Estimating social contact rates for the COVID-19 pandemic using Google mobility and pre-pandemic contact surveys

# Abstract

## Background

During the COVID-19 pandemic, aggregated mobility data (provided by Google, Apple, Facebook, and other corporations) was frequently used to estimate changing social contact rates over time. By taking contact matrices estimated before the pandemic, and transforming these using mobility data obtained during the pandemic, epidemiologists were able to predict the number of contacts individuals were expected to have during large-scale restrictions. This investigation determines the most appropriate method of performing this transformation, using pandemic-era contact surveys as a measure of accuracy for these transformations.

## Methods

We compared four methods for scaling synthetic contact matrices, two methods using fitted regression models to relate mobility to numbers of contacts and two methods using “naïve” mobility or mobility squared models to scale the matrices. The regression models were fitted using CoMix contact survey and Google mobility data from the UK over March 2020 – March 2021. The four models were then used to scale synthetic contact matrices—a representation of pre-pandemic behaviour—using mobility data from the UK, Belgium and the Netherlands to predict the number of contacts in “work” and “other” settings expected for a given mobility level. We then compared partial reproduction numbers estimated from the four models with partial reproduction numbers calculated directly from CoMix contact matrices across the three countries. The accuracy of each model was assessed using root mean squared error.

## Results

The fitted regression models were substantially more accurate in predicting reproduction numbers than the naïve models, even when the regression models, which were fitted only to UK data, were applied to Belgium and the Netherlands. In the UK case, the most accurate regression model included a quadratic term, whereas for Belgium and the Netherlands, the most accurate regression model included a linear term only. Across all countries investigated, the naïve model using mobility alone was the least accurate, followed by the naïve model using mobility squared.

## Conclusion

When attempting to estimate social contact rates during a pandemic without the resources available to conduct contact surveys, using a model fitted to data from another pandemic context is likely to be an improvement over using a “naïve” model based on raw mobility data. If a naïve model is to be used, mobility squared may be a better predictor of contact rates than mobility per se.

# 1. Introduction

The COVID-19 pandemic led to millions of deaths worldwide [(1)](https://www.zotero.org/google-docs/?N9WWig) and motivated the implementation of a huge number of non-pharmaceutical interventions in order to control transmission, alongside the deployment of large-scale contact surveys and measurement of mobility data across the world. Both contact surveys and mobility data can give us insights into changes in interpersonal contact, which drives transmission. Contact rates are a way of measuring transmission potential and contact surveys can be used to ascertain these rates. Contact surveys are conducted to measure who people come into physical or conversational contact with over a period of time, typically a 24-hour period. One such survey was the CoMix survey [(2)](https://www.zotero.org/google-docs/?tEUd1w) which began in the UK [(3)](https://www.zotero.org/google-docs/?XSUkTx), Belgium [(4)](https://www.zotero.org/google-docs/?YJ4sXR) and the Netherlands [(5)](https://www.zotero.org/google-docs/?Q3Kkgz) in 2020 and was subsequently slowly incorporated in many countries in Europe [(6)](https://www.zotero.org/google-docs/?UIN8xF). The CoMix survey aimed to monitor behavioural changes during the unfolding of the pandemic, with a particular attention at estimating the impact of large scale mitigation strategies, such as lockdowns.

Contact surveys provide measures of contact rates, which are used as a proxy of epidemiological risk behaviour, and are often designed so that the results can be directly used to parameterise infectious disease models which take social contact rates as an input [(7,8)](https://www.zotero.org/google-docs/?PLIdXQ). Studies such as CoMix undertook repeated social contact surveys during the pandemic, aiming at population representativity. However, these surveys are expensive and difficult to undertake, so data was collected only in a limited number of countries. In early 2020, several companies started to release aggregated “mobility” data with the stated aim of helping public health professionals to understand behavioural change in response to the pandemic and was considered a valid proxy of risk behaviour [(9)](https://www.zotero.org/google-docs/?MRA60y). In the most common instances, this mobility data quantified the time spent in different locations by mobile phone users, was made freely available, and covered participants in nearly all countries. These advantages led to the heavy use of mobility data during the pandemic to assess the impact of social distancing measures and to parameterise mathematical models, with Google’s “Community Mobility Reports” [(10)](https://www.zotero.org/google-docs/?um4Gim) being particularly widely used [(11–14)](https://www.zotero.org/google-docs/?UoDv4s). However, as the use of mobility data as a proxy for social contact rates has not been formally assessed, it remains unclear how precisely to transform relative changes in mobility in different locations into changes in contact rates, and how well such transformations predict measured contact rates.

# 2. Methods

In this study, we compare trends in mean contacts recorded in the CoMix social contact survey with Google mobility data and assess the performance of different methods used to relate mobility to social contacts. Specifically, we investigated the relationship between specific Google mobility indicators and corresponding contact types from the CoMix dataset for the UK, and developed a series of statistical and mechanistic models to relate the data to one another. We then used these models to estimate contact rates from mobility data for the UK, Belgium, and the Netherlands, in order to determine the predictive accuracy of each model. Our results serve as a tool for epidemiologists and infectious disease modellers to understand better how mobility data might relate to contact rates in a population.

## 2.1 Study Design

The study design and method of informed consent for the CoMix study were approved by the ethics committee of the London School of Hygiene and Tropical Medicine (reference number 21795) in the UK, the Ethics Committee of UZA (reference 3236 - BUN B3002020000054) in Belgium and need for approval was waived by the Medical Research Ethics Committee (MREC) NedMec (research protocol number 22/917) in the Netherlands. All analyses were carried out on anonymised participant data. Additionally, Google mobility data was obtained from publicly available sources [(10)](https://www.zotero.org/google-docs/?Ouv0Cx).

### 2.1.1 Mobility data

The mobility data uses a median value from a 5-week baseline period of 02/01/2020 to 06/02/2020 to compare changes in the usage of specific services/areas. The raw data is expressed as a percentage change relative to the baseline, e.g. -50% for half as many visits as during the baseline period and +100% for twice as many visits as during the baseline period. We transformed this to a scale which expresses the change from baseline as a multiple, e.g. 0.5 for half as many visits as compared to baseline and 2 for twice as many visits as compared to baseline. We included the following mobility indicators in analyses that measure the number of visits:

* Retail and recreation
* Grocery and pharmacy
* Public transit
* Workplaces

Data is available between 23/03/2020 to 13/10/2022. We were unable to find clear information in regard to how Google classified visits into these categories [(15)](https://www.zotero.org/google-docs/?TKZJeK).

### 2.1.2 Survey data

The CoMix survey collected information on contacts weekly in the UK from 23/03/2020 to 01/03/2022 [(2)](https://www.zotero.org/google-docs/?DqR6rj). The present study includes weekly survey results for the entire study period and subsequent statistical analyses were limited to between 23/03/2020 to 31/03/2021 due to poor correlations in the second year. Adult panels at first contained 1500 participants, and increased to include 2500 participants from August 2020. Participants from each panel were surveyed once every 2 weeks, with panels alternating so that each week was covered. Ipsos MORI used quota sampling, for age, gender, and region, to recruit a representative sample of the UK [(16)](https://www.zotero.org/google-docs/?G4YBvH). The survey followed the design of the 2005/2006 POLYMOD survey [(17)](https://www.zotero.org/google-docs/?7H7YKR) with some additional questions. More details on the CoMix survey can be found in Gimma et al. and Jarvis et al. [(3,16)](https://www.zotero.org/google-docs/?YQ4Mn5). Survey participants were asked to report the number of people they met on the day prior to the survey in various settings. Participants had the option of recording any of their contacts in one of two ways: either individually (i.e. reporting details about contact made with one individual), or as mass contacts (i.e. reporting a summary of contact made with a group of several people). We processed these two contact types differently, as detailed below.

Traditionally social contact surveys report aggregated contact data as contacts having occurred at “home”, “work”, “school” or “other”, the CoMix data was also processed to reflect this usual categorisation. We chose to focus on "work" and "other" contacts, as they would have the greatest potential for variability amongst adults. We excluded children's contacts as they were reported by their parents. We did not assess home contacts as that is more plausibly related to household size as opposed to number of visits.

## 2.2 Data Preparation

We included participants aged 18 and over, but did not restrict contact ages. As per previous CoMix analyses [(16)](https://www.zotero.org/google-docs/?rEqrQZ), we excluded survey rounds six and seven due to data collection issues due to an ad-hoc change to the questionnaire resulting in fewer contacts reported for those weeks. Panels were sampled from every other week, with surveys interchanging the panel from which participants were sampled to ensure weekly data was collected. To average out differences in behaviour between panels, we averaged daily “work” and “other” contacts over a two-week moving window to smooth out panel effects. We reweighted the sample based on age, social class and whether the response was given on a weekend or not (see supplementary material for characteristics over time). This was to ensure that changes in the recruitment process over the course of the survey did not influence the investigation, and to improve the generalizability of the study. For the weekend-weekday reweighting, we reweighted the sample so that contacts made on weekends comprised 2/7ths of the total sampling weight. This was done as approximately 25% of observations were on a Saturday or Sunday and the number of contacts on these days was generally lower.

Occasionally, some participants reported mass contacts in extremely high numbers, such as 3000 and higher. These occasional large mass-contact events introduced substantial noise into our estimates of the mean contact rate over time. To stabilise these estimates, mass contacts were capped at 50 contacts per participant, per date, for each contact type, resulting in the reduction of 6% in the total number of “work” contacts and of 7% in the total number of “other” contacts. We capped contacts by randomly sampling 50 contacts in each contact type for each participant for each date, when the contacts were listed as ‘mass’. Other analyses performed on this data implemented a similar approach and hence for consistency it was applied to this investigation [(16)](https://www.zotero.org/google-docs/?IYRip4).

## 2.3 Statistical Analysis

R version 4.2.1 was used for all analyses [(18)](https://www.zotero.org/google-docs/?wr7hFg). Code and data are available on GitHub [(19)](https://www.zotero.org/google-docs/?d2blOz). We conducted two broad analyses: comparing Google mobility to contact rates, and investigating ways to use Google mobility data to scale pre-pandemic contact rates. Through the second analysis we used four methods to scale contacts, comparing methods using Google mobility alone (naïve models) and methods built using CoMix data and Google mobility data.

### 2.3.1 Comparing Mobility Indicators to Contacts

We first visually compared the mobility indicators to the relevant contacts: "work" contacts were compared against the "workplaces" mobility indicator and "other" contacts against the mean of the “retail and recreation”, “transit stations”, and “grocery and pharmacy” mobility indicators.

### 2.3.2 Creating Relative Contact Rates

The aim of this investigation is to determine how to use Google mobility as a proxy for social contact rates during a pandemic, specifically by comparing four different models—two “naïve” models and two regression models—relating mobility to relative contact rates. The two naïve models are commonly encountered in the literature [(12–14)](https://www.zotero.org/google-docs/?5nzTPo), while the regression approach is less commonly used [(20)](https://www.zotero.org/google-docs/?PNIRVH) and the specific regression models we analyse are fitted in this paper.

The two “naïve” models are based on first principles, namely that contacts are either directly proportional to mobility (“mobility” model) or to mobility squared (“mobility squared” model). The “mobility” model [(12)](https://www.zotero.org/google-docs/?a8MQka) assumes that given a visit is made to a particular type of venue, a person makes the same number of contacts. An example of this would be going on a date, your only contact would be the person you were going on a date with. The “mobility squared” model [(13,14)](https://www.zotero.org/google-docs/?u1PkDz) assumes that, given a visit is made to a particular type of venue, the number of contacts depends upon the number of other people making visits to the same type of venue. In other words, the “mobility” model assumes that contacts are made through coordinated activities, and the “mobility squared” model assumes that contacts are made at random with other people at the same venues For each of the naïve models, a mobility value of 1 corresponds to a relative contact rate of 1, and the relative contact rate is then multiplied by the pre-pandemic contact rate estimated for a given setting to produce an estimate of the during-pandemic contact rate.

The two regression models were fitted to contact rates for the UK as measured by CoMix, with Google mobility as the independent variable. We contrasted two regression models, one with an intercept and linear term (the “linear model”) and one with an intercept, a linear term, and a quadratic term (the “quadratic model”). Like the naïve models, the regression models produce an estimate of relative contact rates from an input of mobility, which are then scaled up by a pre-pandemic “baseline” contact rate, or used as a multiplicative factor on a pre-pandemic “baseline” contact matrix, to yield contact rates or a contact matrix. The pre-pandemic “baseline” was provided by estimates for the UK as measured by the 2006 POLYMOD survey [(17)](https://www.zotero.org/google-docs/?XsQ9HK) - see supplementary material for a more detailed description. The naïve models were each fitted separately to “work” contacts from CoMix, using Google mobility "workplace" visits as the predictor, and to “other” contacts from CoMix, using the composite Google mobility "other" visits measure as the predictor, for the UK.

The four models obtained above were used to predict relative contact rates from Google mobility data for the UK, Belgium, and the Netherlands. To assess the performance of the four models in the context of methods people may use during an emergency when there is no contact data available, we used synthetic contact matrices [(21)](https://www.zotero.org/google-docs/?XMxwi9). We chose the synthetic matrices to increase generalisability as these matrices are available for 177 countries and therefore can be used in the majority of the countries of the world. These matrices were then scaled by the relative contact rates yielded by each model. We assumed that the estimated relative contact rate we have estimated for adults can be extended to children without adjustment. As this investigation focuses on “work” and “other” we assumed that transmission drivers would be similar for adults and children.

From the subsequent scaled matrices, we then calculated the dominant eigenvalues for each fortnight in the study period. In order to translate these results into something understandable, we transformed the dominant eigenvalues into partial reproduction numbers. This was done by comparing early (07/03/2020-13/03/2020) estimates of Rt (number of secondary infections generated by one infected person) from the EpiForecast work [(22)](https://www.zotero.org/google-docs/?nywuFi) and the dominant eigenvalue for the synthetic matrix for ‘all’ contacts for the UK [(21)](https://www.zotero.org/google-docs/?spjrjB). This gave us a multiplicative factor to use to transform the dominant eigenvalues we calculated, into partial reproduction numbers. These partial reproduction numbers are only components of the true reproduction number as they are specific to "work" and "other" contacts. This multiplicative factor was calculated as follows:

We assess the accuracy of the four models in predicting partial reproduction numbers both visually and quantitatively, in the latter case using root mean squared error. As a ‘true’ value for the partial reproduction numbers we calculated the partial reproduction number for the CoMix contact matrices for the respective contact type and country.

# 3. Results

Here we present the results of the broad analyses, as well as the characteristics of the participants used in this investigation.

## 3.1 Participant Characteristics

For the UK there were 121,057 surveys completed in the study period, filled out by 17,497 participants. Characteristics of both the surveys completed and the individual participants are provided in table 1. As participants were sampled multiple times we see that characteristic distributions are slightly different between participants and observations. We see that characteristics were broadly representative of the British population. There was a lower percentage of male participants. There was also a slightly larger proportion of middle class and lower middle class participants and a slightly smaller proportion of skilled working class and working class participants, when compared to the population distribution. The proportion of employed versus unemployed is much higher due to the inclusion of older participants who are likely retired. When restricted to participants under 65 (see supplementary material) the proportion of employed versus unemployed matches the population proportions (which is for those of employment age).

