Betydelsen av anställda och anhörigas sociala nätverk för smittspridning av COVID-19 på äldreboenden under 2020

Peter Nilsson

Underlagsrapport till SOU 2021:89 Sverige under pandemin

Stockholm 2021

ISBN 978-91-525-0259-4

Betydelsen av anställda och anhörigas sociala nätverk för smittspridning av COVID-19 på äldreboenden under 2020

Sammanfattning

Denna rapport belyser i vilken utsträckning anställda och nära anhöriga påverkade mortaliteten på äldreboenden (SÄBO) under våren 2020 genom spridning av SARS-CoV-2 i deras sociala nätverk. Med hjälp av registerbaserad individdata på de boende, deras anhöriga, och de anställda på äldreboendet finner vi en stark koppling mellan storleken på och smittspridningen i de anställdas, men inte de anhörigas, sociala nätverk och dödligheten på äldreboendena.

Att mäta hur smittspridningen genom de respektive nätverken påverkar dödligheten på äldreboenden under den första vågen av pandemin är en svår empirisk utmaning av flera skäl. Under det initiala skedet av pandemin testades endast ett fåtal individer för SARS-CoV-2, och de som testades var inte slumpmässigt utvalda. Sannolikt testades anställda i en högre utsträckning än anhöriga. Å andra sidan, om sannolikheten att testa sig är positivt korrelerad med socioekonomisk status så var testbenägenheten sannolikt större i de anhörigas, än i de anställdas, social nätverk. Att använda direkta mått på smittspridning bland de anhöriga, anställda, och i deras respektive sociala nätverk för att utvärdera och jämföra påverkan på dödligheten på äldreboenden riskerar därför att leda till felaktiga slutsatser.

Av primärt dessa skäl analyserar rapporten i huvudsak betydelsen av de sociala nätverkens *storlek*. Under förutsättning att ett större antal sociala kontakter ökar risken att exponeras för smitta, och att måtten på de sociala nätverkens storlek är behäftade med en mer jämförbar felmarginal hos de anhöriga och anställda, så kan analyser av kopplingen mellan dödligheten på äldreboenden och de anställdas och anhörigas nätverksstorlek användas för att runda problemen kopplade till systematiska skillnader i testbenägenheten under våren 2020.

Mer specifikt använder rapporten information om andelen anställda och anhöriga med barn i olika åldrar i hushållet för att approximera den potentiella storleken på de två nätverken. Det finns fyra huvudsakliga skäl till valet av detta mått. För det första är åldern på barnen i hushållet en variabel som mäts med hög reliabilitet i registerdata för båda grupperna. För det andra ökar antalet potentiella kontakter inom skolmiljön systematiskt (klasstorlek, skolstorlek, och skolornas upptagningsområden) med barnens ålder. Under 2019 gick i barn i mellanstadieåldern i genomsnitt i klasser med ca 20 elever i mellanstadium med i genomsnitt ca 150 barn. Förskolebarn kan antas träffa betydligt färre (90 barn i samma förskoleenhet), medan barn i högstadiet (300 i högstadiet) och gymnasiet (500 i gymnasieskolan) i genomsnitt gick i betydligt större skolor (se Figur 1). Som jämförelse är medianantalet anställda ca 90 för Sveriges arbetsplatser under samma period. För det tredje, i linje med detta visas i rapporten att ålder på barnet är en starkt signifikant prediktor för sannolikheten att föräldrarna testas positivt för SARS-CoV2-2 under 2020. Slutligen påverkas hushåll med barn i olika åldrar olika av de restriktioner gällande distansundervisning som infördes i skolan under våren 2020. Som visas i Figur 1 så är andelen positiva test bland föräldrar med barn i gymnasieåldern (som distansundervisades under delar av 2020) signifikant lägre än de med barn i högstadieåldern. Denna policydrivna variation i föräldrarnas potentiella nätverksstorlek ger skarpa prediktioner för hur smittspridningen in på äldreboende ska variera över tid beroende på åldern på de anställdas och anhörigas barn som kan testas i data. Om den potentiella nätverkstorleken bland de anhöriga och anställda är kopplad till dödlighet på äldreboenden så ska t.ex. anställda med barn i

gymnasieåldern ha en signifikant mindre påverkan på mortaliteten på äldreboenden än de med barn i grundskoleåldern, men endast efter att gymnasieskolan övergick till distansundervisning.

I rapporten studeras av dessa skäl först hur andelen anställda och anhöriga med barn påverkar mortaliteten på äldreboenden i Stockholm under våren 2020. Analyserna tar hänsyn till skillnader i de boendes bakomliggande riskfaktorer, så som skillnader i underliggande hälsa och ålder, ursprung och utbildning, och motsvarande typ av karaktäristika hos de anhöriga och anställda. Denna analys ger stöd för att ett större potentiellt nätverk bland anställda, men inte anhöriga, ökar dödligheten på äldreboendena. En högre andel anställda med barn i hushållet är kopplad till en signifikant högre dödlighet på äldreboende. Rensat för skillnader i de observerade egenskaperna hos de boende, anhöriga, och anställda tyder skattningarna på att mortaliteten var ca 25 procent högre på de äldreboenden med högst andel (90 percentilen) anställda med barn i hushållet jämfört med de äldreboenden med lägst andel (10 percentilen) anställda med barn i hushållet. När vi separerar andelen anställda med barn i gymnasieåldern (vars potentiella nätverksstorlek påverkades direkt av fjärrundervisningsrestriktionerna), finner vi att en högre andel anställda med barn i gymnasieåldern är associerad med en signifikant *lägre* mortalitet i nästan samma storleksordning. Resultatet ger stöd för hypotesen att den potentiella nätverkstorleken bland de anställda genom deras barn är det som driver korrelationen mellan barnets ålder och ökningen i mortaliteten på äldreboendena under våren 2020. Skillnaderna i effekten på dödlighet mellan anställda med barn i gymnasieåldern och övriga åldrar är dessutom starkare efter att distansundervisningen har kunnat ge effekt. Som ytterligare stöd för nätverkstorlekshypotsen så finns det före pandemin inga skillnader i dödlighet på äldreboendena beroende på hur stor andel av de anställda som har barn i respektive ålder.

I nästa steg skattas samma modeller med nationella data och ett liknande mönster återfinns: ingen effekt av anhörigas potentiella nätverksstorlek, men en signifikant högre mortalitet på boenden med en högre andel anställda med barn som inte direkt påverkades av fjärrundervisningsrestriktioner. Detta gäller såväl på mortalitet oavsett dödsorsak, som specifikt för mortalitet där COVID 19 bedömts vara huvud- eller bidragande orsak till dödsfallet.

I det sista steget av analysen studeras hur andelen bekräftat smittade anställda och anhöriga påverkar sannolikheten att de boende blir bekräftat smittade eller dör. Även med hjälp av dessa data, med alla dess brister, finner vi ett liknande mönster som i huvudanalysen. De anställdas, men inte anhörigas, bekräftade smitta korrelerar med både risken för bekräftad smitta såväl som dödlighet på äldreboenden. Vi visar dock att en del av denna korrelation sannolikt drivs (åtminstone delvis) av systematiska skillnader mellan boenden som testar mycket och de som testar lite. Slutligen visar vi att andelen bekräftat smittade i de anställdas – men inte anhörigas – bostadsområden är positivt och signifikant korrelerade med dödligheten på äldreboendena och att smittvägen mellan bostadsområdet och äldreboendet verkar uppstå genom de anställda.

Tillsammans ger de samlade resultaten stöd för att SARS-CoV-2 under den första vågen av pandemin, i huvudsak, fördes in via de anställdas sociala kontakter snarare än de anhörigas kontakter. Potentiella bidragande orsak till dessa skillnader är sannolikt en lägre kontaktintensitet mellan anhöriga och boende jämfört med mellan anställda och boende, samt möjligen de besöksförbud som infördes på många äldreboenden vid olika tidpunkter under våren 2020.

1. Introduction

The objective of this report is to shed some light on the role of SARS-COVID-2 infections in the social networks of employees and relatives as determinants of mortality in nursing homes (SÄBO) during the first wave of the COVID-19 pandemic in Sweden. To this end, the report uses matched administrative data on all individuals residing in nursing homes, their close relatives and the employees in the nursing homes. We find a strong association between the size of, and the infection rate in, the employees' – but not the relatives' – social networks and mortality in the elderly homes.

To measure how the transmission of SARS-CoV-2 through the respective social networks affects the mortality in the nursing homes is a daunting empirical challenge during the first wave of the pandemic for several reasons. During the first wave in Sweden, only a limited number of SARS-CoV-2 tests were administered, and those tested were far from randomly sampled. Employees were most likely tested to a higher extent than relatives. On the other hand, to the extent that the testing frequency was positively associated with socioeconomic status (e.g. educational attainments), it is likely that the social contacts (e.g. the neighbors) of the relatives were tested to a higher extent than the social contacts of the employees, due to employees living in less affluent neighborhoods to a much larger extent. Using direct measures of SARS-CoV-2 infections among relatives, employees and their respective social networks to evaluate and compare the impact of each respective group's contribution to mortality in the nursing homes therefore risks leading to biased conclusions.

Primarily for these reasons, the report mainly focuses on the impact of the *size* of the respective social networks. Under the assumption that a larger number of social connections, all else equal, increases the risk of infection, and that the measurement error of the size of the social network is more evenly distributed between the employees and the relatives, we can compare how mortality in the nursing homes is linked to the potential size of the employees' and relatives' social networks to avoid the problems associated with systematic testing differences during the first wave.

