Value of risk-contact data from digital contact monitoring apps in infectious disease modeling

Uiterkamp, M.H.H.S.; Dijk, W.J.V.; Heesterbeek, H.; Hofstad, R. van der; Kiefte-de Jong, J.C.; Litvak, N.

Epidemics 55 (2026) 100920

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Value of risk-contact data from digital contact monitoring apps in infectious disease modeling

Martijn H.H. Schoot Uiterkamp [a,*], Willian J. van Dijk [b,c], Hans Heesterbeek [d], Remco van der Hofstad [e], Jessica C. Kiefte-de Jong [b], Nelly Litvak [e]

[a] Department of Econometrics and Operations Research, Tilburg University, The Netherlands
[b] Department of Public Health and Primary Care, Health Campus The Hague and Leiden University Medical Center, The Netherlands
[c] Department of Clinical Epidemiology, Leiden University Medical Center, The Netherlands
[d] Department of Population Health Sciences, Utrecht University, The Netherlands
[e] Department of Mathematics and Computer Science, Eindhoven University of Technology, The Netherlands

ARTICLE INFO

Dataset link: https://github.com/mhhschootuiterkamp/DCM-data-integration-in-SEIR-models

Keywords:
Digital contact monitoring
Infectious disease transmission
Risk-contact
Compartmental model
Effective reproduction number

ABSTRACT

In this paper, we present a simple method to integrate risk-contact data, obtained via digital contact monitoring (DCM) apps, in conventional compartmental transmission models. During the recent COVID-19 pandemic, many such data have been collected for the first time via newly developed DCM apps. However, it is unclear what the added value of these data is, unlike that of traditionally collected data via, e.g., surveys during non-epidemic times. The core idea behind our method is to express the number of infectious individuals as a function of the proportion of contacts that were with infected individuals and use this number as a starting point to initialize the remaining compartments of the model. As an important consequence, using our method, we can estimate key indicators such as the effective reproduction number using only two types of daily aggregated contact information, namely the average number of contacts and the average number of those contacts that were with an infected individual. We apply our method to the recent COVID-19 epidemic in The Netherlands, using self-reported data from the health surveillance app COVID RADAR and proximity-based data from the contact tracing app CoronaMelder. For both data sources, our corresponding estimates of the effective reproduction number agree both in time and magnitude with estimates based on other more detailed data sources such as daily numbers of cases and hospitalizations. This suggests that the use of DCM data in transmission models, regardless of the precise data type and for example via our method, offers a promising alternative for estimating the state of an epidemic, especially when more detailed data are not available.

1. Introduction

During the outbreak of an infectious disease, monitoring interactions between individuals provides useful information on the spread of the disease within the population. Information on social contacts has been used frequently in mathematical models of infectious disease transmission (Beutels et al., 2006). The two primary methods to collect this information have been diary-based social contact surveys (Hoang et al., 2019) and field experiments involving proximity-measuring sensors (see Anglemyer et al. (2020) and the references therein). Contact data obtained via either method have been successfully used to infer infection dynamics within a population and, in particular, to identify contact networks and heterogeneity in contact behavior (Read et al., 2012). However, most surveys and experiments are only conducted incidentally and not continuously on the daily level (Verelst et al., 2021). As a consequence, data from these sources is of limited use for continuously estimating short-term changes in key indicators during an ongoing epidemic.

Contact data from *digital contact monitoring (DCM)* apps that were in use during an actual epidemic could potentially be used to address this limitation of current methods and sources of contact data. Here, we use the term DCM to refer to both digital contact tracing (DCT) and to public health surveillance via self-reporting of, e.g., symptoms and risk behavior. Most of these apps were developed to replace or complement manual contact tracing, in part because their use is less labor-intensive, more time-efficient, and highly scalable (see, e.g., Ferretti et al. (2020) and the references therein for apps developed and used during the recent COVID-19 pandemic).

Despite the benefits of DCM compared to traditional methods and sources, it is currently unclear to what extent these contact data are

* Corresponding author.
E-mail address: m.h.h.schootuiterkamp@tilburguniversity.edu (M.H.H. Schoot Uiterkamp).

https://doi.org/10.1016/j.epidem.2026.100920
Received 28 March 2025; Received in revised form 9 December 2025; Accepted 13 May 2026
Available online 20 May 2026

useful in transmission modeling. There is evidence that contact data from DCM apps is not only useful for directly reducing the number of infections, but also for large-scale analysis of drivers behind epidemiological growth in (close to) real-time (Kendall et al., 2024). However, comparing data collected via different apps is challenging because of differences in implementation and varying reliability of (different parts of) the contact registration process within these apps (Masel et al., 2023). Furthermore, the few contact data that are stored by most DCT apps are very coarse and typically contain few or even no details on the registered contacts, such as user demographics. In particular, public versions of these DCT data are often limited to counts on positive test results of DCT app users. This is because of the delicate balance between obtaining sufficient information on infectious contacts and aiming to preserve privacy of users (Bradford et al., 2020). For instance, individual location data from GPS-based DCT apps should be anonymized and aggregated so that individual travel patterns cannot be reconstructed (Xu et al., 2017). To comply with such requirements, most DCT apps that were developed during the COVID-19 epidemic had a decentralized design where as few privacy-sensitive data as possible were shared and stored (Shahroz et al., 2021).

Altogether, the question remains to which extent DCM data is useful to inform transmission models. The goal of this paper is first of all to demonstrate the use of such data to initialize a conventional susceptible–exposed–infectious–recovered (SEIR) compartmental model. We then characterize the validity by estimating key indicators such as the effective reproduction number and the number of infectious individuals. We take the COVID-19 pandemic in the Netherlands as our use-case to compare our estimates with published values derived from detailed patient, testing and hospital data.

Compared to other methods for estimating these key indicators such as Kendall et al. (2024), our approach does not require model calibration or statistical estimation of transmission rates. Instead, we directly initialize relevant compartments and parameters from the available contact data. For this, we derive a simple but powerful mechanistic relationship between the infectious compartment and the proportion of contacts that were with infected individuals. One important advantage of this approach is that our approach only requires aggregated contact data as input, instead of also individual-level information. Moreover, our subsequent estimates of the effective reproduction number are robust against any structural multiplicative bias in the contact data that, for instance, may arise from the inevitable and systematic under-reporting of contacts (see also Section 4).

We compare two different approaches to obtain these aggregated data that use different methodologies for data collection. The first approach is based on self-reported data on close contacts and contacts with infected individuals that were collected via the public health surveillance app COVID RADAR (van Dijk et al., 2021), from April 2020 to February 2022. In the second approach, we combine COVID-RADAR data with contact data obtained via the Dutch DCT app "CoronaMelder" (ter Haar et al., 2023), from October 2020 to April 2022.

We validate our approach and assumptions by comparing our estimates to those made by the Dutch National Institute for Public Health and the Environment (RIVM). Finally, we conclude the paper by discussing the limitations of the approach and data, and providing recommendations for improving data collection with DCM apps like COVID-RADAR and CoronaMelder for the benefit of modeling infectious disease transmission in future epidemics.

2. Methods

2.1. Data sources

We use two data sources on contacts with infected individuals obtained from DCM apps:

Self-reported contact data obtained via the COVID RADAR app (van Dijk, 2020)

In summary, during the early phase of the pandemic (April 2020) the COVID RADAR app was launched after a short (social) media campaign. Users could anonymously report COVID-19 related symptoms, contacts and test results on a daily basis. In particular, users were asked to report the number of close interactions (within 1.5 m) and whether they had a known recent contact with COVID-19 patients (in the last 14 days). These are the two variables that we will use in this research (see also Section 2.3). The question of the latter variable did not specify how close the contact must have been (e.g., within 1.5 m). However, focus group interviews with app users did not indicate any feelings of ambiguity towards the phrasing of the question, unlike towards other contact-related questions (Splinter et al., 2022). Thus, we assume here that the reported recent contacts with COVID-19 patients were within 1.5 m and thus could have contributed to transmission. To preserve privacy, each user was assigned a unique non-retraceable ID and data of that user was only linked to this ID. Moreover, the collection of precise personal information was minimized (e.g., instead of via date of birth, age was recorded in broader categories). Users were automatically reminded each other day to submit their responses and were offered insights into the data via in-app news updates and live feedback about personal (risk) behavior compared to the national average. In total over 250,000 unique individuals filled in the app over 8.5 million times, over a period of 750 days. For further details on the collection and validation of data via the COVID RADAR app, we refer to van Dijk et al. (2021).