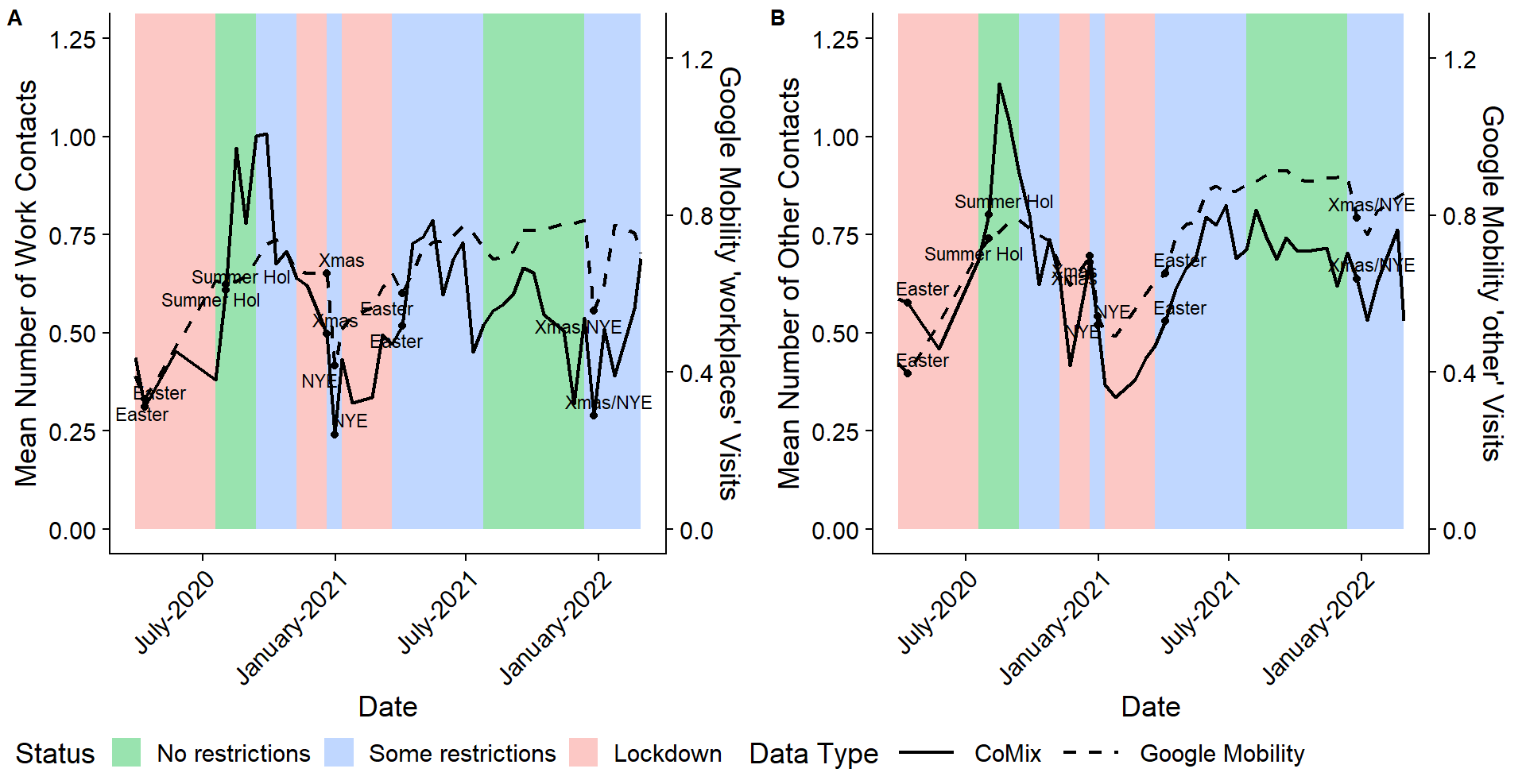
| ***Characteristic*** | ***Number of Participants (%)*** | ***Number of Observations (%)*** | ***Population Proportion %*** |
| --- | --- | --- | --- |
| ***Gender*** | | | |
| Female | 8,591 (49.1) | 49,826 (50.8) | 50.75 |
| Male | 7,832 (44.8) | 45,367 (46.3) | 48.75 |
| Other | 48 (0.3) | 259 (0.3) | 0.5 |
| Missing | 1,026 (5.9) | 2,629 (2.7) | - |
| ***Age-group*** | | | |
| 18-29 | 3,039 (17.4) | 12,552 (12.8) | 18.8 |
| 30-39 | 3,035 (17.3) | 15,401 (15.7) | 17.2 |
| 40-49 | 2,883 (16.5) | 16,279 (16.6) | 16 |
| 50-59 | 3,206 (18.3) | 19,396 (19.8) | 17.3 |
| 60-69 | 3,213 (18.4) | 20,691 (21.1) | 13.6 |
| 70+ | 2,121 (12.1) | 13,762 (14.0) | 17.2 |
| ***Employment Status*** | | | |
| Employed | 7,508 (42.9) | 44,007 (44.9) | 75 |
| Unemployed | 9,989 (57.1) | 54,074 (55.1) | 25 |
| ***Social Class*** | | | |
| A - upper middle class | 806 (4.6) | 4,554 (4.6) | 4 |
| B - middle class | 4,638 (26.5) | 25,820 (26.3) | 23 |
| C1 - lower middle class | 5,626 (32.2) | 32,450 (33.1) | 29 |
| C2 - skilled working class | 2,889 (16.5) | 15,561 (15.9) | 21 |
| D - working class | 2,479 (14.2) | 14,100 (14.4) | 15 |
| E - lower level of subsistence | 1,059 (6.1) | 5,596 (5.7) | 8 |
| ***Area*** | | |  |
| East Midlands | 1,350 (7.7) | 7,634 (7.8) | 7 |
| East of England | 1,664 (9.5) | 9,504 (9.7) | 9 |
| Greater London | 2,248 (12.8) | 12,889 (13.1) | 13 |
| North East | 757 (4.3) | 4,267 (4.4) | 4 |
| North West | 1,205 (6.9) | 6,571 (6.7) | 11 |
| Northern Ireland | 473 (2.7) | 2,410 (2.5) | 3 |
| Scotland | 1,592 (9.1) | 8,799 (9.0) | 8 |
| South East | 2,510 (14.3) | 13,875 (14.1) | 14 |
| South West | 1,632 (9.3) | 9,259 (9.4) | 9 |
| Wales | 922 (5.3) | 4,939 (5.0) | 5 |
| West Midlands | 1,625 (9.3) | 9,165 (9.3) | 9 |
| Yorkshire and The Humber | 1,519 (8.7) | 8,769 (8.9) | 8 |
| ***Household size group*** | | |  |
| 1 | 3,712 (21.2) | 23,211 (23.7) | 30 |
| 2 | 7,152 (40.9) | 42,874 (43.7) | 35 |
| 3-5 | 6,276 (35.9) | 30,629 (31.2) | 30 |
| 6+ | 357 (2) | 1,367 (1.4) | 5 |

**Table 1:** Participant characteristic summary; an observation is a complete survey response; percentages rounds to 1 decimal place so may not add to 100. Most population proportions were from the 2021 census [(23)](https://www.zotero.org/google-docs/?AiLYBV) aside from social class which comes from a 2008 IPSOS social class report [(24)](https://www.zotero.org/google-docs/?Ze2FCg).

Table 1 illustrates why we reweighted samples based on participant age and social class, as these were the characteristics least representative of the British population.

## 3.2 Statistical Analysis

### 3.2.1 Comparing Mobility Indicators to Contacts



**Figure 1:** Contacts against mobility over time, plot A shows mean “work” contacts (solid line) and “workplace” mobility (dashed line) and plot B shows mean “other” contacts (solid line) and “other” mobility (dashed line), over time respectively.

In figure 1 we see similar trends over time between contacts rates and mobility metrics when we compare the mean daily number of "work" contacts from CoMix to the number of visits to "workplaces" in Google Mobility, and compare the mean daily number of "other" contacts from CoMix to the "other" visits (created from a weighted predictor of three mobility indicators - see methods).

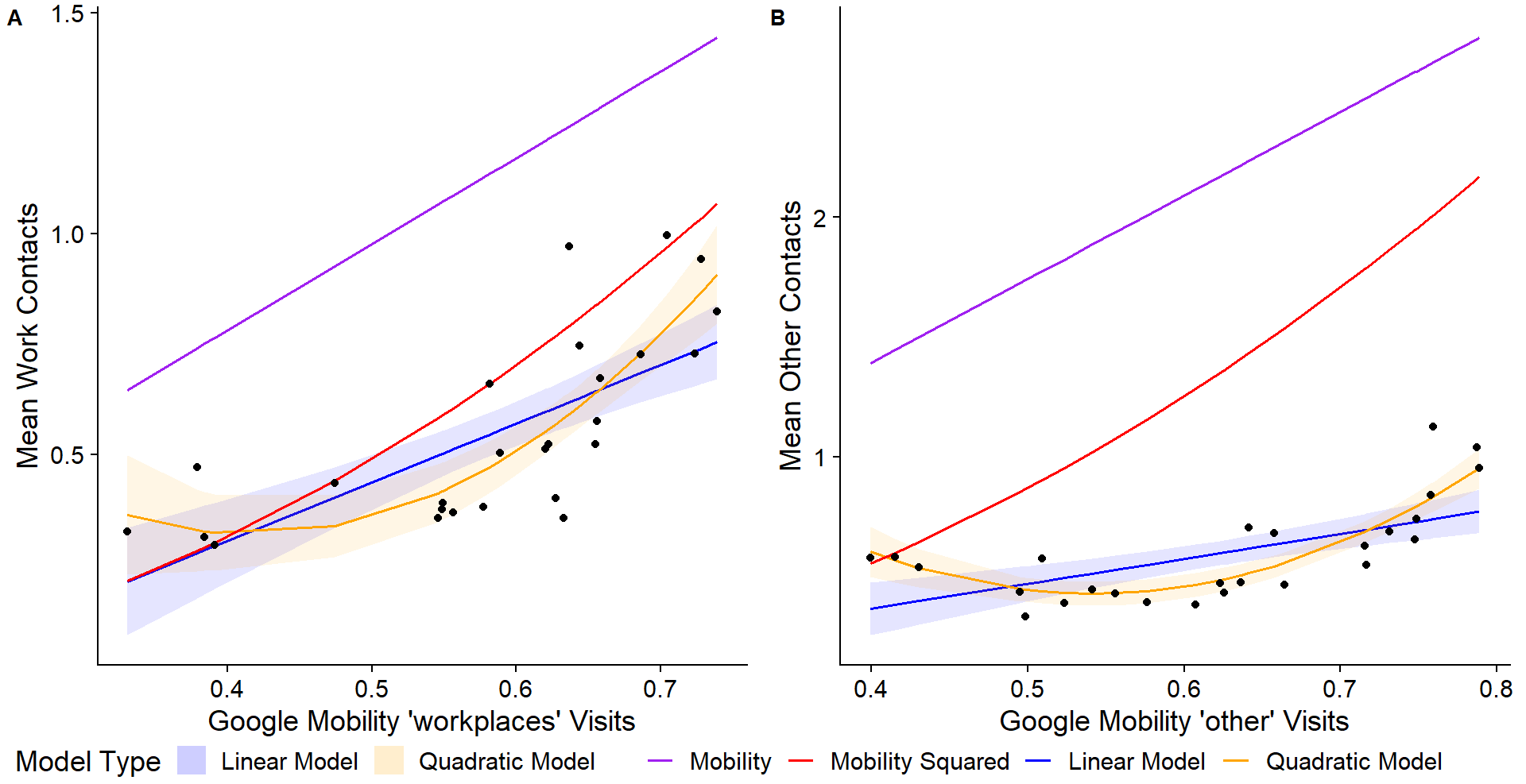
We can see visually that the correlation between mobility and contacts changes over time, with a notable difference between “work” contacts and “workplace” mobility in the second year of data available. The choice to perform subsequent analyses on the first year of data only was epidemiologically informed, more information is available in the supplementary material.

### 3.2.2 Comparing Methods of Scaling Contacts

Here we compare four possible ways of estimating contact rates based on mobility data to determine the most appropriate method to use when contact data is not available.

#### 3.2.2.1 Regression models

For both "work" and "other" contacts, the models which had the most evidence were the quadratic models ("work": p-value = 0.0009, "other": p-value << 0.0001). When we compare AIC values, it is shown that both for “work” contacts (linear regression: -32.43, quadratic regression: -42.64) and “other” contacts (linear regression: -29.94, quadratic regression: -57.81) the quadratic regression model best fits the data when compared to the linear regression model. However, this was evidence of the best model for UK data and hence it was important to compare both linear and quadratic relative contact rates for other countries.

******

**Figure 2:** predicted values for the regression models created using contact data and Google mobility data from the UK. Plot A shows the linear and quadratic regression models for "work" contacts along with “workplace” mobility and mobility squared, plot B shows the linear and quadratic regression models for "other" contacts along with “other” mobility and mobility squared. The shade areas show the respective confidence intervals for the regression model predictions and the black points show the raw data used in the regression models. The naïve curves are calculated by multiplying the mobility for a specific date by the POLYMOD estimate for that contact type (“work” for “workplace” mobility and “other” for “other” mobility).

Figure 2 shows that when using the naïve models (mobility, and mobility squared) the estimated number of mean contacts is larger than the number predicted by the regression models. We see that mobility predicts the highest numbers of contacts and mobility squared appears to closely follow the regression models for "work" contacts but not for "other" contacts.

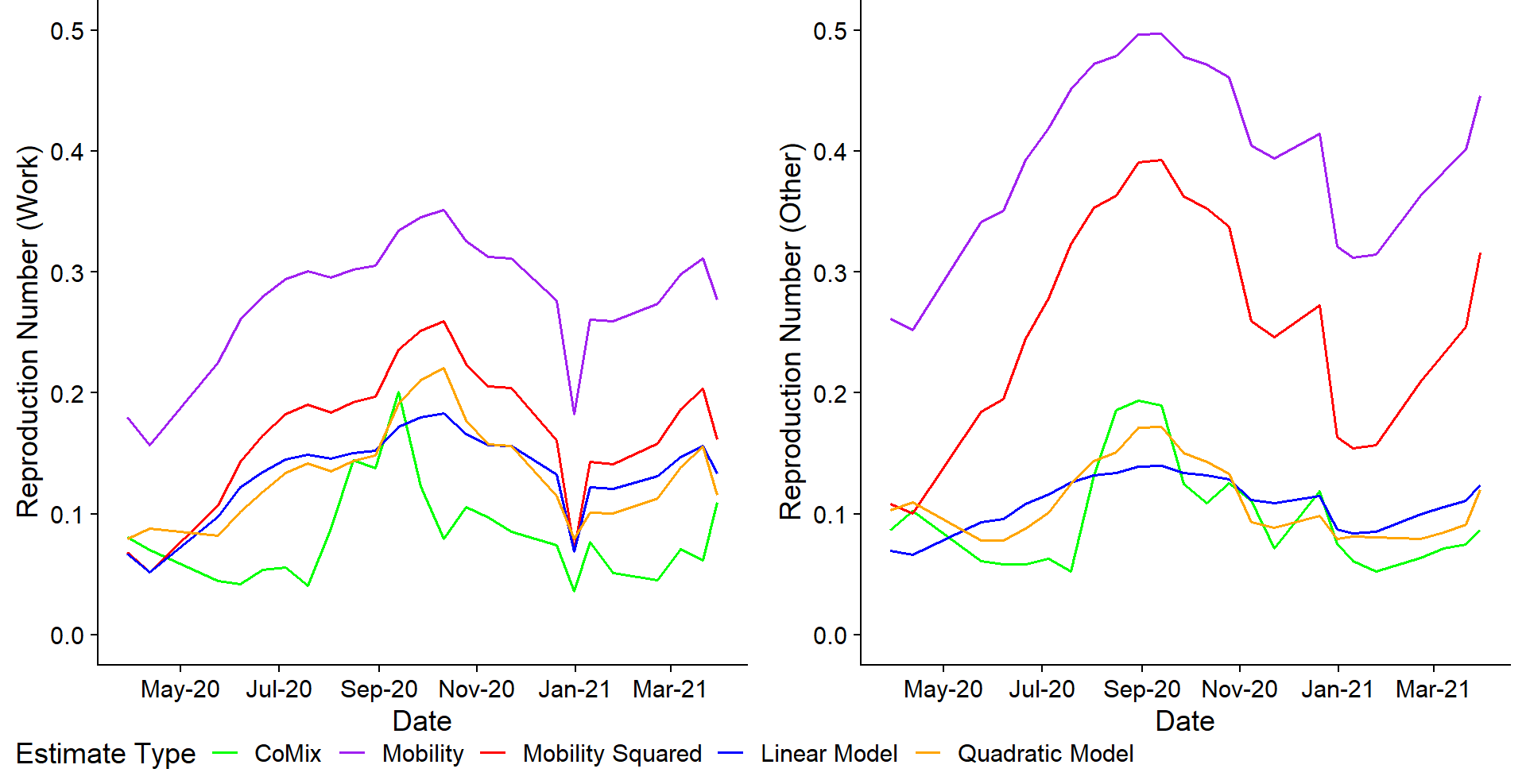
#### 3.2.2.2 Creating the relative contact rate

The parameter estimates from the regression models described above were used with the corresponding mobility data ("workplace" indicator for the "work" models and "other" indicator for the "other" models) to predict a number of contacts for a given mobility. These were then scaled by the relevant POLYMOD number of contacts ("work"/"other"). For "work" contacts this was 1.9529 contacts, and for "other" contacts this was 3.4837 contacts. Parameter estimates are given in the supplementary material (tables S1-S4).

