

<b>POPULATION AND MIGRATION STATISTICS (PAMS) COMMITTEE (SCOTLAND)</b>
--

**Migration Estimate Adjustment Using Record-Level Linkage of HESA and NHSCR Data****Background and overview**

1. Each UK constituent country produces the best possible estimate of the population using the best administrative data sources and methods available. The statistical offices responsible for producing these estimates are continually researching ways of ensuring and improving their quality.
2. Scotland is the only UK constituent country that does not incorporate administrative data on student moves from the Higher Education Statistics Agency (HESA) to account for known issues with GP registration rates among the student population. In England and Wales, and Northern Ireland, HESA data is used to adjust for student moves in the estimation of internal migration, but not for moves between UK constituent countries and to countries overseas.
3. A previous study by National Records of Scotland (NRS) in 2010 investigated the possibility of a migration estimate adjustment for Scotland using HESA data. It was not considered possible for NRS to implement the pre-2012 ONS student adjustment methodology at that time because the necessary record-level data was not available, and analysis indicated that the aggregate method would result in a relatively small adjustment.
4. This paper reports on the improvements in accuracy that record-level linkage of HESA and NHSCR data could provide to current student-age internal migration estimates in Scotland. This work was conducted by Douglas Clark, a student from the University of Edinburgh on placement in the Population and Migration Statistics team.
5. Since mid-2012, the Office for National Statistics (ONS) has compared consecutive years of data from the Patient Register linked with HESA data on date of birth, sex and output area of domicile address. This allows ONS to allocate students to their most recently recorded address (in either the HESA or Patient Register data). All individuals whose local authority (LA) differs from the previous year are identified as migrants. If a student does not appear with an updated GP registration record, the student is estimated to remain at their HESA term-time LA. A proportion of these lapsed students are reallocated to the LA of their most recent GP practice record. This process is repeated each year until all lapsed students are re-allocated.
6. The Northern Ireland Statistics and Research Agency (NISRA) creates a student adjustment by comparing the number of moves recorded by HESA between domicile address and term-time address at the start of study by age and gender to the moves recorded by the medical card register. The higher of the two measurements is then used to create the adjustment.
7. Analysing end-of-study moves requires a time-series of HESA data matched to NHSCR data at the record level. Currently, NRS only has access to 2011 HESA data, although negotiations are on-going to access extracts from other years.

Since a matched time-series is not available this paper can only consider start-of-study moves. Additionally, differences in the higher education system in Scotland such as starting ages, course lengths, and thus the pattern of student moves make analysis of post-study movement factors more complicated. These differences also might mean that the ONS method is not suitable for application in Scotland. For these reasons, this paper also considers how linked HESA-NHSCR data might affect population flows calculated using the NISRA method.

### **Linked Dataset**

8. This paper uses a dataset containing 2011 HESA and NHSCR data linked at the record level. This dataset captures 141,000 (47%) of the 302,000 students in the 2010-11 raw HESA extract. This is partly explained by the fact that the linked dataset is limited to full-time students aged 17-30, while the full extract (in which 57% are listed as studying full-time) is not restricted by age or mode of study. The captured proportion is approximately consistent across both the male and female populations.
9. However, the proportion of students captured by the linking process (completed before this research) is not uniformly distributed among universities. Typically, universities in more urban areas are better represented than those in more rural areas. For example, while 76% of all students studying at Edinburgh Napier University and 62% of all students at the University of Glasgow are captured in the linked dataset, only 10% of all students at the University of the Highlands and Islands and 5% of those at the Scottish Agricultural College (now Scotland's Rural College) are represented. These differences are also likely a result of differences in the method and rigour with which data is collected by individual universities on HESA's behalf.
10. Analysis also indicates that the Fife Campus of the University of Dundee (School of Nursing and Health Science) is not recorded in the HESA dataset as a separate campus. The reason for this cannot be definitively ascertained from the data, although it could be that HESA records the campus as part of the "University of Dundee: Main Campus". This will likely impact estimates of the student population in Fife, and have knock-on effects for start-of-study flows between Dundee and Fife. It could also be that other universities have campuses which are not distinctly recorded in the data.

### **Postcode analysis**

11. The linked dataset contains three postcodes which were used in the analysis:
  - a. HESA-listed term-time postcode,
  - b. HESA-listed domicile postcode,
  - c. NHSCR-listed postcode.
12. The failure of many students to re-register with a GP when they move to begin their studies is shown by the fact that while only 18% of first years are listed as staying in their Parental/Guardian home during term time, 58% have a HESA-listed domicile postcode which matches their NHSCR postcode. 59% of first year students have a term-time postcode which matches their NHSCR postcode. This not only includes students who have stayed at home and not needed to re-register with a GP, but also students who have moved away and neither listed

their term-time address accurately nor re-registered with a GP, as well as students who have done both correctly.