More specifically, the main analysis uses information on the share with children in the household to proxy for the potential network size of the employees and the relatives. There are four main underlying reasons for this choice. First, the age of the children in the households is measured with high reliability in administrative data in both groups. Second, the potential number of social contacts (class-size, school-size, and the school catchment area) in the Swedish school system is increasing systematically with the age of the child. During 2019, children in *mellanstadiet* (ages 10-12) attended schools with around 150 children on average in the same age range with an on average class-size of around 20 children. Kindergarten age children can be assumed to meet significantly fewer children on a daily basis (childcare units have around 90 kids on average, and around 90-95% of the children attend childcare in ages 2-5), while children in *högstadiet* (ages 13-15, on average 300 children in a school), and high schools (on average 500 children) attend schools with a significantly larger number of potential social contacts (see Figure 1). For a comparison, the median workplace in Sweden during the same period had 90 employees. Third, in line with this systematic increase in the potential network size of the children, the share of parents who test positive for SARS-CoV-2 during 2020 is strongly significantly and positively associated with the age of their child (see Figure 2). Finally, households were differentially affected by the remote-teaching mandates affecting high schools that were implemented during the spring of 2020 depending on the age of the child. As shown in Figure 2, the share of parents testing positive for SARS-CoV-2 is a discontinuous function of the age of their child around the high school starting age cut-off (age 16), with a significantly higher share of *högstadie* age child parents testing positive, as compared to high school age child parents. This policy driven variation in the parents' potential network size gives rise to sharp predictions for how the virus should contribute to the mortality rates in the nursing homes depending on the timing and the age of the employees' and relatives' children that can be tested in the data. If the potential network size of the employees and relatives is connected to mortality in the elderly homes then, for example, the share of employees with high school age children should have a significantly lower effect on mortality in the nursing homes than the employees with compulsory school age children, but only after high schools moved to online-teaching.

For these reasons, the report starts by analyzing how the share of employees and relatives with children in the household affects mortality in the nursing homes during the first half of 2020 in Stockholm, the epicenter of the first wave. The analysis takes into account underlying differences in the risk factors of nursing home residents, such as differences in underlying health, age, immigration status, marriage status, and education, as well as similar characteristics of the employees and relatives. This analysis supports that a larger potential network size of employees - but not relatives – increases the mortality in the nursing homes. A larger share of employees with children at home is associated with a significantly higher mortality rate in the nursing homes. Adjusted for differences in the observable characteristics of the nursing home residents, their relatives, and the employees, the estimates indicate that residing in a nursing home with the largest share of employees with children in the household (90 percentile) is associated with a 25 percent higher risk of dying between March and June 2020 as compared to if residing in a nursing home with the lowest (10 percentile) share of employees with children in the household. When splitting out employees with high-school age children in the household (i.e. those directly affected by the remote-teaching restrictions), we find that a larger share of employees with high school age children is associated with a significantly *lower* mortality in the nursing homes of almost the same magnitude. This result gives support for the hypothesis that the potential network size of the employees is driving the correlation between the presence of children in the household and the mortality in the nursing homes during the first wave. The differences in the effects on mortality are furthermore particularly strong after the remote teaching mandates could have had time to reduce the transmission of the virus from children to parents to nursing homes residents. As additional support for the network size hypothesis, we find no indication that the mortality rates in the nursing homes *before* the pandemic are associated with the share of employees with children in different age groups in the household. This result rules out that the observed association *during* the pandemic is caused by preexisting systematic differences in the mortality risk across nursing home units with a larger share of employees with children in the household.

In the next step of the analysis, we estimate the same type of models using national data and find a similar pattern: no impact of relatives' potential network size, but a significantly higher mortality rate in nursing homes with a larger share of employees with children who were not directly affected by remote-teaching mandates. This applies both in terms of overall mortality, and specifically for mortality rates where COVID 19 was judged to be the main, or contributory, cause of death.

In the final step of the analysis, we leverage the direct measures of confirmed infections made available to us, and study how confirmed infections among the employees and relatives are associated with confirmed infections and mortality among the nursing home residents. With the help of the confirmed infection data, keeping all of its limitations in mind, we find a similar pattern as in the baseline analysis. The employees', but not the relatives', confirmed infections are correlated with both the risk of confirmed infections and mortality in the nursing home population. However, we show that this correlation, at least in part, is likely driven by systematic differences across nursing homes that test their employees a great deal vs those with limited testing among their employees. Finally, the report shows that the share of confirmed infections among the neighbors of the employees – but not relatives' neighbors– is positively and significantly associated with mortality in the nursing homes, and that the link between infections in the neighborhood and the nursing homes seems to be mediated via the employees.

Together, the collected pieces of empirical evidence suggests that SARS-CoV-2 during the first wave of the pandemic, which was so devastating for the nursing home population,[1] mainly entered the nursing homes via the social connections of the employees rather than via the social connections of the close relatives. Likely contributing to these findings are the closer and more frequent interactions between the employees and the nursing home residents, as well as the ban on visits from relatives that was introduced in many nursing homes at different times during the first-wave.

[1] By mid-June 2020, around of 50% of the Covid19 death toll in Sweden people were in elderly care services (Szebehely, 2021).

The rest of the report is structured as follows. Section 2 outlines a framework for the identification challenges of contagion effects from an individual's social network on individual outcomes. Section 3 describes the data available to us and how we link the nursing home population to employees and relatives. Section 4 outlines the empirical model. Section 5 provides a first set of descriptive results linking observable characteristics to individual confirmed SARS-CoV-2 infection data. Section 6 provides the main results for nursing homes in the Stockholm and the national samples. Section 7 summarizes and concludes the report.

2. The identification problem of social network effects in observational data

We want to shed some light on the relative role of employees and relatives and their respective social networks in explaining the transmissions of SARS-CoV-2 into the nursing homes in Sweden. In order to do so, we first need to choose how to measure infection in the nursing homes. For residence in the nursing homes (SÄBO), it is well-documented that far from all residents were tested and even if testing positive, the vast majority of the deceased aged above age 80 never went to intensive care units (Stern and Klein, 2020). As a result, the best measure available of SARS-CoV-2 infections, and perhaps the most interesting in the SÄBO population, is mortality.

Second we need to choose the relevant social connections of the elderly. Compared to many other situations, the relevant social network in the present context is unusually well defined since the nursing home population to a large extent is confined to the nursing home and the social interactions occurring there. Hence, we are relatively comfortable in assuming that employees in the elderly homes and the close relatives constitute, if not the complete social network of the elderly, the lion share of the social connections with whom the elderly in the nursing homes interact during the pandemic. Therefore, we define the employees and the close relatives and their respective social network as the key social networks of the elderly in the homes.[2] We return to the more difficult task of how to define and measure the relevant social network of the relatives and employees in the following sections of the paper.

Third, we need to pick a variable that reliably measures infections or the risk of infections among the employees and relatives and their social network. To motivate our choice of measures and to fix ideas, we revisit the challenges in the estimation of contagion effects using observational data as outlined by Manski (1993). Manski (1993) synthesizes the problems associated with the identification of contagion/social interaction/spill-over effects from social networks on individual outcomes using observational data. In the linear model analyzed, Manski provides formal

[2] Of course this is not strictly true, but during the pandemic this is likely a reasonable simplification of the actual social connections of the elderly in the nursing homes. Other contacts include various service deliveries and consultants from the outside (food deliveries, hairdressers, etc.)

expressions for three hypotheses for why individuals belonging to the same social network are observed to behave similarly, or in the current context to contract SARS-CoV-2. Adapted to the current context, the hypotheses are (a) *Endogenous effects*: the probability that the elderly contract SARS-CoV-2 is a function of the infection rates of their social network. (b) *Contextual effects:* the propensity to contract SARS-CoV-2 varies with the exogenous characteristics of the social network. (c) *Correlated effects*: the probability of contracting SARS-CoV-2 among individuals in the same social network (the elderly, the employees and the relatives) does co-vary, not because they are more or less likely to infect each other, but simply because they face similar individual characteristics or institutions.

To clarify the distinctions in the present context, and focusing on the employee network, according to Manski there is an *endogenous* effect if, all else equal, elderly infections are increasing in the average infection rate in the employees' social network. There is a *contextual* effect if infections tend to co-vary with, for example, the socio-economic composition of the employees' social network. There is a *correlated* effect if both employees and nursing home residents become infected directly through other "non-social" channels,[3] or more directly relevant for the main outcome variable used in this report, that nursing homes with a persistent higher average mortality rate even before the pandemic simply are those that happen to also have a larger share of infected employees and/or relatives. In cross-sectional observational data, all three of these hypothesized channels can generate correlations between the average infection rates in the three groups (elderly, employees, and relatives), and the mortality rates in the nursing homes.

A first-order question is whether we can distinguish the social effects (the endogenous and contextual effect) from the non-social effect (i.e. the correlated effect). The present context provides its own challenges and possibilities in this regard. We next discuss the rich data made available for this report and then describe our preferred estimation procedure given the constraints of the data and the framework outlined above.

3. Data

3.1 Linking the elderly to their relatives and the employees of the nursing homes.

The data we use derives from population-based register data and was made available to us by the Stockholm University Covid-19 program, which has sprung up with support from the Corona Commission as described in Appendix A. (in Swedish). From the National Board of Health and Welfare (Socialstyrelsen), we get information on everyone assigned to a Särskilt boende (SÄBO). Access and

[3] For example, food deliveries or via ventilation in the elderly homes, or visits to hospitals (or through any other channel which we do not observe), rather than by each other.

assignment to these homes are based on the need for full-time assistance and can be run by public and private providers, but they are administered by the local governments (municipalities). The SÄBO data contains information on the start and end date of the spell. End dates occur due to changes in SÄBO homes, but in practice almost all cases are due to the fact that the patient dies.

