Data Quality Statement

NHS Vacancy Statistics England

February 2015 – March 2018,
Provisional Experimental Statistics

Published 26 July 2018
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>3</td>
</tr>
<tr>
<td>Data Sources</td>
<td>3</td>
</tr>
<tr>
<td>NHS Job Data</td>
<td>3</td>
</tr>
<tr>
<td>Electronic Staff Record (ESR) vacancy data</td>
<td>3</td>
</tr>
<tr>
<td>Points to Note</td>
<td>4</td>
</tr>
<tr>
<td>NHS Jobs Data</td>
<td>4</td>
</tr>
<tr>
<td>ESR Data</td>
<td>8</td>
</tr>
<tr>
<td>Assessment of statistics against quality dimensions and principles</td>
<td>9</td>
</tr>
<tr>
<td>Accuracy</td>
<td>9</td>
</tr>
<tr>
<td>Relevance</td>
<td>10</td>
</tr>
<tr>
<td>Comparability and Coherence</td>
<td>10</td>
</tr>
<tr>
<td>Timeliness and Punctuality</td>
<td>11</td>
</tr>
<tr>
<td>Accessibility</td>
<td>11</td>
</tr>
<tr>
<td>Performance cost and Respondent Burden</td>
<td>11</td>
</tr>
<tr>
<td>Confidentiality, Transparency and Security</td>
<td>11</td>
</tr>
</tbody>
</table>
**Introduction**

This document constitutes a background data quality report for the NHS Vacancy Statistics publication.

The NHS Vacancy Statistics is a biannual publication of NHS vacancy statistics created from administrative data, related to published vacancy adverts obtained from NHS Jobs. It also includes a high-level summary of the vacancies advertised in NHS Jobs and those vacant posts recorded in the Electronic Staff Record (ESR).

As with any new data source there are a number of data quality issues which require further investigation as the data is exploited further. As far as possible these have been accounted for in the processing undertaken by NHS Digital, but some issues still remain within the data as published.

**Data Sources**

There are two data sources used in this publication, data extracted from NHS Jobs related to job advertisements and data extracted from ESR related to vacant posts created in the system.

**NHS Job Data**

NHS vacancy statistics have been created from administrative data related to job advertisements published on NHS Jobs, the main recruitment website for the NHS. This information is a proxy data source for NHS vacancies and, therefore, should be treated with caution. A job advertisement can be used to fill one vacancy, multiple vacancies or an ongoing recruitment programme and so it is not possible to accurately state the number of vacancies in a period. Job advertisements can also be placed by NHS sub-contractors and local authorities so not all adverts will be for jobs in the NHS, although NHS Digital have attempted to allow for this in their processing and constrain the adverts to those related to NHS organisations.

Information using NHS Jobs data can be found in Tables 1-15 of the main report. To access to full set of data tables please click [here](#).

**Electronic Staff Record (ESR) vacancy data**

ESR is a payroll and human resources system which, since April 2008, has contained staff records for over 99% of the directly employed NHS staff in England. ESR is designed to support the entire HR process, including recruitment; thus, data related to vacancies may also be captured within ESR. NHS Digital has produced a high-level summary of the vacancies advertised in NHS Jobs and those recorded in ESR (Table 13a), and provided a table showing vacant posts recorded in ESR by NWD Staff Group and HEE Region (Table 13b). Both have been done with the aim of increasing our understanding of how the two data sources relate to each other, and to expand our understanding of how they could be used to complement each other in future analysis.

Results for a selection of organisations were considered in greater detail to identify whether their ESR information corresponded with information from NHS Jobs. There was a large variation in results, with data for some organisations matching exactly,
others partially matching and some not matching at all as no data was recorded in ESR and vice versa. This analysis tallies with the information we have received regarding variation in recruitment practice across the NHS in England and the range of systems which are used to manage and advertise vacancies. As neither ESR nor NHS Jobs is mandated for use by the NHS, separate organisations are free to use as little, or as much of the functionality of each system available to them. This investigation was therefore to give a measure of understanding of the variation in practice of different organisations, rather than to understand any particular feature of the two systems whose data was being compared.

Further investigation is required to better understand the issues and recruitment practices underlying the differences in the full-time equivalent figures for the two data sources. This investigation may allow the inclusion of additional analysis within future publications and may also contribute to potential data quality improvements related to vacancy information for secondary uses across both NHS Jobs and ESR. It also raises the need to consider whether it could be possible to investigate data from other systems which sit between ESR and NHS Jobs and may explain the lack of information for some organisations using ESR. These issues and more are being considered by a Vacancy Information sub-group, which is described in more detail in the Future Developments section of the main report.

