

# **Subnational Population Projections Accuracy Report**

**Office for National Statistics Centre for Demography  
August 2008**

## **1. Executive Summary**

This report considers the accuracy of the official subnational population projections (SNPPs) made for England since 1996. These projections are produced by the Office for National Statistics Centre for Demography (ONSCD).

Each set of projections is compared to the mid-year population estimates (MYE) for 2006, and consideration is given to the MYE revisions made following the 2001 census and in 2004. The 2007 MYEs became available shortly before publication of this report and there is a brief comparison of these to the two most recent sets of projections.

Total populations of each area have been projected reasonably accurately, particularly in the shorter term. The degree of accuracy decreases for lower levels of geography, but even at local authority level the projections can be considered as sufficiently accurate for a wide range of uses. The average differences (measured using root mean square error) between projections two years ahead and the MYEs for LA district level is less than one per cent.

However some outlier areas were noted and these inaccuracies were linked to the projections methodology around handling of student migration, and migration of highly mobile populations within London. These aspects of the methodology are due to be reviewed before the next update of the projections. In the meantime it is planned to introduce subnational projection variants, including 'high' and 'low' variants, and this has the potential to improve understanding of the level of uncertainty around the principal projections.

## **2. Introduction**

The purpose of the subnational population projections (SNPPs) is to show how the size and structure of the population, for each local authority, would change over the next 25 years if recent demographic trends were to continue. They are intended to help local users plan ahead, and are one of the factors used to determine how funding will be distributed to local government and health authorities.

The projections are not forecasts, and make no attempt to account for planned developments (such as house-building or economic growth) that may affect the future population size in a given area. This is partly because reliable and consistent data across all the local authorities in England on their plans, and how these are likely to affect their population, are very difficult to collect. ONSCD is however considering whether to produce a projection variant that takes regional house building plans into account.

So, strictly speaking, even if the projections differ from the subsequent annual mid-year population estimates they are not wrong as they simply extrapolate

recent trends forwards. However as the projections are used for planning purposes it is correct to assess their level of accuracy.

This paper measures the accuracy of the historic SNPPs by comparing their projections for 2006 to the mid-year population estimates (MYEs) for the same year. We are therefore defining accuracy in terms of being an accurate forecast of the relevant MYE. As the SNPPs are based upon the MYEs, both in terms of getting an initial population from which to project forwards and for data on recent trends, this is the only sensible accuracy test to apply to them.

This report is accompanied by an Excel file containing the relevant data.

## **2.1 Issues when measuring accuracy of historic SNPPs**

The SNPPs examined in this paper are based upon the MYEs for the years 1996, 1998, 2000, 2003, 2004 and the revised MYEs for 2004 (plus a shorter section comparing the 2006-based SNPPs with the 2007 MYEs). It becomes increasingly difficult to compare earlier SNPPs against the 2006 MYEs due to:

- Area boundary changes, meaning we cannot compare like with like
- Revisions to the MYEs, meaning that the figures upon which a set of projections were based are no longer considered to be the best available
- Significant changes to the projections model, meaning that older sets of projections are of little relevance when assessing the likely accuracy of new ones
- Availability of data, as we only have hard copies of older projections and these only include selected areas

So in this paper, which is the first accuracy report on the SNPPs, we have got back only as far as the 1996-based projections. In future it would be possible to produce an expanded analysis, tackling some of the issues above. However, such a study would be of limited value and may not justify the resource required to undertake it.

There were two sets of revisions to the MYEs that must be taken into account when assessing the SNPPs back to the 1996-based.

The biggest one occurred following the 2001 census, when it was found that the estimates rolled forwards from the 1991 census were too high for many areas. Revised MYEs were then produced back to 1992, which were in line with the 2001 census results. This means that the 1996, 1998 and 2000-based SNPPs were based upon MYEs that are not consistent with the 2006 MYE.

There was a second set of revisions following the 2004 MYEs, due to a change in the methodology used to estimate international migration flows to and from each area. Revised MYEs were then produced back to 2002, meaning that the 2003-based and 2004-based projections are based on MYEs that are not consistent with the 2006 MYE.

So the only set of SNPPs that are based on MYEs consistent with the 2006 MYE are the revised 2004-based. With all of the previous sets of SNPPs it has to be taken into account that, however good the projections were, a degree of inaccuracy is inevitable due to the MYEs upon which they were based being subsequently revised.

### **3. Summary of the subnational population projections (SNPP) methodology**

Before looking at the accuracy of the historic SNPPs it will be useful to briefly go over the methodology used to produce them. This will be relevant when we consider what factors might lead to projections that differ from later mid-year estimates.

The projections take as their starting point the mid-year population estimates for the base year. The projections for each year are calculated by ageing on the population from the previous year, applying local fertility and mortality rates to calculate the projected number of births and deaths, then adjusting for migration into and out of the area. This process is repeated for each of the twenty-five years of the projection period.

The sum over all areas of the projected population and births, deaths and migration components are constrained to the totals for England in the national population projections with the same base year. The national projections were produced by the Government Actuary's Department (GAD) up to the 2004-based, and are now produced by ONS.

