Regional multilevel modelling of under-5 mortality in countries with few data

Richard J Silverwood & Simon Cousens

April 2009

Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT

E-mail: Richard.Silverwood@lshtm.ac.uk

Introduction

Child mortality has been the focus of a great deal of attention in recent years due to the Millennium Development Goal 4 (MDG4) which states the aim to reduce under-5 mortality by two-thirds between 1990 and 2015 [1]. The requirement for national and global estimates of current levels and trends in under-5 mortality with which to assess progress towards MDG4 has in turn spawned much methodological debate.

Recently, Murray et al [2] proposed a new approach to the estimation of under-5 mortality, since used by others [3], which aims to be reproducible and transparent. This approach uses data derived from different sources (e.g. survey data, vital registration data) at different time points, and applies loess regression to obtain smoothed estimates of past levels of child mortality and to predict future levels based on past and recent trends.

The method proposed by Murray et al is described in more detail elsewhere [2], but is briefly reviewed here. The general approach is to use loess regression [4, 5], a type of local regression, to model under-5 mortality within a given country. The basic linear model underlying the loess approach is

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \tag{1}$$

where y_i is log under-5 mortality, x_i is year and $\varepsilon_i \sim N(0, \sigma_\varepsilon)$ is an error term. Although the original formulation of the model also includes a term denoting vital registration data, this is ignored here for simplicity as our example includes no vital registration data.

The loess function is fitted using weighted least squares regression, with the weights corresponding to each observed under-5 mortality value calculated using a separate weighting function. This weighting function is tuned by a single parameter, α. By fitting the loess model using a variety of values of α, it is possible to vary the sensitivity of predictions to recent data trends.

For each time point of interest the loess model is fitted for each of a pre-determined set of α values, and the multivariate normal distribution defined by the estimated regression coefficients and their variance-covariance matrix is used to simulate 1000 random draws for the parameters in (1). For each of these sets of random draws under-5 mortality is estimated. The 1000 estimates per α value are then pooled across the set of α values, and the final

estimated under-5 mortality calculated as the median value, with an uncertainty interval corresponding to the 2.5^{th} and 97.5^{th} centiles.

In countries where data are plentiful this approach may provide appropriate estimates of under-5 mortality, but in countries where there are few data the reliability of the estimates may be questioned. The scarcity of data means that only larger α values can be used, which means that the estimates are effectively assumed to follow a linear trend (i.e. a loglinear trend on the original scale) across all time points considered. Additionally, in countries with fewer data it seems intuitive that there is greater uncertainty surrounding any estimates, but the reduced set of α values often results in very narrow uncertainty intervals if the data display little variability.

We consider an extension to Murray et al's loess-based approach for use in countries with few data. For a given country with few data (the 'target country'), we investigate the use of data from additional countries which are deemed similar to the target country in a multilevel extension of the loess model (the 'regional multilevel loess model'). Country-specific under-5 mortality estimates may still be obtained, but now the estimates for the target country will be formed by 'borrowing' information from the additional countries in the model.

We describe this method in more detail and illustrate its use in Congo, where only 12 data points from a single survey (DHS 2005) are available, using two different approaches to the selection of the set of additional countries. We additionally perform sensitivity analyses to assess the extent to which the under-5 mortality estimates obtained for the target country are dependent on the set of additional countries included in the regional multilevel loess-based approach.

Although a loess-based approach to under-5 mortality estimation has recently been proposed, alternative smoothing methods could be employed within the same regional multilevel framework. We thus also investigate the use of penalised splines.

Methods

Our proposed extension to the loess-based approach of Murray et al involves using the data for several other 'similar' countries in addition to the data for the target country itself. These countries should be chosen so that their past, current and, ideally, future trends and levels of child mortality are similar to those in the target country. The data can then be considered as a hierarchy, with individual data points nested within countries.

Hierarchical data of this nature can be described through the use of multilevel models [6]. The linear model underlying the conventional loess-based approach (1) can be extended to become a random slopes and intercepts model,

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0i} + u_{1i} x_{ij} + \varepsilon_{ij} \tag{2}$$

where

$$\begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u0}^2 & \sigma_{u01} \\ \sigma_{u01} & \sigma_{u1}^2 \end{pmatrix} \right) \text{ and } \varepsilon_{ij} \sim N(0, \sigma_\varepsilon),$$

with y_{ij}, x_{ij} and z_{ij} now the jth values within the ith country. This formulation allows both the intercept and the slope of the fitted model to differ between countries. The multilevel model can again be fitted using a weighted approach via (restricted) maximum likelihood. By using the same weighting function as for the conventional loess model, (2) becomes the 'multilevel loess model'.

For a given time point x the estimated average under-5 mortality across all included countries (the 'population average' estimate) y^{PA} can be seen from (2) to be

$$y^{PA} = \hat{\beta}_0 + \hat{\beta}_1 x, \tag{3}$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimates of the fixed effects β_0 and β_1. y^{PA} has variance

$$\text{var}(y^{PA}) = \text{var}(\hat{\beta}_0) + x^2\text{var}(\hat{\beta}_1) + 2x\text{cov}(\hat{\beta}_0, \hat{\beta}_1). \tag{4}$$

Similarly, the country-specific estimate for country i is

$$y_i^{CS} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{u}_{0i} + \hat{u}_{1i}x, \tag{5}$$

where $\hat{u}_{0i}$ and $\hat{u}_{1i}$ are estimates of the random effects u_{0i} and u_{1i}. y_i^{CS} has variance

$$\begin{aligned} \text{var}(y_i^{CS}) \quad = \quad & \text{var}(\hat{\beta}_0) + x^2\text{var}(\hat{\beta}_1) + 2x\text{cov}(\hat{\beta}_0, \hat{\beta}_1) + \text{var}(\hat{u}_{0i}) + x^2\text{var}(\hat{u}_{1i}) + 2x\text{cov}(\hat{u}_{i0}, \hat{u}_{i1}) \\ & +2\left(\text{cov}(\hat{\beta}_0, \hat{u}_{0i}) + \text{cov}(\hat{\beta}_0, \hat{u}_{1i}) + \text{cov}(\hat{\beta}_1, \hat{u}_{0i}) + \text{cov}(\hat{\beta}_1, \hat{u}_{1i})\right). \end{aligned} \tag{6}$$

It seems reasonable to expect that an increase in the fixed (or random) intercept would be associated with a decrease in the random (or fixed) slope, and that the two intercept terms and the two slope terms would be similarly negatively related. This would lead to each of the covariances between fixed and random effects ($\text{cov}(\hat{\beta}_0, \hat{u}_{0i})$, $\text{cov}(\hat{\beta}_0, \hat{u}_{1i})$, $\text{cov}(\hat{\beta}_1, \hat{u}_{0i})$ and $\text{cov}(\hat{\beta}_1, \hat{u}_{1i})$) usually being negative, meaning that

$$\text{var}_{UB}(y_i^{CS}) = \text{var}(\hat{\beta}_0) + x^2\text{var}(\hat{\beta}_1) + 2x\text{cov}(\hat{\beta}_0, \hat{\beta}_1) + \text{var}(\hat{u}_{0i}) + x^2\text{var}(\hat{u}_{1i}) + 2x\text{cov}(\hat{u}_{i0}, \hat{u}_{i1}) \tag{7}$$

will generally provide an upper bound for the variance of the country-specific estimate.