The national SÄBO data made available to us includes an individual index number, a by Statistics Sweden deidentified personal number, as well as the address of the SÄBO home and the start-end date of the elderly home residence spell. In addition, in this report we also use a more detailed data set on nursing home residents directly delivered to us from Stockholm municipality.

The index number allows us to link each individual to information from Statistics Sweden on their relatives and a wide range of observable background characteristics (age, immigration status, education, etc.). The index number also allows us to link each individual to their biological children and grandchildren, and their respective background characteristics (including information on their age, and neighborhood of residence etc.).

The SÄBO address allows us to link each SÄBO to firm registers collected by Statistics Sweden which allows for linking each resident to all employees that work in the particular SÄBO. As for the relatives, the employees' index number allows us to link each employee to their children and partners, their place of residence etc. The linking between nursing home residents and the (workplace identifier of the) employees is based on the address of the elderly homes. We used an automated matching procedure by first restricting the data to the municipality of the nursing home, and then using a fuzzy matching procedure in STATA for the SÄBO address and the address of the workplace in the elderly care sector in the same municipality. Manual checking of the Stockholm municipality data matches suggests that the matching procedure works relatively well; however there are inevitably matching errors and, as a result, erroneous assignments of employees to the specific nursing homes. Other sources of errors in the matching procedure include those where the address of the nursing home residents cannot be matched with the employer register. To avoid errors from this type of matching error, we restrict the analyses to nursing home residents where there are at least five employees assigned to the specific address. At other times, the procedure matches nursing homes with a similar, but not identical workplace address in the same municipality, to the same workplace identifiers. For such cases, we assign the same worker to both nursing homes and then collapse the worker data at the nursing home level.

We used the same matching procedure in the national data that we used and manually checked for errors in the Stockholm data and end up with a matched analysis sample of approximately 3500 nursing homes with around 50 000 residents who resided in the nursing home at the end of 2019 and were still alive at the beginning of March 2020.

3.2. Health data

The health data stems from various registers. The inpatient data covers all individuals ever admitted overnight in a hospital. The outpatient registers contain data on individuals visiting a health care specialist. Both patient registers include information on the main diagnosis code (ICD10) for the visit as well as secondary contributory codes. These data sources first of all allow for information on the general health status, and the number of risk factors linked for severe instances of SARS-CoV-2 (the data construction of these measures is described elsewhere). In addition, we use data from the SMINET database covering all individuals that tested positive for SARS-CoV-2 along with the index number and the date of the test. As noted above, the testing capacity was low for the general population at the beginning of the pandemic (spring 2020), and was only progressively expanded. The testing was, and still is, far from complete or random. Moreover, we do not have access to any information on how many individuals that were tested in a specific social network.

The main outcome variable used is mortality from any cause derived from the cause of death register, but we also use data on mortality where COVID-19 was deemed to be the primary cause of death or contributed to the death (data construction described elsewhere). We choose to primarily focus on the deaths due to any cause for two reasons. During the first wave, very few of the nursing home residents ended up in intensive care units, and hence focusing on deaths due to COVID-19 will only likely understate the true death toll from SARS-CoV-2 infections in the nursing home population. However, we do also report results for the COVID-19-specific mortality. Second, using overall mortality irrespective of cause of death allows us to check the underlying assumptions of our estimations by testing if the measure of infections/network size is correlated with mortality rates also *before* the pandemic, which would cast some doubt on the interpretation of any correlations found *during* the pandemic.

4. Empirical Specification

This section outlines the empirical model that we use to estimate the impact of employees' and close relatives' social network on mortality in the nursing homes. Suppose that the true model governing SARS-CoV-2 mortality in the elderly homes is given by

$$Pr\left(Death_{ij}\right) = \alpha + \beta_1 Netw_{EMPLOYEE\,j} + \theta_1 Netw_{RELATIVE\,j} + X_i{}'\delta + Z_j{}'\pi + \varepsilon_{ij}, \quad (1)$$

where *i* indexes individuals, *j* elderly homes, $Death_{ij}$ is a dummy indicating death occurring within a specified period of time (e.g. Spring 2020), $Netw\_EMPLOYEE_j$ and $Netw\_RELATIVE_j$ measure the SARS-CoV-2 infections from employees and close relatives, X_i are observed and unobserved personal characteristics, Z_j are observed and unobserved network and elderly home characteristics, and ε_{ij} is the error term.

Measuring $Netw_{EMPLOYEE\,j}$ and $Netw_{RELATIVE\,j}$ is difficult. First we do not know exactly which employees come in contact with the elderly, nor do we know which of the close relatives that visit the elderly. Moreover, the omitted variable bias is exacerbated since individuals with many contacts in either of the networks may be systematically different from one with few contacts (small elderly home, few close relatives). Finally, especially during the first part of the pandemic, SARS-CoV-2 infections in the network are measured with error, and the measurement likely differs between the two networks since employees in the homes likely had greater access to testing facilities and were more closely monitored for infections.

Putting these measurement issues aside, one could estimate

$$Pr(Death_{ij}) = \alpha + \beta_1 \overline{INFECTION}_{EMPLOYEE\,j} + \theta_1 \overline{INFECTION}_{RELATIVE\,j} + X_i{}'\delta + \varepsilon_{ij}, \quad (1)$$

where $\overline{INFECTION}_{EMPLOYEE\,j}$ and $\overline{INFECTION}_{RELATIVE\,j}$ represent the mean infection rates in the employee and relative networks, respectively, and $X_i{}'\delta$ reflects individual characteristics such as relevant co-morbidities. This regression model suffers from what Manski (1993) dubbed the "reflection problem". Is mortality in the elderly homes due to the relative infection rates and the behavior of the respective social contacts (network effects), or are e.g. employees vectors for SARS-CoV-2 infections to both elderly and relatives, or are the infection rates in all three groups driven by some third unobserved channel (e.g. via visits to hospitals or by in home external health care providers)? Omitted individual or elderly home characteristics may be correlated with the infection rates in either of the social networks. For example, nursing homes located in less affluent areas may have a more fragile elderly population as well as a higher infection rate among the relatives/employees in general. In general, any shock that directly affects the infection rates in the whole network (elderly, employees and/or relatives) will result in a simultaneity problem, in turn resulting in a positive correlation between the infection rates in the social networks and mortality in the nursing homes, and potentially biasing the estimates of β_1 and θ_1 in uncertain directions. Thus, even in the absence of measurement errors in the actual infection rates of the two networks, a comparison of the size of the two estimates of β_1 and θ_1 cannot be clearly interpreted as clearly informative of the relative importance of the two social networks' influence on mortality in the elderly homes.[4]

[4] A simple solution to the simultaneity problem is to use a dynamic specification, i.e. specify with which lag the elderly are infected by the employees and the relatives. For example, in this setting, it could be natural to

An alternative approach to get around the measurement error problems is to instead focus on the role of contextual network effects. In other words, we need to find observable predetermined characteristics of the employees and the relatives that are predicting infection rates in these two groups. To the extent that they are strong predictors of infections, observable for both groups and well measured, a model of the following form could potentially allow for an easier comparison of the relative importance of the two networks

$$Pr(Death)_{ij} = \alpha + \gamma\overline{PRED\_INFECT\_EMP_j} + \theta\overline{PRED\_INFECT\_RELA_j} + X'\beta + Z'\beta + \varepsilon_{ij} \quad (2)$$

In general, in this model, $\overline{PRED\_INFECT\_EMP_j}$ and $\overline{PRED\_INFECT\_RELA_j}$ are the average predicting characteristics of the infection rates in the respective network. An advantage of this specification in the specific context is that it also captures unconfirmed/asymptomatic infections in the networks that, in theory, once they reach the nursing homes, could wreak havoc in the much more sensitive nursing home population. A disadvantage of this reduced form specification is, however, that it is necessary to ensure that the predicting characteristics solely influence the mortality rates in the elderly homes via their proximate role for SARS-CoV-2 infections. With the data at hand, we address this concern in several ways in the results section.

To measure $\overline{PRED\_INFECT\_EMP_j}$ and $\overline{PRED\_INFECT\_RELA_j}$, we primarily focus on how the *potential network size*, as measured by the share of employees and relatives with children in their homes, influences parental infection rates and in the end the mortality rates in the nursing homes. Figure 1 and Figure 2 serve to illustrates the key idea behind this approach. Irrespective of whether measured by the average number of children in the same class or by school size, Figure 1 shows that the size of the potential network to which the parents are connected is increasing with the child's age. Figure 2 shows that there is a strong positive correlation between confirmed SARS-CoV-2 infections during 2020 among parents and the age of their child in months. However, there is a significantly lower share of confirmed infection rates among parents of high school students relative to parents of slightly younger children. This pattern is consistent with Vlachos, Hertegård, and Svaleryd (2020) who study how the introduction of remote teaching for high school students and thus, a sharp reduction in the potential network size, affected parents and teachers in

---

assume that the infection rates in period *t* in either of the social groups take e.g. two weeks to transmit to the elderly and another two weeks before an elderly person dies. With such a dynamic structure, the reflection problem is solved, under the maintained assumption of the specific lag structure being valid. In the present context, the dynamic framework is a seemingly attractive solution to achieve identification, and hence allows for a direct comparison of the relative importance of the two social groups connected to the individual. However, the fundamental concern that $\overline{INFECTION}_{EMPLOYEE\,jt-1}$ and $\overline{INFECTION}_{RELATIVE\,jt-1}$ are measured with error remains, and that the degree of measurement error is likely to differ between the two social groups.

the spring of 2020. Here we show that the same pattern seems to hold for the entirety of 2020. Adjusting for the age of the parents and the region of residence, if anything, reinforced the pattern in Figure 2, thus suggesting that the correlation in Figure 2 is not simply due to a higher sensitivity (and hence more testing) of older parents.[5] The discontinuity at the high school starting age also suggests that the pattern in Figure 2 is not likely solely due to an on average increasing viral load by the age of the child. Figure 3 plots the age distribution of nursing home employees with children and the age distribution of the children of the residents in the elderly homes in Stockholm, and the overlap of the two distributions suggests that there is plenty of variation that can be used to compare the influence of the two networks.

