Preface

Since the onset of the COVID-19 pandemic in early 2020, there has been a proliferation of software packages that make inference about the current state of an infectious disease outbreak based on daily counts of disease cases. An important and widely used parameter is the instantaneous or effective reproduction number, R(t), defined in Gostic et. al. (2020) 1 as follows:

"The effective reproductive number, denoted as Re or Rt, is the expected number of new infections caused by an infectious individual in a population where some individuals may no longer be susceptible."

Defined as such, R(t) is an unobserved quantity that captures the aggregated combination of disease characteristics (e.g., infectiousness under controlled conditions, mode of transport) and extrinsic factors (e.g., lockdowns that reduce person-to-person contact). An $R(t) = 1$ indicates a "stable" epidemic at time t, where each infected person infects on average one additional person; R(t) values above or below 1 represent a growing or a shrinking epidemic respectively. We focus on R(t) given its relevance for time-sensitive decision-making, and summarize the currently used inputs, data, methods and assumptions in R(t) estimation across the following categories:

- How the relationship between R(t) and infections is defined
- How R(t) is constrained using distributions for key variables
- How R(t) is constrained over time
- Additional data and distributions that are used to constrain R(t)
- Inference frameworks that are used to estimate R(t) ** is constrained the right word?

In this paper we provide a theoretical comparison of the current field of methods for estimating R(t), with the goal of informing user decision-making about which package to choose and in interpreting package outputs. For reference, Table 1 lists packages with an accompanying peer-reviewed journal manuscript, Table 2 lists packages without a peer-reviewed journal manuscript, and Table S1 contains R(t) packages that calculate R(t) but were excluded from this summary. In the text, the package citation is given the first time each package is referenced.

We limit the methods discussed here to those for estimating historical to present-day R(t) values using daily case count data, where a case can be flexibly defined as an individual with a reported positive test (either through healthcare-seeking behavior, routine surveillance, or a hospital admission). Other methods not discussed here include inference of R(t) exclusively

from alternative data sources (e.g., genetic data,2 behavioral data,3 or viral loads in wastewater4), or calculations from compartmental, agent-based models, or network.5–7 We also limit the discussion to packages in the statistical software R,8 which may exclude some packages in other software programs that combine many of the methodological considerations discussed below.9 We do not discuss any packages for now-casting or forecasting, though a number of R(t) estimation packages can be used for this purpose. The methods discussed below and references to specific R packages are current as of December 1, 2024. We attempt to harmonize the mathematical choices between each package using terminology from each.

An evaluative comparison of the performance of these methods would be highly complex, given the following challenges. Some of the most widely-used packages are not accompanied with a peer-reviewed manuscript that describes or evaluates the theory behind modeling choices. Each package contains a subset of the methods below for constraining R(t) in time, but with subtle variations in implementation that are often not well-documented. Most packages have not been recently updated, and even those that have are not maintained on CRAN, instead leaving updates on a development version on GitHub. The combination of differing model frameworks associated with each package make it challenging to easily compare the impacts on estimated R(t), especially when considering additional factors like ease of implementation and computational time