Supplementary appendix for "Sick-leave due to COVID-19 during the first pandemic wave in France, 2020"

David R. M. Smith, Sofía Jijón, Ajmal Oodally, George Shirreff, Karim Aït Bouziad, Pearl Anne Ante-Testard, Jonathan Bastard, Hanifa Bouziri, Oumou Salama Daouda, Tom Duchemin, Anne-Sophie Godon-Rensonnet, Paul Henriot, Yasmine Houri, Hélène Neynaud, Anne Perozziello, Frédérique Thonon, Pascal Crépey, William Dab, Kévin Jean, Laura Temime

a. Probability of symptomatic sick-leave

The daily probability of taking symptomatic sick-leave for a given age group a, region of residence r, and time t was computed as:

$$P_{a,r,t}(\text{symptomatic sick-leave}) = P_{a,r}(\text{employed}) \times P_{a,r,t}(\text{symptomatic COVID-19} \mid \text{employed}) \times P_{a,r}(\text{taking sick-leave} \mid \text{symptomatic COVID-19})$$

For simplicity, the probability of developing symptomatic COVID-19 infection was assumed to be independent of employment status,

$$P_{a,r,t}(\text{symptomatic COVID-19} \mid \text{employed}) = P_{a,r,t}(\text{symptomatic COVID-19})$$

and reflects the daily incidence of symptomatic SARS-CoV-2 infection as estimated in the model published previously by Massonnaud *et al*.(1) This model was fitted retrospectively to SARS-CoV-2 hospitalisation data over time across regions and age groups in France, and age- and region-stratified data corresponding to the first wave of COVID-19 were extracted for use in the present work.

The probability of taking sick leave when employed and infected with symptomatic COVID-19 was assumed to be independent of age and region:

$$P_{a,r}(\text{taking sick leave} \mid \text{symptomatic COVID-19}) = P(\text{taking sick leave} \mid \text{symptomatic COVID-19})$$

Using Bayes' theorem, this latter probability was estimated as:

$$P(\text{taking sick-leave} \mid \text{symptomatic COVID-19}) = P(\text{symptomatic COVID-19} \mid \text{sick-leave}) \times P(\text{taking sick-leave during survey period}) / P(\text{symptomatic COVID-19 during survey period})$$

where $P(\text{symptomatic COVID-19 during survey period})$ was computed from model simulations as the cumulative person-days infected divided by the cumulative person-days in

the population over the period covered by the occupational health survey (1 March to 31 March), and $P(\text{taking sick-leave during survey period})$ was computed from the occupational health survey.(2) Together, this gives the daily probability:

$$\begin{aligned} P_{a,r,t}(\text{symptomatic sick-leave}) \\ &= P_{a,r}(\text{employed}) \times P_{a,r,t}(\text{symptomatic COVID-19}) \\ &\times P(\text{COVID-19 symptoms} \mid \text{sick-leave}) \\ &\times P(\text{taking sick-leave during survey period}) \\ &/ P(\text{symptomatic COVID-19 during survey period}) \end{aligned}$$

Cumulative symptomatic sick-leave incidence among employed individuals over the study period, Inc_{SSL}, was then calculated by multiplying the above probability by the total population size $N_{a,r}$ and summing over all a, r and t,

$$Inc_{SSL} = \sum_{a,r,t} \left(P_{a,r,t}(\text{symptomatic sick-leave}) \times N_{a,r} \right).$$

To calculate the attack rate of symptomatic sick-leave over the study period, the cumulative incidence of symptomatic sick-leave incidence was divided by the size of the working population, both stratified by age group and region.