For each time point of interest and each given α value we therefore fit the multilevel loess model and independently use the multivariate normal distributions defined by i) the estimated fixed effects and their variance-covariance matrix, and ii) the predicted random effects and their variance-covariance matrix to simulate 1000 random draws for the fixed and random effects in (2). For each of these sets of random draws under-5 mortality is estimated. The 1000 estimates per α value are then pooled across the set of α values, and the final estimated under-5 mortality and its uncertainty interval calculated as in the conventional loess-based approach.

We illustrate the regional multilevel loess-based approach using data from Congo, where only 12 data points are available. We select the additional countries used in the model using two different approaches. In the first, we use the four countries which share a border with Congo (Cameroon, Central African Republic, Democratic Republic of the Congo and Gabon) — these are referred to as the 'geographical neighbours'. In the second, we use the four countries in sub-Saharan Africa which have the most similar under-5 mortality rates to Congo over the period for which data are available — these are referred to as the 'mortality neighbours'. More explicitly, the period of time for which under-5 mortality data are available for Congo is split into five year periods. Within each period the mean under-5 mortality in each country in sub-Saharan Africa is calculated. Each country is then given a ranking based on the proximity of their mean under-5 mortality to the mean in Congo (with 1 being the closest). These rankings are then averaged over all the five year intervals and the four countries with the lowest overall ranking used in the model.

The data used by Murray et al are available from the Institute for Health Metrics and Evaluation (IHME) website [7], and we use these data in our analyses. Within each country we exclude the same data points as were excluded by Murray et al, though for many of these it is unclear what the precise basis for exclusion was. Again to maintain comparability with Murray et al, we exclude any α values which result in an annual rate of decline in under-5 mortality more than 3 standard deviations from the mean rate across all countries, as reported in their paper [2].

We additionally perform sensitivity analyses to assess the extent to which the under-5 mortality estimates obtained for the target country are dependent on the set of additional countries included in the regional multilevel loess-based approach. To do this we exclude each country in turn from the model, re-estimate under-5 mortality in Congo at a set of time points using the restricted dataset, and compare the results.

We also investigate the use of penalised splines [8] as an alternative modelling approach to loess for use within the same regional multilevel framework.

All analyses are conducted using R version 2.6.1 [9].

Results

Geographical neighbours

Table 1 summarises under-5 mortality data availability in each country included in the regional multilevel loess-based analysis using geographical neighbours. Congo, the target country in the model, can be seen to have fewer data points than the other countries, derived from a single survey and covering a relatively narrow interval of time. This scarcity of data in Congo may lead to questionable estimates of under-5 mortality if the conventional loess-based approach of Murray et al is employed.

Country	Number of data points	Range of years with available data
Congo	12	1982–2004
Cameroon	64	1963–2003
Central African Republic	27	1960–1996
Democratic Republic of the Congo	16	1970–2003
Gabon	17	1977–1999

Table 1: Summary of under-5 mortality data availability in each country included in the regional multilevel loess-based analysis using geographical neighbours.

The under-5 mortality data for each country are plotted in Fig. 1. Although there is much variability in the data points, those for Congo generally lie within the range of those for the surrounding countries.

We fit the regional multilevel loess model as described above. The estimated population average curve across all countries and its uncertainty interval, calculated based on (3) and (4), are plotted in Fig. 2. The population average curve is seen to follow a similar trend to that evident in most countries, though there is a wide uncertainty interval due to between-country variability.

Fig. 3 shows the estimated country-specific under-5 mortality curves for each country included in the model. It can be seen that in regions of each plot where data are available the estimated curves fit the data well. When data are not available in a given country the curves more closely resemble the population average curve, illustrating that information from other countries is being used to inform the country-specific trends.

The estimated country-specific under-5 mortality curve for Congo in Fig. 3 is reproduced in Fig. 4 along with the corresponding uncertainty interval upper bounds calculated based on (6). Although the upper bounds for the uncertainty interval are relatively wide across the entire range of the plot this is particularly so for periods during which there are no data available and hence the estimates are largely informed by data from other countries.

Also shown in Fig. 4 are the equivalent estimates and uncertainty intervals from the conventional loess-based approach of Murray et al [2]. The conventional estimates can be seen to follow an approximately linear trajectory due to the scarcity of data meaning that only larger α values can be used. Although both the conventional and regional multilevel estimates show similarly increasing under-5 mortality during the middle of the period for which empirical estimates are available (1982 to 2004), outside this period there are marked differences. Prior to 1982 the conventional

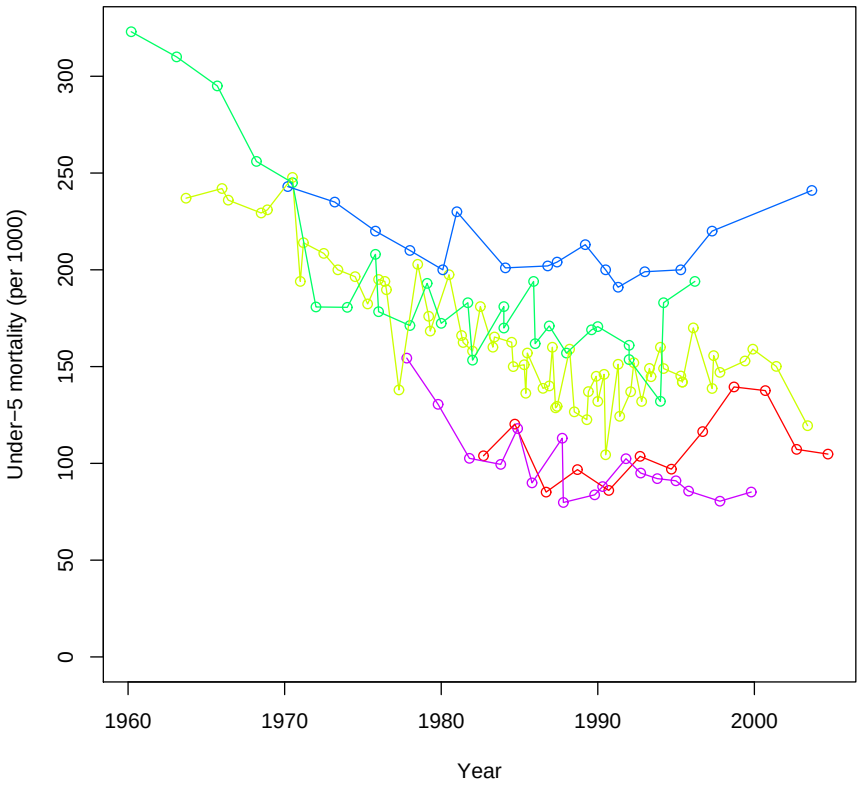

Fig. 1: Under-5 mortality data for Congo and geographical neighbours. Red points and lines represent Congo (the 'target country'), yellow represent Cameroon, green represent Central African Republic, blue represent Democratic Republic of the Congo and purple represent Gabon.

