

Team led by Fulvia Mecatti, Professor of Statistics at University of Milano-Bicocca
Dr. Gaia Bertarelli, Post-doc at University of Pisa
Dr. Maeregu Woldeyes Arisido, Post-doc at University of Milano-Bicocca

Evaluation Grid

1) Is the model scientifically sound (statistically or otherwise)? Describe and justify the chosen modelling approach.

Section **Brief description of our SAE-LM model** of our Report describes the rational and the basic theoretical foundations of our modelling approach. Essential methodological details about model specification, basic assumptions and probability distributions, are given in a separate section at the end of the report. Complete details are given in Bertarelli et al. 2018, Small area estimation for unemployment using latent Markov models, *Survey Methodology*, 44, 167-1928. With respect to our specific application, Trace plots (available on demand) of Monte Carlo simulations to fit the model, show clear stability and good performance of our fitting algorithm for all estimated model parameters

2) Is the approach replicable? Provide documented code or pseudo-code.

All R-code for model fitting and selection we used to produce results illustrated in our Final Report, is provided alongside.

3) Do model input data have known associations with TB prevalence? Describe the input data used and justify their function in the model in terms of hypothesised relationship with TB

In Section **Background** of our Final Report a list of all available auxiliary variables is given; in Section **Brief description of our SAE-LM model** details are provided concerning variable selection and model fitting.

4) Does the model have a sound methodology to derive credible/confidence intervals (CI) for the estimates (or other measures of sampling variability)? Describe and justify chosen methodology and provide the CI.

In Section **SAE-LM model Validation** familiar measures of sampling variability (Coefficient of Variation) are provided and discussed. In addition two methods to derive uncertainty intervals are illustrated: 1) Interval estimates, with chosen confidence level, are given directly by posterior distribution of small area parameters of our best SAE-LM model, as Bayesian credible interval; and 2) familiar 95% Confidence Intervals around the area prevalence estimate $\pm 1.96 \times SE$ (SE= Standard Error) can be computed as well. A complete list of all district 95% interval estimates is provided in a separate table at the end our Report.

5) What is the model's estimated predictive power based on 2010 data? Provide mean squared error (MSE) and relative square error (RSE) score from leave-one-out-cross-validation comparing actual and predicted 2010 cluster-level prevalence point estimates.

Cross- validation methods, including the recommended leave-one-out cross-validation, are not practical options for SAE-LM model validation, essentially due to its resource consuming nature. In Section **SAE-LM model Validation** several alternative diagnostics is provided, which we reckon considerably supports the predictive power of our SAE-LM model. Further diagnostics as been performed and it is available on demand. We also remark that there is a theoretical connection between leave- one-out cross-validation and Bayesian Information Criterion (BIC) we used for model selection (as well as with other information criterion methods based on maximum likelihood. For technical details see Klaenskens, 2008, *Model selection and Model averaging* Cambridge University Press, Section 2.9)

6) Do the estimates produce a distinction between high and low prevalence districts in 2018 in line with the local understanding of the epidemiology? We will conduct an opinion poll with apx 20-30 Pakistan TB experts to get consensus on top 3 highest districts and bottom 3 in terms of TB burden. Models should have those above their 75% percentile and below their 25% respectively.

see Section **SAE-LM Results** of our Final Report. Maps of both district-wise indirect estimates and a classification into three classes of TB burden are provided. A complete list of all district-wise estimates computed and time evolution of indirect estimates is provided in a separate table at the end our Report.

7) Is the model overfitting the prevalence survey data? Reflect on potential overfitting of the model.

We think the potential overfitting of our SAE-LM model is limited, We can bring two arguments: 1) SAE-LM methodology have been derived to improve over a previous method (You, Y., Rao, J.N.K. and Gambino, J. (2003). *Survey Methodology*, 29, 1, 25-32.), which offered the advantage to be able to combine coss-sectional and longitudinal data under a hierarchical Bayes modelling; at the same time it had to be improved over its tendency to possible oversmoothing. 2) Covariates have been carefully selected from the set of available auxiliary variable, according to a two steps procedure: first, hightly correlated auxiliary variables have been detected and reduced. Second, step-wise regression has been applied to identify sub-sets of relevant covariates.