Towards real-time monitoring of social contacts via participatory disease surveillance

Kathleen N. Kelley[1], Pietro Coletti[2], Nicolò Gozzi[1], Lisa Hermans[3], Mattia Mazzoli[1], Albert Jan van Hoek[4], and Daniela Paolotti[1]

[1] ISI Foundation, Turin, Italy
[2] Universitè Catholique de Louvain, Institute of Health and Society (IRSS), Brussels, Belgium
[3] Data Science Institute, Hasselt University, Hasselt, Belgium
[4] Dutch National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands

Abstract

Social contacts are key drivers of infectious disease transmission, yet most available data on contact behavior come from stand-alone surveys that are resource-intensive, infrequent, and limited in scope. Digital participatory surveillance offers a promising alternative, enabling continuous collection of health-related data with fewer resources. In this study, we analyzed the first year of contact data collected through Influweb, the Italian branch of the European InfluenzaNet network, starting in February 2024. We linked $1,393$ contact diaries with self-reported symptoms and participant characteristics, applying a zero-inflated negative binomial regression model to assess demographic, temporal, and health-related correlates of contact behavior. Based on weighted sample estimates, symptomatic participants reported higher median contacts. In the regression model, mild symptoms were associated with a borderline-significant 20% increase in contacts compared to asymptomatic individuals. Contacts were also higher on weekdays and for employed participants, lower during holidays, and varied by age. In parallel, we constructed age-stratified contact matrices describing contact rates across settings and months. Comparisons with prior studies showed concordance in structural features, including age-assortative mixing in schools and workplaces and intergenerational mixing at home, while overall contact volumes were lower, consistent with recent post-pandemic evidence. These findings demonstrate the potential of digital participatory surveillance for real-time monitoring of social contacts, highlighting its added value in linking contact behavior with health and sociodemographic information, and its potential as a scalable complement to traditional contact surveys.

1 Author summary

Infectious diseases spread through the social contacts we make in daily life, but collecting detailed information on these contacts is usually done through specialized surveys that are costly, infrequent, and limited in scope. In this study, we tested whether digital participatory surveillance—where volunteers regularly share health

information online—can also be used to monitor contact behavior. We invited participants in Influweb, the Italian participatory platform, to report their social contacts and analyzed the first year of responses. By linking contact reports with participants' symptoms and characteristics, we found that people with mild symptoms reported more contacts than those without symptoms. Contact numbers were also higher on weekdays, lower during holidays, and varied by age and employment status. We also constructed contact patterns by age and setting, which reproduced well-known features of social mixing, while overall volumes were lower than in pre-pandemic surveys and consistent with more recent studies. Our findings show that participatory platforms can provide timely and flexible insights into how social behavior and health interact, offering a scalable complement to traditional surveys for understanding disease transmission.

2 Introduction

Understanding how people interact is central to modeling and predicting the spread of infectious diseases [1]. Social contacts provide opportunities for disease transmission, and integrating realistic contact patterns into epidemic models improves their realism, accuracy, and interpretability [2–8]. Such data is also crucial for evaluating the impact of non-pharmaceutical interventions (e.g., social distancing) on transmission dynamics [9–11], for capturing behavioral changes linked to disease spread [12], and, more recently, contact data has been used to assess how socio-economic factors shape disparities in disease burden [13–15].

Social contacts can be measured in several ways, such as using proximity sensors [16–19], time use surveys [20–22], or inferred from large-scale synthetic models [23;24]. Nonetheless, contact diary surveys remain the primary tool used for collecting data on social interactions [25]. In these studies, participants report their encounters on a given day, recording attributes such as (perceived) age and gender of each contacted individual, and setting for each contact (i.e., workplace, school, etc). The main output of these studies is the social contact matrix, i.e. a matrix detailing the average number of contacts between individuals (usually on a daily basis) depending on their age. The landmark POLYMOD study (2006) collected this information via a paper-based survey across eight European countries [26], providing age-structured contact matrices that have since been widely used in transmission models of various infectious diseases [27–29].

Almost fifteen years later, during the COVID-19 pandemic, contact patterns from studies conducted under stable public health conditions, such as POLYMOD, no longer reflected the behavior of individuals living through a public health crisis marked by widespread restrictions. The introduction of non-pharmaceutical interventions aimed at curbing the spread of SARS-CoV-2 significantly impacted both the overall volume and the structure of social contacts. For this reason, several novel data collections were rapidly deployed to capture contact patterns under evolving public health landscapes and social distancing restrictions [12;30;31]. A notable example is the CoMix digital survey launched in March 2020 that collected social contacts data over multiple waves for over 20 European countries between 2020 and 2021 [32], with additional rounds later conducted in several countries in 2022 [33].

While highly informative and valuable, traditional contact surveys face significant challenges: they are indeed logistically demanding, costly, and resource-intensive, as they require extensive recruitment, training, and data collection. As a result, they are typically conducted infrequently and with a cross-sectional approach. Additionally, due to the level of required detail in the contact-related questions (i.e. participants must record every encounter), surveys often restrict the scope and duration of data collection to reduce participant burden. This limits the opportunity to capture richer information about non-explored determinants of contacts and about longitudinal patterns that could provide deeper insights into the drivers of

contact behavior and enable extended representations of contact matrices beyond the commonly considered dimension of age[34].

Digital participatory surveillance platforms, by leveraging the voluntary participation of citizens, can help address the challenges faced by traditional contact surveys. Indeed, they offer direct access to a pool of individuals that can be routinely probed to capture health and behavioral data in a cost-effective manner. The InfluenzaNet network, present in 12 European countries, is a consortium of digital participatory surveillance platforms, that since 2009 continuously collect data through voluntary, web-based participation[35;36]. Over the years, the InfluenzaNet platforms have become reliable sources of high-quality, longitudinal data on influenza-like illness (ILI) prevalence, vaccination rates, and health-related behaviors, among other indicators[37–40].

In this work, we aim to test the feasibility of using participatory surveillance platforms to collect contact data for epidemiological applications. To this aim, we developed a digital contact survey and administered it to the users of Influweb, the Italian branch of the European InfluenzaNet network[38], since February 2024. Users were invited to submit bi-monthly contact diaries where they reported the number of contacts, in the previous day, with individuals in different age groups in four settings. By building on the existing infrastructure, this required minimal additional logistics and substantially fewer resources than a new stand-alone survey. The platform's integration of concurrent surveys allowed us to link participants' contact responses with their questionnaires on socio-demographics, self-reported symptoms, and health-related behaviors, providing a more comprehensive picture of health and behavior while only minimally increasing respondent burden. Moreover, the platform's capacity for continual data collection means that longitudinal trends and shifts in contact patterns, well documented during COVID-19[31;41] and earlier pandemics[42], can be monitored as they occur, providing insights not possible with intermittent surveys.

We enriched a total of $1,393$ contact diaries from 493 unique users, with information regarding self-reported symptoms, socio-demographics, and health-related behaviors of each participant. We constructed overall and setting-specific age-stratified contact matrices, which showed strong age-assortative mixing and high contact intensity observed among minors. We also applied a zero-inflated negative binomial regression model to explore associations between these features and the reported number of contacts. Our work demonstrates the potential of digital participatory surveillance for the monitoring of social contacts and illustrates its added value in linking contact behavior with self-reported symptoms, socio-demographics, and health-related behaviors.

3 Results

3.1 Descriptive overview of Influweb sample and reported contacts

The sample included 1393 survey responses from 439 participants, from February 1, 2024 to February 1, 2025. Participants were invited to complete a contact survey every other month; however, because recruitment was conducted on a rolling basis, data collection timing does not align across individuals. For example, a participant who completed their first survey in March would have been invited to respond again in May, while another who began in February would have their next opportunity in April. The goal of this approach was to sample the entire cohort every month without burdening too frequently the single individuals. In each questionnaire, participants were asked to report all interpersonal contacts in the previous day, the perceived age of each contacted individual, and the setting of each contact. More details on the contact survey design

are provided in the Materials and Methods section 5.2.

The sample was composed by 55% of men. The largest age group consisted of individuals aged 45–64, comprising 51% of the participant pool, followed by those aged 65 and older and those aged 19–45. The smallest age group, individuals aged 18 and under, made up 4% of the sample, with surveys in this group completed by caregivers on behalf of minors. In terms of employment status and education, 56% of participants were employed and 48% held a university degree. Table 1 shows further details of sample characteristics.

On average, 116 responses were collected each month and most participants completed four or more surveys during the study period (see Figs S1 and S2 in the Supporting Information for the distributions of the number of surveys returned each month and the number of responses per participant, respectively).

Regarding symptom reporting, 7% of survey responses (8%of the participant sample) were associated with of symptoms consistent with the ECDC case definition of influenza-like illness (ILI) [43] occurring within 30 days prior to the contact diary entry. More broadly, 26% of responses (30% of participants) were linked with at least one symptom registered within the 30 days prior to completing a contact survey. Cough and sore throat were the most commonly reported symptoms, followed by malaise.