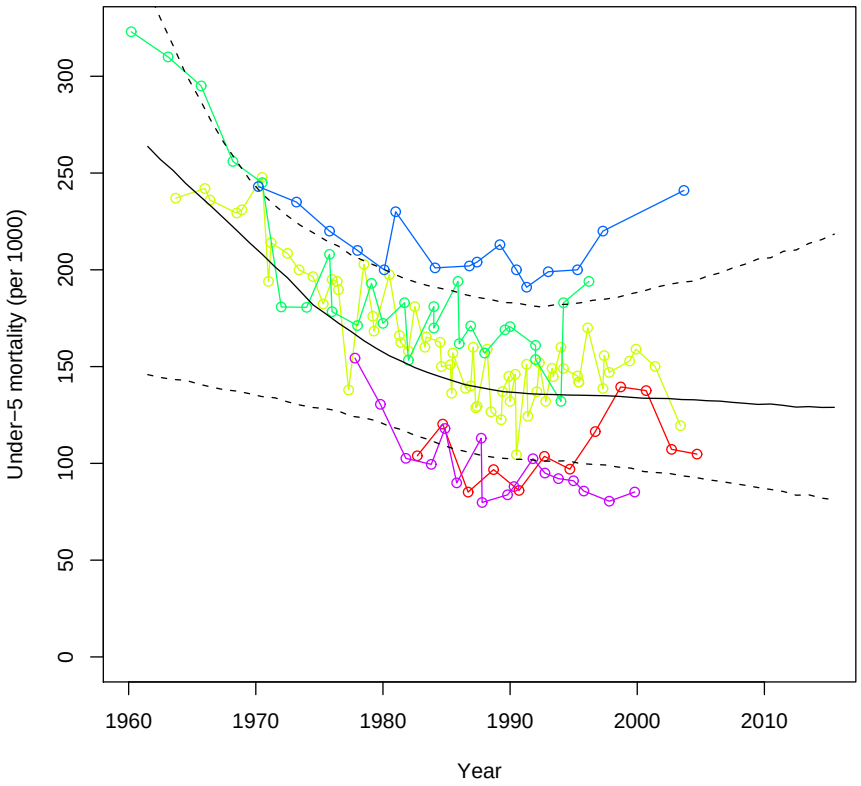

Fig. 2: Estimated under-5 mortality population average curve (solid black line) and uncertainty interval (dashed black lines) for Congo and geographical neighbours using regional multilevel loess-based approach. Other points and lines are under-5 mortality data for Congo and geographical neighbours as described in Fig. 1.

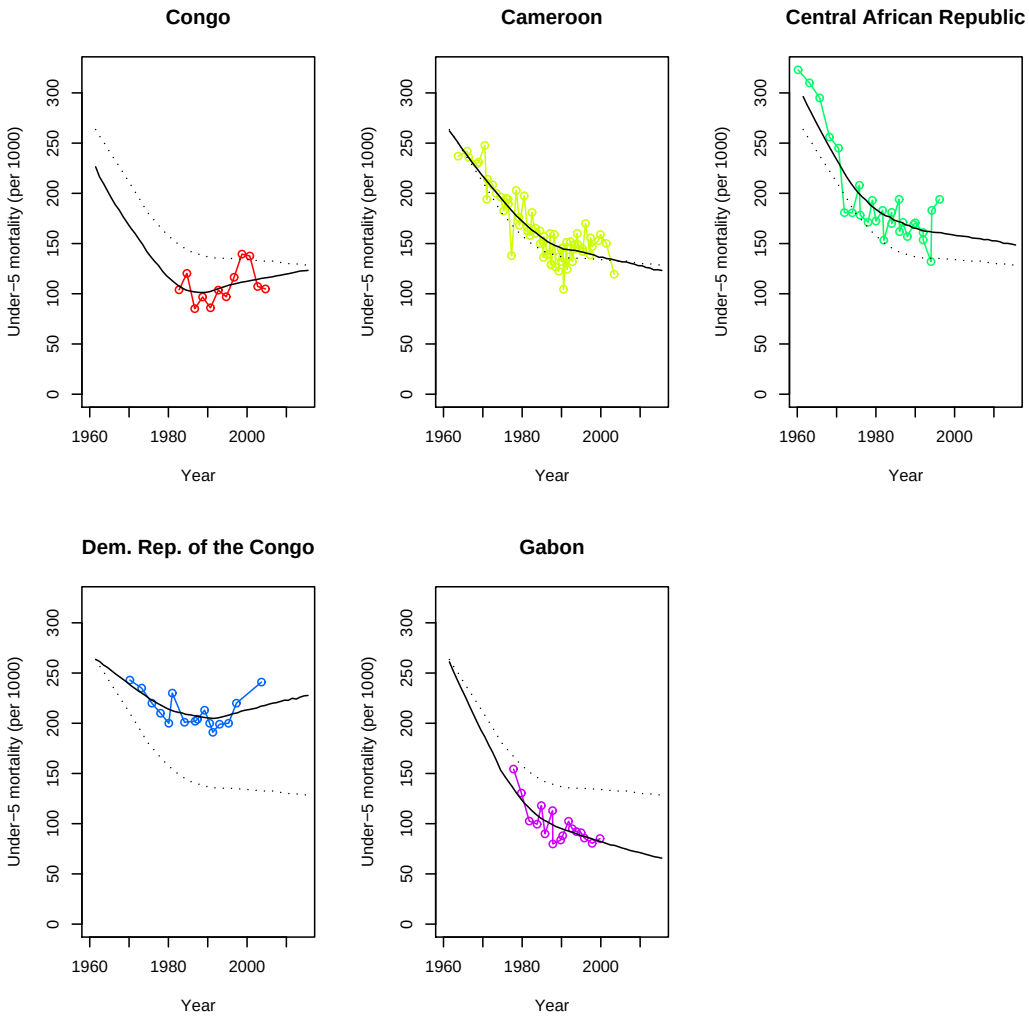

Fig. 3: Estimated country-specific under-5 mortality curves (solid black lines) and population average curve (dashed black line) for Congo and geographical neighbours using regional multilevel loess-based approach.

