

South Yorkshire Cohort Validation Report (for CLAHRC SY)

Executive Summary

- The South Yorkshire Cohort contains information on 27,806 individuals from the South Yorkshire and surrounding regions (2010-2012).
- The cohort is over-representative of the elderly, females and individuals from affluent areas, with smaller differences by ethnicity and health compared to expected patterns from the 2011 Census.
- Those who gave consent to be contacted again are not too different to those who did not give consent.
- A series of weights will be created to accompany the dataset to improve its representative coverage of South Yorkshire.

Introduction

The South Yorkshire Cohort (SYC) is a longitudinal observational study. It was set up to collect information on the health of residents of South Yorkshire. A range of variables were collected including demographic characteristics, current and long-standing health, well-being, health care usage and health-related behaviours. The first wave of data collection was between 2010 and 2012.

This report examines the representativeness of the SYC data collected in the first wave to validate and indicate how useful it is. The distribution of participants in the SYC across a range of characteristics will be compared against those for the total underlying population of South Yorkshire. The variables used to compare datasets include age, gender, ethnicity, deprivation and health.

Data is mainly taken from the 2011 UK Census, since this provides near perfect coverage on residents in South Yorkshire across a range of variables. It is also useful since the 2011 Census fell in the middle of the first wave of the SYC. Other information drawn on from outside the 2011 Census are detailed in their relevant sections.

Coverage

A two stage sampling procedure was used to collect data. GP surgeries were contacted and asked if they wished to participate in the study, with 43 agreeing (50% acceptance). GP surgeries were selected to achieve representative coverage of the South Yorkshire region. Consenting GP surgeries then mailed out letters of invitation to all their patients aged 16 to 85. Included with the letter of invitation was an eight page questionnaire for which data was collected with. 156,866 letters were mailed out, with 27,806 questionnaires returned (a response rate of 15.9%). Participation between GP surgeries ranged from 32.43% to 9.09%. Of the 27,806 participants, 22,179 (81.7%) gave consent to be contacted again and 22,150 (79.7%) gave consent for researchers to access their health records. 25,862 participants (93.2%) resided in South Yorkshire and the rest of the analyses focuses upon these.

Age and Gender

The proportion of individuals by individual years of age captured in the SYC, as well as for the whole of South Yorkshire, is presented in Figure 1. The SYC contains a greater proportion of individuals aged above 50 than would be expected given the population characteristics of residents of South Yorkshire. This is contrasted by individuals younger than this age range, who are under-represented in the SYC.

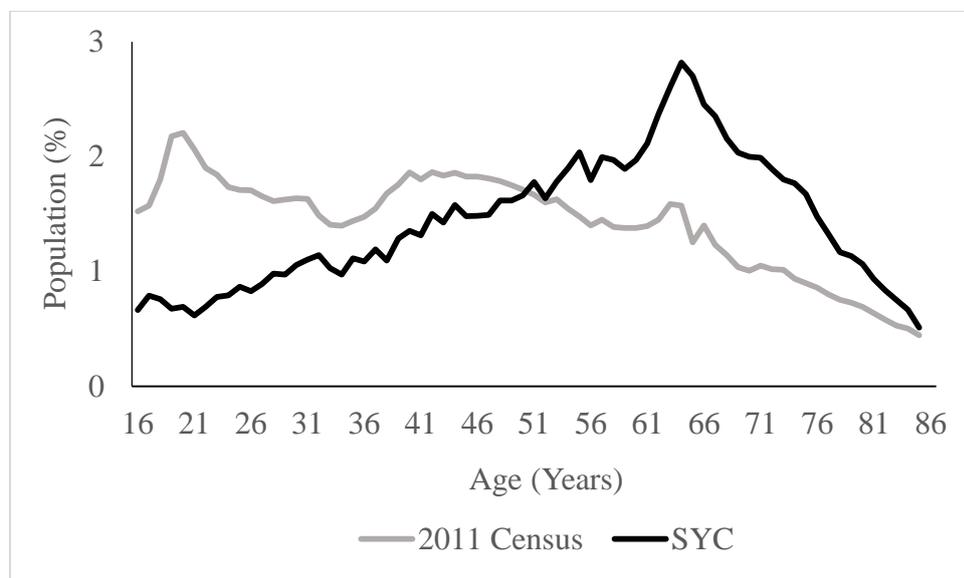


Figure 1: The proportion of ages as a total of all individuals aged 16-85 in the South Yorkshire Cohort (SYC) and the 2011 Census for South Yorkshire.