To examine how social contact patterns vary across individual characteristics and time, we calculated the weighted median number of daily contacts stratified by month and setting (Fig 1), demographic attributes (Fig 2a), and by chronic condition status and self-reported symptom status (Fig 2b). Estimates and their 95% confidence intervals were derived using bootstrap resampling (10,000 iterations), reflecting sampling variability.

Fig 1 presents the monthly proportion of reported contacts by setting (left axis), alongside the weighted median number of contacts each month with 95% confidence intervals derived from bootstrap sampling (right axis). Notably, July and August show a decline in the median number of reported contacts, coinciding with a marked reduction in work and school contacts. During this same period, in August, the proportion of contacts attributed to leisure increases substantially, suggesting seasonal shifts in contact behavior.

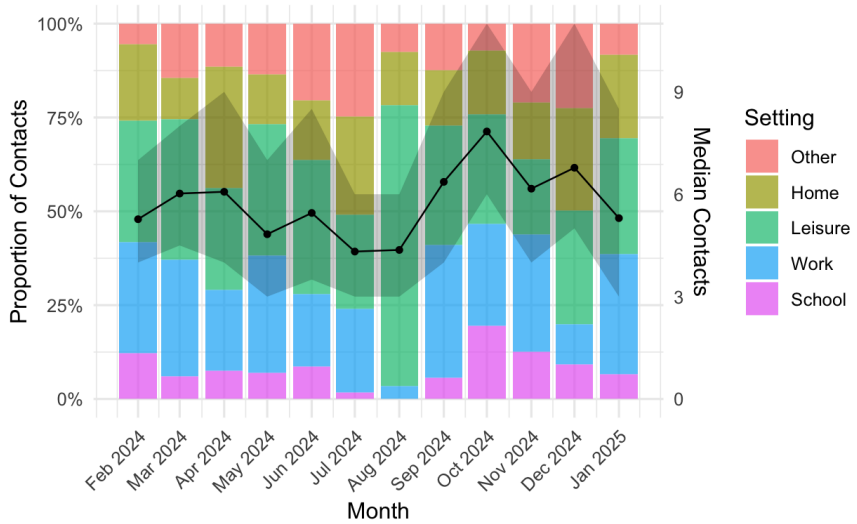
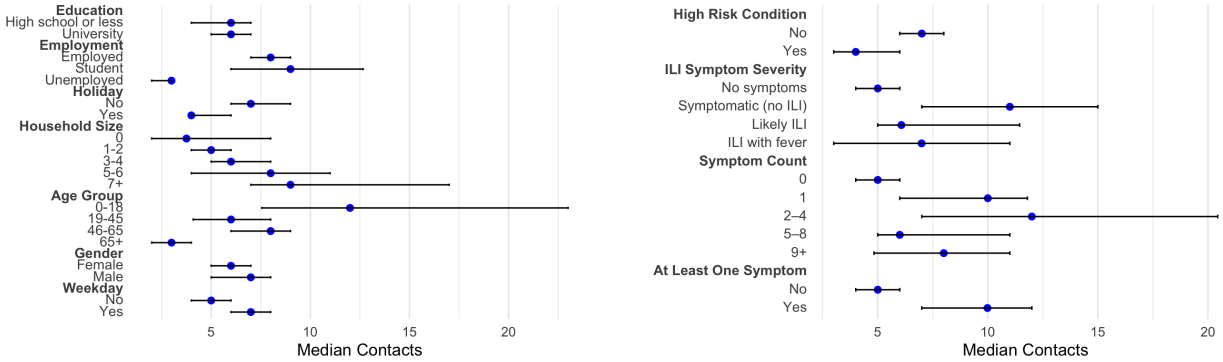


Fig 1. Proportion of contacts by setting and monthly median contact counts. Stacked bars show the proportion of total monthly contacts attributed to each setting (left axis). The overlaid line shows the weighted median number of contacts per participant each month (right axis), with 95% bootstrap confidence intervals shaded in grey.

	No. of participants (%)	No. of responses (%)
Education		
High school or less	227 (52%)	695 (50%)
University	212 (48%)	698 (50%)
Employment status		
Employed	244 (56%)	756 (54%)
Student	39 (9%)	116 (8%)
Unemployed	156 (36%)	521 (37%)
High-risk condition		
No	314 (72%)	964 (69%)
Yes	125 (28%)	429 (31%)
Household size		
0	18 (4%)	45 (3%)
1–2	168 (38%)	577 (41%)
3–4	160 (36%)	500 (36%)
5–6	57 (13%)	141 (10%)
7+	36 (8%)	130 (9%)
Age group		
0–18	18 (4%)	51 (4%)
19–45	82 (19%)	239 (17%)
46–65	224 (51%)	707 (51%)
65+	115 (26%)	396 (28%)
Sex		
Female	196 (45%)	587 (42%)
Male	243 (55%)	806 (58%)
Survey day		
Weekend	—	441 (32%)
Weekday	—	952 (68%)
Holiday period		
Non-holiday	—	943 (68%)
Holiday	—	450 (32%)
Symptoms reported		
None	—	1031 (74%)
Symptoms (no ILI)	—	265 (19%)
Likely ILI	—	51 (4%)
ILI with fever	—	46 (3%)
Symptom count		
0	—	1031 (74%)
1	—	100 (7%)
2–4	—	175 (13%)
5–8	—	55 (4%)
9+	—	32 (2%)
At least 1 symptom	—	362 (26%)

Table 1. Overview of study participants and survey responses. Descriptive statistics for study participants and survey responses. Note percentages may not sum to 100% due to rounding.

Fig 2a shows the median number of contacts by demographic groups and highlights clear heterogeneity



(a) **(b)**

Fig 2. Median daily contacts by participant characteristics and health status. Weighted median daily contacts by (a) demographic characteristics and (b) comorbidity and symptom status, with 95% bootstrap confidence intervals.

across demographic groups. Median contacts tend to be higher among children and students compared to older adults and unemployed individuals, and lower among those living alone relative to participants in multi-person households.

Fig 2b presents weighted median contacts stratified by symptoms reported within the 30 days prior to survey submission. Participants reporting any symptoms tended to have higher median contacts than asymptomatic individuals, although those with a self reported high-risk chronic condition (e.g., asthma, diabetes, chronic lung disease, cardiovascular disease, renal disease, or immunosuppression due to treatment or underlying illness) reported fewer median contacts than those who did not report a chronic condition.

3.2 Exploring correlates of contact behavior

We fit a negative binomial regression model with a zero-inflation component to estimate correlates of the number of contacts of each participant, exploring social and demographic features, household size, day of the week, holiday period and reported symptoms. Post-stratification weights for age, sex, and day of week were applied in model estimation, and random intercepts were included to account for repeated measures within participants. Estimated coefficients are reported in Table 2.

Weekday responses were associated with 32% higher contact counts compared to weekends (IRR = 1.32; 95% CI: $1.09 - 1.61$), consistent with increased work- and school-related contacts during the week. Similarly, the holiday period indicator was significantly associated with contact behavior: participants reported 22% fewer contacts during holiday periods compared to non-holiday times (IRR = 0.78; 95% CI: $0.68 - 0.91$).

Participants who were unemployed reported 41% fewer contacts compared to those employed (IRR = 0.59; 95% CI: $0.42 - 0.81$). Similarly, contact counts were lower among individuals aged 19–44 and those aged 65+ compared to the reference group aged 45–64, representing 38% (IRR = 0.62; 95% CI: $0.47 - 0.83$) and 40% (IRR = 0.60; 95% CI: $0.43 - 0.86$) fewer contacts, respectively. No significant differences were observed by sex, education level, or household size at the 5% significance level.

Finally, we considered the symptoms that participants reported in the 30 days prior to each contact survey entry. Participants who reported being symptomatic but without qualifying as ILI had 20% more contacts than those without symptoms, an effect that was borderline significant at the 5% level (IRR = 1.20; 95% CI: $0.99 - 1.44$). Those reporting symptoms consistent with ILI or ILI with fever did not significantly

Variable	Coefficient	IRR	Std. Error	P-value
Employment (ref: Employed)				
Student	-0.132	0.876	0.228	0.563
Unemployed	-0.534	0.586	0.165	0.001 **
Weekday vs Weekend	0.278	1.320	0.100	0.005 **
Holiday period vs Non-holiday	-0.243	0.784	0.076	0.001 **
Age group (ref: 45–64)				
0–18	0.056	1.057	0.323	0.863
19–45	-0.475	0.622	0.147	0.001 **
65+	-0.504	0.604	0.179	0.005 **
Symptoms (ref: No symptoms)				
Symptomatic, no ILI	0.178	1.195	0.094	0.059 †
ILI	0.169	1.184	0.177	0.341
ILI with fever	-0.191	0.826	0.197	0.332
Chronic Condition	0.049	1.050	0.125	0.698
Household Size	0.025	1.026	0.019	0.180
Male vs Female	0.127	1.136	0.112	0.255
Education: University vs High School or less	0.156	1.169	0.114	0.172

† $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2. Zero-inflated negative binomial regression model results. Results from the zero-inflated negative binomial regression of contact counts, including random intercepts for participants. Estimates were obtained using post-stratification weights for age, sex, and day of week. The incidence rate ratio (IRR) is the exponentiated coefficient.

differ in contact counts from those without symptoms.

