Opening of hotels and ski facilities: impact on mobility and Covid-19 outcomes

Krzysztof Zaremba*

June 2021

Abstract

This paper investigates how the opening of hotels and ski facilities in Poland impacted mobility and Covid-19 outcomes. Leveraging geolocation data from Facebook, we show that ski resorts experienced a significant influx of tourists, increasing the number of local users by up to 50%. Furthermore, we show that there was an increase in the probability of meetings between pairs of users from distanced locations and pairs of users from touristic and non-touristic areas. As the policy impacted travels and gatherings, we then analyze its effect on the diffusion of Covid-19. We find a significant association between touristic movements and the severity of a major pandemic wave in Poland. In particular, we observe that counties with ski facilities experienced more infections after the opening. Moreover, counties strongly connected to the ski resorts during the opening had more subsequent cases than weakly connected counties.

1 Introduction

Tourism creates long-distance movements of the population and as such can contribute to the spread of infectious diseases. Simultaneously, tourists generate significant income for local economies. Hence, any decisions concerning touristic activity require balancing the trade-off between economic and public health outcomes. Yet, there is currently no evidence quantifying the impact of touristic mobility on the diffusion of infections. This paper aims to fill this gap by analyzing the effects of the reopening of Polish tourism on mobility and the spread of Covid-19.

Our study design takes advantage of a unique policy that reopened all hotels and ski lifts in Poland. On the 12th of February 2021, the Polish government reopened ski lifts and hotels at 50% capacity and with food supplied through room service only. At the same time, authorities reopened only cinemas, theatres, and operas at 50% capacity with mandatory masks[1]. The hotels and ski lifts remained open until the 20th of March, when the second wave of infections ravaged the country.

This policy caused significant movements of tourists. Using aggregated and anonymized geolocation data from Facebook (FB), we show that the policy's implementation increased the number of FB users in ski resorts by 25%-50%. Moreover, there has been a surge in the probability that users from non-touristic and touristic areas meet and in the probability that users living far away meet.

While the number of travels and tourist gatherings increased due to the policy, its effect on Covid-19 outcomes is a priori not obvious. On the one hand, visitors could carry the disease from their homes to touristic locations and back. On the other hand, they could only gather in their rooms or outdoors as restaurants were closed. To learn more about the impact of the policy on Covid-19 cases, we derive theoretical predictions from Suscetpible-Infected-Recovered-Deceased model. Next, we test models implications by leveraging granular infection data and the uncofoundedness approach from Callaway and Li (2021) that

*We thank Douglas Almond and Daniel Mangrum for invaluable comments. We thank Facebook Data for Good Initiative for making their data available. We thank Alex Pompe for assistance with data. Acknowledgments: Map data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org

[1]Such entertainment facilities rarely operate in ski resorts

we extend to a network setting. We find that counties with ski resorts see additional cases already in the third week after the opening. Such an early effect is absent in other counties. Moreover, counties strongly interacting with ski resorts during the policy have a higher incidence of infections than counties with weak interactions. This is consistent with a secondary spread from tourists bringing the virus back home.

In the remainder of the paper, we first relate our work to the existing literature (section 2). Then, in section 3, we explain what data we use throughout the paper. The following section (4) discusses empirical methods and results regarding impact of the policy on mobility. Analogously, section 5.1 examines the effect of the policy on Covid-19 outcomes. The conclusion closes the paper.

2 Related literature

The epidemic has sparked a large body of literature related to Covid-19, however relatively little research has been dedicated to the effect of tourism on infections. Researchers have shown that industry closures and stay-at-home orders have a limiting impact on both mobility and subsequent Covid-19 outcomes (Fang et al. (2020), Gupta et al. (2020), Lyu and Wehby (2020), Beria and Lunkar (2021), Courtemanche et al. (2020), Abouk and Heydari (2020), Badr et al. (2020), Lau et al. (2020), Morley et al. (2020), Xu et al. (2020), Goolsbee and Syverson (2021)). Fewer papers analyzed reopening. An exception is Nguyen et al. (2020) who found that lifting restrictions led to 6-8% increase in mobility.

Opening the tourism industry can lead to travels and large gatherings, and there has been some evidence relating these phenomena to the viral spread. One of the seminal papers on this topic is Adda (2016), which shows that school vacation and transportation strikes disrupt viral transmission. More recently, Chernozhukov et al. (2021), Andersen et al. (2021), Courtemanche et al. (2021), Bravata et al. (2021) , Goldhaber et al. (2021) link school and colleges operating modality to the local prevalence of infections. However, the opening of schools can have substantially different effects than tourism opening as children are less likely to suffer severe consequences of Covid-19 (Castagnoli et al. (2020); Dong et al. (2020)).

Closer to our population of interest (adults) are studies analyzing sport, social, and political gatherings. Large sporting events such as hickey, basketball and football game can lead to higher Covid-19 prevalence (Carlin et al. (2021), Alexander et al. (2020), Breidenbach and Mitze (2021)). Similarly, smaller gathering such as birthdays or bar meetings increase likelihood of subsequent infections (Harris (2020), Whaley et al. (2021)). The evidence in the case of political gatherings is mixed. Palguta et al. (2021) find an increase in the growth rate of Covid-19 in areas with elections, while Dave et al. (2020) conclude that a political rally in Tulsa did not affect local Covid cases. They note, however, that the local population enhanced social distancing, which could offset the effect of the gathering. This compensatory behavior is unlikely to have taken place for tourism opening because local population populations will interact with incoming tourists by providing them hospitality services.

Tourism can also encourage long-distance travel, and it has been shown that travelers contribute to the diffusion of infections. Mangrum and Niekamp (2020) provide evidence that college students returning from spring break trips accelerated the local spread of Covid-19 and the related mortality. Burlig et al. (2021) analyzes how the length of the travel ban matters for subsequent Covid-19 outcomes and shows an empirical association between migrants traveling home and the following number of cases in their home area. Finally, people attending large events with little protective behavior such as Capitol Riot (Dave et al. (2021b)) or Sturgis Motorcylce Rally (Dave et al. (2021a)) bring the disease when traveling back to their home counties.

As the literature shows, long-distance travel and large gatherings are both associated with the diffusion of Covid-19. The opening of tourism can encourage both travels and gatherings, and hence it is potentially highly relevant for the viral spread. To understand opening consequences and guide future policy, our paper firstly establishes whether the opening of ski resorts and hotels contributed to the long-distance travel and gatherings. Secondly, it aims to answer whether the reopening of tourism accelerated the spread of Covid-19

3 Data

We compile a unique dataset featuring mobility patterns of Facebook users and administrative data on Covid-19 related outcomes from the Polish Ministry of Health.

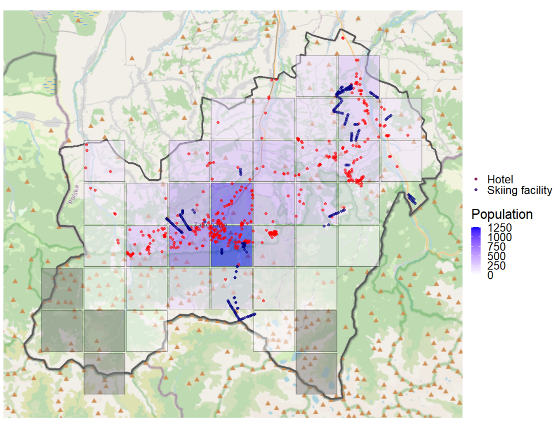
3.1 Mobility

The data on mobility comes from Facebook's project Data for Good Initiative[2]. It originates from Facebook's users who enabled the location services on their devices. We use two measures of mobility: population and collocation probabilities.

Population The population at time window t and tile A is defined as the number of users who were logging mostly from the tile A during the time window t. There are three time windows per 24 hours (with breaks at 00:00, 08:00, and 16:00 UTC) and tiles are approximately 3km x 3km. Observations with less than ten users are omitted for privacy reasons. We assign tiles to counties based on their centroids. As an example, consider the map on the figure 1a. It represents the population logging on the tiles covering the Tatrzanski county - a popular tourist destination - on the afternoon (17:00 - 01:00 ETC) of the 14th of February 2021. The red dots and navy dots indicate the location of hotels and ski facilities, respectively. We use hotels and ski facilities' locations [3] to classify tiles as touristic or non-touristic. Figure shows the time series of the number of users in Tatrzanski county. There is a clear uptick in the FB population after hotels' opening. See appendix table 1a and figure 12a for summary statistics of the population data.

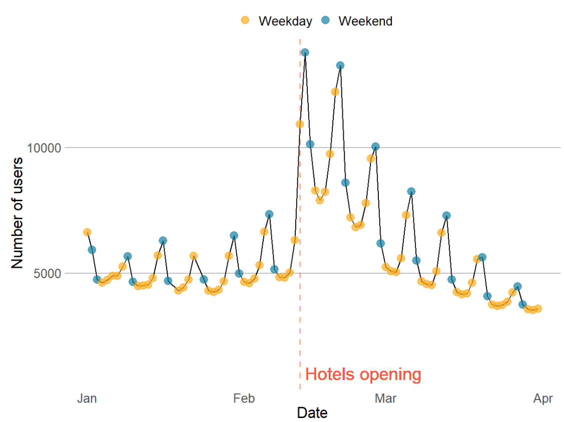
Figure 1: FB Population data

(a) Hotels and population in Tatrzanski county on the afternoon of the 14th February 2021



Note: The color of a tile represents the population, i.e., the number of users logging from the tile. Grey tiles correspond to no records. Red dots represent coordinates of hotels, and navy dots represent coordinates of skiing facilities. Source: OpenStreetMap and own elaboration based on Facebook data

(b) Number of FB users logging in between 17:00 and 01:00 in Tatrzanski county



Note: The count of users in Tatrzanski county represents the sum of users from tiles with centroids in the county. Users logging in multiple tiles during the 8-hour window are assigned to the modal tile. Source: Own elaboration based on Facebook data