The Census reports that there are a fairly even spread of males and females in South Yorkshire, with slightly more females (Table 1). The distribution of genders in the SYC is biased, with a greater proportion of females. Figures 2 and 3 show how the distribution of ages vary by gender. Females are more representative of the age profile of South Yorkshire in the SYC than compared to males, although there is some bias for both young and elderly females.

	Males (%)	Females (%)
2011 Census	49.3	50.7
SYC	43.8	56.2

Table 1: The proportion of genders of individuals aged 16-84 in South Yorkshire using the South Yorkshire Cohort and the 2011 Census.

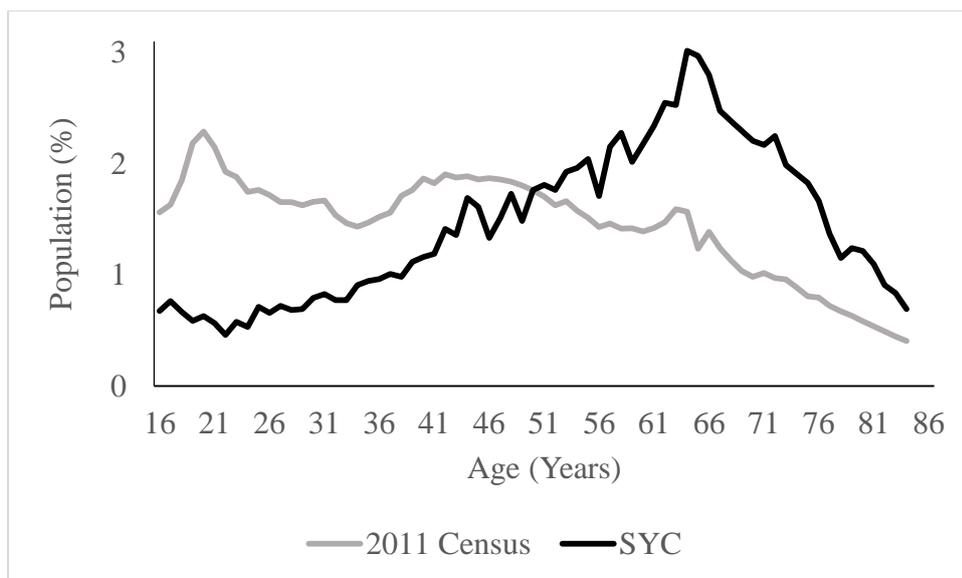


Figure 2: The proportion of ages for males aged 16 to 84 in both the South Yorkshire Cohort (SYC) and the 2011 Census for South Yorkshire.

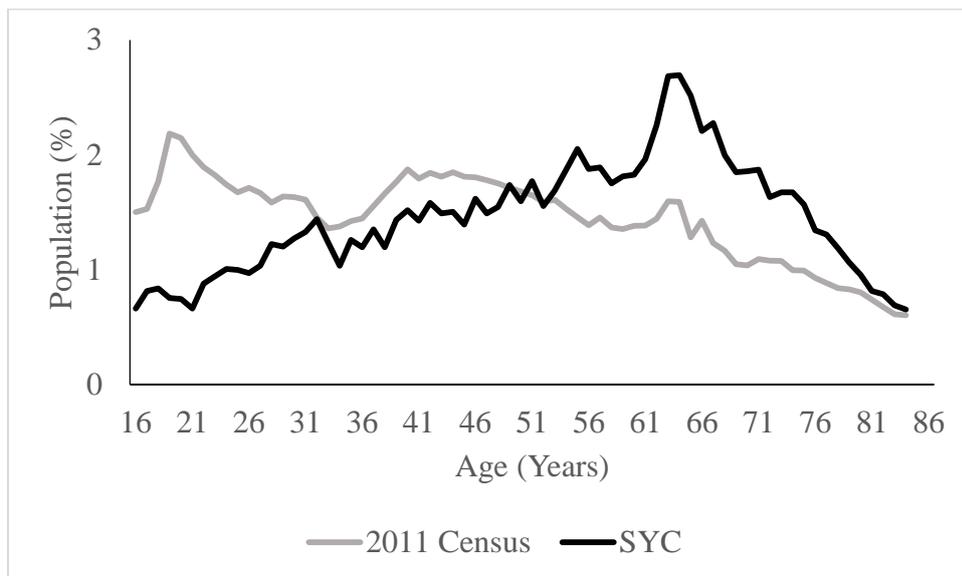


Figure 3: The proportion of ages for females aged 16 to 84 in both the South Yorkshire Cohort (SYC) and the 2011 Census for South Yorkshire.