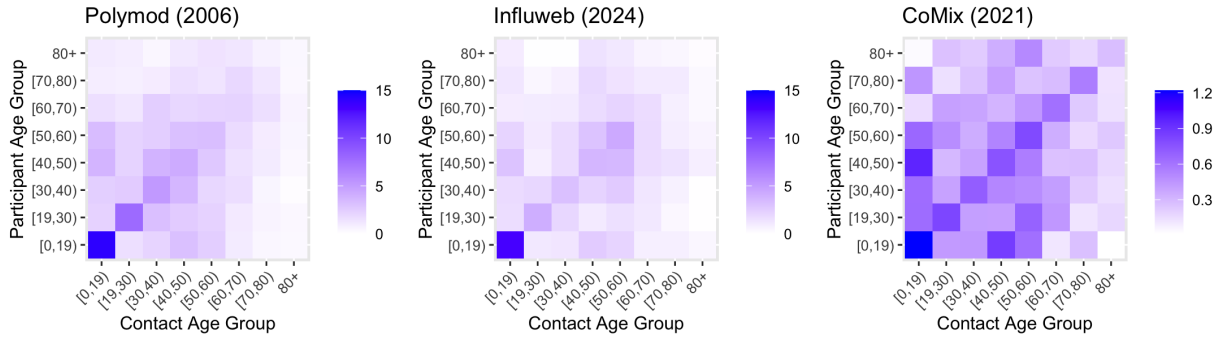
3.3 Contact matrix comparison across studies

To assess how contact matrices from our study compare to prior work, we examined both overall and setting-specific contact matrices alongside those from the POLYMOD[26] and CoMix[31] studies, chosen among other contact studies for their Italian data availability and relevance for contrasting pre- and post-pandemic contact patterns. All contact matrices were constructed using post-stratification weights for age, sex, and day-of-week.

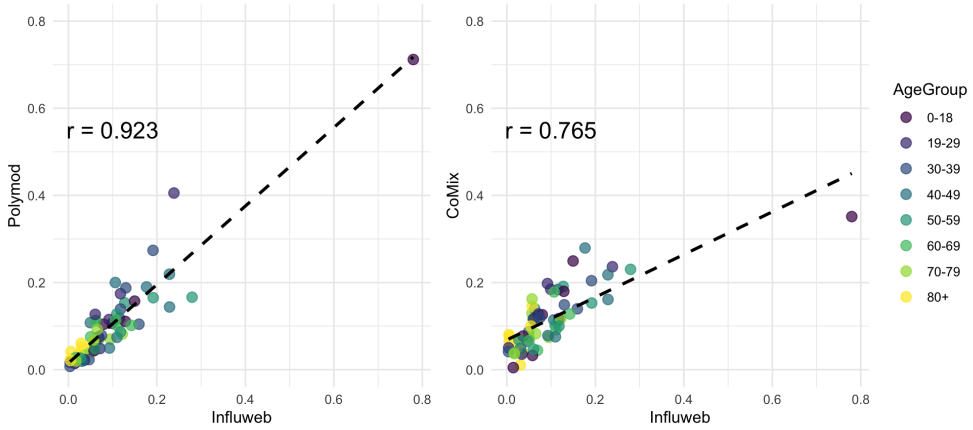
We assessed the similarity between contact matrices using three complementary metrics: cosine similarity, the Sørensen–Dice index, and the Frobenius norm, to provide a view of both relative and absolute agreement between studies. Cosine similarity captures the alignment of contact patterns by assessing whether the same matrix positions tend to have high or low values relative to other age pairs, the Sørensen–Dice index measures the overlap in contact volumes between matrices, with greater weight on shared presence, and the Frobenius norm quantifies the overall magnitude of differences between matrices, being sensitive to total contact volume. Further details are provided in the Materials and Methods section 5.6.

We first present results for the overall matrices, followed by comparisons stratified by setting (home, school, work, and leisure).

3.3.1 Overall contact matrix comparison



(a) Age-structured contact matrices for the POLYMOD, Influweb, and CoMix studies. Each cell represents the per capita daily contact rate between individuals in age group i and those in age group j. Contact matrices were constructed using post-stratification weights for age, sex, and day of week, and symmetrized using the Italian national population structure for the year of data collection. Note the difference in the scale for CoMix relative to the other two studies.



(b) Comparison of age-coupled values from the Influweb contact matrix to those from the POLYMOD (left) and CoMix (right) matrices. Each point represents a unique age group pairing. Matrix values are normalized such that the dominant eigenvalue equals one, resulting in unitless relative contact intensities. The dashed line is a fitted regression line; x-axes show values from Influweb, while the y-axes show corresponding values from POLYMOD or CoMix.

Fig 3. Comparison of age-specific contact patterns across studies. Panel (a) shows age-structured contact matrices from the three studies, and panel (b) compares normalized contact intensities between Influweb and the other datasets.

Fig 3a shows the overall contact matrices for the three studies considered. We note the difference in scale between CoMix and the other two studies. POLYMOD and Influweb appear visually similar, with the highest number of contacts occurring among those aged 18 and under (14 in POLYMOD and 13 in Influweb), and a diagonal pattern indicating frequent interactions within age groups. In contrast, the CoMix matrix has a noticeably lower overall scale and a more muted diagonal, a pattern observed in other pandemic-era studies and linked with reduced inter-generational mixing and more heterogeneous off-diagonal values [31].

The quantitative comparison reported in Table 3 supports these visual observations. The cosine similarity and Sørensen–Dice index between Influweb and POLYMOD indicate strong alignment in both the age-mixing

structure and the overlap in contact volume. The Frobenius norm suggests a moderate difference in contact intensity. In contrast, the comparison with CoMix reveals a lower Sørensen–Dice index and higher Frobenius norm, pointing to greater differences in both the distribution and total volume of reported contacts. The cosine similarity remains relatively high, suggesting that the patterns of high- and low-contact cells are similar, though the number of contacts in those positions differs.

	Cosine Similarity		Sørensen–Dice Index		Frobenius Norm	
Setting	vs POLYMOD	vs CoMix	vs POLYMOD	vs CoMix	vs POLYMOD	vs CoMix
Overall	0.959	0.868	0.822	0.404	7.485	15.032
Home	0.900	0.921	0.739	0.784	1.955	1.272
Work	0.895	0.880	0.761	0.257	2.588	3.313
School	0.982	0.409	0.824	0.211	1.949	8.837
Leisure	0.869	0.695	0.790	0.005	2.998	5.553

Table 3. Contact matrix similarity metrics between Influweb, POLYMOD, and CoMix. Matrix similarity metrics for both overall and setting-specific contact matrices from Influweb compared with the other two studies. Higher values of cosine similarity and Sørensen–Dice indicate greater similarity; lower Frobenius norms indicate smaller differences.

Fig 3b further illustrates the comparison by plotting the values of each cell from the normalized contact matrices. The POLYMOD and Influweb matrices show clustering around the regression line, indicating strong agreement in normalized contact intensity (Pearson correlation of 0.92, $p < 0.001$). In contrast, the CoMix data show greater vertical dispersion, suggesting more variability and weaker agreement in specific age-pair interactions. Despite this, the overall Pearson correlation is 0.77 ($p < 0.001$), indicating moderate similarity of contact structure even under pandemic restrictions.

3.3.2 Setting-specific contact matrices comparison

We examined setting-specific contact matrices, also constructed using post-stratification weights for age, sex, and day of week, to explore how contacts in different social environments compare across studies.

Table 3 shows higher cosine similarity and Sørensen–Dice values in the Influweb–POLYMOD comparison compared to the Influweb–CoMix comparison across all settings except for home, where both prior studies show strong agreement with Influweb. School contacts show the strongest alignment, with the highest cosine similarity and Sørensen–Dice index metrics, as well as visual agreement in the strong 0–18 to 0–18 block (see S6 Fig in Supporting Information). The work matrices from all studies also align well, with high cosine similarity scores, and visible contact clustering among ages 30 to 60 in their work-setting matrices (see S3 and S4 Figs for Influweb and POLYMOD, respectively). Home contacts show more variation between the two studies but still reflect common patterns, such as strong within-minor and minor-to-30–50-year-old contacts. Leisure settings show a pronounced diagonal in both studies, indicating age-assortative mixing, though contact volume is lower in Influweb.

The comparison in Table 3 between the school and leisure settings in Influweb versus CoMix show the lowest Sørensen–Dice indices and highest Frobenius distances, reflecting pandemic-era restrictions such as school closures and limits on social activities. The correlations for these settings, shown in S7 Fig, are markedly lower compared to the home and work settings. Both studies showed peak work interactions between individuals aged roughly 30–50, as seen in the setting specific contact matrices for Influweb (S3 Fig) and CoMix (S5 Fig).

