**Supplemental Digital Content 3**

**LOESS modelling**

**Background**

As the longitudinal demand for hospital services is intrinsically episodic and non-linear, we used non-parametric regression, based on locally weighted polynomial regression (LOESS), to provide an accurate estimate of mean hospital demand. LOESS is a fitting technique that does not specify a relationship between the dependent and independent variables prior to the analysis and focuses on the fitted curve. LOESS combines simple regression over small subsets of the data with nonlinear smoothing techniques similar to a least-squares method for data analysis.1 The LOESS smoothing technique additionally allows 95% confidence intervals to be simply generated using bootstrap techniques.

The major advantage of this method is that it does not require any assumptions about the parametric relationship between demand for hospital services and time, although it can be computationally intensive. Only through advances in computational power has it been possible to look at such large data sets this way.2 The technique is widely used in social sciences, environmental analysis and election and voting behaviour.1,3,4

**What is locally weighted polynomial regression (LOESS) ?**

LOESS is a kernel smoothing method that essentially fits local polynomial regressions and joins (smooths) them together.5  The regression fit at point x is made using points in the neighbourhood of x, weighted by their distance from x. Half the size of the smoothing window is called the span or bandwidth, which is defined based on a nearest neighbour strategy. As the span increases the estimated curve becomes smoother. LOESS models are fitted based on the following steps:

**Step 1:** Define the window width (determined by the span, α)

This defines the nearest neighbours (number or percentage of data observations) used in each local regression. The nearest neighbour method overcomes any problems that may be experienced due to sparse data. It is generally chosen based on a visual inspection of the final curve, taking into account the trade-off between bias and variance (see limitations of LOESS below ).6 The LOESS algorithm will select a series of focal points equally spaced across the range of x-values (in the hospital example these are equally spaced in time) for the steps below.

**Step 2:** Weight the data

The weight function gives the greatest weight to observations that are closest to each focal point x. For α<1, the neighbourhood includes proportion α of the points, and these have tricubic weighting (proportional to (1 - (distance/max\_distance)3)3). For α>1, all points are used, with the ‘maximum distance’ assumed to be α(1/p) times the actual maximum distance for p explanatory variables. In our analysis we did not adjust for additional covariates.

**Step 3:** Fit a local regression about each focal point x in the data

For each focal point x a fitted value is obtained using polynomial regression based on weighted least squares (with tricubic weights), using observations in the neighbourhood.

**Step 4:** Connect fitted focal points by line segments

Adjacent fitted focal points are connected by line segments. The fitted focal points are located close together along the x-axis (in the hospital demand example this is the time axis), so the series of line segments appears to be a smooth curve. The relationship between x and y can only be graphed.

**Example code for R** (free software for Statistical Computing)

The “loess” function in the R statistical software is simple to use. To generate and plot a fitted curve we used the following code, where “hospsum” is the total hospital usage on a given day, “day” is the time relative to the cardinal event and “pdaysum” is the total number of individuals in the database for a given day:

model <- loess(hospsum ~ day, span=1.4, weights=pdaysum)

day.range <- seq(1,4,by=0.1)

plot(day, hospsum)

lines(day, predict(model, data.frame(day=day.range)))

**Limitations of LOESS**

There is a balance between bias and variance in all LOESS fitted curves. This can be influenced by:

1. A relatively small span produces a curve characterised by short, rapid increases and decreases (or “noise”), meaning that there is insufficient data that falls within the neighbourhood for the local regression and results in large variance. In contrast, a relatively large span may oversmooth the curve, potentially resulting in bias if the local polynomial does not fit the data well.
2. Higher degree polynomials will have less bias as they provide a better estimate of the underlying mean, however, since there are more coefficients to estimate they also have higher variability. In general a lower order polynomial with an appropriate bandwidth will provide the best trade-off between bias and variance.

**References**

1. Jacoby WG. Loess: a nonparametric, graphical tool for depicting relationships between variables. Electoral Studies 2000. 577-613
2. Simonoff JS (1993) Smoothing methods in statistics. Springer, Berlin
3. Schwartz J. Harvesting and Long Term Exposure Effects in the Relation between Air Pollution and Mortality. Am J Epidemiol 2000; 151:440-8.;
4. [Wang](http://link.springer.com/search?facet-author=%22Xiaogang+Wang%22) X, [Wang](http://link.springer.com/search?facet-author=%22Jiafeng+Wang%22) J, [Russell](http://link.springer.com/search?facet-author=%22Curtis+Russell%22) C, [Proctor](http://link.springer.com/search?facet-author=%22Paul+Proctor%22) P, [Bello](http://link.springer.com/search?facet-author=%22Richard+Bello%22) R, [Higuchi](http://link.springer.com/search?facet-author=%22Kaz+Higuchi%22) K, [Zhu](http://link.springer.com/search?facet-author=%22Huaiping+Zhu%22) H. Clustering of the abundance of West Nile virus vector mosquitoes in Peel Region, Ontario, Canada. [Environmental and Ecological Statistics](http://link.springer.com/journal/10651). January 2014;
5. Cleveland, William S.; Devlin, Susan J. (1988). "Locally-Weighted Regression: An Approach to Regression Analysis by Local Fitting". Journal of the American Statistical Association 83 (403): 596–610.
6. Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome. The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Second Edition). Springer (2009), New York. ISBN: 978-0-387-84857-0