Ethnicity

Figure 4 compares the proportion for each ‘broad’ ethnic group. The SYC over-represents White categories, with relative differences between the Census and the SYC for the non-White groups being high (although absolute differences remain low).

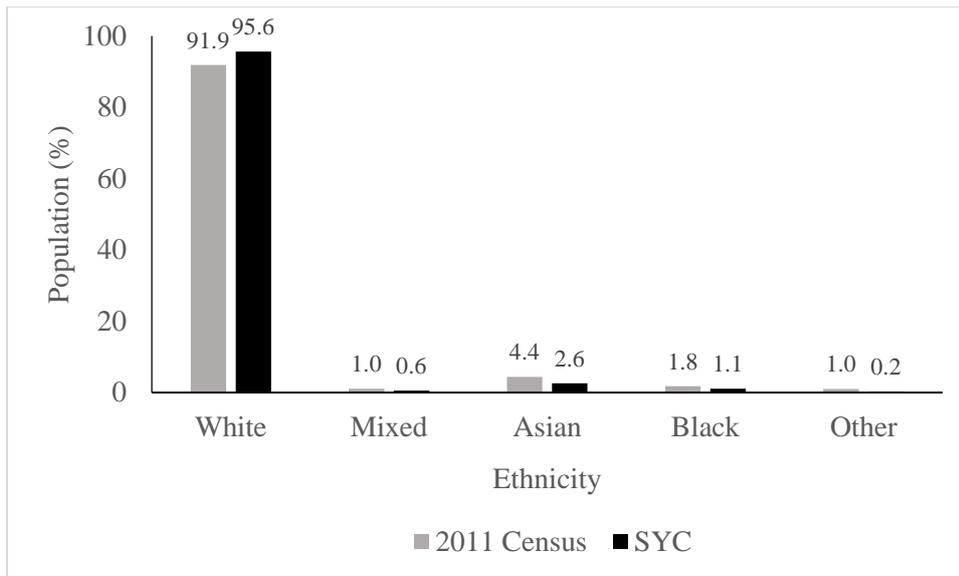


Figure 4: The proportion of ethnic groups in both the South Yorkshire Cohort (SYC) and the 2011 Census (individuals aged 16-84).

Deprivation

To measure deprivation across South Yorkshire, the Indices of Deprivation 2010 measure was used since it provides a multi-dimensional measure of deprivation. The index scores Lower Super Output Areas, which can be ranked and national quintiles were assigned based upon the rankings. Figure 5 shows the proportion of the population of South Yorkshire that lives in each quintile, as well as the proportions of each quintile in the SYC. The SYC has a higher proportion of participants in the least deprived areas than compared to the underlying population. The SYC is under-representing the deprived areas of South Yorkshire.

