National Audit Office

REPORT

NHS England's modelling for the Long Term Workforce Plan

NHS England

SESSION 2023-24
22 MARCH 2024
HC 636

We are the UK's independent public spending watchdog.

We support Parliament in holding government to account and we help improve public services through our high-quality audits.

The National Audit Office (NAO) scrutinises public spending for Parliament and is independent of government and the civil service. We help Parliament hold government to account and we use our insights to help people who manage and govern public bodies improve public services.

The Comptroller and Auditor General (C&AG), Gareth Davies, is an Officer of the House of Commons and leads the NAO. We audit the financial accounts of departments and other public bodies. We also examine and report on the value for money of how public money has been spent.

In 2022, the NAO's work led to a positive financial impact through reduced costs, improved service delivery, or other benefits to citizens, of £572 million.

National Audit Office

NHS England's modelling for the Long Term Workforce Plan

NHS England

Report by the Comptroller and Auditor General

Ordered by the House of Commons
to be printed on 20 March 2024

This report has been prepared under Section 6 of the National Audit Act 1983 for presentation to the House of Commons in accordance with Section 9 of the Act

Gareth Davies
Comptroller and Auditor General
National Audit Office

19 March 2024

HC 636 | £10.00

013886 03/24 NAO

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).

Part One

Introduction to health workforce modelling

1.1 This part introduces health workforce modelling and the contribution it can make to health workforce planning in England. It also gives an overview of NHS England's (NHSE's) Long Term Workforce Plan (LTWP).

England's health workforce

1.2 Hospitals and community health services in England directly employ around 1.5 million people with another 194,000 working in general practice (**Figure 1** overleaf). Around half of these workers are professionally qualified clinical staff, who have completed years of education and training to achieve professional registration. The other half provide support to clinical staff or work in infrastructure and support roles.

Health workforce modelling

1.3 Modelling combines information about the current workforce and its training pathways with information on potential future staffing requirements and availability. In broad terms, the information falls into two categories: matters affecting the supply of staff and matters affecting the demand for staff.

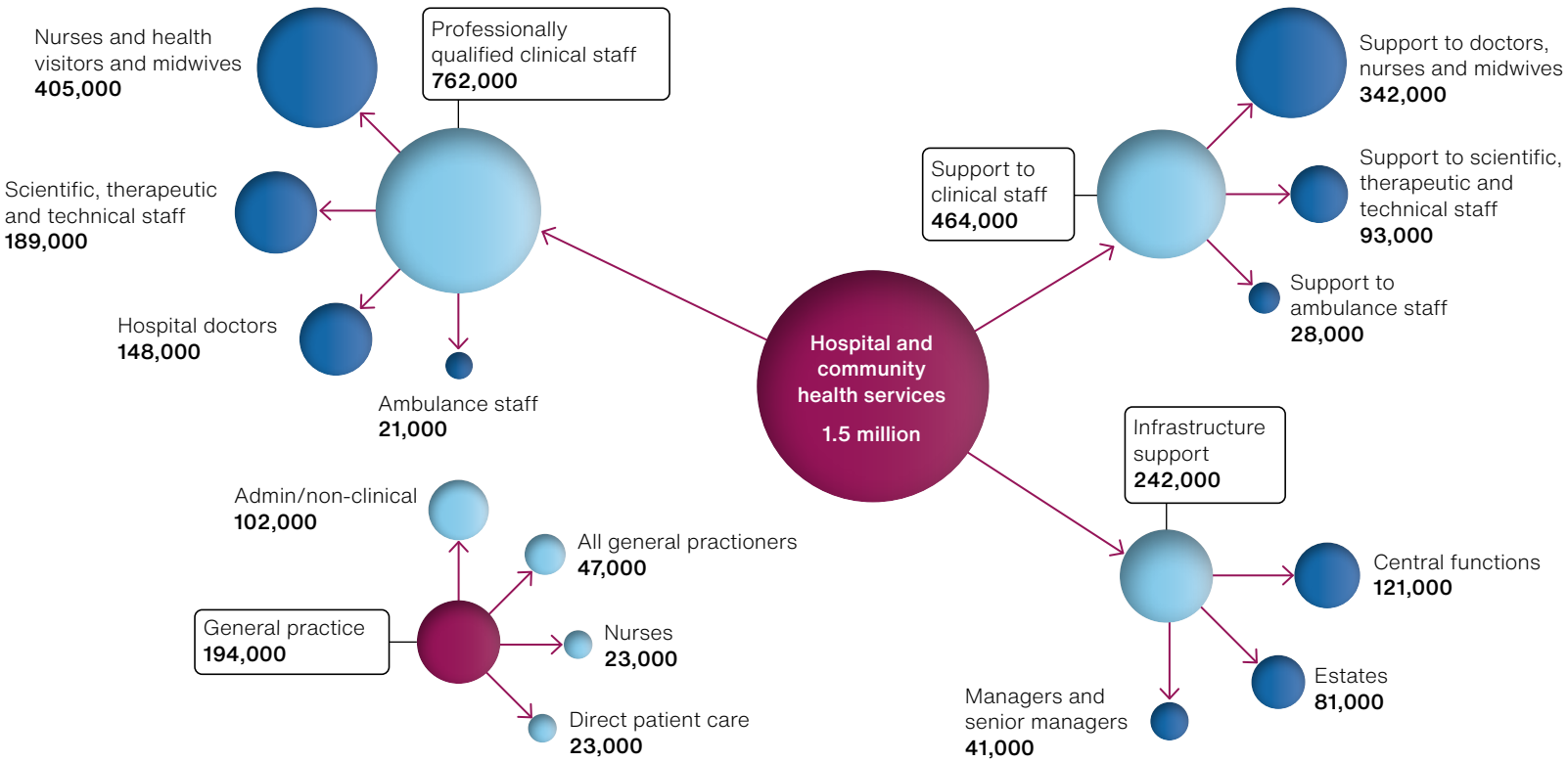
Supply

1.4 The 1.7 million staff employed in the NHS system in England (its total 'stock') change over time as people move in and out of the workforce (movements known as 'flows'). A 'stock and flow' model can be used to project the size of the future workforce by estimating movements over time, including new workers completing training and existing workers resigning or retiring. As with all projections, there is uncertainty about how variables will change in future. Modellers need to make assumptions about how flows in and out will change – typically informed by historical data – and what they know about future policies and aims, in consultationwith relevant experts and stakeholders.

Figure 1

The NHS workforce in England, September 2023

Of the 1.7 million people employed in NHS healthcare services, 1.5 million work directly for hospital and community health services (HCHS) and 194,000 work in general practice



Notes

1 Numbers reported are from published General Practice Workforce statistics, which includes 'traditional general practice' settings, and from published HCHS statistics, which includes 'staff in NHS trusts and other core organisations'.

2 Accounting for people who work part time, the full-time equivalent number of staff who work in HCHS is 1.3 million and in general practice is 146,000.

3 In addition to the 1.7 million people employed in NHS healthcare services, some clinicians work outside of HCHS and general practice in settings including social care, local authorities, third sector and independent sector services. There are less robust data on these areas.

4 Headcount totals are unlikely to equal the sum of components, due in part to individuals working across multiple roles and areas.

Source: National Audit Office analysis of NHS Digital, General Practice Workforce and Hospital and Community Health Service workforce statistics for September 2023; NHS Digital Independent Healthcare Provider Workforce Statistics for September 2022 and Skills for Care Workforce Estimates for 2022-23

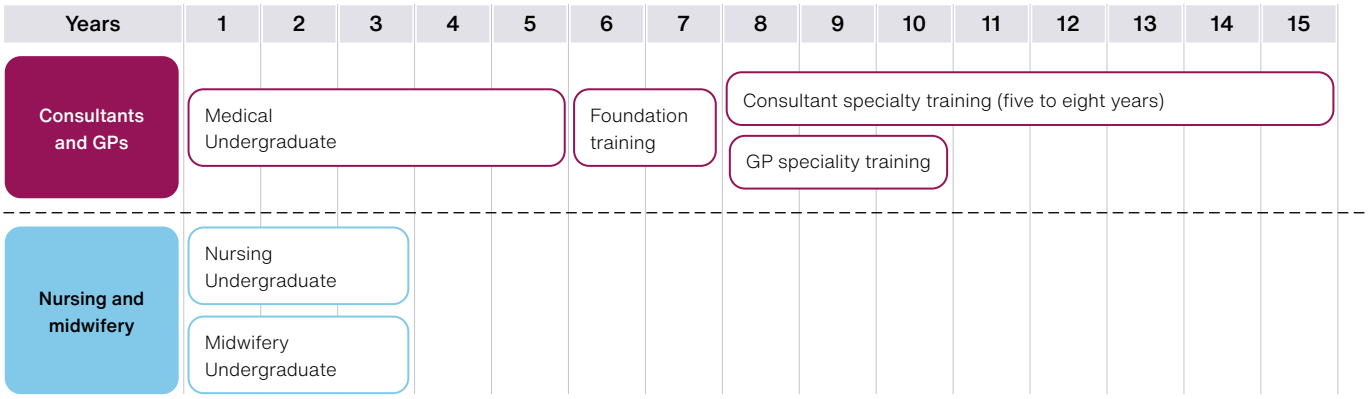
1.5 The education and training of different types of professionally qualified health workers take different lengths of time (**Figure 2**). Workforce modelling can be helpful in projecting how many new people will start training and how many of them will subsequently join the workforce in various years. For this, modellers need estimates of how many people will leave training prior to completion (known as 'training attrition'). Overall, the variables for supply modelling are well established and there are data collections about stocks and flows that can be used to inform supply model assumptions. An additional complexity is that policymakers may choose to model policy choices, such as alterations to the responsibilities of staff in future by broadening the types of workers that can carry them out.

Demand

1.6 Demand for health workers is often modelled by reference to the amount of health activity that has happened in the past and a projection of what will be necessary in future. Such projections are inherently challenging because of the number and range of factors affecting a population's health, the changing services a health service provides, and the changing ways in which services are delivered. There are many potential future demand scenarios. Workforce planners may consult with a range of stakeholders and experts to understand how health workforce demand is likely to change.

Figure 2
Examples of health worker training lengths

Nursing and midwifery training takes three years, and it takes around 10 years to become a fully qualified General Practitioner (GP) and between 12 and 15 years, depending on the specialty, to become a consultant



Notes

1 Following medical training, some doctors may become locally-employed doctors or specialist and associate specialist doctors rather than enter a consultant or GP specialty training pathway. This figure does not show all possible pathways.

2 Full-time nursing undergraduate degree courses in a single field of nursing (adult, children's, learning disability or mental health nursing) last for three years. There are other entry routes into nursing, including apprenticeship and graduate routes.

Source: National Audit Office analysis of NHS health careers information

1.7 When considering future demand, modellers also have to decide whether to take account of the potential future funding and costs of healthcare. In reality, decisions relating to affordability will affect the amount of education and training that a country funds and the quantity of potential future demand a health service will meet, including the size of workforce it can afford to employ.

Why workforce modelling matters

1.8 Workforce modelling is carried out to support the development of long-term workforce strategies in complex organisations such as the NHS in England. It is particularly necessary in healthcare because of the length of time it takes to train health professionals. Where, over the long term, national-level workforce supply is projected to be below national-level workforce demand, planners may need to increase supply through more training to narrow the gap. If long-term supply were to exceed demand, actions may be needed to reduce workforce supply and avoid training too many staff.

1.9 There are other options. Where historical workforce planning decisions have not anticipated the level of workforce demand accurately, changes to workforce supply have tended to be achieved by increasing or reducing recruitment of already-qualified health professionals, primarily from abroad. The NHS also fills gaps in workforce supply through extensive use of temporary staff, which includes bank and agency staff and through overtime arrangements.