13. It is also possible to analyse the distance between the three postcodes listed in the linked dataset and the postcode of the campus at which the student is listed as studying. Such analysis indicates that HESA-listed term-time postcodes are more likely accurate than NHSCR postcodes: 85% of HESA-listed term-time postcodes for 1<sup>st</sup> years are situated within 20 km of the student's listed campus, compared to only 77% of NHSCR postcodes.
14. Many students also move address within the same city during their time at university. Using linked HESA data and distance analysis allows more of these moves to be captured in the small area population estimates, which might be particularly useful for tracking the moves of students studying at institutions situated near council boundaries.

### **Communal Establishments**

15. The potential for linked HESA-NHSCR data to increase the accuracy of population flows is further shown by distance analysis of the postcodes of students listed by HESA as living in Communal Establishments (CEs) i.e. student halls. This analysis is based on the assumption that the term-time postcodes of students living in halls will be geographically close to their listed university campus. Indeed, 95% of the HESA term-time postcodes of all students listed as living in halls are situated within 5 km of the student's university campus. In contrast, only 78% of the NHSCR postcodes of students living in halls are situated within 5 km of the student's university campus.
16. Furthermore, the availability of a Communal Establishments (CEs) look-up file means that it is possible to measure the accuracy of term-time postcodes of students living in halls by examining the rate at which they match the postcodes of existing CEs. This allows verification of the accuracy of the HESA term-time postcodes in a way which is not possible for other accommodation types for which a similar look-up file does not exist.
17. 69% of students living in halls had a HESA term-time postcode which matched a CE postcode, while only 36% had an NHSCR postcode which matched. Overall, 81% of students living in halls had at least one out of the three postcodes in the linked dataset match a CE postcode. Given that CE postcodes only apply to students during term time, this indicates that the linked dataset offers a higher rate of plausibly accurate term-time postcodes for use in measuring migration flows.

### **Flows**

11. Using the NISRA method with the linked dataset, it is only possible to generate an estimation of start-of-study student moves. Linked data can track moves following the completion of university through GP registrations, but without a time-series it is very difficult to make an accurate end-of-study adjustment. In any case, estimating start-of study moves using HESA data still gives a useful illustration of the scale of student moves that are missed by estimations which rely on NHSCR data only.

12. The flows of students moving from one council area to another at the start of study are measured using the linked HESA-NHSCR dataset, and then compared to the flows measured in the 2011 Mid-Year Population Estimates (MYEs), which rely on GP registration data to record student moves. The gross number of student in-flows was 17,900 higher in the linked dataset, meaning that at least 17% of the total within-Scotland moves are missing from the MYEs. This disparity is greatest for areas where the largest student populations are concentrated i.e. city LA areas and Fife. The highest percentage of missing moves in the MYE data was found in Aberdeen City, where 38% of within-Scotland moves are unrecorded in the patient register data compared with the combined patient register and HESA data. Out-flows are more evenly distributed throughout all council areas, with the highest unrecorded proportion of within-Scotland outflows being in Orkney Islands (41%).
13. These flows are likely to still undercount the true number of young people moving from one LA area to another. For example, if the estimated flows from Council A to Council B are higher using HESA data, the NHSCR estimate would be disregarded. However, some migrants (e.g. young people moving for work and not to study) would be captured in the NHSCR estimate but not the HESA estimate, and therefore would not be counted. Thus, the student-age population would still be undercounted, but far less so than it currently is.

#### **Overall assessment and recommendation**

14. The use of linked HESA-NHSCR data would help NRS mitigate the inaccuracies in term-time postcodes presented by the organization's current reliance on NHSCR data. This is particularly true of special populations resident in communal establishments where the accuracy of term-time postcodes can be verified.
15. Crucially, HESA data can improve the method of estimating start-of-study population flows, which, as the above analysis suggests, requires a sizeable adjustment to account for currently missing LA flows. While new population estimates incorporating HESA data would still undercount the student-age population, the estimates would be undercounted far less than they currently are. The above analysis concerns only start-of-study moves, and further work is needed to account for end-of-study moves before any overall adjustment can be made using HESA data.
16. It is therefore recommended that NRS acquires a time-series of HESA data matched to NHSCR data at record level so that end-of-study moves can be estimated. No adjustment for student migration can be made to the MYEs until end-of-study moves are accounted for in this way. It is also recommended that NRS builds a strong business case to justify the release of this data.
17. PAMS are asked to endorse these recommendations, and to recognise the value of research to improve the MYEs. Comments should be sent to Denise Patrick ([Denise.Patrick@nrscotland.gov.uk](mailto:Denise.Patrick@nrscotland.gov.uk)).