Motivated by Figure 1-3, and under the maintained assumption that children act as connections between (even otherwise completely isolated) parental networks and that a higher number of social connections is linked to a higher risk of infection, all else equal, we can write the predicted infection rates among e.g. employees as

$$\overline{PRED\_INFECT\_EMP}_{ja} \approx \left(\begin{array}{c} \text{Share of employees} \\ \text{with children in age group } a \\ \text{in elderly home } j \end{array} \right) \times \left(\begin{array}{c} \text{Potential network size} \\ \text{of children in age group } a \end{array} \right)$$

With access to information on the number of children in each child's daycare group/class or daycare center/school, one could have tested directly how the actual size of the potential network of the employees and relatives influences the mortality rates in the elderly homes. However, a potential problem with such an approach would be that using the actual size of the children's school network could introduce omitted variable biases. Parents choosing smaller schools for their children could be systematically different than parents placing their children in larger schools/daycare centers.

Absent such data, we do instead build on the information provided in Figure 1 and 2 and simply proxy for the infection rates of the employees (and the relatives) using information on the age of the children residing in the homes of the employees and close relatives. In practice, the data at hand allows us to construct the proxys using the share of employees/relatives with children at home in predefined age bins that roughly correspond to the shifts in the potential network sizes as illustrated in Figure 1

$$\overline{PRED\_INFECT\_EMP}_j = \overline{CHILD_{0-5}}_j + \overline{CHILD_{6-15}}_j + \overline{CHILD_{16-18}}_j + \overline{CHILD_{19-25}}_j \quad (3)$$

[5] Not reported here but available upon request.

and plug these measures into equation (2). As noted in the introduction, in addition to capturing shifts in the potential network size, the heterogeneity in the student geographic residence also increases with age as the school catchment areas are expanding, potentially also playing a role for our results since the parents of the older age group of children are not only exposed to more connections but also to more connections from geographically remote neighborhoods and even different municipalities. As an indication, around 3% of children in grade 1 commute from a school in a different municipality than the one in which the school is located, a share that increases to around 7% of students in grade 9, to around 30% in high school), which may potentially facilitate the spread of SARS-CoV-2 across larger areas via the school networks.

Of particular interest in equation (3) is the influence of having high school children at home, $\overline{CHILD_{16-18}}_{\ j}$, since the potential size of the social network of these children varies with the remote teaching mandate in high schools, and confirmed infection rates of the parents. This allows us to assess the validity of our empirical design, and address some of the remaining empirical concerns discussed above. The following sections present further descriptive results to motivate our choice of predictor of infections in the social networks and then present the main results and robustness checks.

Finally, that children can get infected by and transmit SARS-CoV-2 is by now well established. To this date, there is, however, a debate about the *extent* to which children become infected with SARS-CoV-2 relative to adults, the symptoms they present once infected, and to what *extent* they transmit the virus to adults. For much of the early stages of the pandemic in Sweden, the official stance was reminiscent of the conclusions in an early literature review by Ludvigsson (2020) that concluded that "children are unlikely to be the driver of the pandemic", and that opening up schools was "[…] unlikely to impact covid-19 mortality rates". As noted by Ludvigsson, the majority of studies in that review was based on Chinese data, where lockdowns were strict, and therefore arguably difficult to generalize to the Swedish setting. While it is out of scope of the current report to assess whether children were in fact driving the pandemic or not, it is important for the interpretation of the results in the main analysis to first confirm that the correlation in shown Figure 1 is not simply due to factors correlated with both child age and confirmed infections among their parents. Before reporting the findings from the main analysis, we take a first important step in this direction in the next section.

5. A descriptive analysis of determinants of positive SARS-CoV-2 tests during 2020

We start with a descriptive analysis of the observable individual factors that contributed to a positive SARS-CoV-2 test during 2020. The workers and relatives of the patients in the nursing homes constitute the focal point of our subsequent analysis. However, for generalizability and comparability, we start by examining a set of individual predictors for the full sample of working age individuals (25-65) residing in the Stockholm region, the epicenter of the first wave.

Table 1 column (1) provides estimates from the baseline multivariate regressions. Table 1 column (1) shows the correlations between testing positive and the presence of children in the household, immigration status, educational attainments (years of schooling), wage income in 2019, and indicator variables for whether the individual had zero income in 2019 and whether the subject is male, and married. On top of these explanatory variables, the baseline model also accounts for year of birth, neighborhood of residence and occupation (3 digit) fixed effects. In other words, the models estimated account for systematically different infection rates and the testing capacity/willingness across individuals residing in different neighborhoods, occupations, and ages.

Columns (2) and (3) use the infection rates during the spring and fall as the outcome variable instead of the whole of 2020. Column (4) restricts the full sample to those working in workplaces with between 20-100 employees, for an even closer comparability between this full Stockholm sample and nursing home workers. Column (5) and (6) restrict the sample to the workers employed in the nursing homes. Columns (7) and (8) restrict the sample to the close relatives (children) of the nursing home residents. Columns (6) and (8) include nursing home fixed effects. That is, columns (6) and (8) compare how the observables are correlated with a confirmed test for employees (relatives) who work (have parents), conditional on working in/having relatives in a specific nursing home, on top of accounting for e.g. systematic differences across individuals residing in different neighborhoods.

Overall the table provides a relatively consistent and clear pattern: variables that are associated with having a larger social network are consistently related to a higher risk of testing positive for SARS-CoV-2 during 2020. Conditional on the neighborhood, occupation, age-specific factors as well as the other controls, being married, having a job, and having school age children at home increase the probability of having tested positive for SARS-CoV-2 during 2020.[6] These potential network size related factors stand out as individual factors that potentially contribute to an increased risk of contracting and transmitting SARS-CoV-2 into the elderly homes. Due to the policy relevant nature of the child related connections, we focus our analysis on this particular network and now proceed with the main findings of this report.

[6] Other interesting patterns emerge from this simple initial analysis. For example, immigration status or being male seems to have no correlation with testing positive in Stockholm during 2020. However, as seen from columns (2) and (3), this seemingly null effect stems from a *positive* and significant correlation during the first wave and a *negative* correlation with these observables during the second wave. One possible explanation for this pattern could be that higher exposure during the spring resulted in higher immunity and hence less infections during the fall in these subgroups. Irrespective of cause, this finding highlights the problems of using observed infections in the fall (when the testing capacity improved) to predict the infections rates during the first wave.

6. The link between the potential network size of employees and relatives as proxied by the presence and age of children in the household and mortality in the nursing homes

6.1 The Stockholm Sample

Building on the above results, we start by proxying for the potential network size of the employees and relatives using the presence and age of the children in the households of the employees and relatives. A key concern with the initial descriptive analysis presented in Table 1 is that e.g. families with older children could, for example, be more informed about the risks of SARS-CoV-2 infections and therefore potentially be more willing to take a test. As discussed above, our main analysis sidesteps that concern by testing the link between a larger potential network associated with the social connections of the children in their daycare and school environment and mortality in the nursing homes.

The unit of analysis sample in the first part of this analysis consists of individuals that are residing in nursing homes in Stockholm municipality, at the end of 2019, who have not died before March 1, 2020. The outcome variable is, to start with, whether the individual died during the first wave (March-June), or not. Column (1) of Table 2 shows that individuals living in a nursing home with a larger share of employees with children at home are associated with a higher risk of dying during the first wave, but that the share of close relatives with children at home is not linked to mortality in the nursing homes. All explanatory variables reported in this and the following tables are standardized (mean 0, std.dev. 1), implying that a standard deviation increase in the share of employees with children at home is associated with a 2.3 percentage point, or 10% (relative to the mean mortality rate during the spring) increase in the risk of dying from any cause in the nursing homes in Stockholm. This result is the first to suggest that our crude measure of the size of the potential social network of employees is significantly more related to mortality in the elderly homes than the size of the social network of relatives; a result that holds throughout the rest of the report.

Column (2) adds individual elderly controls (age and month of birth, gender, immigrant status, married, and the number of health risk factor fixed effects), worker/workplace controls (wages, immigration status, education, married, age, male, number of co-workers, somatic/dementia nursing home), and relatives' controls (wages, immigration status, education, married, age, male). Adding these controls does not change the pattern reported in column (1).

