2015/16 National Tariff Payment System: A consultation notice

Annex 5b: Data cleansing method

26 November 2014

Annex 5b: Data cleansing method

Reference costs cleansing methodology

Admitted Patient Care (APC) 2011/12 reference costs data were cleansed using this methodology:

1. Outliers were removed from the raw reference cost data set based on the Grubbs method, also known as the 'maximum normed residual test'.[1]

   The Grubbs test is defined for the hypothesis:

   - H_0: the sample doesn't have outliers

   - H_1: the sample has at least one outlier

   The Grubbs score is calculated using the following formula:

   $$G = \frac{\max |X_i - \mu|}{sd}$$

   where G is the Grubb's score; X_i is the unit cost after the Market Forces Factor (MFF) is removed in a specific sample, μ is the sample mean and sd is the standard deviation.

   The outliers are then identified (and subsequently removed) by comparing the Grubbs score for each observation in the sample with the Grubbs critical value for the sample.[2]

   The test detects one outlier at a time. This outlier is temporarily deleted from the dataset and the test is reiterated until no outliers are detected. This test is undertaken across the natural logarithm of the unit cost after the MFF is removed for each healthcare resource group (HRG) and department.[3]

2. Providers are removed if they submit reference costs that are more than 50% lower than the national average for more than 25% of the HRGs submitted, and at the same time submit reference costs that are 50% higher than the national average for more than 25% of the HRGs submitted for each HRG and department.

3. Providers are removed if they report reference costs that include more than 75% duplicate costs across HRGs and departments.

---

[1] For further explanation of the method see Grubbs F. (1969), Procedures for Detecting Outlying Observations in Samples, *Technometrics*, 11(1), 1-21; Stefansky W. (1972), Rejecting Outliers in Factorial Designs, *Technometrics*, 14, 469-479.

[2] The values of Grubbs statistic can be found at: pages.towson.edu/rsours/docs/210/Statistics_Tables.pdf

[3] The following departments are considered in APC: day case, elective inpatients, elective inpatients excess bed days, non-elective inpatients short stay, non-elective inpatients long stay and non-elective inpatients long stay excess bed days.