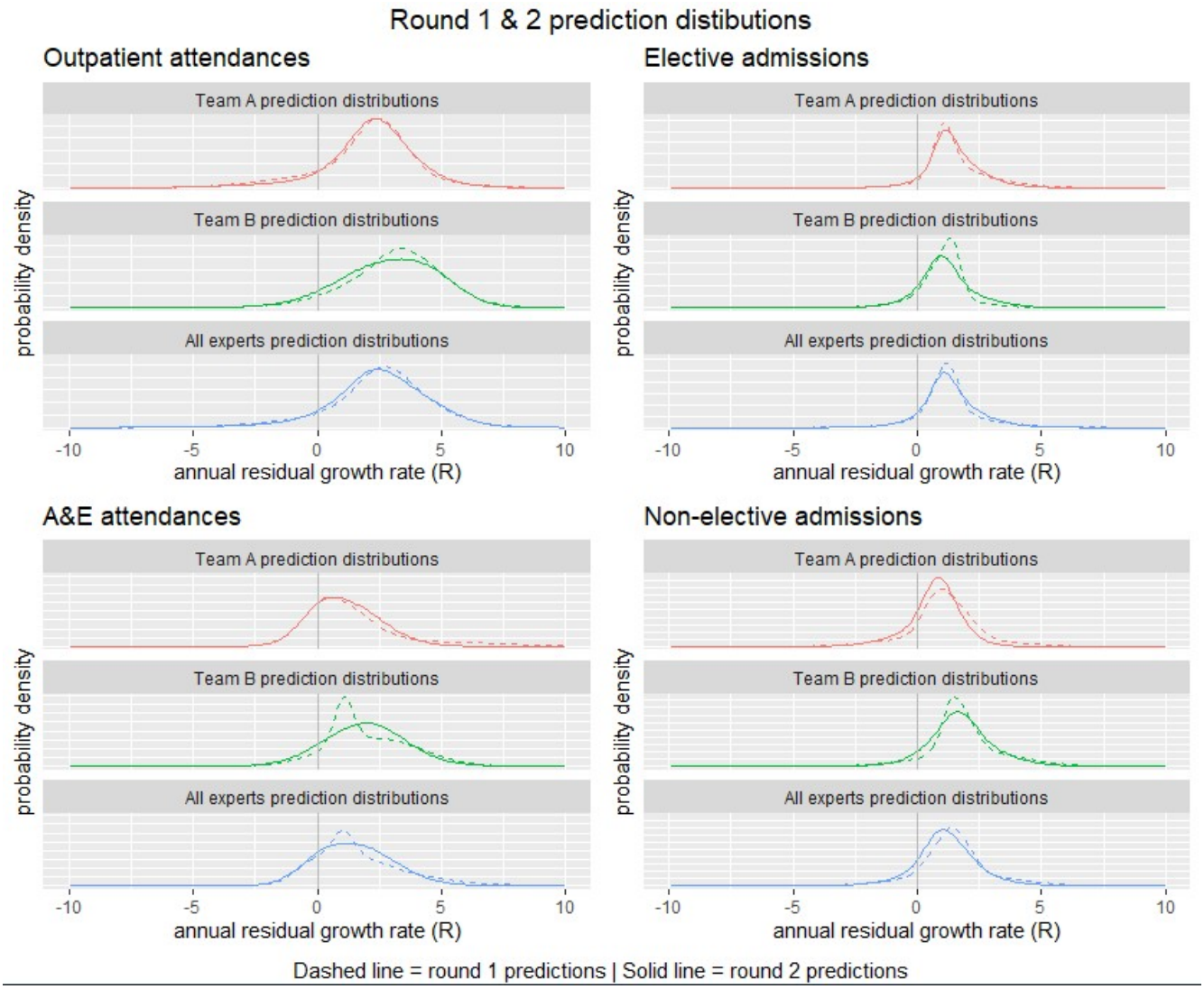
In this section we explore the extent to which individual SMEs shifted their responses from round one to round two. As shown in the figure below, in general, there were minor shifts in responses from round one to round two across most SMEs with some exceptions. SME e02(A) and e09(B) increased the length of their prediction intervals from round one to round two for outpatient attendances whilst decreasing their prediction intervals for elective admissions, A&E attendances and non-elective admissions. SME e05(A) decreased the prediction intervals from round one and round two.

Figure 4 Comparison of round 1 vs round 2 results for each expert by points of delivery



The impact of these shifts from round one to round two are shown in the figure overleaf. The greatest impacts are seen in respect of the distributions of team B which are more symmetrical in round two.