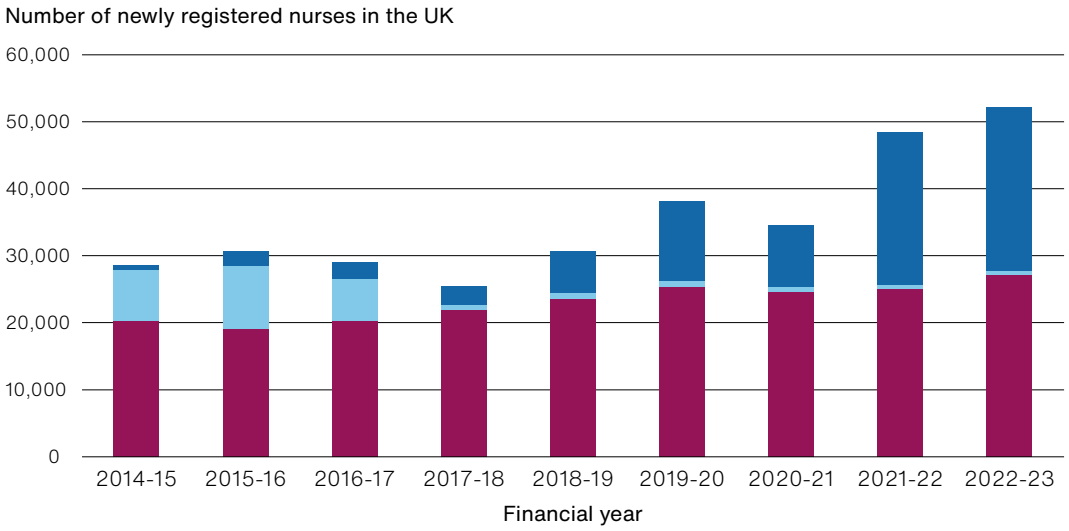
1.10 Training is a long-term investment. The government pays universities and hospitals for elements of healthcare professionals' education and training. In England, this includes the medical and nursing professions, as well as dentistry, allied health professionals, healthcare scientists, and pharmacists. The Department of Health & Social Care (DHSC) estimated that in 2017 it cost around £230,000 to put an individual through medical school. Of this, £163,000 was paid by the government in grants to providers (£151,000) and bursaries to students (£12,000), with the remainder forming a student loan, which typical newly qualified doctors repay over their working life. The government sets student intake levels for medical and dental student intakes. In total, the education and training budget for the NHS's future workforce was £4.5 billion in 2022-23.

1.11 Typically, domestic workforce supply has not satisfied the level of workforce demand of the NHS in England, and the NHS has had a long-term reliance on recruiting health professionals from abroad. The UK overall has the highest share of foreign-trained doctors and foreign-trained nurses of the six reporting G7 countries: 30% for doctors and 15% for nurses in 2019, compared with an overall Organisation for Economic Co-operation and Development average of 18% and 6%. **Figure 3** shows the UK's reliance on nurses trained elsewhere. In 2021-22 and 2022-23, around half (48%) of all new nursing entries on the Nursing and Midwifery Council's professional register were nurses trained outside the UK.

Figure 3

Place of training of newly registered nurses in the UK, 2014-15 to 2022-23

Large numbers of newly registered nurses were trained overseas, particularly in 2021-22 and 2022-23



	2014-15	2015-16	2016-17	2017-18	2018-19	2019-20	2020-21	2021-22	2022-23
UK	20,333	19,114	20,240	21,931	23,498	25,381	24,611	25,028	27,142
European Economic Area	7,518	9,389	6,382	805	968	913	810	663	651
Rest of World	665	2,135	2,403	2,724	6,157	11,933	9,156	22,745	24,355
Total	**28,516**	**30,638**	**29,025**	**25,460**	**30,623**	**38,227**	**34,577**	**48,436**	**52,148**

Notes

1 Geographies are based on where newly-registered nurses with the Nursing and Midwifery Council completed their training.

2 The European Economic Area includes all 27 European Union countries and Iceland, Liechtenstein, and Norway.

Source: National Audit Office analysis of Nursing and Midwifery Council's annual reports

The NHS Long Term Workforce Plan

1.12 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. In June 2023, NHSE published the LTWP, which looks nationally at all major parts of the health workforce over a 15-year time horizon through to 2036-37.[3] The chief executive of NHSE stated that this was the start of an ongoing programme of strategic workforce planning. 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 full-time equivalent (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%. The LTWP estimated a starting shortfall between workforce supply and demand of approximately 150,000 FTE NHS workers.

1.13 Publication of the LTWP was welcomed. The Health Foundation described it as "a major milestone for the NHS". The King's Fund said it was "well overdue" and "an essential first step in solving the NHS workforce crisis". The National Audit Office (NAO) itself had previously called for a long-term strategic plan. In our 2021 report on *The NHS nursing workforce* we identified "central coordination and oversight", "understanding of the future need", and "setting out an overall workforce plan" as challenges facing all those who have to manage large and complex public sector workforces.[4]

1.14 We have not audited in detail the LTWP's policy commitments and aspirations as part of this work. However, we note as positives the fact that it recognises explicitly the pressure many current NHS staff feel, the interdependencies between NHS workforce planning, wider capital investment and the adult social care system, and that NHS performance is currently well below the standards patients have a right to expect.

1.15 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 plan. 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.

The NAO's independent assessment

1.16 The published LTWP is underpinned by modelling that NHSE carried out to establish potential future scenarios for the supply of and demand for NHS workers. As with the plan itself, NHSE has committed to updating the modelling regularly, and publishing the results, around every two years. It has recognised that there is potential to improve the modelling and it plans to incorporate improvements into future versions.

3 NHS England, *NHS Long Term Workforce Plan*, June 2023.
4 Comptroller and Auditor General, *The NHS nursing workforce*, Session 2019–21, HC 109, National Audit Office, March 2020.

1.17 After a request from HM Treasury, DHSC and NHSE, the Comptroller and Auditor General agreed that the NAO would carry out an independent assessment of the modelling. The rest of this report looks in detail at the modelling, with the aim of providing context for users of the current LTWP and assisting NHSE as it updates its models during 2024 and 2025. Overall, our work considers whether the modelling provided a reasonable basis for regular strategic workforce planning and also points to improvements NHSE could make. The rest of this report is organised as follows:

- Part Two examines NHSE's modelling approach to determine if the modelling was logical, accurate and appropriate, and was built and developed robustly; and
- Part Three examines key assumptions the modellers made.

1.18 Our audit approach is set out in detail in Appendix One.

Part Two

NHS England's workforce modelling pipeline

2.1 This part describes NHS England's (NHSE's) production of its workforce modelling. The modelling took the form of a pipeline: a structured sequence of steps involving a series of distinct models. We then provide the National Audit Office's (NAO's) assessment of that pipeline.

NHSE's approach to workforce modelling

Background to NHSE's workforce modelling

2.2 The Secretary of State for Health and Social Care commissioned NHSE to produce the Long Term Workforce Plan (LTWP) in January 2022. NHSE was asked to set out future workforce demand and supply scenarios and include the workforce interventions needed to support the NHS's overall service strategy.

2.3 NHSE's modelling built on *Framework 15: shaping the future workforce*, a working paper setting out analysis of the current NHSE workforce situation and the range of available options to improve it. Produced by NHSE, the Department of Health & Social Care (DHSC) and Skills for Care, it provided useful guidance to the modelling team and facilitated dialogue between modellers and the senior leadership of NHSE as the modelling developed. *Framework 15* was originally intended for publication in July 2022 but ultimately remained unpublished, which removed an opportunity for transparency about how NHSE was approaching the LTWP.

Timeline of the LTWP modelling

2.4 After being commissioned, NHSE agreed the terms of reference for the modelling project in March 2022. At that point, it aimed to complete the modelling within seven months, including iterating policy options. The core modelling team comprised 15 people, including periods when up to six people worked full time on the modelling.

2.5 NHSE engaged the healthcare consulting and data company Carnall Farrar to build a central workforce model to its specifications. In May 2022, Carnall Farrar handed over an initial version of a model to project workforce supply and demand, which was written in the Python programming language (the Python model). NHSE's core modelling team then refined the Python model and updated the input data. It used this together with further modelling, described below, to project supply and demand and to test policy options.

2.6 NHSE's modelling project continued for over 15 months (**Figure 4** overleaf). Although the modellers had substantially completed their work after a year, discussions between NHSE, DHSC and HM Treasury, including refining the modelling of policy options, continued until NHSE published the LTWP in June 2023. It is possible that some of the weaknesses we have identified in the modelling pipeline could have been avoided if NHSE modellers had been able to plan for longer at the outset.

Overview of model components and approach

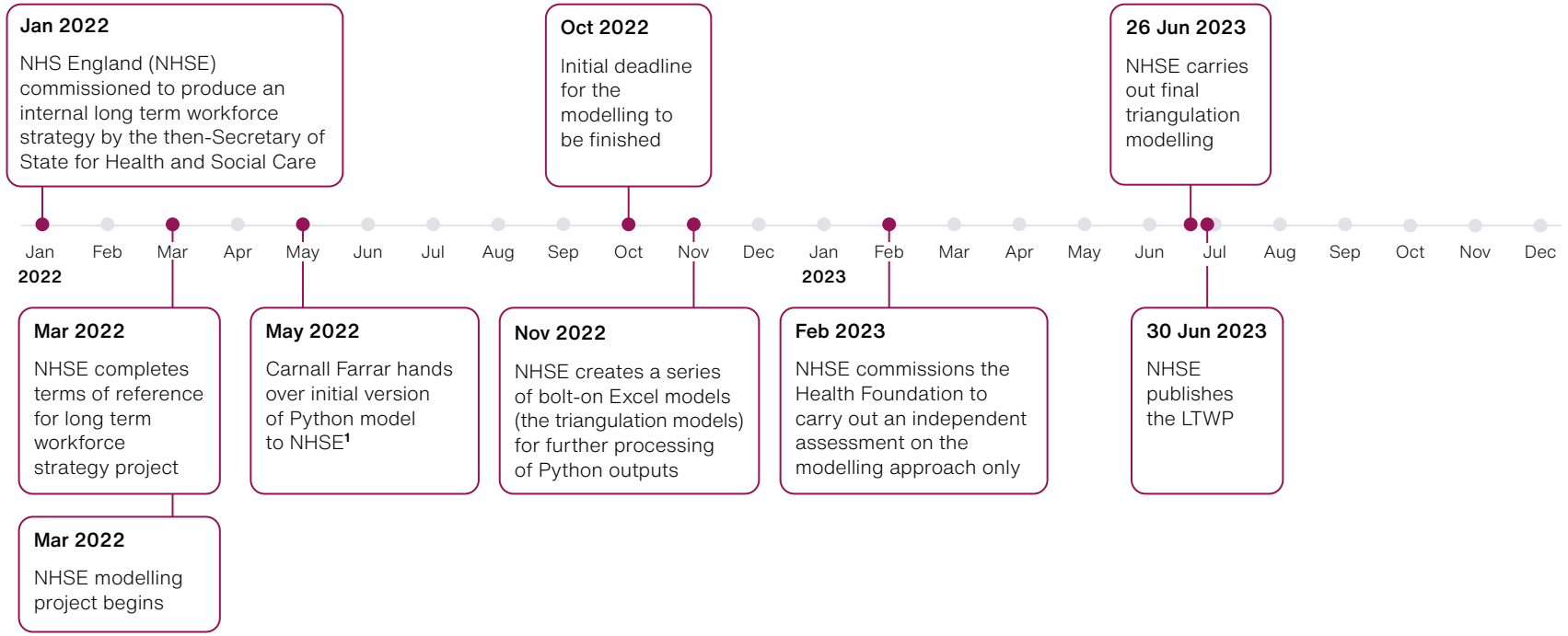
2.7 NHSE's Python model aimed to understand and project workforce supply and demand for each year between 2021-22 and 2036-37 to estimate staffing shortfalls. It did this across 52 professions and five care settings (acute, ambulance, community, mental health, and primary care). NHSE varied modelling assumptions to generate supply and demand projections under different service and workforce scenarios. The modelling was not designed to precisely forecast shortfalls in NHS staffing, but instead to have a level of accuracy sufficient to inform the policy levers and initiatives required to reduce the shortfall.

2.8 The modelling is a series of models with four distinct component parts along a pipeline, shown in **Figure 5** on page 21. NHSE used inputs from 18 input models to populate the Python model that produced the supply and demand projections. The Python model projected a shortfall in workforce supply and the modellers used it to calculate the impact of different policy interventions that might reduce the shortfall. These included interventions to improve productivity, increase retention and change the mix of health staff who perform certain work. NHSE referred to this exercise as 'shortfall analysis'. Outputs from the shortfall analysis were then processed further in separate Excel models to calculate the impact of additional interventions relating to training, education and recruitment to close the remaining gap between supply and demand, a process NHSE refers to as 'triangulation'. Finally, NHSE used a separate Excel workbook to prepare the published ranges in the LTWP, manually combining the different scenario outputs generated from both the Python model and the triangulation process. We now provide more detail on each of these stages.