One major limitation of the data is that the user population is not a random sample of the Dutch population. In particular, the age distribution among users is not proportional to that of the Dutch population, with a disproportionately large share of middle-aged and elderly (60+) users. Not addressing this issue might bias the results since contact rates are known to differ between age groups and in particular are below average for people aged 60 or older (Backer et al., 2021). To account for this and improve the representativity of the user population for the entire Dutch population, user's responses were weighted in the rate calculations according to the proportion of their age group within the user population and within the Dutch population. Furthermore, to further reduce bias, users who listed "healthcare professional" as occupation were excluded. This is because people in these occupations are generally more in contact with COVID-19 patients but usually have a lower risk of infection due to the use of more elaborate protective equipment and a stricter adherence to measures.

Contact tracing information obtained via the Dutch DCT app "CoronaMelder" (Ministerie van Volksgezondheid, Welzijn en Sport, 2022b)

The CoronaMelder app was developed by the Dutch Ministry of Health, Welfare, and Sport and officially launched in October 2020 after several regional field tests in the summer of 2020. Users who downloaded and activated the app received a notification whenever another user that recently was nearby received a positive test result from an official test location of the Dutch municipal health service, i.e., whenever they were recently in close contact with a potentially infectious other user. The data contains the daily number of app downloads and the number of users who received a positive test and gave permission to send out notifications to other users. To improve users' privacy, no data was recorded on which users have received such a notification, meaning that no individual user data on risk-contacts is stored. As a consequence, and in contrast to the data from

COVID RADAR, no other personal information is known on the user base of CoronaMelder in general and on those that gave permission in particular. Thus, unlike for the COVID RADAR data, we cannot correct for any bias in user characteristics such as age or occupation. The app was downloaded over 5,8 million times over a period of 606 days. We refer to ter Haar et al. (2023) for more technical details and an empirical evaluation of the effectiveness of the app.

In our approach, we require average numbers of notifications among active app users. However, daily estimates of this number have only been made from May 2021 onward and are not publicly available on the daily level (Ebbers et al., 2021). Furthermore, the number of app downloads is not representative of the number of active app users because many users stopped using the app throughout its run (van der Laan and de Wit, 2022). To approximate the daily number of active users, we interpolate monthly published statistics on the number of active users (Ministerie van Volksgezondheid, Welzijn en Sport, 2022a). These reports indicate that, over time, the number of active users fluctuated between 1,9 and 3,4 million. To initialize the interpolation, we assume that the number of active users at the launch of CoronaMelder in October 2020 equals the cumulative number of app downloads at that time.

To demonstrate the value of the contact data from COVID RADAR and CoronaMelder, we will make comparisons with two other publicly available data sources that were used and published by RIVM during the COVID-19 epidemic (Rijksinstituut voor Volksgezondheid en Milieu):

Effective reproduction number. The national-level effective reproduction number was estimated based on state-reported numbers on hospital admissions as reported by the Dutch National Intensive Care Evaluation COVID-19 registry (Dongelmans et al., 2022) (until June 12, 2020) and on the daily number of positive test results registered at the official test locations of the Dutch municipal health service (after June 12, 2020). For details on the methods used to estimate the reproduction number from these data, we refer to Wallinga and Lipsitch (2007), van de Kassteele et al. (2019).

Number of currently infectious individuals. The number of individuals that were infectious on a given day was estimated using age-stratified number of daily hospitalized COVID-19 patients in combination with serological data from a nationwide population based study (Vos et al., 2021). Here it is assumed that the infectious period ranges from two days before symptoms to four to eight days after symptoms.

2.2. Model

In our approach, we model the spread of infection using the following standard deterministic SEIR-compartmental model:

$$\frac{d}{dt}S(t) = -\frac{\beta(t)S(t)I(t)}{N} + \frac{R(t)}{\omega},$$
$$\frac{d}{dt}E(t) = \frac{\beta(t)S(t)I(t)}{N} - \frac{E(t)}{\alpha},$$
$$\frac{d}{dt}I(t) = \frac{E(t)}{\alpha} - \frac{I(t)}{\tau},$$
$$\frac{d}{dt}R(t) = \frac{I(t)}{\tau} - \frac{R(t)}{\omega}.$$

Here, $S(t)$, $E(t)$, $I(t)$, and $R(t)$ denote the numbers of susceptible, exposed, infectious, and recovered individuals in the population at time t, respectively. Moreover, we denote by ω, α, and τ the average immunity, latent, and infectious period, respectively. We assume that the total size of the population is N and remains constant through time.

Finally, we decompose the transmission rate $\beta(t)$ as $\beta(t) = \varepsilon c(t)$, where ε is the time-independent probability that a contact between a susceptible and an infectious individual leads to transmission and $c(t)$ is the contact rate at time t.

Our aim is to integrate data on contacts with infected individuals in this model. Therefore, it is crucial to explicitly focus in the SEIR model on the contacts with infectious individuals (in Section 2.3, we describe how we estimate contact rates with infectious individuals from contact rates with infected individuals). We denote this infectious contact rate by $c^I(t)$. The goal is to initialize the compartments of the SEIR model, given the infectious contact rate $c^I(t)$ and the overall contact rate $c(t)$.

We focus on first initializing the number of infectious individuals $I(t)$. For this, we exploit the assumption of random mixing within the SEIR model, i.e., contacts between individuals happen at random and thus each pair of individuals is equally likely to meet. Under this assumption, the infectious and overall contact rate are related via $c^I(t) = c(t)\frac{I(t)}{N}$. This means that, given $c^I(t)$ and $c(t)$, we may initialize $I(t)$ as

$$I(t) = N\frac{c^I(t)}{c(t)}. \tag{1}$$

This relationship implies that as long as the number of infectious individuals is nonzero, a positive fraction of the total number of contacts within the population is between a susceptible and an infectious individual. This fraction is represented by the fraction of contacts that were with an infectious person, which is modeled from the data as $c^I(t)/c(t)$. When this fraction is close to 1, almost all contacts were with infectious persons, meaning that within the SEIR model the share of infectious individuals in the total population of size N is close to 1 as well. On the other hand, when the fraction is close to 0, hardly any contacts were with infectious persons, meaning that in the SEIR model the share of infectious individuals is close to 0 as well. We stress that this relation is invalided when there is a clear case of non-random mixing, for instance when measures of quarantine are in place. These are the moments where we expect our results to be less reliable and deviate most from other models and data sources that do take non-random mixing into account.

We initialize the remaining compartments of the model in the following way, based on the now known expression for $I(t)$ and direct manipulation of the differential equations that describe the model:

$$E(t) = \alpha\left(\frac{d}{dt}I(t) + \frac{I(t)}{\tau}\right),$$
$$R(t) = e^{-\frac{t}{\omega}}R(0) + \int_0^t e^{-\frac{t-s}{\omega}}\frac{I(s)}{\tau}ds,$$
$$S(t) = N - E(t) - I(t) - R(t).$$

Note that this implies that, to initialize the model at time t, we require the entire time series of $I(t)$ and thereby also of $c^I(t)$ and $c(t)$.