There are two Foundation Trusts (Moorfields Eye Hospital NHS Foundation Trust and Chesterfield Royal Hospital NHS Foundation Trust), which do not use the ESR and submit workforce data to NHS Digital via the wMDS (workforce Minimum Dataset). NHS Digital will investigate their vacancy data and compare it against NHS Jobs vacancy data in a future publication.

**Points to Note**

**NHS Jobs Data**

The following points regarding the NHS Jobs data need to be taken into consideration alongside the figures provided in the tables that make up this publication.

1. Despite this publication being called Vacancy Statistics the figures in the tables provided do not represent vacancies, although they relate to vacancies.

This report focuses on data from NHS Jobs, providing an indication of the advertised vacancy FTE, based upon the data relating to vacancy adverts placed on NHS Jobs whose published date falls within the quarter/month specified and is limited to known NHS organisations within the stated regions for published vacancies advertised as being of Fixed Term or Permanent type only. This differs slightly from the first publication, which was based on the number of unique vacancy adverts and did not attempt to account for the FTE advertised. Although the FTE field is completed for all vacancy adverts and did not attempt to account for the FTE advertised. Although the FTE field is completed for all vacancy adverts, there remain some issues with regards to the way in which this field is completed. For example, the inclusion of hours rather than FTE, the use of a default number for rolling adverts and FTE inflation related to system interface issues for reused adverts etc. NHS Digital has sought some direct input from users of the system and corrected those records where it has gained additional information. Based upon this information,
the feedback to the consultation and cross comparison of different fields within the data, it has also been necessary to undertake some additional cleansing of the FTE data as received.

2. Despite the changes to the underlying data and the processing undertaken, as one vacancy advert can be used to fill multiple vacancies it is still not possible to accurately state the number of vacancies in a period. Therefore, the only accurate statement remains that the number of advertised vacancy FTE shows the minimum number of vacancies advertised. Further feedback on this development is very much welcomed to enhance our understanding – both of the usefulness of the figures provided and also in relation to the completing of FTE information related to vacancy adverts within NHS Jobs.

3. As with any administrative data source there are concerns about completeness of the fields; what the available fields relate to; the classification of vacancy Staff Groups, Area of Work and Occupation Codes and consistency of practice at trusts. There are also inconsistencies within the data with some records containing contradicting information. In these cases, the fields were investigated and those we had a greater confidence in were selected.

4. NHS Digital has attempted to update any potentially unrealistic FTE values. Unrealistic FTE values are defined as those with an FTE $\geq 10$ and could be due to hours or sessions being stated instead of the actual FTE. NHS Digital has attempted to reduce overestimating the number of vacancies advertised, by dividing all FTE $\geq 10$ by 37.5, the standard weekly contracted hours. For the ESR data we were able to refine these conditions to be more specific for the Medical and Non-Medical staff groups but for the NHS Jobs data this has not been possible.

5. Where possible NHS Digital has removed test adverts from the NHS Jobs data. However, there may be a small number of test adverts which have not been possible to remove and apparent contradictions between different fields remain within the data for a small number of advertisements where it has not been possible to resolve the apparent conflicts.

6. Vacancies advertised are classified in various ways, with some ambiguity around the staff groups. NHS Jobs uses a set of Staff Groups based on the National Workforce Data Set (NWD) Staff Groups. These are shown in Tables 2a to 8, 9d, 14 and 15. This grouping is also available in ESR, which is the main HR and payment system for the NHS, NWD Staff Group is also shown in table 13b based on data extracted from ESR. This form of categorisation is not used in the majority of standard workforce statistics published by NHS Digital.

   The NWD Staff Group is a mandatory field within NHS Jobs and therefore is complete for all adverts. Although there may be some minor data quality issues, it provides a more complete way of categorising Staff Group for this data than Occupation Codes.

   Another staff grouping is shown in Tables 10 and 11. These are based on Occupation Codes which allows categorisation at several levels of detail. Table 10 shows the main functional grouping, whilst Table 11 is based upon the staff groups used in standard NHS Digital workforce publications to give more detail.
The staff groups presented in Tables 10 and 11 are not directly comparable with the Staff Groups in the other staff group tables in this publication.

NHS Digital was advised that the Occupation Code classification mechanism within NHS Jobs allows for adverts to be created without specifying a particular code, and that the means of inputting a code when creating an advert may contribute to issues with the classification of adverts in this way. This is evidenced in the fact Tables 10 and 11 show there are a significant number of published vacancy adverts where Occupation Code has not been specified. Furthermore, there are a number of vacancy adverts where a value which is not recognised has been included.