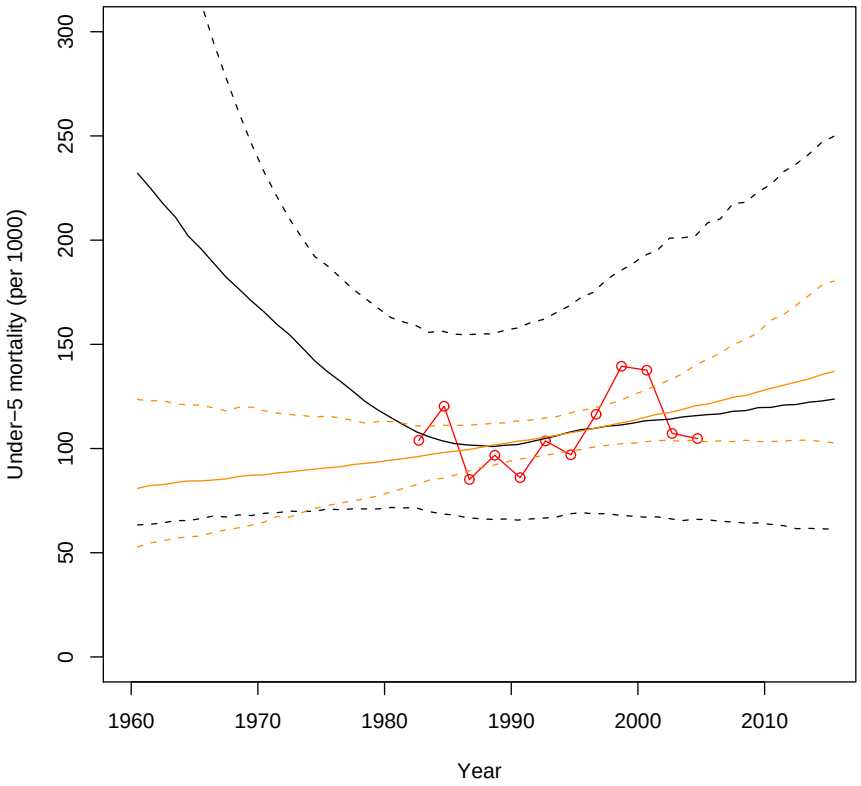

Fig. 4: Estimated under-5 mortality curves (solid lines) and uncertainty intervals/uncertainty interval upper bounds (dashed lines) for Congo using regional multilevel loess-based approach with geographical neighbours (black lines) and conventional loess approach (orange lines). Red points and lines are under-5 mortality data for Congo.

loess-based estimates show steadily increasing under-5 mortality — a back-extrapolation of the trend seen in the data for Congo – whereas the regional multilevel loess-based estimates suggest decreasing under-5 mortality, as was evident in other countries in the region (see Fig. 2). Post-2004, whilst both sets of estimates show increasing under-5 mortality, the more rapid increase in the the conventional loess-based estimate is again a continuation of the trend seen in the data for Congo while the smaller increase in regional multilevel loess-based estimates is closer to the relatively stable trend seen on average in the region (see Fig. 2). However, it must be remembered that not only do the sets of estimates differ in terms of the statistical models used: the increased number of data points used in the multilevel approach means that a much wider range of α values can be used — the minimum α value used for the conventional loess-based approach is 1.70, whereas for the multilevel equivalent it is 0.45. This leads to the conventional estimates being forced to lie on a near-linear trajectory.

The uncertainty intervals/upper bounds for uncertainty intervals also clearly differ between the approaches. The uncertainty intervals for the conventional loess-based approach are very narrow, especially within the period for which data are available. This is again partly due to the restrictive set of α values used. The upper bounds for the regional multilevel loess-based uncertainty intervals on the other hand are very wide, including many possible values for a given point in time. It must be remembered, however, that i) they incorporate uncertainty not just relating to the data in Fig. 4 but also to the fitting of the population average curve, and ii) they are upper bounds.

Table 2 more explicitly compares the estimates and uncertainty intervals obtained using the conventional and regional multilevel loess-based approaches. Although the regional multilevel loess-based estimate is almost twice the conventional loess-based estimate in 1970, at time points within the range of years for which data are actually available for Congo (1990 and 2000) the two sets of estimates are almost identical. When comparing the uncertainty interval widths between the two approaches it is preferable to consider the *relative* uncertainty interval widths (i.e. the uncertainty interval width as a percentage of the corresponding estimate) due to the different underlying trends in the point estimates. The increase in relative uncertainty interval width under the regional multilevel loess-based approach is sizeable at all time points considered, peaking at almost a five-fold increase in 1990 and 2000. However, it must again be remembered that the uncertainty intervals corresponding to the regional multilevel loess-based approach are upper bounds.

	Estimated under-5 mortality (per 1000) (UI)			
Year	Conventional loess	Regional multilevel loess	Ratio of estimates (multilevel:conventional)	Ratio of relative UI widths (multilevel:conventional)
1970	88 (65, 118)	165 (69, 236)	1.88	1.68
1980	95 (79, 112)	115 (72, 162)	1.21	2.26
1990	103 (95, 114)	103 (66, 156)	1.00	4.75
2000	115 (103, 128)	113 (69, 186)	0.98	4.77
2010	129 (103, 161)	120 (65, 224)	0.93	2.94
2015	137 (102, 182)	124 (64, 245)	0.91	2.50

Table 2: Comparison of estimated under-5 mortality in Congo using conventional loess-based approach and regional multilevel loess-based approach with geographical neighbours. UI, uncertainty interval.

It is also informative to compare the uncertainty intervals obtained under the two approaches with the calculated confidence interval corresponding to the original estimate of under-5 mortality from the DHS, which is 103–130 for the period 2001–5. This range of values is clearly closer to the conventional loess-based uncertainty interval than to

the regional multilevel equivalent in 2000. Although the data points used in fitting the models relate to only two year periods so will have somewhat wider confidence intervals, this is unlikely to make sufficient difference to affect this observation.

Table 3 presents the results from the sensitivity analysis in which each country in turn is excluded from the model and under-5 mortality in Congo is re-estimated at a set of time points using the restricted dataset. When comparing the results it must be borne in mind that as well as the data points used in the model fitting differing, the set of α values used may also change as a result. For example, when excluding Gabon the minimum α value used is 0.55, whereas when excluding Central African Republic it is 0.40.

Excluded country	Year	Estimated under-5 mortality (per 1000) (UI)	Ratio of estimates (Excluding given country:Using all data)	Ratio of relative UI widths (Excluding given country:Using all data)
Cameroon	1970	151 (65, 232)	0.92	1.10
	1980	112 (65, 177)	0.97	1.28
	1990	102 (57, 186)	0.99	1.45
	2000	114 (55, 231)	1.01	1.50
	2010	121 (51, 286)	1.01	1.46
	2015	125 (49, 322)	1.01	1.49
CAR	1970	167 (67, 264)	1.01	1.17
	1980	115 (66, 176)	1.00	1.23
	1990	101 (58, 177)	0.98	1.36
	2000	114 (56, 230)	1.01	1.49
	2010	120 (50, 282)	1.00	1.45
	2015	124 (46, 315)	1.00	1.49
DRC	1970	160 (70, 278)	0.97	1.29
	1980	114 (70, 161)	0.99	1.03
	1990	102 (66, 158)	0.99	1.03
	2000	113 (69, 189)	1.00	1.03
	2010	121 (66, 227)	1.01	1.00
	2015	125 (64, 250)	1.01	1.02
Gabon	1970	154 (61, 217)	0.93	1.00
	1980	110 (65, 165)	0.96	1.17
	1990	103 (66, 159)	1.00	1.03
	2000	113 (73, 177)	1.00	0.89
	2010	123 (77, 197)	1.03	0.74
	2015	128 (79, 209)	1.03	0.70

Table 3: Sensitivity analysis for estimated under-5 mortality in Congo using regional multilevel loess model with geographical neighbours. UI, uncertainty interval; CAR, Central African Republic; DRC, Democratic Republic of the Congo.