Daily prevalence of symptomatic sick-leave, $Prev_{SSL}$, was estimated by multiplying the daily probability of symptomatic sick-leave by $D(\text{sick-leave})$, the average sick-leave duration, and $N_{a,r}$, and summing over all a and r,

$$Prev_{SSL} = \sum_{a,r} \left(P_{a,r,t}(\text{symptomatic sick-leave}) \times D(\text{sick-leave}) \times N_{a,r} \right)$$

b. Probability of contact sick-leave

We accounted for daily contact behaviour in our model by incorporating data from a contact survey conducted in France by Bosetti *et al*.(3) Between 10 April and 28 April 2020, survey respondents were asked to report their demographic characteristics and employment status, as well as their contact behaviour from the preceding day across a range of different settings, accounting for the number of contacts and the ages of the individuals they contacted. In the survey, contacts were defined as close contacts within a distance of one metre (e.g. a face-to-face conversation), and were classified as "prolonged contacts" if occurring for a duration >30 minutes.

Based on these data, we defined close contacts of COVID-19 cases ("COVID-19 contacts") as persons who had contact with a symptomatic COVID-19 case and who could potentially be notified of that contact. We only included individuals in the survey who reported active employment during the first wave of COVID-19, and included their contacts that occurred across three different settings K: contact of any duration at home, prolonged contacts in the workplace, and prolonged contacts with relatives outside the home (see Figure S1). Other settings included in the contact survey (in shops, on public transport) were excluded as they were assumed to occur in contexts where subsequent contact notification was unlikely, particularly in early 2020 prior to widespread use of digital mobile contact tracing applications. Contacts (and calculation of contact-related sick-leaves, described below) were considered in individuals' region of primary residence, although we note that 8% of survey respondents were confined to a secondary residence during the survey period, which may have been situated in a different region than their primary residence.

The daily probability $P_{a,r,t}(\text{contact sick-leave})$ of taking contact sick-leave during the first COVID-19 wave, for age group a, region r and time t, was computed as

$$\begin{aligned} P_{a,r,t}(\text{contact sick-leave}) \\ = P_{a,r}(\text{employed}) \times P_{a,r,t}(\text{COVID-19 contact} \mid \text{employed}) \\ \times P_{a,r}(\text{taking sick-leave} \mid \text{COVID-19 contact}) \end{aligned}$$

The daily probability $P_{a,r,t}(\text{COVID-19 contact} \mid \text{employed})$ of an employed person having a COVID-19 contact was estimated for each contact setting K as the mean daily number of contacts in that setting from the contact survey in the region r, in the age group of the employed participant a, the age group of the reported contact b, and working status h, $C^K_{a,r,b,h}$, multiplied by the incidence of symptomatic COVID-19 in the age group of the contacted individual two days prior to the contact taking place, $P_{b,r,t-2}(\text{symptomatic COVID-19})$, and then summed over all contact age groups

$$P^K_{a,r,t,h}(\text{COVID-19 contact} \mid \text{employed}) = \sum_b C^K_{a,r,b,h} \times P_{b,r,t-2}(\text{symptomatic COVID-19})$$

Working status was coded as either $h = 1$ if the participant worked outside the home, or $h = 2$ if they were engaged in remote work. We assumed the same proportion of individuals are infected with COVID-19 in each contact setting K. For the edge cases in $t = \{1,2\}$, symptomatic COVID-19 prevalence two days prior was estimated by adjusting prevalence according to the daily rate of change in $t = \{3,4,5\}$ at the national level.

In many instances there were few or no respondents for a combination of setting, region, participants age, contact age and working status $C^K_{a,r,b,h}$. If there were fewer than 2 respondents, the value was replaced by the expected value based on the marginal totals for participants' age and region.

In addition, it was noted that the proportion of contact survey respondents engaging in remote working (53%) was higher than in a recent Barometer survey of the French population (34%) (4). We therefore corrected this bias by integrating the probability of an infectious contact $P^K_{a,r,t,h}(\text{COVID-19 contact})$ over the national proportion of age group a having working status h according to the Barometer survey $B_a(\text{working status} = h)$,

$$P^K_{a,r,t}(\text{COVID-19 contact} \mid \text{employed}) = \sum_h B_a(\text{working status} = h) \times P^K_{a,r,t,h}(\text{COVID-19 contact} \mid \text{employed})$$

where

$$\sum_h B_a(\text{working status} = h) = 1$$

The probability of taking a contact sick-leave subsequent to contact with someone infected with COVID-19 was assumed to be independent of age and region:

$$P_{a,r}(\text{taking sick leave} \mid \text{COVID-19 contact}) = P(\text{taking sick leave} \mid \text{COVID-19 contact})$$