Figure 5 Pooled probability density functions for round 1 and round 2 by team (A vs B) across points of delivery



Complexity of the residual

Expert discussions highlighted the intricate relationships between the residual (R) and other growth components in the growth equation. The elicitation process has helped reveal these complex interactions - elements that were previously hidden within the residual term, which has traditionally functioned as an opaque 'black box' in growth modelling. There were two instances where there were clear indications of contamination with other terms in the growth model (see below).

Point of Delivery	Rationale	Contamination
OP attendances	e08(B) states in their P90 rationale "*Obesity keeps increasing in the population*."	This indicates a possible contamination of R with H - here the expert is predicting a change to health status which is driving the increase in activity.
Non-elective admissions	e01(A) states in their P10 rationale, "*Increase in prevention and primary care reduce avoidable admissions.*"	This indicates a possible contamination of R with M – "prevention" being one of the mechanisms of activity mitigation which can be factored into locally by schemes.

During the discussions, experts emphasised two key challenges in assessing the residual (R) component: First, they highlighted the importance of understanding how R values correlate across different points of service delivery under various future scenarios. Second, they noted ambiguity in the scope of the residual (R) with respect to mitigation (M). For instance, new surgical technology may no longer require hospital inpatient stay so it's not clear if this is captured under mitigation (M) of inpatient admissions or appears under the residual (R) as technological change driving down inpatient admissions. Moreover, innovations like virtual wards may already be captured in local assessments of M making it difficult to establish consistent national estimates of R.

These methodological challenges underscore the need for more clarity and improved correlation analysis in future elicitation exercises.

Workshop evaluation

Below are the evaluations of the elicitation workshop from each SME and a summary of their free text comments. The responses were anonymous.

Content (Depth & Clarity)	Structure & Facilitation	Pace of the session	Overall
Good	Excellent	Excellent	Excellent
Fair	Good	Good	Good
Excellent	Excellent	Excellent	Excellent
Good	Excellent	Excellent	Excellent
Excellent	Good	Good	Excellent
Excellent	Excellent	Fair	Excellent
Good	Excellent	Excellent	Excellent
Good	Good	Good	Good
Fair	Good	Good	Good

1. Overall Positive Experience: Participants felt the discussion was well-controlled and productive, with an appreciation for the facilitator's efforts.
2. Desire for Technical Detail: There is a request for more technical details in advance of discussions, particularly regarding the SU model and the relationship between R and M.
3. Suggestions for Improvement: Specific suggestions include displaying overall growth rates on a sliding scale for better calibration and breaking down broader activity groups (e.g., separating OP and SDEC from NEL) to provide clearer insights.
4. Facilitator Praise: The facilitation was praised, though participants felt more team involvement in the discussion would enhance the process.
5. Complexity of Formula Components: Participants noted the difficulty in separating components of the formula while considering their combined impact, ultimately influencing hospital size and cost.
6. Diverse Perspectives: There is recognition of the value of diverse viewpoints in enriching the discussion, though it raises concerns about the comparability of estimates.

Discussion

We conducted an in-person exercise to gather long-term probabilistic forecasts from experts on potential growth in various types of hospital activity driven by residual factors. Experts recognized the complexity and nuances in predicting the residual (R) and provided valuable recommendations to enhance future elicitation efforts. The workshop was well-received and has now equipped the NHP with a consolidated set of probabilistic forecasts that clearly outline the distribution of uncertainty related to residual (R) growth in future hospital activity.

Experts provided rationales for all four points of delivery, with varying levels of detail. Some gave brief key factors, while others described future scenarios to explain their P10/90 predictions. Both team A and team B had similar rationale themes. For P10, experts cited low funding and out-of-hospital care limiting future growth. For P90, they focused on inadequate primary care and higher funding driving demand. Some experts noted that growth in one area could affect another, and technology was seen as influencing growth either way. A key difference was that Team B experts referenced workshop analyses, while Team A did not. Despite some revisions to P10/P90 predictions in round two, few rationales changed. However, we did find that a few of the rationales appeared to show some evidence of contaminating R, which is to some extent unsurprising given the complexity of decomposing R. The growth equation that underpins the NHP demand model assumes that each term (P,H,M and R) are independent. This is a necessary simplifying assumption and SMEs pointed out that these factors interact with each other. Moreover, some SMEs indicated that their predictions distributions should be assumed to covary over points of delivery. They suggested that a high value for outpatient growth might be associated with a low growth in elective admissions and vice versa.

In this elicitation exercise we had inside (NHS, team A) experts and outside (non-NHS, team B) experts in the same workshop. The results from these two teams were broadly comparable and since there were no apparent outliers in either round, it appears that the methodology is robust to the selection of individual experts. Nevertheless, we acknowledge that our recruitment criteria are largely pragmatic and the extent which different experts would produce materially different forecasts remains an open question. The evidence does suggest the best approach to selecting experts for forecasting exercises is whether the participant can understand the questions being asked and that multiple experts almost always perform as well as, or better than, the best regarded expert(s).