Migration data are the most complex and require careful restructuring for use in the projection calculations. This partly reflects the limitations of the data sources available. While there is a legal requirement to register a birth or a death there is no such requirement to register a change of address. The subnational projections model uses migration data for single year of age and sex for each area, as with the other components. Estimates are required for in and out migration flows for both internal (defined as migration within England for these projections) and international migration.

The projections are based on past demographic trends, generally as observed over the five years prior to the base year. So for example, the 2004-based projections will be based on trends observed in the mid-year population estimates from 1999-2004.

Potential issues that could affect the accuracy of the projections include:

- The mid-year population estimates for the base year (by local area, age and gender) are not accurate in every respect
- The data sourced from the mid-year estimates for the past five years on births, deaths and migration are not all accurate. This is more likely to be an issue for migration data, as the sources are not so good.

- The projection is based on accurate demographic trends but these trends change, at least in some areas. For example, substantial economic and/or housing development in an area might lead more people to migrate there from other areas.
- The national population projection differs from subsequent mid-year estimates, meaning that the subnational projections are constrained to the wrong total.

Further detail on the current SNPP methodology can be found in the methodology guide on the web page:

<http://www.statistics.gov.uk/snpp>

#### **4. Comparison of historic subnational population projections (SNPPs) against the 2006 mid-year estimates (MYEs)**

In this section we compare the subnational projections, from 1996-based through to the revised 2004-based, against the mid-year population estimates for 2006.

##### **4.1 Accuracy of the SNPPs by level of geography**

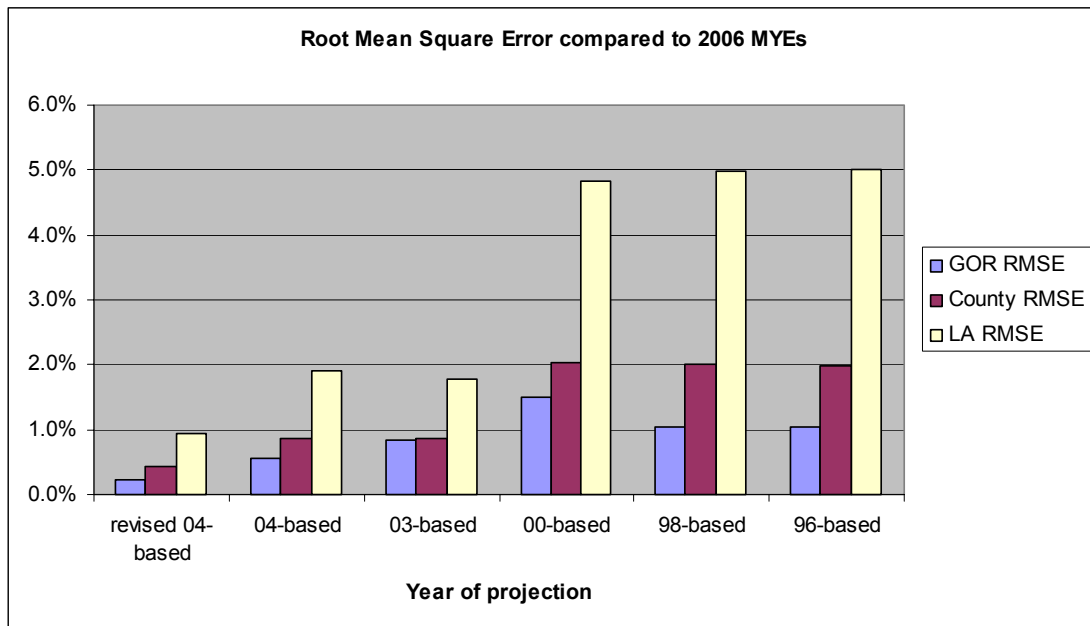
To begin with we look at the root mean square error (RMSE), also known as the root mean square deviation<sup>1</sup>, for the projections of total population at the level of government office region (GOR), county and local authority (LA).<sup>2</sup> The results are shown in the chart.

---

<sup>1</sup> The RMSE is a frequently-used measure of the differences between values predicted by a model or an estimator and the values actually observed from the thing being modelled or estimated. The RMSE is a means to aggregate individual differences into a single measure of predictive power. In the usage in this paper it provides a measure of the average difference between a projection and the mid-year estimate.

<sup>2</sup> London boroughs and unitary authorities are included as local authorities for the purpose of compiling these figures.

## Accuracy of historic SNPPs compared to 2006 MYEs



	revised 04-based	04-based	03-based	00-based	98-based	96-based
GOR RMSE	0.2%	0.6%	0.8%	1.5%	1.0%	1.0%
County RMSE	0.4%	0.9%	0.9%	2.0%	2.0%	2.0%
LA RMSE	0.9%	1.9%	1.8%	4.8%	5.0%	5.0%

As would be expected, the RMSE is higher for the lower levels of geography and as the period between the projection base and the reference year (2006) increases. In particular, the projections at LA level show a higher level of error than those for GORs and counties.