We use our model and the relation between the number of infectious individuals and the contact rates in (1) to estimate the effective (instantaneous) reproduction number, where we apply the method of Cori et al. (2013). This method requires information on the incidence, i.e., the number of new infections, and the generation interval, i.e., the time between the infection of an infected person and of their infector. The incidence at time t is given in our model by $I(t) = \beta(t)S(t)I(t)/N$ and the generation interval is assumed to be a random variable with a given probability density function w (we discuss a suitable choice for the distribution of the generation interval in Section 2.3). We estimate the reproduction number at time t as

$$\mathcal{R}(t) = \frac{I(t)}{\int_0^\infty I(t-i)w(i)di}.$$

We compare two approaches for evaluating this expression, leading to two estimates $\mathcal{R}^1(t)$ and $\mathcal{R}^2(t)$ for the effective reproduction number:

1. In the first approach, using the relation in (1), we rewrite the original expression for $\mathcal{R}(t)$ in terms of only the infectious contact rate, number of susceptible individuals, and the generation interval. This results in our first estimate $\mathcal{R}^1(t)$ of the effective reproduction number that can be evaluated directly using the time series of the size of the susceptible compartment up until time t:

$$\begin{aligned}\mathcal{R}^1(t) &= \frac{\mathcal{I}(t)}{\int_0^\infty \mathcal{I}(t-i)w(i)di} \\ &= \frac{\frac{\beta(t)S(t)I(t)}{N}}{\int_0^\infty \frac{\beta(t-i)S(t-i)I(t-i)}{N} w(i)di} \\ &= \frac{c(t)S(t)I(t)}{\sum_{i=0}^\infty c(t-i)S(t-i)I(t-i)w(i)} \\ &= \frac{c^I(t)S(t)}{\int_0^\infty c^I(t-i)S(t-i)w(i)di}. \end{aligned} \tag{2}$$

2. In the second approach, we approximate $\mathcal{R}(t)$ by integrating a simplifying assumption on $S(t)$ in the earlier obtained expression for $\mathcal{R}^1(t)$ in (2). More precisely, we assume that the relative daily change in $S(t)$ is small, meaning that $S(t-i) \approx S(t)$ for small values of i. Substituting this relationship in (2) yields our second estimate $\mathcal{R}^2(t)$ of the effective reproduction number:

$$\mathcal{R}^2(t) = \frac{c^I(t)}{\int_0^\infty c^I(t-i)w(i)di}.$$

In particular, this estimate of the effective reproduction number only depends on the infectious contact rate and on the generation interval.

In general, note that neither $\mathcal{R}^1(t)$ nor $\mathcal{R}^2(t)$ depend on the unknown transmission probability ε. Moreover, both estimates are robust against any structural multiplicative bias in the (proxies used for the) contact rates $c^I(t)$. Sources for such bias include structural under-reporting of contacts with, e.g., asymptotic individuals or incidental contacts with strangers (for self-reported data) and limited accuracy of the specific technology used to detect close contacts (for proximity-measuring apps) (we discuss this in more detail in Section 4). The robustness of the estimates is because $c^I(t)$ occurs in each term of the division in $\mathcal{R}^1(t)$ and $\mathcal{R}^2(t)$ and thus any multiplicative bias in $c^I(t)$ is divided out. This means that, when applying this approach, it is sufficient that the used proxy for $c^I(t)$ has the same temporal structure as $c^I(t)$ rather than the same volume.

In Section 3, we compare both these approaches with each other and with the approach underlying the estimates produced by RIVM (Wallinga and Lipsitch, 2007; van de Kassteele et al., 2019). Finally, we note that, although we focus in this section on SEIR compartmental models, the core ideas and analyses presented in this section are also applicable to other types of compartmental models with a similar interaction between susceptible and infectious individuals.

2.3. Data integration

Within the compartmental model, the values for the average latent, infectious, and immunity periods α, τ, and ω are chosen based on the literature as 5.5 days (Xin et al., 2021), 9.5 days (Byrne et al., 2020), and 90 days (Stein et al., 2023), respectively. The generation interval is assumed to follow a Gamma distribution with a mean of 4 days and a standard deviation of 2 days (Backer et al., 2022). To initialize the contact rate $c(t)$, we use the variable "numberpersons150cm" from the COVID RADAR dataset. This variable, here denoted by $C_i(t)$, states the number of persons that were within 150 cm of user i on day t. We directly initialize $c(t)$ as the average of $C_i(t)$ over all N users, i.e., $c(t) = \frac{1}{N}\sum_{i=1}^N C_i(t)$.

Finally, to initialize the infectious contact rate $c^I(t)$, we employ a two-step approach. First, we initialize an intermediate rate $z(t)$ of contacts with *infected* individuals. In Sections 2.3.1 and 2.3.2, we present two approaches for this using COVID RADAR data and CoronaMelder data, respectively. Second, we discuss how we transform this rate $z(t)$ of infected contacts to the desired rate $c^I(t)$ of infectious contacts. For this, note that we do not know the status of the infected persons at the time of the contact, i.e., whether they were actually infectious and could have caused a transmission. Therefore, we assume that the infected persons were either exposed or infectious and that this division is proportional to the relative difference between the latent and infectious period. This means that we assume a fraction $\frac{\alpha}{\alpha+\tau}$ of these contacts to be with an exposed individual and the remaining fraction $\frac{\tau}{\alpha+\tau}$ to be with an infectious individual. Thus, we initialize the infectious contact rate at time t by appropriately discounting the rate $z(t)$ of contacts with infected individuals:

$$c^I(t) = \frac{\tau}{\alpha+\tau} z(t).$$

2.3.1. Estimating the infected contact rate using COVID RADAR data

We use the variable "contact" from the COVID RADAR dataset. This binary variable, denoted here by $C_i^{14}(t)$, represents whether or not user i has had contact with an infected person within the 14 days before t. Note that the variables $C_i^{14}(t)$ individually do not provide precise information on when the contact with an infected person occurred. To extract this information from the data, we consider differences in reported values of $C_i^{14}(t)$ between subsequent days as follows. Suppose that a user i reports on day t that they had contact with an infected person within the last 14 days. If they report the next day $t+1$ that they did *not* have such a contact within the last 14 days, then any contact that was reported on day t must have occurred on day $t-14$. Based on this reasoning, we construct a new intermediate variable $z_i(t)$ that is 1 when user i had contact with an infected person on day t and 0 otherwise:

$$z_i(t) = \begin{cases} 1 & \text{if } C_i^{14}(t+14) = 1 \text{ and } C_i^{14}(t+15) = 0; \\ 0 & \text{otherwise.} \end{cases}$$

We assume that users report at most one contact per day with an infected person. This means that the average number of contacts with infected persons at day t is given by $z(t) = \frac{1}{N}\sum_{i=1}^N z_i(t)$.

One limitation of the data is that users did not always submit a response daily. Therefore, we calculate the daily contact rates only over those users who submitted a response over that given day. Moreover, when a user has reported a recent infected contact at time t but has not reported such a contact in any of the subsequent 14 days (either reporting not having had such a contact or not submitting a report at all), we assume that on day $t+1$ the user did not have an infected contact in the last 14 days. This means that we then initialize $z_i(t-14) = 1$. This case occurred in 22.6% of the surveys where a recent infected contact was reported and thereby in 0.523% of all surveys.

2.3.2. Estimating the infected contact rate using both COVID-RADAR and CoronaMelder data

In this approach, we first estimate the daily number of people that have received a notification in the app and calculate from this the expected time series of total number of contacts with app users that recently tested positive, i.e., $Nz(t)$. Unfortunately, no direct information is available on the number of received notifications due to the design of the app to preserve user privacy. Therefore, we instead use the variable "Reported positive tests through app authorised by GGD (daily)" from the CoronaMelder app data. This count variable, denoted here by $M(t)$, represents the number of app users who received a positive test result on day t and agreed that other app users that were within 1.5 m contact in the last 14 days receive a notification. To obtain a proxy for the daily number of users that receive a notification, we multiply $M(t)$ by the average number of contacts within 1.5 m in the last 14 days according to the COVID RADAR data, i.e., with $\sum_{s=1}^{14} c(t-s)$.

