# IMMUNITY SCORE CALIBRATION BASED ON VACCINATION REBALANCING

**Ferdous Nasri**[1,*]
Ferdous.Nasri@hpi.de

**Simon Cyrani**[1,*], **Lukas Wenner**[1]
{firstname.lastname}@student.hpi.de

**Bernhard Y. Renard**[1]
Bernhard.Renard@hpi.de

[1]Hasso Plattner Institute, Digital Engineering Faculty, University of Potsdam, 14482 Potsdam, Germany
*These authors contributed equally.

## ABSTRACT

Having fine-grained vaccination data enables authorities to discover and target deficiencies in vaccination endeavors more easily and focused, leading in the best case to better containment of outbreaks. In many countries there are two factors that hinder more detailed breakdowns of vaccination data: Firstly, vaccinations are registered solely by place of vaccination, instead of the actual residency of the vaccinated person. Additionally, significant proportions of the population in rural areas travel to neighboring, more urban counties, to receive vaccinations. For related vaccination efforts for SARS-CoV-2, these factors resulted in very distorted vaccination numbers, with some counties having theoretical vaccination rates exceeding 100%. Furthermore, a lack of vaccination registry which records exact days and vaccine types for each person decreases retractability. It is well established, though, that for SARS-CoV-2 the effectiveness of vaccinations decreases over time. Therefore, exact calculations of actual effectiveness depending on the day and type of the last vaccination are not possible, either. Our work presents a two-part approach to tackling this problem and delivering more reliable county-wide data, using Germany as an example. In the first step, we reverse the aforementioned effect of vaccination tourism using a flow-based linear programming model. In the second step, we approximate a vaccination registry per county using daily vaccination data of each county's health authority. Combining recent insights in the effectiveness of SARS-CoV-2 vaccines regarding contagiousness over time, we calculate lower and upper bounds of the current effectiveness of vaccinations per county via linear programming and show that these are in fact close to each other.

## 1 INTRODUCTION

Vaccinations play a crucial role in decreasing global disease spread. This was highlighted with the recent SARS-CoV-2 pandemic, which caused speedy innovation of a variety of vaccines with different effectiveness. Many states and their corresponding health authorities thus established infrastructure to collect data on their vaccinated doses. Sufficiently tracking the number of shots administered has several benefits: First, this yields a basis to communicate its effectiveness to the public more easily. Second, this data enables decision-makers to discover and target deficits in vaccination endeavors more easily and focused, in the best case leading to better containment of outbreaks (Mathieu et al., 2021). However, different countries approached this task differently. In Germany, for instance, the data was collected based on where the vaccination dose was administered which may not necessarily be synonymous with the living location for that individual. This raised the problem that vaccination and therefore immunity rates could not be calculated from the raw vaccination numbers.

In this paper, we focused on reversing the effects of vaccination tourism and rebalancing the data to provide vaccination numbers based on residence. Furthermore, we calculate the immunity provided by the vaccinations, taking into account the vaccine effectiveness and deliver upper and lower immunity bounds for each area.

## 2 METHODS

### 2.1 COUNTY-WIDE VACCINATION REBALANCING WITH LINEAR OPTIMIZATION

In order to rebalance the county vaccination data to be based on residency instead of vaccination location, we use the official vaccination data released by the German public health institute, Robert Koch Institute (RKI) (Koch-Institut, 2022b). We focus on protection levels 1 to 3 here, which are defined as the first three vaccinations for most used vaccines except the Janssen vaccines which were counted as a protection level 2 after the first dose was administered.

Taking into account residency records, we limit the residency-based vaccination rate at 100% per protection level. Furthermore, we use the gravity law of human mobility (Zipf, 1946) and assume people do not travel far to get vaccinated. The problem motivates modelling with a graph, $G = (V, E)$, representing all counties as nodes and edges between close counties, defined as the shortest distance being less than 10 km between all county border points. For every county $i \in V$, let $v_i$ be its amount of vaccinations given by our data for a specific protection level, and let $p_i$ be its total population.

For every edge $(i, j) \in E$ from county $i$ to $j$, we find the number of people $x_{(i,j)}$ that live in $j$ but got vaccinated in $i$. Further assuming people only travel to close counties: We transfer for every edge $(i, j) \in E$ all $x_{(i,j)}$ of shots from $i$ to $j$. Formally, this yields for every county $i \in V$ new vaccination amounts, $\tilde{v}_i(x) := v_i + \sum_{j:(j,i)\in E} x_{(j,i)} - \sum_{j:(i,j)\in E} x_{(i,j)}$. Assuming a smooth distribution between counties, we can make the sum of movements $M$ as small as possible, $M(x) := \sum_{e\in E} x_e$.