There is a very noticeable increase in accuracy for the 2003-based projections onwards. This is likely to be mainly due to the mid-year estimates being significantly revised following the 2001 census. In fact the RMSEs for the 2000-based projections, shown in the chart above, are similar to those we get when we compare mid-year estimates rolled forward from the 1991 Census with the final 2001 population estimates based upon the 2001 Census. The two sets of values are shown in the table.

## Accuracy of 2000-based population projections compared to accuracy of the 2001 MYEs Rolled Forward from 1991

	2000-based projection vs 2006 MYE	2001 Rolled Forward MYEs vs 2001 Census-based MYEs
GOR RMSE	1.5%	1.7%
County RMSE	2.0%	1.7%
LA RMSE	4.8%	4.3%

We can never expect the projections to be more accurate than the mid-year estimates upon which they are based, so in fact these older projections are as accurate for 2006 as could be hoped.

The revised 2004-based projections are also noticeably more accurate than the original 2004-based projections and the 2003-based projections. The mid-year estimates were revised again in 2004, less significantly this time, to incorporate improvements to the methodology for estimating international migration. So in fact the revised 2004-based projections are the only ones to be based on a set of mid-year estimates that are entirely consistent with those for 2006.

The accuracy of the revised 2004-based projections is encouraging, although in this case the projection period is only two years. The RMSE at LA level is less than 1%, with county level at 0.4% and GOR level at 0.2%.

Another external factor that affects the subnational population projections is the accuracy of the national projections. The subnational population projections are constrained to sum to the most recent national projection, so inaccuracy at national level will mean that the subnational projections are scaled to the wrong total. The national projections are also based on the mid-year population estimates, so revisions to the mid-year estimates will affect their accuracy too.

#### Accuracy of national projections for 2006

	Projection for 2006	Difference from 2006 MYE
1996-based	50,526	-0.5%
1998-based	51,052	0.6%
2000-based	51,165	0.8%
2003-based	50,483	-0.6%
2004-based	50,714	-0.1%
<b>2006 MYE</b>	<b>50,763</b>	

The 2001 mid-year population estimate for England was revised downwards by 1.5% following the census. The effect of this can be seen in the national projection for 2006 dropping by nearly 0.7 million between the 2000-based and 2003-based projections. The 2004-based national projection is easily the most accurate, and this is reflected in the accuracy of the revised 2004-based subnational projections.

#### **4.2 Average accuracy of the SNPPs within each geography**

We have so far looked at the level of accuracy by geography and seen that, as would be expected, the projections tend to be more accurate for GORs and for counties than they are for LAs. But what level of variation is there within each geography, and do the projections tend to be more accurate for some areas than for others?

#### 4.2.1 Average accuracy of the SNPPs by GOR compared to the 2006 MYEs

	revised 04-based	04-based	03-based	00-based	98-based	96-based	average	max error	min error
NORTH EAST	-0.5%	-0.1%	-1.0%	-0.5%	0.0%	0.0%	<b>-0.4%</b>	<b>0.0%</b>	<b>-1.0%</b>
NORTH WEST	0.1%	0.3%	-0.3%	0.4%	0.4%	-0.2%	<b>0.1%</b>	<b>0.4%</b>	<b>-0.3%</b>
YORKSHIRE AND THE HUMBER	-0.3%	-0.7%	-1.8%	-1.1%	-0.9%	-0.9%	<b>-0.9%</b>	<b>-0.3%</b>	<b>-1.8%</b>
EAST MIDLANDS	-0.2%	-0.7%	-1.0%	-1.3%	-1.1%	-1.2%	<b>-0.9%</b>	<b>-0.2%</b>	<b>-1.3%</b>
WEST MIDLANDS	-0.1%	0.1%	-0.2%	-0.1%	0.3%	-0.2%	<b>0.0%</b>	<b>0.3%</b>	<b>-0.2%</b>
EAST	0.0%	-0.7%	-0.6%	1.0%	0.7%	-0.4%	<b>0.0%</b>	<b>1.0%</b>	<b>-0.7%</b>
LONDON	0.0%	1.1%	0.1%	3.0%	1.3%	-2.3%	<b>0.5%</b>	<b>3.0%</b>	<b>-2.3%</b>
SOUTH EAST	-0.1%	-0.4%	-0.3%	2.6%	2.3%	1.3%	<b>0.9%</b>	<b>2.6%</b>	<b>-0.4%</b>
SOUTH WEST	0.0%	-0.2%	-0.5%	0.3%	0.1%	-0.5%	<b>-0.2%</b>	<b>0.3%</b>	<b>-0.5%</b>

Starting at the GOR level, we can see that there has been a tendency for the SNPPs to produce projections that are too low for East Midlands, Yorkshire and the Humber and to a lesser extent for the North East. In the former two cases all of the projections have under-estimated the total population in 2006, while for the North East the 2000-based projections onwards have been under-estimates.

The revised 2004-based projections for Yorkshire and the Humber and East Midlands show an improved level of accuracy, though are still noticeably too low for 2006. The situation for the North East is more concerning, with the revised 2004-based projection actually being a bigger under-estimate than the original 2004-based projection (based on the MYEs prior to their revision of international migration figures).

