

Discussion of the papers by Bekker-Nielsen Dunbar & Held and by Pellis et al

Gianpaolo Scalia Tomba
University of Rome Tor Vergata, Italy.

First of all, many thanks to presenters and organizers for a very stimulating meeting and session. The two papers presented in this session can be described as time series analysis mixed with a "mechanistic" component and "mechanistic" analysis mixed with a time series component, respectively, which emphasises that work intended to apply to reality needs to combine modelling and statistics...

Formally, the two papers cannot be compared, since the Bekker-Nielsen Dunbar & Held paper aims at counterfactual inference about school closures (although estimation of R_t would be possible) while the Pellis et al paper is a review and discussion of modelling/statistical COVID related work, mainly in GB.

Each paper contains many interesting points, for instance the Pellis et al paper discusses the concept of R_c , the reproduction number of an infection in the presence of interventions but not accounting for the level of immunity of the population.

A related question is whether such a number exists for an isolated intervention or must interventions be evaluated in combination or in sequence or in a certain epidemiological situation or in a given state of public sentiment? However, in the aftermath of epidemic peaks, the evaluation of relative worth and cost of various interventions will certainly occur...

The Bekker-Nielsen Dunbar & Held paper shows the importance of incorporating changes in population contacts into the description and prediction of disease spread. In the model, these changes are modelled on top of synthetic age-structured population matrices, but an obvious question is whether the age-group stratification is enough or if finer distinctions should be made, such as "responding" and "non-responding" individuals, or other subclasses... This question relates to the more general one how much aggregation can be made in these matrices, since we know that the sum in a multitype process does not simply and directly correspond to a single group model...

There are also more general points, many of which are also made in the papers, but that I would like to emphasize.

A general problem seems to be that not enough emphasis is given to methods of delay correction in data and to under-reporting of cases.

Especially the latter will have varied during the epidemic, when the incidence and prevalence has changed in and between age groups with different degrees of symptomaticity. The time variation could itself be subject to modelling...

There are also several "theoretical problems" that one can pick up on for discussion, for instance the use of the "generation time distribution" in "model-free" estimation of R_t , since this distribution will also have varied over time because of interventions, changed abruptly at certain times and been different in various population groups. How should this be estimated and integrated in the estimation of R_t ?

In model-based estimation, there is a more fundamental doubt... To my knowledge, there is no "micro to macro" theory of how the mass action-like dynamics of large populations should arise from such dynamics in smaller groups, i.e. do we know that this kind of dynamics is a correct representation of real infection spread? There is no doubt that these models can fit data, but do they possess the right properties to predict effects of intervention or future development?

Estimating R_t or short-term behaviour of number of cases is based on continuity, but long-term prediction must be based on "knowledge"...

Finally, I feel that there is almost consensus among us (modellers, statisticians,...) that more information than one or two R or r -values, both out of models and from data, is needed to understand the progress of an epidemic and to propose topical interventions...

We know that an R is not really "the average number of secondary cases caused by one infective right now" because there are a number of distinctions to this statement, both model-dependent and data-dependent...

However, this seems to get lost when communicating with others, politicians and general public, probably because admitting complexity would seem less knowledgeable (not only by scientists, politicians have quickly picked up the style...). The emphasis on R is one of the symptoms of this tendency...

This leads to an important point, the "political" value of simple summaries... Is simplification, which, in its limit, will almost be falsification, unavoidable/avoidable, necessary/unnecessary, good or bad?

Maybe the famous quote "Models should be as simple as possible, but not more..." (Einstein, perhaps...) should be valid also for information about epidemics...