**Population Estimates Comparison Project Closure Document**

**Executive Summary**

## The Population Estimates Comparison Project was set up in December 2012 to investigate the differences between population estimates in Scotland calculated by NRS (the Census and the Mid-Year Estimates) and estimates derived directly from administrative data sources, such as the CHI, NHSCR, Higher Education Statistics Authority, Local Authority (LA) data, etc.

## A review and comparison of existing datasets by NRS (National Records of Scotland) was carried out and concluded that the MYEs (Mid-Year Estimates) were the most accurate available measure of the population. A final recommendation was given to TAGRA in April 2015 to continue using the NRS population estimates as the most accurate count of the population. .

Work continues within NRS to continuously improve the quality of the MYEs. Much of the work carried out as part of this project is also on-going and has been absorbed within the 2021 Census programme which aims to make best use of technology and administrative data in its design. Approval from the Privacy Advisory Committee, CHI Advisory Group and NHS Caldicott Guardian Forum has been received for NRS to gain access to limited health data indicating signs of activity for use in the production of statistics. A data access agreement has also been signed to allow access to specific Glasgow City Council council tax data. Work is on-going to secure these data and carry out research into how the data might best be used to inform the production of population estimates.

## Further steps are also being taken forward by NSS Practitioner Services Division to continuously improve the quality of the CHI data.

# **Background**

## **Health budget resource allocation**

## The total annual monetary budget for the NHS in Scotland is agreed in the Scottish Government’s Spending Review. The Scottish Government then divides and distributes this budget across the fourteen territorial health boards in Scotland using a resource allocation formula.

## The formula does not calculate the individual need of each health board but uses statistical modelling to decide what percentage share of the overall budget should be assigned to each health board. The aim of this system is to provide a transparent and fair mechanism for the dividing of available resources and to ensure equity amongst health boards in terms of enabling them to meet the real needs of their resident populations.

## Between 1978 and 2000, a relatively simple formula called the SHARE (Scottish Health Authority Resource Equalisation) formula was used to allocate the NHS budget across Scotland. The formula placed strong emphasis on standardised mortality ratios (0-64 years) using these as a proxy for the pull on NHS resource in each health board and allocating resource in proportion to this. This remained untouched for more than 20 years until technological developments enabled a more sophisticated approach to be developed.

## Between 2000 and 2008, the Arbuthnott formula was introduced to calculate the percentage shares of the NHS budget in Scotland, for the first time introducing the concept of adjusting the allocation by health service supply factors. In March 2005, the [NHS Scotland Resource Allocation Committee (NRAC)](http://www.nrac.scot.nhs.uk/) was established with a remit to “improve and refine” the Arbuthnott Formula. The NRAC formula was introduced in 2009-10 and continues to use the same weighted capitation approach as the Arbuthnott formula (i.e. based on the potential number of patients within a catchment area rather than the actual volume of healthcare delivered), but incorporates a number of recommended improvements.

## The NRAC formula is responsible for allocating the funds for hospital and community health services (HCHS) and for General Practitioner (GP) prescribing, which account for approximately 70% of the total NHS Scotland budget. The [Technical Advisory Group on Resource Allocation](Following%20the%20publication%20of%20the%20NRAC%20final%20report%2C%20the%20responsibility%20for%20maintaining%2C%20updating%2C%20and%20refining%20the%20formula%20for%20dividing%20the%20NHS%20budget%20among%20the%20NHS%20Boards%20has%20been%20transferred%20to%20the%20Technical%20Advisory%20Group%20on%20Resource%20Allocation%20%28TAGRA%29.%20Full%20details%20on%20the%20work%20of%20TAGRA%20are%20available%20at%3A%20http%3A/www.tagra.scot.nhs.uk/) (TAGRA) is responsible for overseeing the maintenance and development of the formula. TAGRA provides a platform to review the NRAC formula and ensures that it is kept up-to-date.

## The NRAC formula follows a weighted capitation approach, with the principle objective being to provide equity of access to healthcare services.  Resources are distributed among Health Boards on the basis of the population share, weighted according to the relative needs in terms of; (i) the demographic profile of the population, and (ii) relative levels of morbidity and mortality. The formula uses costed utilization as a proxy for need in determining these adjustments. The formula also includes a weighting to account for the relative need for resources in each Board according to the unavoidable excess costs of service delivery. The Acute Morbidity Life Circumstance (MLC) adjustment within the formula is presently subject to on-going review by a Subgroup of TAGRA.