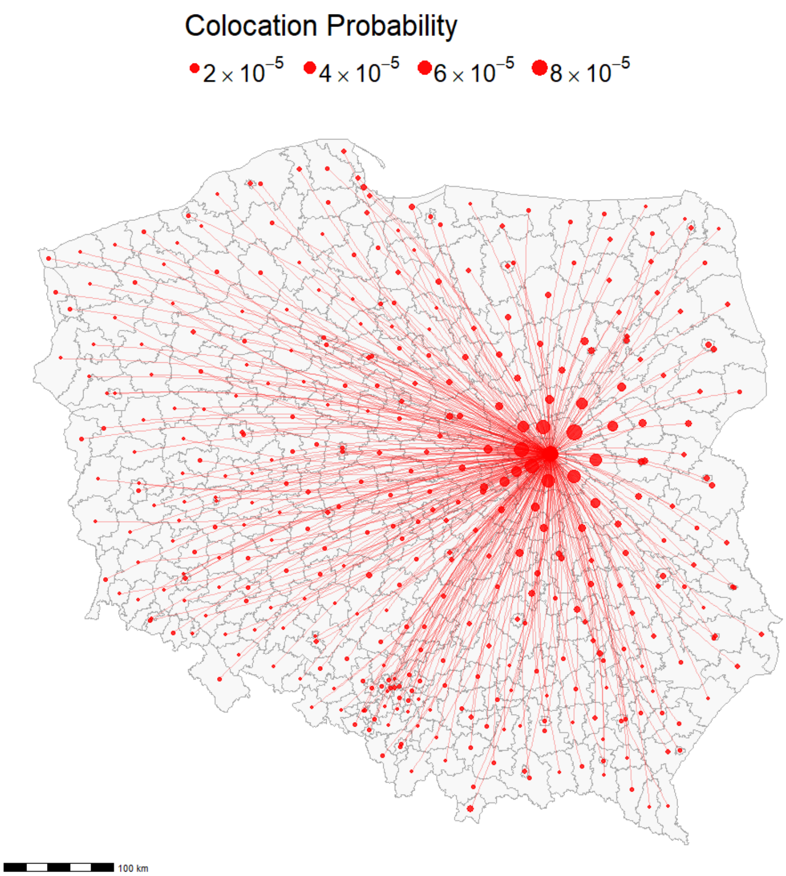
Colocation Colocation data aims to approximate how often users from different regions meet. Technically, it measures the probability that two randomly chosen users from county i and j were within the same location, i.e., the same 0.6km $\times$ 0.6km tile at a randomly chosen 5-minute interval of a given week (Wednesday to

[2] Data provided by the Facebook's Data for Good Initiative: https://dataforgood.fb.com/

[3] We use the name "hotel" for any accommodation facility. We find hotels and ski facilities' coordinates from the OpenStreetMaps project (?). The location of ski facilities were scraped from www.narty.pl and validated through own search

Tuesday) [4] We do not consider collocation between users from the same county. Note that the colocation probabilities are small in magnitude because the denominator is the number of possible pairs of users from the two counties multiplied by the number of 5 minutes intervals in a week. The user's county of residence is derived from a consistent history of night-time locations. The map in figure 2 illustrates the collocation probabilities of users from Warsaw with users from other counties in the week ending on the 16th of February 2021. The size of a dot and the transparency of a link is proportional to the collocation probability between Warsaw and the given county. See appendix table 1b and figure 12b for summary statistics of the colocation data.

Figure 2: Colocation of users from Warsaw and other counties in the week ending on the 16th of February 2021



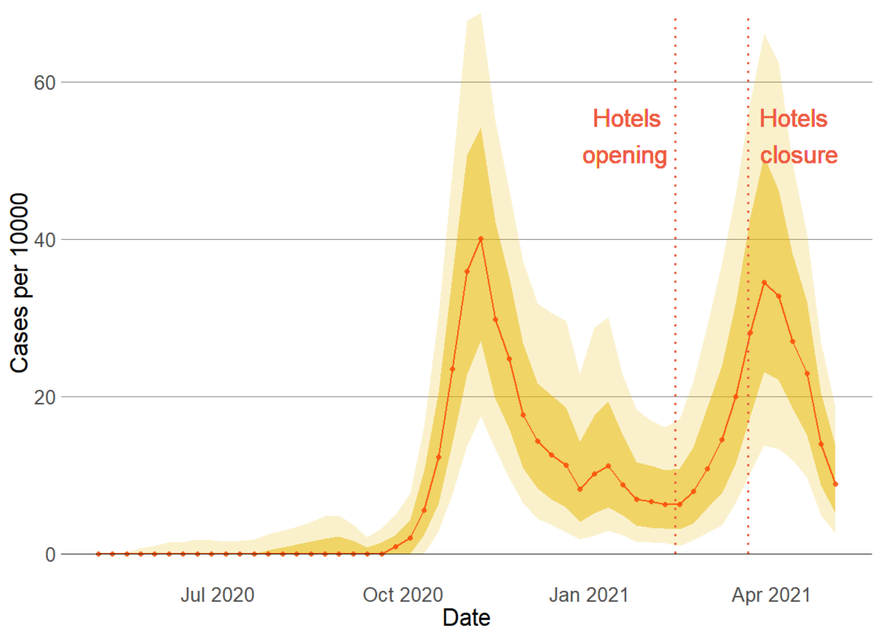
Note: Size of red dots and transparency of curved links are proportional to the colocation probabilities between users from Warsaw and given county Source: Own elaboration based on Facebook data

[4]Formally, and ignoring the week index, let X_{tir} be the number of users from region r at tile t in the 5 minute time interval i. Then let m_{rs} to be the sum of meetings between pairs of individuals from region r and s across all tiles and time intervals, that is $m_{rs} = \sum_{ti} X_{tir} X_{tis}$. The colocation probability is then the ratio of all actual meetings and all potential meetings, that is: $Pr(Colocation_{rs}) = \frac{m_{rs}}{2016 n_r n_s}$, where n_r is the number of users from region r and 2016 is the number of all five-minute intervals in a week. See Iyer et al. (2020) for more details

3.2 Health outcomes

The Polish Ministry of Health provided the data on health outcomes. It contains weekly observations at the commune (*gmina*) level for the number of Covid-19 cases, deaths, tests taken, and people vaccinated with two doses. Commune is the lowest administrative unit in Poland, and they usually correspond to a town or a couple of villages. There are 2477 communes with a median population of 7486. Our data covers May 2020-April 2021 for cases and tests, while remaining variables are available for the period January 2021-April 2021. During this period, Poland experienced its second major wave of infections. Figure 3 visualizes the the evolution of the pandemic in Poland by showing the median, 10^{th}, 25^{th}, 75^{th} and 90^{th} quantiles of weekly cases per 10 000 inhabitants. As one can see, the number of new infections varied considerably in the temporal and cross-sectional dimensions. Interestingly, the timing of the second wave coincided with the opening of hotels.

Figure 3: Quantiles of weekly cases per 10 000 inhabitants in Poland



Note: The lighter shaded area corresponds to the 10^{th} and 90^{th} quantiles. The darker area corresponds to the 25^{th} and 75^{th} quantiles. The red line and points represent the median. Source: Own elaboration based on the data from the Ministry of Health

4 Mobility outcomes

This section shows that the opening of hotels significantly increased mobility, especially at long distances. A sharp influx to touristic areas raised the frequency of meetings between inhabitants of touristic and non-touristic counties.

4.1 Empirical framework: the impact of the policy on population movements

Using differential tourist accommodation capacity and proximity to ski resorts, we conduct an event study to learn whether the reopening of hotels increased the inflow of tourists. In particular, we analyze whether the number of users on tiles with many hotels increased more after the policy's implementation compared to tiles with no hotels. Furthermore, we stratify the analysis by whether the tiles are close to ski facilities. We hypothesize that the policy induced a large influx of tourists into ski resorts, noting that places with many hotels attract more tourists due to their grater capacity. Moreover, stratifying the analysis by the proximity to ski resorts partly alleviates the concern that the number of hotels proxies high urbanization as ski resorts are usually located within small towns.

To implement this strategy, we locate any accommodation and ski facilities and assign them to tiles. We then calculate the number of hotels in each tile. We bin tiles into five categories: 0 hotels, 1 hotel, 2 to 9 hotels, 10-19 hotels, and 20 or more hotels. Next, we define tiles as *in proximity to a ski resort* if they are within 25km of the closet ski lift in the mountains. Then we estimate the following regression separately for the tiles in proximity and not in proximity to ski resorts:

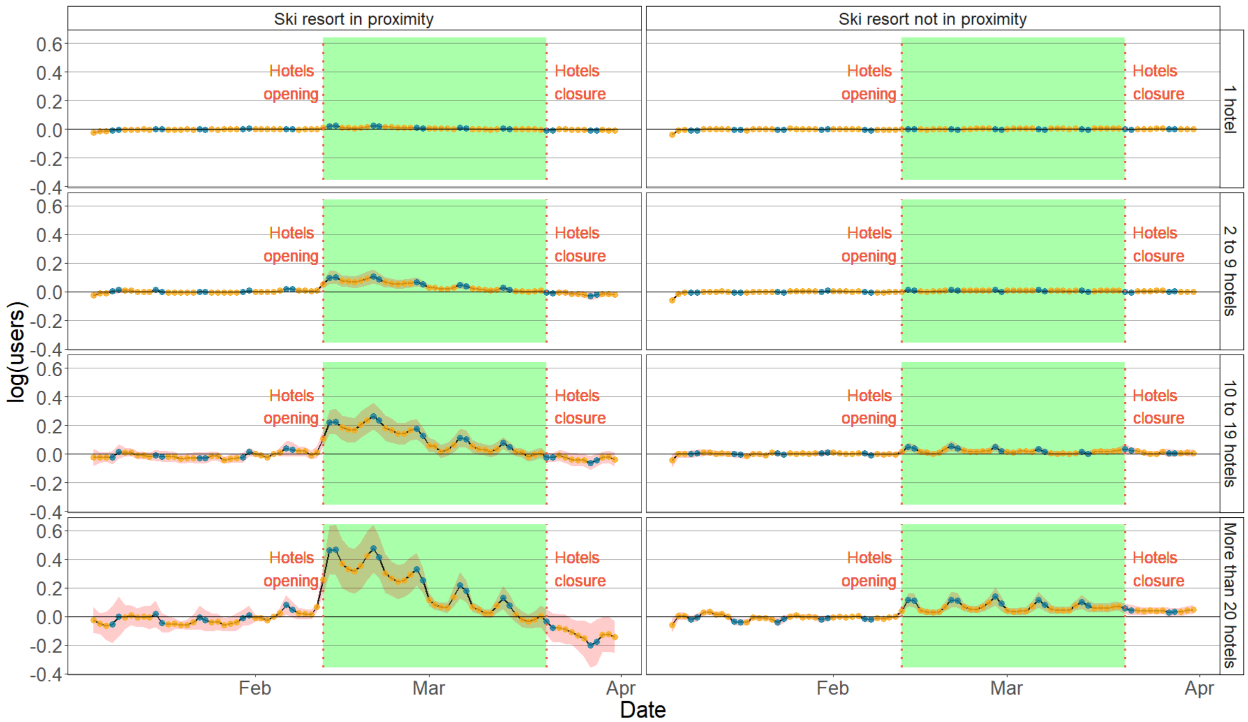
$$log(population)_{jtp} = \sum_{\substack{h \in \{\{1\},\{2:9\} \\ \{10:19\},\{20+\}\}}} \sum_{\substack{T \in \{\{01/06:02/03\}, \\ \{02/05:03/31\}\}}} Hotels_h^j I(t=T)\beta_h^T + \lambda_{jp} + \sum_{dw \in DW} \alpha_{i(j)}^{dw} + \gamma_{tp} + \epsilon_{jtp} \tag{1}$$

The event study in equation 1 analyzes the change in the population on tiles with h hotels at date t compared to an analogous change in tiles with 0 hotels. The baseline period is the 4th of February, which is excluded from time dummies. The outcome variable is the natural logarithm of the population at tile j, date t, and time window p. The dummy $Hotels_h^j$ takes value 1 if the tile j has a number of hotels in the bin h. The parameter of interest is β_h^t, and we expect it to be 0 before the policy (12th of February), and positive after the policy. Moreover, β_h^t should increase with the number of hotels. The increase should be considerably larger in proximity to ski resorts if the movements are tourism-related. We allow for the tile×time-window fixed effects $\lambda_j p$, weekday×county fixed effects $\alpha_{i(j)}^{dw}$, and date×time-window fixed effects $\gamma_t p$. We cluster the standard errors at the county level.

