National Audit Office

REPORT

NHS England's modelling for the Long Term Workforce Plan

NHS England

SESSION 2023-24
22 MARCH 2024
HC 636

Contents

This report can be found on the National Audit Office website at www.nao.org.uk

If you need a version of this report in an alternative format for accessibility reasons, or any of the figures in a different format, contact the NAO at enquiries@nao.org.uk

The National Audit Office study team consisted of:

Laura Cole, John Fellows, Hannah Kreczak, Natalie Low, Sumayyah Mian, Shivam Sood and Paul Wright-Anderson, under the direction of Ruth Kelly and Tim Phillips.

For further information about the National Audit Office please contact:

National Audit Office
Press Office
157–197 Buckingham Palace Road
Victoria
London
SW1W 9SP

☎ 020 7798 7400

www.nao.org.uk

@NAOorguk

Summary

1 In June 2023, NHS England (NHSE) published its Long Term Workforce Plan (LTWP).[1] Based on an extensive modelling exercise, the LTWP estimated a starting shortfall between workforce supply and demand of approximately 150,000 full-time equivalent (FTE) NHS workers. The LTWP projected NHSE's estimate of the NHS's health workforce needs and identified ways to meet them over the next 15 years, from a supply of 1.4 million FTE workers in 2021-22 to between 2.3 million and 2.4 million FTE workers in 2036-37, an increase of 65% to 72%. NHSE has committed to continue to develop its modelling and the LTWP, publishing a refreshed projection every two years, or aligned with fiscal events as appropriate.

Scope of this report

2 After a request from HM Treasury (HMT), the Department of Health & Social Care (DHSC) and NHSE, the Comptroller & Auditor General agreed that the National Audit Office (NAO) would carry out an independent assessment of the modelling underpinning the LTWP. Our scope was as follows:

- to consider whether NHSE constructed its models effectively and whether they operated correctly in a technical sense to generate the projections and other outputs required of them; and
- to consider whether NHSE's approach to workforce modelling and the models themselves are a reasonable basis for regular strategic workforce planning.

3 When NHSE published the LTWP, the government also announced additional cumulative funding of £2.4 billion up to 2028-29. It said this would be used to pay for initial increases in domestic training places, in line with the overall LTWP. Our review is only of the modelling underpinning the LTWP. We have not assessed the value for money of the funding decisions that accompanied the LTWP.

4 We have taken a structured approach to model review, based on the NAO's *Framework to review models.*[2] In doing this we have had full access to NHSE's modelling and have augmented our own review with interviews with the modellers themselves.

[1] NHS England, *NHS Long Term Workforce Plan*, June 2023.

[2] National Audit Office, *Framework to review models,* January 2022.

5 We considered NHSE's overall methodology and process for the LTWP modelling. The LTWP modelling took the form of a modelling pipeline, which is a structured sequence of steps involving a series of distinct models. We reviewed NHSE's pipeline to determine if the modelling was logical, accurate and appropriate, and had been constructed with appropriate controls. We considered the quality assurance processes used for a subset of the input data and examined whether the input data were processed correctly. We conducted a detailed code review to check for errors within the central model and attempted to validate numbers published in the LTWP. We examined whether the modelling was replicable and if the documentation surrounding it would allow for regular updating. Finally, we assessed the underlying assumptions. Our methods and evidence base are described in further detail in Appendix One.

6 The report is organised in three parts, which cover:

- an introduction to health workforce modelling (Part One);
- NHSE's workforce modelling pipeline (Part Two); and
- NHSE's key modelling assumptions (Part Three).

The rest of this summary contains our high-level findings and conclusion, and a list of our summary recommendations.

High-level findings

Description of NHSE's modelling

7 **In January 2022, the government asked NHSE to produce a long-term workforce strategy that would set out a range of future demand and supply scenarios.** After being commissioned, NHSE agreed the terms of reference for the modelling project in March 2022. From this point it aimed to do all the modelling within seven months, including iterating policy options, but the project ultimately continued for over 15 months. Although the modellers had substantially completed their work after a year, discussions between NHSE, DHSC and HMT, including refining the modelling of policy options, continued until NHSE published the LTWP in June 2023 (paragraphs 1.12, 2.4 and 2.6 and Figure 4).

8 **NHSE's workforce modelling was complex and took the form of a pipeline.** NHSE used inputs from existing workforce supply models and workforce service activity projections to populate a central model, written in the Python programming language (the Python model). The Python model produced supply and demand projections for 52 professions across five care settings, for each year between 2021-22 and 2036-37. By combining existing workforce trends with funded interventions already underway, the Python model estimated a projected staffing shortfall. NHSE then used the Python model to test the impact of additional policy interventions that might reduce the shortfall by 2036-37, the end of the modelling period. NHSE referred to this exercise as 'shortfall analysis'. Outputs from the shortfall analysis were further processed in separate models, known in NHSE as 'triangulation' models, to calculate the impact of additional interventions to close any remaining gaps between supply and demand. Finally, NHSE manually combined outputs from both the Python model and the triangulation process to prepare the published ranges in the LTWP (paragraphs 2.5, 2.7 to 2.10 and Figure 5).