In a review paper of 25 models to forecast healthcare expenditure[14] (which also considered healthcare activity), the authors identified several key drivers of healthcare activity which serve to highlight the complex mechanisms that the residual (R) term is composed off. These included

[14] Astolfi R, Lorenzoni L, Oderkirk J. Informing policy makers about future health spending: a comparative analysis of forecasting methods in OECD countries. Health Policy. 2012 Sep 1;107(1):1-0.

demographic factors and health status, funding for healthcare, consumers' behaviour, treatment practices (intensity of care), technological progress (medical advances), health productivity (efficiency) and health care system organisation. Across a spectrum of modelling approaches, a consistent general finding emerged: the residual (R) is a complex factor that plays a key role in shaping healthcare activity beyond demographic changes alone.

To optimise engagement, we designed an in-person facilitated workshop supported by multiple learning modalities: synthesised evidence, structured discussions, and interactive data science tools that combined quantitative data, textual explanations and pooled results. This multi-faceted approach helped participants navigate the complexity of the elicitation exercise. Nevertheless, the elicitation process has helped reveal the complexities involved in eliciting the residual (R) that were previously opaque. Moreover, participant feedback highlighted several opportunities for enhancement in future exercises, particularly:

- Clearer explanation of the growth equation and the interaction between its components
- The correlation of the residual (R) across the different points of delivery.
- Consideration of scenario-based elicitation.

These suggested refinements would strengthen future elicitation methodologies and improve participants' ability to make more informed assessments.

The NHP demand model initially used data from 2011-2019 to set assumptions for residual growth (variant 1), but this had limitations. A second variant (variant 2) was developed for non-elective admissions with variant 1 used for sensitivity testing. The use of variant 1 and variant 2 were seen as interim positions pending this national elicitation exercise. The process for elicitation adopted here represents a formal data collection exercise which is systematic, transparent, and subject to scrutiny and continual improvement. This is in marked contrast with the not uncommon "black box" approach to assumption setting which has attracted criticism.

Conclusions & recommendations

A national elicitation exercise has provided long term pooled forecasts across England that make explicit the variation and uncertainty associated with residual (R) growth from an outside perspective. There was considerable overlap between team A and team B and so we recommend using the pooled prediction intervals. These forecasts may now be incorporated into the NHP demand model, providing a more robust foundation for planning and an opportunity for learning how probabilistic data enhances the quality of decision making for NHP.

We recommend that outside view elicitation exercises become a standard component of large-scale NHS project planning. This approach would necessary involve ongoing methodological refinement to address the limitations of this and previous exercises.

Acknowledgements

Participants

The authors would like to express their gratitude and thanks to everyone who supported this project from the New Hospital Programme, especially the participants (see below) who took part in the elicitation workshop on the 9 Oct 2024.

Name	Job Title
Rob Shaw	Assistant Director of Analytics, NHS England
Ed Kendall	Director, NHS England
Andrew Jackson	Director of Operational Insights, NHS England
Ben Jupp	Director of Strategy, NHS England
Thomas Mason	Assistant Director of Analytics, Activity, Capacity & Planning, NHS England
Anita Charlesworth	Director of Research and REAL Centre, The Health Foundation
Chris Kypridemos	Senior Lecturer, University of Liverpool
Max Warner	Research Economist, Institute for Fiscal Studies
Panos Kasteridis	Senior Policy Analyst, Centre for Health Economics, University of York

We also thank Steven Rocks, Toby Watt and Juney De Aguilar (at the REAL centre in the Health Foundation) for supporting the workshop. We are grateful Anita Charlesworth, Director of the REAL centre for hosting the workshop on 9 October 2022.

Funding

The project was funded by NHS England as part of the New Hospitals Programme. The funder approved the study protocol and assisted with recruitment of participants.

The Strategy Units' Elicitation Project team

Mohammed A Mohammed, Emeritus Professor & Principal Consultant m.mohammed2@nhs.net
Josh Cook, Consultant josh.cook@nhs.net
Jennifer Wood, Managing Consultant jennifer.wood43@nhs.net
Steven Wyatt, Head of Research and Policy swyatt@nhs.net
Rihan Davies, Senior Data Scientist rhian.davies25@nhs.net
Peter Spilsbury, Director peter.spilsbury1@nhs.net

The Strategy Unit
Email: strategy.unit@nhs.net
Web: www.strategyunitwm.nhs.uk