4.2 Hotels opening led to increase in mobility in touristic areas

The reopening of hotels increased the population present at ski resorts considerably. Figure 4 displays β_h^t coefficients. We see that β_h^t is 0 before the policy date (dashed line), confirming that trends were parallel before the reopening in all types of tiles. However, after the 12th of February, we see high growth in the number of users present in tiles with hotels, especially in the proximity to ski resorts (left panel). The growth is also higher for tiles with more hotels. For tiles with more than 20 hotels and in proximity to ski resorts, we see about a 50% increase in the population during weekends and a 30% increase during weekdays. The effects subside with time as Poland enters its second wave of the pandemic. While there is a significant increase in population at tiles with many hotels and not in proximity to ski resorts (right bottom panel), its magnitude is small. We conclude that there was a large influx of tourists to ski resorts after the opening.

Figure 4: Event study: Hotels and ski resorts reopening and Facebook users



Note: Lines and points correspond to the estimates of β_h^T from equation 1. The excluded category is *tiles with 0 hotels*, and the excluded date is the 4th of February. The panel on the left represents the estimates for tiles in the proximity to ski resorts, the panel on the right shows estimates for the remainder of the tiles. Estimates of β_h^T for each bin h are plotted separately, starting with the lowest bin h at the top and the highest bin h at the bottom. Blue points correspond to weekends and yellow to weekdays. Red shaded area plots 95% confidence bands, which allows for clustering at the county level. The green rectangle represents the time when the hotels and ski resorts were open. Source: Own elaboration based on Facebook data

4.3 Empirical framework: the impact of the policy on the frequency of meetings

In this section, we investigate whether the policy affected the frequency of meetings between users from different counties. Such meetings are essential from an epidemiological perspective because they can transform local outbreaks into a national wave. Our hypothesis is that the reopening policy made population flows at long distances and flows from non-touristic to touristic counties more likely. To test this, we perform two analyses. The first analysis investigates whether the frequency of long-distance meetings increased relative to short-distance meetings after the policy was enacted. We classify each link into five distance bins based on the distance between centroids of the counties. The bins are $d \in D = 0-100km, 100-200km, 200-300km, 300-400km, 400+km$. Next, we regress the log of colocation probabilities on the interaction of the week dummies with the distance bins:

$$log(P(collocation))_{klw} = \sum_{\substack{W \in \{\{01/05:02/02\},\\ \{02/16:04/13\}\}}} \sum_{d \in \underline{D}} Distance_{kl}^{d} I(w=W)\beta_d^W + \phi_{kl} + \chi_w + \upsilon_{klw} \quad (2)$$

Where $log(P(collocation))_{klw}$ is the log of the probability of collocation between users from county k and l in the week w. A dummy $Distance_{kl}^{d}$ is equal to 1 if the distance between counties l and k is in the bin d.

The bin with the shortest distance is excluded as a reference. The dummy $I(w = W)$ is equal to one if the week of the observation corresponds to the week W. The excluded week is the last week before the opening, that is, the week of 02/09. We allow for link ϕ_{kl} and week fixed effects χ_w, and we cluster the standard errors at the link level. The parameter of interest is β_d^W which is a difference-in-differences estimator: first difference measures the percentage change in the collocation probabilities for counties at a distance d in the week w compared to the week of 02/09. The second difference takes this change and compares it to an analogous change for counties at a distance within 0-100km. We expect β_d^W to be positive for weeks when the policy was in place as people started to travel long distances.

The second analysis tests the hypothesis that meetings between non-touristic and touristic counties increased after the policy was implemented. We rely on the assumption that counties with large skiing and accommodation capacities have greater touristic appeal (see the derivations of the theoretical impact of policy on colocation in the appendix). Hence, we define the exposure to tourists by using the total length of skiing trails and the number of hotel beds in the county[5]. In particular, we first classified counties as below ($few\ hotels$) or above ($many\ hotels$), the third quartile of the distribution of hotel beds. Second, we classify counties with 0 skiing trails as 0 $trials$. Finally, for counties with some skiing trails, we divide them by whether they are below ($few\ trails$) or above (($many\ trails$) the third quartile of the distribution of the total length of skiing trails [6]. In this way, we obtain five possible exposure statuses $es \in ES$={*0 trails & few hotels beds, 0 trails & many hotel beds, Few trails & few hotel beds Few trails & many hotel beds, Many trails & Many hotel beds*} [7]. See Figure 15 for the spatial distribution of the exposures.

We expect that the probability of collocation between inhabitants of touristic and non-touristic regions to have increased after the policy's implementation. The most substantial effect should exist for pair of counties with and without ski resorts. To measure such effect, we conduct our analysis on the link level (links between counties) by running the following regression:

$$
\begin{aligned}
log(P(collocation))_{klw} = & \sum_{\substack{W \in \{\{01/05:02/02\},\\ \{02/16:04/13\}\}}} \sum_{s \in \underline{ES}} \sum_{q \in \underline{ES}} Exposure_k^s \times Exposure_l^q I(w = W)\beta_{sq}^W \\
& + \delta_{kl} + \gamma_w + \epsilon_{klw}
\end{aligned} \tag{3}
$$

A dummy $Exposure_k^s$ is equal to 1 if the county k belongs to the exposure category s and 0 otherwise. The excluded combination of classes is the one between counties which both belong to *0 trails & few hotels beds*. Analogously to equation 2, we exclude the week of 02/09, and we allow for link δ_{kl} and week fixed effects γ_w, and we cluster the standard errors at the link level. The parameter of interest β_{qs}^W estimates the percentage change in the collocation probabilities between users from counties of types s and q in the week w compared to the week of 02/09 and relative to an analogous change for users from two different counties both belonging to the type *0 trails & few hotels beds.* Note that the links are undirected, hence we use β_{qs}^W independently of which county is in q and which in s. We expect β_{qs}^W to be positive after policy implementation for pairs s and q such that one has ski trails and hotels while the other does not. Moreover, the effect should grow in the difference between the counties touristic appeals and hence should see the most prominent effect for pair *0 trails & few hotels beds* and Many trails & Many hotel beds}.

4.4 Policy increased likelihood of long distance meetings and meetings between locals and tourists

The frequency of long-distance meetings and meetings between non-touristic and touristic counties increased after the opening. Figure 5 shows the parameter of interest from the equation 2. We see a clear increase in colocations at long distances compared to counties within 100km after the policy was enacted. Moreover,

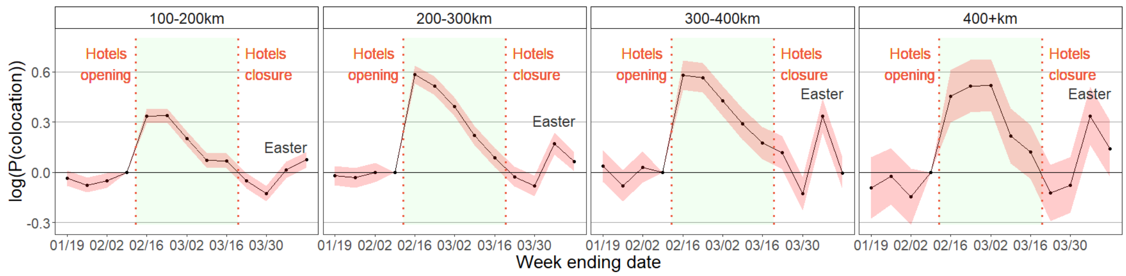
[5]Data from the Polish Statistical Agency for the year 2019

[6]Among counties with any skiing trails

[7]All counties with long skiing trails have many hotels beds, hence exposure *Few hotel beds & Many trails* is missing

the increase was greater for distances above 200km, which is consistent with tourists going further as they can stay for the night in a hotel. The parameter of interest decreases after the initial surge, which may be related to the rising number of Covid cases in late March. Furthermore, we see a spike in the first week of April, which likely corresponds with Easter festivities.

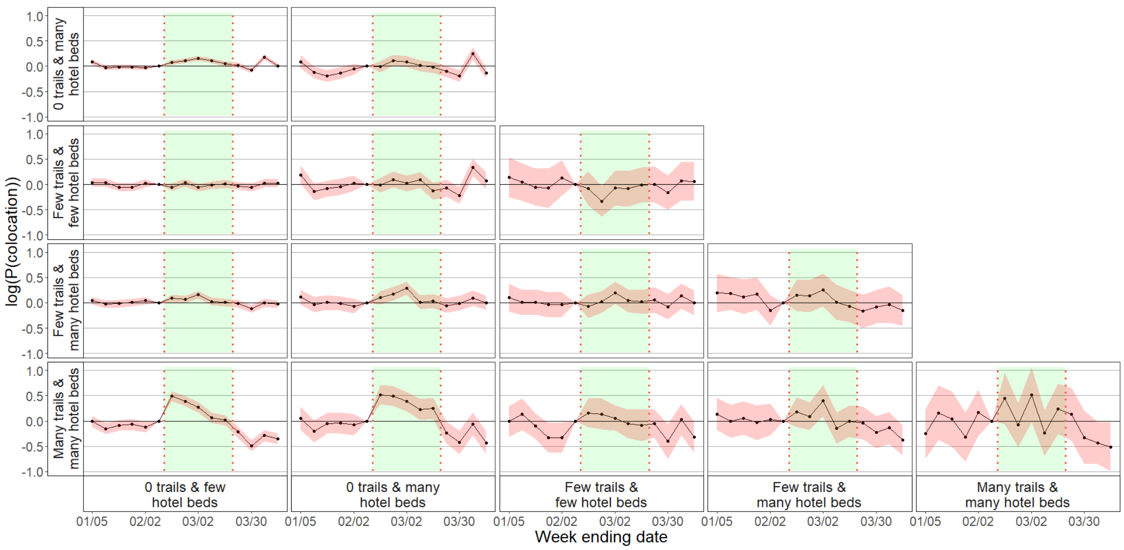
Figure 5: Event study: Hotels and ski resorts reopening and long distance colocation



Note: Lines and points correspond to the estimates of β_d^W from equation 2. The excluded category is *Distance < 100km* and the excluded week is 02/09. Each panel represents estimates of β_d^W for a different distance bin b starting with the lowest distance on the left. Red shaded area plots 95% confidence bands, which allows for clustering at the link level. The green rectangle represents the time when the hotels and ski resorts were open. Additionally, an annotation is added to mark the week of Easter. Source: Own elaboration based on Facebook data

Turning our attention to equation 3, figure 5 shows an increase in colocation between tourists and locals. Each panel in figure 5 corresponds to estimates of parameters β_{qs}^W for different s and q. The row label represents category s, and the column label represents category q. For example, the top-left panel represents the change in the colocation probabilities between counties with *0 trails & few hotels beds* and with *0 trails & many hotels beds*. As expected, we do not see significant changes in the frequency of meetings between pairs of counties that are either both non-touristic (that is, have 0 or few trails) or both touristic. While the null effect among non-touristic counties is accurately estimated, we obtain noisy estimates for links between counties that both have trails. This is due to a lower number of touristic counties and, hence, fewer connections among them. Importantly, we see a significant increase in the probability of meetings between counties with 0 trials and counties with many trials immediately after the opening. The magnitude of this increase is around 50% in the week following the the opening and stays positive for three weeks.