To obtain an estimate of the daily number of infected contacts, and thus of $z(t)$, we delay the estimate of the daily number of received

notifications by the time between contact with an infected individual and receiving a notification in the app. We do this via deconvolution, i.e., we find $z(t)$ as a solution of

$$M(t)\sum_{s=1}^{14} c(t-s) = \int_0^t N z(u) f(t-u)du, \tag{3}$$

where f denotes the probability density function of the time between contact and notification. We model this time as a Gamma distribution with a mean and standard deviation of 3.1 and 3.6 days, respectively, which is in line with the distributions chosen in Leung et al. (2024) and the data analysis in Dolman (2021) that both used a more detailed version of the CoronaMelder data. To obtain $Nz(t)$ from (3), we directly apply inverse filtering, i.e., we express $Nz(t)$ in terms of the Fourier transformations of the other functions in (3). First, we apply the Fourier transform to both sides of (3) to obtain $FT\left(M(t)\sum_{s=1}^{14} c(t-s)\right) = FT(Nz(t))\cdot FT(f(t))$, where $FT(\cdot)$ denotes the Fourier transform operation. This implies that $FT(Nz(t)) = FT\left(M(t)\sum_{s=1}^{14} c(t-s)\right)/FT(f(t))$. Subsequently, we apply to both sides the inverse Fourier transformation, denoted by $IFT(\cdot)$, and find that

$$z(t) = \frac{1}{N} IFT\left(\frac{FT\left(M(t)\sum_{s=1}^{14} c(t-s)\right)}{FT(f(t))}\right).$$

2.4. Evaluation setup

We estimate the effective reproduction number and number of infectious individuals on the daily level, and not continuously, because all data sources we used are reported at this level of granularity, as described in Section 2.1. This means that, when initializing the compartmental model, we use a time-discretized version of the SEIR model and of the probability density functions of the relevant distributions, i.e., those of the generation interval and the time between contact with an infected individual and receiving a notification in the CoronaMelder app.

For the comparison of the effective reproduction number estimates, we select the time periods based on the range of the COVID RADAR and CoronaMelder data. This means that for the estimates purely based on COVID RADAR data, we consider the time period from April 2, 2020 to February 14, 2022 and for the estimates based partly on CoronaMelder data, we focus on the period from October 10, 2020 to February 14, 2022. The end dates of these periods are the same because the data collection for the COVID RADAR app ended on February 28, 2022. For the comparison of the estimates of the number of infectious individuals, we use the same time periods as for the estimates of the effective reproduction number but note that the estimates from RIVM are available only until June 28, 2021. Within the selected time periods, we focus in particular on a cluster of national-wide super-spreading events at the start of July 2021 that became known as the "dancing with Janssen" event (van der Veer et al., 2023). This event cluster occurred after a change in policy that allowed people to directly visit bars and nightclubs without any distancing restrictions after receiving one dose of the Janssen vaccination.

We first compare the two different expressions $\mathcal{R}^1(t)$ and $\mathcal{R}^2(t)$ to estimate the effective reproduction number. More precisely, we validate the assumption that $S(t+i) \approx S(t)$ for small i and thus that the simpler expression $\mathcal{R}^2(t)$ is to be preferred. In a second step, we compare our estimates of the effective reproduction number and the number of infectious individuals with those made by RIVM. Here, we calculate Spearman correlation coefficients between the estimates to assess their agreement. To reduce the impact of noisy data, we smooth the infectious and overall contact rates by fitting for each rate a cubic smoothing spline function, where we select the smoothing parameter via generalized cross validation (Wahba, 1990).

All source code underlying the evaluation is written in Python version 3.9.13 and is publicly available at https://github.com/mhhschootuiterkamp/DCM-data-integration-in-SEIR-models.

3. Evaluation

3.1. Results

Fig. 1 shows the relative difference between our two estimates $\mathcal{R}^1(t)$ and $\mathcal{R}^2(t)$ of the effective reproduction number, i.e., $\mathcal{R}^2(t)/\mathcal{R}^1(t) - 1$. Regardless of the used data source (COVID RADAR or CoronaMelder), this difference is generally below 0.4%, with a single outlier at the very last day for the estimate based on only COVID RADAR. This suggests that the estimates $\mathcal{R}^1(t)$ and $\mathcal{R}^2(t)$ are very close together and thus that the assumption used to derive $\mathcal{R}^2(t)$ is justified. Therefore, in the remainder of this section, we only use the approximation $\mathcal{R}^2(t)$ when comparing our estimates to those of RIVM because this estimate is robust against misspecifications in the overall contact rate $c(t)$ and the average immunity, latent, and infectious period and the specific method used to initialize the compartmental model.

Fig. 2 shows the effective reproduction numbers as estimated via our method ($\mathcal{R}^2(t)$, using either COVID RADAR or CoronaMelder contact data) and as estimated by RIVM (using hospitalization and positive test data). The Spearman correlations between our two estimates and the estimate from RIVM are 0.545 (when using COVID RADAR contact data) and 0.787 (when using CoronaMelder contact data). When shifting our estimates over time, these correlations respectively peak at 0.551 with a delay of 1 day and at 0.903 with a lead time of 4 days. Overall, all three estimates show the same general trend and, as shown in Fig. 3, our estimates generally deviate from those of RIVM by at most 25%. In particular, all three estimates reflect the strong increase in new cases around the start of July 2021 as a consequence of the "dancing with Janssen" cluster of super-spreading events mentioned before. However, Figs. 2 and 3 show that the three estimates disagree on the distribution of infections over time during these events. The approach based on COVID RADAR estimates these infections to occur in two peaks, of which one occurs earlier than the single peak estimate by RIVM. On the other hand, the approach based on both COVID RADAR and CoronaMelder data estimates the single peak in infections to occur later than the RIVM estimate.

Fig. 4 shows the number of infectious individuals as estimated by our methods and by RIVM, and Fig. 5 shows the ratios between our estimates and those of RIVM. The results in Fig. 4 indicate that all three estimates display the same general patterns and waves of infections. However, Fig. 5 shows a distinct behavior between different time periods in the magnitude of the estimates. From April to September 2020, the estimate based on COVID RADAR data is in the same order of magnitude as the RIVM estimate and often larger. However, from September 2020 on, both estimates increase and from October 2020 on, their proportion remains stable at 0.41 on average. The ratio between the RIVM estimate and that based on both COVID RADAR and CoronaMelder data fluctuates a bit more and is on average 0.34. The sudden increase in this ratio a at the end of June 2021 is due to the apparently early identification of the "dancing with Janssen" cluster of super-spreading events in July 2021.

3.2. Discussion of the results

The observation in Fig. 2 that all three estimates of the effective reproduction number follow the same trend underscores the added value of models that do not require fitting transmission rates to observed output. More precisely, population-level insight in the status of the epidemic that is obtained directly from the available population-aggregated self-reported and proximity-based contact data is similar to insight obtained using more detailed data and statistically estimated transmission parameters. This is also supported by the calculated Spearman coefficients, which suggest a moderate (only COVID RADAR contact data) or even strong (both contact data sources) association between our estimates and those of RIVM. We did observe, however,

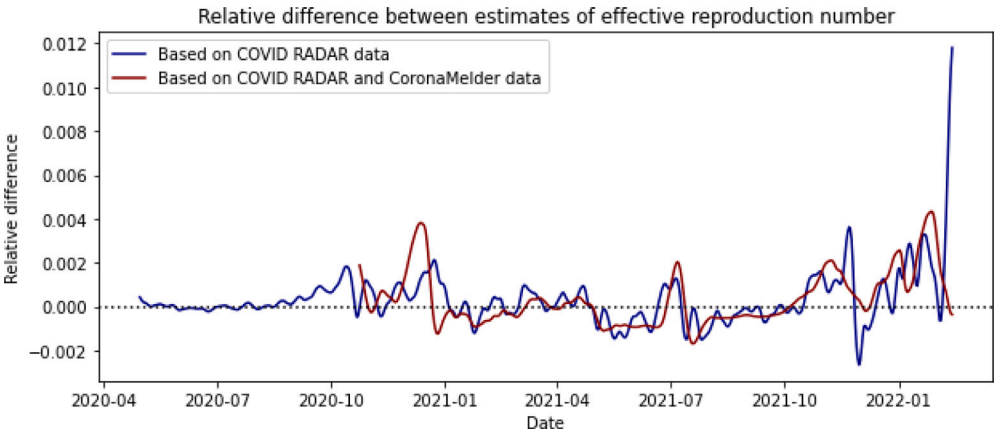

Fig. 1. Relative difference between our estimates of the effective reproduction number ($\mathcal{R}^2(t)/\mathcal{R}^1(t) - 1$).