7. Occupation Codes are a set of codes used to define and identify the NHS workforce. Further information on Occupation codes can be found here

Where possible NHS Digital has identified and cleaned any invalid Occupation Codes in the NHS Jobs Data. These are Occupation Codes which had not been completed, had only been partially completed or transposed. Previously this was done manually but is now part of the automated checks, which identify these codes in the data and update them using supporting information from the Job Title and Agenda for Change fields. For the latest data (data from 1 October 2017 – 31 March 2018) there were 3,251 invalid codes of which we were able to assign 2,421 with a valid code.

8. The Area of Work information presented in Tables 9a, 9b, 9c, 9d and 15 gives an indication of the areas which the adverts relate to, for example in Accident and Emergency or Child and Adolescent Psychiatry. Further information on Area of Work can be found here

9. Not all trusts will use NHS Jobs in the same way, or even at all in some cases. If they use it to publish a vacancy then the data will show that and the system will record the number of web hits on the advert and the applications submitted through NHS Jobs, although organisations may choose to close an advert early if the number of applications is particularly high.

However, if they do not use NHS Jobs to shortlist applicants then we will not have the full statistics on numbers shortlisted, and if they do not update the site with the details of applicants appointed then we will only have partial data on that as well. In general, the completion of data gets less further along the recruitment process and this is visible in Tables 3-6.

Data presented in tables 4-6 does not represent the total number of applicants, people shortlisted or appointed to vacant posts and should not be interpreted as such. The data is provided as an indication of the information potentially available within NHS Jobs, though variation in behaviour mean this data is incomplete. More information regards the previous misinterpretation of this data is available on the NHS Digital website here.

10. Currently the data does not clearly specify a date of recruitment, so it is not possible to quantify how long it has taken to recruit or whether an advert has failed to recruit. This means that we cannot replicate previously published vacancy statistics which showed a vacancy rate for vacancies outstanding for 3
months or more. In future it may be possible to give an indication of which adverts have been re-opened, although this may include both adverts which had previously failed to lead to a successful recruitment and also successful adverts which had simply been reused.

11. NHS Jobs does not include the standard list of NHS Organisation Codes upon which the majority of NHS Digital workforce publications are based. Whilst the system does contain organisation information relating to those bodies placing adverts and some indication of their region, it is not suitable for secondary analysis directly and NHS Digital has undertaken some processing to assign the adverts standard Organisation Data Service (ODS) organisation codes, names and regions.

Due to the assumptions made in this processing the information has not previously been suitable for analysis below Health Education England (HEE) Region level. Previous improvements in classification and processing allowed the production of tables based on organisation functional cluster groupings in addition to the regional data and also different regional aggregations in response to direct requests, to provide a more granular level of detail. The organisation information in this latest publication has been enhanced further, as the data is better understood and the data is now regarded as fit for publication at organisation level. Table 14 of the publication includes organisation level advertised vacancy information by NWD Staff Group for the latest two quarters added in this publication, expanding the time series to a year. Due to the level of reprocessing which would be required to recreate the existing data, it has not been possible to extend the historic timeseries for this table, though it will grow with the publication in common with the majority of the information in future.

The data continues to be limited to only include English NHS organisations which, as far as possible, reflect the organisations included in the standard quarterly NHS Digital workforce publications, incorporating NHS Trusts, CCGs, Support Organisations and Central Bodies. Some organisations have been excluded, for example Hospices, Local Authorities, Charities, Private Organisations, Universities, Schools and GP Practices. There is only partial coverage in these sectors so analysis on these organisations would be incomplete.

In this publication, ‘Other NHS organisations’ refers to organisations which use one combined organisation record in NHS Jobs, for example a group of local CCGs. Although they are from the same HEE region, so can be included in high level HEE regional figures, it is not possible to break figures down into individual organisations to link to standard NHS Digital organisation reference data to allow them to be shown separately in this analysis.

12. Organisations use NHS Jobs in different ways, or they may not use NHS Jobs at all. Data is only provided for those organisations that have used NHS Jobs over the period to which the data relates and where NHS Digital has been able to match their records to known NHS organisations.

Those organisations which do use NHS Jobs to advertise job vacancies may do so differently for different Staff Groups, for example an organisation may choose to advertise clinical roles through NHS Jobs but administrative roles through another means. These differences in practice are visible in the wide variation in
the figures presented in table 14 and therefore a zero FTE may reflect non-use of
the system for a particular Staff Group, rather than there being no advertised
vacancies for that Staff Group. Similarly, an organisation may be absent because
it has not advertised any vacancies over the time period that the data relates to,
or it may have used a means other than NHS Jobs to advertise any vacancies
that it has had.