The point estimates are seen to be very robust to the exclusion of individual countries, with the exception of Cameroon and Gabon in 1970 when estimates are reduced by 7–8%. However, relative uncertainty interval widths

may be noticeably increased or decreased. For example, the exclusion of Cameroon or Central African Republic results in around a 50% increase in relative uncertainty interval width from 2000 onwards. As these countries both provide a large number of data points towards the centre of the distribution across all the countries included in the model (see Fig. 1) it seems reasonable that their exclusion increases uncertainty. The exclusion of Gabon, a country providing data points towards the extreme of the distribution across all countries, results in a reduction of relative uncertainty interval width of around 30% from 2010 onwards.

Mortality neighbours

The four countries in sub-Saharan Africa with under-5 mortality most similar to that in Congo over the period for which empirical estimates are available for Congo are found to be Gabon, Kenya, Lesotho and Swaziland. Table 4 summarises under-5 mortality data availability in each of these countries. Congo can again be seen to have fewer data points than the other countries.

Country	Number of data points	Range of years with available data
Congo	12	1982–2004
Gabon	17	1977–1999
Kenya	97	1955–2002
Lesotho	62	1953–2003
Swaziland	20	1950–1995

Table 4: Summary of under-5 mortality data availability in each country included in the regional multilevel loess-based analysis using mortality neighbours.

The under-5 mortality data for each country are plotted in Fig. 5. In general, there is relatively little between-country variability in the data points for a given year, especially if considered relative to the within-country variability seen in some countries.

The estimated population average curve across all countries and its uncertainty interval are plotted in Fig. 6. The population average curve is follows a similar trend to that evident in most countries, and the uncertainty interval is relatively narrow over the period where empirical estimates are available due to the low between-country variability.

Fig. 7 shows the estimated country-specific under-5 mortality curves for each country included in the model. The estimated curves are again seen to fit the data well in periods where data are available for that country and more closely resemble the population average curve in periods where data are not available. However, due to the low between-country variability in under-5 mortality, the differences between the country-specific curves and the population average curve are never great.

The estimated country-specific under-5 mortality curve for Congo in Fig. 7 is reproduced in Fig. 8 along with the corresponding upper bounds for the uncertainty interval. The upper bounds for the uncertainty interval are relatively narrow in the period where data are available for Congo, but rapidly become much wider outside this period.

Also shown in Fig. 8 are the equivalent estimates and uncertainty intervals from the conventional loess-based approach of Murray et al [2] and the estimates and upper bounds for the uncertainty intervals from the regional multilevel loess-based approach using geographical neighbours. The estimates using mortality neighbours are similar to those when using geographical neighbours, showing a fairly rapid decrease prior to the period where empirical estimates are available and a slow increase after this.

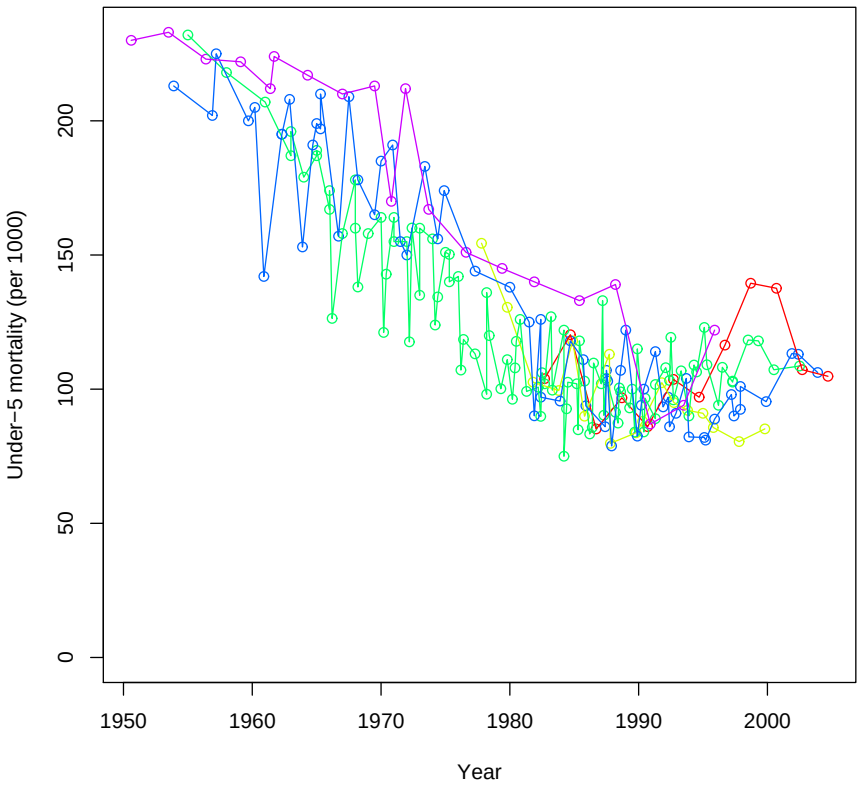

Fig. 5: Under-5 mortality data for Congo and mortality neighbours. Red points and lines represent Congo (the 'target country'), yellow represent Gabon, green represent Kenya, blue represent Lesotho and purple represent Swaziland.

The upper bounds for the uncertainty intervals when using mortality neighbours, whilst still wider than those when using the conventional approach across the majority of the plot, are much narrower than those when using geographical neighbours. In the middle of the period for which empirical estimates are available, the upper bounds for the uncertainty intervals when using mortality neighbours are a similar width to those when using the conventional approach.

Table 5 compares the estimates and uncertainty intervals obtained using the conventional loess-based approach and the regional multilevel loess-based approach with mortality neighbours. Whilst the regional multilevel estimate is markedly greater than the conventional estimate in 1970, at time points within the range of years for which empirical estimates available for Congo (1990 and 2000) the two sets of estimates are almost identical. The relative uncertainty interval widths are similar under the two approaches up to 1990, but from 2000 onwards there are more obvious differences.

Again, the confidence interval for 2001–5 from the DHS (103–130) is more similar to the conventional loess-based uncertainty interval than to the regional multilevel equivalent in 2000, though this is less marked than when using geographical neighbours.

Table 6 presents the results from the sensitivity analysis when using mortality neighbours. There is again some variability in the set of α values used in each model — when excluding Kenya the minimum α value used is 0.50, when excluding Swaziland it is 0.30.