We assumed that that the probability of taking a contact sick-leave depends on both the probability of being notified of the symptomatic contact, which may depend on the contact's setting, and the probability of then taking sick-leave after being identified as a COVID-19 contact,

$$P^K(\text{taking sick-leave} \mid \text{COVID-19 contact}) = P(\text{taking sick-leave} \mid \text{notified}) \times P^K(\text{notified} \mid \text{COVID-19 contact}).$$

The probability of taking sick-leave after being identified as a COVID-19 contact was calculated directly from the Malakoff Humanis occupational health survey.(2) However, data were not available to inform probabilities of notification for different contact settings. In the baseline analysis, we assumed 90% notification from contacts in the same household, 60% from prolonged contacts during family visits outside the household, and 60% from prolonged workplace contacts. Reflecting uncertainty about notification rates, we performed sensitivity analyses considering rates of 25%, 50%, 75% and 100% across all settings K.

Together, combining these equations gives the daily probability,

$$
\begin{aligned}
&P_{a,r,t}(\text{contact sick-leave}) \\
&\quad = P_{a,r}(\text{employed}) \\
&\quad \times \sum_{K} \sum_{b} \sum_{h} \Big(B_a(\text{working status} = h) \times C^{K}_{a,r,b,h} \\
&\quad \times P_{b,r,t-2}(\text{symptomatic COVID-19}) \times P(\text{taking sick-leave} \mid \text{notified}) \\
&\quad \times P^{K}(\text{notified} \mid \text{COVID-19 contact}) \Big)
\end{aligned}
$$

Cumulative contact sick-leave incidence over the study period, Inc_{CSL}, was then calculated as:

$$Inc_{CSL} = \sum_{a,r,t} \big(P_{a,r,t}(\text{contact sick-leave}) \times N_{a,r} \big)$$

To calculate the attack rate of contact sick-leave over the study period, the cumulative incidence of contact sick-leave incidence was divided by the size of the working population, both stratified by age group and region.

Finally, daily prevalence of contact sick-leave, $Prev_{CSL}$, was estimated by multiplying the daily probability of contact sick-leave by $D(\text{sick-leave})$ and by $N_{a,r}$, and summing over all a and r,

$$Prev_{CSL} = \sum_{a,r} \big(P_{a,r,t}(\text{contact sick-leave}) \times D(\text{sick-leave}) \times N_{a,r} \big)$$

Table S1. Parameter values.

Parameter	Value	Reference
$N_{a,r}$	Variable	INSEE (5)
$C^{K}_{a,r}$	Variable	Bosetti et al. (3)
$D(\text{sick-leave})$	9.33 days	MH (2)
$D(\text{COVID-19 symptoms})$	8.53 days	MH (2)
$P_{a,r}(\text{employed})$	Variable	INSEE (3)
$P_{a,r,t}(\text{symptomatic COVID-19})$	Variable	Massonnaud et al. (4)
$P(\text{COVID-19 symptoms}\mid\text{sick-leave})$	12.3%	MH (2)
$P(\text{taking sick-leave during survey period})$	2.8%	MH (2)
$P(\text{taking sick-leave}\mid\text{notified})$	67%	MH (2)
$P(\text{symptomatic COVID-19 during survey period})$	0.54%	Massonnaud et al. (1)
$P(\text{notified}\mid\text{prolonged workplace contact})$	60%	{25%, 50%, 75%, 100%} in SA
$P(\text{notified}\mid\text{prolonged family contact outside household})$	60%	{25%, 50%, 75%, 100%} in SA
$P(\text{notified}\mid\text{household contact})$	90%	{25%, 50%, 75%, 100%} in SA
$B_a(\text{working status} = h)$	Variable	Datacovid (4)

Abbreviations: INSEE = National Institute of Statistics and Economic Studies, MH = Malakoff Humanis, SA = sensitivity analysis.