NHS Digital will continue to investigate what more can be done with a view to
publishing more information at organisation level in January 2019.

13. This report includes the full time series of advertised vacancy FTE data for the
period of 1 February 2015 to 31 March 2018 inclusive. Future publications are
intended to extend this time series and to ensure consistency, any analysis of this
data should always refer to the latest publication.

14. Due to the time lag associated with the completion of the recruitment process, it
has been necessary to stage the time series in the data presented in relation to
applicant behaviour, as opposed to advertised vacancies. This applies to the
number of shortlisted applicants and appointed posts related to advertised
vacancies (tables 5 and 6) and data for the number of web hits and applicants
(tables 3 and 4).

In this publication we publish the latest six months of data, (i.e. data from 1
October 2018 – 31 March 2018), however for these four tables we will publish
data from 1 July 2017 – 31 December 2017 only. This three-month lag in this
element of the time series data ensures that the majority of the publication is as
timely as possible, but that this information is complete when first published.

**ESR Data**

As with any data source there are a number of data quality issues within the ESR data
which require further investigation as it is exploited further. As far as possible these
have been accounted for in the processing undertaken by NHS Digital, but some issues
still remain within the data as published. For example:

As one opened vacant post can be used to fill multiple vacancies it is not possible to
accurately state the number of vacancies within a specified period. Therefore, the only
accurate statement remains that the number of created vacancy full-time equivalents
shows the minimum number of vacancies. It is not possible to state the precise level of
undercounting, but it is possible to say that it will vary for different staff groups – for
example the undercount for nurses is likely to be greater than for other staff groups
because of a number of issues including the high-level of rolling vacancies linked to
generic posts within ESR used for that staff group.

Although it has not been possible to identify job type information within the data (i.e.
whether a vacancy is for a permanent, fixed term, bank or honorary post etc.), NHS
Digital has been able to make some exclusions around job type based upon different
fields within the data. For example, where possible bank vacancies have been excluded
from results based on records having a Tertiary Area of Work of ‘Bank’, a blank FTE or
an FTE <0.1. It has not been possible to exclude all bank vacancies, so there may be
some which remain in the data. It should also be noted that ESR vacancy related data
cannot be said to be fully constrained to fixed term and permanent job types only.
After further investigation, some rolling vacancies are assigned a default FTE = 99. Although it is not possible to tell how many FTE vacancies will be filled using one vacant post for a rolling vacancy, NHS Digital has assigned an FTE = 1. On two occasions vacant posts had been created where Assignment numbers had been used in place of FTE, these significantly inflated the overall FTE value until accounted for. These have also been replaced with an FTE = 1.

Following feedback from the ESR NHS Central team, we amended the conditions used to update potentially unrealistic FTE values in the ESR data. Unrealistic FTE values are defined as those with an FTE >=10 and could be due to hours or sessions being used instead of the actual FTE. Previously NHS Digital attempted to reduce overestimating the number of vacancies advertised, by dividing all FTE >=10 by 37.5, the standard weekly contracted hours for staff on Agenda for Change terms and conditions. We have refined these conditions to be more specific for the Medical and Non-Medical staff groups. Unrealistic FTE values in the ‘Medical and Dental’ Staff Group were divided by 40, the standard weekly contract hours for this Staff Group. The only exceptions were those with a Job Role of ‘Consultant’, ‘Speciality Doctor’ and ‘Trust Grade’, here unrealistic FTEs were divided by 10 as the data here was more likely to represent the number of sessions per week rather than the FTE. All other unrealistic FTE values were divided by 37.5 as is the standard weekly contracted hours for staff on Agenda for Change terms and conditions, which covers the vast majority of staff in other Staff Groups.

Assessment of statistics against quality dimensions and principles

Accuracy

These statistics are experimental and provisional. They are provided as an indication of the potential of the data available and as a guide to the number of published vacancies advertised and other related information. However, they should be treated with caution, particularly with respect to what they actually refer to and their interpretation. Continued improvements within the completion of the source data and the processing carried out by NHS Digital increases confidence in the accuracy of the data. The increased consistent time series now makes internal comparisons possible and therefore increasing the scope for relative comparisons within the series to be undertaken.