The point estimates are seen to be even more robust to the exclusion of individual countries than when using geographical neighbours, with estimates generally differing by 1% or less and never more than 3%. Variability in relative uncertainty interval width is also reduced, though the exclusion of Kenya and Lesotho lead to large increases

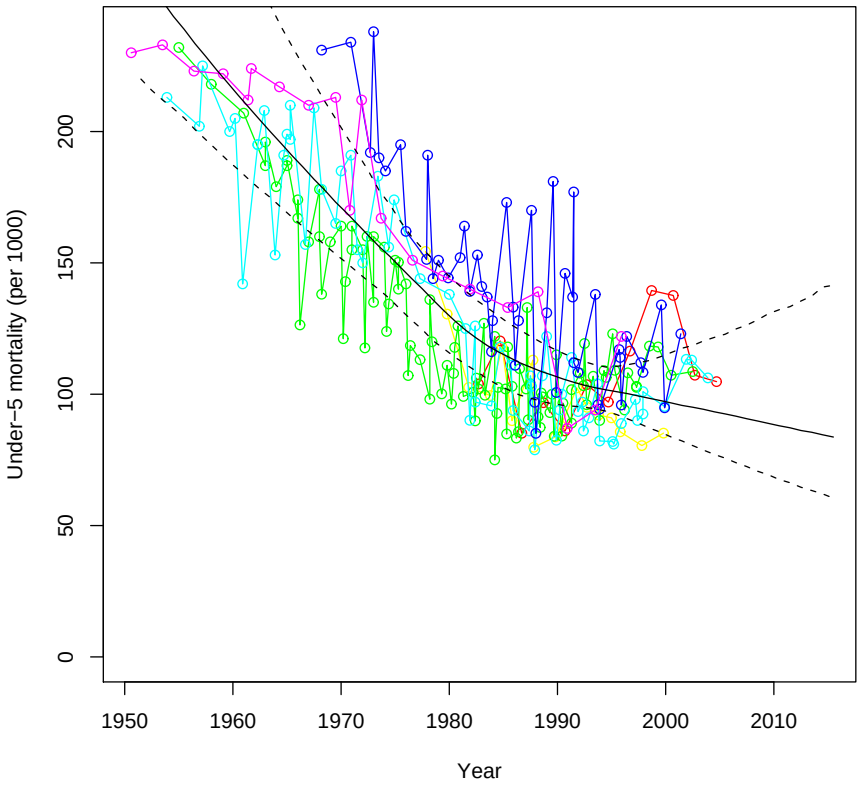

Fig. 6: Estimated under-5 mortality population average curve (solid black line) and uncertainty interval (dashed black lines) for Congo and mortality neighbours using regional multilevel loess-based approach. Other points and lines are under-5 mortality data for Congo and mortality neighbours as described in Fig. 5.

	Estimated under-5 mortality (per 1000) (UI)			
Year	Conventional loess	Regional multilevel loess	Ratio of estimates (multilevel:conventional)	Ratio of relative UI widths (multilevel:conventional)
1970	88 (65, 118)	154 (88, 184)	1.75	1.03
1980	95 (79, 112)	111 (85, 135)	1.17	1.29
1990	103 (95, 114)	103 (92, 111)	1.00	1.00
2000	115 (103, 128)	112 (91, 135)	0.97	1.77
2010	129 (103, 161)	122 (82, 173)	0.95	1.67
2015	137 (102, 182)	126 (77, 196)	0.92	1.62

Table 5: Comparison of estimated under-5 mortality in Congo using conventional loess-based approach and regional multilevel loess-based approach with mortality neighbours. UI, uncertainty interval.

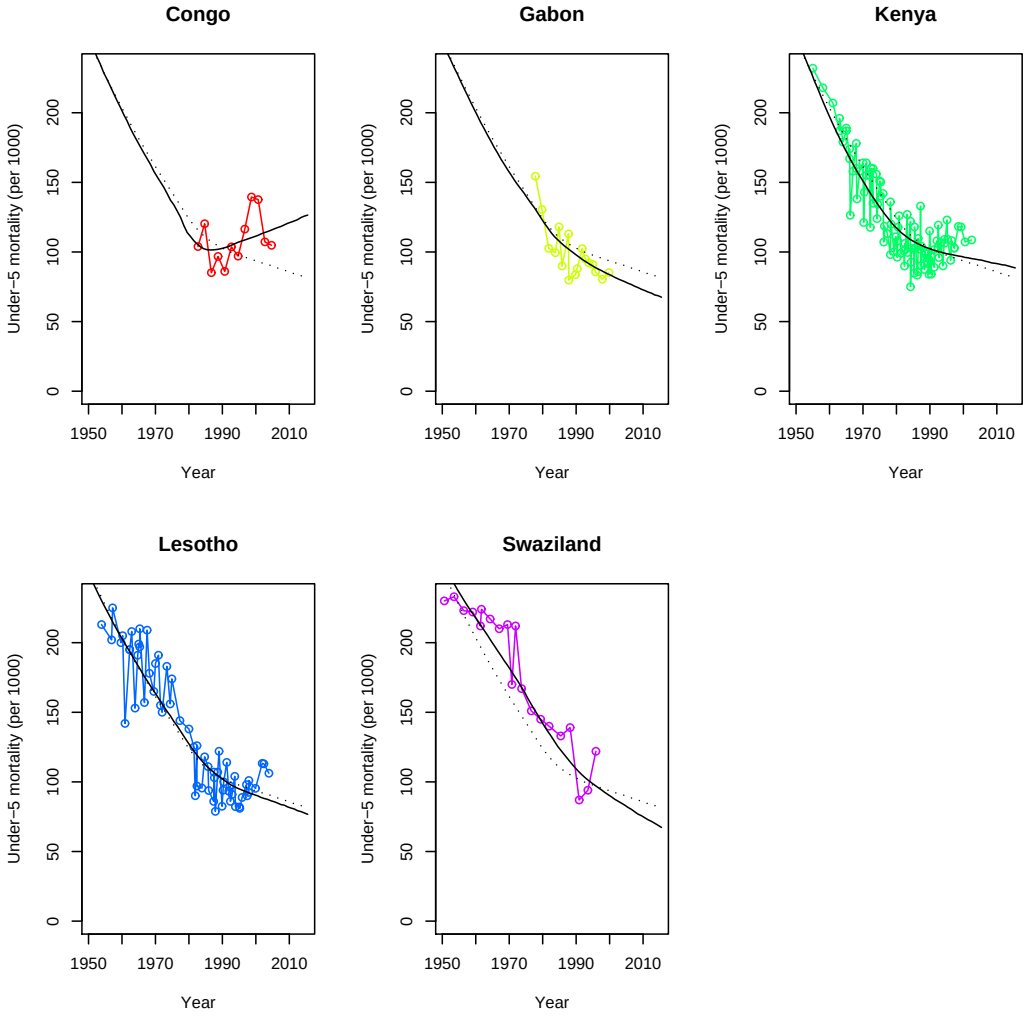

Fig. 7: Estimated country-specific under-5 mortality curves (solid black lines) and population average curve (dashed black line) for Congo and mortality neighbours using regional multilevel loess-based approach.