#### 3.2.2.3 Comparing the relative contact rates to Google mobility

After scaling the synthetic matrices [(21)](https://www.zotero.org/google-docs/?WYj699), we calculated the dominant eigenvalues of these scaled matrices. We then transformed the dominant eigenvalues into partial reproduction numbers using a multiplicative factor of 0.1582 calculated using the formula given in the methods section.

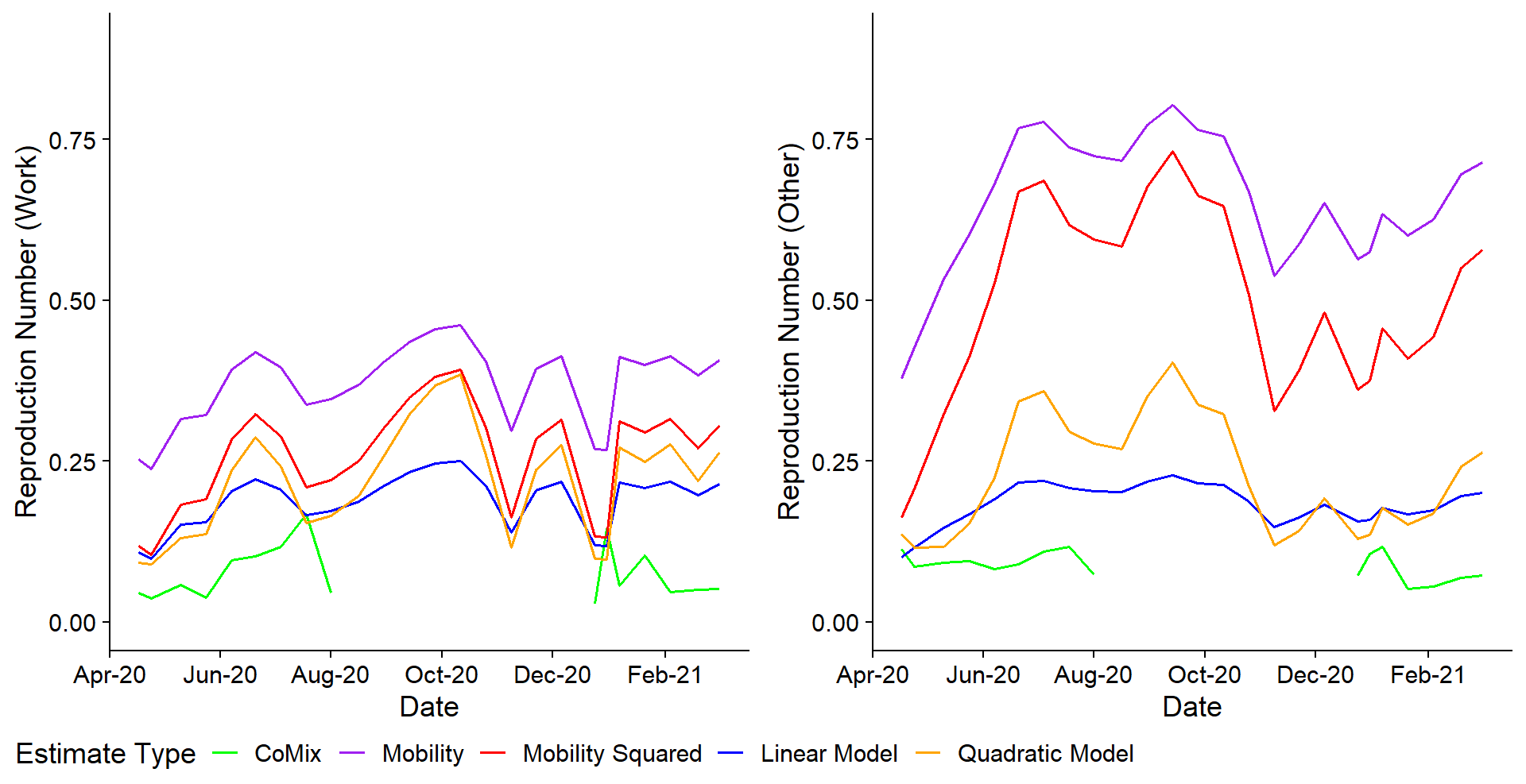
#### 3.2.2.4 Reproduction numbers: UK

**

**Figure 3:** partial reproduction numbers over time for each estimate type, green line indicates the reproduction number calculated from the CoMix data. The purple line indicates the mobility scaled estimates, the red line indicates the mobility squared scaled estimates, the blue line indicates the linear model scaled estimates and the orange line indicates the quadratic model scaled estimates.

Figure 3 shows that the model estimates appear to more accurately approximate the estimates from the CoMix data. Comparing root MSE can determine the best approximation quantitatively. The approximation with the smallest root MSE was for the quadratic model relative contact rate ("work": 0.0626, "other": 0.0264) and the approximation with the largest root MSE was for Google mobility ("work": 0.2031, "other": 0.3070). (See tables S5 and S6 for the root MSEs for all approximations).

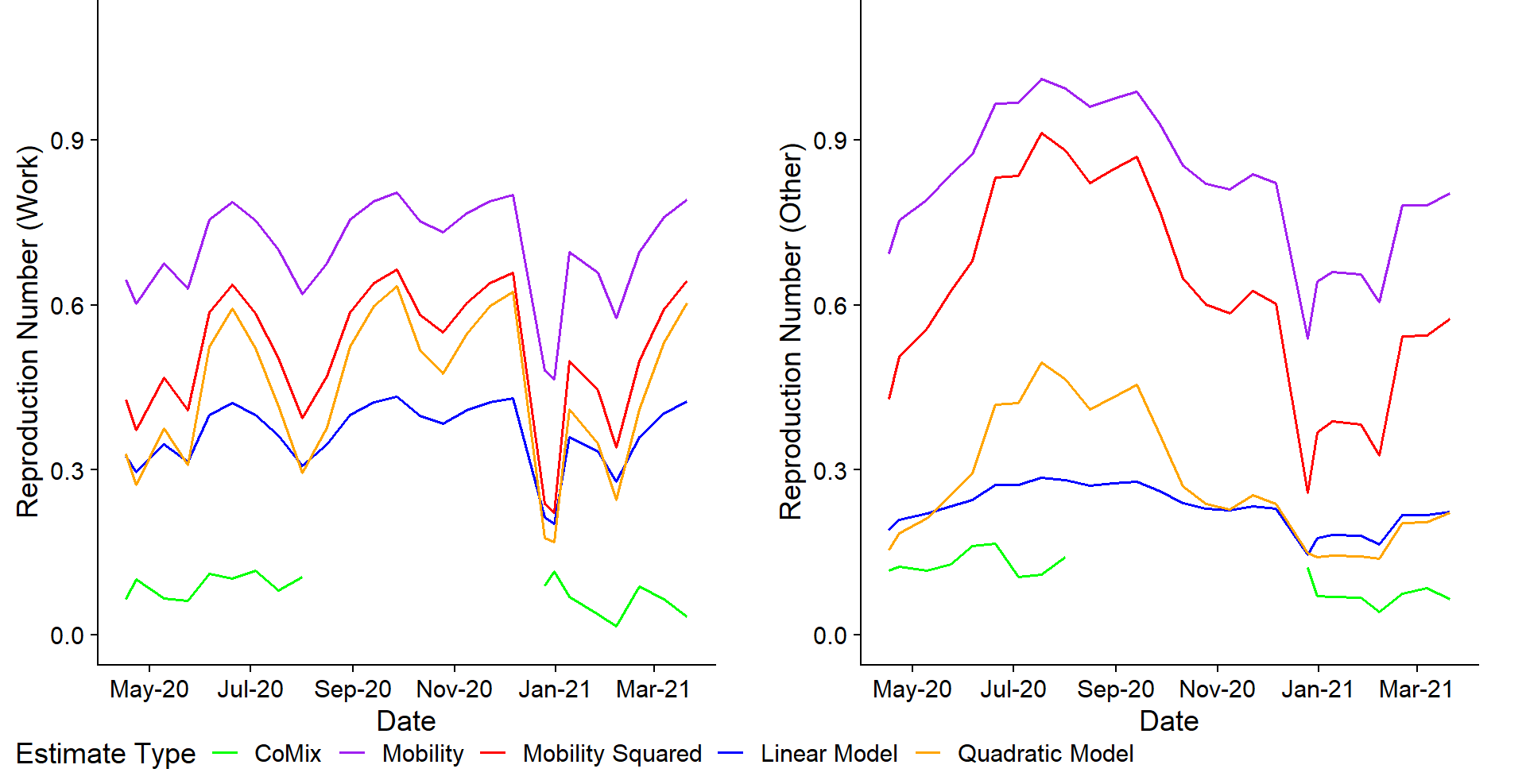
#### 3.2.2.5 Reproduction numbers: BE

**

**Figure 4:**  partial reproduction numbers over time for each estimate type, green line indicates the reproduction number calculated from the CoMix data. The purple line indicates the mobility scaled estimates, the red line indicates the mobility squared scaled estimates, the blue line indicates the linear model scaled estimates and the orange line indicates the quadratic model scaled estimates.

Figure 4 shows that where a comparison can be made, the model scaled estimates appear closer than the mobility scaled estimates. When looking at the root MSE (see tables S7 and S8) the best approximation is the linear model relative contact rate ("work": 0.1130, "other": 0.0962) and the worst is the Google mobility ("work": 0.2825, "other": 0.5520).

#### 3.2.2.6 Reproduction numbers: NL

**

**Figure 5:** partial reproduction numbers over time for each estimate type, green line indicates the reproduction number calculated from the CoMix data. The purple line indicates the mobility scaled estimates, the red line indicates the mobility squared scaled estimates, the blue line indicates the linear model scaled estimates and the orange line indicates the quadratic model scaled estimates.

Figure 5 shows that as with Belgium (figure 4), in the Netherlands where a comparison can be made the estimates using the model relative contact rates are closer to the estimates from the CoMix data. The approximation with the smallest root MSE was for the linear model relative contact rate ("work": 0.2703, "other": 0.1205) and, as with the UK and Belgium, the approximation with the largest root MSE was for the Google mobility ("work": 0.5950, "other": 0.6927).

# 4. Discussion

During the COVID-19 pandemic, mobility data has been widely used to get insights into the spread of SARS-CoV-2 [(25)](https://www.zotero.org/google-docs/?ajb4qq). Two main approaches have been used: some studies have tried to establish a relation between mobility and transmission [(26)](https://www.zotero.org/google-docs/?zNYUSJ), while others have been trying to relate mobility to contact rates [(27,28)](https://www.zotero.org/google-docs/?PStXxd). While the former approach is suitable for identifying the mobility data that better provide insights into epidemic spread (e.g. mobile phone data), the latter approach has the advantage of an easier implementation within mathematical models of infectious diseases [(29,30)](https://www.zotero.org/google-docs/?R7vuKo) and is the one we focus our attention on in this work. When comparing mobility and contact rates, we see that there was an apparent association between both the respective contacts and mobility indicators (see figure 1), although this was stronger for the first year of data available. For both "work" and "other" contacts, the mean number of contacts is lower in the second year of the study relative to the corresponding mobility. This may indicate that there was a change in participant behaviour, a change in the relationship between mobility and contacts or a change in participant recruitment. This was the motivation for using the first year of data for this investigation. Indeed, the relationship between contacts and mobility is expected to be time-varying and although mobility has been found to be very predictive of social contacts during lockdown in China, this was not the case in the post-lockdown scenario [(31)](https://www.zotero.org/google-docs/?GxxbZX). Also, several factors such as risk perception, do affect both mobility and contacts to a different degree [(32,33)](https://www.zotero.org/google-docs/?3xDHw6), therefore potentially affecting their mutual relationship. Starting from this observation, we explored different regression models that could quantify the link between mobility and contacts using UK data. We then used the best-fit model to generate synthetic contact data from mobility alone and compared it with the results coming from collected contact data. We then extended this approach to the Netherlands and Belgium, to see how our regression model could be generalizable to other countries. While the quadratic models were best when looking at the models themselves, there was no guarantee this would be the case for non-UK data. We then used both model types to create relative contact rates in order to determine which best predicted the behaviour across different countries. Comparing models using information criterion (AIC) shows that the quadratic model was the best fit for both “work” and “other” contacts.

When looking at results from the UK, we see that the most accurate approximations to the partial reproduction numbers from the CoMix survey are those produced by the quadratic regression model as seen in figure 3 (see supplementary material for further figures). The root MSE is also smaller for the linear model. However, for both Belgium and the Netherlands we see the most accurate approximations are those produced by the linear regression model, as shown in figures 4 and 5 (see supplementary material for further figures). The tendency for the linear regression model to provide the most accurate approximations out of sample may be due to the fact that there are periods where no CoMix data is available for Belgium and the Netherlands, however, this is difficult to confirm. For all countries and contact types, using just Google mobility produced the worst approximations, with the mobility squared producing the second worst approximations. This is reflected in figures 3-5, with the mobility approximation being clearly less accurate than any other method as well as having the largest root mean squared error. In addition, for "work" contacts in Belgium and the Netherlands the approximations for any method were not as close to those for the UK. Approximations for the UK are generally better as the models used to form the relative contact rates were created using UK data.

As CoMix data was not available for some of the study period outside of the UK it is difficult to determine the accuracy of the predictions made outside of the UK. This also means that residuals (given in the supplementary material) and the root mean squared error do not account for predictions in this period as there is no ‘true’ value to compare to. For the times where data was available for Belgium and the Netherlands the countries were for the most part under some restrictions, this is why the CoMix estimates were often consistently low. Therefore, it is difficult to determine whether the accuracy of the approximations would be better or worse when no restrictions were in place. For the UK, when restrictions were not in place the accuracy of the approximations remained consistent with their accuracy during restriction periods. Assuming this is true outside of the UK, then the recommendation would be to use the linear regression model relative contact rate instead of Google mobility, to scale pre-pandemic contacts. This method could also be used to calculate contact matrices outside of the data collection period.