The South East and London are the regions where the subnational projections have tended to over-estimate. The 2000-based projections made the most significant over-estimate in each case. This will be partly due to the population of both regions, particularly London, being over-estimated by the MYEs prior to the 2001 census. The 2001 MYE for London was revised downwards after the census by 3.4%, and the South East by 1.2%.

The projections for the South East since the 2003-based projection have actually been slight under-estimates, so the MYE revisions due to the 2001 Census seem to have fixed the issue of these projections being too high.

For London the original 2004-based projection was too high, but the revised 2004-based projection was spot on. The revisions to the 2004 MYEs reduced the estimate for the number of international migrants that stay within London. This is probably why the revised 2004-based projection was closer to the 2006 MYE.

Overall London stands out as the GOR with the biggest projection errors on both the positive (3.0%) and negative (-2.3%) sides. This reflects the relative difficulty of producing accurate projections for the highly mobile population in this region.

## 4.2.2 Average accuracy of the SNPPs at county and LA level

In this section we focus on those areas where the SNPPs have not proved to be as accurate as they have for others at the same level of geography.

The following tables show, for counties and LAs, the five areas where the average projection is highest and lowest when compared to the 2006 MYEs.

### The five counties where the average projection is highest compared to the 2006 MYEs

COUNTIES	revised 04-based	04-based	03-based	00-based	98-based	96-based	average	max	min
Cambridgeshire	-0.1%	0.9%	0.2%	4.5%	4.4%	3.3%	2.2%	4.5%	-0.1%
Oxfordshire	0.3%	-0.3%	-0.2%	5.2%	3.7%	2.5%	1.9%	5.2%	-0.3%
West Sussex	0.2%	-0.2%	-0.1%	4.0%	4.1%	2.0%	1.7%	4.1%	-0.2%
Hampshire	0.0%	0.0%	0.1%	3.2%	3.4%	3.0%	1.6%	3.4%	0.0%
Buckinghamshire	-0.6%	-1.3%	-1.6%	4.0%	3.9%	3.9%	1.4%	4.0%	-1.6%

### The five counties where the average projection is lowest compared to the 2006 MYEs

COUNTIES	revised 04-based	04-based	03-based	00-based	98-based	96-based	average	max	min
Northamptonshire	-0.7%	-1.6%	-1.8%	-3.0%	-3.2%	-4.2%	-2.4%	-0.7%	-4.2%
Nottinghamshire	0.3%	-0.2%	-0.7%	-3.4%	-3.8%	-2.5%	-1.7%	0.3%	-3.8%
Lincolnshire	0.9%	0.4%	0.2%	-3.5%	-3.0%	-3.6%	-1.4%	0.9%	-3.6%
Essex	-0.1%	-1.1%	-1.2%	-1.1%	-1.9%	-2.7%	-1.4%	-0.1%	-2.7%
Cornwall and Isles of Scilly	0.3%	0.2%	0.2%	-2.1%	-2.7%	-3.8%	-1.3%	0.3%	-3.8%

### The five LAs where the average projection is highest compared to the 2006 MYEs

Local Authority	revised 04-based	04-based	03-based	00-based	98-based	96-based	average	max	min
Richmond upon Thames	0.6%	6.5%	4.7%	21.8%	18.6%	14.4%	11.1%	21.8%	0.6%
Kensington and Chelsea	4.7%	16.5%	8.7%	17.2%	4.9%	-3.8%	8.0%	17.2%	-3.8%
Forest Heath	-3.5%	-0.6%	-1.0%	18.7%	16.7%	15.3%	7.6%	18.7%	-3.5%
Cambridge	-1.1%	5.3%	-0.8%	14.9%	12.9%	11.7%	7.2%	14.9%	-1.1%
Southend-on-Sea UA	-0.3%	-0.1%	0.9%	14.2%	15.2%	12.7%	7.1%	15.2%	-0.3%

### The five LAs where the average projection is lowest compared to the 2006 MYEs

Local Authority	revised 04-based	04-based	03-based	00-based	98-based	96-based	average	max	min
Southwark	-2.1%	-4.5%	-6.2%	-11.1%	-14.2%	-16.9%	-9.2%	-2.1%	-16.9%
Welwyn Hatfield	-2.4%	-5.9%	-6.5%	-9.4%	-10.0%	-9.2%	-7.2%	-2.4%	-10.0%
Ribble Valley	1.2%	0.3%	0.3%	-8.8%	-13.3%	-15.7%	-6.0%	1.2%	-15.7%
Oswestry	-0.5%	-1.0%	-2.5%	-9.8%	-10.3%	-6.8%	-5.2%	-0.5%	-10.3%
Gravesham	-0.9%	-2.6%	-1.8%	-7.5%	-8.4%	-9.7%	-5.2%	-0.9%	-9.7%

Looking at the 'top five' counties, the more inaccurate projections were all prior to the 2001 census when the MYEs were revised. Four of the five counties (Oxfordshire, West Sussex, Hampshire, Buckinghamshire) are in the South East region. It was noted above that the South East GOR as a whole

saw high projections prior to the census, but that this issue seems to have been corrected by the MYE revisions.