Figure 6: Event study: Hotels and ski resorts reopening and touristic colocation



Note: Lines and points correspond to the estimates of β^{W}_{sk} from equation 3. The excluded category is one with both counties belonging to *0 trails & few hotels beds* and the excluded week is 02/09. Each panel represents estimates of β^{W}_{sk} for a different pair of s and k types. Note that the ordering of types does not matter because links are symmetric. The types are described in the strips on the left and on the bottom. For example, the bottom left panel represents β^{W}_{sk} where one county belongs to *0 trails & few hotels beds* and the other to *many trails & many hotels beds*. Red shaded area plots 95% confidence bands, which allows for clustering at the link level. The green rectangle represents the time when the hotels and ski resorts were open. Additionally, we add an annotation to mark the week of easter. Source: Own elaboration based on Facebook data

We conclude that the policy increased the frequency of long-distance meetings and that of meetings related to tourism. As such, it could have a significant impact on Covid-19 outcomes.

5 Covid-19 outcomes

5.1 Theoretical framework

Theoretically, as the policy encourages gatherings and travel, it should impact the number of infections by increasing the number of contacts between individuals. Suppose that infections follow the Susceptible-Infected-Recovered-Deceased model (SIRD). Let us denote infected population in commune k in week w by $I_{k,w}$. Furthermore, let $S_{k,w}$ be the number of susceptible individuals in commune k in week w and $P_{k,w}$ its population. Then, let d_{kl} be the number of contacts of a person from commune l with people in commune k. Then, the new infections evolve according to the following equation:

$$I_{k,w} - I_{k,w-1} = \underbrace{\frac{S_{k,w-1}}{P_{k,w-1}}\beta d_{kk} I_{k,w-1}}_{\text{New cases from within commune}} + \underbrace{\frac{S_{k,w-1}}{P_{k,w-1}}\beta \sum_{l \neq k} d_{kl} I_{l,w-1}}_{\text{New cases from out of commune}} - \underbrace{(1-\kappa)\gamma I_{k,w-1}}_{\text{Recoveries}} - \underbrace{\kappa\gamma I_{k,w-1}}_{\text{Deaths}} \tag{4}$$

This model assumes that the level of infections can change for three reasons. Firstly, new cases can increase the number of infected individuals. The transmission in any contact between an infected and a susceptible individual occurs with the probability β. Each infected individual from commune k meets d_{kk} other individuals in commune k, but only a share $\frac{S_{k,w-1}}{P_{k,w-1}}$ of them are susceptible. Hence, each infectee in a commune k produces on average $\frac{S_{k,w-1}}{P_{k,w-1}}\beta d_{kk}$ secondary infections within their commune. Analogously, each infectee can produce secondary infections in other communes if they travel outside of their home. Secondly, prevalence decreases because some people recover, where recovery occurs with probability $(1-\kappa)\gamma$ [8]. Finally, infections disappear when people die, which happens with probability $\kappa\gamma$.

We assume that tourism opening affected parameters d_{kl}. Denote d'_{kl} as the new contact rates. In particular, $d'_{kl} > d_{kl}$ and $d'_{lk} > d_{lk}$ where k is a touristic commune, and l is a non-touristic commune. This increase stems from new interactions between tourists and locals. To find the model implications of this policy, I compare the potential outcomes of a commune with and without the policy. In particular, denote $Y(1)_{k,w*}$ the potential outcome that a commune k would experience in the week after the opening had the policy taken place. In our case, potential outcome corresponds to the change in infections $I(1)_{k,w*} - I_{k,w*-1}$. Analogously, let $Y(0)_{k,w*}$ be the potential outcome of the commune k in the same time period had the policy not taken place. The difference between the two potential outcomes measures the treatment effect $\tau_{w*,k}$ of the policy. In particular:

$$\begin{aligned}
\underbrace{Y(1)_{k,w*} - Y(0)_{k,w*}}_{\tau_{k,w*}} &= I(1)_{k,w*} - I_{k,w*-1} - (I(0)_{k,w*} - I_{k,w*-1}) \\
&= \underbrace{\beta \frac{S_{k,*w-1}}{P_{k,w*-1}} \sum_l d'_{kl} I_{l,w*-1} - \gamma I_{k,w*-1}}_{I(1)_{k,w*} - I_{k,w*-1}} \\
&\quad - \underbrace{(\beta \frac{S_{k,*w-1}}{P_{k,w*-1}} \sum_l d_{kl} I_{l,w*-1} - \gamma I_{k,w*-1})}_{I(0)_{k,w*} - I_{k,w*-1}} \\
&= \beta \frac{S_{k,*w-1}}{P_{k,w*-1}} \sum_l (d'_{kl} - d_{kl}) I_{l,w*-1}
\end{aligned} \tag{5}$$

Model implies that the treatment effect *on impact* consists of the direct effect of the change in the interaction rates. The effect will differ by the type of the commune. Suppose there are three types: communes not sending tourists to ski resorts (nt), communes sending tourists to ski resorts (tr) and communes which contain ski resorts (s). The treatment effect for the first type is null in the week after the policy:

$$\tau_{nt,w*} = \beta \frac{S_{k,w*-1}}{P_{k,w*-1}} \sum_l (d'_{nt,l} - d_{nt,l}) I_{l,w*-1} = 0$$

These communes do not experience direct effects of the policy because their contact rates have not changed: $d'_{nt,l} - d_{nt,l}$ $\forall l$. Next, tr communes experience a positive treatment effect:

$$\tau_{tr,w*} = \beta \frac{S_{tr,*w-1}}{P_{tr,w*-1}} (\sum_s d'_{tr,s} - d_{tr,s}) I_{s,w*-1}$$

These communes have worse epidemic outcomes as tourists get infected in ski resorts. Finally, the treatment effect for ski resort communes is:

$$\tau_{s,w*} = \beta \frac{S_{s,w*-1}}{P_{s,w*-1}} \sum_{tr} (d'_{s,tr} - d_{s,tr}) I_{tr,w*-1}$$

It represents new infections brought by tourists [9] . While the treatment effects are simple on impact, they become more complicated in later periods as new infections spread through existing networks. For example, the treatment effect in commune k in the second week after the opening is [10]:

[8] We assume that recovered individuals do not become susceptible again within the short timeframe of our analysis

[9] Note that the model predicts new infections, which are detected with a lag and hence visible in the data later than in the first week of the policy

[10] We assume for simplicity that the susceptible share does not change in the short term after the policy to the extend in which it could affect new infections

$$\tau_{k,w*+1} = \beta \frac{S_{k,*w-1}}{P_{k,w*-1}} \sum_l [d'_{kl}\tau_{l,w*} + (d'_{kl} - d_{kl})I(0)_{l,w*}] - \gamma\tau_{k,w*}$$

This effect is composed of three terms. The first element in the sum represents spillover effects. Additional infections from the previous period spread through the existing networks. The second term represents the direct effect of the policy if it is still in place. That is, new infections occur due to the increased contact rate between tourists and locals in ski resorts. Finally, the last term represents recoveries of additional infections caused by the policy in the previous period. Looking at the effects by commune type, we first notice that the treatment effect is no longer zero in *nt* communes:

$$\tau_{nt,w+1*} = \beta \frac{S_{nt,*w-1}}{P_{nt,w*-1}} \sum_l d'_{nt,l}\tau_{l,w*}$$

An *nt* commune experiences spillovers as long as it is connected to some communes which had direct effects of the policy in the previous period. The structure of the treatment effects is similar in *tr* and *s* communes:

$$\tau_{tr,w*+1} = \beta \frac{S_{tr,*w-1}}{P_{tr,w*-1}} [\sum_l d'_{tr,l}\tau_{l,w*} + (d'_{tr,s} - d_{tr,s})I(0)_{s,w*}] - \gamma\tau_{tr,w*}$$

$$\tau_{s,w*+1} = \beta \frac{S_{s,*w-1}}{P_{s,w*-1}} [\sum_l d'_{s,l}\tau_{l,w*} + \sum_{tr}(d'_{s,tr} - d_{s,tr})I(0)_{tr,w*}] - \gamma\tau_{s,w*}$$

Both types experience spillover effects and direct effects. Note that the spillover effects may dominate direct effects for units with many connections. Similarly, the spillovers effects are higher for communes with high rate of internal contacts d'_{kk}, as this amplifies treatment effects from the previous period.

5.2 Empirical framework

Translating the theoretical framework to an empirical framework poses three main challenges. First, we need to identify communes sending tourists. Second, we need to adjust for differences in the pre-policy pandemic outcomes across communes. Third, we need to find a valid counterfactual in the realm of a nation-wide policy.