## **Use of population data**

## The population data used in the NRAC formula are the Mid-Year Estimates (MYEs) (at both small area and Health Board level) and the population projections (at Health Board level), both of which are produced by the NRS. The MYEs are produced using the demographic cohort component method using the latest Census as their base. The way the method works is that each year the population is ‘aged on’ one year (i.e. all 1 year olds become 2 year olds, and so on), the number of births in the year are added, the number of deaths subtracted and adjustments are made for estimated migration and other changes in special populations. Given that NRAC is a weighted capitation formula, the Health Board population shares directly underlie the final target shares. Given the size of the overall baseline budget, a 0.1% shift in NRAC target share would equate to a c. £8m shift in funding (were all Health Boards actual allocations to be at parity1 with their target shares).

We have a definitive baseline in the Census to work from and definitive counts of births and deaths but rely on various sources to estimate migration as there is no definitive way of counting migration both within Scotland and in and out of Scotland.

## An alternative approach for the NRAC formula would be to use live sources that give counts of health service users on an ongoing basis between Censuses. Candidates to provide this information in this way are the Community Health Index and the NHSCR. The Community Health Index (CHI) is a register of NHS patients in Scotland that exist to ensure patients can be correctly identified and that information pertaining to a patient’s health is available to all providers of care. Each patient is assigned a unique ten-digit CHI number, which helps to preserve confidentiality by reducing the use of patient identifying information. The NHSCR is maintained by NRS and combines birth and death registration data with data drawn from the CHI system. . While both of these source theoretically have sufficient coverage to produce population counts taken independently neither produces counts consistent with the MYEs. The key purpose of the project was to attempt to identify the reason for the differences to identify the best source for the population counts used in the NRAC formula.

## **Discrepancy between CHI and Mid Year Estimate population counts**

## The CHI provides an additional count of the resident population in Scotland, although the count as estimated by the CHI is consistently larger than the MYE for the same year for all health boards. However, a particularly large disparity has been noted between the population estimates taken from the CHI and the MYEs for the NHS GG&C. This issue was identified as a potential concern by NHS GG&C, particularly as the final share of funding is greatly influenced by the crude population share.

## It is generally recognised that the CHI suffers from over-inflation as a result of difficulties in identifying whether a new patient is registered elsewhere and patients not deregistering when they move. . It is possible that the CHI records for NHS GG&C suffer a greater proportion of these problems in comparison to the remaining 13 health boards. However, NHS GG&C expressed concern that this over-inflation between the CHI and the MYEs could be complicated by people not being included in the Census (which acts as the base for the MYEs) but who are health service users. As a result, understanding the reason(s) behind the differences between the MYEs population estimate and the CHI population estimates was of significant interest to both NHS GG&C and TAGRA. There were several lines of inquiry that NHS GG&C had been pursuing. This project complimented these lines of inquiry.

## NHS GG&C subsequently asked NRS to lead an investigation into the differences between the population estimates for the CHI and MYE. By comparing only aggregate population estimates from these two datasets, it was not possible to say with any certainty which source was the most accurate. It was possible that the CHI was over-counting the population or the MYE was undercounting. In order to come to a conclusion, it was proposed that a review and comparison of existing datasets, such as MYEs, Census, CHI and the NHS Central Register (NHSCR), be carried out and a data-linkage exercise of council administrative data sources be conducted in order to create a population estimate through record linkage.

# **Project Initiation**

## This project was set up to investigate the possible disparity in population estimates between different data sources across Scotland, with a view to understanding the differences and highlighting any potential implications for NRAC formula. As such, the project focused primarily on the population of NHS GG&C which had the highest degree of apparent over-inflation of the CHI. However, a wide range of other learning opportunities were envisaged with the potential for any relevant findings to be applied across Scotland as a whole.

## This project also acted as a pathfinder for the comparison of population estimates taken from different administrative data sources. This was an important aspect of the Beyond 2011 programme which was concerned with the future provision of population and socio-demographic information.

## **Objectives**

## The primary objectives of this project were:

## to provide a better understanding of the differences between population estimates in Scotland calculated by NRS (the Census and Mid-Year Estimates) and estimates derived from administrative data sources, such as the CHI, NHSCR, Local Authority (LA) data, etc.