The high projections for Cambridgeshire are partly driven by high projections for Cambridge, which features in the list of 'top five' LAs. The 2001 MYE for Cambridge was adjusted downwards by a very significant 14.4% following the 2001 Census and, as the projections have been considerably more accurate since (particularly following the 2004 MYE revision), the major issue seems to have been the changes to the MYEs.

The East Midlands GOR includes all of the 'bottom three' counties (Northamptonshire, Nottinghamshire, Lincolnshire), where the average projection for 2006 has been too low. It was noted above that the projections have tended to be too low for the East Midlands GOR as a whole, though the situation has improved with the more recent sets of projections.

The under-projections for Cornwall and Isles of Scilly have been reversed in the last three sets of projections (2003-based onwards), all of which projected a slightly higher figure for 2006 than was seen in the MYE.

The Essex projections have been quite consistently too low for the year 2006. The 2003-based and 2004-based projections for the East GOR as a whole were too low, though this was corrected in the revised 2004-based projections. None of the LAs within Essex stand out as being particularly responsible for the low overall projections.

Looking at the individual LAs, the projections for 2006 for the London boroughs of Richmond Upon Thames and Kensington & Chelsea have been consistently too high. Other London boroughs where projections have generally been higher than the 2006 MYEs are Westminster, Barnet and Hammersmith & Fulham. Most of these areas stand out as having had their 2001 MYE revised substantially downwards following the census:

- Westminster, down 40.5%
- Kensington & Chelsea, down 25.8%
- Richmond Upon Thames, down 15.1%
- Barnet, down 11.4%
- Hammersmith & Fulham, down 3.0%

Population estimates for these boroughs were generally revised downwards again in 2004, when changes to the international migration methodology led to a higher number of these inward migrants being distributed outside London.

The remaining LAs in the 'top five' (Forest Heath, Cambridge, Southend-on-Sea UA) also had their 2001 MYEs revised significantly downwards as a result of the 2001 Census:

- Forest Heath, down 14.5%
- Cambridge, down 14.4%
- Southend-on-Sea, down 9.7%

The projections for these areas following the census revisions to the MYEs have been far more accurate.

The LAs in the 'bottom five' (where the average projection for 2006 has been too low) can also be partially explained by the Census revisions to the 2001 MYEs, as these led to increased MYEs for four of the five areas:

- Oswestry, up by 5.7%
- Gravesham, up by 4.1%
- Welwyn Hatfield, up by 2.4%
- Southwark, up by 1.7%

The projections for Welwyn Hatfield and Southwark have remained too low following the 2001 Census revisions to the MYEs, though both were improved by the subsequent 2004 revisions.

Overall, when looking back at the accuracy of the historic projections, the quality of the MYEs for each area seem to be the dominant factor in determining the quality of the projections. Significant MYE revisions, particularly around the 2001 Census, have resulted in some older projections being quite inaccurate for 2006. We have not identified many instances of low quality projections that cannot be at least partially explained by significant MYE revisions in those areas.

#### **4.3 Accuracy of the revised 2004-based SNPPs by county and LA**

As the revised 2004-based SNPPs are the only ones to be based on an MYE that is entirely consistent with the 2006 MYE (other than the 2006-based projections where a short comparison is made with the 2007 MYE in the next section of this document), it is informative to look briefly at the counties and LAs where these projections were furthest off in either direction. This may give the best indication of where the projections methodology itself (independently of the quality of the MYEs) may not be working as well as it could do, though we must be cautious when drawing conclusions from just one comparison for an area.

The five counties where the revised 2004-based projection is highest compared to the 2006 MYEs

<b>County</b>	<b>revised 2004-based</b>
Lincolnshire	0.9%
Wiltshire	0.9%
East Sussex	0.8%
Somerset	0.6%
Bedfordshire	0.6%

The five counties where the revised 2004-based projection is lowest compared to the 2006 MYEs

<b>County</b>	<b>revised 2004-based</b>
Durham	-0.9%
Northamptonshire	-0.7%
Tyne and Wear (Met County)	-0.7%
West Yorkshire (Met County)	-0.6%
Buckinghamshire	-0.6%

The five LAs where the revised 2004-based projection is highest compared to the 2006 MYEs

<b>Local Authority</b>	<b>revised 2004-based</b>
Kensington and Chelsea	4.7%
Westminster	4.2%
Newham	2.7%
Richmondshire	2.4%
West Wiltshire	2.2%

The five LAs where the revised 2004-based projection is lowest compared to the 2006 MYEs

<b>Local Authority</b>	<b>revised 2004-based</b>
Forest Heath	-3.5%
Exeter	-3.4%
Durham	-3.1%
Welwyn Hatfield	-2.4%
Charnwood	-2.3%

It will take further investigation to determine whether there are particular reasons why the above counties and LAs were not projected as accurately as the rest, and if these reasons may apply to the accuracy of future SNPPs too.

There is reason to suspect regular under-projection of the population in some student areas. Clearly Exeter and Durham have large student populations as a proportion of their total population. Charnwood also houses some of the student population of Loughborough University.