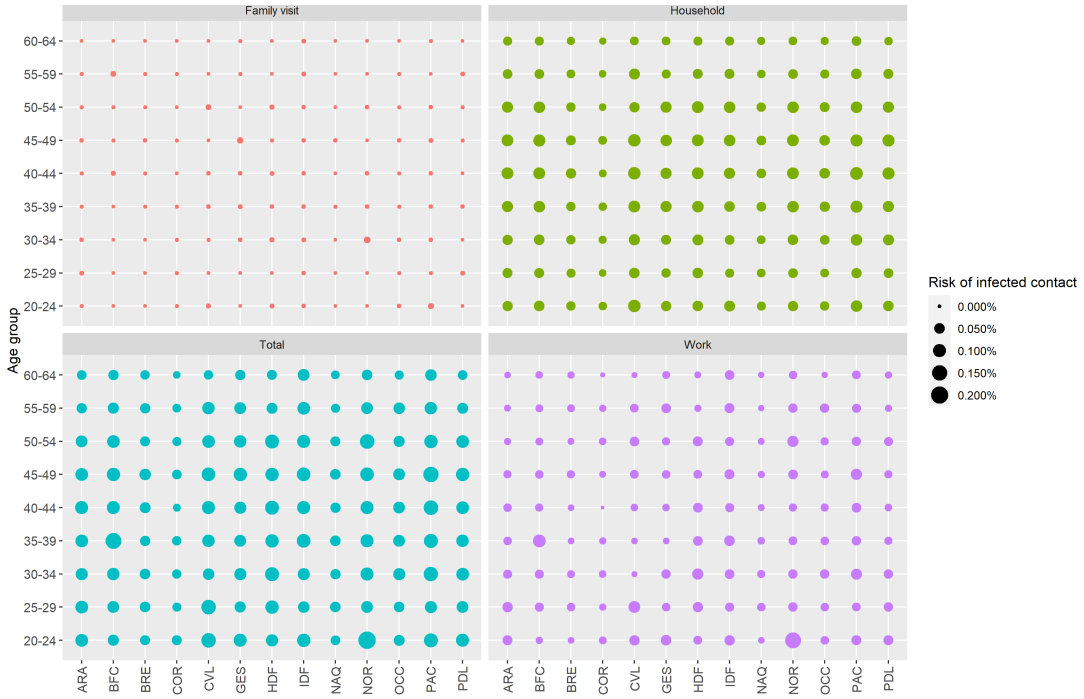

Figure S1. Cumulative risk of being a close COVID-19 contact over the study period (1 March to 31 May 2020) for working-age groups by region. Relevant close contacts included were prolonged contact with family members outside the household (red) contacts within the household (green) and prolonged contacts at the workplace (purple). The total risk across all types of contact (blue) for each age and region was taken as the sum of risks from each individual type of contact. These results were calculating by crossing symptomatic COVID-19 prevalence data from the transmission model by Massonnaud *et al*. and contact data from the SocialCov survey. (ARA = Auvergne-Rhône-Alpes, BFC = Bourgogne-Franche-Comté, BRE = Bretagne, CVL = Centre-Val de Loire, COR = Corse, GES = Grand-Est, HDF = Hauts-de-France, IDF = Île-de-France, NOR = Normandie, NAQ = Nouvelle-Aquitaine, OCC = Occitanie, PAC = Provence-Alpes-Côte d'Azur, PDL = Pays de la Loire.)