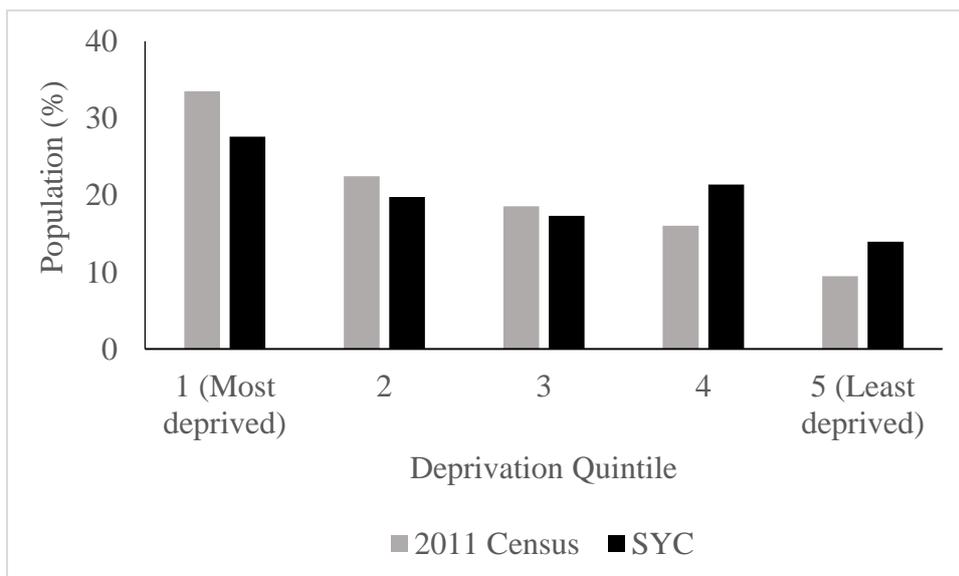


Figure 5: The proportion of the population (aged 16-85) found in each deprivation quintile within the South Yorkshire Cohort (SYC) and the 2011 Census.

Health

There is limited information collected in the Census (and other official sources) to validate the health information collected in the SYC. One question measured asks individuals how limited they are due to their health in performing ‘day-to-day’ activities (‘limited a lot’, ‘limited a little’ or ‘not limited’). This is conceptually similar to one question asked in the SYC. Individuals were asked in the questionnaire whether they had problems with performing their ‘usual activities’ (e.g. work, study, housework, family or leisure activities) and were offered three options (‘none’, ‘some’ or ‘unable to perform’). These match up to the Census variable. The question was included in the questionnaire to calculate the EQ5D measure for quality of life. The values were recoded to match the Census values (i.e. ‘none’ became ‘not limited’, ‘some’ became ‘limited a little’ and ‘unable to perform’ became ‘limited a lot’).

Figure 6 charts the proportions of these two variables. The SYC over-represents individuals with poor health, with a lower proportion not limited in their day-to-day activities than compared to those in the Census. There were differences in the categories for ‘limited a lot’ and ‘limited a little’ as well. However due to the slight inconsistency between these variables, the comparisons should be interpreted carefully.

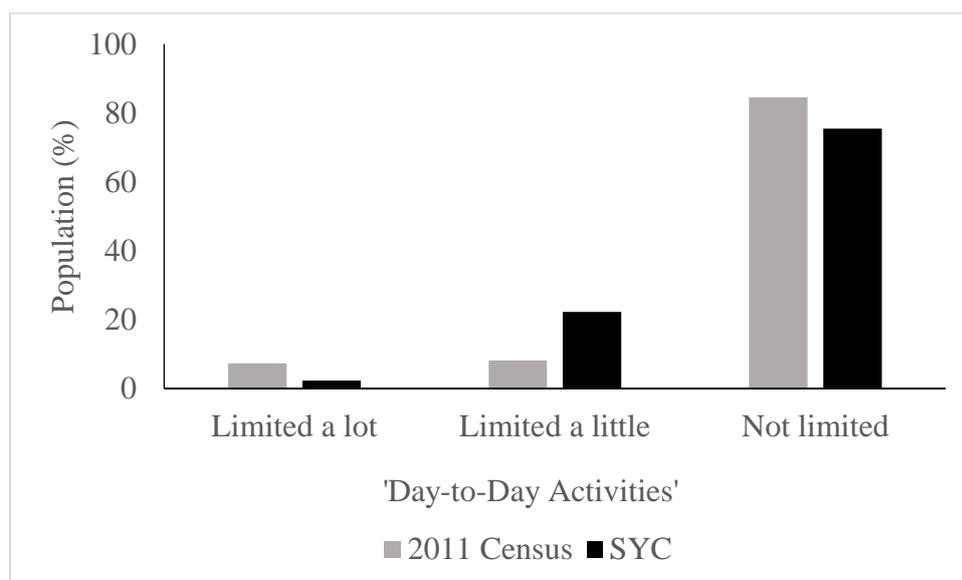


Figure 6: The proportion of individuals (aged 16-64) in the South Yorkshire Cohort (SYC) and the 2011 Census by health status.

Differences in those who wish to be contacted again

21,092 individuals (81.6%) in South Yorkshire gave consent to be contacted again and therefore the analysis compares if they differ from those who chose not to, to evaluate the validity of the previous analyses against those who will be contacted for future waves (Figures 7-10 and Table 2). Although there is some disagreement between those who gave consent to be contacted again and those who did not, the two groups are not too indifferent. Those who did not give consent were slightly older, more likely to be ethnic minorities and from deprived areas. There were no differences by age and health.