Figure 4

Timeline of Long Term Workforce Plan (LTWP) Modelling

Modelling for the LTWP took 15 months compared with an original timeline of seven months



Note

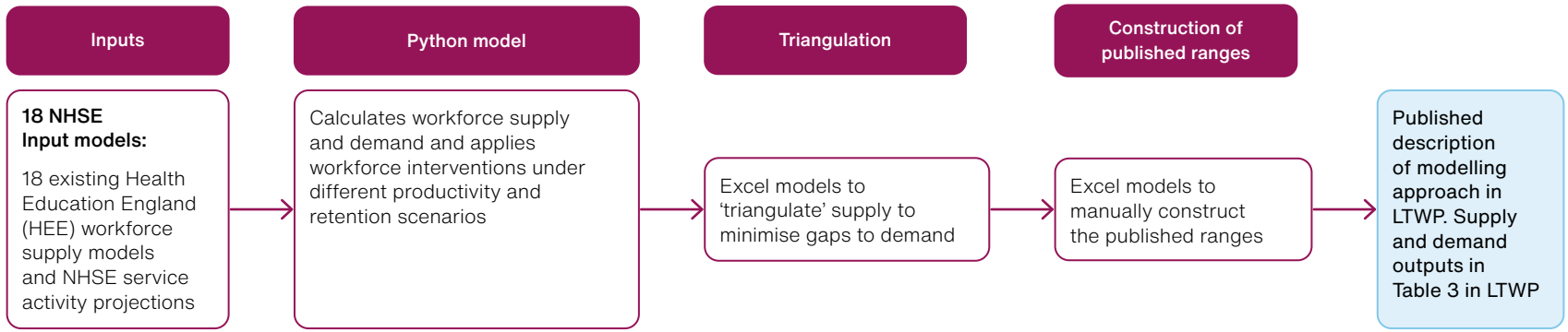
1 After the commission in January 2022 for a long-term workforce strategy, NHSE rescoped an existing engagement with Carnall Farrar to produce a Python model.

Source: National Audit Office analysis of NHS England internal and published documents

Figure 5

Overview of NHS England's (NHSE's) modelling pipeline for the Long Term Workforce Plan (LTWP)

The LTWP modelling is a series of models with distinct elements along a pipeline



Notes

1 On 3 April 2023, NHSE announced the merger of NHSE and Health Education England (HEE). As a result, all input data is owned by NHSE.

2 A 'model run' is defined as producing a new set of outputs from the Python model by adjusting input data and/or assumptions.

3 We define a modelling pipeline as a structured sequence of steps involving a series of distinct models.

4 'Triangulation' is the process NHSE used to manually adjust domestic education places and international recruitment numbers to balance supply and demand by the end of the modelling period (2036-37).

Source: National Audit Office analysis of NHS England's modelling process and unpublished documents

Inputs

2.9 The Python model used outputs from pre-existing NHSE models and other supporting input data to project future supply and demand. The core LTWP modelling team brought together input data from 18 established NHSE models. NHSE engaged with internal teams and external stakeholders to test and refine the input data throughout the modelling project. A list of the input models is provided in Appendix Two.

The Python model

2.10 The Python model combined existing workforce trends with funded interventions already underway to estimate a projected staffing shortfall of between 260,000 and 360,000 staff by 2036-37, known as the 'counterfactual'. The published LTWP reported a counterfactual shortfall of between 260,000 and 360,000 staff by 2036-37.

- To project the number of staff, NHSE took the number of substantive full-time equivalent workers (FTEs) employed in March 2022 as its supply baseline. It then used existing forecasts based on historic growth rates, funded commitments, retention rates, and international recruitment to produce projections for the supply of staff.
- For demand, NHSE used the number of substantive FTEs employed in March 2022 plus the number of temporary staff FTEs in March 2022 as a demand baseline. It then modelled demand projections under two separate time horizons. First, it projected demand up to 2024-25 based on funded service plans and expected trends in performance improvement. Secondly, it used its own estimates for the remaining period up to 2036-37, informed by existing NHSE models of service activity for elective care, urgent and emergency care, mental health, primary and community care, cancer care, and diagnostics.

2.11 The Python model allowed modellers to test the impact of potential policy intervention scenarios on reducing the projected shortfall. To do this shortfall analysis, NHSE worked alongside internal and external stakeholders to identify high-impact policy levers and initiatives. NHSE quantified the impact of proposed interventions relating to changing care pathways, upskilling and expanding new roles, productivity improvements, increasing retention, and reducing reliance on international recruitment and temporary staffing.

2.12 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. The scenario runs assessed the impact of different policy interventions in combination and accounted for future uncertainty in the assumptions on staff retention rates and labour productivity gains, using central, upside and downside values for these assumptions. The cumulative impact of these interventions reduced the gaps between projected supply and demand.

Triangulation

2.13 However, even with the modelled policy interventions, the output projections of the Python model still predicted a shortfall in supply, and consequently NHSE implemented additional modelling to test supply interventions that would close the remaining gap. The modelling team used Excel-based models to manually adjust domestic education places and international recruitment numbers to balance supply and demand by the end of the modelling period. The modellers described this process as 'triangulation'. The triangulation process was generally only applied to supply projections. An exception was the modelling carried out for general practice, which was a separate model incorporated into the LTWP modelling at the triangulation stage and where demand was also manually adjusted (see Part Three).

Constructing ranges

2.14 To get to the published figures in the LTWP, NHSE compared and combined scenario outputs from both the Python and triangulation models in a separate Excel workbook. NHSE constructed ranges for demand and supply projections in the LTWP by calculating the difference between outputs from upside (for example, high productivity and retention) and downside (low productivity and retention) scenarios with those of the counterfactual scenarios.

An illustrative example: modelling the adult nursing workforce

2.15 **Figure 6** on page 25 illustrates this process for adult nursing, the largest single professional workforce group in the LTWP.

- The service demand input models were used to create workforce demand scenarios – this included a demand counterfactual and three scenarios with different productivity adjustments. The range of these scenarios is shown as a light blue area in Figure 6. In the central demand scenario, 0.8% fewer nurses were required each year due to assumed increases in productivity. A higher demand scenario at lower levels of workforce productivity (0.6% per annum) and a lower demand scenario at higher productivity (an average of 0.93% per annum) were also used. The three scenarios also included small adjustments to demand (a reduction of 800 adult nurses in 2036-37 from skill-mix changes). Using these scenarios, the Python model projected a modelled demand range of 313,000 to 326,000 adult nurses by 2036-37.
- The supply projection used multiple scenarios to establish a range, including a counterfactual supply scenario, with no new national interventions to improve supply, and others with lower and higher levels of retention. There were other small adjustments to supply, including the addition of post-registration learners (an additional 600–730 adult nurses in 2036-37). Using these scenarios (shown as a dark blue area in Figure 6), the Python model produced a supply range of 298,000 to 306,000 adult nurses by 2036-37.

- The NHSE modellers then used the separate Excel-based triangulation model to increase manually the number of new adult nursing students. The number of staff recruited internationally was also adjusted: initially increased to fill supply gaps and then reduced in line with growing domestic supply. The range of the three triangulated scenarios is shown as a purple area in Figure 6. NHSE used triangulation to aim for an oversupply that did not exceed 5% between 2028 and 2030 and was then maintained to the end of the modelling period (2036-37), although for adult nursing they aimed for a greater oversupply to allow more workers to move from adult nursing to community nursing. NHSE also assumed a constant supply of 10,300 adult nursing bank staff in each year. This projected a supply range of 317,000 to 327,000 adult nurses by 2036-37. Using this triangulated supply scenario, the gap between central demand estimates and supply was projected to close by 2036-37. The published numbers in the LTWP (table 3, page 123) set out the supply and demand situation as it is projected to be in 2036-37.

NAO assessment of the LTWP modelling

2.16 Our audit approach is based on the NAO's *Framework to review models* (Figure 12, Appendix One). We examined the LTWP modelling pipeline to determine if the models were logical, accurate and appropriate, and had been constructed with appropriate controls. We considered the quality assurance processes in place for a subset of the input data and examined whether the input data were processed correctly in the Python model. We conducted a detailed code review of the Python model and attempted to validate the outputs from the shortfall analysis conducted in the Python model. We reviewed the available evidence and interviewed modellers for the triangulation models, after assessing that there was insufficient documentation to attempt replication. Our methods and evidence base are described in further detail in Appendix One.

Model concept and design

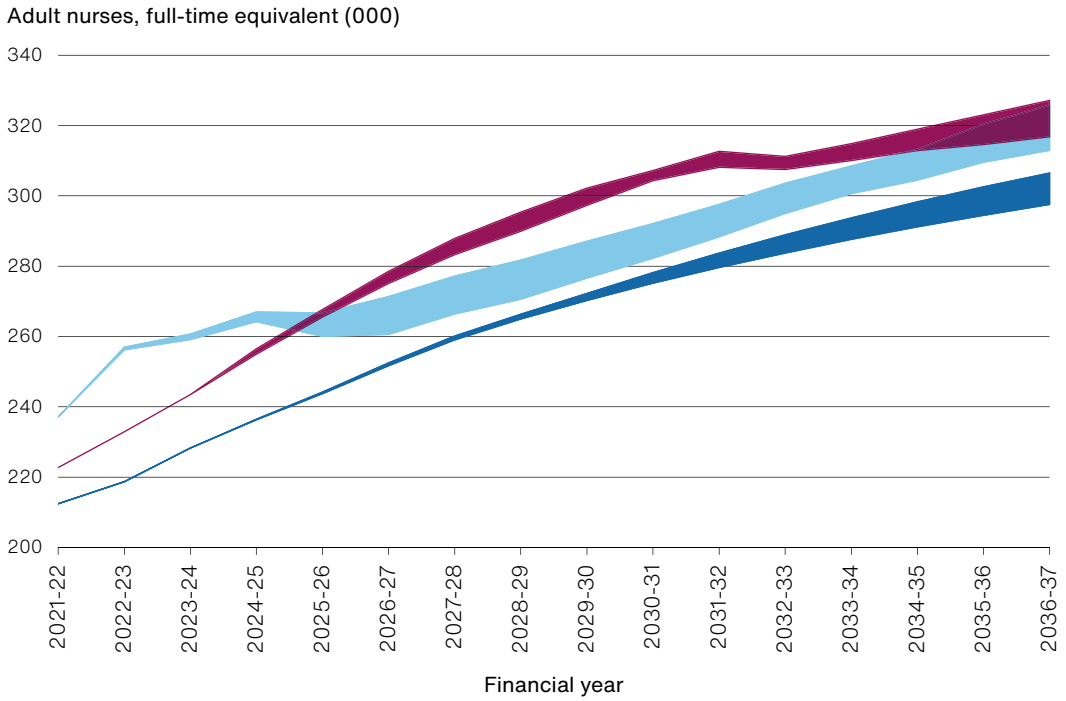
2.17 The methods used by NHSE to project demand and supply and perform shortfall analysis in the Python model represent a reasonable technical approach to health workforce modelling. On the demand side, NHSE has taken a step forward by incorporating its current service plans into workforce modelling and integrating this with its established supply modelling.

2.18 However, it is our assessment that the structure of the Python model and its code created limitations for the efficient testing of additional scenarios. The assumptions underpinning the different workforce scenarios were integrated into the Python model inputs and, therefore, revised sets of model input data had to be produced in order to test each new scenario. NHSE told us it was a relatively quick process to run its chosen scenarios in the Python model, once it had set them up. In our view, alterations to the Python model's design could assist NHSE in testing a wider range of scenarios at speed in future.

Figure 6

NHS England's (NHSE's) supply and demand projections for adult nursing, 2021-22 to 2036-37

NHSE significantly increased supply projections using the triangulation process to address an ongoing shortfall of adult nurses projected by the Python modelling



Notes

1 The Python model projects supply and demand scenarios. 'Triangulation' is the process NHSE used to manually adjust domestic education places and international recruitment numbers to balance supply and demand at the latest by the end of the modelling period (2036-37).

2 The 'triangulation scenario range, plus bank staff' assumes a constant supply of bank (temporary) staff of 10,300 FTE each year.

3 Adult nursing is shown because it is the largest single professional workforce group in the published numbers in table 3, p.123 of the Long Term Workforce Plan, where the difference between multiple scenarios is presented, and supply and demand is reported at 2036-37 only.