3.4 Differences in derived R_0 distributions

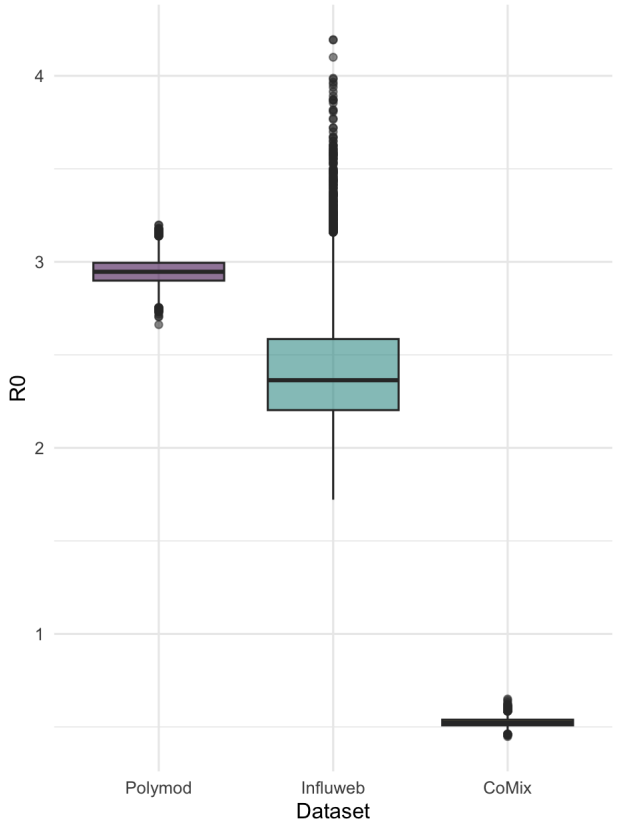


Fig 4. Estimated R_0 distributions from POLYMOD, Influweb, and CoMix contact data. Bootstrapped distributions of R_0 estimates from the POLYMOD, Influweb, and CoMix contact studies. Each distribution is based on 10,000 weighted contact matrices, with R_0 calculated as the dominant eigenvalue of the next-generation matrix.

The basic reproduction number (R_0) represents the average number of secondary infections caused by a typical infectious individual in a fully susceptible population. For each study, we calculated R_0 of a hypothetical respiratory pathogen, such as the SARS-CoV-2 virus, as the dominant eigenvalue of the next-generation matrix[44], derived from the weighted contact matrices (see Materials and Methods subsection 5.7 for details). The resulting distributions of the R_0 values are shown in Fig 4.

The distribution from the POLYMOD dataset was narrow, with a median R_0 of 2.9 (95% CI: $2.8 - 3.1$). The R_0 computed with the Influweb contact matrix exhibited a wider distribution with a median of 2.4 (95% CI: $2.0 - 3.2$). The CoMix matrix, collected under pandemic social restrictions, produced a tight distribution of R_0 centered around 0.5 (95% CI: $0.5 - 0.6$). While the median estimates differ, these values are based on bootstrap-derived distributions, and the confidence intervals reflect uncertainty due to sampling variability rather than results from formal significance testing. A two-sample Kolmogorov–Smirnov test indicated that the R_0 distributions from Influweb differed significantly from those of both POLYMOD and CoMix ($p < 0.001$ for both).

4 Discussion

In this study, we tested the potential of participatory surveillance platforms to collect meaningful social contact data. This objective is motivated by the fact that, unlike stand-alone contact surveys, these platforms offer a cost-effective approach to real-time monitoring and allow the integration of diverse participant data. By linking multiple survey types within the same system, they enable deeper investigation into factors shaping contact behavior that are difficult to study using traditional, isolated survey methods. With established systems already operating in many countries, this approach provides a scalable framework for enhancing our understanding of human-to-human infectious disease transmission.

To this end, we validated the contact data collected through Influweb against large, representative surveys of social mixing. Comparative analyses show strong concordance with two prior large-scale, representative studies, namely POLYMOD[26] and CoMix[32]. Indeed, matrix comparison metrics indicate a high degree of agreement between the Influweb and POLYMOD studies in terms of age-mixing structure and contact volume. Agreement with CoMix was weaker in absolute volume, reflecting its pandemic-era context, though the age-mixing structure was consistent. Influweb's setting-specific contact matrices likewise show expected age-assortative patterns at school and work, and inter-generational mixing at home between minors and adults of likely parental age–patterns already observed in the POLYMOD setting-specific matrices and in recent post-pandemic studies[45;46]. Overall, these results show that participatory surveillance platforms can capture reliable contact patterns, offering complementary, less resource-intensive alternatives to traditional surveys for monitoring social contacts at scale, with infrastructure already established in many countries.

Beyond generating valid contact data, a major strength of Influweb and InfluenzaNet platforms lies in its integration of a broad range of participant data that is typically harder to obtain through traditional contact surveys. All contact study participants contributed to weekly syndromic surveillance, enabling the linkage of contact behavior to symptom history. In our study, descriptive analyses showed that participants reporting symptoms had higher median contacts than those without symptoms, regardless of severity. A similar, although borderline-significant, effect was observed in the regression model, where participants with mild symptoms not qualifying as ILI reported 20% more contacts than individuals not reporting symptoms. To contextualize these findings, we examined an opt-in survey item asking symptomatic participants whether they changed their routine. Overall, 69% of respondents reported no change. Even among those with ILI symptoms, most either made no change (44%) or adjusted their routine without taking time off work/school (20%). Only those reporting ILI with fever showed a majority (51%) who stayed home (see Supporting Information Table S1). These results suggest that many symptomatic individuals in our sample remained socially active, a pattern with implications for disease transmission and control. Other contact studies provide similar perspectives. A recent analysis of aggregated population trends in weekly contacts and respiratory symptoms in the United Kingdom reported that peaks in contacts coincided with peaks in symptoms[47]. In a different design, Ref.[48] observed reduced contacts among symptomatic individuals recruited from antiviral distribution centers. Taken together, these results indicate a meaningful association between symptoms and social behavior, even though no causative conclusions can be drawn.

In addition to health status, temporal context strongly impacts contact behavior. Participants reported fewer contacts on weekends and holidays than on days in which individuals attended work or school, reflecting the role of work and school routines. This was seen both in the setting-specific matrices, where contacts were highest in schools and workplaces, and in our regression results, which showed employed individuals had more daily contacts than those unemployed. These patterns align with prior studies[8;26;45;49] and emphasize the

importance of situating contact behavior within its temporal and social context. With their cost-effective, established infrastructure for the continuous collection of contact, syndromic, and health data, participatory surveillance platforms like Influweb offer unique opportunities to capture and explore these health- and time-dependent trends in future studies.

The same infrastructure enables participatory surveillance platforms to monitor for recurrent disease threats and enable rapid response during evolving public health situations. In addition to routine temporal effects due to work or school attendance, contact patterns also shift abruptly during outbreaks and evolve gradually as behavioral patterns and demographic structures change, such as population aging in Italy in the decades since the POLYMOD study[50]. Our findings show contact patterns shaped by present-day demographic and behavioral contexts. Overall contact volumes estimated by Influweb are lower than those reported in the POLYMOD study. As a result, the median of the R_0 distribution for the same hypothetical respiratory pathogen was lower when derived from Influweb matrices compared to POLYMOD matrices, although our estimates were more variable, likely due to the smaller sample size. This comparison across datasets should be interpreted cautiously, as R_0 estimates are sensitive to sampling method, contact definitions, and the composition of the sample population, as the same disease can spread differently in different populations[51]. Nonetheless, this finding is consistent with recent European studies showing reduced contacts and transmission potential in the post-pandemic period compared to the pre-pandemic[9;33;45]. Together, these results highlight the importance of regularly updating contact data used in epidemic modeling. Abrupt behavioral changes observed during COVID-19[31] and other outbreaks[42] show that contact data collected under normal public health conditions cannot be relied upon during crises, underscoring the need for systems capable of capturing rapid changes in real time. The continuous data collection enabled by platforms like Influweb provides such a resource, allowing contact behavior to be monitored as it evolves.

While these findings highlight the value of participatory surveillance, our approach is not without limitations. Influweb's participants are self-selected, which may over-represent health-conscious individuals[36]. The application of post-stratification weights helps correct demographic imbalances but cannot fully eliminate selection bias. As such, our results should be interpreted as reflecting patterns in an engaged surveillance cohort rather than nationally representative estimates. The observational design precludes causal inference, and reliance on self-reported data introduces risks of recall and social desirability bias, though these were likely reduced by the short recall windows and the relatively stable public health context during data collection. Survey fatigue is a concern in longitudinal designs, but response rates stabilized after the launch, and most participants completed four or more diaries (see Figs S1 and S2).

Despite these limitations, evidence from over a decade of participatory surveillance platforms in multiple countries, including Influweb, show that citizen-sourced data tracks well with sentinel-based systems and are frequently used as complementary sources in population health surveillance[35;36;38;52]. In times of crisis, when traditional data systems may be delayed or overwhelmed, participatory systems offer fast and reliable insights[53;54]. Our study extends this evidence to social contact data, demonstrating how participatory surveillance can complement traditional surveys by enabling continuous monitoring of social contacts while simultaneously integrating symptom, demographic, and behavioral data. Such systems can be deployed with minimal adaptation in other countries where participatory surveillance platforms already exist, making them a scalable and resource-efficient complement to traditional approaches. By linking contact behavior with health information in real time, these platforms hold particular value for both routine surveillance and emergency response. Future applications also hold potential for integrating additional dimensions of data. Spatial granularity could enable analyses of urban–rural differences or north–south divides, while

the incorporation of biological sampling, as piloted by Infectieradar in the Netherlands with participant self-swabs[55], opens the way for combining behavioral and genomic insights.