This investigation is limited by the fact that Google mobility data is difficult to define, not much information is available on how the indicators are defined. In addition, CoMix surveys were not available throughout the entire study period for countries other than the UK, this limited the amount we were able to determine the accuracy of the results outside of the UK. A final limitation is that the relationship between contacts and Google mobility appears to change in the second year of data (March 2021 onwards) particularly in work contacts. It is difficult to determine whether this change was due to mobility data becoming less accurate over time, contact data being less accurate towards the later surveys, or a genuine change in the relationship between mobility in contacts. It was for this reason we decided to focus on the first year of data (March 2020-March 2021).

## 4.2 Conclusions

If given the choice between mobility and mobility squared to scale pre-pandemic contacts, the best approximation comes from mobility squared. However, the relative contact rates produced using the parameter estimates given above do provide better approximations than from using mobility data on its own. The results for Belgium and the Netherlands suggest that the linear regression model provides best approximations for partial reproduction number estimates outside of the UK. This investigation is the first to compare mobility indicators and contact matrices and shows that mobility data can be improved upon as a method to measure transmission potential.

# References

[1. Mathieu E, Ritchie H, Rodés-Guirao L, Appel C, Giattino C, Hasell J, et al. Coronavirus Pandemic (COVID-19). Our World Data [Internet]. 2020 Mar 5 [cited 2023 Jan 13]; Available from: https://ourworldindata.org/covid-deaths](https://www.zotero.org/google-docs/?oSLTLy)

[2. CMMID Repository [Internet]. 2020 [cited 2023 Jan 13]. CoMix study - Social contact survey in the UK. Available from: https://cmmid.github.io/topics/covid19/comix-reports.html](https://www.zotero.org/google-docs/?oSLTLy)

[3. Jarvis CI, Van Zandvoort K, Gimma A, Prem K, Klepac P, Rubin GJ, et al. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. BMC Med. 2020 May 7;18:124.](https://www.zotero.org/google-docs/?oSLTLy)

[4. Coletti P, Wambua J, Gimma A, Willem L, Vercruysse S, Vanhoutte B, et al. CoMix: comparing mixing patterns in the Belgian population during and after lockdown. Sci Rep. 2020 Dec 14;10(1):21885.](https://www.zotero.org/google-docs/?oSLTLy)

[5. Backer JA, Bogaardt L, Beutels P, Coletti P, Edmunds WJ, Gimma A, et al. Dynamics of non-household contacts during the COVID-19 pandemic in 2020 and 2021 in the Netherlands. Sci Rep. 2023 Mar 30;13(1):5166.](https://www.zotero.org/google-docs/?oSLTLy)

[6. Verelst F, Hermans L, Vercruysse S, Gimma A, Coletti P, Backer JA, et al. 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. 2021 Sep 29;19(1):254.](https://www.zotero.org/google-docs/?oSLTLy)

[7. Weidemann F, Remschmidt C, Buda S, Buchholz U, Ultsch B, Wichmann O. Is the impact of childhood influenza vaccination less than expected: a transmission modelling study. BMC Infect Dis. 2017 Apr 11;17(1):258.](https://www.zotero.org/google-docs/?oSLTLy)

[8. Meyer S, Held L. Incorporating social contact data in spatio-temporal models for infectious disease spread. Biostat Oxf Engl. 2017 Apr;18(2):338–51.](https://www.zotero.org/google-docs/?oSLTLy)

[9. Tizzoni M, Bajardi P, Decuyper A, King GKK, Schneider CM, Blondel V, et al. On the Use of Human Mobility Proxies for Modeling Epidemics. PLOS Comput Biol. 2014 Jul 10;10(7):e1003716.](https://www.zotero.org/google-docs/?oSLTLy)

[10. COVID-19 Community Mobility Reports [Internet]. [cited 2023 Jan 13]. Available from: https://www.google.com/covid19/mobility/](https://www.zotero.org/google-docs/?oSLTLy)

[11. Nouvellet P, Bhatia S, Cori A, Ainslie KEC, Baguelin M, Bhatt S, et al. Reduction in mobility and COVID-19 transmission. Nat Commun. 2021 Feb 17;12(1):1090.](https://www.zotero.org/google-docs/?oSLTLy)

[12. Real-time nowcasting and forecasting of COVID-19 dynamics in England: the first wave | Philosophical Transactions of the Royal Society B: Biological Sciences [Internet]. [cited 2023 May 31]. Available from: https://royalsocietypublishing.org/doi/full/10.1098/rstb.2020.0279#d1e3299](https://www.zotero.org/google-docs/?oSLTLy)

[13. Gozzi N, Chinazzi M, Dean NE, Longini IM, Halloran ME, Perra N, et al. Estimating the impact of COVID-19 vaccine allocation inequities: a modeling study [Internet]. medRxiv; 2022 [cited 2023 May 31]. p. 2022.11.18.22282514. Available from: https://www.medrxiv.org/content/10.1101/2022.11.18.22282514v1](https://www.zotero.org/google-docs/?oSLTLy)

[14. Gozzi N, Tizzoni M, Chinazzi M, Ferres L, Vespignani A, Perra N. Estimating the effect of social inequalities on the mitigation of COVID-19 across communities in Santiago de Chile. Nat Commun. 2021 Apr 23;12(1):2429.](https://www.zotero.org/google-docs/?oSLTLy)

[15. Overview - Community Mobility Reports Help [Internet]. [cited 2023 Jan 13]. Available from: https://support.google.com/covid19-mobility/answer/9824897?hl=en&ref\_topic=9822927](https://www.zotero.org/google-docs/?oSLTLy)

[16. Gimma A, Munday JD, Wong KLM, Coletti P, Zandvoort K van, Prem K, et al. Changes in social contacts in England during the COVID-19 pandemic between March 2020 and March 2021 as measured by the CoMix survey: A repeated cross-sectional study. PLOS Med. 2022 Mar 1;19(3):e1003907.](https://www.zotero.org/google-docs/?oSLTLy)

[17. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med. 2008 Mar 25;5(3):e74.](https://www.zotero.org/google-docs/?oSLTLy)

[18. R: The R Project for Statistical Computing [Internet]. [cited 2023 Sep 21]. Available from: https://www.r-project.org/](https://www.zotero.org/google-docs/?oSLTLy)

[19. emiel200. comix\_mobility [Internet]. 2022 [cited 2023 Aug 22]. Available from: https://github.com/emiel200/comix\_mobility](https://www.zotero.org/google-docs/?oSLTLy)

[20. Davies NG, Barnard RC, Jarvis CI, Russell TW, Semple MG, Jit M, et al. Association of tiered restrictions and a second lockdown with COVID-19 deaths and hospital admissions in England: a modelling study. Lancet Infect Dis. 2021 Apr;21(4):482–92.](https://www.zotero.org/google-docs/?oSLTLy)

[21. Prem K, Zandvoort K van, Klepac P, Eggo RM, Davies NG, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, et al. Projecting contact matrices in 177 geographical regions: An update and comparison with empirical data for the COVID-19 era. PLoS Comput Biol. 2021 Jul;17(7):e1009098.](https://www.zotero.org/google-docs/?oSLTLy)

[22. epiforecasts/covid-rt-estimates: National and subnational estimates of the time-varying reproduction number for Covid-19 [Internet]. [cited 2023 Jul 17]. Available from: https://github.com/epiforecasts/covid-rt-estimates/tree/master](https://www.zotero.org/google-docs/?oSLTLy)

[23. Census - Office for National Statistics [Internet]. [cited 2023 Aug 22]. Available from: https://www.ons.gov.uk/census](https://www.zotero.org/google-docs/?oSLTLy)

[24. Ipsos [Internet]. 2009 [cited 2023 Aug 22]. Social Grade. Available from: https://www.ipsos.com/en-uk/social-grade](https://www.zotero.org/google-docs/?oSLTLy)

[25. Alessandretti L. What human mobility data tell us about COVID-19 spread. Nat Rev Phys. 2022 Jan;4(1):12–3.](https://www.zotero.org/google-docs/?oSLTLy)

[26. Rüdiger S, Konigorski S, Rakowski A, Edelman JA, Zernick D, Thieme A, et al. Predicting the SARS-CoV-2 effective reproduction number using bulk contact data from mobile phones. Proc Natl Acad Sci. 2021 Aug 3;118(31):e2026731118.](https://www.zotero.org/google-docs/?oSLTLy)

[27. Tomori DV, Rübsamen N, Berger T, Scholz S, Walde J, Wittenberg I, et al. Individual social contact data and population mobility data as early markers of SARS-CoV-2 transmission dynamics during the first wave in Germany—an analysis based on the COVIMOD study. BMC Med. 2021 Oct 14;19(1):271.](https://www.zotero.org/google-docs/?oSLTLy)

[28. Lajot A, Wambua J, Coletti P, Franco N, Brondeel R, Faes C, et al. How contact patterns during the COVID-19 pandemic are related to pre-pandemic contact patterns and mobility trends. BMC Infect Dis. 2023 Jun 16;23(1):410.](https://www.zotero.org/google-docs/?oSLTLy)

[29. Hoang T, Coletti P, Melegaro A, Wallinga J, Grijalva CG, Edmunds JW, et al. A Systematic Review of Social Contact Surveys to Inform Transmission Models of Close-contact Infections. Epidemiol Camb Mass. 2019 Sep;30(5):723–36.](https://www.zotero.org/google-docs/?oSLTLy)

[30. Keeling MJ, Rohani P. Modeling Infectious Diseases in Humans and Animals [Internet]. Princeton University Press; 2008 [cited 2023 Sep 22]. Available from: https://www.jstor.org/stable/j.ctvcm4gk0](https://www.zotero.org/google-docs/?oSLTLy)

[31. Zhang J, Litvinova M, Liang Y, Zheng W, Shi H, Vespignani A, et al. The impact of relaxing interventions on human contact patterns and SARS-CoV-2 transmission in China. Sci Adv. 2021 May 7;7(19):eabe2584.](https://www.zotero.org/google-docs/?oSLTLy)

[32. Wu S, Grant-Muller S, Yang L. Modelling the Mobility Changes Caused by Perceived Risk and Policy Efficiency. ISPRS Int J Geo-Inf. 2022 Aug;11(8):453.](https://www.zotero.org/google-docs/?oSLTLy)

[33. Wambua J, Loedy N, Jarvis CI, Wong KLM, Faes C, Grah R, et al. The influence of COVID-19 risk perception and vaccination status on the number of social contacts across Europe: insights from the CoMix study. BMC Public Health. 2023 Jul 13;23(1):1350.](https://www.zotero.org/google-docs/?oSLTLy)

[34. Danon L, Read JM, Keeling MJ, House TA, Vernon MC. Social contact study [Internet]. University of Warwick, Department of Biological Sciences; 2009 [cited 2023 Mar 16]. Available from: http://wrap.warwick.ac.uk/54273](https://www.zotero.org/google-docs/?oSLTLy)

Supplementary Material

[**S.1 Regression Models 17**](#_f0cymti2jzqj)

[**S.2 Baseline Contact Levels 18**](#_wuynedladsqb)

[**S.3 Relative Contact Rates 18**](#_fk8p0kv2rwx7)

[**S.4 Supplementary Figures 18**](#_zay2rqbgyw2v)

[S.4.1 Age proportions over time 18](#_g17h68nwuzt0)

[S.4.2 Social class proportions over time 19](#_jdc3opjzy4i3)

[S.4.3 Employment status (employment age) 20](#_n8inpbfpkqn)

[S.4.4 Reproduction numbers: UK "work" contacts 20](#_c5hunu3wj7gd)

[S.4.5 Reproduction numbers: UK "other" contacts 22](#_909jvjgrh99l)

[S.4.6 Reproduction numbers: BE "work" contacts 22](#_af4e5l4yfnpk)

[S.4.7 Reproduction numbers: BE "other" contacts 24](#_l02b6smzhb7b)

[S.4.8 Reproduction numbers: NL "work" contacts 24](#_34g9chsl9nlq)

[S.4.9 Reproduction numbers: NL "other" contacts 26](#_9fmo5fioham5)

[**S.5 Supplementary Tables 26**](#_3g212116cowi)

[S.5.1 Comparing contacts and mobility 27](#_o3nmqm4ih3a1)

[S.5.2 Regression models: "work" contacts 27](#_hhyaxfpky9s0)

[S.5.3 Regression models: "other" contacts 27](#_dc0qnzp6jog4)

[S.5.4 Reproduction numbers: UK 28](#_4rlqug2tyfl8)

[S.5.5 Reproduction numbers: BE 28](#_d5zbaptot2bj)

[S.5.6 Reproduction numbers: NL 29](#_9zdxtaqnqug6)

# S.1 Regression Models

*"Work" contacts against "workplace" indicator*

The linear "work" model is:

And the quadratic "work" model is:

*"Other" contacts against weighted predictor of indicators*

The linear "other" model is:

And the quadratic “other” model is:

# S.2 Baseline Contact Levels

As the CoMix survey was designed from the POLYMOD survey, conducted in 2005/2006 [(17)](https://www.zotero.org/google-docs/?q5BAbv) this was an obvious choice for the baseline contact information. We compared this baseline to the Warwick contact study which was conducted in 2009 [(34)](https://www.zotero.org/google-docs/?q36FAU), in order to determine the suitability of the POLYMOD baseline. We found that both provided similar estimates of baseline contacts.

# S.3 Relative Contact Rates

For "work" contacts the relative contact rate for the linear model was calculated as follows:

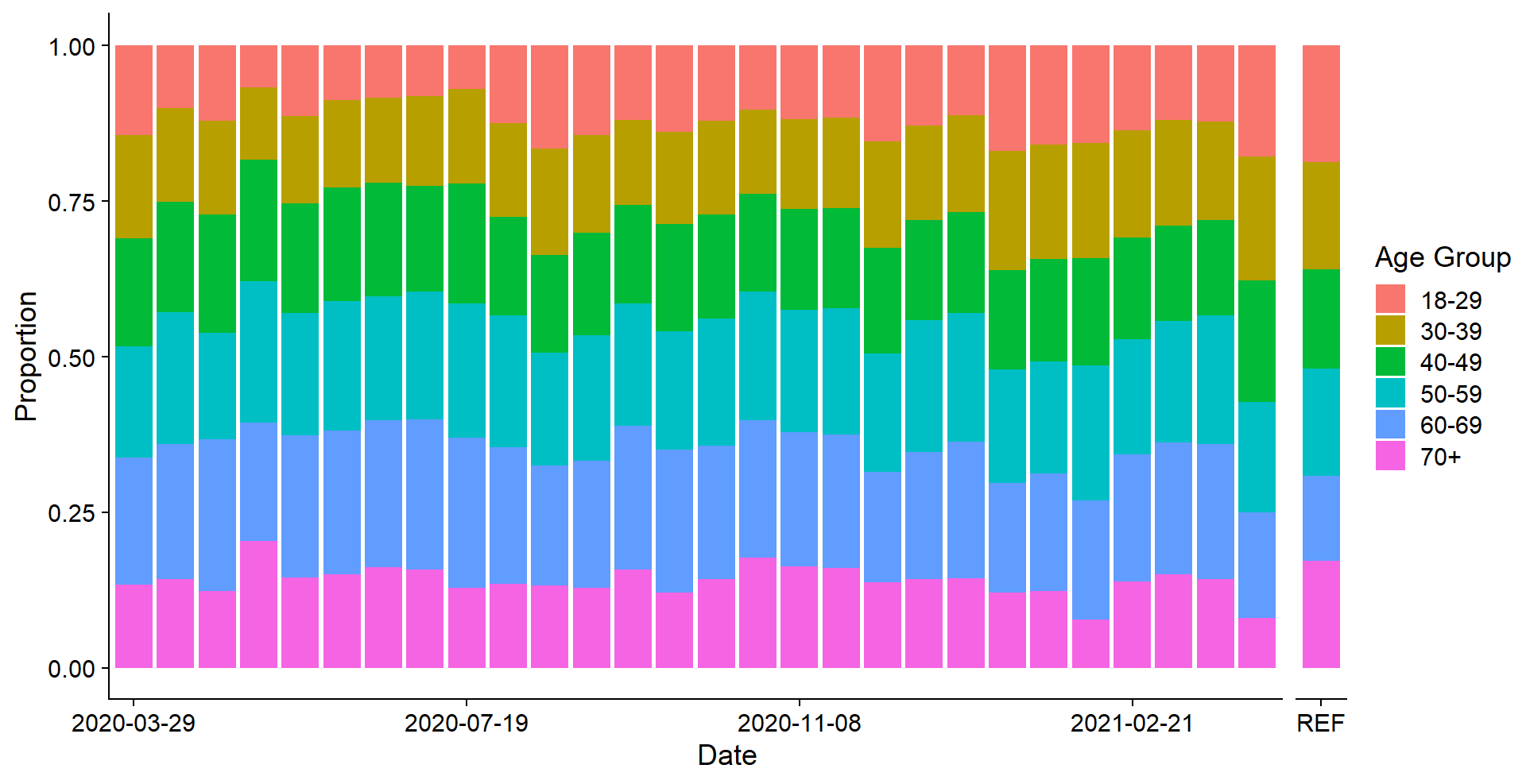
And for the quadratic model:

For "other" contacts the relative contact rate for the linear model was calculated as follows:

And for the quadratic model:

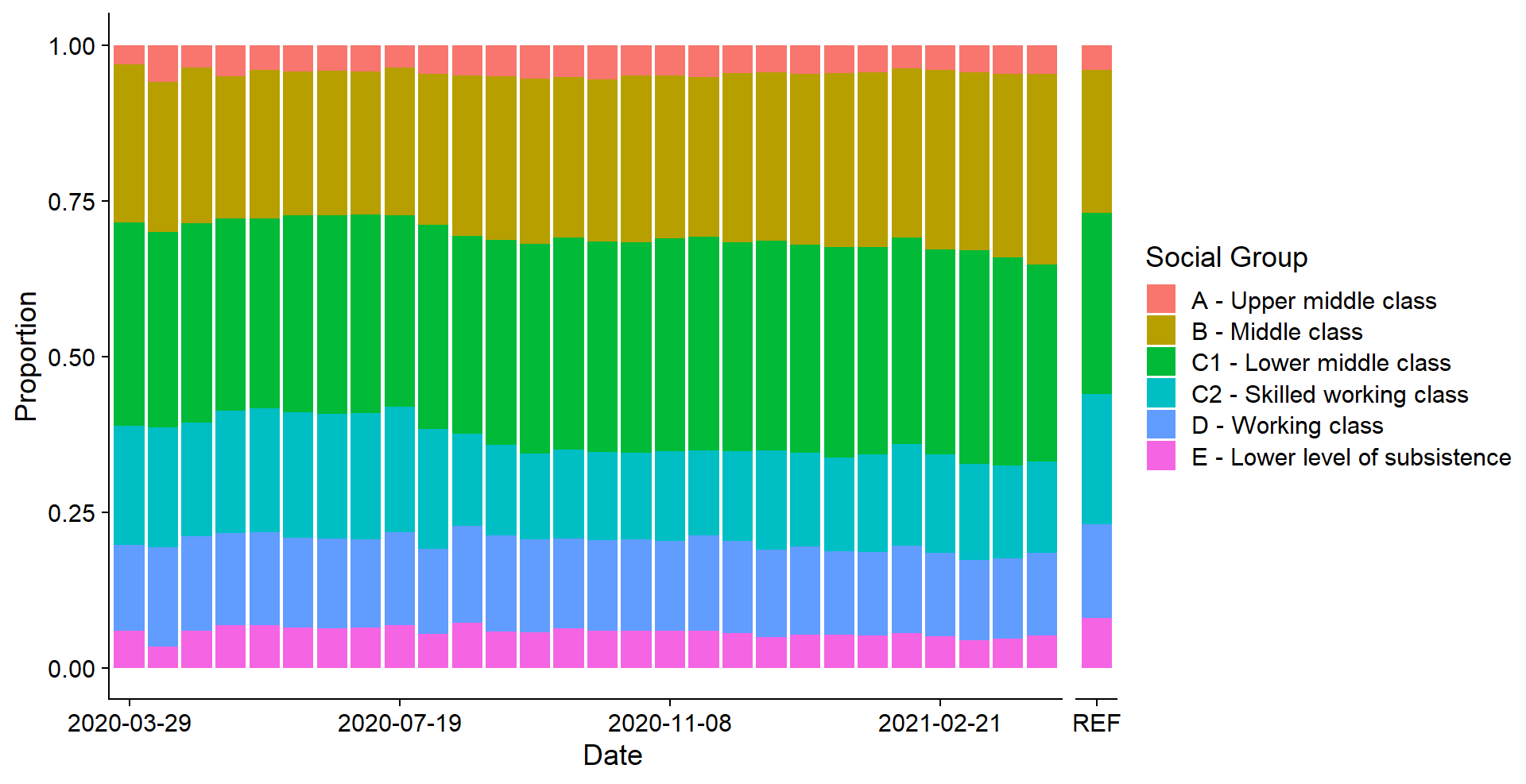
# S.4 Supplementary Figures

## S.4.1 Age proportions over time

**

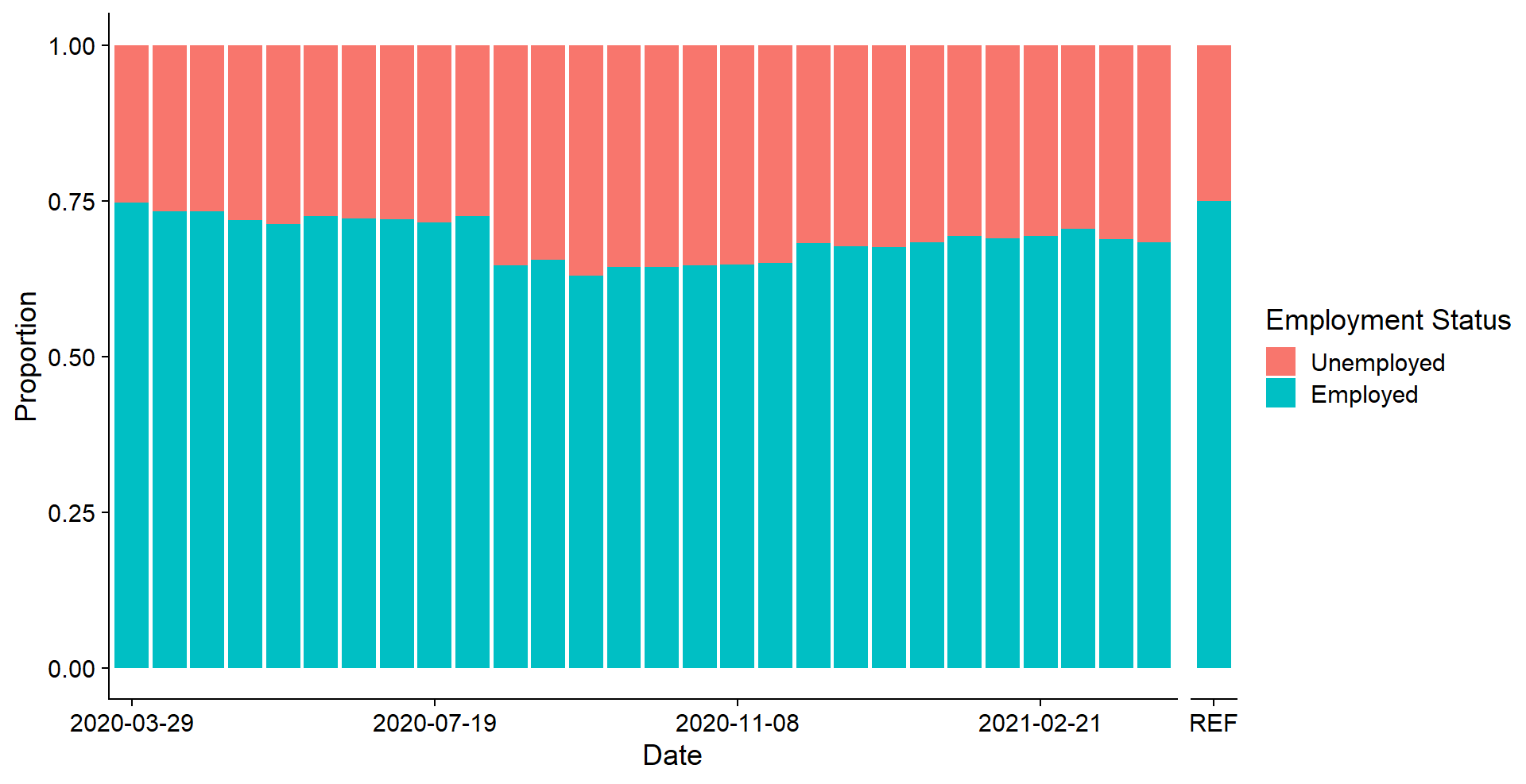
**Figure S1:** proportion of participants per age group by two week periods, REF shows the population proportions

S.4.2 Social class proportions over time

**

**Figure S2:** proportion of participants per social group by two week periods, REF shows the population proportions

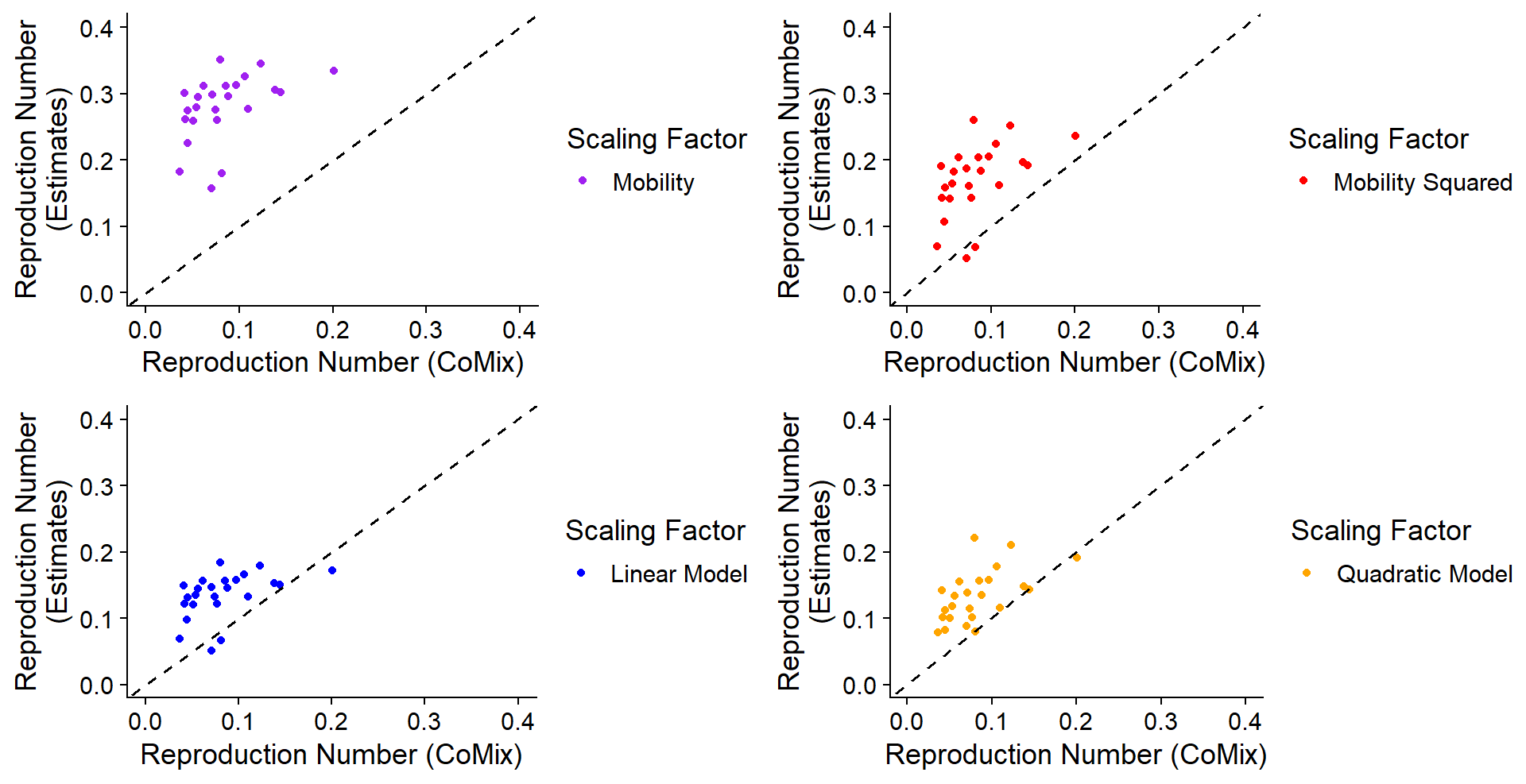
## S.4.3 Employment status (employment age)

**

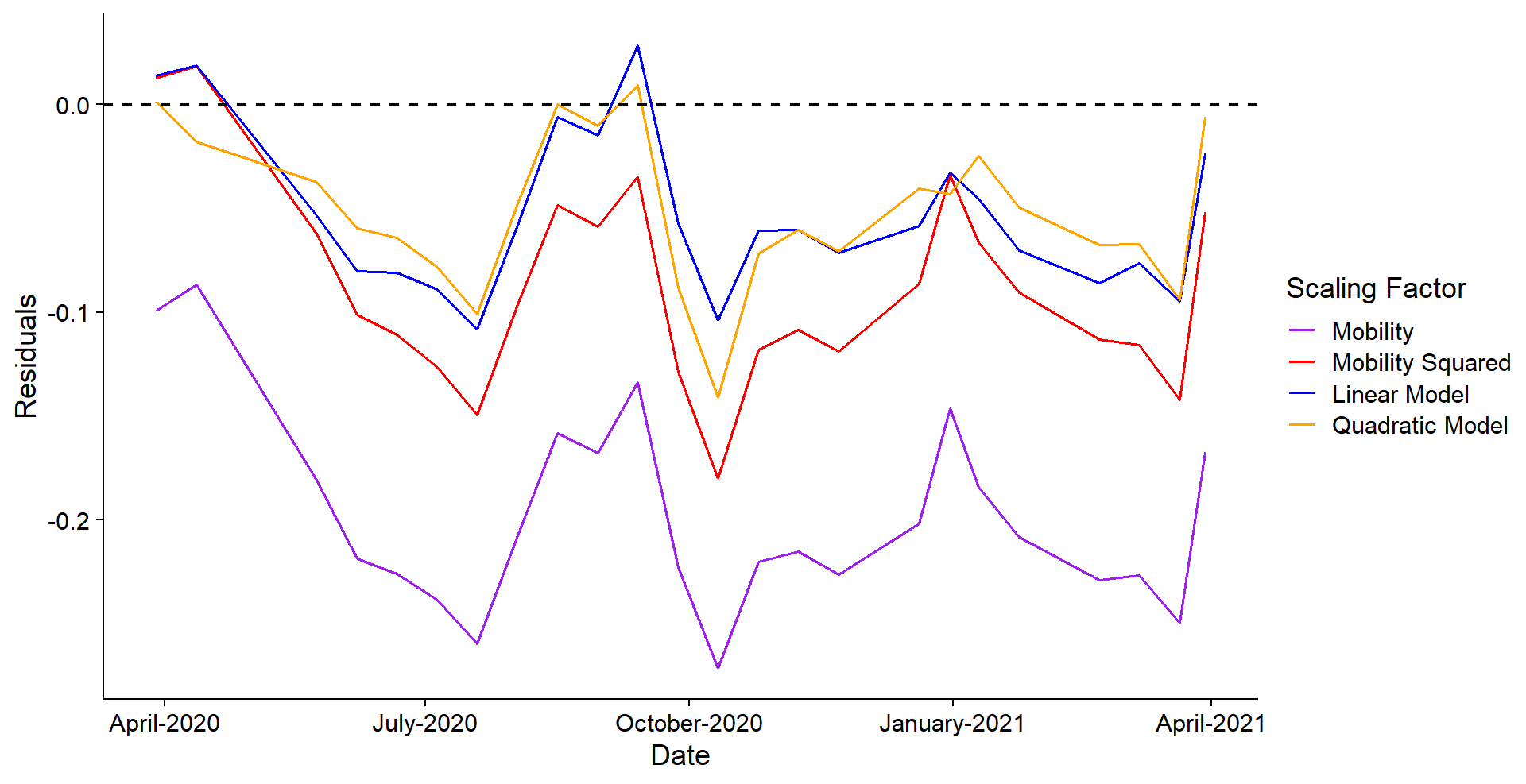
**Figure S3:** proportion of participants per employment status by two week periods, REF shows the population proportions