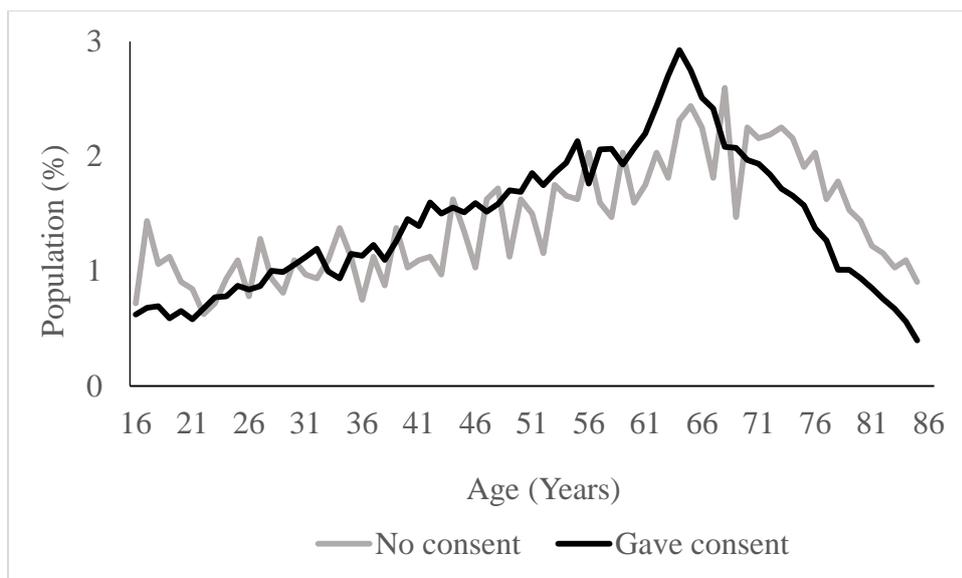


Figure 7: The proportion of ages as a total of all individuals aged 16-85 in the South Yorkshire Cohort by whether they gave consent to be contacted again.

	Females (%)	Male (%)
No consent	13.2	13.9
Gave consent	86.8	86.1

Table 2: The proportion of genders of individuals in the South Yorkshire Cohort by whether they gave consent to be contacted again.

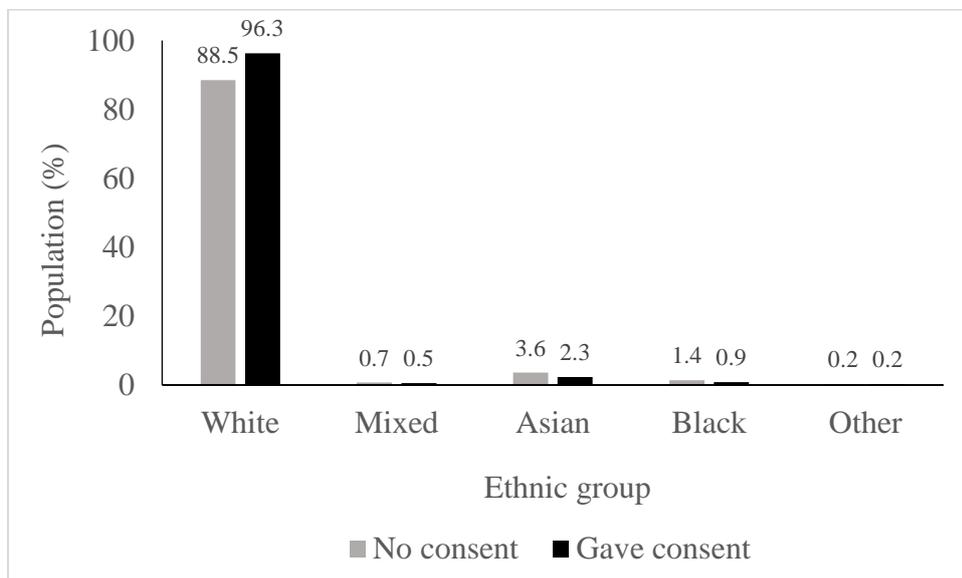


Figure 8: The proportion of ethnic groups in both the South Yorkshire Cohort by whether they gave consent to be contacted again.