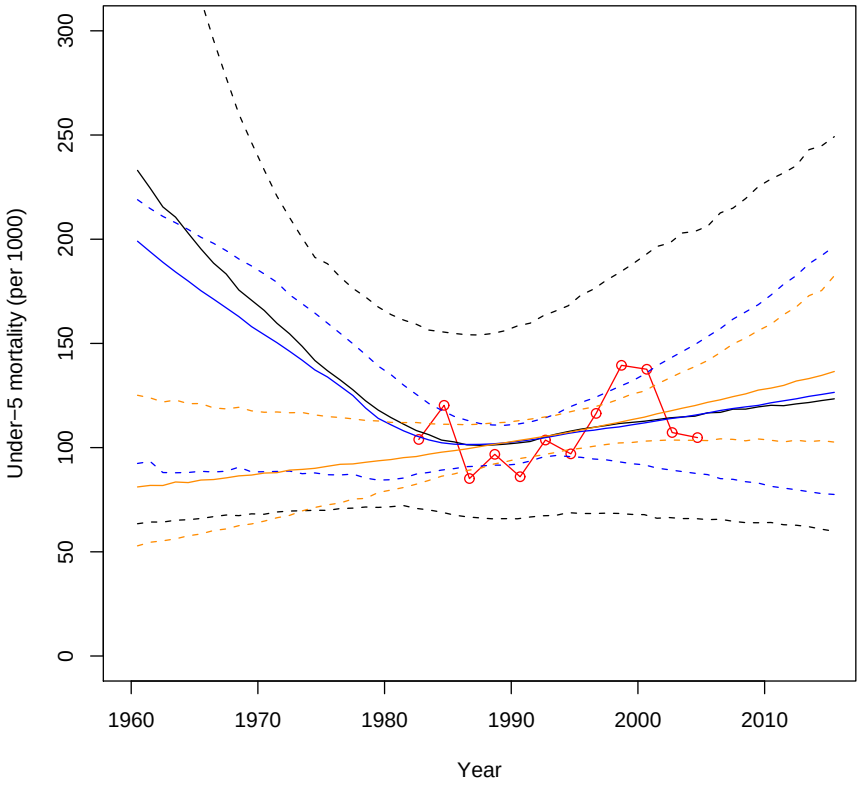

Fig. 8: Estimated under-5 mortality curves (solid lines) and uncertainty intervals/uncertainty interval upper bounds (dashed lines) for Congo using regional multilevel loess-based approach with geographical neighbours (black lines), regional multilevel loess-based approach with mortality neighbours (blue lines) and conventional loess-based approach (orange lines). Red points and lines are under-5 mortality data for Congo.

Excluded country	Year	Estimated under-5 mortality (per 1000) (UI)	Ratio of estimates (Excluding given country:Using all data)	Ratio of relative UI widths (Excluding given country:Using all data)
Gabon	1970	153 (79, 187)	0.99	1.15
	1980	110 (81, 137)	0.99	1.13
	1990	102 (92, 109)	0.99	0.94
	2000	112 (91, 134)	1.00	0.97
	2010	123 (81, 175)	1.01	1.01
	2015	130 (75, 202)	1.03	1.04
Kenya	1970	155 (76, 216)	1.01	1.45
	1980	112 (79, 137)	1.01	1.16
	1990	103 (90, 116)	1.00	1.39
	2000	112 (92, 137)	1.00	1.03
	2010	122 (83, 178)	1.00	1.04
	2015	127 (78, 206)	1.01	1.07
Lesotho	1970	152 (79, 189)	0.99	1.16
	1980	110 (81, 137)	0.99	1.13
	1990	103 (89, 115)	1.00	1.39
	2000	112 (92, 138)	1.00	1.05
	2010	122 (83, 180)	1.00	1.07
	2015	127 (80, 206)	1.01	1.05
Swaziland	1970	151 (76, 180)	0.98	1.11
	1980	112 (86, 131)	1.01	0.88
	1990	104 (93, 110)	1.01	0.89
	2000	113 (90, 142)	1.01	1.18
	2010	119 (81, 176)	0.98	1.07
	2015	122 (76, 197)	0.97	1.05

Table 6: Sensitivity analysis for estimated under-5 mortality in Congo using regional multilevel loess model with mortality neighbours. UI, uncertainty interval; CAR, Central African Republic; DRC, Democratic Republic of the Congo.

in certain years.

Alternative approaches

Alternative smoothing approaches to loess could also be utilised within the regional multilevel framework. One such approach is penalised splines [8]. Briefly, penalised splines are a means of modelling the relationship between two (or more) variables without the imposition of a rigid parametric form (for example linear or higher order polynomial). They are formed by the piecewise connection of polynomial curves between a series of 'knots' subject to certain penalties to encourage smoothness. The fusion between parametric mixed modelling and smoothing approaches such as penalised splines is referred to as 'semiparametric mixed modelling'.

We illustrate this approach in Congo using the set of geographical neighbours. Fig. 9 shows fitted linear, quadratic and cubic regional multilevel penalised splines for Congo. Each model is fitted using five knots, a reasonable number given data availability. Knot locations are spread evenly throughout the time points at which under-5 mortality data are observed across all five countries combined, so that knots are placed at the $\left(\frac{1}{6}\right)^{\text{th}}, \ldots, \left(\frac{5}{6}\right)^{\text{th}}$ quantiles of the unique time points.

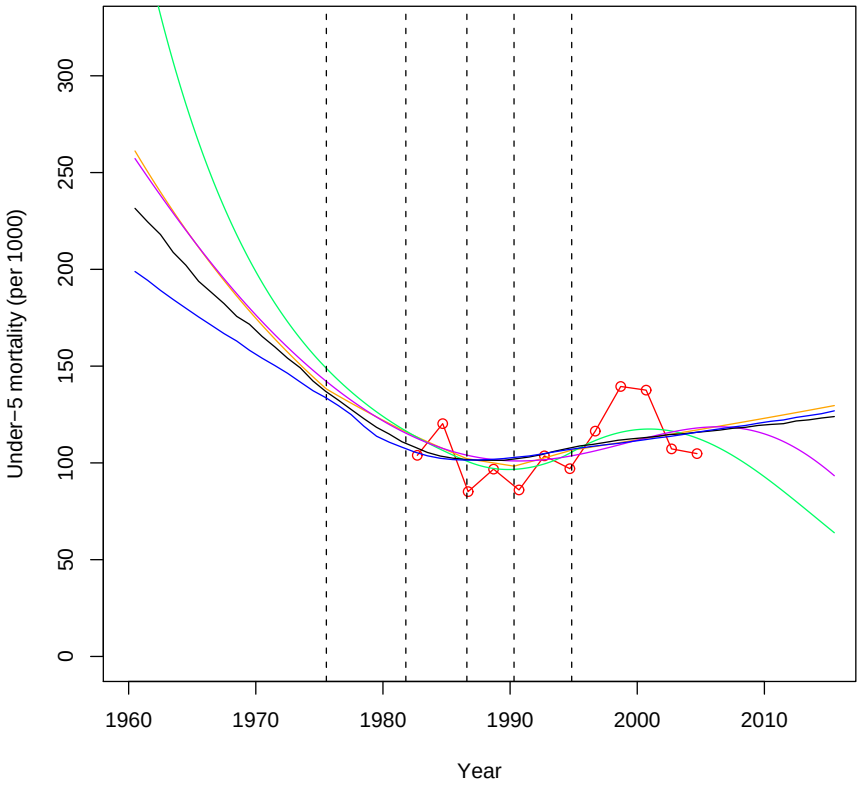

Fig. 9: Estimated under-5 mortality curves for Congo using linear (orange), quadratic (green) and cubic (purple) regional multilevel penalised splines, and using regional multilevel loess-based approach with geographical neighbours (black) and mortality neighbours (blue). Red points and lines are under-5 mortality data for Congo. Vertical dashed black lines are knot locations for penalised splines.

