Corrigendum: paragraph 1, page 96.

Finally, we assess the predictive accuracy of the three logistic regression models in Figs. 1 and 2 to Random Forests on out-of-sample data. The CWD described in Section 3 runs from 1945 to 2000. We updated the CWD for all countries in Africa and the Middle East from 2001 to 2014. The updated data give us an additional 896 observations with nineteen civil war onsets. We trained each model or algorithm on the 1945–2000 CWD and tested on the updated CWD for Africa and the Middle East. We assess each model or algorithm’s predictive accuracy using the AUC score defined earlier. Table 1 reports the AUC scores for each model as well as for Random Forests. Random Forests is superior to all logistic models with an AUC of 0.94. The predicted probabilities of civil war for each model are shown in Table 1.

All logistic regression models fail to specify any civil war onset in the out-of-sample data.

Random Forests correctly predicts nine of nineteen civil war onsets in this out-of-sample data

when the threshold for positive prediction is 0.50. Random Forests correctly predicts the onset

of civil war in Iraq in 2004, Somalia, Uganda, Rwanda, and Angola. It fails to correctly predict the civil wars resulting from the U.S. invasion of Afghanistan, or the civil wars in Syria and Libya that resulted from the Arab Spring. It is possible that civil wars resulting from external intervention or revolutions may have different causes and are thus poorly predicted when civil war is defined only by the number of deaths in battle. We discuss the application of statistical learning methods to the analysis of causality in the following section.

Corrigendum: footnote 10. The Hegre and Sambanis (2006) model has an AUC score of 0.83, Fearon and Laitin (2003) has an AUC score of 0.69, and the Collier and Hoeffler (2004) model has an AUC of 0.90.