Column (3) splits the children of the worker and the relative into two different groups: (i) those directly affected by the reductions in potential network size due to the mandate to move to remote-teaching (high school students), vs (ii) those children not directly affected by such restrictions (pre- and primary schools, and post-secondary school age children). This specification suggests that during the first wave, the overall impact of connections to the workers' children is driven by children not facing restrictions in their potential number of social connections. Relative to a larger share of workers without children at home, a larger share of workers with children directly affected by remote-teaching mandates is associated with a significant *decrease* in the mortality in the elderly homes, while a higher share of employees with children facing no restrictions on social interactions is associated with a significant *increase* in mortality in the nursing homes.[7] Again, no similar pattern is found for the relatives' school related potential network size.

Columns (4) and (5) provide results from a specification where we take this one step further by splitting the first-wave sample period into a pre-remote teaching period and post-period. The size of the potential social network of parents of high school children was affected by the remote teaching restriction from March 18 and onward. The potential network size reduction induced by the remote-teaching reform did likely affect the mortality rates in the nursing home with a lag. Assuming an incubation period of up to 14 days to reduce infections in children, and an additional up to 14 days to reduce infections in their parents, 14 days for reductions in the elderly, and finally an additional 7 days or so before SARS-CoV-2 related deaths started to be affected in the nursing homes (the elderly were seldom moved to intensive care units). Therefore, we split the spring period into two periods March-April (in school-period affected period), and May-June (remote teaching affected period) and assess how the impact of having a larger share of employees with high school children affected the mortality rates in the nursing homes. Columns (4) and (5) show a distinctly different pattern as compared to the baseline full spring model and indicate that the high school closure induced reduction in the potential number of social connections coincides with a significant reduction in the number of deaths in the nursing homes due to employee connections, but not in the period before the regulations had the time to influence the mortality in the nursing homes.

Column (6) shows that restricting the outcome variable to those where SARS-CoV-2 is indicated as the primary cause or contributing cause of death provides similar, albeit less precise, estimates. Given that relatively few of the nursing home residents ended up in intensive care units, the reduced precision is likely due to a measurement error in the outcome due to misdiagnosed causes of death.

---

[7] Note that the (omitted) reference group is the workers in the nursing homes without children at home. A relevant question is why the share workers with high school age children have a negative sign compared to the references group. Note that workers without children are likely systematically different in many important aspects in comparison with employees with children. For example, employees without children on average live in more densely populated residential blocks (lägenheter) than those with children (radhus), and if population density is correlated with virus transmission this could explain the negative coefficient. In general any differences in (risk) behavior across households with and without children will yield a similar pattern when holding potential networks size fixed. The more relevant comparison is therefor the differences between employees with children in different age groups as shown in the tables.

A relevant concern when using the share of children as a proxy for potential network size is that it could reflect other factors than just social connections that are also related to a higher mortality rate in the nursing homes. While the results for restricted and unrestricted children connections address much of such concerns, it could still be the case that parents of high school children in general work in on average systematically different elderly homes than children of younger and older children. To address this concern, column (7) shows that a larger share of employees with children is not associated with higher mortality rates before the pandemic (in the spring of 2019), thus suggesting that the observed pattern during the pandemic is not due to unobserved factors associated with both higher mortality and larger shares of employees with children at home in general.

Finally, the employees' control variables reported at the bottom of Table 2 are interesting for a comparison with the main parameters of interest. We find some indications that a higher average income of the employees is negatively related to mortality during the first wave, that a higher average educational attainment of the employees reduces the mortality rates in the nursing homes during the second half of the spring, and that a higher share of immigrants is not systematically correlated with mortality in Stockholm . However, the significance and the direction of the impact of these variables on nursing home mortality rates are not particularly stable across the specifications.

6.2 The National Sample

We now turn to the full Sweden sample. Columns (1) to (3) in Table 3 reiterate the baseline specifications from Table 2. As in the Stockholm sample, a larger share of employees with children in "unrestricted social interaction" ages is statistically significantly associated with a higher mortality rate relative to having a larger share of employees without children at home, and statistically significantly different from the effect of a larger share of employees with children in "restricted social interaction" age groups. In the full Sweden sample, the effect on COVID-19 mortality is now also clearly significant. Once more, we find no similar pattern for the connections via relatives.

Columns (4-6) compare whether the proxies for workers' social network connections in Stockholm, the epicenter of the first wave, are significantly different for the same connections of workers in other areas of Sweden. The idea behind these specifications is to test whether the potential network size of workers in high infection areas (Stockholm) is significantly different from the same sort of connections in low (or at least lower) infection areas. The simple test of the link shown in columns (4-6) implicitly assumes no infection links in the other parts of Sweden during the first wave. Despite this strong assumption, which will tend to generate estimates biased towards zero for Stockholm, the overall pattern holds. Relative to the other regions, the link between the school related potential network size of the employees and mortality rates is positive in the high exposure region, but again only for unrestricted children and only during the latter part of the first wave. For restricted interaction age children, the sign of the coefficient is negative and large, but not statistically different from having a higher share of employees without children. Overall, the results from the Stockholm municipality analysis and the nationwide analysis yield a similar pattern.

It may also be of interest to compare the main parameters of interest with the three other measures of employee characteristics also reported in the Stockholm sample for the national sample. Due to space constraints, they are omitted from Table 3, but in contrast to the Stockholm sample, the share of employees with immigrant background, and the average income of the employees are both in general positively and significantly associated with mortality in the elderly homes in the full Sweden sample, while the average years of schooling of the employees are typically is still associated with a reduced mortality rate.

6.3 Confirmed Infections in Social Networks and nursing home mortality during the first wave

Testing whether the potential network size/connections are correlated with mortality rates in the nursing homes is motivated by the measurement problems of SARS-CoV-2 infection, and the empirical challenges of relating measured infection rates in social networks to infections in the nursing homes as discussed in section 2. In this section, we take the main analysis one step further, keeping those challenges in mind. In Table 4 we extend the baseline analysis and now test how confirmed infections in the working age population in the residential neighborhoods of the nursing home employees and relatives are linked to SARS-CoV-2 infections in the nursing homes in the national sample.

We start our analysis in Table 4 column (1) by assessing the strength of the link between confirmed infections of the workers and relatives with confirmed infections in the nursing homes using individual data on the nursing home residents. Column (1) shows that confirmed infections among workers, but not relatives, are strongly linked to confirmed infections in the nursing homes. In unreported aggregated nursing home level regressions, a standard deviation increase in the share of employees testing positive is associated with a .38 standard deviation increase in confirmed infections in the nursing homes. Column (2) shows that the share of infected workers is also strongly positively correlated with the mortality rates in the nursing homes. These estimates suggest that a standard deviation increase in the confirmed infections of the employees is associated with a 22 percent increase in mortality among the nursing home residents.

While expected, these correlations could in part be due to other unobserved factors linked both to a higher mortality rate and a higher share of positive tests among the workers. Column (3) shows that the share of workers testing positive during the first wave is linked to higher mortality rates already in January and February. Given that community infections were not confirmed until early March 2020, this finding suggests that the nursing homes with a large share of positive tests among their workers later in the spring are systematically different in terms of mortality than those with a lower share of workers testing positive, already *before* the first wave. For comparison, column (4) tests the same relationship using our baseline proxies for potential network size. Unlike confirmed infections, the share of workers with children in the households is not linked to mortality *before* the pandemic. In other words, the nursing homes with a larger share of employees with children in their households are not systematically different in terms of mortality risk than nursing homes with a low share before the pandemic, thus providing further support for the validity of our main empirical approach.

With these methodological issues in mind, we now turn to the link between confirmed infections among the working age population in the neighborhoods (as defined by Statistics Sweden's "distrikt") where the worker/relative lives and mortality in the nursing homes. Column (5) shows a strongly positive association between infections among neighbors of workers, but not relatives, and mortality in the nursing homes during the first wave. Once we control for the workers' own infections, in column (6), the relationship between confirmed infections in the neighborhood of residence and mortality in the nursing homes vanishes. This finding indicates the key role of the worker as a mediator between the outside world and the nursing home residents, relative to close relatives. However, based on these findings alone we cannot rule out that the difference between relatives and workers is at least in part due to a greater degree of measurement error of infection rates in the relative network, or as indicated by the findings in column (3), that the relationship is caused by other unobserved factors, or even that e.g. non-confirmed infections among the relatives actually started or at least contributed to the confirmed infections among both employees and the nursing home residents.

That unconfirmed infections among the employees can have contributed to the mortality rates, above and beyond confirmed infections, is suggested by the findings in Columns (6), (7) and (8). These final columns of Table 4 show that the correlation between mortality and the school related potential networks size survives when we include the controls for the workers' own confirmed infections, shaving off only around 20% of the estimated link. Again, it is the connections to the children that are not directly affected by the reduction in the potential network size ("unrestricted interaction age children") by remote-teaching mandates that are driving these results. Under the maintained assumption that a larger potential network size is the main reason for the link between differences in the association across children of different ages, these final results indicate that asymptomatic, or at least unconfirmed, infections could have played a role in the link between the employees' network and mortality in the nursing homes. Note also that the neighborhood infection rates are calculated for the working age population, whereas the share of children in the households is primarily intended to capture the link to the school related networks. Ideally, we would have liked to measure the risk of infections transmitted from these, at least partly, separate networks using the same metric (randomly tested adults and children). Absent such data, it is difficult to pin down the exact causes of the different patterns for the two social network measures with certainty, once we control for the confirmed infection rates of the employees.