Student migration can be very difficult to measure and project accurately. The SNPPs currently track migration around the country via the NHS Central Register (NHSCR), but if people don't promptly register with a new doctor following a move (and young people often don't) then the migration statistics will miss them.

There can also be problems calculating accurate migration propensities for areas where students represent a high proportion of the total population in a particular age-group. This is because the recorded number of migrants for a particular year of age (such as 22) can be almost as high, or occasionally even higher, than the MYE for the population of that age in the base year. So

we can end up with very high migration propensities, approaching 100%, that need to be rationalised somehow.

ONSCD is currently undertaking work looking at improving the measurement of student migration using data from the Higher Education Statistics Authority.

Forest Heath is a known problem area for the projections due to the large US military bases there. Accurately capturing and modelling the inflows and outflows of the armed forces dependants has been an issue and may be responsible for the under-estimate.

The three LAs where the revised 2004-based SNPPs were furthest above the actual 2006 MYEs are all in London (Kensington & Chelsea, Westminster, Newham). However the projection for the London GOR as a whole was about right, suggesting that it is the way the model distributed people around London that could have been better. There were under-projections for other London boroughs, in particular Southwark (-2.1%), Waltham Forest (-1.0%) and Merton (-1.0%). As has been noted above, due to the mobility of London's population it is a particularly difficult region for which to produce accurate projections.

Overall it appears to be the areas with high migration flows (student areas, areas containing foreign armed forces, London boroughs) that have been caused problems for the SNPPs. There may be additional issues, for example for counties like Lincolnshire and Wiltshire, but it would require further investigation to determine this.

## **5. Comparison of recent subnational population projections against the 2007 mid-year population estimates**

The 2007 mid-year population estimates (MYEs) became available shortly before publication of this report<sup>3</sup> and it's useful to do a brief analysis of how they compare to the two most recent sets of SNPPs, which are the 2006-based and revised 2004-based. These two sets of projections were based on MYEs that are consistent with those for 2007 (there have been no revisions to the MYEs in between) so this will not be a factor in assessing their accuracy.

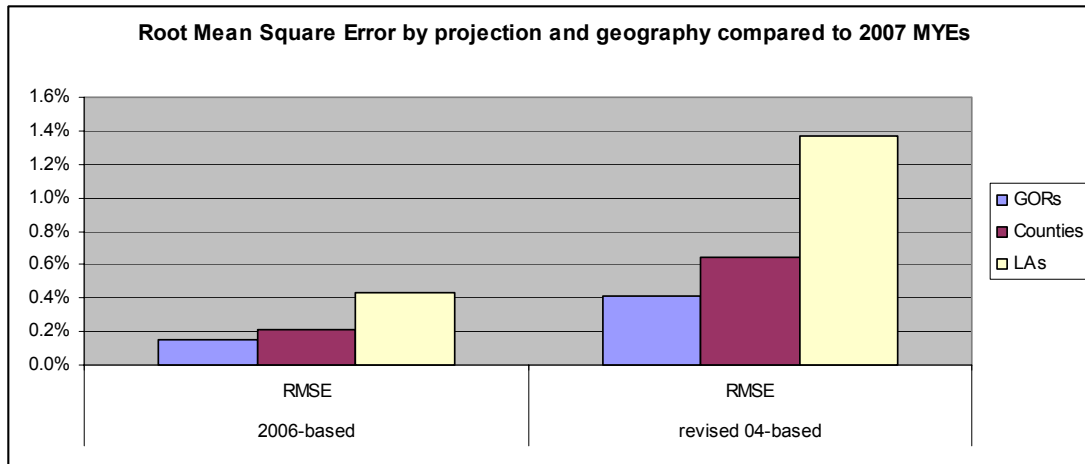
### **5.1 Accuracy by level of geography**

#### **Overall level of accuracy by projection and geography compared to the 2007 MYEs**

As in the previous section, we have used the root mean square error (RMSE) to measure the average accuracy of each projection by level of geography.

---

<sup>3</sup> Early access was granted to the mid-2007 population estimates, for the purposes of writing this report, as this report is a complementary publication to the estimates.



	2006-based RMSE	revised 04-based RMSE
GORs	0.2%	0.4%
Counties	0.2%	0.6%
LAs	0.4%	1.4%

The accuracy of the three year projection from the revised 2004-based figures for 2007 is significantly better at each level of geography than the 2003-based projections for 2006. This is not surprising due to the mid-year estimates being revised in 2004. So the figures above are the best reflection of the SNPPs accuracy when looking three years ahead.

For an unbiased estimator, the root mean square error is equivalent to the standard error. This means that approximately 95% of cases should fall within two RMSEs of the target value.