All three penalised spline models can be seen to provide a good fit in the region of the plot where data are available for Congo. Also included in the plot are the fitted curves from the equivalent loess-based approach using both geographical and mortality neighbours. Prior to the region of the plot where data are available for Congo the linear and cubic penalised spline models are relatively similar to the loess-based version using geographical neighbours. In this region of the plot the models are informed by the data from the other four countries so the fits are unlikely to

be too dissimilar. In the region of the plot after the last data point for Congo, however, there are no data points for the other countries either, meaning that the fitted curves are extrapolations. The result of this differs depending on the order of the model: as the linear version is constrained to be linear in this region the same trend as is observed in the later data points is continued; as the quadratic and cubic versions are constrained to be quadratic and cubic in this region they both veer off and would very soon approach zero. These are examples of the problematic 'edge effects' associated with polynomial (and, in this case, polynomial-based) modelling and illustrate why use of this type of model to predict values in the future is not recommended. As the loess-based approach is linear (within each α value) it is no surprise that it follows a similar trajectory to the linear penalised spline. The latter model, however, has the disadvantage of not being smooth at each knot. A further disadvantage of the regional multilevel penalised spline approach is the reduced tractability with regards to the calculation of uncertainty intervals.

Regional multilevel penalised spline models could also be fitted for Congo using the set of mortality neighbours, but the estimates would change little from those seen when using geographical neighbours and the edge effects would remain.

Discussion

We present a multilevel extension to the loess-based approach of Murray et al for use in countries with few data. We illustrate this method by using it to calculate under-5 mortality estimates for Congo, where only 12 data points are available, using two different approaches to the selection of the set of additional countries to use in the model. The resultant estimates are similar to those obtained under the conventional loess-based approach for time points within the period where empirical estimates are available for Congo, but differ markedly outside this period. Although the point estimates are largely similar across the two different approaches for selecting additional countries, the uncertainty intervals differ greatly. When using mortality neighbours, the uncertainty is similar to under the conventional loess-based approach within the period for which empirical estimates are available, but somewhat greater outside this period. When using geographical neighbours, uncertainty is vastly increased at all time points. Sensitivity analyses show the regional multilevel loess-based point estimates to be largely robust to the exclusion of individual countries, particularly when using mortality neighbours, though the associated uncertainty is less so.

Although our proposed method appears to be a useful extension to the original approach for countries where data are few, there remain important issues to consider. The use of upper bounds for the uncertainty intervals arises from difficulties in quantifying uncertainty in combinations of fixed effect parameters and the best linear unbiased predictors (BLUPs) of random effects. It would, however, be preferable to have more reliable estimates of the country-specific uncertainty corresponding to the estimates from the regional multilevel loess-based point estimates and further work in this area may prove fruitful.

The point estimates for time points within the period for which data are available for Congo appear to be very robust to the exclusion of individual countries or to the use of an entirely different set of additional countries, but at time points a long way outside this period there may be some variability. The estimated uncertainty surrounding each estimate, however, appears to be far more sensitive to the set of additional countries used. In particular, using mortality neighbours (with little between-country variability in mortality) leads to much narrower uncertainty intervals than does using geographical neighbours (with much greater between-country variability). Thus the set of countries selected is clearly of great importance to the final results.

The set of countries should be as similar as possible to the target country in terms of level and trend of under-5 mortality. Similarity in trend will ensure the point estimates follow an appropriate trajectory, whilst similarity in level will minimise the associated uncertainty. We thus suggest choosing the set of additional countries primarily based on

the available under-5 mortality data. By choosing countries with similar mortality levels over short periods of time it is possible to also identify countries with a similar trend, without examining this explicitly. However, for many target countries mortality will only be known over a short period, so this approach may still lead to the selection of countries with vastly differing mortality levels and trends outside this period. Hence assessment of similarity may also have to be based on available information regarding factors such as geography, climate, politics, other health indicators, health system funding, etc. The robustness of estimates to the set of additional countries used in the model should be investigated further through applications in other countries and possibly simulations.

It should also be noted that there remain several outstanding issues regarding the estimation of under-5 mortality using the existing loess-based approach which the regional multilevel extension makes no attempt to solve. These include data derived from the same source (e.g. a given DHS or MICS) within a country currently being assumed to be independent, and direct and indirect estimates derived from the same source often both being utilised. Both of these issues could theoretically be addressed by the addition of further levels in the hierarchy of the current regional multilevel approach, though the scarcity of data may mean that this suggestion is not practicable.

We conclude that the proposed regional multilevel extension to the existing loess-based approach to under-5 mortality estimation provides a useful alternative in countries with few data if a suitable set of countries can be identified. Although there remain issues with the proposed approach which require further exploration and development, it provides an alternative set of estimates for comparison with those resulting from the conventional approach. Where the two sets of estimates closely agree it is reassuring, where they do not this should at least provide the catalyst for further investigation.

Acknowledgements

This work was funded by grants from Unicef and the World Health Organisation.

References

[1] United Nations. *United Nations Millennium Declaration, Resolution 52/2*. United Nations, New York, 2000.

[2] C. J. Murray, T. Laakso, K. Shibuya, K. Hill, and A. D. Lopez. Can we achieve Millennium Development Goal 4? New analysis of country trends and forecasts of under-5 mortality to 2015. *Lancet*, 370(9592):1040–54, 2007.

[3] H. Masanja, D. de Savigny, P. Smithson, J. Schellenberg, T. John, C. Mbuya, G. Upunda, T. Boerma, C. Victora, T. Smith, and H. Mshinda. Child survival gains in Tanzania: analysis of data from demographic and health surveys. *Lancet*, 371(9620):1276–83, 2008.

[4] W. S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc*, 74(368):829–36, 1979.

[5] W. S. Cleveland and S. J. Devlin. Locally weighted regression: an approach to regression analysis by local fitting. *J Am Stat Assoc*, 83:596–610, 1988.

[6] H. Goldstein. *Multilevel Statistical Models*. Hodder Arnold, London, 2003.

[7] Institute for Health Metrics and Evaluation. Worldwide Child Mortality (Mortality Data). `http://www.healthmetricsandevaluation.org/print/mortality.xls`, accessed November 2007.

[8] D. Ruppert, M. P. Wand, and R. J. Carroll. *Semiparametric Regression.* Cambridge University Press, New York, 2003.

[9] *R 2.6.1.* The R Development Core Team, 2007.