## to review and document available data sources and assess whether the data quality of administrative data sources could be improved.

## to establish the suitability and accuracy of relevant data sources and determine whether it was possible to use them to contribute to population statistics.

## to assess whether the population of NHS service users in GG&C Health Board area was significantly different to the population estimates produced by NRS, particularly given the significance of the MYE in determining the allocation of the NHS resources.

## to identify any potential disparities in the population estimates for particular geographical areas or subsets of the population, with the intention of documenting, explaining and assessing the implications of these differences where possible.

## to report on any barriers to data sharing that were identified and to document the steps taken to reduce or alleviate these barriers and to inform the development of the population spine. This would assist the work that was taking place in the Beyond 2011 programme and the Data Sharing & Linkage Service (DSLS).

## **Governance**

A project board was set up with representation from NRS, NHS GG&C HB, Glasgow Council, SG Health Analytical services, NSS Information Services Division, NSS Practitioner Services Division and the Improvement Service.

It was acknowledged that in order that NRS carry out any statistical linkage of data to inform the project, rigorous safeguards needed to be in place to protect personal information. Formal data sharing agreements had already been developed for use of CHI and HESA data for estimation of population and migration purposes . [A report on privacy issues](http://www.nrscotland.gov.uk/files/nhscr/privacy-pop-estimates-report.pdf) and safeguards to protect personal information in relation to the transfer and use of data from the NHSCR in generating population and migration estimates had previously been commissioned by the [NHSCR Governance Board](http://www.nrscotland.gov.uk/statistics-and-data/nhs-central-register/consultations-and-meetings). The NHSCR Governance Board were also kept informed of linkages of the NHSCR and Census records. Privacy issues in relation to statistical linkage of datasets were also considered regularly by the NHSCR extract working group on which the NHS Caldicott Guardian and the Information Commissioner’s Office were represented.

## **Results of analyses**

### **Comparisons between Mid Year Estimates and CHI population counts**

 One of the initial objectives was to gain a better understanding of the differences between population estimates in Scotland calculated by NRS (Census and Mid-year estimates) and estimates derived from administrative data sources, such as the CHI, NHSCR and Local Authority data. The chart below shows a population pyramid for 2011 giving a count of males in Glasgow and Edinburgh from the census and from the CHI currently registered population (CRP).



The difference between the CHI currently registered population (CRP) and census & MYE are more pronounced for males in Glasgow than in other parts of Scotland. These differences are what led to the creation of the Population Estimates Comparison Project. The large differences between the Census and CRP are more pronounced in urban areas and deprived areas. Analysing the sex ratios suggested that these differences were likely to be driven, in part, by list inflation in the CRP and NHSCR as there was an unrealistically high ratio of males to females in the CRP..

### **Comparisons between NHSCR and Census**

As part of the Population Estimates Comparison Project, NRS examined some patterns and relationships in detail between the NHSCR and the Census for a relatively small area in Glasgow in 2011.

For the purposes of the research, detailed analysis of records from the NHSCR and the 2011 Census were conducted and the results used to obtain a picture of the patterns of over and under count. This allowed investigation of the list inflation which is known to exist in the area.

There are many different ways in which list inflation can occur, but the main source is when persons incorrectly, and indefinitely in some cases, remain on the GP List whilst having moved elsewhere. Many students, when they complete their courses of study, will return to their original domicile or will move elsewhere in Scotland, the rest of the UK or abroad.

Analyses showed that the overall number of records for males in the NHSCR which were not linked to the Census was substantially higher than those of females. Furthermore the shape and extent of the age distribution was substantially different. The analyses seemed to indicate that there was little difference in the patterns between males and females who were in the Census but not in the NHSCR. However for those records which were in the NHSCR but not the census, the NHSCR list inflation seems to have a much shorter “age range” and “scale” for females than males. This may indicate that the problem of list inflation for females is “corrected” much earlier, and that this correction must logically be caused by females using health services (re-registering) in Scotland or elsewhere in the UK. For males, it is possible that the large excess extending to 40 - 50 year old males may be the result of a cumulative pattern (over a period of 20 to 30 years) of retaining records which are no longer “live” - there is currently no formal mechanism for removing such inactive records. Males using health services in other areas of Scotland reasonably soon after attending university (and having registered with a GP in the area where their university is), should be reflected as valid re-registrations – however, there may be residual problems with students returning home after their studies but not de-registering.

