Sebastian Funk, Sam Abbott *(London School of Hygiene & Tropical Medicine)* **and Johannes Bracher** (*Karlsruhe Institute of Technology*)

We thank Teh and coauthors for a stimulating paper. The comparison of the different performance characteristics of R_t estimators is to be applauded. The lack of any more general systematic comparison has been a crucial gap in the literature. At the same time, the chosen approach highlights some difficulties in addressing this gap:

1) Reproduction numbers are generally not observable and the ability of a model to reconstruct their trajectories can only be assessed in simulation studies. These may not yield sufficient information to accurately assess the ability of a model to estimate R_t in the real world. Whilst simulation studies can be useful to assess the self-consistency of a model and its sensitivity to potential mis-specification, their use for comparing estimates between models is more limited unless the data generating process is not one of the models used for estimating.
2) It is challenging to make a fair comparison between tools that each have much flexibility and can be used and parameterised in a variety of ways that affect the resulting reconstructed R trajectories; identifying whether any resulting performance differences are due to differences in the methods themselves or specific user choices can be difficult.
3) By definition of the instantaneous reproduction number R_t is the average number of new infections caused by infectious individuals at time t (weighted by their infectiousness). However, this leaves much room for interpretation. It can be defined over any time frame such as daily or weekly, and this choice affects the estimates. Moreover, even within the same model there may be several quantities which could be interpreted as R_t. For example, consider the following branching process model:
   $I_{t+1} \sim \mathrm{Poisson}(\lambda)$
   $\lambda = X_t I_t$
   $X_t \sim \mathrm{gamma}(\mathrm{mean} = R_0, \mathrm{var} = R_0/k)$
   There are three ways one could define a reproduction number of this model:
   i) $R_t = I_{t+1}/I_t$
   ii) $R_t = X_t$
   iii) $R_t = R_0$
   Any assessment of the quality of R_t estimates would thus involve a decision on the timescale at which the reproduction number is considered and what exactly R_t represents and whether models agree that this is being estimated.

We believe that evaluating the performance of published R_t estimates or methods for generating them is valuable, and reiterate our appreciation of the authors' approach to doing so. Further comparisons should be encouraged by both those developing methods and those consuming them. Ideally, this could be led by an independent arbitrer who defines the parameters of the exercise and, in collaboration with the developers, applies the methods to be assessed.