6. Conclusions

This report seeks to shed some light on the relative importance of employees' and close relatives' social networks in explaining the transmission of SARS-CoV-2 into nursing homes in Sweden during the first wave of the COIVD 19 pandemic in Sweden using administrative data. Due to the lack of reliable and comparable measures of infection rates in the two social networks during the first wave, the main analysis focuses on measuring the effects of the relative impact on mortality in the nursing homes using information on the potential network size as proxied by having children in their households (and, by extension, the children's age dependent potential school-related networks). The main findings in the report are consistent with employee connections to a larger social network via their children being a much more important explanation for the transmission of the SARS-CoV-2 virus into the nursing homes, than the same sort of connections of relatives. This result holds in national data as well as in data from Stockholm municipality, the epicenter of the first wave. Despite limited and selective testing during the first wave, the same result emerges when instead using data on confirmed infections in the neighborhood of residence of the employees and relatives. These two different and complementary ways of measuring the role of employees and close relatives' social networks provide a consistent picture. The social network of the nursing home employees seems to have played a large and significantly more important role in explaining the mortality rates in the vulnerable nursing home population than SARS-CoV-2 transmissions through the social networks of close relatives.

References

Manski, Charles (1993). “Identification of Endogenous Social Effects: The Reflection Problem,” *Review of Economic Studies*, LX, 531-542.

Ludvigsson, J. “Children are unlikely to be the main drivers of the COVID-19 pandemic – A systematic review”, Volume109, Issue8, August 2020 Pages 1525-1530.

Stern, Charlotta, and Daniel B Klein. “Stockholm City's Elderly Care and Covid19: Interview with Barbro Karlsson.” *Society*, 1-12. 19 Jul. 2020, doi:10.1007/s12115-020-00508-0.

Skolverket 2019a, Database ”https://www.skolverket.se/skolutveckling/statistik/sok-statistik-om-forskola-skola-och-vuxenutbildning?sok=SokD&niva=S&omr=elever&exp=136&uttag=null&verkform=21&lasar=2019”.

Skolverket 2019b, Database ”https://www.skolverket.se/skolutveckling/statistik/sok-statistik-om-forskola-skola-och-vuxenutbildning?sok=SokD&niva=S&omr=elever&exp=6&lasar=2019&uttag=null”.

Skolverket 2020a, ”Redovisning av uppdrag om gruppstorlekar vid fjärrundervisning”, Dnr U2020/04688/GV)”.

Skolverket 2020b, ”Barn och personal i förskola 2020”, Dnr 2021:435.

Sveriges officiella Statistik Tabell 1: Förskole- och skolenheter/fritidshem. Barn/elever läsåret 2019/20).

Szebehely, M. (2020), “Internationella erfarenheter av covid-19 i äldreboenden”, Underlagsrapport till SOU 2020:80 *Äldreomsorgen under pandemin*

Vlachos J., E. Hertegård, H. B. Svaleryd (2021), ”The effects of school closures on SARS-CoV-2 among parents and teachers”, *Proceedings of the National Academy of Sciences* Mar 2021, 118 (9) e2020834118; DOI: 10.1073/pnas.2020834118.

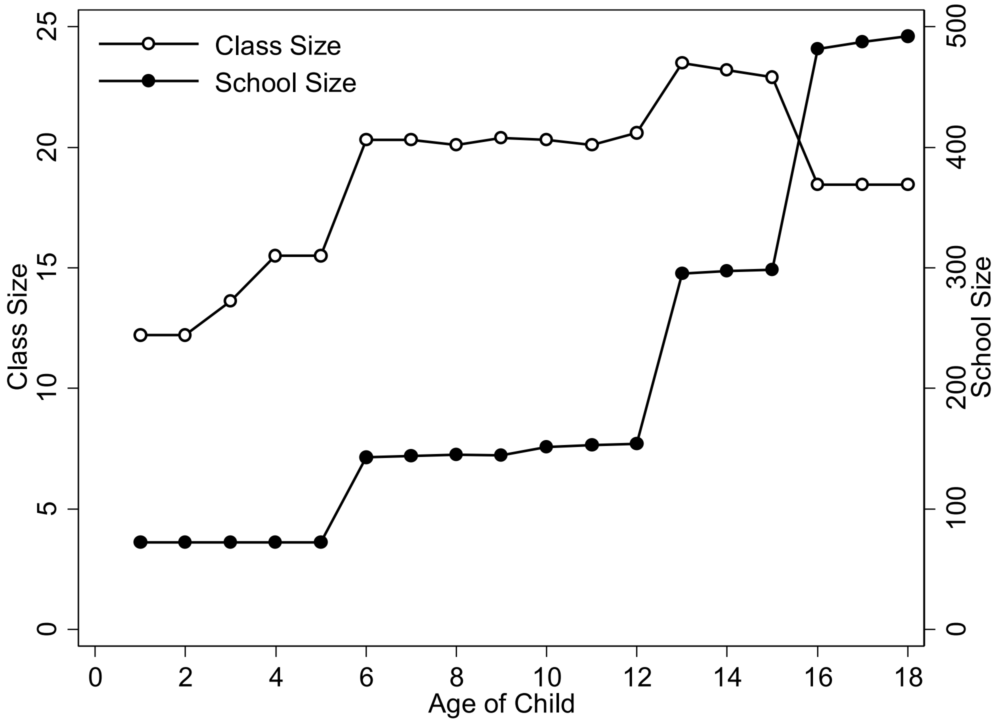

Figure 1: Potential School Network Size of the Child

Note: This figure shows the average school network size of the child by age as proxied by the number of children in the same class and school. The figure shows the national raw average data for average class-sizes by age groups, and the population weighted average school size. The average school size is computed for each age group and uses the number of children in the same school and "stadium" i.e. (childcare unit, lågstadium, mellanstadium, högstadium, and high school) and is weighted by the number of children in the school-by-grade cohort – thereby reflecting the average school size that the average child in each age group experiences. Sources. Skolverket, 2019a; Skolverket 2019b; Skolverket 2020a, Redovisning av uppdrag om gruppstorlekar vid fjärrundervisning Dnr U2020/04688/GV) (Förskola Sveriges officiella Statistik, Tabell 2 A: Antal inskrivna barn, förskoleenheter och antal barn per barngrupp 2009-2019) (Grundskola Skolverket 2020, Redovisning av uppdrag om gruppstorlekar vid fjärrundervisning Dnr U2020/04688/GV). Daycare centers are typically organized into groups based on ages 1-3 and 3-5. The average group/class size for the younger children is 12.2 and for the older children 15. We use the share of daycare centers with mixed groups to assign the share of three-year olds into either the older or the younger group using data in Skolverket (2020b).

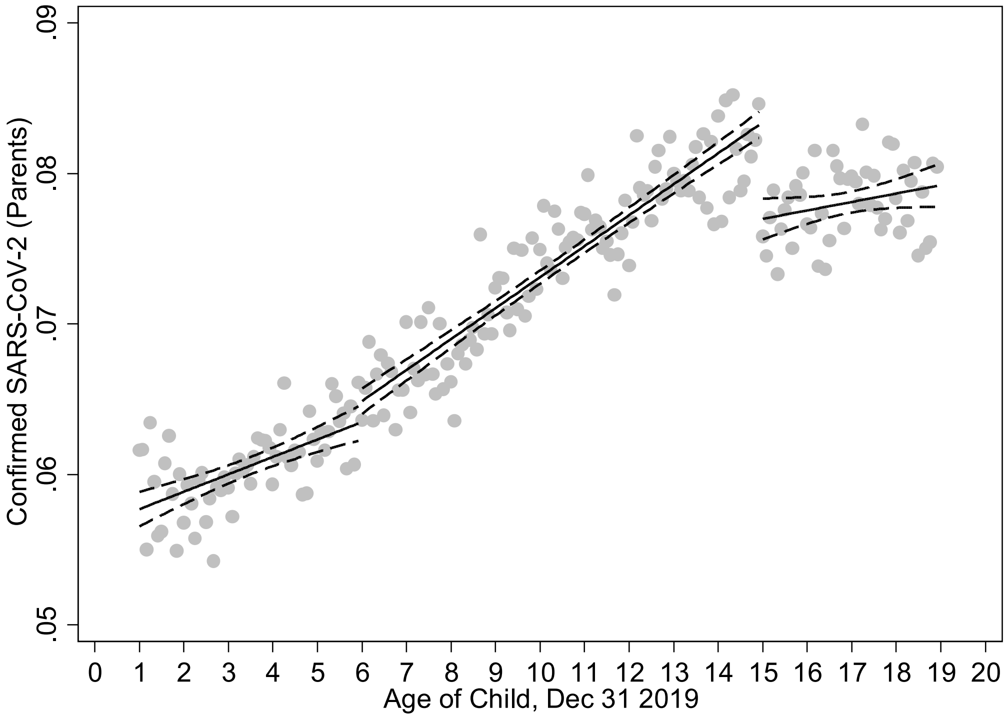

Figure 2 Age of child and Parents' Confirmed Sars-CoV-2 infection during 2020

Note. This figure plots the relationship between the share of confirmed infections of parents and the age of their child. We exclude parents who do not have a child residing in their home. 95% confidence intervals in dashed. The scatter plots show the share of parents with a positive test against the age of their child in months at the end of 2019. The linear regressions have breaks at the time of compulsory school and high school starting ages. On average, 95% of children attend preschools, typically starting in the year when they turn 2.