Looking ahead, this study provides proof of concept for building an international social contact observatory through multi-country collaboration of citizen-science networks, providing timely, context-specific data to strengthen epidemic modeling and public health preparedness.

5 Materials and Methods

5.1 Ethics statement

The Influweb study complies with the Italian Data Protection Authority's regulations and the European Union's General Data Protection Regulation (GDPR). The protocol was reviewed by the ISI Foundation's institutional review board, which waived the need for formal ethics approval, as informed consent was obtained online, and all data are processed in accordance with GDPR. Data are pseudonymized, linked via anonymized IDs, and used only in aggregated form for scientific purposes. The Influweb platform (https://influweb.org/) provides detailed privacy information, and all data are hosted on GDPR-compliant infrastructure with appropriate security measures, including encryption and two-factor authentication.

5.2 Study design and data

Influweb

Influweb is a participatory digital surveillance platform that has been active since 2008. It operates in parallel with traditional public health surveillance systems to monitor influenza-like illness (ILI) in Italy. The signal from Influweb's syndromic surveillance has been shown to align well with those from sentinel-based systems, supporting its utility as a data source[35;36;52].

The data used in this study are based on three types of regular (yearly, monthly and weekly) surveys hosted on the Influweb platform:

1. Intake survey: This collects demographic and health information, including age, sex, region of residence, education level, vaccination status, and the presence of chronic conditions. Participants are invited every year to submit new intake information.

2. Weekly syndromic survey: Participants receive weekly email reminders to report whether they have experienced any symptoms in the past seven days and to describe associated health behaviors. These surveys are available year-round to support continuous monitoring of ILI.

3. Social contact survey: Beginning in February 2024, existing participants were invited to take part in a social contact study. The survey prompts individuals to report all interpersonal contacts that occurred between 5:00 a.m. the day prior and 5:00 a.m. on the current day. Reported contacts are categorized by age group, gender of the contact, and the setting in which the contact took place (home, work, school, leisure).

This is an observational longitudinal study. A total of 446 participants from the syndromic surveillance study accepted the invitation to join the contact survey study. They were invited to complete a contact survey every other month, while continuing to participate in weekly syndromic surveillance surveys. This

bi-monthly schedule was intended to minimize participant burden and reduce the likelihood of survey fatigue. Recruitment was rolling, meaning participants began at different times, so follow-up points do not align across individuals. For example, someone who submitted their first contact survey in March would be re-invited in May, while a participant beginning in February would receive their next invitation in April.

While the contact survey is ongoing at the time of writing, the data included in this analysis covers a 12-month period, resulting in 1405 distinct survey submissions. This enables the capture of contact patterns across seasons and in the context of concurrent health behavior reporting over the course of one year.

POLYMOD and CoMix

To contextualize the findings from the Influweb contact survey, we compared our results with data from two widely used, representative prior studies of social contact behavior: the POLYMOD and CoMix surveys. In both cases, we used only the data collected in Italy. While other datasets on social contacts exist, these studies were chosen because they provide accessible, directly comparable data for the Italian context and capture distinct social mixing situations, pre-pandemic (POLYMOD) and pandemic-era (CoMix), against which to situate our findings.

The POLYMOD study was conducted between May 2005 and September 2006 as a cross-sectional survey across eight European countries, including Italy. There are 849 surveys from Italian participants who reported their social contacts using paper diaries over a single 24-hour period. Contact data were collected alongside demographic information, and participants were recruited to be broadly representative of the population by age, sex, and region. Further methodological details are available in Mossong et. al 2008 [26].

The CoMix study began in March 2020 as a longitudinal, multi-country survey designed to monitor social behavior during the COVID-19 pandemic. In Italy, CoMix collected 7657 survey observations. Participants reported all contacts made during a 24-hour period in repeated waves. The Italian CoMix data used in this study spans the same diary-based structure as POLYMOD but captures contact behavior during a period of active public health interventions. More information on CoMix study design and sampling can be found in Verelst et. al 2021 [32].

5.3 Post-stratification weights

Post-stratification weights were applied to align the study samples with the reference population (i.e., the Italian general public in the year of data collection) in terms of age and sex distribution, as well as the temporal distribution of weekdays and weekends.

First, we accounted for differences between the age and sex distribution of the survey sample and the reference population, a common practice in survey analysis [56].

Let $k = (s, a)$ denote a strata defined by sex $s \in \{M, F\}$ and age $a \in \{0{-}4, 5{-}9, \ldots, 96{-}100\}$. Weights were calculated by taking the ratio of the proportion in stratum k of the Italian population in a given survey year to the proportion of the study sample in the same stratum:

$$w_k = \frac{P_k/P}{N_k/N},$$

where P represents the Italian population and N the sample. Population counts were obtained from national demographic data [57;58].

Then we adjusted for differences in survey response timing. Contact patterns differ by day of the week, with higher rates often seen on weekdays at work and school. To ensure that weekday and weekend responses

reflect their share of the calendar week, we apply day-of-week weights:

$$w_{day.of.week} = \frac{5/7}{N_{weekday}/N} \quad \text{or} \quad \frac{2/7}{N_{weekend}/N},$$

where $N_{weekend(weekday)}$ is the number of surveys returned on the weekend(weekday), and N is the total number of surveys.

The final weight for participant p is the product of the age-sex and day-of-week weights: $w_p = w_k \times w_{day.of.week}$. These weights were applied in all analyses (contact matrix construction, regression modeling, and calculation of median daily contacts by characteristic) to the Influweb sample, as well as in the POLYMOD and CoMix matrix construction, ensuring consistency and comparability across studies.

5.4 Contact matrix construction

Weighted contact rates were computed as the mean number of contacts between respondent and contact age groups. The following age groups were defined for use in the contact matrices: 0–18, 19–30, 30–40, 40–50, 50–60, 70–80, and 80 years and older. For respondent age group i and contact age group j:

$$C_{ij} = \frac{\sum_{p \in i} w_p \, c_{p,j}}{\sum_{p \in i} w_p},$$

where $c_{p,j}$ is the number of contacts reported by respondent p with individuals in group j, and w_p is the participant weight.

Because social interactions are reciprocal, an interaction a person in age group i has with someone in age group j should also be reflected in age group j's contacts with age group i. The total number of contacts from age group i to j should equal the number from j to i, when scaled by the size of each group[59]. This symmetry condition can be written as $C_{ij} \times N_i = C_{ji} \times N_j$.

Survey data rarely satisfy this symmetry, due to variations in sampling and reporting errors, among others[60] . To correct for this we symmetrize matrices using census-based estimates of the Italian population in the survey year[57;58]. Let P_i (P_j) denote the population size of age group i (j). The symmetric contact rate is[15;60]:

$$C_{ij}^{\text{sym}} = \frac{1}{2P_i} \left(C_{ij} P_i + C_{ji} P_j \right).$$

To account for sampling variability and quantify uncertainty, we generate bootstrap replicates of the contact matrices for each study. In each iteration, survey participants are sampled with replacement to create a dataset of the same size as the original, and a weighted, symmetrized contact matrix was constructed as described above. This process is repeated n_{boot} times ($n_{\text{boot}} = 10000$ for overall matrices and $n_{\text{boot}} = 5000$ for the setting-specific matrices of home, school, work, or leisure).

The final contact matrix is the element-wise mean of the n_{boot} replicates, with uncertainty quantified by calculating confidence intervals[61].

All procedures were implemented using the `socialmixr` R package[62;63].

5.5 Modeling number of contacts

5.5.1 Variable description

The modeling analysis included a set of variables commonly identified in the literature as associated with contact behavior. These covariates include age, sex, employment status, day of the week, household size, and educational attainment[14;26;49;64]. Age was categorized into four groups: 0–18, 19–45, 46–65 (reference category), and 66 years and older. Sex was coded as a binary variable (male or female), and employment status was classified into three levels: employed (reference), student, and unemployed. The day of the week on which contacts were reported was included as a binary indicator for whether the day was a weekday (true or false). Household size was included as a continuous variable reflecting the number of people living in the respondent's home. Educational attainment was included as a categorical variable with two levels: high school or less and university education.

Additionally, we included a variable for whether the respondent reported taking medication for a chronic condition, such as asthma, diabetes, cardiovascular or pulmonary disease, renal disease, or conditions involving immunosuppression (true or false). To account for potential temporal variation in contact behavior, we included an indicator variable for whether the contact survey was submitted during a school or public holiday period. Holiday periods were defined based on Italy's national public holidays and the 2023-2024 and 2024-2025 regional academic calendars for Piedmont.

Finally, we made use of the rich data we have on all contact survey participants by matching each contact survey to a corresponding weekly symptom report, a process described in section 5.5.2. Using the ECDC case definition[43], we classified symptom history in the 30 days preceding the contact survey into four categories: no symptoms (reference), symptoms not qualifying as ILI, ILI, and ILI with fever.