The evidence showed that students and young people are being successively added to the GP list each year. However, there appeared to be a lag between people leaving the area and their de-registration with the local health services with an accumulation building up over the last 20-25 years. The evidence also suggested that there could also be a discontinuity between de-registration from Scottish lists and re-registration to lists elsewhere in the UK .

We will be able to investigate this further once we have gained access to the “activity flag” data, when we will be able to observe patterns of recent interactions with health services.

### **Comparisons between NHSCR and HESA**

During the course of the project various different sources of data were used. Some analyses examined the differences between the Census and NHSCR; other analyses examined the Higher Education Statistics Agency (HESA) data and the NHSCR.

The initial part of this task was to identify students who were in both the HESA dataset and the NHSCR dataset. This part was conducted on the basis that the combination of forename, surname and date of birth uniquely identified students. The table below shows the HESA data and match rates with NHSCR. It should be remembered around 6% of the population of Scotland do not respond to the Census and that this figure is higher for student groups. However, the Census operation involves complex processes to estimate the number of missing individuals to ensure the resultant population estimates are a true reflection of Scotland’s population.

|  |  |  |  |
| --- | --- | --- | --- |
| Institution | All students in HESA | Students found in NHSCR  | Percentage of students found in NHSCR |
| Glasgow Caledonian | 10,370 | 8,985 | 86.6% |
| Glasgow School of Art | 1,349 | 1,141 | 84.6% |
| Royal Conservatoire | 770 | 645 | 83.8% |
| University of Glasgow | 18,879 | 16,393 | 86.8% |
| University of Strathclyde | 14,973 | 13,021 | 87.0% |
| **Greater Glasgow Health Board area** | **46,341** | **40,185** | **86.7%** |
|  |  |  |  |
| **All students (Scotland)** | **156,492** | **131,178** | **83.8%** |

Further work was done analysing the data by the postcode of the student in GG&C HB and various other Council areas. The conclusion from this work was that there appeared to be limited scope for using HESA information to accurately update location at postcode level. However, by analysing HESA data at Council area level, we may be able to generate an approximate location for students.

### .

# **Summary of project**

## **Deliverables**

Various papers were produced and published on the web (see Annex A). Other analyses is on-going as part of the continuous improvement of population estimates.

## **Benefits realised**

During the closure of the PECP a questionnaire was circulated to members of the group asking for views on how well the project delivered on its objectives.

All members of the group felt that the PECP project had delivered numerous benefits including;

* Joint working with others in the public sector. The initial focus of the project meant that several organisations were directly involved. Good contacts were made which will continue beyond the scope of the project. It gave all parties a better understanding of the processes and outcomes of population estimation in relation to other data sources.
* Wider understanding of the various data being used and a wider appreciation of the challenges in arriving at accurate numbers.
* Confidence in the existing structure while informing interested parties in areas for future focus and development.
* A much clearer understanding of the extent, and some further understanding of the origins, of discrepancies between population products (notably between the MYE and CHI).
* Reassurance provided to NHS GG&C, through their involvement in the project, regarding the use of MYE population estimates within the NRAC formula.
* Renewed confidence in NRS’s processes to calculate population estimates across Scotland.
* Development of processes for facilitating access to administrative data sets e.g. development of data access agreements, fair processing notices etc.

## **Lessons learnt**

The following were felt to be some of the lessons learnt :-

* The importance and value of close working with stakeholders (NHS GG&C, Health ASD, ISD etc) was emphasised.
* The length of time required to gain approval and funding for particular strands of work (e.g. ISD signs of activity) should not be underestimated
* In developing data access agreements, it should be ensured that fair processing notices are developed and disseminated as early as possible.
* The length of time required to negotiate a data access agreement should not be underestimated
* Invaluable insights are given through case studies on the ground (e.g. NHS GG&C staff liaising with local GP practices)
* Much of the work involved in specifying and bringing together the data required, resulted in valuable capacity building in the partner organisations enabling them to carry out further research

## **Further work**

It is clear that there is still further work that could be undertaken. For example;