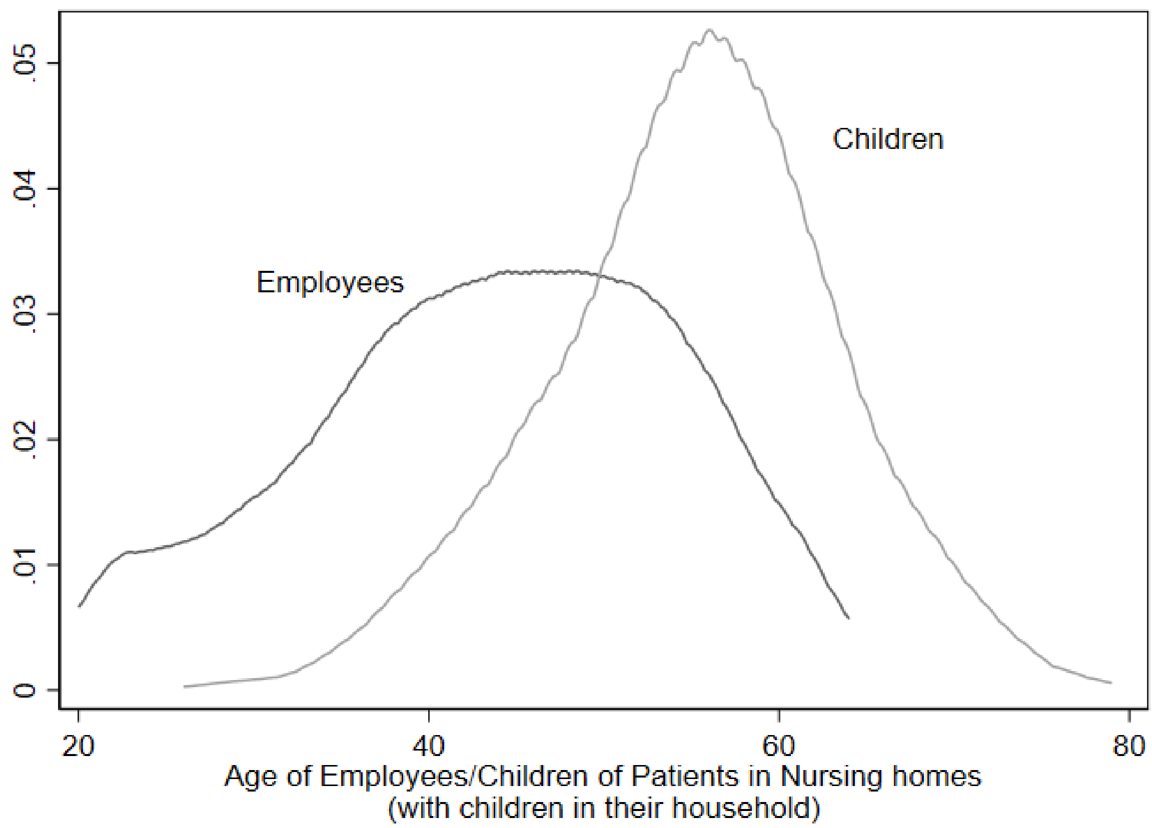

Figure 3 Age distribution of employees and close relatives (children) of the residents in the nursing homes in Stockholm

Table 1: Baseline correlates for confirmed SARS-CoV-2 infections in Stockholm region during 2020 for ages 20-65.

	(I)	(I)	(I)	(II)	(III)	(IV)	(V)	(VI)
Sample:	Full Stockholm (Full 2020)	Full Stockholm (Spring 2020)	Full Stockholm (Fall 2020)	Employees (Workpl.w. 20-100 emp.)	Workers in Elderly Care Homes	Workers in Elderly Care Homes	Children to Patients in Eldery Care	Children to Patients in Eldery Care
1.(Pre-school Age Kid at home)	-.0081***	-.0018***	-.0062***	-.0106***	-.0114	-.0057	.0068	.00002
	(.0016)	(.0004)	(.0015)	(.0022)	(.0166)	(.0162)	(.0219)	(.0223)
1.(Primary School Age kid at home)	.0083***	.0006	.0078***	.0073***	.0397***	.0398***	.0232	.0271
	(.0014)	(.0005)	(.0013)	(.0022)	(.0143)	(.0146)	(.0211)	(.0219)
1.(Secondary School Age)	.0086***	.0016***	.0069***	.0089***	.0244	.0263*	.0539***	.0566***
	(.0013)	(.0005)	(.0011)	(.0022)	(.0186)	(.0143)	(.0200)	(.0207)
1.(Post-Secondary School Age kid)	-.0030**	-.0001	-.0029**	-.0006	-.0021	-.0037	.0191	0206
	(.0014)	(.0004)	(.0014)	(.0020)	(.0191)	(.0156)	(.0159)	(.0165)
1.Immigrant	-.0003	.0034***	-.0037***	.0012	.0073	.0075	.0019	.0019
	(.0009)	(.0005)	(.0009)	(.0013)	(.0096)	(.0091)	(.0107)	(.0103)
Years of Education	-.0005***	-.0004***	-.0001	-.0002	.0001	-.00004	-.0033***	-.0036***
	(.0001)	(.0001)	(.0001)	(.0003)	(.0017)	(.0019)	(.0015)	(.0015)
1.Male	.0010	.0019***	-.0009	.0009	-.0015	-.0022	.0067	.0055
	(.0006)	(.0003)	(.0006)	(.0012)	(.0102)	(.0102)	(.0093)	(.0096)
Wage income	.0005***	.0001*	.0004	.0004**	.0119	.0118	.0018***	.0017***
	(.0001)	(.00004)	(.0001)	(.0002)	(.0038)	(.0030)	(.0003)	(.0003)
1.No-income	-.0188***	-.0012***	-.0175***				-.0062	-.0081
	(.0010)	(.0004)	(.0008)				(.0084)	(.0093)
1.Married	.0069***	.0006*	.0064***	.0090***	.0173**	.0169**	.0044	.0029
	(.0008)	(.0003)	(.0007)	(.0013)	(.0072)	(.0079)	(.0074)	(.0078)
Total number of Children	.0058***	.0013***	.0046***	.0059***	-.0034	-.0049	-.0081	-.0103
	(.0009)	(.0002)	(.0008)	(.0012)	(.0092)	(.0070)	(.0098)	(.0103)
Year of Birth Fixed Effects	yes	yes	yes	yes	yes	yes	yes	yes
Neighborhood Fixed Effects	yes	yes	yes	yes	yes	yes	yes	yes
Occupation (3-digit) Fixed Effects	yes	yes	yes	yes	yes	yes	yes	yes
Eldery Home Fixed Effects						yes		yes
Sample sizes	1 425 304	1 425 304	1 425 304	333 055	10 599	10 599	4 528	4 528
Mean Confirmed SARS-CoV-2 Infection	0.067	0.01	0.057	0.076	0.165	0.165	0.054	0.054

Note: The full model also includes indicator variables for missing values on years of schooling, which is replaced by the average years of schooling in the municipality and the country of origin cell. Standard errors in parenthesis allowing for an arbitrary correlation within the neighborhood of residence (”distrikt”) 139. The occupation codes used are at the 3-digit level. */**/*** indicates statistical significance at the 10/5/1% levels, respectively. The nursing homes include those organized by Stockholm municipality.

Table 2 Social Networks and Patient Mortality in Nursing Homes in Stockholm – Potential Network Size Measured by Presence of Workers and Relatives' Children at Home

Unit of observation:	*Elderly*	*Elderly*	*Elderly*	*Elderly*	*Elderly*	*Elderly*	*Elderly*
Specification:	(1)	(2)	(3)	(4)	(5)	(6)	(6)
Outcome variable	Mortality Spring 2020	Mortality Spring 2020	Mortality Spring 2020	Spring Bef. Remote teaching	Spring After Remote teaching	Covid Diagnosed Mortality Spring 2020	Dead Spring **2019**
Workers:							
Share with Kids at Home (*All age groups*)	.023**	.0214**					
	(.0099)	(.0087)					
Share with Kids (*"Unrestricted Interaction" ages*)			.0243***	.0121	.0162**	.0103	.0016
			(.0090)	(.0088)	(.0052)	(.0079)	(.0119)
Share with Kids (*"Restricted Interaction" ages*)			-.0199**	-.0061	-.0170**	-.0120*	-.0035
			(.0091)	(.0086)	(.0074)	(.0068)	(.0083)
Close Relatives:							
Share with Kids at Home (*All age groups*)	-.0066	.0096					
	(.0088)	(.0131)					
Share with Kids (*"Unrestricted Interaction" ages*)			-.0016	.0019	-.0043	-.0116	.0150
			(.0127)	(.0111)	(.0079)	(.0112)	(.0102)
Share with Kids (*"Restricted Interaction" ages*)			.0153	.0219**	-.0053	.0057	-.0025
			(.0093)	(.0097)	(.0064)	(.0085)	(.0105)
Additional Controls:	*None*	*Ind+Work +Relatives.*	*Ind+Work +Relatives.*	*Ind+Work +Relatives.*	*Ind+Work +Relatives.*	*Ind+Work +Relatives.*	*Ind+Work +Relatives.*
Share of Immigrants		.0039	-.0073	.0095	-.0187**	-.0004	-.0043
		(.0109)	(.0120)	(.0104)	(.0078)	(.0104)	(.0113)
Average Income		-.0199**	-.0153	-.0139	-.0045	-.0219**	-.0139
		(.0099)	(.0095)	(.0126)	(.0103)	(.0099)	(.0099)
Average Education		-.0039	-.0057	.0103	-.0168**	.0004	-.0002
		(.0104)	(.0102)	(.0096)	(.0075)	(.0093)	(.0148)
Sample Size (#Elderly homes)	3,870(109)	3,870(109)	3,870(109)	3,870(109)	3,289 (107)	3,870(109)	3,523 (93)