## 

## S.4.4 Reproduction numbers: UK "work" contacts

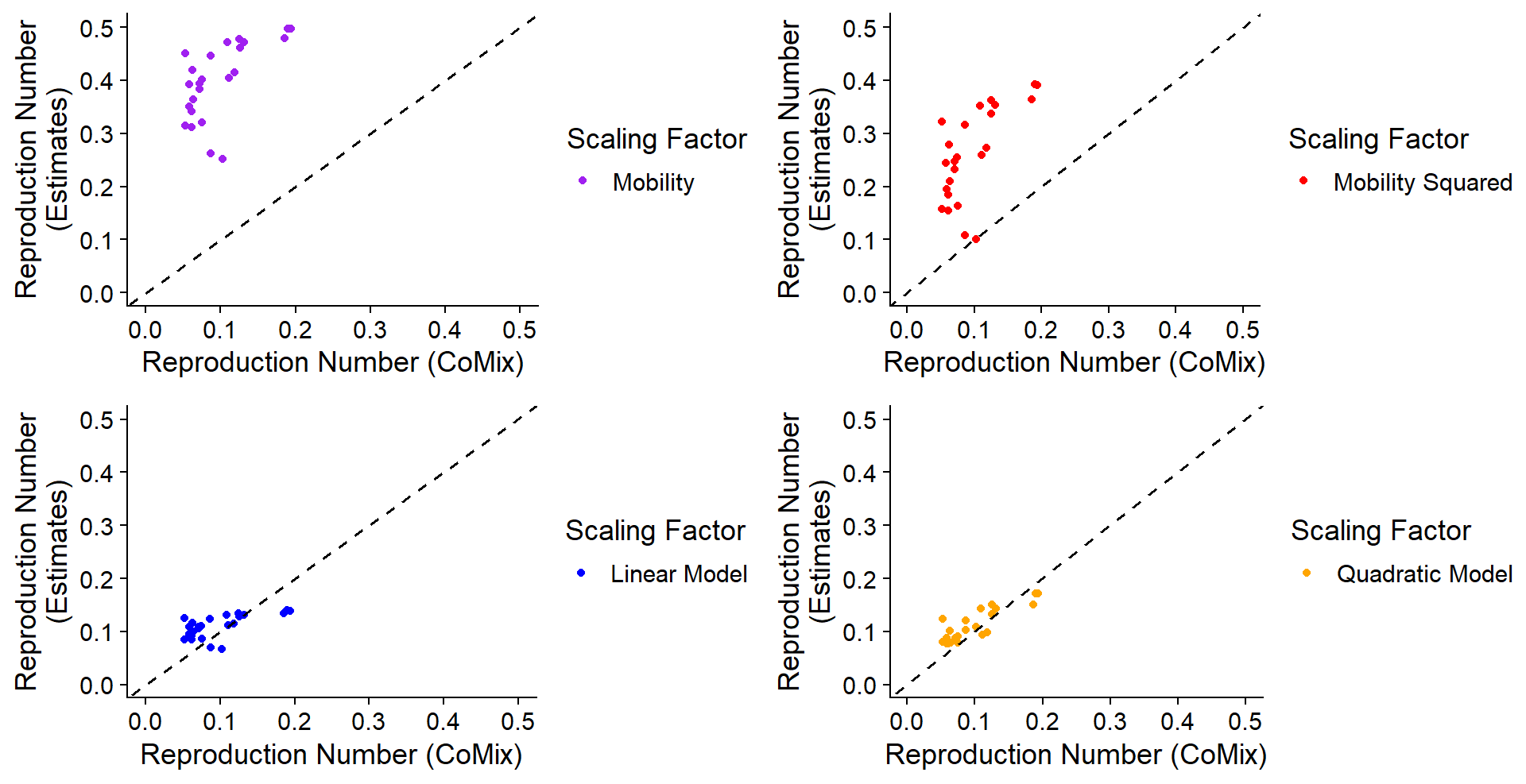


**Figure S4A:** reproduction number estimates against reproduction numbers from CoMix survey, top left shows estimates using mobility, top right shows estimates using mobility squared, bottom left shows estimates using the linear model relative contact rate and bottom right shows estimates using the quadratic model relative contact rate.

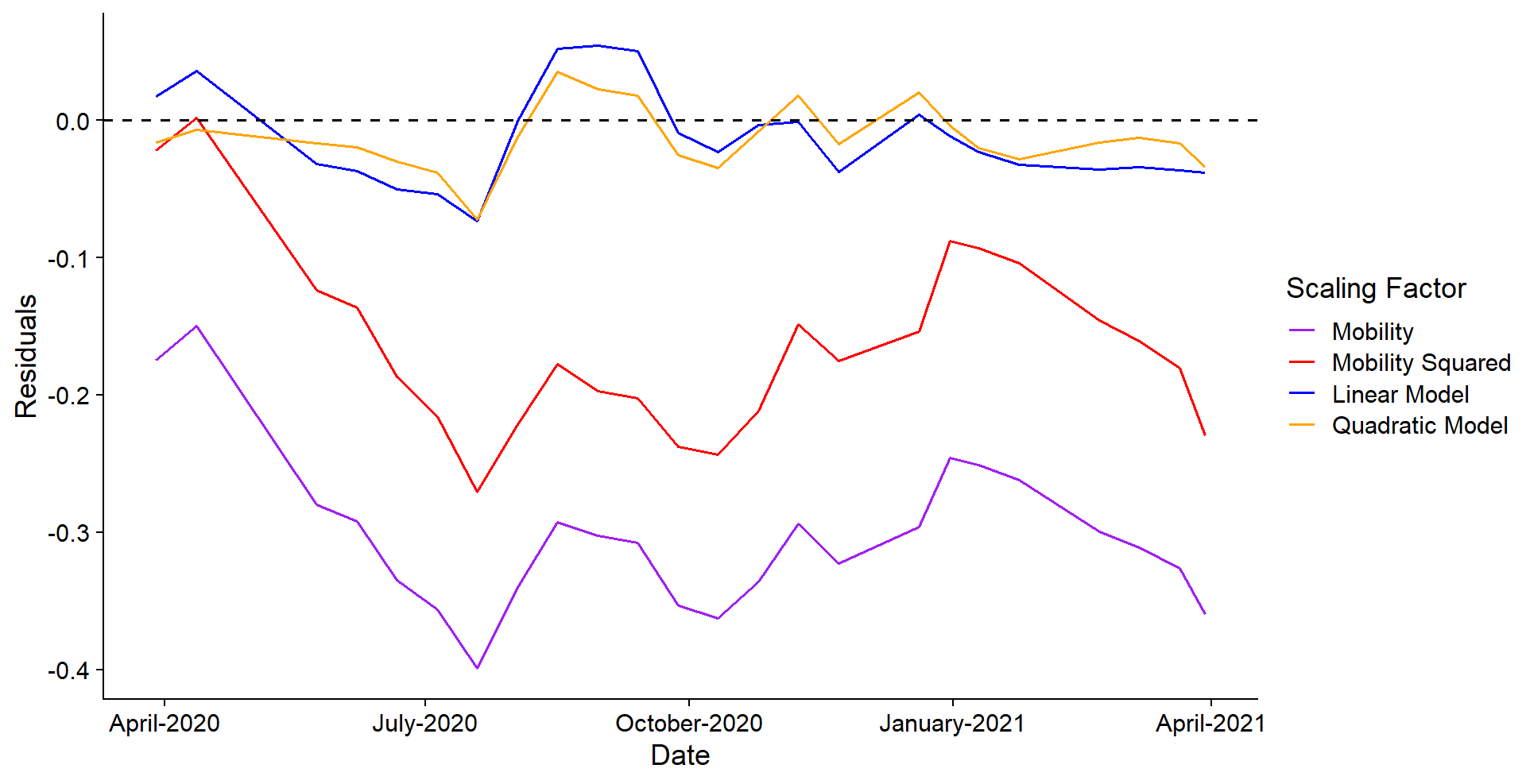


**Figure S4B:** residuals from the reproduction number estimates as compared to the estimates from the CoMix survey, the purple line shows estimates from the mobility scaled matrices, the red line shows estimates from the mobility squared scaled matrices, the blue line shows estimates from the linear model scaled matrices and the orange line shows estimates from the quadratic model scaled matrices.

## S.4.5 Reproduction numbers: UK "other" contacts

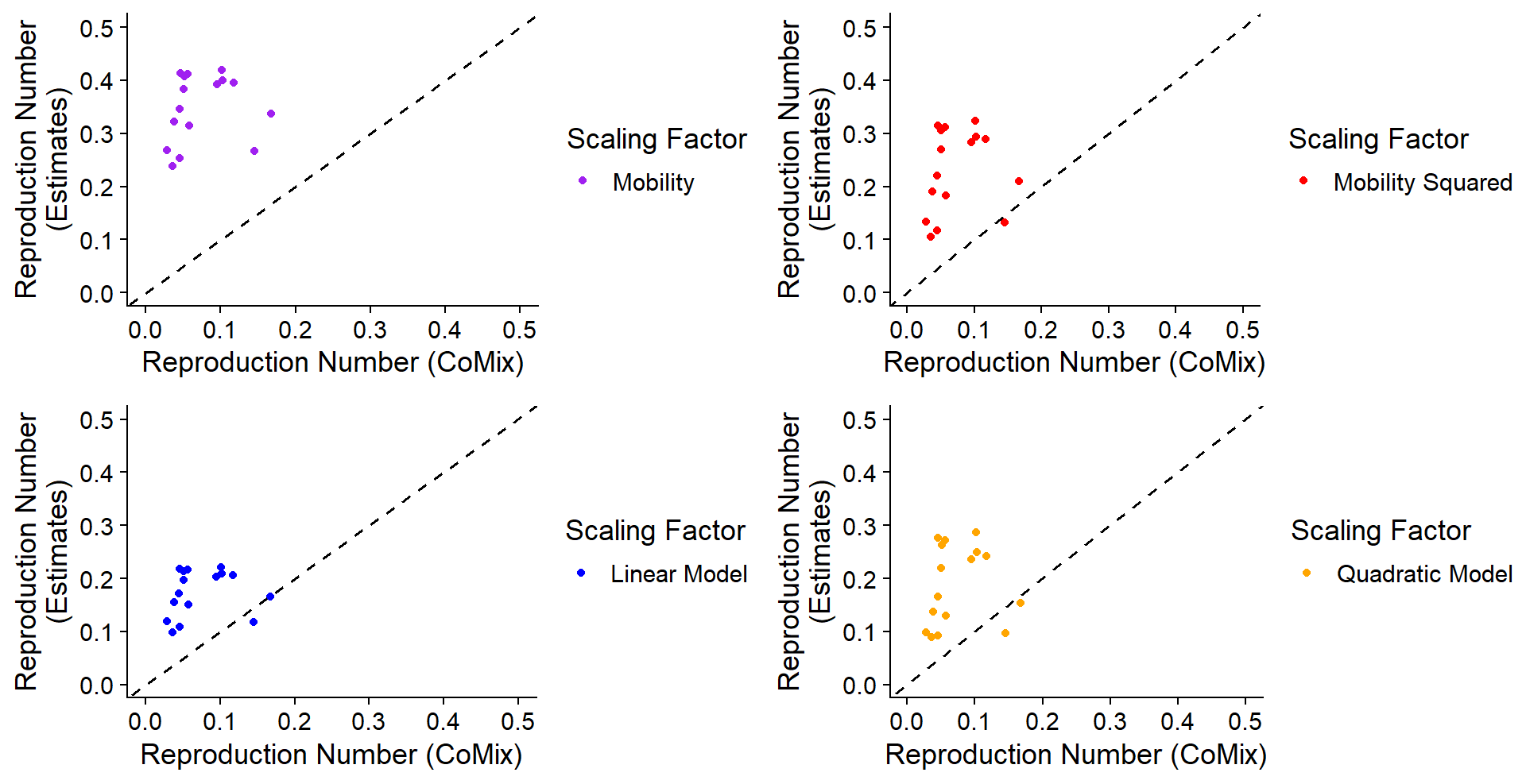


**Figure S5A:** reproduction number estimates against reproduction numbers from CoMix survey, top left shows estimates using mobility, top right shows estimates using mobility squared, bottom left shows estimates using the linear model relative contact rate and bottom right shows estimates using the quadratic model relative contact rate.

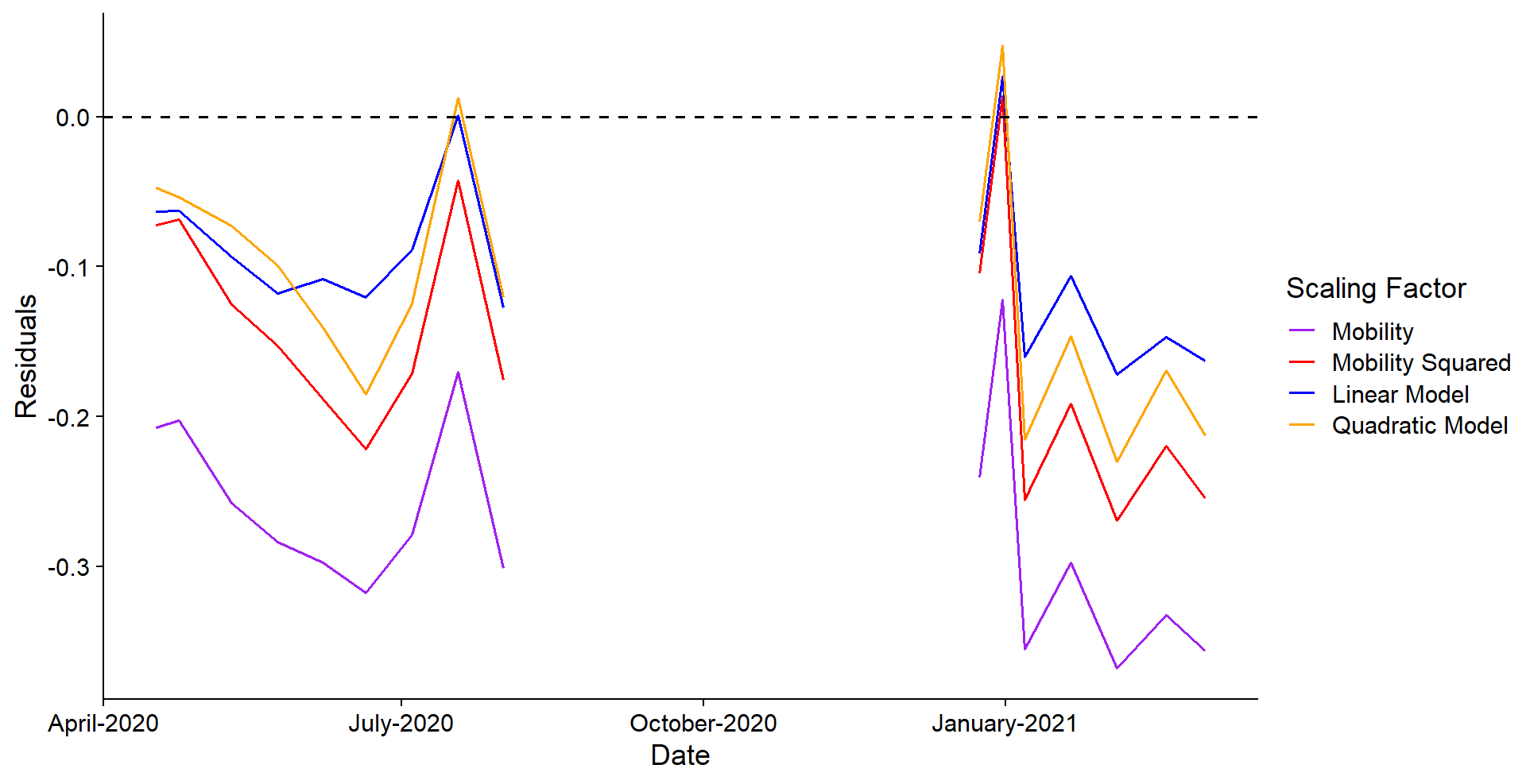


**Figure S5B:** residuals from the reproduction number estimates as compared to the estimates from the CoMix survey, the purple line shows estimates from the mobility scaled matrices, the red line shows estimates from the mobility squared scaled matrices, the blue line shows estimates from the linear model scaled matrices and the orange line shows estimates from the quadratic model scaled matrices.