Source: National Audit Office analysis of NHS England's modelling

2.19 The final two stages of the modelling pipeline involved NHSE analysts using Python model outputs as inputs to the separate Excel spreadsheet-based supply triangulation models and scenario comparisons workbook. Changing between modelling environments introduces a risk of assumptions being inconsistent in the different environments. The Python model makes an automatic record for each scenario run of the input data used and the model outputs generated. NHSE analysts did not use the automatically recorded data for the triangulation modelling and for comparing between scenario runs. Instead, the analysts manually entered the record of inputs and Python model outputs into the Excel workbooks.

2.20 NHSE analysts told us that the complexity of the triangulation process, including the manual adjustments, meant that this part of the modelling pipeline could not be brought into the Python model during the time frame of the model production. This manual process of transferring data between software environments and comparing scenarios introduces the risk of inconsistency to model data and assumptions, increasing the likelihood of error. NHSE told us that it intends to reduce the amount of manual processing in future versions of the model and incorporate more of the modelling done in triangulation into the core Python model.

Recommendation 1: 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.

Reproducibility and model documentation

2.21 Model documentation is the set of records that enables the transfer of knowledge, including how a model works, its quality assurance, its limitations, and what purposes its results are suitable for.[5] Good model documentation is fundamental and gaps in it can make models difficult to interpret, revisit or review.

2.22 NHSE provided documentation to the NAO regarding all four parts of the modelling pipeline, but the documentation varied in quality and level of detail. Good practice guidance indicates that modelling should be repeatable by an independent reviewer based solely on the supporting documentation.[6,7,8] This was not the case for the LTWP modelling.

5 Comptroller and Auditor General, *Financial modelling in government*, Session 2021-22, HC 1015, National Audit Office, January 2022, paragraph 2.14.
6 HM Treasury, *The Aqua Book: Guidance on producing quality analysis for government*, 2015.
7 HM Treasury, *Review of quality assurance of government analytical models* (Macpherson Review), 2013.
8 National Audit Office, *Framework to review models*, January 2022.

2.23 Although the technical documentation was not sufficient on its own to re-run the Python model, we were able to replicate this part of the modelling with the support of additional interviews with the modellers. We were able to reproduce all the scenarios of the shortfall analysis that were carried out by NHSE to inform the LTWP. 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.

2.24 For the triangulation models, we received documentation explaining the high-level approach and assumptions. The analysts working on the triangulation models used an iterative process of adjustments, and told us that, due to the frequency and volume of changes requested by their policy colleagues, they did not document their decisions throughout this manual process. Where model adjustments are made based on expert judgement, it is important that clear records are kept of decisions. The lack of adequate documentation for this part of the process meant that there was insufficient basis for us to determine why these values were selected.

2.25 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 the triangulation phase, setting out the procedure it had used in its iterative approach. This enhanced our understanding of the process but was not in itself sufficient to replicate the outputs that had ultimately informed the published LTWP. NHSE attempted to partially reperform the triangulation and told us it had come within 5% to 8% of its original total education outputs across nursing and midwifery and allied health professionals (this group includes professions such as physiotherapy). However, there were larger variances for medical undergraduate place numbers, at 13% to 17% below the original modelling, as well as for some key individual professions. NHSE told us that the variances were due to adjustments based on policy decisions and expert judgement. As noted above, we found that these had not all been documented at the time.

2.26 The final part of the modelling pipeline was the preparation of the modelling outputs, summarising the supply and demand ranges for each profession, to publish in the LTWP. The outputs of modelled scenarios were single projections for both supply and demand in each profession. To quantify ranges for the demand and supply counterfactuals and for the impact of different policy interventions, NHSE compared combinations of multiple scenarios. The reasoning behind the combinations of scenarios used in each calculation was not documented and was difficult to interpret from the scenario descriptions. Added to the problems we identified with the triangulation models, this meant it was difficult to trace the logic from Python model outputs to final published values and reproduce the values for verification. This is inadequate for the purposes of an independent review. As described in paragraph 2.6, it is possible that some of these limitations could have been avoided if NHSE modellers had been working to longer original time frames.

2.27 The absence of documentation describing all components of the modelling pipeline increases the risk of error, as model outputs cannot be wholly replicated and validated through independent quality assurance. The lack of documentation will also make it difficult for future analysts to repeat and modify analysis when NHSE reviews the LTWP in the future. In addition to limitations in the technical documentation, the quality of overall documentation varied, lacking a single document on the input data changes used to run different scenarios, and a written succession plan.

2.28 We are concerned that the complexity of the whole modelling pipeline, from input models through to the numbers published in the LTWP, created a risk to the validity of the outputs of the model because any dependencies or inconsistencies between different parts of the modelling may not be fully understood. To generate a new model scenario, the model design required input from at least four analysts across different components, including both shortfall analysis and triangulation. The Python model could only be adjusted by a small number of analysts as the LTWP modelling team had limited familiarity with Python. The triangulation process was carried out by a further two analysts, separate from those working on the Python model. This means that the reproducibility of the modelling process is vulnerable to loss of this expertise.

Recommendation 2: NHSE should improve documentation across the whole of the modelling pipeline to make it easier for future analysts to understand the modelling and any limitations, reproduce model outputs, and update the model when producing future versions.

Recommendation 3: Given NHSE's commitment to update the LTWP at regular intervals, it should assign responsibility for understanding the entire modelling pipeline to a small team and ensure that there is sufficient knowledge of the Python and triangulation models available in the team. NHSE should complete a succession plan for the modelling.

Quality assurance of the modelling pipeline

2.29 Following receipt of the Python model from Carnall Farrar, NHSE built a replica of the Python model in the Alteryx software environment to validate the outputs. In August 2022, it made comparisons between the Python and Alteryx models, identifying some issues, which it then corrected in the Python model. NHSE then arranged internal quality assurance of the Python model code in January 2023, as part of wider peer review on the modelling pipeline and documentation. The Python model was found to be fit for purpose, though with room for improvement of its design. NHSE implemented some of the recommendations in the version of the modelling used for the LTWP, but told us it had decided to defer other recommendations on model design to future versions of the modelling.

2.30 In December 2022, NHSE requested that the analysts of the pre-existing NHSE input and triangulation models review their quality assurance arrangements, including recording the models on NHSE's business-critical register. The quality of inputs is a key driver of the robustness of any modelling. While we did not review technical documentation for the input models, NHSE provided us with quality assurance documentation for the NHSE models that fed into the LTWP modelling. Quality assurance documentation for some of the input models identified as high impact for the LTWP modelling was still incomplete or lacked evidence of independent scrutiny at the time of our review. If analytical assurance activities are not carried out independently or wholly recorded, confidence in the analysis is reduced.

2.31 In February 2023, NHSE asked the Health Foundation REAL (Research and Economic Analysis for the Long Term) Centre to carry out an independent assessment of the modelling approach for the LTWP. The REAL Centre did not review the modelling pipeline itself. The assessment took place over two weeks, with the REAL Centre provided with documentation on methods for various components of the models, a log of assumptions, and quality assurance documentation. The REAL Centre found the data sources and methods used in the model appropriate, and the assumptions used to develop the counterfactual projections plausible, but it said that the modelling of interventions was subject to greater uncertainty due to incomplete documentation.

Recommendation 4: NHSE should ensure quality assurance practices take place in a timely manner, so analysts have sufficient time to respond accordingly.

Recommendation 5: NHSE should revisit internal quality assurance arrangements for existing models when they are used for a new purpose, such as the models used to provide input data for the workforce modelling, and ensure independent scrutiny is evidenced accordingly.

Part Three

NHS England's key modelling assumptions

3.1 NHS England's (NHSE's) workforce modellers had to identify and give values to a number of assumptions to produce their projections. This part examines NHSE's overall approach to testing assumptions and explores how the following specific assumptions have been applied in the modelling:

- demand for services;
- productivity;
- retention of existing staff;
- expansion of education and training;
- international recruitment; and
- general practitioners (GPs) and workforce substitution.

This part concludes by looking at modelling assumptions used elsewhere in the Department of Health & Social Care (DHSC) and NHSE.

NHSE's overall approach to testing assumptions

3.2 NHSE's modellers showed a good understanding of the range of variables that will affect the future size and shape of the NHS workforce. They also understood that there is uncertainty about how each of these variables will develop in future and that, in many cases, one variable will impact on others. The modellers tested some alternative scenarios of how the future might develop. However, as described in Part Two, the Python model's design created limitations for the efficient testing of additional scenarios.

3.3 In some cases – productivity and retention – a small number of different assumptions were tested and NHSE communicated information about this in the LTWP. In other instances, such as demand for health service activity, only one future scenario was communicated. The total range of uncertainty that was communicated is low. Overall, it would have been better if NHSE could have provided a more thorough assessment of the uncertainty of its assumptions, and of how those uncertain assumptions affect one another in combination. In practice, some of the modelling assumptions relate to historical trends, which NHSE thinks will continue, while others are more akin to targets, which will require policy changes, the agreement of professional bodies, and significant investment. Modelling can help to expose and quantify differences between these kinds of assumptions.

Recommendation 6: In the next version of the model, NHSE should communicate outputs for a wider range of plausible future scenarios. NHSE should consider different methods and processes which would allow it to better capture uncertainty in its modelling.

Recommendation 7: Modelling of this kind requires gathering assumptions about the future of the NHS in one place. This presents an opportunity to expose those assumptions widely to scrutiny and challenge, both internally and externally. Assumptions should be generated in transparent and systematic consultation with external stakeholders. NHSE should make more of this opportunity the next time it repeats its modelling.

Demand for services

3.4 As described in Part One, future demand for healthcare workers may be calculated using a range of factors. These may include the projected health status of a population or an assessment of the amount of healthcare a health service will be able to provide in future. The calculation may or may not be constrained by an assessment of affordability.

3.5 NHSE's modellers created projections through to 2036-37 for each of the NHS's main lines of service, including elective care, mental health, and accident and emergency (A&E). These were based on existing activity forecasts where available. In all cases, the modellers expected increasing demand for services. The rates of assumed service growth were different for different areas. The average annual growth from 2025-26 to 2036-37 was 1.4% for admitted care, 3% for primary care, 3.1% for A&E attendances, 4.2% for mental health services and 5.1% for outpatient first attendances.

3.6 We looked in detail at NHSE's assumptions about A&E services to understand how the projections had been created. The modellers' core judgement was that demand for A&E attendances would grow substantially over the 15 years of the plan, settling at 3.1% annually from 2025-26 to 2036-37. At this rate the modelling projected the number of A&E attendances rising from 23 million in 2021-22 to 41 million in 2036-37. This activity–demand scenario contains underlying assumptions, including that all additional service activity demand is affordable and that there will be an NHS estate of sufficient size to accommodate the growth – in this case, a 78% increase in A&E attendances over 15 years.

3.7 Examining a range of funding scenarios for the demand for workers or, alternatively, calculating the likely cost of different demand scenarios would make future modelling more useful. Transparency about the cost of future demand scenarios is desirable because of their public spending implications. The activity-demand scenarios may be addressed with different types of workforce supply, for example through different combinations of domestically trained workers, international recruits and agency staffing. Transparency on the relative cost of different kinds of workforce supply may help contextualise the long-term commitment of resource funding that is implied by decisions to train additional health workers.

Recommendation 8: In future versions of the modelling, NHSE should model and communicate a wider range of different scenarios for the demand NHS services may face. In our view, this could include scenarios in which different amounts of total funding are available, and upside and downside scenarios for demand for particular services.

Productivity

3.8 In the modelling, projected increases in service demand are translated into a need for additional staff by taking account of the number of workers currently required to deliver a particular activity and increasing this number proportionally. However, the future relationship between these variables is not fixed, because NHSE makes an adjustment for productivity improvements. The workforce productivity assumption has a major effect on how many health workers it takes to deliver a given level of activity over the long term. We examined how productivity assumptions operated within NHSE's models.

3.9 NHSE expects the NHS workforce to become more productive each year. In its modelling, NHSE therefore adjusted downwards the total number of workers it will need in future. For early years of the modelled period, the modellers took productivity assumptions from pre-existing service plans, for instance the Elective recovery plan and the *Delivery plan for recovering urgent and emergency care services*. For later years, it modelled scenarios with 0.6%, 0.8% and 0.93% annual workforce productivity improvements, before opting to use 0.8% as its preferred central projection. This means that, at the end of a ten-year period, a fixed level of activity is planned to be delivered with 92% of the workforce needed at the beginning. 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.