Note: Unrestricted interaction ages are defined as children 0-15 and 20+ residing at the home of the parent. Restricted interaction are ages 15-19 residing at home. Additional individual controls not show in table: male, immigrant, medical risk factors, 5 year age bins, month of birth – indicator variables. Additional controls for nursing home employees' characteristics not shown in the table: (averages) : number of employees, married, male, income and age of the employees. All models are estimated with OLS, standard errors (in parenthesis) are clustered at the nursing home level. */**/*** denotes significance at the 10/5/1% levels Mean mortality Full 22%, Early 15%, Late 8.3%, Covid 10%;2019 11%;

Table 3 Additional Results Mortality in the Elderly Homes (Full sample Sweden)

Unit of observation:	*Elderly*	*Elderly*	*Elderly*	*Elderly*	*Elderly*	*Elderly*
Specification:	(1)	(2)	(3)	(4)	(5)	(6)
Period:	Spring 2020	Spring 2020	Spring 2020 COVID mortality	Spring 2020	Spring Before Remote teaching	Spring After Remote teaching
Workers in Elderly Homes						
Share w. Kids at Home (*All age groups*)	.0050**					
	(.0022)					
Share w. Kids *("Unrest. Interaction" ages*)		.0053**	.0043**			
		(.0022)	(.0015)			
Share w. Kids (*"Restricted Interaction" ages*)		-.0020	.0019			
		(.0024)	(.0016)			
Close Relatives						
Share w. Kids at Home (*All age groups*)	-.0009					
	(.0022)					
Share w. Kids *("Unrest. Interaction" ages*)		-.0011	-.0009			
		(.0022)	(.0014)			
Share w. Kids (*"Restricted Interaction" ages*)		-.0019	-.0002			
		(.0019)	(.0012)			
Workers in Elderly Homes outside Stockholm						
Share w. Kids *("Unrest. Interaction" ages*)				.0100	.0021	.0005
				(.0066)	(.0018)	(.0017)
Share with Kids (*"Restricted Interaction" ages*)				-.0014	-.0013	-.0002
				(.0025)	(.0019)	(.0017)
Workers in STHLM (relative to other regions)						
Share w. Kids (*"Unrest. Interaction" ages*)				.0025	.0033	.0082*
				(.0024)	(.0056)	(.0044)
Share w. Kids (*"Restricted Interaction" ages*)				-.0019	.0049	-.0074
				(.008)	(.0067)	(.0048)
Sample Size (#Elderly homes)	49,806 (3,689)	49,806 (3,689)	49,806 (3,689)	49,806 (3,689)	49,806	45,921
Mean outcome variable	.137	.136	.031	.137	.078	.064

Note: Unrestricted interaction ages are defined as children 0-15 and 20+ residing at the home of the parent. Restricted interaction are ages 15-19 residing at home. Additional individual controls not shown in the table: male, immigrant, medical risk factors, 5 year age bins, month of birth, and region of residence – indicator variables. Additional controls for elderly home employees' characteristics not shown in the table: (averages) : number of employees, married, male, income and age of the employees. All models are estimated with OLS, standard errors (in parenthesis) and are clustered at the elderly home level. */**/*** denotes significance at the 10/5/1% levels

Table 4 Exploring Contagion Networks in More Detail

Unit of observation:	*Patients*	*Patients*	*Patients*	*Patients*	*Patients*	*Patients*	*Patients*	*Patients*	*Patients*
Specification:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Period:	Infections Spring 2020	Mortality Spring 2020	Mortality Jan-Feb 2020 (Before Covid)	Mortality Jan-Feb 2020 (Before Covid)	Mortality Spring 2020	Mortality Spring 2020	Mortality Spring 2020	Mortality Spring 2020	SARS-CoV-2 Mortality Spring 2020
Share Employees Conf. Infected Spring	.0646***	.0281***	.0024**			.0284***	.0280***	.0284***	.0273***
	(.0027)	(.0024)	(.0011)			(.0025)	(.0024)	(.0025)	(.0019)
Share Relatives Conf. Infected Spring	.0035*	.0009	-.0009			.0011	.0009	.0011	.0031**
	(.0018)	(.0019)	(.0009)			(.0019)	(.0019)	(.0019)	(.0012)
Share Employees w. Kids (*Any age*)				.0005					
				(.0013)					
Share Relatives w. Kids (*Any age*)				-.0009					
				(.0013)					
Social Networks									
Worker Neighbors Conf. Infections					.0109**	.0006		.0008	-.0008
					(.0044)	(.0043)		(.0043)	(.0022)
Relatives Neighbors Conf. Infections					-.0018	-.0011		-.0013	.0006
					(.0040)	(.0039)		(.0039)	(.0021)
Share Emp. w. Kids ("Unrestricted ages")							.0043*	.0044**	.0033**
							(.0022)	(.0022)	(.0013)
Share Emp. w. Kids ("Restricted ages")							-.0027	-.0030	.0012
							(.0023)	(.0023)	(.0014)
Sample Size	49,806	49,806	52,949	52,949	49,589	49,589	49,589	49,589	49,589
Mean of Outcome Variable	.068	.137	.059	.059	.136	.136	.136	.136	.031

Note: See Table 3.

Appendix A: Datatillgång

Ett forskningsprogram om covid-19 i Sverige: Smittspridning, bekämpning och effekter på individer och samhälle

Kommissionen bedömer att det krävs genuint ny kunskap för att på ett tillfredställande sätt uppfylla Kommissionens direktiv. Mot bakgrund av detta har kommissionen initierat *Ett forskningsprogram om covid-19 i Sverige: Smittspridning, bekämpning och effekter på individer och samhälle.* Programmet har sin hemvist på Stockholms Universitet (SU). Det leds av Torsten Persson (som också är ledamot i kommissionen) och koordineras av Adam Altmejd, Evelina Björkegren samt Olof Östergren (som också är anställda på deltid i kommissionens sekretariat).

Programmet har två övergripande frågeställningar:

1. Vilka är konsekvenserna av covid-19 – samt av de smittskyddsåtgärder myndigheterna satt in för att hantera utbrottet – för folkhälsan, i termer av mortalitet, morbiditet, samt psykisk och fysisk ohälsa i bred bemärkelse?
2. Vilka är konsekvenserna av pandemin – samt av åtgärderna för smittskydd och för att mildra skadeverkningarna på samhället – för centrala sociala och ekonomiska utfall, som till exempel jobb, inkomster och ojämlikhet?

Resultaten från SU:s Covid-19 program kommer att utgöra ett oberoende vetenskapligt underlag och syftar till att komplettera – men även bekräfta och utvärdera – existerande rapporter från myndigheter, samt andra forskare och aktörer.

Programmet har en kvantitativ ansats och bygger på ett omfattande datamaterial som huvudsakligen består av registeruppgifter om enskilda individer, företag samt vårdinrättningar från ett tjugotal leverantörer, inklusive Statistiska Centralbyrån, Socialstyrelsen, Folkhälsomyndigheten, Försäkringskassan, Skatteverket samt Inera/1177. Uppgifter som kan identifiera specifika individer och företag, till exempel namn, personnummer eller organisationsnummer, ingår inte i materialet. Programmet har sökt och fått etiskt tillstånd från Etikprövningsmyndigheten (Dnr 2020-06492 och Dnr 2021-01115) för att hantera de ingående registeruppgifterna för att svara på ett antal specifika frågeställningar.

För närvarande deltar ett trettiotal forskare i programmet. Dessa får, enligt etablerad praxis i svensk registerforskning, tillgång till relevanta delar av datamaterialet via säkra anslutningar till SCB:s fjärrsystem Microdata online access (MONA). Forskningsprogrammet kommer att fortsätta behandla frågor kring pandemin under ett antal år efter det att Kommissionen har levererat sitt slutbetänkande.

Relationen mellan SU:s Covid-19 program och Kommissionen är specificerad i ovannämnda etiktillstånd. Forskarna i programmet samarbetar med Kommissionen på två sätt. För det första erbjuder programmet tillgång till registermaterialet åt ett antal forskare som Kommissionen också har anlitat för att skriva underlagsrapporter. Dessa forskare har enbart tillgång till de uppgifter som är nödvändiga för att besvara de frågor som skall behandlas i respektive underlagsrapport.

För det andra producerar SU:s Covid-19 program beskrivande statistik och analyser till grund för Kommissionens betänkanden (samt för ett par underlagsrapporter där författarna inte har direkt tillgång till registermaterialet). Programmet levererar sålunda färdiga tabeller eller figurer, för publikation i Kommissionens betänkanden eller underlagsrapporter. Statistikens specifika innehåll bestäms i samråd mellan sekretariatet och programmets koordinatorer. De senare arbetar således deltid som sekreterare på Kommissionen och deltid som forskare på Stockholms universitet. I linje med rådande lagstiftning har Kommissionen (som är en myndighet) inte direkt tillgång till några registeruppgifter, trots att den delvis bekostat beställningen och leveransen av dessa data.

Kommissionens betänkanden innehåller tabeller och figurer som bygger på statistik från en rad olika källor. I de fall en tabell eller figur har tagits fram inom SU-programmet anger källhänvisningen ”Beräkningar inom SU:s Covid-19 program” såväl som den myndighet vars register ligger till grund för beräkningarna, (till exempel SCB eller Socialstyrelsen).