We also assume there is an inherent truth in our data, i.e., a county with a much higher vaccination rate compared to another county before rebalancing would unlikely have a reverse relation after, modeled as:

$$S(x) := \sum_{e\in E} \left| \frac{\tilde{v}_i(x)p_j - \tilde{v}_j(x)p_i}{p_i + p_j} \right|, \quad \forall e \in E, \tag{1}$$

$S$ and $M$ evenly act against each other (given the property of equation 1, explained in appendix A.1.1). Note that the absolute terms in equation 1 are only piece-wise linear convex in $x$. In order for our model to be solvable utilizing linear optimization, we substitute every summand in $S$, $S_e$, by an additional variable $z_e$. Finally, we want to ensure that any county's total flow is at most their original vaccination amount. This restricts a possible "pass-through" effect of vaccinations across several counties.

Combining all the above-mentioned clauses, we have the following linear optimization model:

$$\begin{aligned} \min_{x,z} \quad & \sum_{e\in E} z_e + \sum_{e\in E} x_e \\ \text{s.t.} \quad & \sum_{j:(i,j)\in E} x_{(i,j)} \le v_i, \quad \forall i \in V, \\ & \tilde{v}_i(x) \le p_i, \quad \forall i \in V, \\ & z_e \ge +S_e(x), \quad \forall e \in E, \\ & z_e \ge -S_e(x), \quad \forall e \in E, \\ & x \ge 0 \end{aligned} \tag{2}$$

In order to keep the dates of the vaccination, we optimize the models for every day independently, using only the new vaccinations for this day. Additionally, we round the numbers to integers for each day.

### 2.2 INTRA-COUNTY IMMUNITY CALCULATION

It is known that the immunity gained by different SARS-CoV-2 vaccines decays over time (Nordström et al., 2022; Pormohammad et al., 2021; Cohn et al., 2021). Hence, only knowing the vaccination rate is insufficient to properly represent the vaccination landscape. Therefore, we gather vaccine effectiveness information from long-term studies and apply them to our rebalanced data , explained in more detail in the appendix A.1.2. By assuming that most people get all their vaccinations in the same county, we continue with the results of the rebalanced vaccination counts combined with the vaccine effectiveness to calculate immunity achieved by vaccines per county.

Assume $M$ is the maximum number of possible vaccinations and $D$ is the number of registered days in the vaccination data. For each $s \in \{1, 2, \ldots, M\}$ we find the number of people $x_{i,j}^{s-1,s}$ that got their $s-1$-th shot on day $i$ and their $s$-th shot on day $j$. Additionally, for $s \in \{0, 1, \ldots, M\}$, we encode with $x_i^{s,\mathsf{F}}$ the number of people whose $s$-th shot, received on day $i$, was their final shot until day $D$. Notice in Germany the special role of Janssen vaccines which are considered second shots, skipping the first shot. Analogously we introduce $x_{i,j}^{0,\mathsf{J}}$, $x_{i,j}^{\mathsf{J},3}$, and $x_{i,j}^{\mathsf{J},\mathsf{F}}$. For further simplicity we define $\mathsf{J}+1 = 3$ and $\mathsf{J}-1 = 0$.

Given for each $s \in \{1, 2, \ldots, M, \mathsf{J}\}$ functions $f_s : \{1, \ldots, D\} \to [0, 1]$ that map $d$ to the effectiveness after $d$ days of receiving shot $s$ as the most recent shot, we can calculate the county-wide effectiveness as

$$\sum_{d=1}^{D} \sum_{s \in \{1,\ldots,M,\mathsf{J}\}} f_s(D-d) x_d^{s,\mathsf{F}}. \tag{3}$$

Furthermore, we introduce for $s \in \{0, \ldots, M-1, \mathsf{J}\}$ minimal number of days $\Delta_s$ that must have passed between shots $s$ and $s+1$ for each person. That is, we enforce

$$x_{i,j}^{s} = 0 \quad \text{for } i + \Delta_s > j. \tag{4}$$

Assume $v_i^s$ is the number of vaccinations of type $s$ on day $i$ according to the input data. We add for each day $d$ and vaccination type $s$ two kinds of constraints to assure consistency with our input data and between consecutive state transitions:

$$\sum_{i=1}^{D} x_{i,d}^{s-1,s} = v_d^s, \quad \text{and} \tag{5}$$

$$x_d^{s,F} + \sum_{i=1}^{D} x_{d,i}^{s,s+1} = \sum_{i=1}^{D} x_{i,d}^{s-1,s} \tag{6}$$