3.10 Our assessment of the evidence is that in recent years it has become harder to rely on steadily improving NHS productivity. The Office for National Statistics (ONS) produces a time series of non-quality adjusted productivity changes for healthcare in England (**Figure 7** overleaf). The ONS calculates this productivity measure by comparing growth in all healthcare outputs with growth in the total inputs used. The ONS measure of productivity shows that long-term productivity gains before the pandemic (1996-97 to 2018-19) averaged around 0.7% annually in the non-quality-adjusted measure. The productivity average is sensitive to the time period used. Productivity declined after 2017-18 before reducing by around 26% in the first full year of the pandemic, 2020-21. More recently, we know from NHSE data that, as input costs and workforce numbers continued to rise, the total amount of output did not return to pre-pandemic levels.

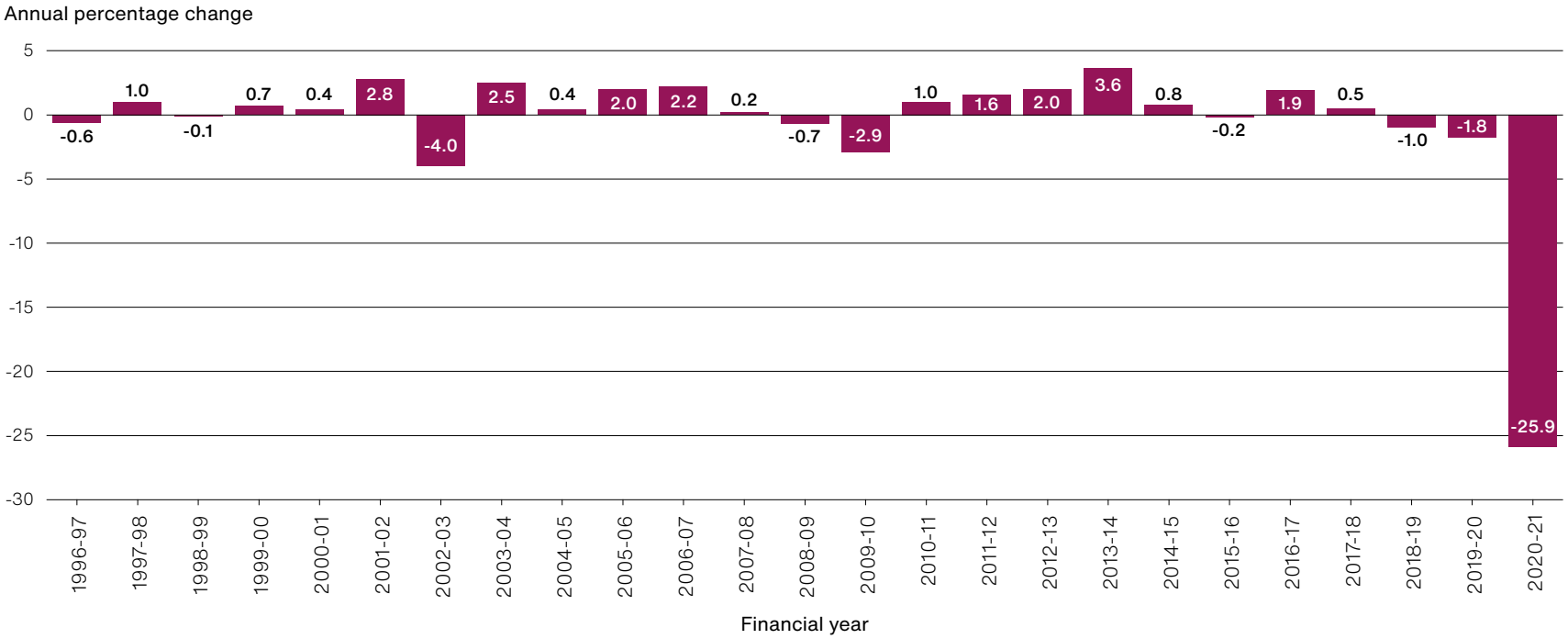
3.11 NHSE's main assumption of 0.8% annual workforce productivity improvement is not directly comparable with the ONS's long-term average of 0.7%, which is a broader measure of total productivity. While the ONS does not calculate a separate measure of workforce productivity in isolation, we think it would be prudent for NHSE to understand more about the relative contribution of workforce (including management), physical infrastructure and technology so that it is applying the most appropriate indicator for workforce capacity modelling. Achieving consecutive years of annual productivity improvements of more than the long-term average for total productivity appears challenging in the current context.

3.12 We believe there may be a specific issue for NHSE to explore with regard to the large expansion the LTWP foresees in the number of new workers entering the NHS each year. The modelling currently assumes that an additional health worker always carries out the same amount of activity as an existing worker. In our view, this may not be the case in a future where the NHS is comprised of a higher proportion of newly qualified workers. Unless the NHS is ready to use each of its workers optimally from soon after they commence service, this model assumption is likely to be optimistic.

Figure 7

Public service healthcare productivity, annual percentage change, England, 1996-97 to 2020-21

Total healthcare productivity varies between years and has been affected by the COVID-19 pandemic



Note

1 Healthcare productivity can be quality or non-quality adjusted. Data shows non-quality adjusted values.

Source: National Audit Office analysis of data in Office for National Statistics, *Public service productivity, healthcare, England*, 2023

3.13 In the published LTWP, NHSE described productivity in a different way, stating that a long-term productivity gain of 1.5% to 2% each year would be possible. NHSE told us that this assumption captures the overall service changes that are implicit in its projections of the future as described in the input models to the modelling pipeline. This is comprised of:

- improvements in technology and the NHS's physical infrastructure (an estimated 0.5% to 1% annual FTE saving compared with not having improvements in technology and the estate);
- pathway improvement in acute care (an estimated 0.4% annual FTE saving);
- delivering the same value of care in cheaper settings (0.2%);
- a reduction in the cost per FTE from introducing new roles (0.2%); and
- adjusting to living with COVID-19 (0.2%).

3.14 In internal documents, NHSE has stated that the 1.5% to 2% annual productivity improvements it is targeting are "highly ambitious". It noted that they will only be possible with sustained investment in technology and the estate. The following are some of the factors that NHSE has said will determine whether the assumptions can be delivered:

- forty new hospital schemes will be completed by 2030 (the NAO report *Progress with the New Hospital Programme* found that only 32 new schemes are now planned to be completed by 2030);
- wider improvements in the acute estate will reduce lost capacity due to backlog maintenance;
- the primary care estate receives investment; and
- the NHS doubles its spending on digital services.

3.15 On the face of it, this measure is even more stretching than both the one applied directly in the modelling and the ONS's measure of historical productivity. NHSE told us that the difference was partly because it was counting some kinds of productivity improvement that are not currently captured by the ONS. NHSE is in discussions with the ONS about future changes to the measurement approach.

Recommendation 9: The modelling (at 0.8% per annum productivity increase) and the LTWP (which describes 1.5% to 2% per annum productivity increase) present productivity differently. In future, NHSE should more clearly explain both internally and in public how it has calculated different measures of productivity and how these relate to the National Statistics produced by the ONS.

Recommendation 10: Ahead of producing the next version of the models, NHSE should review measures of productivity, and other workforce capacity models, to ensure that it is applying the most appropriate measure for its modelling.

Recommendation 11: NHSE's modellers should consider whether it is reasonable to expect the same rate of increasing productivity from a workforce increasingly composed of newly qualified staff.

Retention

3.16 There is a close relationship between the need for new NHS workers and the service's ability to hold on to those it already has. We looked at the assumptions NHSE modellers made about retention.

3.17 NHSE explored three retention scenarios, all of them based on the application of historical data. It used unpublished data about staff turnover, which measure the proportion of staff who leave the NHS each year (the leaver rate), excluding those who subsequently join another NHS organisation (for example, consultants moving between hospitals). **Figure 8** shows how the published leaver rate, which counts staff moving between organisations as leavers, has changed over time. NHS England's internal measure is around 3% lower than the published rate.

3.18 All NHSE's modelled scenarios take a 2022 NHSE leaver rate for each profession as their starting point. By way of example, they assume the following changes in future for adult nursing, beginning with a leaver rate of 6.8%:

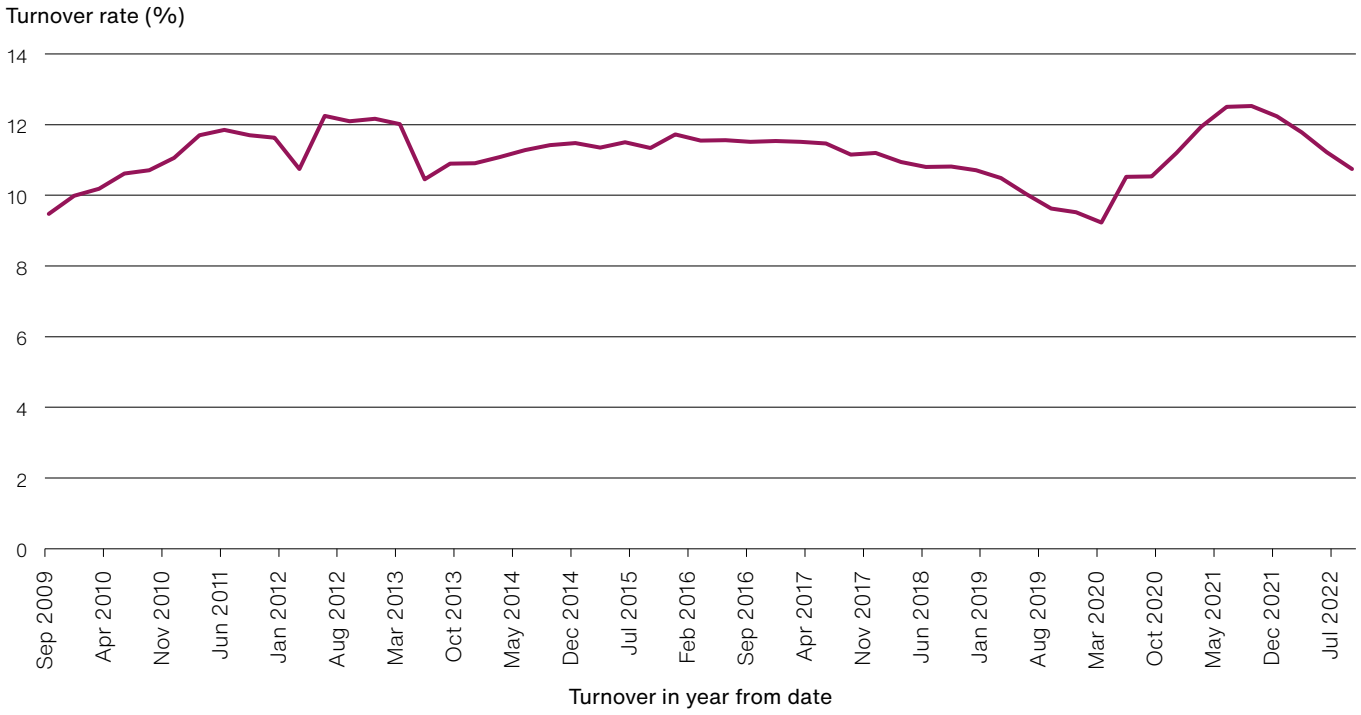
- a change to reach a leaver rate of 6.7% in 2036-37, between the pre-COVID minimum and the pre-COVID average leaver rate (the moderate scenario);
- no change to reach the pre-COVID average leaver rate of 6.8% (average scenario); and
- a change to reach the pre-COVID minimum leaver rate of 6.4% (stretch scenario).

3.19 As with productivity improvements, a key question for the modelling is how to account for uncertainty, particularly because of unusual volatility in data indicators during the pandemic. NHSE expects the leaver rate to stabilise in 2023-24 and for the impact of retention interventions to begin to be felt in 2024-25. Differences in retention have significant effects on the supply projections: NHSE reports that by 2036-37 the long-term difference between its highest and lowest retention scenarios is equivalent to a difference of around 73,000 full-time equivalent (FTE) staff in the workforce, or 3% of all staff.

Figure 8

NHS published staff turnover rate 2009–2022

The long-term average of overall published NHS staff turnover is around 11% pre-pandemic (September 2009 to December 2019)



Note

1 The published leaver rate from NHS Digital includes all turnover from 'NHS Trusts and other core organisations in England' and so will include all leavers. Turnover data are based on headcount and shows people leaving active service, including those going on maternity leave or a career break, for example.