## S.4.6 Reproduction numbers: BE "work" contacts

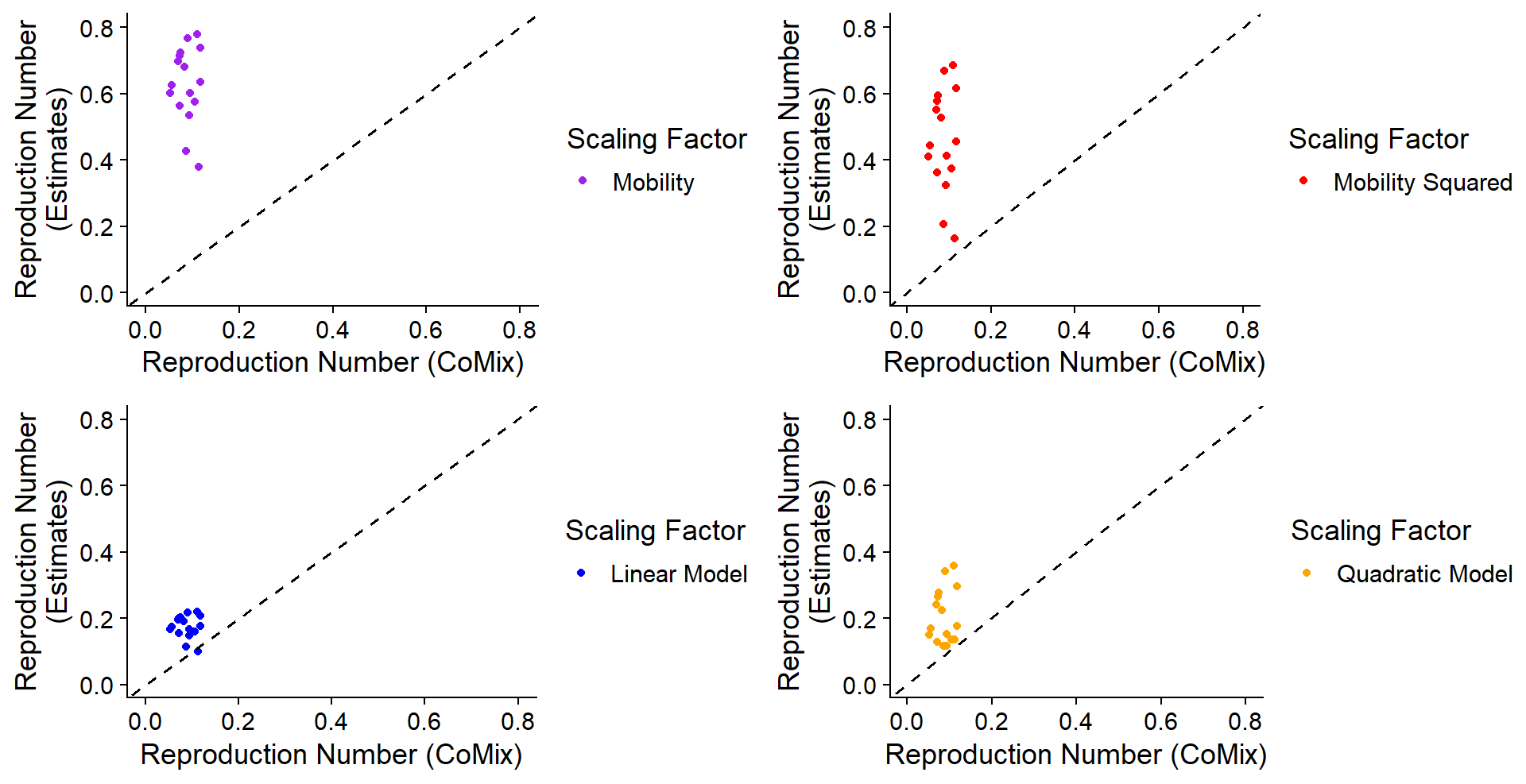


**Figure S6A:** reproduction number estimates against reproduction numbers from CoMix survey, top left shows estimates using mobility, top right shows estimates using mobility squared, bottom left shows estimates using the linear model relative contact rate and bottom right shows estimates using the quadratic model relative contact rate.

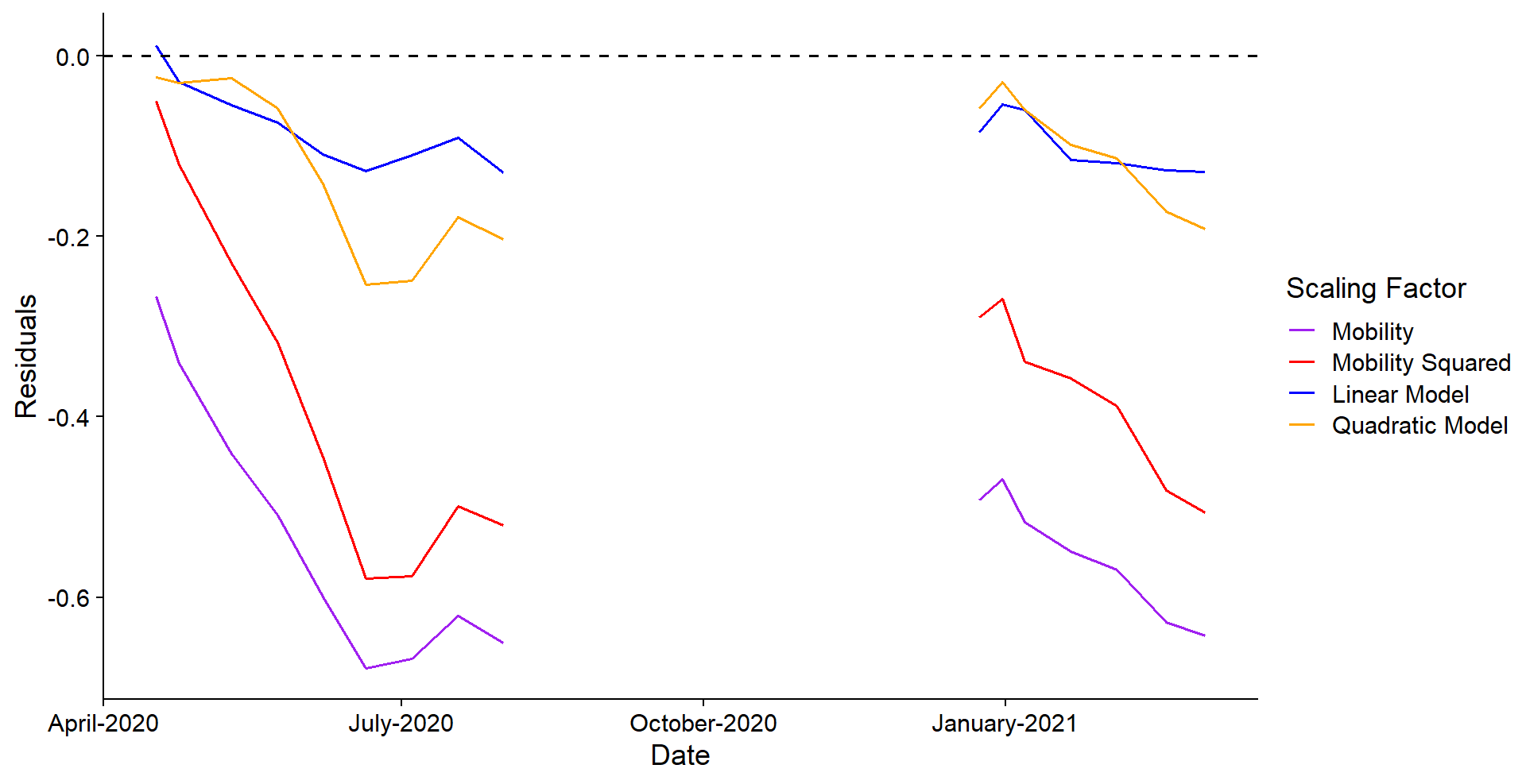


**Figure S6B:** residuals from the reproduction number estimates as compared to the estimates from the CoMix survey, the purple line shows estimates from the mobility scaled matrices, the red line shows estimates from the mobility squared scaled matrices, the blue line shows estimates from the linear model scaled matrices and the orange line shows estimates from the quadratic model scaled matrices.

## S.4.7 Reproduction numbers: BE "other" contacts

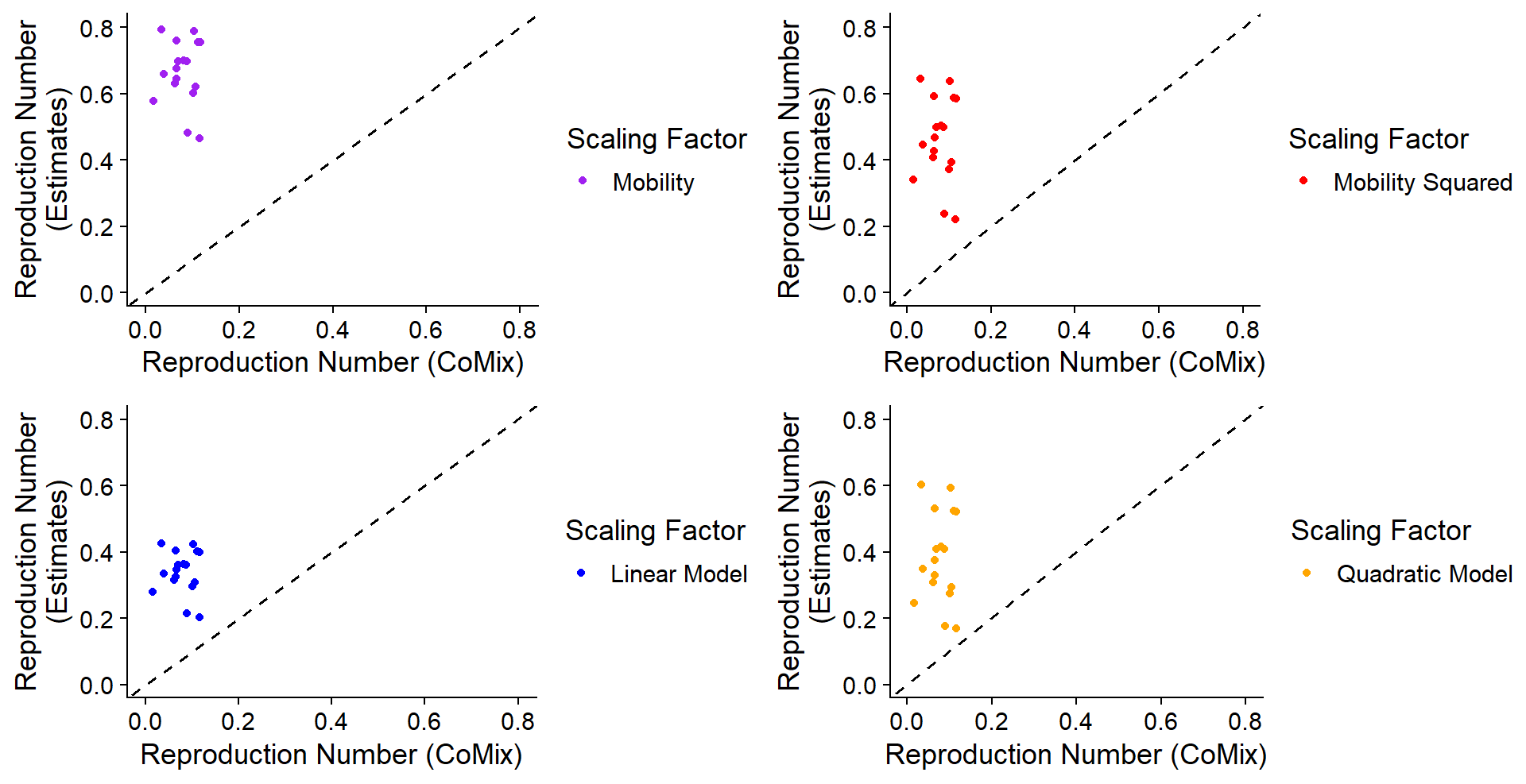


**Figure S7A:** reproduction number estimates against reproduction numbers from CoMix survey, top left shows estimates using mobility, top right shows estimates using mobility squared, bottom left shows estimates using the linear model relative contact rate and bottom right shows estimates using the quadratic model relative contact rate.