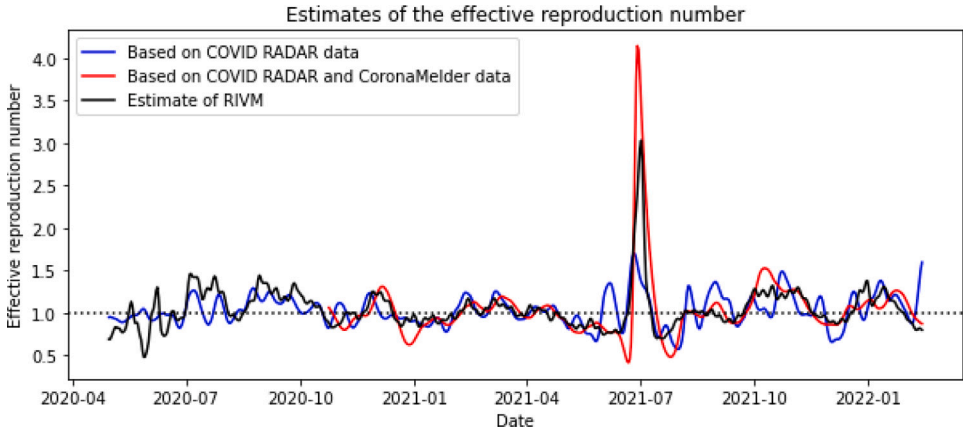

Fig. 2. Estimates of the effective reproduction number.

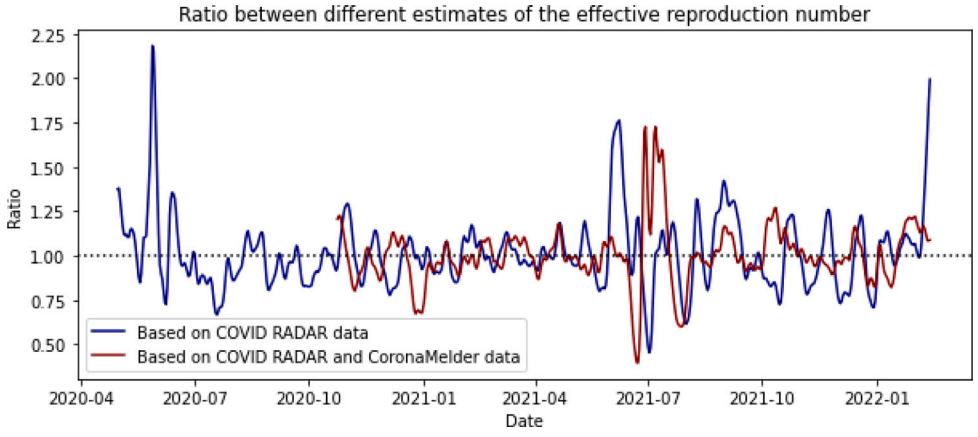

Fig. 3. Ratios between our effective reproduction number estimates and the estimate of RIVM.

that smoothing was indeed necessary to obtain interpretable results, i.e., the contact data is sparse and noisy.

The results of the cross-correlation analysis suggest that our estimates generally also agree over time, i.e., with a delay or lead time of only a few days. The lead of 4 days in the estimate based also on CoronaMelder data suggests that this estimate identifies changes in transmission dynamics earlier than the other two estimates. This supports the findings of Kendall et al. (2024), who observe similar lead times when estimating reproduction numbers for England and Wales using DCT data from the National Health Service COVID-19 app. Thus, our results provide further evidence that the use of contact data from DCT apps allows for earlier detection of changing dynamics and thus for more timely interventions to contain these changes.

Regarding the effective reproduction number, both our estimates and those of RIVM detected the increase in infections as a consequence of the super-spreading events in July 2021, but disagree on the precise magnitude and distribution of infections over time. One reason for the relatively smaller peak estimates based on COVID RADAR data is that the rise in new cases during this super-spreading event was much stronger among younger age groups. Because these groups are strongly underrepresented in the COVID RADAR user base, the users within these groups are likely to be unrepresentative of the entire age

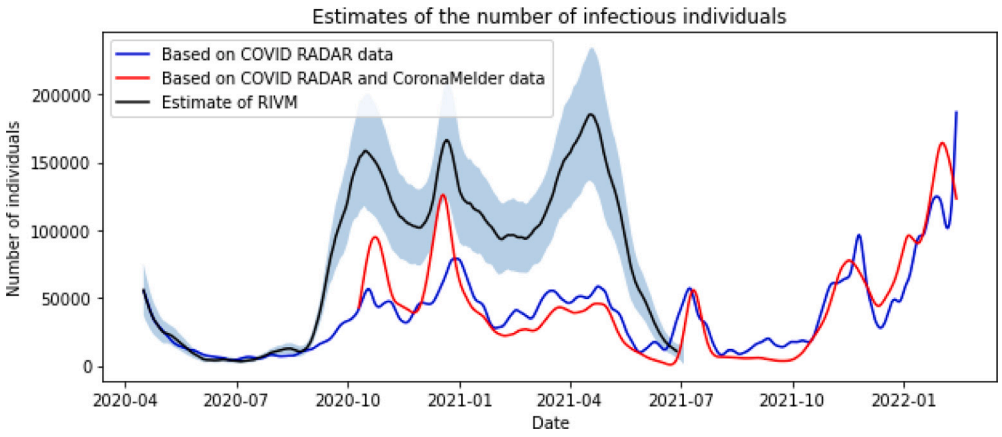

Fig. 4. Estimates of the number of infectious individuals, including the 95% confidence interval for the estimate of RIVM.

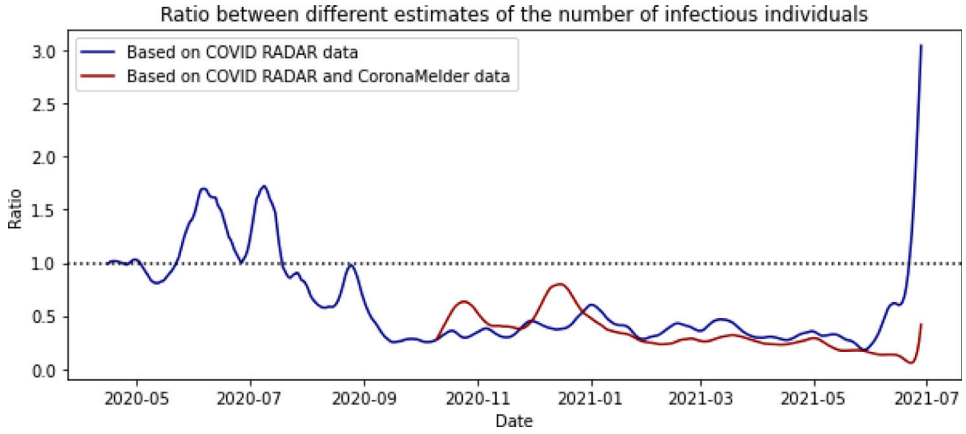

Fig. 5. Ratios between our estimates of the number of infectious individuals and the estimate of RIVM.

group (van Dijk et al., 2021) and the reweighing does not account for this type of bias.