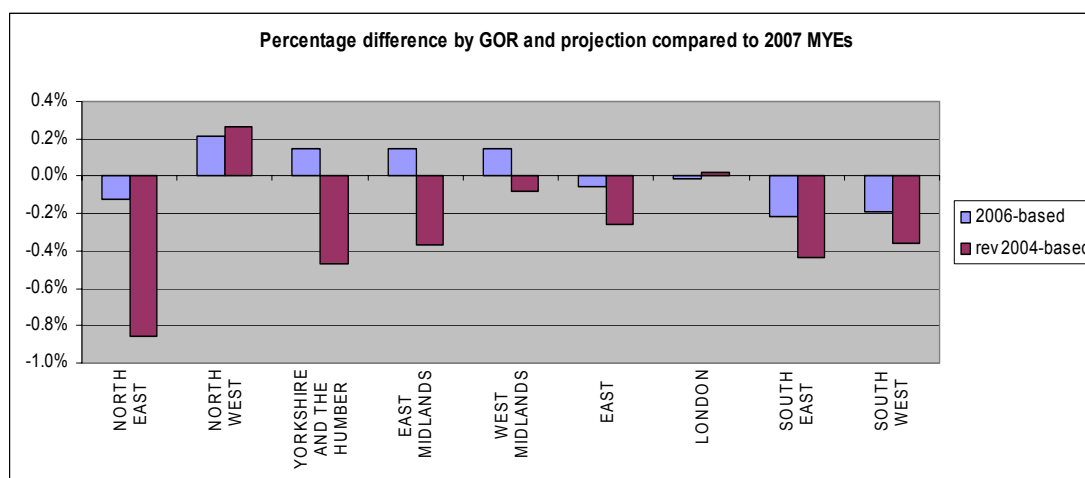
On this basis, when looking three years ahead from the projections base year, roughly 95% of LAs will be within 2.8% of their projection, counties within 1.2% and GORs within 0.8%.

Judging from the RMSEs for the 2006-based projections after one year, they look as though their accuracy after three years (for 2009) will be similar to those for the revised 2004-based projections.

## 5.2 Accuracy within each geography

Similarly to section 4, we will now look at the level of accuracy for the SNPPs within each geography to see if the projections are more accurate for some areas than others.

Comparison of 2006-based and revised 2004-based SNPPs to the 2007 MYEs at GOR level



Positive values indicate the projections were higher than the estimates.

<b>GOR</b>	<b>2006-based % difference</b>	<b>revised 2004-based % difference</b>
NORTH EAST	-0.1%	-0.9%
NORTH WEST	0.2%	0.3%
YORKSHIRE AND THE HUMBER	0.1%	-0.5%
EAST MIDLANDS	0.1%	-0.4%
WEST MIDLANDS	0.2%	-0.1%
EAST	-0.1%	-0.3%
LONDON	0.0%	0.0%
SOUTH EAST	-0.2%	-0.4%
SOUTH WEST	-0.2%	-0.4%

The 2006-based projections are within plus or minus 0.2% of the 2007 MYEs for each GOR.

In the previous section it was noted that projections for Yorkshire & Humberside, East Midlands and the North East had tended to be too low since 1996. However the 2006-based projections for Yorkshire & Humberside and East Midlands were marginally too high in 2007. The 2006-based projection for the North East was marginally too low again, and the revised 2004-based projection for the North East is now looking significantly too low.

The 2006-based and revised 2004-based SNPPs for the South East and South West regions were slightly low for 2007. The North West region is the only one where both sets of projections are too high.

### Range of differences and outliers at LA level

The table shows the extreme differences, both high and low, for the 2006-based and revised 2004-based projections at LA level when compared with the 2007 MYEs. The row 'all areas' includes every LA, while the row 'middle 95%' excludes the top and bottom nine LAs to show the range when the most extreme 5% of cases are removed (there are 354 English LAs).

	06-based Max	06-based Min	Revised 04-based Max	Revised 04-based Min
All areas	1.3%	-1.3%	7.5%	-5.3%
Middle 95%	0.7%	-0.9%	2.1%	-3.0%

The range of difference when these outliers are excluded is roughly as expected from the root mean square error figures shown earlier in this section.

The following tables show the top and bottom nine LAs, based upon the percentage error of their revised 2004-based projections compared to the 2007 MYEs, which were treated as outliers for the analysis above. The percentage difference of their 2006-based projection is included for information.

	<b>2006-based % difference</b>	<b>revised 2004-based % difference</b>
Kensington and Chelsea	1.3%	7.5%
Westminster	1.3%	6.6%
Richmondshire	0.8%	3.3%
South Holland	0.8%	3.0%
West Wiltshire	0.4%	2.9%
Boston	0.9%	2.9%
Newham	-0.8%	2.5%
Bedford	0.9%	2.3%
Bridgnorth	0.2%	2.1%

	<b>2006-based % difference</b>	<b>revised 2004-based % difference</b>
Durham	-1.3%	-5.3%
Exeter	-0.5%	-4.8%
Forest Heath	-0.2%	-4.1%
Southwark	-0.6%	-3.4%
Norwich	-0.6%	-3.3%
Charnwood	-0.1%	-3.3%
Rushmoor	-1.1%	-3.1%
Colchester	-0.6%	-3.0%
West Berkshire UA	-0.7%	-3.0%

It is quite noticeable that the nine LAs with the biggest over-projection (in percentage terms) from the revised 2004-based SNPPs also received a projection that was too high in 2006, with the exception of Newham. All nine of the LAs receiving the biggest under-projection in the revised 2004-based SNPPs also received a projection that was too low in 2006. This suggests there are reasons why at least some of these areas are consistently getting projections that are not as accurate as the majority of LAs.