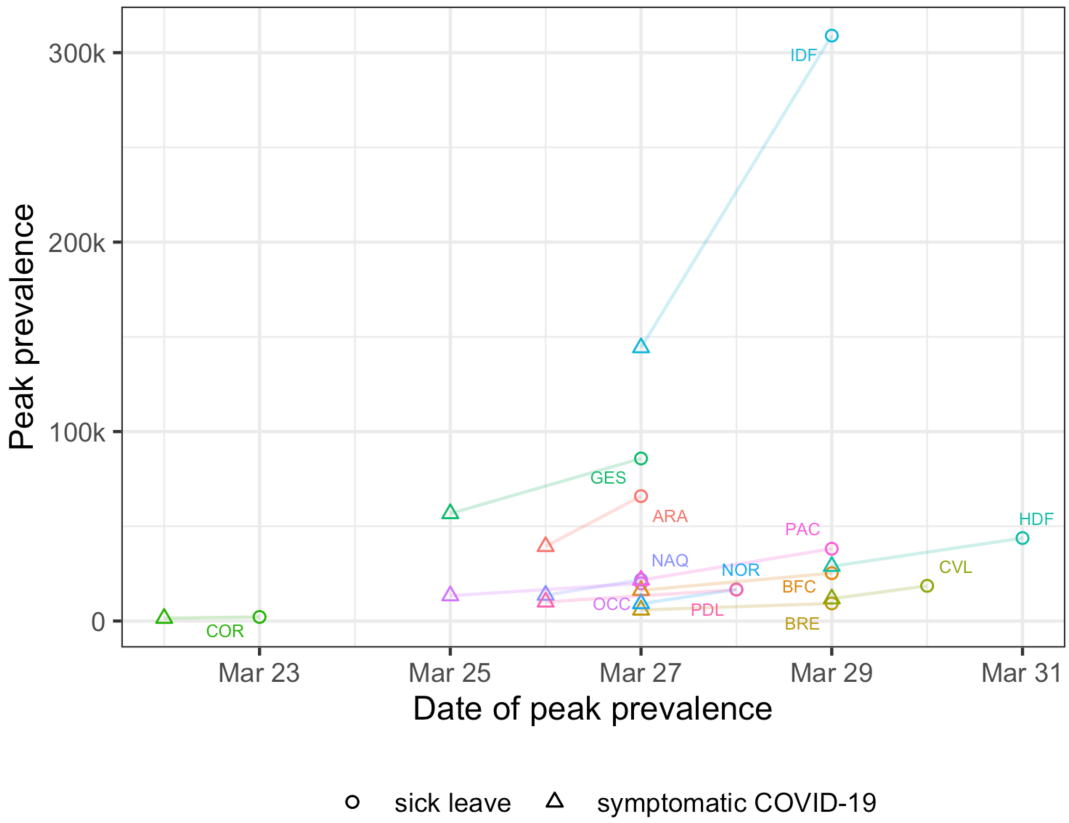

Figure S2. Temporal and regional variation in estimated peak prevalence of sick-leave (circles) and symptomatic COVID-19 among working-age adults (triangles). (ARA = Auvergne-Rhône-Alpes, BFC = Bourgogne-Franche-Comté, BRE = Bretagne, CVL = Centre-Val de Loire, COR = Corse, GES = Grand-Est, HDF = Hauts-de-France, IDF = Île-de-France, NOR = Normandie, NAQ = Nouvelle-Aquitaine, OCC = Occitanie, PAC = Provence-Alpes-Côte d'Azur, PDL = Pays de la Loire.)