Regarding the number of infectious individuals, both our estimates generally agree on both the timing and magnitude. Given that in both estimates the average number of contacts in (1) is estimated using purely COVID RADAR data, also the estimates of the infected contact rates generally agree. This suggests that COVID RADAR and CoronaMelder detect a similar volume of infected contacts. One exception where the estimates differ is the two peaks between October and December 2020. One reason for this could be the method for approximating the number of active users of CoronaMelder during this period. As mentioned in Section 2.1, only from May 2021 onward estimates of the number of active users have been made. Between October 2020 and May 2021, our approximation is a linear interpolation between the number of downloads at October 10, 2020 and the first publicly reported number of active users on April 30, 2021. The strong deviation of our estimates in the period October–December 2020 could thus indicate a disproportional change in active users that is not captured by the interpolation.

From October 2020 on our estimates of the number of infectious individuals are much lower than those of RIVM. One possible explanation for this is that the RIVM estimates are obtained by extrapolating the relations between age-stratified serological data and hospitalization in the period before October 2020. However, these relations could alter during the course of the pandemic due to, e.g., vaccination and hospital crowding and available capacity. In particular, new SARS-CoV-2 variants may have changed these relations due to a change in the share of and heterogeneity among COVID-19 patients who require hospitalization (de Boer et al., 2023) and increased transmissibility (Campbell et al., 2021). Another potential reason is that the RIVM estimate is less biased against asymptomatic cases. These cases are underrepresented in positive test results and thus contacts with asymptomatic infected persons are less likely to be present in both the COVID RADAR data and CoronaMelder data. In particular, the ratios in Fig. 5 are of comparable size to the estimated share of symptomatic COVID-19 cases in the Netherlands (McDonald et al., 2021).

Finally, we note that, generally, a strong violation of the random mixing assumption within the compartmental model may cause our estimate to deviate from the true number of infectious individuals. Recall that we initialize the number of infectious individuals as the share of contacts that are with infectious individuals, scaled up by the population size (1). This share can change when at the same time the total number of infectious individuals remains the same, for instance when there is a change in the proportion of infectious individuals that quarantine themselves.

4. Discussion on methods

Below we discuss limitations regarding the accuracy, sparsity, and user group bias in both the self-reported contact data from COVID RADAR and the DCT data from CoronaMelder, and regarding the modeling of transmission that may have affected the results.

Data accuracy. The data obtained via the COVID RADAR app is self-reported, meaning that the accuracy of the data depends on user's awareness and recollection of contacts and on their truthfulness and precision in reporting these contacts. Asymptomatic cases of COVID-19 often remain untested and users are therefore less likely to be aware

of contacts with infected but asymptomatic individuals (McDonald et al., 2021). Moreover, short contacts with, e.g., infected strangers are less likely to be reported because these encounters are more likely to be forgotten (Smieszek et al., 2012). Regarding truthfulness, focus group interviews with participants of the app indicate that users were generally motivated to participate out of altruism, collectivism, and a desire to help science in developing useful public health surveillance tools (Splinter et al., 2022). Thus, it seems unlikely that users purposefully submit untruthful responses on a large scale. However, regarding precision, these interviews also suggest that the formulation of the questions could be simplified or made more precise. In particular, users may interpret questions differently, which reduces the reliability of the given answers. The CoronaMelder app, like many other DCT apps, used BlueTooth Low Energy to detect other nearby devices that had the app installed (ter Haar et al., 2023). However, the proximity detection range for this type of BlueTooth is not fixed and may range from 2 to 10 m, where the strength of the signal depends on device orientation and obstacles such as indoor walls (Leith and Farrell, 2020). This results in an unknown number of both false positives and negatives that influences the number of app users that receive a notification.

Data sparsity. Variance is introduced due to sparsity in both contact data sources. The COVID RADAR data is sparse due to the relatively low app uptake rate (less than 2% of the Dutch population) and low survey submission rate (less than 5% of the maximum number of surveys that could have been submitted by all unique users). The sparsity in the used CoronaMelder data is in the reported numbers of active users, i.e., we only used monthly published statistics instead of daily numbers. To arrive at our results, we therefore interpolated missing daily data on active CoronaMelder app users and smoothed the contact rates. However, interpolation and smoothing of contact data is not desirable when using our approach to obtain timely daily estimates during an ongoing epidemic because they require future data. Thus, the added value of our approach for real-time analyses during an ongoing epidemic would greatly improve when such data transformations are unnecessary, i.e., when the app uptake and reporting frequency increases. This is also evidenced by similar studies such as Kendall et al. (2024) that do not require these transformations and can produce near real-time estimates of, for instance, effective reproduction numbers.

User group bias. For both COVID RADAR and CoronaMelder, app users were not selected via a random sampling procedure. This complicates statistical analysis of the questionnaire results and DCT notification data. For the COVID RADAR data, we were able to reduce some bias by applying post-stratification with regard to age in a pre-processing step. Although this is already more than can and has been done in similar studies (Kendall et al., 2024), a more sophisticated sampling and stratification procedure would be required to fully quantify the uncertainty and demographic bias in our estimates. Furthermore, bias in both DCM data sources and those used by RIVM may also occur because of factors that are known to influence a person's willingness to get tested for COVID-19 or to use DCM apps. Some of these factors impact both aspects, such as nationality or socioeconomic status (see McDonald et al., Ritsema et al. (2022) and the references therein), but others are more complementary. For instance, on the one hand, the level of literacy (Splinter et al., 2022) and attitude towards technology in general (Jansen-Kosterink et al., 2021) specifically impact the uptake of DCM apps. On the other hand, the availability of testing locations near one's home and the ability to visit test locations located further away via, e.g., public transport specifically impact the willingness to test (see Vink et al. (2022) and the references therein). The observation that the different estimates of the reproduction number generally agree suggests that the methods underlying these estimates are robust against such complementary biases in the input data. We do stress, however, that our approach does not quantify, and thereby explicitly account for, the influence of these biases.

Model. For the modeling of infection transmission, we used a standard SEIR compartmental model that assumes random mixing and does not include any population heterogeneity or underlying contact network structure. This is a strong simplification of existing heterogeneity in contacts (Mossong et al., 2008) and of the over-dispersed nature of COVID-19 transmission in general and of super-spreading events in particular (Sneppen et al., 2021). However, our goal was to show the added value of using coarse contact data that lacks this heterogeneity, which motivated our choice for a transmission model that does not require this. Moreover, our population-level estimates of the effective reproduction number generally agree with those of RIVM, which are based on more complex age-stratified compartmental models. Thus, our results suggest that simple models might be sufficient to monitor the population-level progression of an epidemic, despite the biases and lack of heterogeneity in the used contact data. We do believe that including overdispersion in the modeling of population-level contact behavior might improve the reliability of our estimates during super-spreading events and leave this as a direction for future research.

Data integration. When estimating the infected contact rate using COVID RADAR data (Section 2.3.1), we assumed that users report at most one contact per day with an infected person. As a consequence, the intermediate variable $z_i(t)$ represented the number of contacts with infected persons that user i had on day t. However, this variable might not capture all known contacts that user i had with infected persons. For instance, given that a user reports contact with an infected person, we do not know with how many different infected persons such a contact occurred. Moreover, if a user is aware of multiple contacts with infected persons within 14 days, the variable $z_i(t)$ might not capture the first of these contacts. This cannot be remedied by considering reverse difference in subsequent reported values, i.e., when a user does not report contact with an infected person in the last 14 days on day t but does report this on day $t+1$. Such a situation does not necessarily mean that a contact with an infected person occurred at time $t-14$ because at the time of the contact, the user might not know yet that the other person was infected and only learn about this several days later.

Despite these limitations, we believe that this did not affect our results too much because of the sparsity of the available data. The frequency with which contacts with infected persons are reported in the COVID RADAR data is already very low (2.31% of all submitted surveys), as well as the overall contact rate. This suggests that, in this particular dataset, users were less likely to be aware of multiple contacts with infected persons. However, in general, we do believe that more information on the number of contacts and the precise contact date would improve the reliability of our approach.