5.2.1 Tourists origin communes

We use colocation data to identify communes sending tourists to ski resorts. In particular, we sum all meetings between individuals from county i and any county containing ski resort [11] during the first three weeks of the policy. Mathematically, the strength of connections, which we call *exposure* is:

$$exposure_i = \frac{\sum_{w \in \{w*, w*+1, w*+2\}} \sum_{j \in ski\_resort} Meetings_{ijw}}{N_i}$$

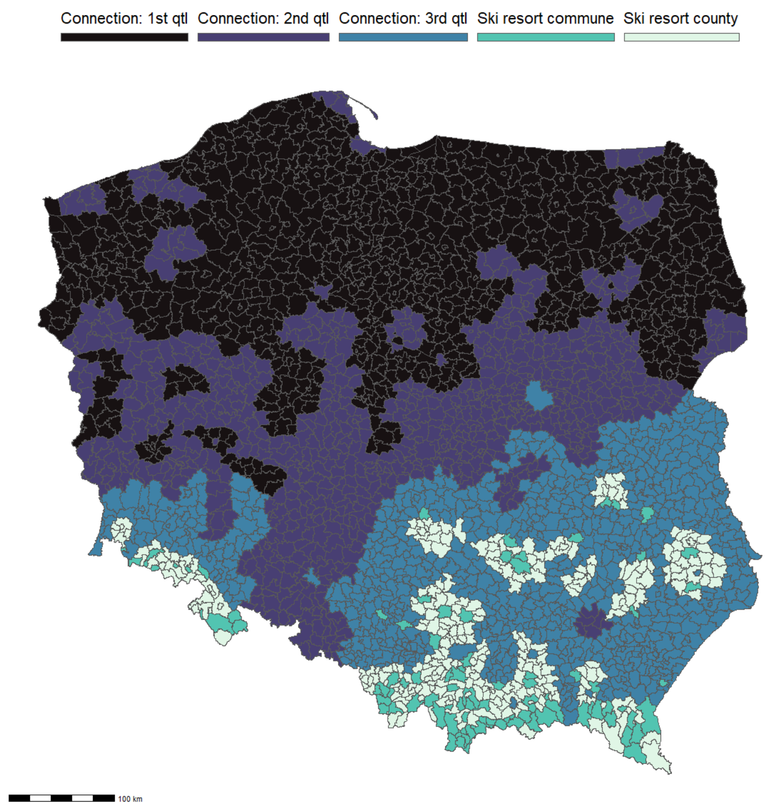
where N_i is the number of FB users in county i and *ski_resort* is the set of counties with ski resorts. Note that the connection measure is at a lower geographic granularity (county) than the outcome (commune). Hence, we assume that all communes in a county have the same *exposure*. We divide communes into 5 categories which correspond to types. Communes with ski resorts are classified as *Ski resort commune*. Communes without ski facilities but in the counties with ski resorts are classified as *Ski resort county* [12]. The rest of the communes are classified according to the tertiles of the *exposure* of their county. We assume that the counties in the first tertile did not send any tourists because they had fewest interactions with inhabitants

[11] We exclude ski facilities not located in the mountains

[12] This distinction is necessary because we do not know connections within the county

of ski resorts during the policy. We further assume that counties in the third tertile sent more tourists than counties in the second tertile. The map on figure 7 shows the spatial distribution of the treatments [13].

Figure 7: Exposure to ski resorts



Note: Each color corresponds to a different exposure category. Communes within the same county belong to the same exposure unless the county contains a ski resort. Communes containing ski facilities are in the category *Ski resort commune* and communes without ski facilities but in counties containing ski resort communes are in the category *Ski resort county.* Source: Own elaboration

For the preliminary empirical evidence on the impact of the policy, we compare new cases in communes with different treatments. Figure 8a plots the average number of new cases per 10 000 inhabitants in communes by their exposure to the policy. Firstly, one notes that the pre-policy infection trends differ between the treatment arms. Secondly, we do not see a significant difference in the number of new infections right after the enacted policy (first dotted line). However, such a simple comparison may be insufficient to uncover the impact of the policy. As shown in equation 4, the diffusion is affected not only by the number of contacts but also by the number of previous infections and susceptible individuals. Different treatments seem indeed to be at a different stage of the pandemic before the policy, as evidenced in figure 8a. This by itself is enough to render a simple comparison unreliable.

5.2.2 Balancing communes by pre-policy outcomes

To analyze the policy, we need to compare locations at a similar pandemic stage before the policy. As SIRD has markovian properties, the current pandemic situation should depend only on the indices in the previous period. The variables representing number of infections, share of susceptibles and number of

[13]Note that the connections are at the county levels, but the outcomes are at the commune level. Hence, communes in the same counties will have the same exposure

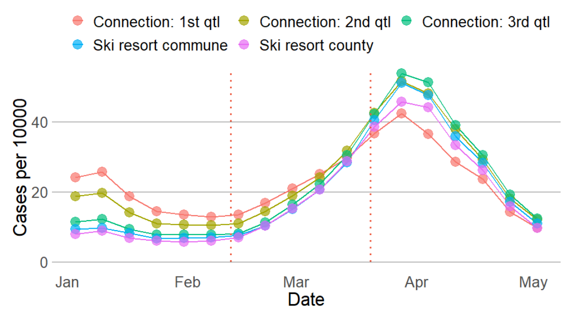
infections among connected units in the current period act as sufficient statistics for the outcome in the next period. Hence, the communes with similar pre-pandemic outcomes and characteristics should evolve in the same way in the absence of the policy. Hence, our analysis conditions on a set of pre-policy variables $F_{k,w*-1} = \{X_{k,w*-1}, Z_k\}$. In particular, $X_{k,w*-1}$ represents pandemic outcomes in the last period before the policy. It includes the number of cases, deaths, and tests per 10 000, their growth rates, and their cumulative numbers since the beginning of the pandemic to proxy for share susceptible[14]. It also contains the weighted sum of cases among neighbors where the weights correspond to the number of commuters per capita. Moreover, $X_{k,w*-1}$ includes the squares and interactions of all these variables. The controls Z_k represent observed location-specific characteristics which may influence the diffusion. In particular, Z_k includes population, population density, the type of the commune (urban or rural), and unemployment rate at the end of 2020. The remainder analysis relies on the assumption that in the absence of the policy and conditional on these variables, the potential outcomes would evolve in the same way across all the types:

Assumption 1 *-Ignorability :* $Y(0)_{k,w \geq w*} \perp\!\!\!\perp Treatment_k^{tp} | X_{k,w*-1}, Z_k$ *for* $tp \in \{nt, tr, s\}$

Figure 8b plots the average number of new cases by type in the sample matched on the conditioning variables. We notice a considerable improvement in the similarity of the pre-policy outcomes, suggesting that the matching was successful. However, the outcomes diverge in the weeks after the opening. In particular, the number of cases is higher in the first weeks after the opening in locations with ski resorts and communes strongly connected to ski resorts.

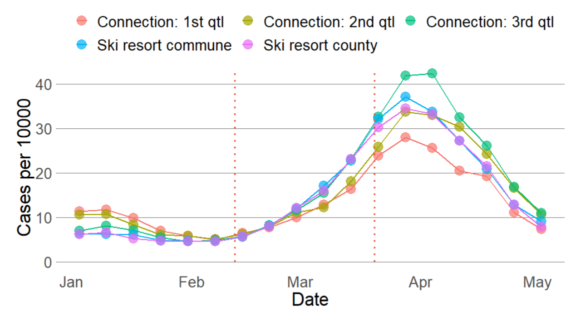
Figure 8: Covid cases in ski resorts

(a) Covid cases: full sample



Note: The averages were calculated on the full sample. Dotted lines represent the opening and closure of hotels. The date corresponds to the last day of the week. Source: Own elaboration based on the Ministry of Health Data

(b) Covid cases: matched sample



Note: The averages were calculated on the matched sample. Each treated unit was matched with one unit from the first tertile of connection. Units were matched by the distance in the propensity scores computed based on the conditioning variables. Dotted lines represent the opening and closure of hotels. The date correspond to the last day of the week. Source: Own elaboration based on the Ministry of Health Data

5.2.3 Identifying control communes

The policy is nationwide, however the exposure to the policy differed between the communes. Using the model and ignorability assumption, we argue that the least exposed communes (nt) can act as controls to identify the *on impact* effect of the policy. Taking the difference between actual outcomes in s and nt communes we obtain:

$$Y(1)_{s,w*} - Y(1)_{nt,w*} = Y(1)_{s,w*} - Y(0)_{nt,w*} = Y(1)_{s,w*} - Y(0)_{s,w*} = \tau_{s,w*}$$

[14]Since the beginning of 2021 in case of deaths, as the number of deaths is only available for 2021

where the first equality follows from the null treatment effect on impact in nt communes and the second equality follows from the ignorability assumption. Hence, we are able to identify on impact treatment effect in communes with ski resorts and analogously in communes sending tourists. The identification is more challenging in future periods because the treatment effect is no longer null in nt communes:

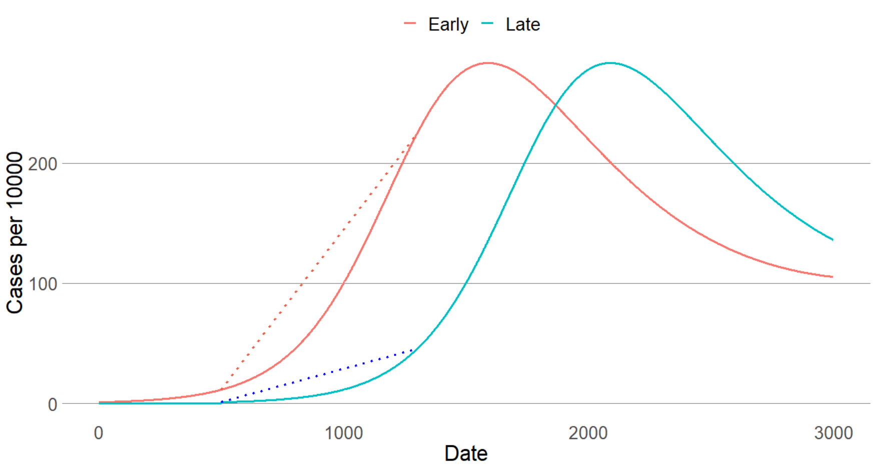
$$Y(1)_{s,w*+1} - Y(1)_{nt,w*+1} = \tau_{s,w*+1} - \underbrace{\tau_{nt,w*+1}}_{0}$$

Hence the comparison is biased in later periods and the bias is larger if nt communes are well connected and hence experience large spillovers. However, model implies that $\tau_{nt,w*+1}$ is non-negative. Consequently, if the difference between potential outcomes is positive, we can identify lower bounds on the treatment effect.

5.2.4 Estimation

To estimate the effects, we turn to the unconfoundedness approach suggested by Callaway and Li (2021), because it performs better than fixed effects estimation in the context of the SIRD model[15]. It has been shown that fixed effects estimation (Gauthier (2021),Callaway and Li (2021)) is unlikely to produce reliable results if outcomes are generated by a non-linear model such as SIRD. This occurs because treated units can be at a different pandemic stage than control units. For example, suppose that treated units experienced their first case earlier than the control units. Then, their outcomes will not evolve in parallel to the outcomes in the control areas, even in the absence of any policy. Figure 9 illustrates this problem. It plots results of a simple simulation of the SIRD model in two areas which are identical except for the timing of their first case. The area represented with the red curve started the pandemic earlier. Dotted lines show linear trends in the number of cases between two points in time. Despite identical parameters, the trends are not parallel because these areas are at different pandemic stages. Hence, differencing trends would introduce a bias rather than remove it.