1. Given the nature of the conclusion arrived at by the PECP, there would seem to be an implicit action for further work to be undertaken to clean and improve the CHI database. Whilst the CHI populations are not used as the basis of resource allocation in NRAC, they are used as the basis of Global Sum funding allocations made to primary care providers (mainly GP practices). Ownership of the CHI maintenance and improvement work would seem to fall to PSD. However, it would also seem important that PSD are supported by other PECP partners in utilising the learning from the PECP, especially in terms of understanding the nature and origins of the discrepancies between the CHI and Mid-year estimates, in order to ensure that such work is progressed as effectively as possible.
2. Replicating the case study which was carried out by a member of GG&C HB to examine whether records in a GP practice were live or not, could be undertaken in a sample of other areas to better inform our understanding of regional differences. The work could be carried out by the health boards in a similar manner to the case study in this project.
3. Further analyses of the differences and the complex relationship between health related data (CHI and NHSCR) and Census based population estimates could be taken forward. The work involved in this kind of project is extremely important, and should be continued in some form or another. NRS rely on several different sources of data to enable the production of population and demographic data that are accurate, reliable and timely, and a good understanding of the NHSCR and CHI data is crucial for this reason.

The overall consensus of the members of the PECP was that the project was very worthwhile and highlighted some interesting and positive outcomes in areas they were not expecting. Although not the outcome that some members anticipated, the conclusion of the project confirmed that NRS’s methodology was robust and the best available at present.

# NRS

November 2015

# **Annex A**

This section contains links to all papers published with regards to the Population Estimates Comparison Project.

Papers relating to the TAGRA meetings - <http://www.tagra.scot.nhs.uk/meeting-papers/>

13th meeting – 17th April 2012:

 [TAGRA – 13th meeting – 17 April 2012 – paper TAGRA(2012)03 – Population work](http://www.tagra.scot.nhs.uk/wp-content/uploads/docs/meeting%20papers/meeting13/TAGRA%20-%2013th%20meeting%20-%2017%20April%202012%20-%20paper%20TAGRA%282012%2903%20-%20Population%20work.doc)

16th meeting – 16th April 2013:

 [TAGRA – 16th Meeting – 16 April 2013 – TAGRA(2013)03 – Population estimates update](http://www.tagra.scot.nhs.uk/wp-content/uploads/docs/meeting%20papers/meeting16/TAGRA%20-%2016th%20Meeting%20-%2016%20April%202013%20-%20TAGRA%282013%2903%20Population%20estimates.docx)

17th meeting – 6th August 2013:

 [TAGRA – 17th Meeting – 6 August 2013 – TAGRA(2013)14 – Population estimates](http://www.tagra.scot.nhs.uk/wp-content/uploads/docs/meeting%20papers/meeting17/TAGRA-%2017th%20Meeting%20-%206%20August%202013%20-%20TAGRA%282013%2914%20Population%20estimates.docx)

18th meeting – 11th December 2013:

 [TAGRA – 18th Meeting – 11 December 2013 – TAGRA(2013)11 – Population estimates comparisons project](http://www.tagra.scot.nhs.uk/wp-content/uploads/docs/meeting%20papers/meeting18/TAGRA%20-%2018th%20Meeting%20-%2011%20December%202013%20-%20TAGRA%282013%2911%20-%20Population%20estimates%20comparisons%20project.doc)

21st meeting – 15th December 2014:

 [TAGRA – 21st Meeting – 15 December 2014 – Paper for information – Population Project Update](http://www.tagra.scot.nhs.uk/wp-content/uploads/2014/03/TAGRA-%E2%80%93-21st-Meeting-%E2%80%93-15-December-2014-%E2%80%93-Paper-for-information-Population-Project-Update.doc)

22nd meeting – 30th April 2015:

 [TAGRA – 22nd meeting – 30 April 2015 – TAGRA(2015)07 – Population Estimates Comparison Project](http://www.tagra.scot.nhs.uk/wp-content/uploads/2014/03/TAGRA-22nd-meeting-30-April-2015-TAGRA201507-Population-Estimates-Comparison-Project.doc)

Papers available on the NRS website;

[Population Estimates Comparison Project: Data Sources Review](http://www.nrscotland.gov.uk/files/census/2021-census/reports-publications/pop-est-proj-data-review.pdf) – This report documents and reviews the data sources that have contributed to the work of the National Records of Scotland’s (NRS) Population Estimates Comparison Project.

Other relevant papers - <http://www.nrscotland.gov.uk/statistics-and-data/beyond-2011-the-future-of-census-statistics/reports-and-research>

## Barriers to data sharing – (<http://www.nrscotland.gov.uk/files//dsls/pop-estimates-comparison-data-sharing.pdf>)