Model design and operation

9 **We were able to replicate the outputs of the shortfall analysis from the Python model and the code was of good quality.** Although the technical documentation was not sufficient on its own to replicate the outputs from the Python model, we were able to do so with the support of additional interviews with the modellers. Overall, we found the Python code to be logically structured and of good quality, although some sections of the code would benefit from being broken down into smaller components to improve readability and reduce the risk of error (paragraph 2.23).

10 **Aspects of the modelling pipeline that NHSE designed are inherently risky and the modellers made considerable manual adjustments.** NHSE ran the Python model multiple times, changing the input data between runs, to generate supply and demand projections under different service and workforce scenarios. Subsequent components of the pipeline – the triangulation models and the preparation of ranges for public presentation – were carried out in Excel spreadsheets. NHSE modellers moved data manually between the different components in the pipeline. The triangulation process required NHSE analysts to manually adjust domestic education places and international recruitment numbers through an iterative approach to balance supply and demand. NHSE analysts told us that the complexity of the manual adjustments in the triangulation process meant that this component of the modelling pipeline could not be brought into the Python model within the timetable they had been given. The manual adjustments made in the final components of the modelling pipeline and manual transfers of data introduce a risk of model data and assumptions being inconsistent across the modelling pipeline and increase the likelihood of error. NHSE told us that it intends to reduce the amount of manual processing in future versions of the model (paragraphs 2.12 to 2.14, 2.19, 2.20, 2.24 and 2.28).

Model governance

11 **NHSE provided documentation to the NAO regarding all parts of the modelling pipeline, but it varied in quality; this and the use of manual processing limited the replicability of NHSE's analysis.** Some quality assurance documentation for input models was incomplete or lacked evidence of independent scrutiny. For the triangulation models we received documentation explaining the high-level approach and assumptions, but the modellers used a manual and iterative approach in reaching the modelling outputs. These iterations were not documented due to the frequency and volume of changes as NHSE considered different policy options. We were not able to attempt to replicate the triangulation process and test whether the calculations in the model were implemented as intended due to the lack of technical documentation and the complex and undocumented manual adjustments made in this part of the modelling pipeline. This means that we were not able to replicate fully the numbers in the published LTWP. After the end of our fieldwork, NHSE improved its documentation of triangulation. This enhanced our understanding of the process but was not in itself sufficient to replicate the outputs that had ultimately informed the published LTWP. Limitations in documentation increase the risk of error as model outputs cannot be wholly validated through independent quality assurance. This may also make it harder for analysts to repeat and modify analysis when NHSE refreshes the LTWP in the future. We are concerned that the complexity of the whole modelling pipeline, from input models through to the published numbers, meant that dependencies or inconsistencies between different parts of the modelling may not be fully understood (paragraphs 2.22 to 2.30).

Assumptions

12 **NHSE's modellers showed a good understanding of the range of variables that will affect the future size and shape of the NHS workforce but NHSE only communicated a limited range of uncertainty in the published LTWP.** NHSE's modellers understood that there is uncertainty about how key variables will develop in future, such as demand for NHS health services, productivity and staff retention, and also that in many cases one variable will impact on others. However, the modelling pipeline made it difficult for them to produce and share outputs for a full range of plausible future scenarios and potential policy options. For some key assumptions, such as future health service activity, only one future scenario was communicated. Furthermore, NHSE communicated a limited assessment of the uncertainty of its assumptions and how those uncertain assumptions might affect one another in combination (paragraphs 2.18, 2.20, 2.28, 3.2 and 3.3).

13 **Some of the modelling assumptions may be optimistic, given the amount of change they imply.** In practice, some assumptions relate to historical trends, which NHSE thinks will continue, while others are more akin to targets, which will require policy changes and significant investment.

a Workforce productivity is applied annually in the modelling to reduce the projected number of health workers required to deliver the same amount of activity. The modelling assumes that over the 15 years of the LTWP workforce productivity will improve by more than the long-term productivity average. The central assumptions are that there will be high workforce productivity gains in the first three years up to 2024-25 (NHSE told us 1.5% per year), then 0% is applied in the modelling for two years, and then annual improvements of 0.8% from 2027-28 onwards. The Office for National Statistics (ONS) long-term average for healthcare productivity improvement is 0.7% per year, although it should be noted that this is a measure of total healthcare productivity, which is a broader measure than workforce productivity. The ONS measure showed a pre-pandemic decline in healthcare productivity after 2017-18 (paragraphs 3.8 to 3.15 and Figure 7).