Figure 9: Non linearities implied by SIRD model



Note: figure plots the results of a simulation of SIRD model in two areas. The parameters are identical in both cases, however the timing of the first case is different. The area represented by red curve had the first case earlier than the area represented by the blue curve. Dotted lines measure linear trends in cases between two points in time. Despite identical parameters, areas experience non-parallel trends

The unconfoundedness approach alleviates the above issue in two ways. Firstly, it does not rely on fixed effects. Secondly, it ensures that the control units are at a similar pandemic stage before the policy. This

[15] We perform traditional event studies on the full and matched samples, see the results in the appendix

is achieved by conditioning on the pre-treatment covariates related to the pandemic. Thus, this approach is compatible with a case in which pandemic-related parameters vary over time and with commune-specific characteristics. It is, however, not compatible with a general unobserved heterogeneity in parameters by location. Hence, the estimation presented below is valid under the assumption that the parameters change across locations only due to the variation in the conditioning controls $F_{k,w*-1}$.

Intuitively, the unconfoundedness approach computes a weighted difference of outcomes in treated communes versus control communes which have similar pre-treatment characteristics to the treated units. The larger the similarity is, the higher is the weight for the control unit. In particular, denote the pre-treatment characteristics as $F_{k,w*-1} = \{X_{k,w*-1}, Z_k\}$ and $C_{w,k}$ as the number of new cases in week w and commune k. Following the notation from Callaway and Li (2021), we estimate:

$$ATT_w^c = E[w(treatment_k^c, F_{k,w*-1})(C_{w,k} - m_{0,w}^C(F_{k,w*-1}) \tag{6}$$

where ATT_w^c corresponds to the average effect on the treated by treatment c, and the weights correspond to:

$$w(treatment_k^c, F_{k,w*-1}) = \frac{treatment_k^c}{E[treatment_k^c]} - \frac{\frac{p(k,F_{w*-1})}{1-p(k,F_{w*-1})}(1 - treatment_k^c)}{\left[\frac{p(F_{k,w*-1})}{1-p(F_{k,w*-1})}(1 - treatment_k^c)\right]}$$

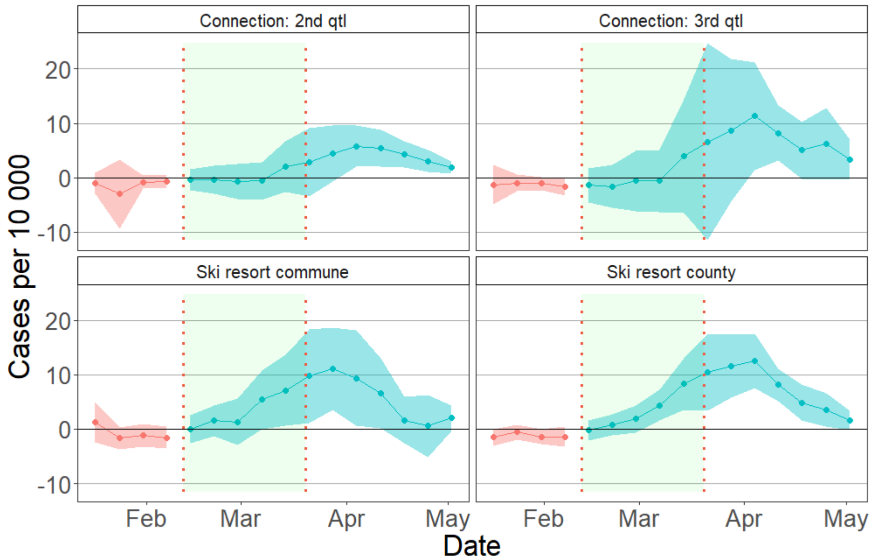
Finally, the untreated potential outcomes correspond to

$$m_{0,w}^C(F_{w*-1}) = E[C_w | F_{k,w*-1}, treatment_k^c = 0]$$

.

The untreated potential outcome of a treated unit k in a week w comes from this weeks' outcomes of untreated units similar to k. Their expectation is unbiased for the control potential outcome under the assumption 1. Concretely, we implement this method by estimating propensity scores $p(F_{k,w*-1})$ with logit and outcome regression $m_{0,w}^C(F_{k,w*-1})$ with OLS. The method is double robust because it is robust to the misspecification in either propensity scores or conterfactual regression of potential outcomes. Note that we estimate four effects: one for each treatment versus the control group of communes in the first tertile of *exposure*. Hence, we estimate the effect in four samples where each sample contains units from the control and one of four treatments. Figure 10 plots the results of the estimation together with 95% confidence bands calculated with multiplier bootstrap.

Figure 10: Event Stydy: Covid-19 and tourism opening



Note: The estimates come from the unconfoundedness approach by Callaway and Li (2021). Red points correspond to the estimates for pre-treatment periods. The reference point in a pre-treatment period t is the previous period $t-1$. Blue points correspond to the estimates for post-treatment periods. The post-treatment periods' propensity scores and outcome regression are based on the last period before the policy $w*-1$. The shaded area represents simultaneous 95% confidence bands with clustering at the commune level. The date corresponds to the last day of the week.

The results are consistent with the hypothesis that the policy precipitated the arrival of the infections' wave, and that tourists brought the disease back to their home counties. Namely, communes in counties with ski resorts experienced a higher number of infections after the opening. They have about five additional infections per 10 000 in the third week after the opening comparing to the control. The effect is statistically and economically significant as it represents a 70% increase with respect to the pre-policy average in these communes. We interpret this as the on impact treatment effect. It is not obvious why communes without ski facilities but in counties with ski resorts experienced impact equivalent to ski resort communes. There may be many interactions between the communes with ski resorts and without ski resorts in the same county. For instance, people may be going skiing or to work in nearby ski resorts. Figure 13 in the appendix shows that commuting to ski resorts communes is stronger than across no ski-resort communes, but magnitude of commutes are not large.

The early increase in cases is absent in communes in the second and the third tertile of $exposure$. However, communes with medium and strong connections see an increase in infections compared to control starting in the fourth week after the opening. This could be the result of secondary infections from tourists bringing the disease back home and spillovers from other communes. Moreover, we see a monotonicity of this effect in the strength of connection: communes in the third tertile have a higher increase in cases than communes in the second tertile. While this is consistent with the story of tourists contributing to the diffusion of the virus, these differences are not statistically significant.

The above analysis provides suggestive evidence that the opening of hotels contributed to the diffusion of COVID-19 through touristic gatherings and travels. In particular, the wave of infections arrived earlier in the counties with ski resorts. Moreover, there were more infections in counties with ski resorts and counties strongly connected to ski resorts than counties weakly connected to ski resorts. Note, however, that this exercise does not allow us to conclude whether the opening of tourism caused or did not cause the second wave of the pandemic in Poland. While we find differential trends in infections by exposure to tourism, there are no units that would not be affected by the policy at medium or long term. Hence, there is no plausible counterfactual which would allow for the evaluation what would happen in the absence of the policy at medium ot=r long term.

6 Conclusion

Tourism plays a vital role in providing income for many local economies. However, while important for economic reasons, it also encourages long-distance travel and gatherings. Moreover, touristic services often require in-person interactions. These features make tourism a transmission vector for various infectious diseases. We hope that our analysis will provide some guidance for policymakers struggling to balance the trade-off between economic and public health goals related to the opening of the tourism industry.

In this paper, we investigated how the opening of hotels and ski facilities impacted mobility and Covid-19 outcomes. The opening was followed by large movements of tourists to locations with ski facilities. Areas with many hotels in proximity to the ski trails experienced an exceptionally high influx of visitors. Travels often originated from distant locations, and hence the probability of meetings between individuals residing far from each other increased after the policy. Additionally, there has been an increase in meetings between pairs of individuals such that one person lives in a touristic and one in a non-touristic location. These observations point out the strong impact of tourism opening on mobility.

Travelers have a high potential to carry the disease between distant locations. This is particularly dangerous when they also participate in gatherings. Visitors of ski resorts could only gather in their hotel rooms and on trails as the restaurants were closed. Nonetheless, there is suggestive evidence that they impacted the Covid-19 trajectory. We showed that having a ski facility in a county is correlated with an increase in infections after the policy. Moreover, counties with frequent meetings with ski resorts during the opening had more infections than counties with few such meetings.

We note that the policy was enacted before the full distribution of Covid-19 vaccines. These could potentially mitigate the impact of tourism on Covid-19. Nonetheless, long-distance travellers have a high potential to carry novel variants to new locations. Future research could explore whether tourism activity is associated with a faster arrival of new variants.

References

Abouk, R. and Heydari, B. (2020). The Immediate Effect of COVID-19 Policies on Social Distancing Behavior in the United States. *medRxiv*, page 2020.04.07.20057356.

Adda, J. (2016). Economic Activity and the Spread of Viral Diseases: Evidence from High Frequency Data *. *The Quarterly Journal of Economics*, 131(2):891–941.

Alexander, A., Martin, H., and Lackner, M. (2020). Mass gatherings contributed to early COVID-19 spread: Evidence from US sports. Working Paper 2003, Working Paper.

Andersen, M. S., Bento, A. I., Basu, A., Marsicano, C. R., and Simon, K. (2021). College Openings in the United States Increased Mobility and COVID-19 Incidence. Technical report. Type: article.

Badr, H. S., Du, H., Marshall, M., Dong, E., Squire, M. M., and Gardner, L. M. (2020). Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet Infectious Diseases*, 20(11):1247–1254.

Beria, P. and Lunkar, V. (2021). Presence and mobility of the population during the first wave of Covid-19 outbreak and lockdown in Italy. *Sustainable Cities and Society*, 65:102616.

Bravata, D., Cantor, J. H., Sood, N., and Whaley, C. M. (2021). Back to School: The Effect of School Visits During COVID-19 on COVID-19 Transmission. Working Paper 28645, National Bureau of Economic Research.

Breidenbach, P. and Mitze, T. (2021). Large-scale sport events and COVID-19 infection effects: evidence from the German professional football 'experiment'. *The Econometrics Journal*, (utab021).

Burlig, F., Sudarshan, A., and Schlauch, G. (2021). The Impact of Domestic Travel Bans on COVID-19 is Nonlinear in Their Duration. Working Paper 28699, National Bureau of Economic Research.

Callaway, B. and Li, T. (2021). Policy Evaluation during a Pandemic.

Carlin, P. R., Minard, P., Simon, D. H., and Wing, C. (2021). Effects of large gatherings on the COVID-19 epidemic: Evidence from professional and college sports. *Economics & Human Biology*, 43:101033.

Castagnoli, R., Votto, M., Licari, A., Brambilla, I., Bruno, R., Perlini, S., Rovida, F., Baldanti, F., and Marseglia, G. L. (2020). Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infection in Children and Adolescents: A Systematic Review. *JAMA Pediatrics*, 174(9):882–889.

Chernozhukov, V., Kasahara, H., and Schrimpf, P. (2021). The Association of Opening K-12 Schools and Colleges with the Spread of COVID-19 in the United States: County-Level Panel Data Analysis. Technical report. Type: article.

Courtemanche, C., Garuccio, J., Le, A., Pinkston, J., and Yelowitz, A. (2020). Strong Social Distancing Measures In The United States Reduced The COVID-19 Growth Rate. *Health Affairs (Project Hope)*, 39(7):1237–1246.