Source: National Audit Office analysis of NHS Digital, 2023

Expanding education and training

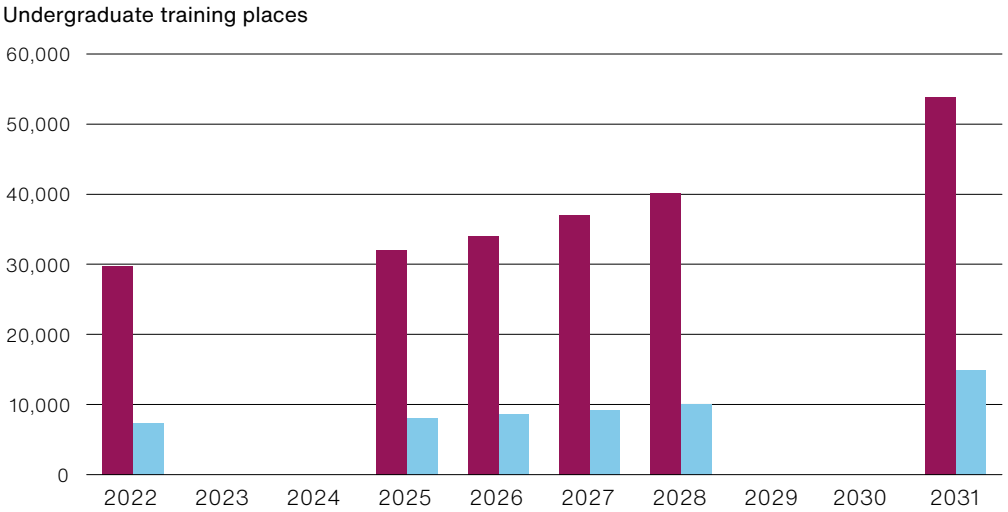
3.20 The LTWP intends to significantly expand domestic education, training and recruitment. The relevant assumption in the modelling relates to the speed and extent of the expansion in domestic education.

3.21 To inform its modelling, NHSE carried out analysis of student applications and of the numbers trained in other countries. Its analysis identified 'maximum theoretical' student intakes in 2030 of 9,360 to 15,000 for medicine and around 19,360 to around 38,370 for adult nursing. According to the plans set out in the LTWP, by 2031-32, the number of places for new medical undergraduates will have doubled to 15,000. Similarly, the number of undergraduate adult nursing places will almost double to nearly 38,000, as part of a general increase in nursing training places to nearly 54,000 (**Figure 9**). For both, this is at the top end of the range that NHSE's analysis suggested could be possible.

Figure 9

Expansion of undergraduate medical and nursing places, 2022 to 2031

Medical school intakes are funded to increase by around a third by 2028-29 and the Long Term Workforce Plan proposes that they double by 2031-32



	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031
Undergraduate intake nursing training places	29,860			32,124	33,981	36,999	39,999			53,858
Undergraduate intake medical places	7,500			8,200	8,700	9,350	10,000			15,000
Nursing percentage increase (%)				8	14	24	34			80
Medical percentage increase (%)				9	16	25	33			100

Notes

1 Percentages are increases in student intake compared with 2022 intake.

2 There are gaps in years 2023, 2024, 2029 and 2030 because NHSE has not published the anticipated growth between 2022 and 2025 and between 2028 and 2031.

3 Adult nursing is the largest branch within nursing studies, accounting for 70% of the total planned intake of around 54,000 in 2031.

Source: National Audit Office analysis of NHS England's Long Term Workforce Plan

3.22 Increasing the number of UK students being educated and trained for careers as clinicians, nurses and other healthcare staff will depend on sufficient applicants to fill the places and sufficient capacity to train them to the appropriate standard. NHSE's analysis of how many students the education and training system could train was based on its estimates of how many people would apply and its assumptions, which included that:

- many more applicants, up to three-quarters for some courses, would be successful in obtaining a place; and
- education and training providers would be able to increase their capacity to accommodate NHSE's estimated number of successful applicants.

3.23 In NHSE's analysis, the attrition rate of students was unaffected by expansion, meaning the same proportion of new starters are assumed to go on to complete their training as is currently the case. NHSE's modelling also assumed the proportion of successful students who then start work in the NHS remains at the current level.

3.24 Such a rapid expansion of capacity will require a number of challenges to be overcome. Higher education institutions (HEIs) are independent and have considerable freedom in determining how many places they wish to offer in particular subjects. For HEIs that already provide relevant courses, NHSE can encourage them to offer more, and the government may fund them to do so. But HEIs might choose to expand at a slower pace, or not at all. NHSE can also support other HEIs to open medical schools or provide relevant courses. While the higher education sector generally welcomed the LTWP, it highlighted that there were real-world constraints to growth, including a need for investment in new facilities, educational technologies and clinical academics. We have confirmed that NHSE's analysis of the theoretical maximum number of places did not include assessing the scale of these constraints or the cost of overcoming them. NHSE recognised that its projections were not based on a detailed analysis of the practical feasibility of expansion and told us this was something it would now assess following the publication of the LTWP.

3.25 Other potentially important factors that NHSE's analysis did not examine but which we think deserve further consideration include:

- the potential interaction between aiming to expand training rapidly and also seeking to maintain quality standards; and
- whether accepting applicants with lower levels of attainment could have an adverse effect on student attrition rates.

3.26 Significant parts of undergraduate and all postgraduate clinical training require practical on-the-job experience in healthcare settings. Existing NHS staff teach and supervise these clinical placements. NHSE factored in its modelling the time students will need for this. However, it is yet to fully consider the following points:

- the capacity of existing NHS staff to provide on-the-job and other training to a large number of additional students;
- whether so much additional training would reduce the quantity of frontline services those NHS staff can provide; and
- the likely need for investment in clinical placement infrastructure. Universities UK reports that this is one of the main barriers to increasing health education system capacity. Similarly, the Royal College of General Practitioners has stated that primary care does not have the space and resources even to support current medical and other trainees.

3.27 We consider that NHSE's assumptions for the rapid expansion of training are optimistic. It should be noted that the main increases in places are planned to happen from 2025, with NHSE not needing to reach the theoretical maxima until over five years later. This allows some time for adjustments to its plans as it revisits them every two years.

Recommendation 12: Future versions of the modelling would be improved by greater consultation and by taking better account of the factors that could limit the expansion of education and training, and of any reductions in service that might result from rapid increases in the amount of on-the-job training.

International recruitment

3.28 NHSE's intention is that increases in domestic education will reverse the NHS's historical and growing reliance on recruiting professionals trained abroad. We explored how, during triangulation, NHSE had adjusted assumptions about international recruitment to reflect the increasing domestic supply that is projected.

3.29 Decisions on domestic training take many years to affect the number of fully qualified health professionals. Individuals in the cohort of medical students who begin studying in 2028-29 will start to be fully qualified general practitioners in 2038-39 and consultants between 2040-41 and 2043-44 at the earliest.

3.30 In broad terms, the modelling showed the relationship we would expect between international and domestic supply: the former decreasing as the latter grows. However, with regard to doctors, NHSE told us that the modelling reflected an aim to reduce reliance on international recruitment. This has resulted in an endpoint that was extreme: no international recruitment at all by the final years of the modelled period. In our view, this is not a reasonable modelling assumption. Few countries do not recruit any doctors from abroad. There is international competition for health workers because their skills are in demand globally. Additionally, many constituent parts of the NHS, including hospital trusts, are free to recruit from overseas if they wish.

3.31 If the assumption about zero international recruitment of doctors is wrong, then the risk within the model's own projections is of an oversupply of doctors in England in future. This would mean that, under the LTWP, too many medical students were being educated from the early 2030s onwards.

Recommendation 13: NHSE's modelling should incorporate more realistic assumptions on the minimum level of international recruitment of doctors in future.

Recommendation 14: Given the long time frames required for medical training, NHSE should stress test their assumptions on training places for medical students by running their modelling over a longer time period than that covered by the LTWP, to ensure that assumptions about training places do not create long-term risks of excess supply of doctors.

General practice and workforce substitution

3.32 In the course of our assessment, we noted differences in the projected growth of consultants and fully qualified GPs by the end of the modelling period. In the central triangulated scenario, the number of consultants is projected to grow by 49% between 2021-22 and 2036-37 (54,800 to 81,600 FTE). By comparison, in the same scenario, the number of fully qualified GPs is expected to grow by 4% (27,800 to 28,900 FTE) (**Figure 10** overleaf, which also shows the trend in actual FTE between 2015 and 2023). We sought to understand the assumptions that explained this difference.

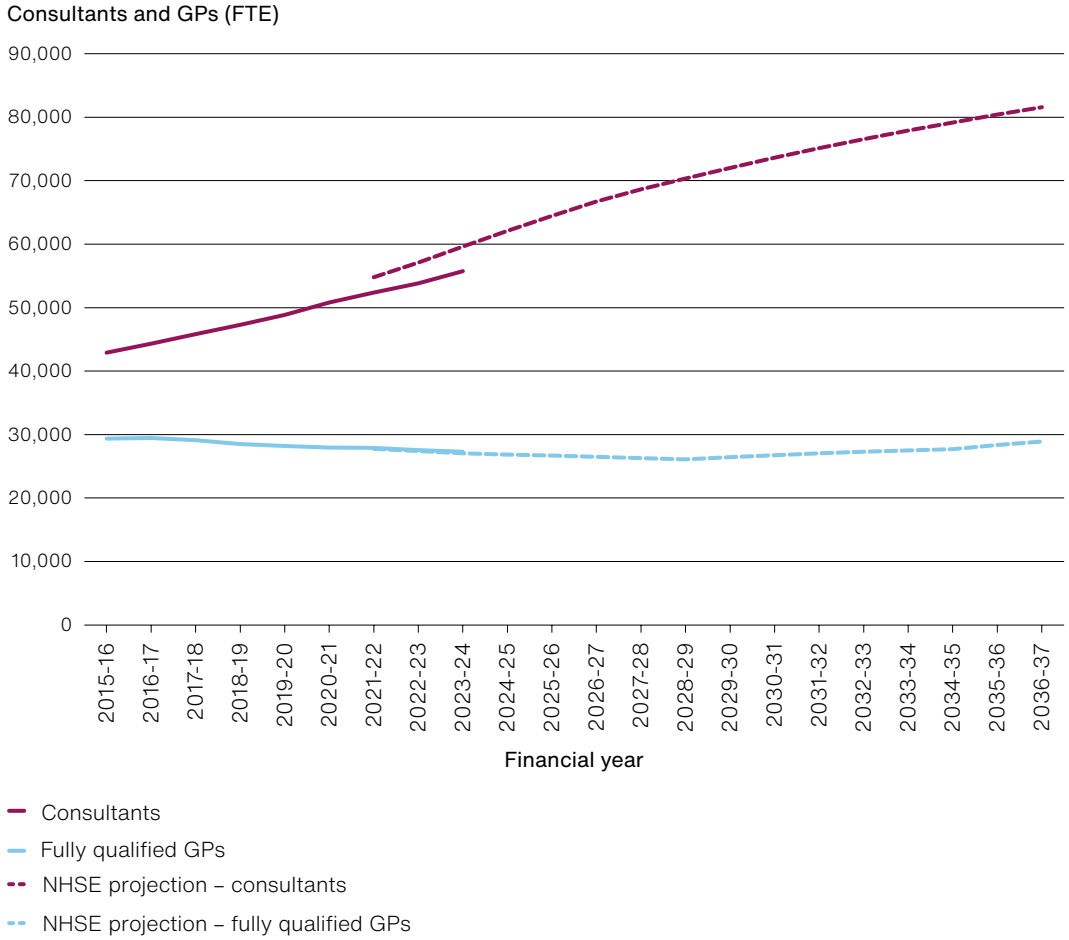
3.33 NHSE modellers treated general practice in a different way to hospital-based professions. Their initial modelling of future demand for, and supply of, doctors for general practice was not carried out in the Python model but in a pre-existing primary care model. We have not reviewed that model but have looked at its outputs and how they were incorporated into the modelling at the triangulation stage.

