APPENDIX 2
Technical Details

As far as possible, this report has aimed to analyse and present data in a consistent way, despite the differences in quality and availability of data. This appendix attempts to explain some of the techniques used, very briefly. For more detail statistical textbooks such as those referred to here should be consulted.

A2.1 Rates

Rates are used throughout this report to enable comparisons between areas and over time. The populations used as denominators originate from ONS, as discussed in A1.1 above, but sub-district populations had to be estimated for years between censuses.

Primary Care Group/Trust (PCG/PCT) populations have been calculated on the basis of geographical areas. The populations were calculated, for each health authority and each age/sex stratum, by taking the proportions of residents in each PCG/PCT at the time of the 1991 Census and applying those proportions to the ONS mid-year estimate for the health authority.

Townsend quintiles have been used to compare relatively affluent groups of the population of each health authority with the most deprived groups. The Townsend Index (Townsend *et al*, 1988) is a categorisation of all enumeration districts (EDs - the smallest areas available for analysis of Census data, containing about 500 people) in the country, using four variables from the Census (unemployment, housing tenure, overcrowding and car ownership) to attribute a 'deprivation' level to those areas. The enumeration districts in each health authority have then been ranked and grouped into five 'quintiles' of roughly equal population. Hence quintile 1 can be said to include the most affluent areas in each district and quintile 5 the most deprived areas. Several HImP targets refer to these quintiles in setting targets to 'narrow the gap' between rich and poor.

The Townsend quintile populations were calculated in the same way as the PCG/PCT populations, ie. by taking the proportions of residents in each quintile in the 1991 Census, applying those to the ONS mid-year estimates for each age/sex stratum.

A2.2 Directly Standardised Rates

Rates for many of the targets presented in this report vary greatly across the range of age groups and between the sexes. We would expect, on the whole, to see higher mortality and hospital admission rates for a PCG with a large proportion of old people than in another PCG where the population is much younger. To enable comparisons between different areas and over time it is necessary to compensate for the differences between age/sex distributions of the populations being compared.

In order to do this we calculate directly standardised rates, using the European Standard Population (Breslow & Day, 1987; 54). In effect, we use the rates in each age group to calculate what the overall rate would be if the population were the same as the European Standard Population. Hence any differences which still exist must be down to factors other than the age and sex structure of the populations in question.

These calculations are have been carried out in the Microsoft Excel spreadsheet software.

A2.3 Forecasting – Holt's Exponential Smoothing

In this report, the focus has been on targets, and every effort has been made to produce forecasts which are as accurate as possible using the data available.

There are various different ways of forecasting future values of a variable. One, all too frequently used, is to fit a regression line or curve and extrapolate. This assumes a consistent trend with independent errors; something which is clearly not the case in many of the examples in this report, and any 'prediction intervals' produced tend not to reflect the uncertainty about the future. Regression should not be used in this way (Altman, 1991; 316).

In cases where data are available routinely on the factors that influence rates, it may be more accurate to forecast the rates using the historic values of the predictor variables (this is called multivariate forecasting, using causal models). These data on predictors do not exist for the health outcomes we have looked at in this report.

Since we have at our disposal only past observations (called time series) of the rates themselves this report presents forecasts produced using a univariate method. Several such methods exist, but here again we are limited by the length of time series available – at most 20 years – which precludes more complex forms of forecasting such as Box-Jenkins methods (Chatfield, 1996; 75).

Given the need to be able to produce univariate forecasts from fairly short time series, work done at Yorkshire Region concluded that the method most appropriate is Holt's Exponential Smoothing (Haward, 1993). This method applies a linear trend to the data, but gives more weight to recent points than to points at the beginning of the time series. Hence the forecasts 'start from where we are'.

It is important to give an indication of the precision of forecasts by giving prediction intervals, which are simply confidence intervals for the forecasts. The intervals are calculated such that, if current trends continue, there is less than a 5% chance that the rate will be outside the 95% prediction interval. There are instances where the prediction intervals are extremely wide. This simply indicates that, because the past observations are so widely scattered, there is no way of knowing whether the rates will go up or down or by how much. In a few cases, particularly where the most recent point is out of kilter with previous values, the forecasts can appear anomalous. The forecasts in this report were all produced in Forecast Pro XE (Business Forecast Systems Inc., 1999).

In order to preserve the mathematical integrity of the forecasting models, the rates are transformed prior to forecasting using the logit function [$\text{logit}(p) = \ln(p/(1-p))$] which, amongst other things, ensures that the rates do not go below zero or above one (100%). For rates close to zero this is effectively the same as log-transforming the rates, but for rates which approach one (such as percentage coverage of cervical cytology programmes) the logit transformation is necessary.

A2.4 Targets

The targets are stated in varying forms in the HImP documents, but are presented here in one of two ways. Firstly, and most simply, where the target is to maintain a certain standard (eg. 80% coverage of cervical screening programme) the target is simply shown as a red horizontal line representing an ongoing standard. Where there is a target reduction or increase (either a percentage reduction or a target figure) these have been translated into a constant rate of change between the baseline period and the target deadline.

For example, the target 'to reduce mortality by at least a fifth by 2010 from a baseline of 1995-7' can be stated as a reduction of 1.58% each year from 1996 to 2010. The baseline is calculated as the average of the 1995, 1996 and 1997 figures. Hence these targets are shown as a constant rate of decline or increase. At any point in the future new data may be compared with the relevant point on the target curve to assess progress being made.

References.

Altman DG (1991). *Practical Statistics for Medical Research*. London: Chapman & Hall.

Breslow NE & Day NE (1987). *Statistical Methods in Cancer Research; Volume II – The Design and Analysis of Cohort Studies (IARC Scientific Publications No 82)*. Lyon: International Agency for Research on Cancer.

Business Forecast Systems Inc. (1999). *Forecast Pro.* Belmont, MA.

Chatfield C (1996). *The Analysis of Time Series: An Introduction*. London: Chapman & Hall.

Haward RA (1993). *'The Health of the Nation' Targets – First Steps. Public Health Report: Statistical Review.* Harrogate: Yorkshire Regional Health Authority.

Townsend P, Phillimore P, Beattie A (1988). *Health and Deprivation. Inequality and the North*. London: Croom Helm.