Courtemanche, C. J., Le, A. H., Yelowitz, A., and Zimmer, R. (2021). School Reopenings, Mobility, and COVID-19 Spread: Evidence from Texas. Working Paper 28753, National Bureau of Economic Research.

Dave, D., McNichols, D., and Sabia, J. J. (2021a). The contagion externality of a superspreading event: The Sturgis Motorcycle Rally and COVID-19. *Southern Economic Journal*, 87(3):769–807.

Dave, D. M., Friedson, A. I., Matsuzawa, K., McNichols, D., Redpath, C., and Sabia, J. J. (2020). Risk Aversion, Offsetting Community Effects, and COVID-19: Evidence from an Indoor Political Rally. Working Paper 27522, National Bureau of Economic Research.

Dave, D. M., McNichols, D., and Sabia, J. J. (2021b). Political Violence, Risk Aversion, and Non-Localized Disease Spread: Evidence from the U.S. Capitol Riot. Working Paper 28410, National Bureau of Economic Research.

Dong, Y., Mo, X., Hu, Y., Qi, X., Jiang, F., Jiang, Z., and Tong, S. (2020). Epidemiology of COVID-19 Among Children in China. *Pediatrics*, 145(6).

Fang, H., Wang, L., and Yang, Y. (2020). Human Mobility Restrictions and the Spread of the Novel Coronavirus (2019-nCoV) in China. Technical Report w26906, National Bureau of Economic Research.

Gauthier, G. (2021). On the Use of Two-Way Fixed Effects Models for Policy Evaluation During Pandemics.

Goldhaber, D., Imberman, S. A., Strunk, K. O., Hopkins, B., Brown, N., Harbatkin, E., and Kilbride, T. (2021). To What Extent Does In-Person Schooling Contribute to the Spread of COVID-19? Evidence from Michigan and Washington. Working Paper 28455, National Bureau of Economic Research.

Goolsbee, A. and Syverson, C. (2021). Fear, lockdown, and diversion: Comparing drivers of pandemic economic decline 2020. *Journal of Public Economics*, 193:104311.

Gupta, S., Nguyen, T. D., Rojas, F. L., Raman, S., Lee, B., Bento, A., Simon, K. I., and Wing, C. (2020). Tracking Public and Private Responses to the COVID-19 Epidemic: Evidence from State and Local Government Actions. Technical Report w27027, National Bureau of Economic Research.

Harris, J. E. (2020). Geospatial Analysis of the September 2020 Coronavirus Outbreak at the University of Wisconsin – Madison: Did a Cluster of Local Bars Play a Critical Role? Working Paper 28132, National Bureau of Economic Research.

Iyer, S., Karrer, B., Citron, D., Kooti, F., Maas, P., Wang, Z., Giraudy, E., Dow, P. A., and Pompe, A. (2020). Large-Scale Measurement of Aggregate Human Colocation Patterns for Epidemiological Modeling. Technical report. Type: article.

Lau, H., Khosrawipour, V., Kocbach, P., Mikolajczyk, A., Schubert, J., Bania, J., and Khosrawipour, T. (2020). The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. *Journal of Travel Medicine*, 27(taaa037).

Lyu, W. and Wehby, G. L. (2020). Shelter-In-Place Orders Reduced COVID-19 Mortality And Reduced The Rate Of Growth In Hospitalizations. *Health Affairs (Project Hope)*, 39(9):1615–1623.

Mangrum, D. and Niekamp, P. (2020). JUE Insight: College Student Travel Contributed to Local COVID-19 Spread. SSRN Scholarly Paper ID 3606811, Social Science Research Network, Rochester, NY.

Morley, C. P., Anderson, K. B., Shaw, J., Stewart, T., Thomas, S. J., and Wang, D. (2020). Social Distancing Metrics and Estimates of SARS-CoV-2 Transmission Rates: Associations Between Mobile Telephone Data Tracking and R. *Journal of Public Health Management and Practice*, 26(6):606–612.

Nguyen, T. D., Gupta, S., Andersen, M., Bento, A., Simon, K. I., and Wing, C. (2020). Impacts of State Reopening Policy on Human Mobility. Working Paper 27235, National Bureau of Economic Research.

Palguta, J., Levínský, R., and Škoda, S. (2021). Do Elections Accelerate the COVID-19 Pandemic? Evidence from a Natural Experiment. Working Paper 891, GLO Discussion Paper.

Whaley, C. M., Cantor, J., Pera, M., and Jena, A. B. (2021). Assessing the Association Between Social Gatherings and COVID-19 Risk Using Birthdays. *JAMA Internal Medicine*, 181(8):1090–1099.

Xu, J., Hussain, S., Lu, G., Zheng, K., Wei, S., Bao, W., and Zhang, L. (2020). Associations of Stay-at-Home Order and Face-Masking Recommendation with Trends in Daily New Cases and Deaths of Laboratory-Confirmed COVID-19 in the United States. *Exploratory Research and Hypothesis in Medicine*, pages 1–10.

7 Appendix

Impact of opening on colocation Suppose that the number of tourists who would come from a county s to a county r is proportional to the length of the trails in the county r (LP_r) and the number of people in the county s. Hence, we have $\alpha LP_r n_s$ tourists from s potentially coming to visit r (where α is a proportionality factor). Now, we want to know the number of additional meetings that will occur once the trails are open. In order to have a meeting, individuals need to be in the same space within a five-minute interval. For the moment, assume that every visitor from s to r stays in r for the same amount of time and that space aspect does not matter. That is, suppose that the probability that a tourist meets a local during their stay is δ. So a tourist meets on avarege δn_r locals during their stay. Consequently, the additional number of colocation events is $\alpha\delta LP_r n_s n_r$. Hence, the probability of colocation after the opening is:

$$Pr(Colocation_{rs}|After) = \frac{m_{rs0} + \alpha\delta LP_r n_s n_r}{2016 n_r n_s} = \frac{m_{rs0}}{2016 n_r n_s} + \alpha\delta LP_r$$

where m_{rs0} is the default number of meetings before the opening captured by the fixed effects. Taking logs we have that

$$log(Pr(Colocation_{rs}|After)) = log(m_{rs0} + \alpha\delta LP_r n_s n_s) - log(2016 n_r n_s)$$

Taking the difference between after and before the policy implementation we obtain:

$$\begin{aligned}
&log(Pr(Colocation_{rs}|After)) - log(Pr(Colocation_{rs}|Before)) = \\
&log(m_{rs0} + \alpha\delta LP_r n_s n_r) - log(2016 n_r n_s) - \\
&(log(m_{rs0}) - log(2016 n_r n_s)) = \\
&log(\frac{m_{rs0} + \alpha\delta LP_r n_s n_r}{m_{rs0}}) \approx \frac{\alpha\delta n_s n_r}{m_{rs}}
\end{aligned} \tag{7}$$

Now let us add the hotel beds to the analysis. Assume that the hotel beds attract some additional tourists from county s and that tourists stay longer in ski resorts. In particular, suppose that the number of new visitors is proportional to the number of beds available for them. Hence we have $\tau H_r n_s$ new visitors from s to r (in addition to those who would come just for open trails) where H_r is the number of hotel beds in r and τ is a proportionality constant. Additionally, visitors coming for skiing can now stay longer. Assume again that a share of them proportional to the number of beds stay longer. Hence more meetings can take place. Suppose that the share $_r$ of tourists who stay longer produce κ more meetings than a tourist who does not stay in a hotel. Let us some up all the new terms. First, we have tourists who come skiing but don't stay for the night: $(1 - \zeta H_r)\alpha LP_r n_s$. Second, we have tourists who come because hotels opened: $\tau H_r n_s$. Third, we have tourists who come skiing and stay for the night: $(\zeta H_r)\alpha LP_r n_s$. In total, we obtain the following expression for the colocation probability after the opening of hotels and trails:

$$\begin{aligned}
Pr(Colocation_{rs}|After) &= \frac{m_{rs0} + (\delta\alpha LP_r n_s n_r)\zeta H_r + \delta\tau H_r n_s n_r + (\kappa\delta\alpha LP_r n_s n_r)(1 - \zeta H_r)}{2016 n_r n_s} \\
&= (\delta\alpha LP_r)\zeta H_r + \delta\tau H_r + (\kappa\delta\alpha LP_r)(1 - \zeta H_r)
\end{aligned} \tag{8}$$

Taking again the difference of logs before and after the policy we obtain:

$$\begin{aligned}
&log(Pr(Colocation_{rs}|After)) - log(Pr(Colocation_{rs}|Before)) \approx \\
&\frac{(\delta\alpha LP_r n_s n_r)\zeta H_r + \delta\tau H_r n_s n_r + (\kappa\delta\alpha LP_r n_s n_r)(1 - \zeta H_r)}{m_{rs}}
\end{aligned} \tag{9}$$

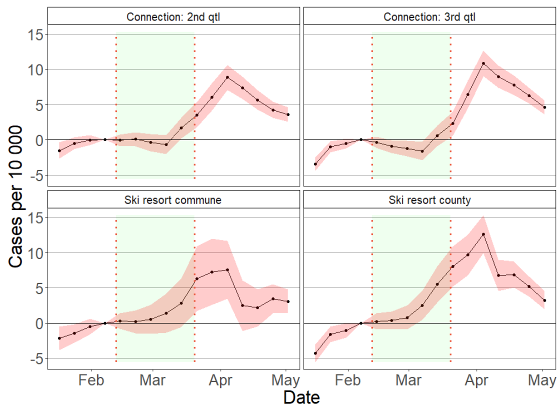
Event studies in the number of Covid-19 Cases Figures 11a and 11b show the coefficients from the health outcomes event study in the full and matched sample, respectively. We estimated the following regression to obtain the coefficients:

$$y_{kw} = \sum_{\substack{W \in \{\{01/17 : 01/31\}, \\ \{02/14 : 05/02\}\}}} Ski\_resort_k I(w = W)\beta^W + X_{kw}\delta + \zeta_k + \pi_w + \epsilon_{klw} \tag{10}$$

Where y_{kw} represents the number of cases per 10 000 in a commune k and a week ending at date w. The dummy $Ski\_resorts_k$ takes value 1 if the commune k contains a ski resort, and the indicator $I(w = W)$ is one if the week at hand is equal to W. The interaction between these two terms measures the differential trend in the cases per 10 000 in communes with versus. without ski resorts. X_{kw} contains controls for the share of fully vaccinated two weeks ago and the number of negative tests per 10 000. We allow for time π_w and commune ζ_k fixed effects, and we cluster the errors at the commune level. Figures 11a and 11b plot β_W coefficients from either estimation on the full sample or matched sample. The results are consistent with the unconfoundedness approach and suggest that the opening of hotels sped up the arrival of second-wave to communes with ski resorts.