On the high end, Kensington & Chelsea and Westminster stand out in both sets of projections. At the low end, Durham and Exeter are also at the extreme in both sets of projections.

Potential reasons for the inaccuracy of the projections for these areas were discussed in the section 4. With Durham and Exeter it is likely that the handling of student migration is a factor, as they have a high ratio of students to general population in the 18-25 age-group. With Kensington & Chelsea and Westminster the handling of migration is again the most likely issue due to the highly mobile population in those areas. Further investigation is needed for these areas, and for others in the tables above, to improve the quality of future projections.

## **6. Future development of the subnational population projections (SNPPs)**

### **6.1 Variant subnational population projections**

The national population projections include a principal projection and variant projections. The principal projection shows what is considered to be most likely to happen, while the variant projections show what would happen in alternative scenarios. For example, a 'high' variant may show what would happen if fertility rates are at the top of their expected range while mortality rates are at the bottom of theirs. This provides users with a feel for the degree of uncertainty around the principal projection.

Up to the current time the SNPPs have only provided principal projections for each English district level local authority (LA), of which there are currently 354. It is obviously a much more complex exercise to produce variants for 354 LA projections than just for the national projection. Issues that arise include:

- Should subnational variants be constrained to sum to national variants, in the same way that the current subnational projections are constrained to sum to the national principal projection?
- Is it possible to have a variant that reflects planned local housing and/or economic development?
- Should ONS produce the variants or enable users to produce their own?
- How should the assumptions be set up for each of the local authorities to produce the variant scenarios?

Prior to the 2006-based SNPPs publication, ONS ran a consultation where all interested parties were invited to comment upon the possibility of producing variant subnational projections and to give their thoughts on the issues listed above. The consultation is now closed but the relevant document is still available on the ONS website:

<http://www.ons.gov.uk/about/consultations/closed-consultations/2006-based-subnational-population-projections/part-2-questionnaire.pdf>

We plan to produce a summary of responses received and to publish this on the SNPP webpage on 30 September 2008:

<http://www.statistics.gov.uk/snpp>

The production of subnational high and low variants in particular would illustrate the degree of uncertainty around the projections and allow users to take this into account.

## **6.2 Review of methodology**

The SNPP team review the methodology between each publication to keep improving the quality of the projections. We will be looking at whether the following areas can be improved before the 2008-based SNPPs are produced:

- Handling of students, school boarders and prisoners
- Handling of foreign armed forces dependants
- Calculation of migration propensities for internal migrants
- Modelling of international migration
- Extending the upper end of the age range beyond '85+'
- Development of the Consultation Panel for reviewing feedback on our initial migration assumptions in the provisional projections
- Working more closely with the national projections team

This accuracy report has noted that it is the areas with relatively high migration, particularly those with special populations such as students, that can cause difficulty for the projections. Progress on the research areas above will help to address that.

## **7. Conclusions**

Assessing the accuracy of the historic subnational population projections (SNPPs) has been complicated by external factors, particular two significant sets of revisions to the population mid-year estimates (MYE) series.

It has therefore been particularly useful to compare the revised 2004-based SNPPs against the 2007 MYEs. There were no MYE revisions in this period,

which allows us to assess the accuracy of the SNPP methodology over a three year period. This is important as it is projections around three years ahead of the base that are used as an input into local area funding settlements.

We found that, as expected, accuracy decreased for lower levels of geography. The revised 2004-based projections, compared to the 2007 MYEs, had root mean square errors (RMSEs) of 0.4% for GORs, 0.6% for counties and 1.4% for LAs. If these RMSEs are typical for a three year projection period then about 95% of the three year projections for GORs would be within 0.8% of the MYE, for counties within 1.2% and for LAs within 2.8%. So for example, in the 2006-based projections for LAs (of which there are 354) we would expect about 336 to be within 2.8% of the 2009 MYEs when these become available.

However we saw that there are recurring outliers amongst the LAs, where the projection is consistently inaccurate in the same direction. The extreme cases were Kensington & Chelsea and Westminster for projections that are too high, and Durham and Exeter for projections that are too low. Further analysis of how the projections handle student migration and migration around London is required to address this issue.

At regional level there was no consistent pattern of a particular region having projections that were too high or low, and in many cases historical error was roughly in line with what would be expected due to MYE revisions. However there is an indication that the revised 2004-based and 2006-based projections for the North East may be too low, and this would merit investigation.

The projections methodology around migration and handling of students is due to be reviewed, and this should be taken forwards with particular attention paid to the effect on the outlier areas referred to above. The planned introduction of 'high' and 'low' variant projections will provide a useful measure of uncertainty, and should reflect the range of error we have found in this analysis.

Overall we have found aspects of the projections that should be improved, but the general level of accuracy in the shorter term is sufficient to make them fit for purpose.