5.5.2 Data preparation

To incorporate symptom-related variables into the analysis of contact patterns, it was necessary to merge data from two sources: the contact survey dataset and the weekly syndromic surveillance dataset. We restricted the analysis to contact survey entries from participants who had also submitted at least one symptom survey in the 30 days prior to a contact survey submission, in order to avoid incorrectly assuming that the absence of a symptom report implied an absence of symptoms.

Each contact survey entry was matched with symptom reports from the same participant. Because participants could submit multiple symptom surveys, we linked each participant's contact entry to all of their symptom surveys completed within the 30 days prior to the contact survey date. For each participant and contact date combination, if multiple, we selected the single symptom survey entry that reported the highest number of symptoms within the window, prioritizing the most severe symptom presentation even if it occurred further in the past (e.g., a report of 10 symptoms 15 days prior was preferred over 1 symptom reported 2 days prior). Linking symptoms based on temporal proximity rather than severity did not meaningfully affect descriptive analyses of reported contacts, and we ultimately hypothesized that more severe symptom states would have a stronger influence on contact behavior.

This matching approach ensured that each contact survey record had one corresponding symptom profile, based on the most informative recent symptom report. Participants without a symptom survey in the prior 30 days were excluded from the symptom-based analysis. This exclusion criterion did not substantially affect the sample size, as the majority of contact survey respondents had at least one qualifying symptom survey within the required timeframe. The sample was reduced from 1405 entries from 446 participants to 1404

entries from 443 participants in this step.

Next, information from the intake surveys, described in section 5.2, was linked to the contact survey data. For each contact survey entry, we matched the participant's intake survey that was returned closest in time to the contact survey date. While the dataset was largely complete, there were some missing values in the demographic explanatory variables from the intake survey. To address this, we first attempted to resolve missingness by filling in missing values using participant information from prior intake submissions. After this procedure, a complete case deletion approach was applied to exclude any remaining entries that still had missing values in one or more model covariates. This process reduced the dataset from an original 1405 contact entries from 446 participants to a final analysis set of 1393 entries from 439 participants. This corresponds to an exclusion of approximately 1% of contact entries and 1.5% of participants.

5.5.3 Negative binomial regression model

We modeled the total number of contacts reported in each survey using a zero-inflated negative binomial (ZINB) regression model with participant-level random intercepts, to account for repeated measures. This approach was implemented using the `glmmTMB` package in R[65].

The outcome variable y_{it}, the number of contacts reported by participant i on occasion t, was truncated at 100 to reduce the influence of extreme outliers. Any values greater than 100 were recoded to 100. This affected only a small number of observations and was applied as a data preprocessing step.

To account for the excess of zero-contact days observed in the data, we included a zero-inflation component. A zero-inflated model assumes that observed zeros may arise from two distinct processes. In the context of social contacts, the first is a "structural" process—such as an individual having no opportunity or choosing not to engage in social interaction on a given day. If a person is isolated for a day their probability of having any contacts is zero. The second is a count process, in which a participant is "at risk" of making contacts but may still report zero by chance.

These two components are modeled separately: a binary model estimates the probability of a structural zero, and a count model estimates the number of contacts conditional on not being a structural zero. The log of the expected count is a linear combination of covariates and a random intercept:

$$\log(\mu_{it}) = \beta_0 + \beta_1 x_{1,it} + \beta_2 x_{2,it} + \cdots + \beta_p x_{p,it} + b_i$$

where β_0 is the intercept, $x_{j,it}$ are the covariates, and b_i is a normally distributed participant-specific random intercept.

The zero-inflation component, or the probability of a structural zero, was modeled as a constant, such that:

$$\text{logit}(\pi) = \gamma_0$$

where π is the probability of a structural zero and γ_0 is the intercept for the zero-inflation process.

In a negative binomial model, coefficients are interpreted on the log scale. Exponentiating a coefficient yields a rate ratio, representing the multiplicative change in the expected count associated with a one-unit change in the predictor, holding all other variables constant. For continuous variables, the exponentiated coefficient reflects the expected change in count for each one-unit increase in the variable. For categorical or binary variables, coefficients represent the expected difference between a given category and a reference category. For example, if the coefficient for male (with female as the reference) is 0.2, then males are expected

to have 22% more contacts than females ($e^{0.2} \approx 1.22$), all else equal.

5.6 Matrix comparison metrics

The overall and setting-specific contact matrices from Influweb were compared to those from the POLYMOD and CoMix studies to assess the consistency of Influweb with established prior studies. Direct comparisons between POLYMOD and CoMix were not performed, as these have already been explored in previous work [31].

Matrix similarity was evaluated using complementary metrics, each capturing a different aspect of comparison: structural overlap, pattern similarity, and overall magnitude.

5.6.1 Generalized Sørensen–Dice Index

$$\mathrm{SD}(\mathbf{A}, \mathbf{B}) = \frac{2 \sum_{ij} \min(a_{ij}, b_{ij})}{\sum_{ij} a_{ij} + \sum_{ij} b_{ij}}$$

This metric quantifies overlap between two contact matrices. For each cell, the smaller value is taken and summed across the matrix, then doubled. The result is divided by the total contact volume of both matrices. The index ranges from 0 (no overlap) to 1 (perfect agreement) and reflects the proportion of shared contact volume, independent of absolute scale [66–68].

5.6.2 Cosine Similarity

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \, \|B\|} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2} \, \sqrt{\sum y_i^2}}$$

Cosine similarity measures the directional alignment between two matrices, independent of their magnitude [69;70]. Each matrix is flattened into a vector, and the cosine of the angle between the two vectors is calculated. The result ranges from 0 to 1 for non-negative data, with values closer to 1 indicating that the same matrix positions tend to have relatively high or low values, or both matrices share a similar age-mixing pattern.

5.6.3 Frobenius distance

$$\|A - B\|_F = \sqrt{\sum_{ij} (A_{ij} - B_{ij})^2}$$

Frobenius distance captures the overall magnitude of difference between two contact matrices. It is calculated as the square root of the sum of squared differences between corresponding cells, producing a single non-negative number. Unlike the previous metrics, it is sensitive to the absolute scale: large differences in high-contact cells contribute more to the distance than small differences elsewhere. A value of 0 indicates identical matrices, while larger values reflect greater divergence in contact frequencies.

5.6.4 Comparison Plots

In a similar method to that used in Prem et. al 2021 [24], each contact matrix was normalized so that the scaled matrix has a dominant eigenvalue of 1. After normalization, corresponding cell values, c_{ij}, from the two matrices were plotted: the normalized value from matrix A corresponds to the x-axis and the normalized value from matrix B the y-axis. A regression line was fit through these points and Pearson's correlation coefficient was computed to assess linear agreement in cell pairs across matrices.

5.7 Estimation of bootstrapped R_0 distributions

For each study, we calculated a distribution of R_0 values for a hypothetical respiratory pathogen, such as the SARS-CoV-2 virus. To derive these distributions, we calculated R_0 for each of the bootstrap contact matrices generated during the construction of the final matrices (see section 5.4). For each matrix, R_0 was calculated as:

$$R_0 = \frac{\beta}{\mu} \cdot \rho(\tilde{C}),$$

where β is the transmission rate, μ is the recovery rate, and $\rho(\tilde{C})$ is the spectral radius of the adjusted contact matrix $\tilde{C}$. The transmission (β) and recovery rates (μ) rates were set to 0.03 and 0.2, respectively, to represent a general respiratory pathogen such as SARS-CoV-2, based on values reported in prior literature[71;72].

The matrix $\tilde{C}_{ij}$ is derived by adjusting each matrix element according to the number of susceptible individuals at time zero in age group i ($(S_i(0))$ and the population size of age group j (N_j): $\tilde{C}_{ij} = \left(\frac{S_i(0)}{N_j}\right) C_{ij}$.

We set initial conditions to zero infections in all age groups to minimize applied assumptions. Age group population sizes were drawn from census data corresponding to each study's year of data collection[57;58].

Finally, we conduct a Kolmogorov-Smirnov test to assess whether the underlying distributions of the R_0 values derived from each study differ on a statistical level[73]. While this analysis facilitates comparison of transmission potential across studies, we note that differences in survey methodologies mean we cannot interpret the resulting R_0 values as directly comparable or as absolute measures of infectiousness. This analysis solely serves as a scenario-based comparison, exploring how estimated transmission potential varies under different contact conditions observed across studies.

6 Data availability

The raw data used in this study cannot be shared publicly as they contain individual-level information that could compromise participant privacy. Aggregated data products derived from the raw data, full analysis code, and figure generation scripts are available on GitHub and archived on Zenodo (DOI: 10.5281/zenodo.17434728).

7 Funding

The authors KNK, NG, MM, and DP acknowledge support from the Lagrange Project of the Institute for Scientific Interchange Foundation (ISI Foundation) funded by Fondazione Cassa di Risparmio di Torino (Fondazione CRT). MM, DP, LH and AJVH acknowledge support by the Horizon Europe grant VERDI (101045989). MM, DP and LH acknowledge support by the ESCAPE project (101095619), funded by the European Union. MM and DP acknowledge support from the EU Commission in the framework of the EU Horizon Project SIESTA (101131957). Views and opinions expressed are however those of the authors' only and do not necessarily reflect those of the European Union or the Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

8 Author contribution

All authors contributed to the design of the study. PC, LH, DP, AJVH participated in the design of the original survey. Data were processed and prepared for analyses by NG, KNK, and MM. KNK performed the statistical analysis, with oversight from PC, DP, MM and NG. The first draft of the manuscript was produced by KNK and all authors reviewed, edited, and approved the final version.