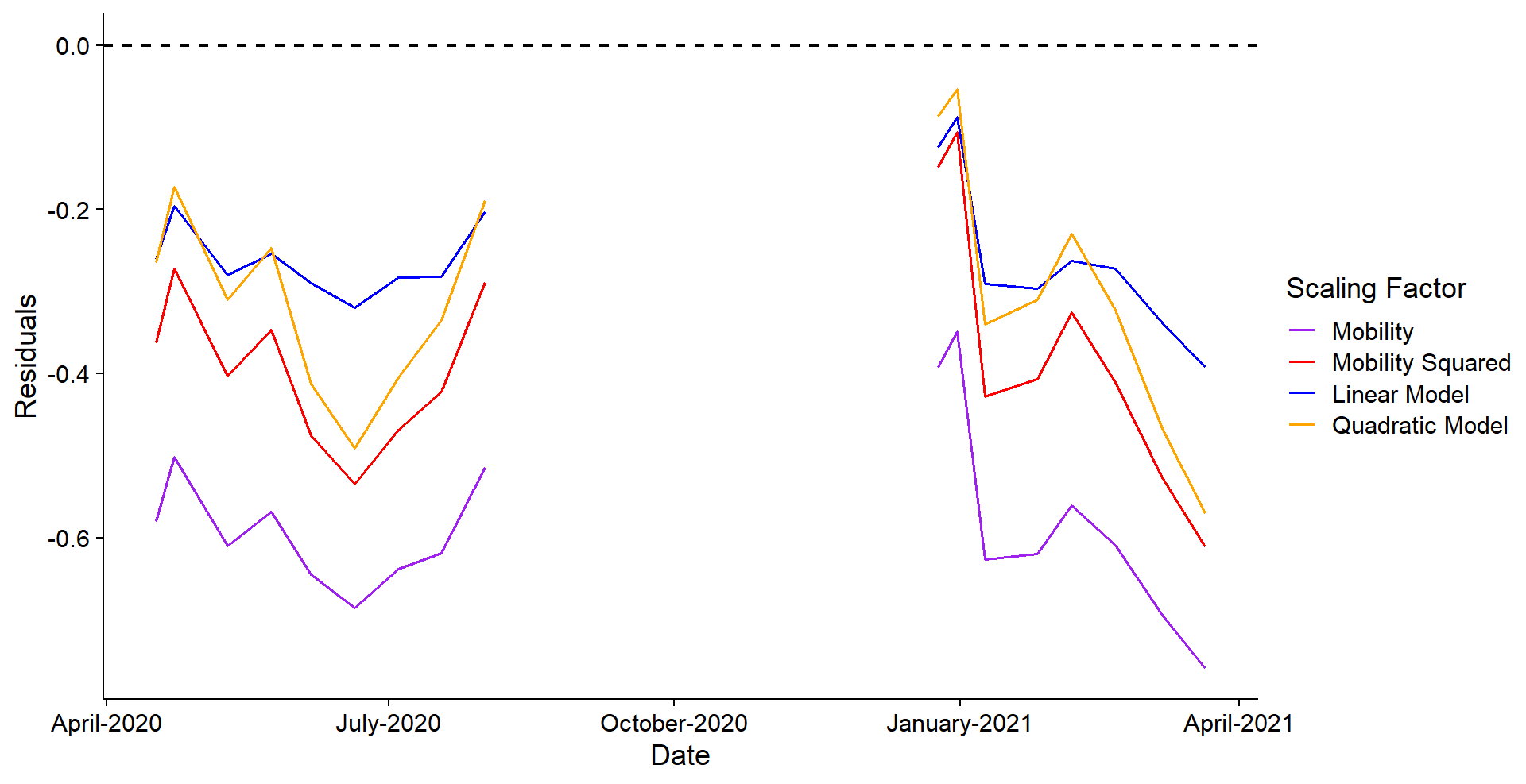


**Figure S7B:** residuals from the reproduction number estimates as compared to the estimates from the CoMix survey, the purple line shows estimates from the mobility scaled matrices, the red line shows estimates from the mobility squared scaled matrices, the blue line shows estimates from the linear model scaled matrices and the orange line shows estimates from the quadratic model scaled matrices.

## S.4.8 Reproduction numbers: NL "work" contacts

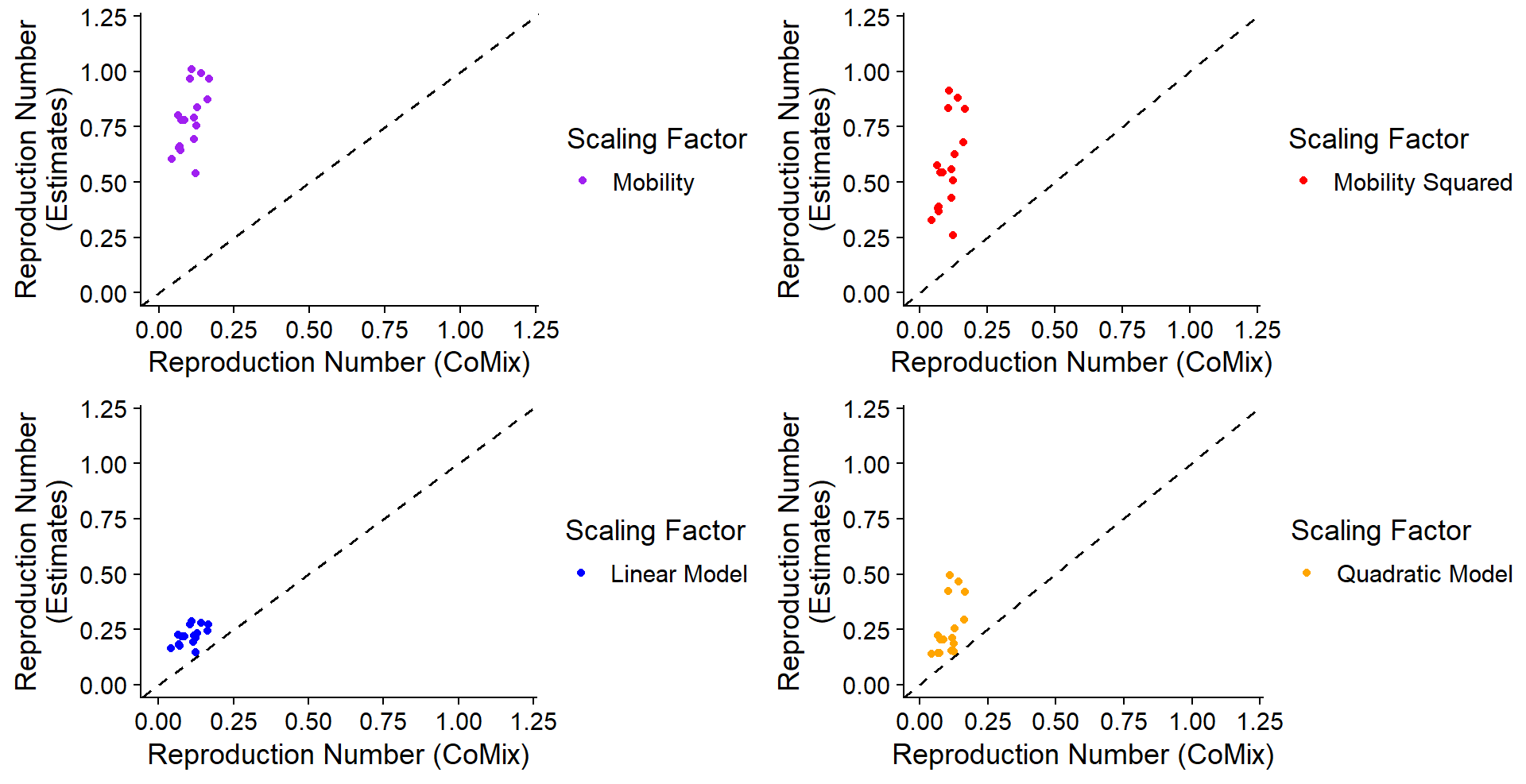


**Figure S8A:** reproduction number estimates against reproduction numbers from CoMix survey, top left shows estimates using mobility, top right shows estimates using mobility squared, bottom left shows estimates using the linear model relative contact rate and bottom right shows estimates using the quadratic model relative contact rate.

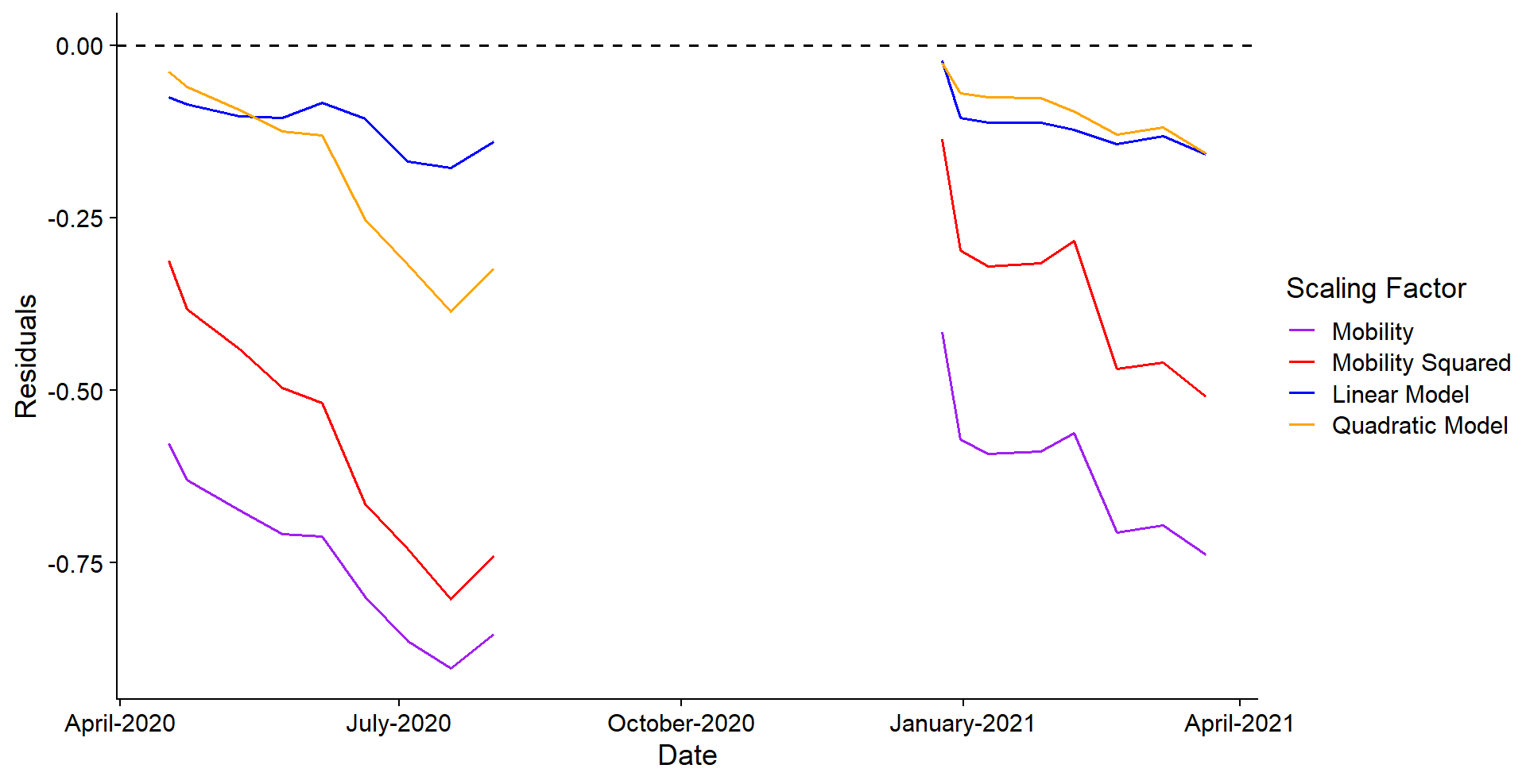


**Figure S8B:** residuals from the reproduction number estimates as compared to the estimates from the CoMix survey, the purple line shows estimates from the mobility scaled matrices, the red line shows estimates from the mobility squared scaled matrices, the blue line shows estimates from the linear model scaled matrices and the orange line shows estimates from the quadratic model scaled matrices.

## S.4.9 Reproduction numbers: NL "other" contacts



**Figure S9A:** reproduction number estimates against reproduction numbers from CoMix survey, top left shows estimates using mobility, top right shows estimates using mobility squared, bottom left shows estimates using the linear model relative contact rate and bottom right shows estimates using the quadratic model relative contact rate.



**Figure S9B:** residuals from the reproduction number estimates as compared to the estimates from the CoMix survey, the purple line shows estimates from the mobility scaled matrices, the red line shows estimates from the mobility squared scaled matrices, the blue line shows estimates from the linear model scaled matrices and the orange line shows estimates from the quadratic model scaled matrices.

# 

# S.5 Supplementary Tables

## S.5.1 Comparing contacts and mobility

| ***Year*** | ***Contact Type*** | ***Correlation*** | ***P-Value*** |
| --- | --- | --- | --- |
| 1 | Work | 0.7110 | <0.0001 |
| 1 | Other | 0.7268 | 0.0001 |
| 2 | Work | 0.3222 | <0.001 |
| 2 | Other | 0.6713 | 0.0001 |

**Table S1:** correlations between contacts and mobility for the first and second years of data, Year 1 is defined as March 2020 to March 2021 and Year 2 is defined as April 2021 to April 2022

## S.5.2 Regression models: "work" contacts

| ***Parameter*** | ***Estimate*** | ***95% Confidence Interval*** |
| --- | --- | --- |
| Intercept | -0.2273 | (-0.4882, 0.0335) |
| "Workplace" mobility | 1.3280 | (0.8890, 1.7670) |

**Table S2:** parameter estimates (and confidence interval) for linear regression model for "work" contacts

| ***Parameter*** | ***Estimate*** | ***95% Confidence Interval*** |
| --- | --- | --- |
| Intercept | 1.3169 | (0.4275, 2.2063) |
| "Workplace" mobility | -4.7718 | (-8.1921, -1.3514) |
| "Workplace" mobility2 | 5.7062 | (2.5262, 8.8862) |

**Table S3:** parameter estimates (and confidence interval) for quadratic regression model for "work" contacts

## S.5.3 Regression models: "other" contacts

| ***Parameter*** | ***Estimate*** | ***95% Confidence Interval*** |
| --- | --- | --- |
| Intercept | -0.0488 | (-0.3279, 0.2304) |
| Weighted predictor | 1.0398 | (0.5971, 1.4824) |

**Table S4:** parameter estimates (and confidence interval) for linear regression model for "other" contacts

| ***Parameter*** | ***Estimate*** | ***95% Confidence Interval*** |
| --- | --- | --- |
| Intercept | 2.9441 | (1.9893, 3.8988) |
| Weighted predictor | -9.2762 | (-12.5159, -6.0366) |
| Weighted predictor2 | 8.5566 | (5.8811, 11.2322) |

**Table S5:** parameter estimates (and confidence interval) for quadratic regression model for "other" contacts

## S.5.4 Reproduction numbers: UK

| ***Estimate Type*** | ***Root MSE*** |
| --- | --- |
| Quadratic Model | 0.0626 |
| Linear Model | 0.0649 |
| Mobility squared | 0.1001 |
| Mobility | 0.2031 |

**Table S6:** root mean squared error for all “work” estimates for the UK, this is the square root of the mean of the squared differences between the CoMix survey reproduction number and the estimated reproduction number

| ***Estimate Type*** | ***Root MSE*** |
| --- | --- |
| Quadratic Model | 0.0264 |
| Linear Model | 0.0365 |
| Mobility squared | 0.1762 |
| Mobility | 0.3070 |

**Table S7:** root mean squared error for all “other” estimates for the UK, this is the square root of the mean of the squared differences between the CoMix survey reproduction number and the estimated reproduction number

## S.5.5 Reproduction numbers: BE

| ***Estimate Type*** | ***Root MSE*** |
| --- | --- |
| Linear Model | 0.1130 |
| Quadratic Model | 0.1384 |
| Mobility squared | 0.1757 |
| Mobility | 0.2825 |

**Table S8:** root mean squared error for all “work” estimates for Belgium, this is the square root of the mean of the squared differences between the CoMix survey reproduction number and the estimated reproduction number

| ***Estimate Type*** | ***Root MSE*** |
| --- | --- |
| Linear Model | 0.0962 |
| Quadratic Model | 0.1420 |
| Mobility squared | 0.4025 |
| Mobility | 0.5520 |

**Table S9:** root mean squared error for all “other” estimates for Belgium, this is the square root of the mean of the squared differences between the CoMix survey reproduction number and the estimated reproduction number

## S.5.6 Reproduction numbers: NL

| ***Estimate Type*** | ***Root MSE*** |
| --- | --- |
| Linear Model | 0.2703 |
| Quadratic Model | 0.3346 |
| Mobility squared | 0.4050 |
| Mobility | 0.5950 |

**Table S10:** root mean squared error for all “work” estimates for the Netherland, this is the square root of the mean of the squared differences between the CoMix survey reproduction number and the estimated reproduction number

| ***Estimate Type*** | ***Root MSE*** |
| --- | --- |
| Linear Model | 0.1205 |
| Quadratic Model | 0.1793 |
| Mobility squared | 0.4971 |
| Mobility | 0.6927 |

**Table S11:** root mean squared error for all “other” estimates for the Netherland, this is the square root of the mean of the squared differences between the CoMix survey reproduction number and the estimated reproduction number