

7 Refining the methods

In the course of undertaking this study, a number of areas where improvements in data availability or structure could enhance the utility and accuracy of analyses were identified. These are discussed below.

1. Data items included in the analysis

When obtaining permissions to use the NHMD and ACCMIS data sets in data linkage studies, the specific data items to be included have to be specified. For this project, a parsimonious approach was taken to selecting data items for inclusion in the analysis. Since analysis of linked NHMD and RAC data had not previously been carried out a number of data items were omitted which, in hindsight, could have enhanced the analysis. For example, in the hospital data only hospital sector (public versus private) was included in the analysis data set whereas analysis by sector of both the hospital and the patient (that is, patient election status) is of interest. Also, in the RAC data, living arrangements prior to admission could have proved useful in the analysis of propensity to move from hospital into permanent RAC.

The experience from this study therefore suggests that a more inclusive approach should be taken when specifying data items to be included in analysis data sets from the NHMD and ACCMIS. Such an approach would make the linked data set useful for a wider range of investigations.

2. History of use of RAC

In the current analysis, data on the use of RAC was restricted to the year of interest (2001–02). However, information on prior and later use of RAC would both improve data available for modelling behaviour and improve data quality for looking at short-term and medium-term outcomes. For example, having data on later use of RAC would allow a full year's data to be used when investigating short-term use of RAC (Section 5.2), and would reduce the level of censoring in the data for the survival analysis (Section 5.3).

3. Mortality data

In analysis of person outcomes, knowledge of whether a person has died can be important. For example, in the analysis of short-term use of RAC presented in Section 5.2, a number of people 'Left reported going to RAC' but were not identified as later returning to RAC. Knowing whether or not these people had died would add a further dimension to the story. In addition, in the analysis of survival following admission into permanent RAC, better death information – in particular, for those who were discharged to hospital – would increase the quality of the data for analysis (by reducing the level of censored data) and so would lead to more reliable models.

Better identification of death could be achieved by linking the National Death Index (NDI) with the RAC data on ACCMIS. As name and date of birth information are on both data sets the linkage would be name based. As well as identifying date of death, such linkage would allow cause of death data to be included in the analysis data set. Note, that such name-based linkage could only be undertaken for people using RAC as the NHMD does not contain name information. While, it may be feasible to link deaths in hospital with the NDI using event-based linkage, date of death and reason of death data could not be obtained for people

who returned to the community after their period in hospital as there are insufficient data to link the NHMD and NDI data sets for these people.

4. Identification of complete hospital stay

In the current NHMD, a person can have several hospital episodes relating to the one hospitalisation. Using episodic data means that for some people their total length of stay in hospital is underestimated and their in-hospital care history incomplete. For the current analysis of transitions, hospital episodes that ended with a movement within the hospital system were excluded from the analysis as they should not have related to movement from hospital to RAC. Consequently, the information for analysis was, of necessity, that associated with the final episode before discharge from hospital. This shortcoming can only be overcome if data on all episodes relating to a single period in hospital can be combined.

Currently, in the NHMD there is neither a name nor a person identifier, nor even some sort of hospital stay identifier, to facilitate joining together contiguous episodes in hospital for the same person. However, such linkage within the hospital data may be possible using event-based linkage similar to that used for this project. The effectiveness of this approach has yet to be investigated, but is expected to be high – especially when linking episodes within the same hospital. Undertaking this within-NHMD linkage before linking to the RAC data would both increase the accuracy of the between-sector data linkage and provide improved information on the period in hospital for analysis.

5. Identification of bi-directional movement

The linkage process used for this publication allows the identification of uni-directional movement from hospital into RAC. While this analysis has provided some information on the use of hospital by people living in RAC, a fuller investigation could be undertaken if transitions in both directions could be identified. To do this, both the data preparation and linkage processes would have to be adjusted to enable identification of:

- discharge from RAC directly into hospital
- discharge from RAC into hospital while on hospital leave from RAC.

In particular, hospital episodes starting with an admission into the hospital system, irrespective of whether they ended with discharge from the sector, would need to be included in the NHMD data set for linkage. For identification of uni-directional movements from RAC to hospital, data linkage could be undertaken using a similar approach to that taken here, except that episodes starting – rather than ending – with a statistical admission or transfer would be excluded. If integrated analysis of movements in both directions were to be required, linkage of contiguous hospital episodes (as above) to obtain hospitalisation dates covering the total period in hospital would be necessary.

Of the above developments, the first two are the most readily achievable, with the remaining three requiring considerable investment in linking – either between or within data sets.