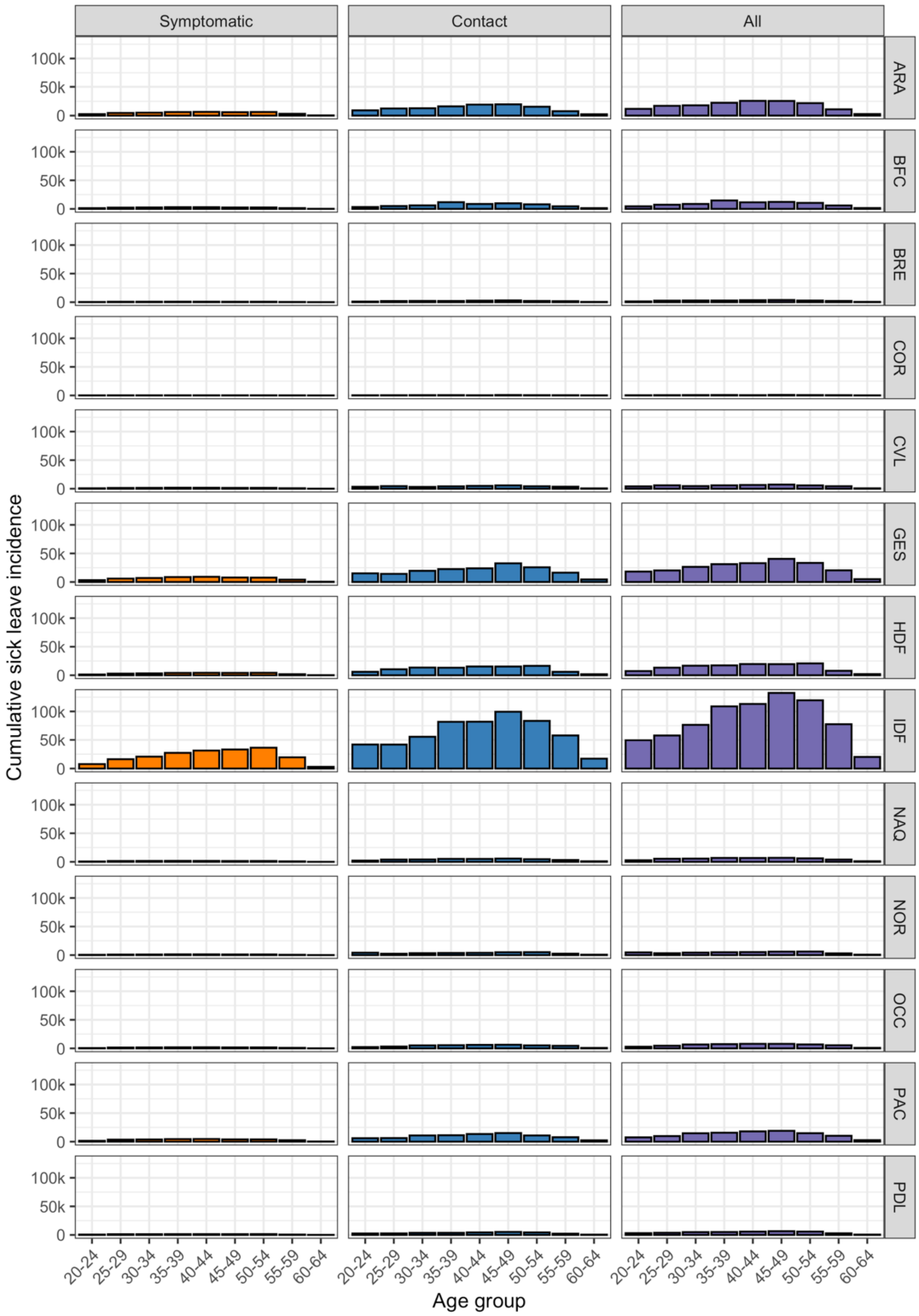

Figure S3. Cumulative number of symptomatic sick-leaves (left column), contact sick-leaves (middle column) and total sick-leaves (right column) between 1 March and 31 May 2020, by

region and age group. (ARA = Auvergne-Rhône-Alpes, BFC = Bourgogne-Franche-Comté, BRE = Bretagne, CVL = Centre-Val de Loire, COR = Corse, GES = Grand-Est, HDF = Hauts-de-France, IDF = Île-de-France, NOR = Normandie, NAQ = Nouvelle-Aquitaine, OCC = Occitanie, PAC = Provence-Alpes-Côte d'Azur, PDL = Pays de la Loire.)