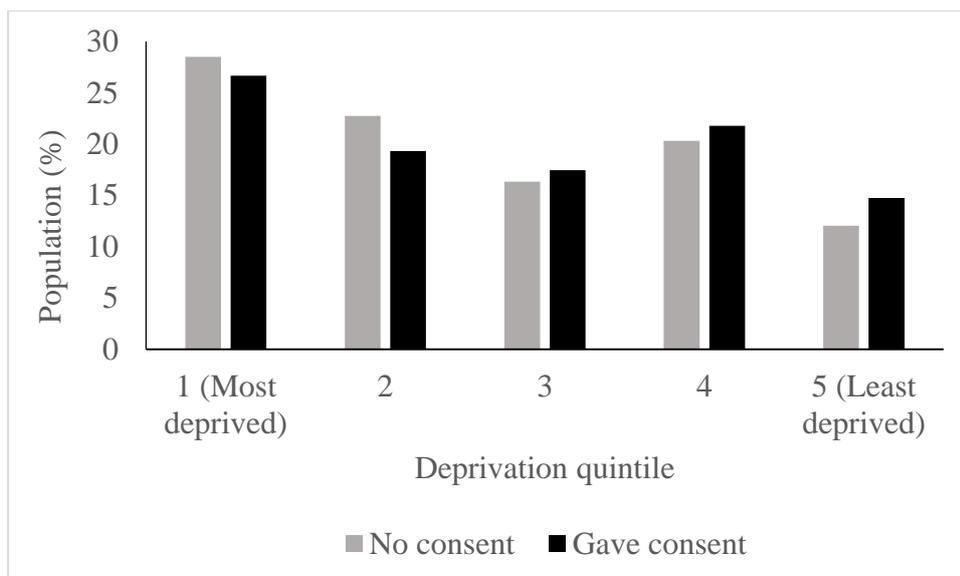


Figure 9: The proportion of the sample found in each deprivation quintile within the South Yorkshire Cohort by whether they gave consent to be contacted again.

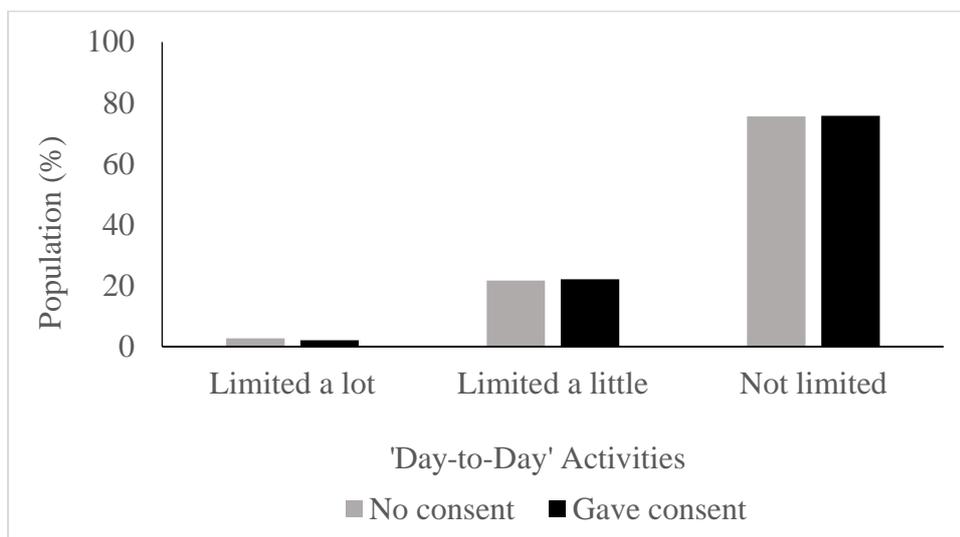


Figure 10: The proportion of individuals by health status in the South Yorkshire Cohort by whether they gave consent to be contacted again.

Conclusion

The validation of the SYC has shown particular population sub-groups to be over- and under-represented through the data collection process with respect to the underlying population of South Yorkshire. Therefore it is recommended that sample weights are created and attached for use in further analyses.

Acknowledgements

Collaboration for Leadership in Applied Health Research and Care for South Yorkshire (CLAHRC SY) acknowledges funding from the National Institute for Health Research (NIHR). The views and opinions expressed are those of the authors, and not necessarily those of the NHS, the NIHR or the Department of Health. CLAHRC SY would also like to acknowledge the participation and resources of our partner organisations.