3.34 The initial modelled outputs showed a growing supply and demand gap for fully qualified GPs. In the LTWP, NHSE combined fully qualified GPs with GPs in training, from a starting supply of 36,000 FTE in 2021-22 to between 50,000 and 53,000 FTE in 2036-37. As shown in **Figure 11** on page 43, the supply growth in the modelled period consists mainly of trainee GPs – who account for 93% of additional supply reported in the LTWP – as well as making increased use of specialist and associate specialist doctors in primary care (SAS doctors). 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.

Figure 10

Actual and projected consultants and fully qualified general practitioners (GPs), 2015-16 to 2036-37

The number of fully qualified (full-time equivalent, FTE) GPs has remained relatively flat since 2015 while the number of FTE consultants has steadily increased



Notes

1 NHSE projection for consultants assumes a constant supply of bank (temporary) staffing of 1,700 each year.

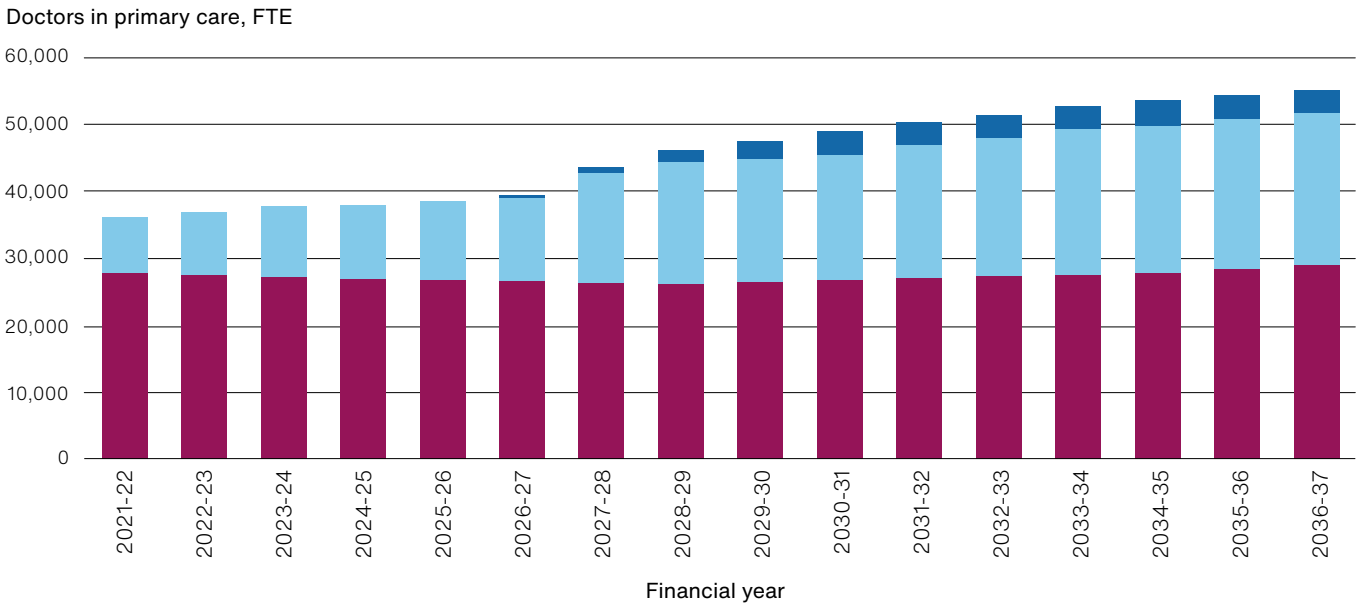
2 Actual values are from September of each year.

Source: National Audit Office analysis of NHS Digital's workforce statistics, 2023 and NHS England's modelling

Figure 11

NHS England (NHSE) supply projection of total doctors in primary care, 2021-22 to 2036-37

The Long Term Workforce Plan counts fully qualified general practitioners (GPs), GPs in training and Specialist, Associate Specialist and Specialty (SAS) doctors in primary care towards its workforce capacity in general practice



Notes

1 Fully qualified GPs have completed full training in general medicine. They can practice independently without supervision and will supervise GPs in training.

2 SAS doctors in primary care are qualified doctors who do not hold a Certificate of Completion of Training in General Practice. Primary care services provide the first point of contact in the healthcare system.

Source: National Audit Office analysis of NHS England's modelling

3.35 NHSE told us that general practice had already been changing in recent years to make use of a wider range of medical and clinical professionals, and that the plans in the LTWP would continue in that direction. The LTWP also states that NHSE wants trainee GPs to spend more of their time in primary care settings in future. Taken together, these changes constitute one of the most significant examples of workforce substitution in the modelling, whereby some of the work typically done by one type of health professional today is moved to another type of health professional in future. When modelling such shifts, the choice of substitution rates is significant. The modelled substitution rate that NHSE applied for general practice is that one GP in training is equivalent to 19% of a fully qualified GP (FTE). NHSE told us that this substitution rate includes an offset in fully qualified GPs' time due to increased supervision and the time that GPs in training spend in other care settings. For SAS doctors, NHSE modelled that one SAS doctor in primary care was equivalent to 70% of a fully qualified GP (FTE).

3.36 NHSE accepts that these assumptions will need to be tested and evaluated. The LTWP notes the need for investment in the primary care estate, which could be a constraint with such a large expansion in training. The next version of the modelling, if it is also done on a 15-year timeframe (or longer, as in recommendation 14), should show the extent to which the large increase in trainee GPs up to 2036-37 eventually converts into increased numbers of fully qualified GPs.

Recommendation 15: Ahead of the next version of the LTWP, NHSE modellers should consult further with a wide range of stakeholders, including Royal Colleges, healthcare providers and other health bodies about the appropriate substitution rates to use when meeting demand for one type of healthcare worker by employing another and should attempt to understand the real-world workload effects on general practice of changes in roles in primary care. NHSE should also be clear about how any proposed changes in skill mix fit with its wider agenda of progressively moving care out of hospital settings and into the community.

Assumptions elsewhere in DHSC and NHSE

3.37 DHSC and NHSE have modellers and analysts who look at other aspects of the NHS's long-term future, including the future design of hospitals and future investment in digital and medical technology. There is an opportunity to improve planning overall if these modellers use a common set of core assumptions where possible, in particular with regard to productivity and service configuration. There are also important crossovers between the NHS workforce and the adult social care workforce, and, therefore, opportunities for analysts and modellers in the NHS and DHSC to share their understanding of how workers flow between these sectors and their wider labour market assumptions.

Recommendation 16: NHSE analysts should document and share their long-term assumptions, for each iteration of the LTWP, within NHSE and with HM Treasury and DHSC. NHSE analysts across the organisation should use common assumptions where possible.

Appendix One

Our audit approach

Our scope

1 After a request from HM Treasury (HMT), the Department of Health & Social Care (DHSC) and NHS England (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 as a part of developing and regularly updating the LTWP projections; and
- to consider whether NHSE's approach to workforce modelling and the models themselves are a reasonable basis for regular strategic workforce planning.

2 To reach our overall conclusion, we considered the extent to which the modelling:

- is designed and implemented in a way that produces reliable outputs with minimal risk of error;
- includes realistic, transparent assumptions;
- is replicable and allows for regular updating; and
- supports consideration of strategic decisions about the health workforce.

3 Our findings and judgements are based on the analysis of evidence we collected between July 2023 and February 2024. We were provided with presentations on the modelling by NHSE in August 2023 and given access to the modelling in September 2023. In forming our findings, we drew on a range of study methods and a variety of evidence sources, which are set out in the paragraphs below. The rest of this appendix describes the methods we used and our approach to model review.

Our methods

Interviews with officials

4 We conducted 25 virtual interviews with NHSE staff, selected to participate because of their job roles and relevance to the audit. This included staff responsible for (or involved in):

- workforce planning;
- workforce modelling;
- analytics and insight;
- strategy; and
- education and training.

5 In addition, we held two interviews with officials from the Workforce Strategy and Insights team at DHSC and two interviews with officials from the health and social care team at HMT.

Fieldwork

6 Fieldwork took place between July 2023 and February 2024. Interviews were carried out online. They typically lasted for one hour.

7 To provide an overview of the modelling process, we attended four online teach-ins in August 2023 with officials from NHSE. They presented on the following topics: the conceptual phase of modelling, the modelling environment (Databricks), workforce high impact areas and retention assumptions, supply modelling, and the overall model approach.

8 We held a further 12 interviews focused on the following topics:

- the modelling pipeline;
- productivity;
- higher education;
- model input data and assumptions; and
- supply-side triangulation.

9 A large part of our audit work was spent reviewing and running NHSE's Python model. To help us run the model, we held nine meetings with NHSE analysts. These interviews were carried out online, typically lasting 30 minutes.

Analytical approach

10 We analysed notes of interviews to:

- inform further lines of enquiry that were followed up with NHSE;
- inform our understanding of the LTWP modelling and its underlying assumptions;
- triangulate evidence from other sources (including document review and evidence from wider stakeholders and service-users); and
- report on NHSE's perspective of the LTWP.

Interviews with external stakeholders

11 We interviewed external stakeholders to inform our study approach and capture a range of views and perspectives on the LTWP. We identified stakeholders via desk research and discussions with NHSE and invited them to participate in an interview. We carried out seven virtual interviews. We spoke to the British Medical Association, the Council of Deans of Health, the General Medical Council, the Health Foundation, the Medical Schools Council, National Voices, the Nursing and Midwifery Council, the Royal College of General Practitioners and Universities UK.

Fieldwork

12 Fieldwork interviews with external stakeholders took place between July 2023 and November 2023. Interviews typically lasted one hour.

13 We provided a brief outline of our study and the issues we wished to discuss. Each interview was tailored to each stakeholder's area of expertise. We sought views on:

- how NHSE engaged with them on the modelling and production of the LTWP; and
- areas to explore further through our assessment.

Analytical approach

14 Interviews were carried out online and typically lasted for one hour. We analysed the interview notes for common themes and issues. We used this qualitative information to inform our study approach and to develop our conclusions and narrative.

15 As the sample was small, the prevalence of views and experiences arising from the stakeholder interviews are not reported. The views expressed allowed us to understand areas in more detail and gave us lines of enquiry to explore.

Document review

16 We reviewed a range of documents from NHSE as well as from other sources. These assisted us with defining the parameters of the assessment and deepened our understanding of the LTWP and its underlying assumptions. This included a review of:

- board and committee meeting minutes and slide packs;
- model control documentation;
- model concept and design documentation;
- analytical assurance documentation;
- higher education sector responses to the LTWP;
- health policy think tank and research institutions' analyses of NHS workforce trends and assumptions within the LTWP; and
- reports on the state and trend of NHS workforce and related dependencies, such as estates, infrastructure, productivity and primary care; and NHSE analyses of individual assumptions underlying the model, such as productivity and training expansion.

17 Our review was carried out between July and February 2023. Most of the documents reviewed were produced from January 2022 onwards.