However, it might be impossible to satisfy all constraints perfectly because of errors already existing in the original data source or introduced by our rebalancing step. To ensure feasibility, we introduce a slack variable $\delta$ that broadens the range of acceptable sums. Note that we do not need the slack for the highest vaccination state $M$ as there are no further dependencies. Adapting equation 5 we get:

$$\left.\begin{aligned} \sum_{i=1}^{D} x_{i,d}^{s-1,s} &\geq v_d^s - \delta \\ \sum_{i=1}^{D} x_{i,d}^{s-1,s} &\leq v_d^s + \delta \end{aligned}\right\} \quad \forall 1 \leq d \leq D, \forall s \in \{1, \ldots, M-1\} \tag{7}$$

Notwithstanding this definition, for $s = 2$ we consider $x^{1,2} + x^{0,\mathsf{J}}$ for the left-hand side summands, and for $s = 3$ we consider $x^{2,3} + x^{\mathsf{J},3}$ instead. In order to contain the maximum possible error, we also introduce a constraint for the sum of vaccinations over all days and a corresponding total slack $\delta_T$:

$$\left.\begin{aligned} \sum_{d=1}^{D} \sum_{i=1}^{D} x_{i,d}^{s-1,s} &\geq \sum_{d=1}^{D} v_d^s - \delta_T \\ \sum_{d=1}^{D} \sum_{i=1}^{D} x_{i,d}^{s-1,s} &\leq \sum_{d=1}^{D} v_d^s + \delta_T \end{aligned}\right\} \quad \forall s \in \{1, \ldots, M-1\} \tag{8}$$

We choose the slack dynamically by starting with small values and increasing them until our system is feasible to ensure the smallest possible deviance from our data.

Lastly, for all days $i$ we use the corresponding state-level ratios $p_i \in [0, 1]$ of used Janssen vaccines with regard to each county's amount of second shots and add a respective constraint:

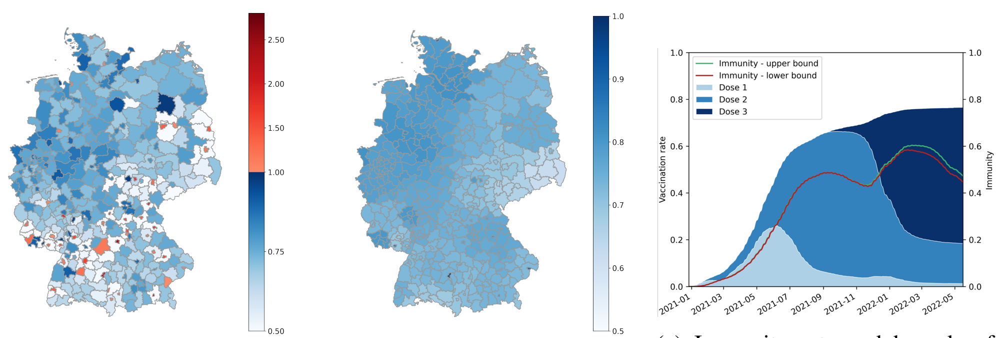
$$\sum_{d=1}^{D} x_d^{\mathsf{J},F} = \sum_{d=1}^{D} p_d \cdot v_d^2 \tag{9}$$

Casting everything into a linear optimization model, we attain the lower and upper bounds (formally described in the appendix equation 10 and 11.

## 3 Results

### 3.1 County-wise rebalancing

The lower and upper bounds for the ratio of people in a county with at least a certain protection level after applying our linear optimization rebalancing method were $[61.4\%, 95.4\%]$ for the first protection level, $[61.7\%, 97.7\%]$ for the second, and $[43.5\%, 78.4\%]$ for the third. Furthermore, figure 1 visualizes the rebalancing results. The first protection level maps are very similar and the third protection level is shown in the appendix figure 2.




(a) Raw vaccination rates (b) Rates after rebalancing (c) Immunity rate and bounds of Berlin.

Figure 1: Subplots 1a and 1b show vaccination rates per county on maps of Germany before and after rebalancing the data, with at least a protection level of 2 on 2022-05-16. Using our rebalanced data, on subplot 1c, we showcase the bounds of immunity of Berlin as an example. Here, immunity represents the total estimated protection against infections.

### 3.2 Intra-county immunity calculations

We assume minimal time distances between two shots and the calculated vaccination effectiveness listed in tables in the appendix B.2. In figure 3a (appendix), we visualize our calculated bounds for each county and illustrate the distance between the lower and upper bound. The difference between the bounds is most often around three percentage points. In figure 1c, as an example for Berlin, we can see in the background the rate of persons exclusively belonging to a protection level and in the foreground the immunities of our rebalancing method. It is visible that when vaccination numbers stagnate between September 2021 and December 2021, immunity falls, as expected, and when many people receive vaccination upgrades in January 2022, immunity increases accordingly.