b The assumption on increasing domestic education and training – so that medical undergraduate numbers double, and nursing undergraduate numbers nearly double, between 2022 and 2031 – is at the top end of the maximum expansion NHSE thought theoretically possible. NHSE's analysis did not include an assessment of the capacity constraints to an expansion of training on this scale, or the costs required to overcome any constraints. NHSE told us this was something it now planned to assess following the publication of the LTWP (paragraphs 3.21 to 3.27 and Figure 9).

c The modelling assumed that international recruitment would continue to be used to fill gaps until the supply of domestic workers increased. However, the modelling reflects an aim to reduce the reliance on international recruitment and assumes that no international doctors are to be recruited from the mid-2030s. In our judgement, this is not a reasonable modelling assumption (paragraphs 3.28 to 3.31).

d A gap between modelled demand for fully qualified general practitioners (GPs) and the supply of these GPs is to be filled by transferring more work from fully qualified GPs to GPs in training and to specialists and associate specialists in primary care. The total supply of doctors in primary care is projected to increase substantially over the modelled period but the total number of fully-qualified GPs is not. At the end of the LTWP period, NHSE projects only 4% more fully qualified GPs than there were in 2021. In contrast, the number of consultants is expected to grow by 49% (paragraphs 3.32 to 3.36 and Figure 10).

Conclusion

14 NHSE has rapidly, and for the first time, produced modelling that brings together its planning of future NHS health services with its longer-term assessment of the workforce it thinks will be required to deliver them. This is a significant achievement, which provides a foundation for NHSE to build on. We welcome its commitment to improve the modelling as part of the regular planned updates to the LTWP. In our model assessment, we found that the methods used in the Python model in NHSE's modelling pipeline represented a reasonable technical approach to health workforce modelling, and we were able to replicate the outputs from this part of the modelling.

15 However, this first version of the modelling pipeline as a whole has significant weaknesses, including the lack of integration between different parts of the pipeline and the manual adjustments to balance supply and demand gaps in the triangulation models. We found that limitations in documentation and the use of manual processing meant we were not able to fully replicate the results of the modelling as an independent reviewer. Some of the assumptions used in the modelling may be optimistic and the model outputs were weakened by the limited extent to which future uncertainties were communicated. NHSE needs to address these issues in order for the modelling to be a reasonable basis for regular strategic workforce planning.

16 Workforce modelling is highly unlikely to produce a single "correct" answer on how many health professionals will be needed in future. In this context, modelling is really an evidence-based and transparent tool for beginning a conversation, including with external stakeholders, about the desirability and feasibility of different approaches and policies. Our recommendations below and throughout this report are intended to assist NHSE in making its modelling more useful and providing ministers and officials with a better basis for reaching decisions in future. While the decisions taken as part of the LTWP are out of the scope of our review, we note that government has only committed funding up to 2028-29 and that NHSE plans to make changes in stages. This gives NHSE a built-in opportunity to make adjustments, for example to the number of training places, after it has revisited the modelling.

Recommendations

17 Three summary recommendations are listed below. Given the technical nature of many of our specific recommendations, we have placed these at the relevant paragraph in the main body of the report. All our recommendations are listed together in Appendix Three.

a **Modelling pipeline:** NHSE should develop a modelling pipeline whose different parts are fully integrated to avoid manual processing, and which creates the capability to more easily test and produce outputs for a wider range of policy options. The pipeline, or a simplified version of it, should also be more easily shareable to allow for greater scrutiny outside NHSE. (This recommendation can be found at paragraph 2.20).

b **Quality controls:** Before it publishes outputs from the modelling in future, NHSE should ensure that the entire modelling pipeline is documented so that all key decisions are clear, and it can be reproduced independently. NHSE should assign a small team responsibility for understanding the entire modelling pipeline and should ensure quality assurance activities take place in a timely manner. Regarding inputs to the workforce modelling, NHSE should revisit quality assurance arrangements for pre-existing models when they are used for a new purpose. (Detailed recommendations on this subject can be found at paragraphs 2.28 and 2.31).

c **Assumptions about the future:** NHSE should improve its documentation of assumptions. It should be clearer with itself and others, inside and outside government, about how stretching its policy aims are and what would happen if any of them were missed. NHSE can improve confidence in its modelling by producing ranges that communicate more of the uncertainty inherent in its assumptions. This is particularly relevant for NHSE's assumptions on productivity, GPs and the interrelationship between domestic training and international recruitment. (Detailed recommendations on this subject can be found at paragraphs 3.3, 3.7, 3.15, 3.27, 3.31, 3.36 and 3.37).