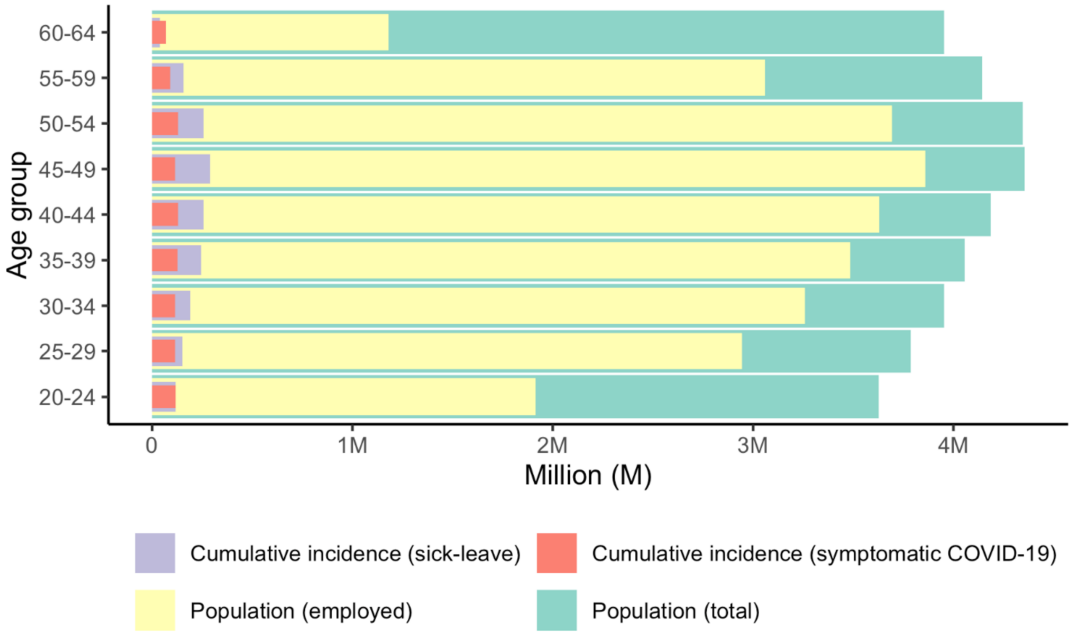

Figure S4. Age distributions of metropolitan France in millions of inhabitants (M) in terms of the total population (green), the actively employed population (yellow), and the cumulative incidence of sick-leave (purple) and symptomatic COVID-19 (red) between 1 March and 31 May 2020.