Data presented in tables 4-6 does not represent the total number of applicants, people shortlisted or appointed to vacant posts and should not be interpreted as such. The data is provided as an indication of the information potentially available within NHS Jobs, though variation in behaviour mean this data is incomplete. More information regards the previous misinterpretation of this data is available on the NHS Digital website here.

Consideration will be given to removing the experimental and provisional label for future publications.
Relevance

**NHS Jobs Data** - The statistics exploit developments in the NHS Jobs website, and the administrative data it contains, which allow the production of information to answer some of the frequent requests for information from users relating to recruitment in the NHS.

The content of this publication has been updated further in response to feedback received by NHS Digital. These changes, including producing an updated time series; increased information on a monthly basis; greater regional comparability within the supplementary Power BI visualisation tool; analysis of information related to Tertiary Area of Work and Staff Group; pay scheme/band tables and information at individual organisation level, should increase the relevance of the series to users. A summary of NHS Jobs data related to advertised vacancies and data extracted from ESR related to vacant posts created in the system is also included in this publication.

**ESR vacancy data** - These statistics explore the vacancy data held within ESR and it is hoped that with further investigation and refinement this information can be used to produce an answer to some of the frequent requests for information from users relating to recruitment in the NHS.

So far, a HEE region level investigation comparing information extracted directly from ESR for recently opened vacant posts and corresponding advertisement information extracted from NHS Jobs has been undertaken. For the first time, ESR vacancy related data has also been provided by National Workforce Data Set (NWD) Staff Group and HEE region.

As far as possible the same conditions were applied to each data set including organisation exclusions, over the same time period, though some differences in processing were required as described in the Points to Note section.

For a wider consideration of the data quality issues related to the workforce information extracted from ESR and published by NHS Digital on a monthly basis, please refer to the latest monthly NHS Workforce statistics publication on the NHS Digital website.

Comparability and Coherence

The figures included in this publication are not directly comparable to the first two publications in the NHS Vacancy Statistics series. The first publication differs in that the original methodology was a count of the number of adverts published, whereas all subsequent publications have used advertised full time equivalent (FTE) information provided by NHS Jobs relating to published adverts. Revised organisational information after the second publication made it non-comparable to future publications, but all more recent publications are directly comparable and include the full time series of comparable base data available from NHS Jobs.

Despite these further enhancements it still remains difficult to link figures in this publication to other workforce publications by NHS Digital due to differences in data source, methodology and coverage. The inclusion of a summary of the data from NHS Jobs and related vacancy data extracted from ESR (table 13a) is intended to be part of the process to attempt to bridge this gap. This publication aims to help establish a set of statistics which will become an accepted standard.
Timeliness and Punctuality

The results of the consultation which related to the timeliness and punctuality of the statistics suggested that users would like to see the data produced on a more frequent basis, perhaps monthly or quarterly, although there were also responses that such frequent refreshes of the data were not required. With the feedback in mind, NHS Digital will consider increasing the frequency of the publication in future, provided that the availability of data will allow this change. In the meantime, the publication will remain biannual until it has completed its initial developmental stage to provide sufficient time to allow for developments and improvements to the analysis provided. NHS Digital will work with NHS Jobs to ensure that the timeliness and punctuality of the statistics are maximised, and the frequency of the publication will become fixed as part of an on-going series. The publication has been brought forward by a month to increase timeliness. To ensure data we publish is complete the timeframe for tables 3 – 6 has been staggered to include a three-month time lag.

Accessibility

The feedback function of this publication is to allow all users to understand the data, to provide input into our understanding of the underlying data source and to determine what useful statistics should be produced from it in the future.

This publication includes a Power BI visualisation tool, replacing the Excel Graphing Tool for the first time, which is intended to increase the accessibility and utility of the publication for users and will be developed further in response to feedback received.

Performance cost and Respondent Burden

The availability of this data was part of the enhancement of NHS Jobs service envisioned by the Department of Health and Social Care to provide a proxy for vacancy information from administrative data to avoid any burden on the NHS. Several of the responses to the consultation propose that a direct data collection should be reinstated in order to gain access to the level of data required to supplement the information available from NHS Jobs. If such an approach were taken, full consideration would be given to minimising the burden of the additional data collection and focusing on the information which is available to and needed by those organisations involved in healthcare workforce recruitment. Further consideration of potential additional data sources is included in the Future Developments section.

Confidentiality, Transparency and Security

The standard NHS Digital data security and confidentiality policies have been applied in the production of these statistics, though the information upon which they are based relates to published vacancy adverts and to vacant posts created within ESR, not to individual applicants or staff members and therefore no identifiable data has been used in the production of these statistics.