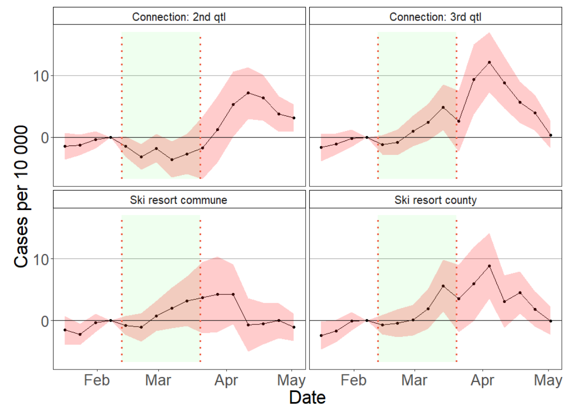
Figure 11: Event study: Covid-19 cases and hotels opening

(a) Covid cases: full sample



Note: The regression coefficients were calculated on the full sample. The date corresponds to the last day of the week. Source: Own elaboration based on the Ministry of Health Data

(b) Covid cases: matched sample



Note: The regression coefficients were calculated on the matched sample. Each treated unit was matched with one untreated units (first quantile of connection). Units were matched by the distance in the propensity scores computed on characteristics in the last period before the policy. The date corresponds to the last day of the week. Source: Own elaboration based on the Ministry of Health Data

Table 1: Summary statistics

(a) Population data

Number of unique tiles	34592
Number of tiles with any hotels	4203
Number of tiles of in proximity to ski resorts	3316
Minimum number of users on tile	10
Maximum number of users on tile	8693
Average number of users per tile 01:00-09:00	70
Average number of users per tile 09:00-17:00	81
Average number of users per tile 17:00-01:00	78

Note: Each observation counts the number of FB users on a tile in an 8-hour window. Data is omitted for privacy reasons if there are fewer than 10 users. If a user was on several tiles during the 8-hour period, they are assigned to the tile from which they were logging the most often (modal tile). The period covered is January 6th 2021-March 31st 2021 Source: Facebook Data for Good

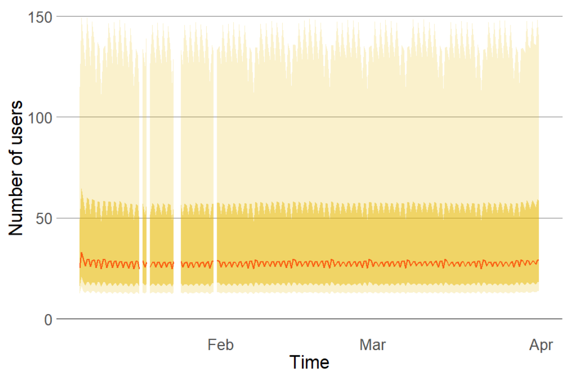
(b) Colocation data

Number of unique links	68820
Average number of users with a consistent home per county	768
Average colocation probability	$2.84 * 10^{-6}$
Minimum colocation probability	$9.89 * 10^{-11}$
Maximum colocation probability	0.00156

Note: Each observation corresponds to the probability that two randomly drawn users from two chosen counties meet in a randomly drawn 5-minute interval in a given week. A meeting is defined as being present on the same tile (0.6km × 0.6km) during a 5-minute interval. A user's home county is defined as one where she/he spent at least 6 nights in 10 days intervals around the date considered. User is discarded from computations if there is no consistent night location. Data is omitted for privacy reasons if there are fewer than 10 users. The period considered begins on the third week of January (with the last day 01-09-2021) and ends with the second week of April (with the last day 04-13-2021). Source: Facebook Data for Good

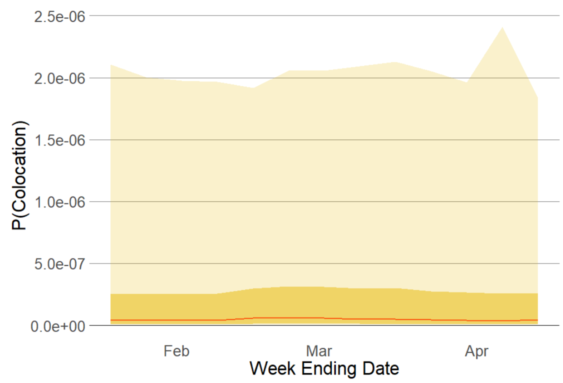
Figure 12: Data summary statistics

(a) Quantiles of FB tile population



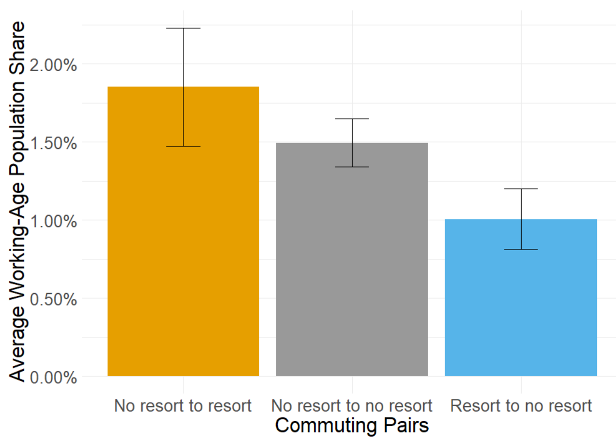
Note: The lighter shaded area corresponds to the 10^{th} and 90^{th} quantiles of the population on tiles in a given 8 hour period. The darker area corresponds to the 25^{th} and 75^{th} quantiles. The red line represents the median. White breaks show missing data. The X-axis shows time which comprises the date and the 8-hour window. Source: Own elaboration based on Facebook data

(b) Quantiles of colocation probabilities



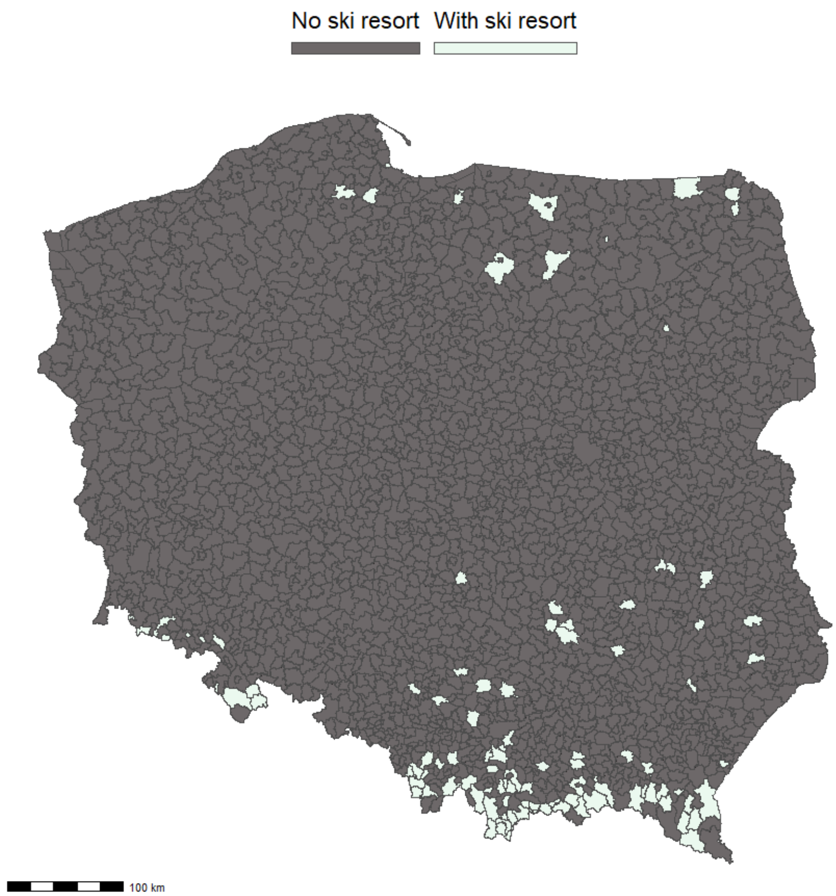
Note: The lighter shaded area corresponds to the 10^{th} and 90^{th} quantiles of colocation probability in a given week. The darker area corresponds to the 25^{th} and 75^{th} quantiles. The red line represents the median. The X-axis shows the last day of the week. Source: Own elaboration based on Facebook data

Figure 13: Commuting in counties with ski resorets



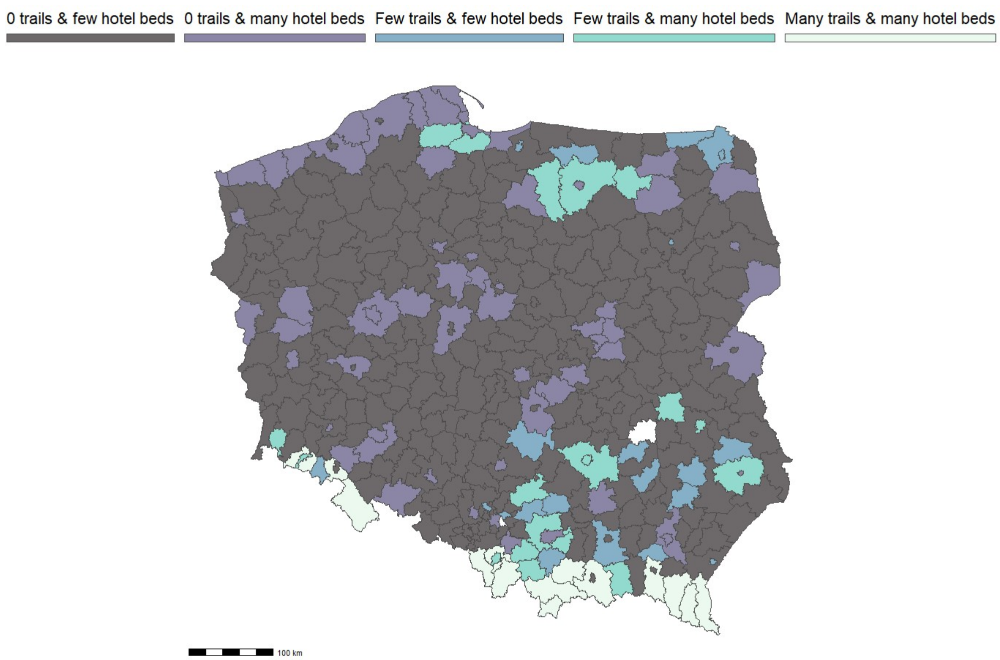
Note: The shares are calculated from data on commuting between communes in 2014 (the most recent available data). A share of working age commuters corresponds to the share of working age population in a commune j commuting to work in a commune i. The averages are taken across types of pairs. No resorts communes are communes without ski resorts and resort communes are communes with ski resorts. The sample is restricted to counties containing ski resorts. Source: Own elaboration based on data from Polish Statistical Office

Figure 14: Spatial distribution of communes with ski resorts



Note: The communes colored in white contain ski resorts. Source: Own elaboration based on data collected from internet

Figure 15: Spatial distribution of the touristic appeal



Note: Colors correspond to the touristic appeal. Source: Own elaboration based on data from Polish Statistical Office and own data