5. Conclusions and recommendations

The goal of this paper was to demonstrate how aggregated data on contacts with infected individuals, obtained from digital contact monitoring (DCM) apps, can inform mathematical transmission models and can be used to estimate key indicators such as the effective reproduction number. For this, we integrated both self-reported contact data from COVID RADAR (van Dijk et al., 2021) and contact tracing information from the Dutch digital contact tracing (DCT) app CoronaMelder (Ministerie van Volksgezondheid, Welzijn en Sport, 2022b) in a standard susceptible–exposed–infectious–recovered compartmental model. We showed that, despite any biases in the underlying contact data, the estimates that our approach produces agree with those made using more detailed data on hospitalizations and seroprevalence research.

We conclude this paper with several recommendations for improvements in DCM apps such as COVID RADAR and CoronaMelder that may increase their value in monitoring and modeling the state of an epidemic. Our approach requires only two daily aggregated types of contact information, namely the average number of overall contacts

that could facilitate transmission for the infectious disease in question and the average number of such contacts that were with an infected individual. The success of our approach depends on the availability and quality of these two quantities. Thus, our recommendation is to include the collection of this information in future apps.

The COVID RADAR app already provides information on overall contacts directly. However, information on contacts with infected individuals had to be inferred by comparing consecutive reports from the same individual on whether such a contact occurred within the last 14 days. This means that only after 14 days we can infer the exact time of a contact with an infected person. As a consequence, we can only estimate the effective reproduction number with a delay of 14 days, which limits the usability of the method during an ongoing epidemic where there is a high urgency to obtain timely estimates. To obtain more timely data on the time of infected contacts and thus allow for more timely estimates of the effective reproduction number, we therefore recommend that a future version of the app includes an additional question that asks for the specific date on which the contact with the infected individual occurred.

Regarding the DCT app CoronaMelder, no data was available on the daily number of close contacts that a given user had with other users who installed the app. However, this information was actually recorded and temporarily stored on individual users' devices. In fact, every day, each device compared these contacts with a daily distributed list of positively tested app users and notified their user if there was a match. In the current app design, the "local" lists of contacts stayed on the individual device and were deleted after 14 days. However, we recommend that in a next version of the app the number of contacts is communicated back and aggregated on a daily basis to provide an estimate for the overall average contact rate. This has also been done in other DCM apps that used the same decentralized design (Wymant et al., 2021). Moreover, also the daily number of matches should be communicated and the date on which the corresponding contact occurred. Using these aggregated data would improve the estimation of the daily number of received notifications, which we now resolved by incorporating data from COVID RADAR. The current design of the app did not record and store this information but could, similarly to the COVID RADAR app, be adapted to keep track of this information without violating individual users' privacy. In particular, no information on individual users other than their number of contacts would need to be shared, which is already less information than what is shared in the COVID RADAR app.

CRediT authorship contribution statement

Martijn H.H. Schoot Uiterkamp: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Willian J. van Dijk:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Hans Heesterbeek:** Writing – review & editing, Validation, Methodology, Funding acquisition, Conceptualization. **Remco van der Hofstad:** Writing – review & editing, Validation, Methodology, Funding acquisition, Conceptualization. **Jessica C. Kiefte-de Jong:** Writing – review & editing, Validation, Methodology, Funding acquisition, Conceptualization. **Nelly Litvak:** Writing – review & editing, Validation, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Co-author was founding editor and editor-in-chief (2008–2019) of the journal to which this work is submitted - H.H. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the editors and reviewers for their thorough assessment of the manuscript and helpful comments and feedback. This work is supported by Netherlands Organisation for Scientific Research (NWO) through ZonMw grants no. 10430032010011, 10430372310021, and 10710062310008. R.v.d.H. and N.L. are also supported by NWO through Gravitation NETWORKS grant no. 024.002.003.

Data availability

All used data is publicly available via the corresponding cited references. The code to reproduce the results is available at https://github.com/mhhschootuiterkamp/DCM-data-integration-in-SEIR-models.

References

Anglemyer, A., Moore, T., Parker, L., Chambers, T., Grady, A., Chiu, K., et al., 2020. Digital contact tracing technologies in epidemics: a rapid review. Cochrane Database Syst Rev. (8).

Backer, J.A., Eggink, D., Andeweg, S.P., Veldhuijzen, I.K., van Maarseveen, N., Vermaas, K., et al., 2022. Shorter serial intervals in SARS-CoV-2 cases with omicron ba.1 variant compared with delta variant, the netherlands 13 to 26 december 2021.. Eurosurveill 27 (6).

Backer, J.A., Mollema, L., Vos, E.R.A., Klinkenberg, D., van der Klis, F.R.M., de Melker, H.E., et al., 2021. Impact of physical distancing measures against COVID-19 on contacts and mixing patterns: repeated cross-sectional surveys, the netherlands, 201617, april 2020 and june 2020. Eurosurveill 26 (8), 2000994.

Beutels, P., Shkedy, Z., Aerts, M., van Damme, P., 2006. Social mixing patterns for transmission models of close contact infections: exploring self-evaluation and diary-based data collection through a web-based interface. Epidemiol. Infect. 134 (6), 1158–1166.

Bradford, L., Aboy, M., Liddell, K., 2020. COVID-19 contact tracing apps: A stress test for privacy, the GDPR, and data protection regimes. J Law Biosci. 7 (1), 05, lsaa034.

Byrne, A.W., McEvoy, D., Collins, A.B., Hunt, K., Casey, M., Barber, A., et al., 2020. Inferred duration of infectious period of SARS-CoV-2: Rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases. BMJ Open. 10 (8).

Campbell, F., Archer, B., H, Laurenson-Schafer, Jinnai, Y., Konings, F., Batra, N., et al., 2021. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at june 2021. Eurosurveill. 26 (24).

Cori, A., Ferguson, N.M., Fraser, C., Cauchemez, S., 2013. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol. 178 (9), 1505–1512, 09.

de Boer, PT, van de Kassteele, J., Vos, E.R.A., van Asten, L., Dongelmans, D.A., van Gageldonk-Lafeber, AB, et al., 2023. Age-specific severity of severe acute respiratory syndrome coronavirus 2 in february 2020 to june 2021 in the netherlands. Influ. Other Respir Viruses. 17 (8), e13174.

Dolman, T.L., 2021. Tabellenrapport CoronaMelder GGD GHOR Nederland. Available from: https://www.rijksoverheid.nl/documenten/publicaties/2021/05/28/rapporten-evaluatie-coronamelder-9-maanden. (Accessed on 16 October 2024).

Dongelmans, D.A., Termorshuizen, F., Brinkman, S., F, Bakhshi-Raiez, Arbous, M.S., de Lange, DW, et al., 2022. Characteristics and outcome of COVID-19 patients admitted to the ICU: A nationwide cohort study on the comparison between the first and the consecutive upsurges of the second wave of the COVID-19 pandemic in the netherlands. Ann Intensiv. Care. 12 (5).

Ebbers, W., Hooft, L., van der Laan, N., Metting, E., 2021. Evaluatie CoronaMelder: Een overzicht na 9 maanden. Ministerie van Volksgezondheid, Welzijn en Sport, Available from: https://www.rijksoverheid.nl/documenten/publicaties/2021/05/28/rapporten-evaluatie-coronamelder-9-maanden. (Accessed on 11 January 2025).

Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., L, Abeler-Dörner, et al., 2020. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Sci. 368 (6491), eabb6936.

Hoang, T., Coletti, P., Melegaro, A., Wallinga, J., Grijalva, C.G., Edmunds, J.W., et al., 2019. A systematic review of social contact surveys to inform transmission models of close-contact infections. Epidemiol. 30 (5), 723–736.

Jansen-Kosterink, S., Hurmuz, M, den Ouden, M., L., van Velsen, 2021. Predictors to use mobile apps for monitoring covid-19 symptoms and contact tracing: survey among dutch citizens. JMIR Form Res. 5 (12), e28416.