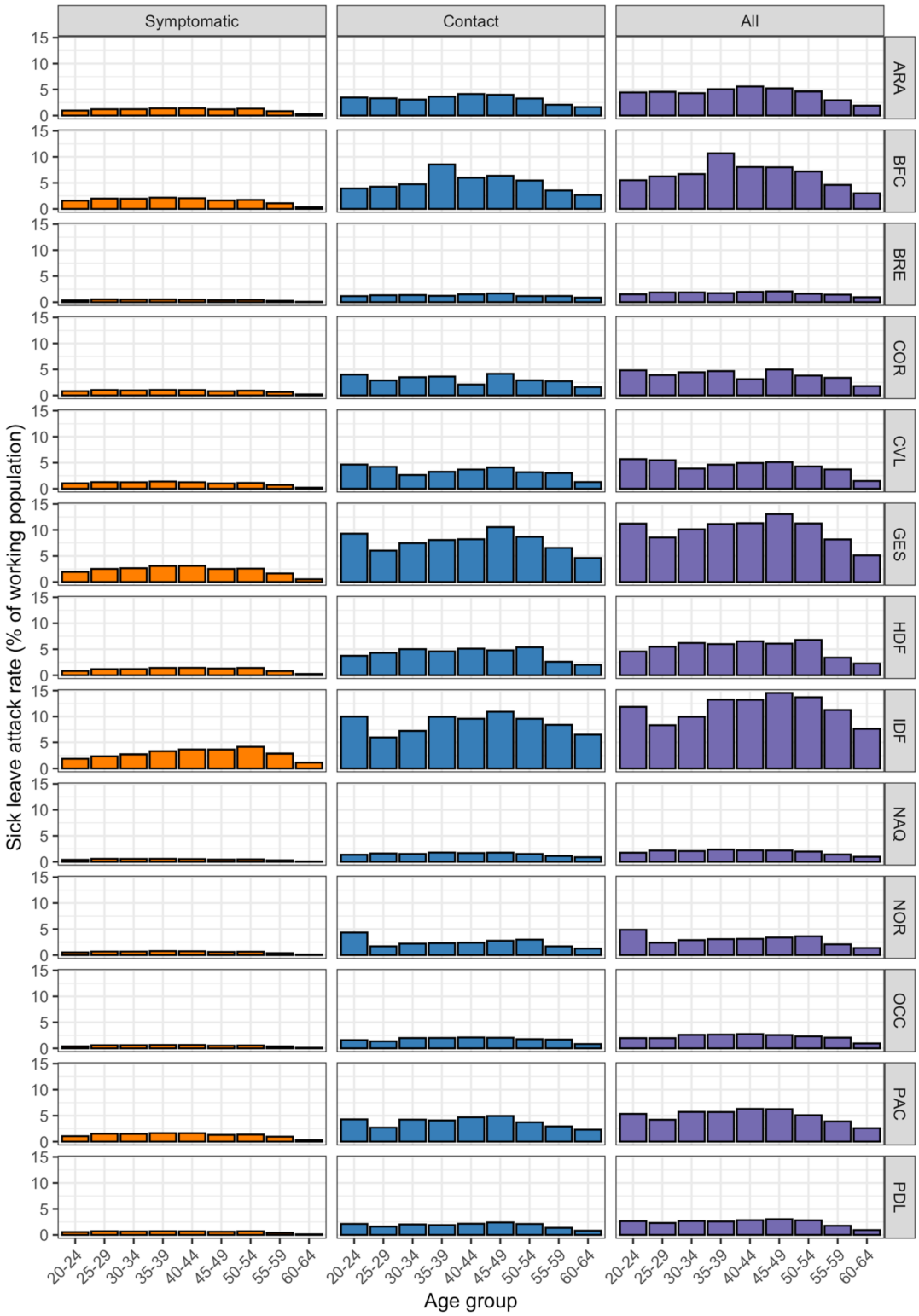

Figure S5. Attack rate of symptomatic sick-leave (left column), contact sick-leave (middle column) and total sick-leave (right column) during the first pandemic wave: the proportion of

the working population that took each type of sick-leave between 1 March and 31 May 2020, by region and age group. (ARA = Auvergne-Rhône-Alpes, BFC = Bourgogne-Franche-Comté, BRE = Bretagne, CVL = Centre-Val de Loire, COR = Corse, GES = Grand-Est, HDF = Hauts-de-France, IDF = Île-de-France, NOR = Normandie, NAQ = Nouvelle-Aquitaine, OCC = Occitanie, PAC = Provence-Alpes-Côte d'Azur, PDL = Pays de la Loire.)

References

1. Massonnaud C, Roux J, Colizza V, Crépey P. Evaluating COVID-19 Booster Vaccination Strategies in a Partially Vaccinated Population: A Modeling Study. *Vaccines*. March 2022;10(3):479. doi:10.3390/vaccines10030479.
2. Baromètre annuel Absentéisme Maladie 2020 - Malakoff Humanis [Internet]. November 2020 [Accessed 28 juin 2021]. Available at: https://newsroom.malakoffhumanis.com/actualites/malakoff-humanis-presente-les-resultats-2020-de-son-barometre-annuel-absenteisme-maladie-545f-63a59.html
3. Bosetti P, Huynh B-T, Abdou AY, et al. Lockdown impact on age-specific contact patterns and behaviours, France, April 2020. *Euro Surveill*. 2021;26(48):2001636. doi:10.2807/1560-7917.ES.2021.26.48.2001636
4. Datacovid, Wave 4, COVID 19 Barometer [Internet]. 4 May 2020 [Accessed 1 March 2021]. Available at: https://datacovid.org.
5. INSEE. Emploi-Population active en 2017 – Emploi-Activité en 2017 | Insee [Internet]. [Accessed 24 November 2021]. Available at: https://www.insee.fr/fr/statistiques/4515500?sommaire=4516095