Analytical approach

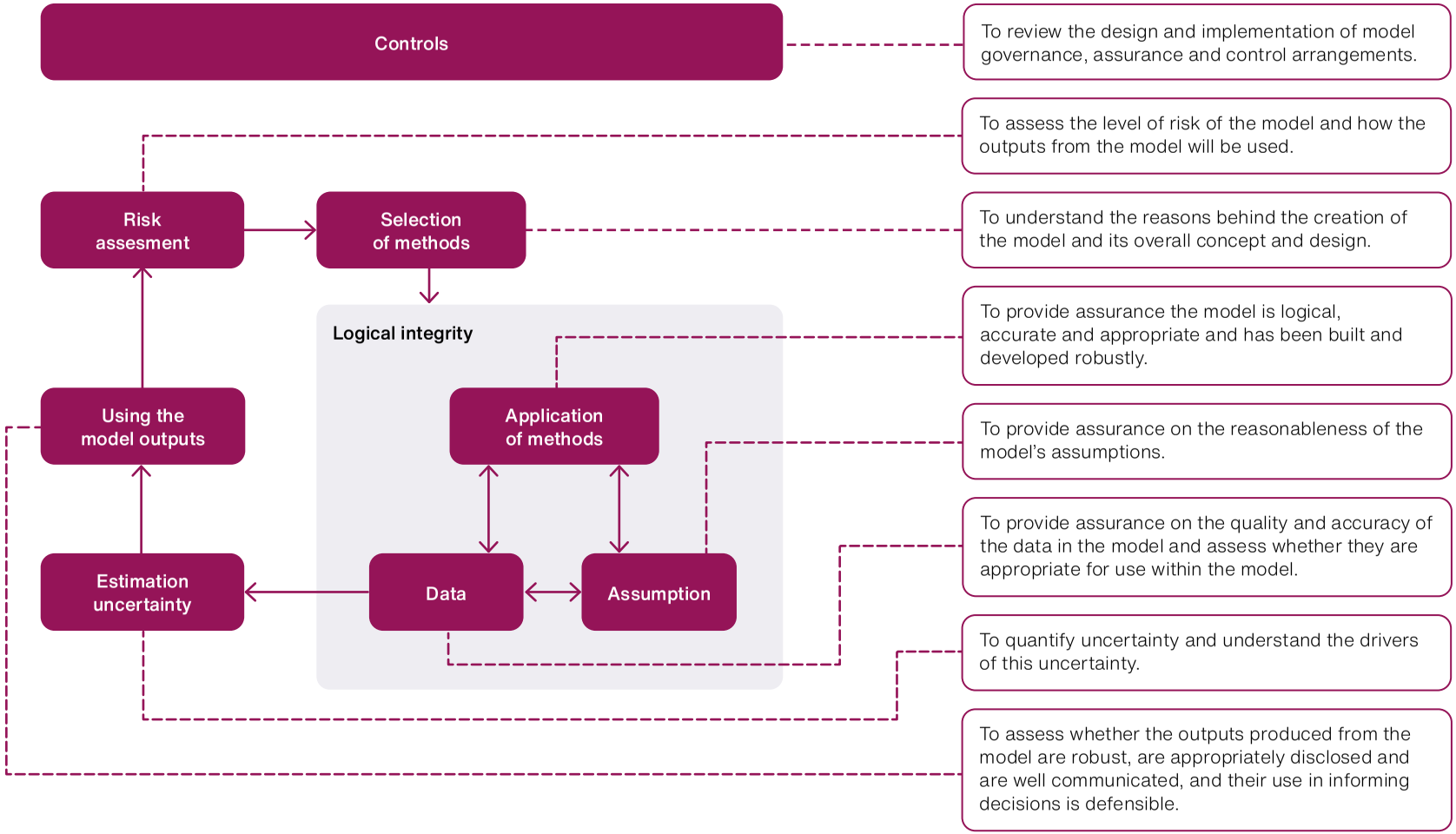
18 We reviewed each modelling document against our overarching audit questions. The review was also used to refine the scope of our work, including defining our audit questions and methods.

Model review

19 Our review drew on our experience of auditing many models across different parts of government in recent years. To conduct our assessment, we adapted our *Framework to review models* to create a bespoke framework to review NHSE's modelling. This framework divided our review into eight areas including model controls, application of methods, and model outputs (**Figure 12**).

Figure 12
National Audit Office's *Framework to review models*

We used our *Framework to review models* to assess NHS England's modelling pipeline for the Long Term Workforce Plan



Source: National Audit Office, *Framework to review models*, January 2022

20 To aid our model review, we also drew on the following good practice guidance:

- the Macpherson *Review of quality assurance of government analytical models*;
- the *Aqua Book: Guidance on producing quality analysis*;
- the *Green Book: appraisal and evaluation in central government*;
- the Department for Business, Energy & Industrial Strategy's *Quality assurance guidance for models*; and
- the government's Analysis Function's *Quality assurance of code for analysis and research*.

21 To verify the outputs of the Python model, we independently ran it and compared our outputs for all scenarios used in the final publication of the LTWP with those produced by NHSE.

22 We completed a comprehensive code review of NHSE's Python model. Our code review focused on assessing the use of appropriate Python code, and whether the selected methodology and assumptions were reasonable and had been applied correctly. To do this, we read through the methodology in the technical documentation and ran the model to check it produced expected outputs (based on the technical guidance) and ran without error. We then carried out a detailed inspection of the code by reviewing and running individual code chunks to check whether the calculations and the resulting outputs were as expected based on the technical documentation and script comments. The code review was documented in Microsoft Word and quality assured by a second reviewer with knowledge of Python.

23 We carried out sensitivity analysis to understand how the model reacts to a change in the input parameters. Due to the large number of input models used to populate the central Python model, we were not able to complete sensitivity analysis on every input model feeding into the Python model. Instead, we conducted sensitivity analysis on the following input models: non-medical supply modelling, medical supply modelling, waiting list interventions modelling, and leaver rate modelling. We adjusted relevant input data from the input models by 1% within the range of ±10% while holding all other variables constant. We found that the model behaved as expected for each sensitivity run.

24 For the triangulation models, we reviewed documentation that we received on the high-level modelling approach and assumptions, and interviewed the analysts who carried out the triangulation. The analysts used an iterative process of adjustments, and told us that, due to the frequency and volume of changes requested by their policy colleagues, they did not document their decisions throughout the triangulation process. 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 inadequate technical documentation and the lack of manual adjustment logs. At the end of our fieldwork, we reviewed NHSE's additional documentation of the triangulation process and attended a demonstration of NHSE's attempt to replicate the outputs from the triangulation process. This was carried out by analysts independent of the original triangulation modelling that informed the LTWP.

25 We reviewed NHSE's quality assurance processes in place for the input data and assumptions that feed into the Python model. We did this by assessing whether the input models had been produced and assured in line with NHSE's internal guidance for models.

26 We did not conduct model review of the input data as this was out of scope for our assessment. We conducted checks to ensure that the input data and assumptions listed in the model documentation and version control logs were present in, and processed correctly by, the Python model.

27 We reviewed model control documentation using a red-amber-green (RAG) rating to assess the quality and appropriateness of model control documentation. We compared NHSE documentation to guidance included in the NAO's *Framework to review models*, and good practice documents. We requested and, where available, reviewed the following documentation:

- specified roles and responsibilities such as the appointment of appropriate and named senior responsible owner and analytical assurer;
- concept and design of the model;
- technical guide for users;
- version control log;
- data and assumptions book;
- analytical assurance plan;
- analytical assurance log;
- model output reports;
- proof of SRO sign-off; and
- succession plan.

Appendix Two

Input models for the NHS Long Term Workforce Plan modelling

1 See **Figure 13**.

Figure 13
Input models

Input model name	Description
Additional roles reimbursement scheme (ARRS) and nursing intervention model	This model assesses the impact of increasing primary care ARRS roles on demand for general practice services.
Ambulance response time model	This model describes ambulance resource requirements and forecast response times.
Community diagnostic centres (CDCs) business case modelling	This workforce modelling is to support the business case for CDCs. It takes account of the capacity required to recover diagnostic waiting time performance.
Community intermediate care model	This model estimates the number of care packages an intermediate care service could provide each year and the workforce requirement to deliver the care packages.
Community pharmacy workforce model	This model estimates community pharmacy workforce demand and supply from 2021-22 to 2036-37.
Community virtual wards workforce model	This model projects virtual wards workforce demand and supply from 2021-22 to 2036-37.
Dental workforce model	This model estimates dental workforce demand and supply from 2021 22 to 2036-37.
Elective activity projections	This model builds on the waiting list interventions tool (below) and extends the projections for elective activity beyond 2024-25.
General practitioner (GP) workforce supply model	Projection of GP workforce in England, including fully qualified, those in training, but excluding GPs in alternative settings.
Leaver rate modelling	Retention interventions that aim to increase workforce supply by reducing the leaver rate. The reduction is modelled at profession level, with each profession having its own "target" rate.
Medical supply modelling – consultants	Counterfactual supply of 'domestic' consultants projected from training supply estimates, average full-time-equivalents of new consultants and observed trends for leavers.
Medical supply modelling – other doctors	Counterfactual supply of doctors other than consultants using assumptions on post-foundation year 2 trainees and international recruitment.
Mental health implementation plan model	This model projects 10-year profiles for the mental health workforce and sets out demand.
Non-medical supply modelling	Counterfactual supply of staff using observed trends on leavers, inflow from the wider labour market, international recruitment, and new supply from education and training.
Temp staffing reallocation model	Temporary staffing data including the number of bank and agency staff procured by providers during the year.
Urgent and emergency care 10-year bed modelling	This model allows users to forecast general and acute bed requirements at national and regional level, based on long-term activity and length-of-stay growth assumptions.
Waiting list interventions tool model	This model is used by NHSE's elective recovery team to model NHSE commitments on waiting list recovery and estimates activity levels required under various scenarios up to 2024-25.
Workforce-activity model (patient level information and costing system ratios)	This model estimates full-time-equivalents for activity in the acute sector.

Source: National Audit Office analysis of NHS England internal documents

Appendix Three

Recommendations

Summary recommendations

a **Modelling pipeline:** NHS England (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.

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.

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

Recommendation 1: 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.

Recommendation 2: NHSE should improve documentation across the whole of the modelling pipeline to make it easier for future analysts to understand the modelling and any limitations, reproduce model outputs, and update the model when producing future versions.

Recommendation 3: Given NHSE's commitment to update the LTWP at regular intervals, it should assign responsibility for understanding the entire modelling pipeline to a small team and ensure that there is sufficient knowledge of the Python and triangulation models available in the team. NHSE should complete a succession plan for the modelling.

Recommendation 4: NHSE should ensure quality assurance practices take place in a timely manner, so analysts have sufficient time to respond accordingly.

Recommendation 5: NHSE should revisit internal quality assurance arrangements for existing models when they are used for a new purpose, such as the models used to provide input data for the workforce modelling, and ensure independent scrutiny is evidenced accordingly.

Recommendation 6: In the next version of the model, NHSE should communicate outputs for a wider range of plausible future scenarios. NHSE should consider different methods and processes which would allow it to better capture uncertainty in its modelling.

Recommendation 7: Modelling of this kind requires gathering assumptions about the future of the NHS in one place. This presents an opportunity to expose those assumptions widely to scrutiny and challenge, both internally and externally. Assumptions should be generated in transparent and systematic consultation with external stakeholders. NHSE should make more of this opportunity the next time it repeats its modelling.

Recommendation 8: In future versions of the modelling, NHSE should model and communicate a wider range of different scenarios for the demand NHS services may face. In our view, this could include scenarios in which different amounts of total funding are available, and upside and downside scenarios for demand for particular services.

Recommendation 9: The modelling (at 0.8% per annum productivity increase) and the LTWP (which describes 1.5% to 2% per annum productivity increase) present productivity differently. In future, NHSE should more clearly explain both internally and in public how it has calculated different measures of productivity and how these relate to the National Statistics produced by the ONS.

Recommendation 10: Ahead of producing the next version of the models, NHSE should review measures of productivity, and other workforce capacity models, to ensure that it is applying the most appropriate measure for its modelling.

Recommendation 11: NHSE's modellers should consider whether it is reasonable to expect the same rate of increasing productivity from a workforce increasingly composed of newly qualified staff.

Recommendation 12: Future versions of the modelling would be improved by greater consultation and by taking better account of the factors that could limit the expansion of education and training, and of any reductions in service that might result from rapid increases in the amount of on-the-job training.

Recommendation 13: NHSE's modelling should incorporate more realistic assumptions on the minimum level of international recruitment of doctors in future.

Recommendation 14: Given the long time frames required for medical training, NHSE should stress test their assumptions on training places for medical students by running their modelling over a longer time period than that covered by the LTWP, to ensure that assumptions about training places do not create long-term risks of excess supply of doctors.

Recommendation 15: Ahead of the next version of the LTWP, NHSE modellers should consult further with a wide range of stakeholders, including Royal Colleges, healthcare providers and other health bodies about the appropriate substitution rates to use when meeting demand for one type of healthcare worker by employing another and should attempt to understand the real-world workload effects on general practice of changes in roles in primary care. NHSE should also be clear about how any proposed changes in skill mix fit with its wider agenda of progressively moving care out of hospital settings and into the community.

Recommendation 16: NHSE analysts should document and share their long-term assumptions, for each iteration of the LTWP, within NHSE and with HM Treasury and DHSC. NHSE analysts across the organisation should use common assumptions where possible.

This report has been printed on Pro Digital Silk and contains material sourced from responsibly managed and sustainable forests certified in accordance with the FSC (Forest Stewardship Council).

The wood pulp is totally recyclable and acid-free. Our printers also have full ISO 14001 environmental accreditation, which ensures that they have effective procedures in place to manage waste and practices that may affect the environment.

National Audit Office

Design and Production by NAO Communications Team
DP Ref: 013886-001

£10.00
ISBN: 978-1-78604-545-4