References

[1] Wallinga J, Edmunds WJ, Kretzschmar M. Perspective: human contact patterns and the spread of airborne infectious diseases. TRENDS in Microbiology. 1999;7(9):372-7.

[2] Munday JD, Abbott S, Meakin S, Funk S. Evaluating the use of social contact data to produce age-specific short-term forecasts of SARS-CoV-2 incidence in England. PLoS Computational Biology. 2023;19(9):e1011453.

[3] Ogunjimi B, Hens N, Goeyvaerts N, Aerts M, Van Damme P, Beutels P. Using empirical social contact data to model person to person infectious disease transmission: an illustration for varicella. Mathematical biosciences. 2009;218(2):80-7.

[4] Abrams S, Wambua J, Santermans E, Willem L, Kuylen E, Coletti P, et al. Modelling the early phase of the Belgian COVID-19 epidemic using a stochastic compartmental model and studying its implied future trajectories. Epidemics. 2021;35:100449.

[5] Leung K, Jit M, Lau EH, Wu JT. Social contact patterns relevant to the spread of respiratory infectious diseases in Hong Kong. Scientific reports. 2017;7(1):7974.

[6] Wallinga J, Teunis P, Kretzschmar M. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. American journal of epidemiology. 2006;164(10):936-44.

[7] Rohani P, Zhong X, King AA. Contact network structure explains the changing epidemiology of pertussis. Science. 2010;330(6006):982-5.

[8] Eames KT, Tilston NL, Brooks-Pollock E, Edmunds WJ. Measured dynamic social contact patterns explain the spread of H1N1v influenza. PLoS computational biology. 2012;8(3):e1002425.

[9] Lucchini L, Marziano V, Trentini F, Chiavenna C, D'Agnese E, Offeddu V, et al. Post-pandemic social contacts in Italy: implications for distancing measures on in-person school and work attendance. 2025.

[10] Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. The lancet public health. 2020;5(5):e261-70.

[11] House T, Baguelin M, Van Hoek AJ, White PJ, Sadique Z, Eames K, et al. Modelling the impact of local reactive school closures on critical care provision during an influenza pandemic. Proceedings of the Royal Society B: Biological Sciences. 2011;278(1719):2753-60.

[12] Koltai J, Vásárhelyi O, Röst G, Karsai M. Reconstructing social mixing patterns via weighted contact matrices from online and representative surveys. Scientific reports. 2022;12(1):4690.

[13] Di Domenico L, Reichmuth ML, Althaus CL. Individual-based and neighbourhood-based socio-economic factors relevant for contact behaviour and epidemic control. medRxiv. 2025:2025-03.

[14] Mousa A, Winskill P, Watson OJ, Ratmann O, Monod M, Ajelli M, et al. Social contact patterns and implications for infectious disease transmission: A systematic review and meta-analysis of contact surveys. eLife. 2021;10:e70294. Available from: `https://elifesciences.org/articles/70294`.

[15] Manna A, Koltai J, Karsai M. Importance of social inequalities to contact patterns, vaccine uptake, and epidemic dynamics. Nature Communications. 2024;15(1):4137. Supplementary Information available at `https://www.nature.com/articles/s41467-024-48332-y#Sec33`.

[16] Holme P. Objective measures for sentinel surveillance in network epidemiology. Physical Review E. 2018;98(2):022313.

[17] Amico L, Kleynhans J, Gauvin L, Tizzoni M, Ozella L, Wolter N, et al. Estimating household contact matrices structure from easily collectable metadata. 2024.

[18] Kiti MC, Tizzoni M, Kinyanjui TM, Koech DC, Munywoki PK, Meriac M, et al. Quantifying social contacts in a household setting of rural Kenya using wearable proximity sensors. EPJ data science. 2016;5(1):21.

[19] Failla A, Citraro S, Rossetti G. Attributed Stream-Hypernetwork Analysis: a SocioPatterns Case Study. In: SEBD; 2022. p. 383-91.

[20] Zagheni E, Billari FC, Manfredi P, Melegaro A, Mossong J, Edmunds WJ. Using time-use data to parameterize models for the spread of close-contact infectious diseases. American journal of epidemiology. 2008;168(9):1082-90.

[21] Hoang TV, Willem L, Coletti P, Van Kerckhove K, Minnen J, Beutels P, et al. Exploring human mixing patterns based on time use and social contact data and their implications for infectious disease transmission models. BMC Infectious Diseases. 2022;22(1):954.

[22] Moore J, Carrasco JA, Tudela A. Exploring the links between personal networks, time use, and the spatial distribution of social contacts. Transportation. 2013;40(4):773-88.

[23] Mistry D, Litvinova M, Pastore y Piontti A, Chinazzi M, Fumanelli L, Gomes MF, et al. Inferring high-resolution human mixing patterns for disease modeling. Nature communications. 2021;12(1):323.

[24] Prem K, Zandvoort Kv, Klepac P, Eggo RM, Davies NG, for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group C, et al. Projecting contact matrices in 177 geographical regions: an update and comparison with empirical data for the COVID-19 era. PLoS computational biology. 2021;17(7):e1009098.

[25] Hoang T, Coletti P, Melegaro A, Wallinga J, Grijalva CG, Edmunds JW, et al. A systematic review of social contact surveys to inform transmission models of close-contact infections. Epidemiology. 2019;30(5):723-36.

[26] Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS medicine. 2008;5(3):e74.

[27] van Leeuwen E, Group PJM, Sandmann F. Augmenting contact matrices with time-use data for fine-grained intervention modelling of disease dynamics: A modelling analysis. Statistical Methods in Medical Research. 2022;31(9):1704-15.

[28] Meyer S, Held L. Incorporating social contact data in spatio-temporal models for infectious disease spread. Biostatistics. 2017;18(2):338-51.

[29] Iozzi F, Trusiano F, Chinazzi M, Billari FC, Zagheni E, Merler S, et al. Little Italy: an agent-based approach to the estimation of contact patterns-fitting predicted matrices to serological data. PLoS computational biology. 2010;6(12):e1001021.

[30] Feehan DM, Mahmud AS. Quantifying population contact patterns in the United States during the COVID-19 pandemic. Nature communications. 2021;12(1):893.

[31] Tizzani M, De Gaetano A, Jarvis CI, Gimma A, Wong K, Edmunds WJ, et al. Impact of tiered measures on social contact and mixing patterns of in Italy during the second wave of COVID-19. BMC Public Health. 2023;23(1):906.

[32] Verelst F, Hermans L, Vercruysse S, Gimma A, Coletti P, Backer JA, et al. SOCRATES-CoMix: a platform for timely and open-source contact mixing data during and in between COVID-19 surges and interventions in over 20 European countries. BMC medicine. 2021;19:1-7.

[33] Jarvis CI, Coletti P, Backer JA, Munday JD, Faes C, Beutels P, et al. Social contact patterns following the COVID-19 pandemic: a snapshot of post-pandemic behaviour from the CoMix study. Epidemics. 2024;48:100778.

[34] Manna A, Dall'Amico L, Tizzoni M, Karsai M, Perra N. Generalized contact matrices allow integrating socioeconomic variables into epidemic models. Science Advances. 2024;10(41):eadk4606.

[35] Paolotti D, Carnahan A, Colizza V, Eames K, Edmunds J, Gomes G, et al. Web-based participatory surveillance of infectious diseases: the Influenzanet participatory surveillance experience. Clinical Microbiology and Infection. 2014;20(1):17-21.

[36] Guerrisi C, Turbelin C, Blanchon T, Hanslik T, Bonmarin I, Levy-Bruhl D, et al. Participatory syndromic surveillance of influenza in Europe. The Journal of infectious diseases. 2016;214(suppl_4):S386-92.

[37] van Noort SP, Codeco CT, Koppeschaar CE, Van Ranst M, Paolotti D, Gomes MGM. Ten-year performance of Influenzanet: ILI time series, risks, vaccine effects, and care-seeking behaviour. Epidemics. 2015;13:28-36.

[38] Perrotta D, Bella A, Rizzo C, Paolotti D. Participatory online surveillance as a supplementary tool to sentinel doctors for influenza-like illness surveillance in Italy. PloS one. 2017;12(1):e0169801.

[39] Gozzi N, Perrotta D, Paolotti D, Perra N. Towards a data-driven characterization of behavioral changes induced by the seasonal flu. PLoS computational biology. 2020;16(5):e1007879.

[40] Kelley K, Gozzi N, Mazzoli M, Paolotti D. Exploring influenza vaccination determinants through digital participatory surveillance. BMC Public Health. 2025;25(1):1-14.

[41] Wong KL, Gimma A, Coletti P, Faes C, Beutels P, Hens N, et al. Social contact patterns during the COVID-19 pandemic in 21 European countries–evidence from a two-year study. BMC infectious diseases. 2023;23(1):268.