Kendall, M., Ferretti, L., Wymant, C., Tsallis, D., Petrie, J., Francia, A.D., et al., 2024. Drivers of epidemic dynamics in real time from daily digital COVID-19 measurements. Sci. 385 (6710), eadm8103.

Leith, D.J., Farrell, S., 2020. Coronavirus contact tracing: evaluating the potential of using bluetooth received signal strength for proximity detection. SIGCOMM Comput. Commun Rev. 50 (4), 66–74.

Leung, K.Y., Metting, E., Ebbers, W., Veldhuijzen, I., Andeweg, S.P., Luijben, G., et al., 2024. Effectiveness of a COVID-19 contact tracing app in a simulation model with indirect and informal contact tracing. Epidemics. 46, 100735.

Masel, J., Petrie, J.I.M., Bay, J., Ebbers, W., Sharan, A., Leibrand, S.M., et al., 2023. Combatting SARS-CoV-2 with digital contact tracing and notification: Navigating six points of failure. 9, e49560,

McDonald, S.A., Miura, F., ERA, Vos, van Boven, M., de Melker, H.E., van der Klis, F.R.M., et al., 2021. Estimating the asymptomatic proportion of SARS-CoV-2 infection in the general population: Analysis of nationwide serosurvey data in the netherlands. Eur. J. Epidemiol. 36, 735–739.

McDonald, S.A., Soetens, L.C., Schipper, C.M.A., Friesema, I., van den Wijngaard CC, Teirlinck A, et al., Testing behaviour and positivity for SARS-CoV-2 infection: insights from web-based participatory surveillance in the netherlands. BMJ Open. 2021 (11), e056077.

Ministerie van Volksgezondheid, Welzijn en Sport, 2022a. Coronamelder data dashboard. Via Archive.today. Available from: https://archive.ph/2022.04.06-092154/https:/coronamelder.nl/nl/faq/1-13-coronamelder-data-dashboard/. (Accessed on January 11 2025).

Ministerie van Volksgezondheid, Welzijn en Sport, 2022b. Coronamelder statistics. Dataset available from https://github.com/minvws/nl-covid19-notification-app-statistics.

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., et al., 2008. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLOS Med. 5 (3), e74, 03.

Read, J.M., Edmunds, W.J., Riley, S., Lessler, J., Cummings, D.A.T., 2012. Close encounters of the infectious kind: methods to measure social mixing behaviour. Epidemiol. Infect. 140 (12), 2117–2130.

Rijksinstituut voor Volksgezondheid en Milieu, COVID-19 dataset. Dataset available from. https://data.rivm.nl/covid-19/.

Ritsema, F., Bosdriesz, J.R., Leenstra, T., Petrignani, M.W.F., Coyer, L., Schreijer, A.J.M., et al., 2022. Factors associated with using the COVID-19 mobile contact-tracing app among individuals diagnosed with SARS-CoV-2 in amsterdam, the netherlands: Observational study. JMIR Mhealth Uhealth. 10 (8), e31099.

Shahroz, M., Ahmad, F., Younis, M.S., Ahmad, N., MN, Kamel Boulos, R, Vinuesa, et al., 2021. COVID-19 digital contact tracing applications and techniques: A review post initial deployments. Transp Eng. 5, 100072.

Smieszek, T., Burri, E.U., Scherzinger, R., Scholz, R.W., 2012. Collecting close-contact social mixing data with contact diaries: reporting errors and biases. Epidemiol. Infect. 140 (4), 744–752.

Sneppen, K., Nielsen, B.F., Taylor, R.J., Simonsen, L., 2021. Overdispersion in COVID-19 increases the effectiveness of limiting nonrepetitive contacts for transmission control. Proc Nat Acad Sci. 118 (14), e2016623118.

Splinter, B., Saadah, N.H., Chavannes, N.H., Kiefte-de Jong, J.C., Aardoom, J.J., 2022. Optimizing the acceptability, adherence, and inclusiveness of the COVID radar surveillance app: Qualitative study using focus groups, thematic content analysis, and usability testing. JMIR Form Res. 6 (9), e36003.

Stein, C., Nassereldine, H., Sorensen, R.J.D., Amlag, J.O., Bisignano, C., Byrne, S., et al., 2023. Past SARS-CoV-2 infection protection against re-infection: a systematic review and meta-analysis. Lancet. 401 (10379), 833–842.

ter Haar, W., Bosdriesz, J., Venekamp, R.P., Schuit, E., van den Hof, S., Ebbers, W., et al., 2023. The epidemiological impact of digital and manual contact tracing on the SARS-CoV-2 epidemic in the netherlands: Empirical evidence. PLOS Digit. Health. 2 (12), 1–17, 12.

van de Kassteele, J., Eilers, P.H.C., Wallinga, J., 2019. Nowcasting the number of new symptomatic cases during infectious disease outbreaks using constrained p-spline smoothing. Epidemiol. 30, 737–745.

van der Laan, L.N., de Wit, J.M.S., 2022. Factoren relevant voor het stoppen met gebruik van de coronamelder. Tilburg University, Available from: https://research.tilburguniversity.edu/en/publications/factoren-relevant-voor-het-stoppen-met-gebruik-van-de-coronamelde.

van der Veer, B.M.J.W., Gorgels, K.M.F., den Heijer, C.D.J., Hackert, V., van Alphen, L.B., Savelkoul, P.H.M., et al., 2023. SARS-CoV-2 transmission dynamics in bars, restaurants, and nightclubs. Front. Microbiol 14.

van Dijk, W.J., 2020. COVID RADAR app. DANS. Dataset available from https://doi.org/10.17026/dans-zcd-m9dh.

van Dijk, W.J., Saadah, N.H., Numans, M.E., Aardoom, J.J., Bonten, T.N., Brandjes, M., et al., 2021. COVID RADAR app: Description and validation of population surveillance of symptoms and behavior in relation to COVID-19. PLoS One 16 (6), 06.

Verelst, F., Hermans, L., Vercruysse, S., Gimma, A., Coletti, P., Backer, J.A., et al., 2021. SOCRATES-CoMix: A platform for timely and open-source contact mixing data during and in between COVID-19 surges and interventions in over 20 European countries. BMC Med. 19 (254).

Vink, M., Z, Iglói, Fanoy, E.B., van Beek, J., Boelsums, T., de Graaf, M., et al., 2022. Community-based SARS-CoV-2 testing in low-income neighbourhoods in rotterdam: Results from a pilot study. J Glob Health. 1, (12:05042).

Vos, E.R.A., den Hartog, G., Schepp, R.M., Kaaijk, P., van Vliet, J., Helm, K., et al., 2021. Nationwide seroprevalence of SARS-CoV-2 and identification of risk factors in the general population of the netherlands during the first epidemic wave. J Epidemiol Community Health. 75, 489–495.

Wahba, G., 1990. Estimating the smoothing parameter. In: Spline Models for Observational Data. Philadelphia, PA:. Society for Industrial and Applied Mathematics, pp. 45–65.

Wallinga, J., Lipsitch, M., 2007. How generation intervals shape the relationship between growth rates and reproductive numbers. Proc R Soc B Biol Sci. 274 (1609), 599–604.

Wymant, C., Ferretti, L., Tsallis, D., Charalambides, M., L, Abeler-Dörner, Bonsall, D., et al., 2021. The epidemiological impact of the NHS COVID-19 app. Nat. (594), 408–412.

Xin, H., Li, Y., Wu, P., Li, Z., Lau, E.H.Y., Qin, Y., et al., 2021. Estimating the latent period of coronavirus disease 2019 (COVID-19). Clin Infect Dis. 74 (9), 1678–1681, 09.

Xu, F., Tu, Z., Li, Y., Zhang, P., Fu, X., Jin, D., 2017. Trajectory recovery from ash: User privacy is NOT preserved in aggregated mobility data. In: Proceedings of the 26th International Conference on World Wide Web. WWW'17. Republic and Canton of Geneva. CHE: International World Wide Web Conferences Steering Committee, pp. 1241–1250.