

ONE NUMBER CENSUS STEERING COMMITTEE**Transformation of the Hard to Count variables**

1. Paper ONC(SC)00/15 recommended the methodology for deriving the Hard to Count (HtC) index for use in the 2001 One Number Census (ONC). It was suggested at the June meeting of the Steering Committee that the variables used to derive the Hard to count index should be normalised.
2. This paper explains the rationale behind not using normalisation, and explores whether standardising the variables provides a better prediction of coverage. The paper concludes that as the distribution is not changed greatly, standardisation should not be used.

The Steering Committee are asked to note the paper.

**Owen Abbott
Census Division
Office for National Statistics
Room 4200W
Segensworth Road
Titchfield
Fareham
HANTS
PO15 5RR**

January 2001

Transformation of the Hard to Count variables

Introduction

1. Paper ONC(SC)00/15 recommended the methodology for deriving the Hard to Count (HtC) index for use in the 2001 One Number Census (ONC). It was suggested at the June meeting of the Steering Committee that the variables used to derive the Hard to count index should be normalised. This paper explains the rationale behind not using normalisation, but as an alternative explores whether standardising the variables provides a better prediction of coverage.

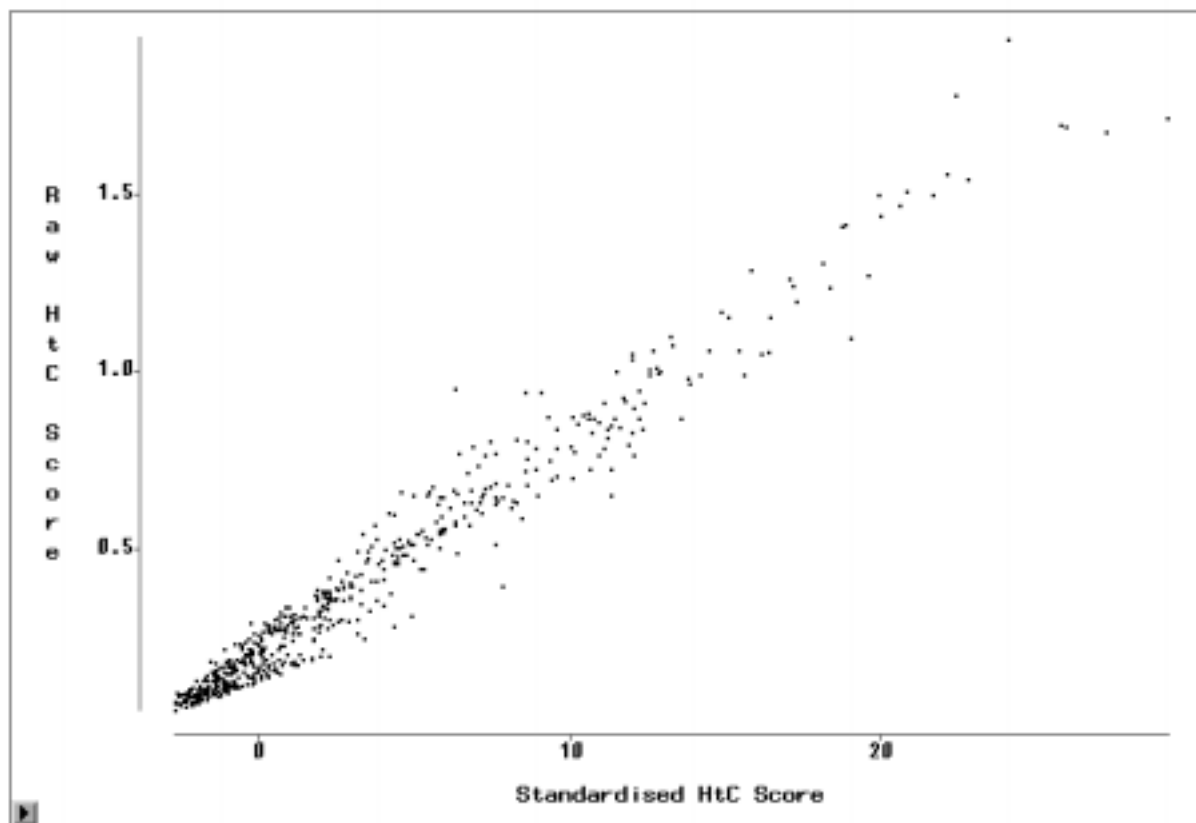
Normalisation

2. A discussion paper arising from the December 1998 meeting of the ONC Project Board suggested that normalising the variables that make up the Hard to Count index has the unwanted consequence of not identifying some hard to count areas. This is mainly because all of the variables are extremely skewed, and it is this skewness that should determine the hardest to count areas. For instance, take two EDs:
 - ED A - An ED which is at the median on all indicators; and;
 - ED B - An ED which is very high on one variable but low on the others.
3. A normalised HtC score would result in ED A being in the middle of the index even though its values are very low for all indicators (given the skewness). Although ED B has more hard to count people, it will be lower on the index than ED A because its low values on most indices balance the high value on one variable.
4. The skewness of the variables is their most important feature, as it highlights those areas that have large populations of hard to count people. Therefore it makes sense to maintain the geographical skewness of the variables. The normalisation process should therefore not be used for this reason.

Standardisation

5. Other potential transformations include standardisation and centralisation. Of these, only standardisation of the variables by subtracting the mean and dividing by the standard deviation will be considered. This is because centralisation will not alter the relative weights of the variables since the HtC score is a simple sum of each variable.
6. To investigate whether standardisation makes a big difference, a HtC score for all 1991 EDs in England and Wales was derived using the standardised proportions of the HtC variables. This was then compared with the score derived from the raw proportions. **Figure 1** (overleaf) shows the scatterplot of the two scores. It shows that the standardisation process has not greatly altered the score distribution.

Figure 1 -Scatterplot of standardised HtC score and non-standardised HtC score



7. The main differences are due to the decreased weight of the rented and language difficulty variables, as these have the highest standard deviations and means. The number of EDs that changed to the hardest to count group from the not so hard to count group as a result of the standardisation was 2156 which is roughly 10% of the strata. This is shown in Annex A. However, there is no clear way of evaluating whether this provides a better hard to count index.

Conclusions and recommendation

8. The first part of this paper explained why normalisation of the variables is not sensible. The second examined whether standardisation of the variables has a big effect on the resulting HtC score. Because the score distribution has not been altered greatly, there will be little difference in the resulting HtC indexes. It therefore does not seem sensible to overcomplicate the methodology for deriving the index if there is little change.
9. **It is therefore recommended that the variables used to derive the Hard to Count index are neither normalised nor standardised.**

**Owen Abbott
One Number Census
ONS**

July 2000 (Annex A added September 2000)

ANNEX A

Table A1 - Number of EDs that changed between strata

Non Standardised Index	Standardised Index		
	1	2	3
1	36636	7021	0
2	7021	34480	2156
3	0	2156	19673