[42] Bansal S, Read J, Pourbohloul B, Meyers LA. The dynamic nature of contact networks in infectious disease epidemiology. Journal of biological dynamics. 2010;4(5):478-89.

[43] European Centre for Disease Prevention and Control. EU case definitions for communicable diseases; 2012. Accessed: 2025-05-15. `https://www.ecdc.europa.eu/en/all-topics/eu-case-definitions`.

[44] Angeli L, Caetano CP, Franco N, Abrams S, Coletti P, Van Nieuwenhuyse I, et al. Who acquires infection from whom? A sensitivity analysis of transmission dynamics during the early phase of the COVID-19 pandemic in Belgium. Journal of Theoretical Biology. 2024 Mar;581:111721. Available from: `https://www.sciencedirect.com/science/article/pii/S002251932400002X`.

[45] Goodfellow L, Quilty BJ, van Zandvoort K, Edmunds WJ. Post-pandemic social contact patterns in the United Kingdom: the Reconnect survey. medRxiv. 2025:2025-08. Preprint.

[46] Tizzani M, Gauvin L. Socioeconomic determinants of protective behaviors and contact patterns in the post-COVID-19 pandemic era: A cross-sectional study in Italy. PLOS Computational Biology. 2025;21(8):e1013262.

[47] Dietz E, Pritchard E, Eyre DW, Peto TE, Stoesser N, Matthews PC, et al. Social behaviours and contact patterns across the 2020/21, 2021/22 and 2022/23 winter seasons in the UK, and associations with symptoms. medRxiv. 2025:2025-08. Preprint.

[48] Van Kerckhove K, Hens N, Edmunds WJ, Eames KT. The impact of illness on social networks: implications for transmission and control of influenza. American journal of epidemiology. 2013;178(11):1655-62.

[49] Coletti P, Wambua J, Gimma A, Willem L, Vercruysse S, Vanhoutte B, et al. CoMix: comparing mixing patterns in the Belgian population during and after lockdown. Scientific Reports. 2020;10(1):21885. Available from: `https://www.nature.com/articles/s41598-020-78540-7`.

[50] Capacci G, Rinesi F. An overview of demographic ageing in Italy. Ageing, lifestyles and economic crises: The new people of the mediterranean. 2017.

[51] Dietz K. The estimation of the basic reproduction number for infectious diseases. Statistical methods in medical research. 1993;2(1):23-41.

[52] Koppeschaar CE, Colizza V, Guerrisi C, Turbelin C, Duggan J, Edmunds WJ, et al. Influenzanet: Citizens among 10 countries collaborating to monitor influenza in Europe. JMIR public health and surveillance. 2017;3(3):e7429.

[53] Bolt K, Gil-González D, Oliver N. Unconventional data, unprecedented insights: leveraging non-traditional data during a pandemic. Frontiers in Public Health. 2024;12:1350743.

[54] Bokányi E, Vizi Z, Koltai J, Röst G, Karsai M. Real-time estimation of the effective reproduction number of COVID-19 from behavioral data. Scientific Reports. 2023;13(1):21452.

[55] Smit T, Carstens G, Han W, Bulsink K, de Bakker J, Elahi M, et al. Flexible and scalable participatory syndromic and virological surveillance for respiratory infections: Our experiences in The Netherlands. PLoS One. 2025;20(8):e0303230.

[56] Greenacre ZA. The importance of selection bias in internet surveys. Open Journal of Statistics. 2016;6(3):397-404.

[57] Italian National Institute of Statistics (ISTAT). Intercensal Register Population Estimates: 2002–2018; 2024. Intercensal population estimates by sex, age, and municipality based on the 2001 and 2011 censuses. Data used for demographic analysis, projections, and socio-economic surveys. Accessed: 2025-05-15. https://siqual.istat.it/SIQual/visualizza.do?id=8888959&refresh=true&language=EN.

[58] Italian National Institute of Statistics (ISTAT). Resident Municipal Population by Age, Sex, and Marital Status: 2020–2024; 2024. Annual resident population estimates by age, sex, and marital status based on the Permanent Census. Includes methodological constraints and aggregation rules for municipalities under 20,000 inhabitants. Accessed: 2025-05-15. https://esploradati.istat.it/databrowser/#/it/dw/categories/IT1,POP,1.0/POP_POPULATION/DCIS_POPRES1/IT1,22_289_DF_DCIS_POPRES1_1,1.0.

[59] Arregui S, Aleta A, Sanz J, Moreno Y. Projecting social contact matrices to different demographic structures. PLoS computational biology. 2018;14(12):e1006638.

[60] Hamilton MA, Knight J, Mishra S. Examining the influence of imbalanced social contact matrices in epidemic models. American journal of epidemiology. 2024;193(2):339-47.

[61] Davison AC, Hinkley DV. Bootstrap methods and their application. 1. Cambridge university press; 1997.

[62] Funk S. socialmixr: Social Mixing Matrices for Infectious Disease Modelling; 2024. R package version 0.1.9. Available from: https://CRAN.R-project.org/package=socialmixr.

[63] Funk S. Introduction to socialmixr; 2025. Accessed via https://epiforecasts.io/socialmixr/articles/socialmixr.html. R package vignette, socialmixr.

[64] Wambua J, Loedy N, Jarvis CI, Wong KLM, Faes C, Grah R, et al. The influence of COVID-19 risk perception and vaccination status on the number of social contacts across Europe: insights from the CoMix study. BMC Public Health. 2023 Jul;23(1):1350. Available from: https://doi.org/10.1186/s12889-023-16252-z.

[65] Brooks ME, Kristensen K, van Benthem KJ, et al. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. The R journal. 2017;9(2):378-400.

[66] Sorensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. Biologiske skrifter. 1948;5:1-34.

[67] Lenormand M, Huet S, Gargiulo F, Deffuant G. A universal model of commuting networks. PLoS ONE. 2012.

[68] Lenormand M, Bassolas A, Ramasco JJ. Systematic comparison of trip distribution laws and models. Journal of Transport Geography. 2016;51:158-69.

[69] Salton G. Associative document retrieval techniques using bibliographic information. Journal of the ACM (JACM). 1963;10(4):440-57.

[70] Schneider JW, Borlund P. Matrix comparison, Part 1: Motivation and important issues for measuring the resemblance between proximity measures or ordination results. Journal of the American Society for Information Science and Technology. 2007;58(11):1586-95.

[71] Zeng Y, Guo X, Deng Q, Luo S, Zhang H. Forecasting of COVID-19: spread with dynamic transmission rate. Journal of Safety Science and Resilience. 2020;1(2):91-6.

[72] Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science. 2020;368(6493):860-8.

[73] Schröer G, Trenkler D. Exact and randomization distributions of Kolmogorov-Smirnov tests two or three samples. Computational Statistics & Data Analysis. 1995;20(2):185-202.

9 Supporting information

S1 Fig. Monthly number of completed survey responses (February 2024–February 2025).

S2 Fig. Distribution of survey responses per participant, showing how many participants returned a given number of surveys.

S3 Fig. Age-structured contact matrices from Influweb, disaggregated by setting. Age-structured contact matrices from the Influweb study, disaggregated by setting: home, work, school, and leisure. Each cell shows the per capita daily contact rate between age groups i and j. Contact matrices were constructed using post-stratification weights for age, sex, and day of week, and symmetrized using the Italian national population structure for the year of data collection. Scales are specific to each setting.

S4 Fig. Age-structured contact matrices from POLYMOD, disaggregated by setting. Age-structured contact matrices for the POLYMOD study, disaggregated by setting: home, work, school, and leisure. Each cell shows the per capita daily contact rate between age groups i and j. Contact matrices were constructed using post-stratification weights for age, sex, and day of week, and symmetrized using the Italian national population structure for the year of data collection. Scales are specific to each setting.

S5 Fig. Age-structured contact matrices from CoMix, disaggregated by setting. Age-structured contact matrices for the CoMix study, disaggregated by setting: home, work, school, and leisure. Each cell shows the per capita daily contact rate between age groups i and j, symmetrized using the Italian national population structure for the year of data collection. Scales are specific to each setting.

S6 Fig. Normalized comparison of Influweb and POLYMOD contact matrices by setting. Comparison of cell values from Influweb and POLYMOD contact matrices by setting (home, work, school, and leisure). Each point represents a unique age group pairing. Matrix values are normalized such that the dominant eigenvalue equals one, resulting in unitless relative contact intensities. The dashed line is a fitted regression; x-axes show Influweb values and y-axes show corresponding POLYMOD values.

S7 Fig. Normalized comparison of Influweb and CoMix contact matrices by setting. Comparison of cell values from Influweb and CoMix contact matrices by setting (home, work, school, and leisure). Each

point represents a unique age group pairing. Matrix values are normalized such that the dominant eigenvalue equals one, resulting in unitless relative contact intensities. The dashed line is a fitted regression; x-axes show Influweb values and y-axes show corresponding CoMix values.

S1 Table. Responses to optional survey question on daily routine changes from illness. Distribution of responses to the optional survey question "Did you change your daily routine because of your illness?".