## 4 Conclusion

We presented a rebalancing method for regional vaccination data. We showed how we used vaccination effectiveness data from table 1 to obtain an estimated regional vaccination coverage. Additionally, when determining vaccination coverage, we observed that the upper and lower bounds were not far apart, reducing the necessity of a vaccination registry for this use case. Overall, we have created methods for Germany and other countries without vaccination registries to generate more realistic vaccination data and determine an approximation of vaccination protection per region. Our approach is easily updatable and adaptable to other countries. Regional immunity scores provide valuable information to further research and disease modeling as well as aiding officials in making decisions regarding vaccination efforts, among many other use cases. We believe our approach is a solution to a crucial need for better spatially resoluted calibration of immunity which in turn can have a substantial impact on global health.

## REFERENCES


Barbara A. Cohn, Piera M. Cirillo, Caitlyn C. Murphy, Nickilou Y. Krigbaum, and Arthur W. Wallace. Sars-cov-2 vaccine protection and deaths among us veterans during 2021. *Science — AAAS*, November 2021. URL `https://doi.org/10.1126/science.abm0620`.

Robert Koch-Institut. Beschluss der stiko zur 20. aktualisierung der covid-19-impfempfehlung. *Epidemiologisches Bulletin*, (21), 2022a. URL `https://www.rki.de/DE/Content/Infekt/EpidBull/Archiv/2022/Ausgaben/21_22.pdf?__blob=publicationFile`.

Robert Koch-Institut. Covid-19-impfungen in deutschland, 2022b. URL `https://github.com/robert-koch-institut/COVID-19-Impfungen_in_Deutschland`.

Edouard Mathieu, Hannah Ritchie, Esteban Ortiz-Ospina, et al. A global database of covid-19 vaccinations. *Nature Human Behaviour*, 5(7):947—-953, 2021. URL `https://doi.org/10.1038/s41562-021-01122-8`.

Peter Nordström, Marcel Ballin, and Anna Nordström. Risk of infection, hospitalisation, and death up to 9 months after a second dose of covid-19 vaccine: A retrospective, total population cohort study in sweden. *The Lancet*, February 2022. URL `https://doi.org/10.1016/S0140-6736(22)00089-7`.

Ali Pormohammad, Mohammad Zarei, Saied Ghorbani, Mehdi Mohammadi, Saeideh Aghayari Sheikh Neshin, Alireza Khatami, Diana L. Turner, Shirin Djalalinia, Seied Asadollah Mousavi, Heydar Ali Mardani-Fard, et al. Effectiveness of covid-19 vaccines against delta (b.1.617.2) variant: A systematic review and meta-analysis of clinical studies. *MDPI*, December 2021. URL `https://doi.org/10.3390/vaccines10010023`.

George Kingsley Zipf. The $\frac{P1P2}{D}$ hypothesis: On the intercity movement of persons. *American Sociological Review*, 11(6):677–686, 1946. ISSN 00031224. URL `http://www.jstor.org/stable/2087063`.


# A APPENDIX

## A.1 METHODS

### A.1.1 COUNTY-WISE REBALANCING

Term equation 1 has following property, which gives evidence that $S$ and $M$ will evenly act against each other: Suppose $i, j \in V$ are counties such that the currently estimated residency-based vaccination rate is higher in $i$ than in $j$ that is

$$\frac{\tilde{v}_j(x)}{p_j} < \frac{\tilde{v}_i(x)}{p_i}.$$

Then,

$$\frac{d}{dx_{(i,j)}} M(x) = 1 = -\frac{d}{dx_{(i,j)}} S_{(i,j)}(x) = -\frac{d}{dx_{(j,i)}} S_{(j,i)}(x).$$

In particular, increasing $x_{(j,i)}$ by $\delta > 0$ decreases both $S_{(i,j)}(x)$ and $S_{(j,i)}(x)$ by $\delta$ each, as long as equation A.1.1 still holds for the new $\tilde{v}_i(x), \tilde{v}_j(x)$.

### A.1.2 VACCINE EFFECTIVENESS

Combining a few studies to gather this information, given the lack of sole resource, we primarily use the Nordström et al. (2022) for Comirnaty, Moderna, and AstraZeneca, supplemented with Cohn et al. (2021) for Janssen effectiveness. For first-dose vaccinations, we use Pormohammad et al. (2021) as the baseline effectiveness for 14 days after vaccination. Because there are no long-term studies for first-dose vaccinations, we assume that effectiveness changes as in Nordström et al. (2022) after second-dose vaccinations. For third vaccinations, due to lack of data, we assume that they provide the same effectiveness as second vaccinations, i.e., vaccine protection is renewed during the third vaccination.

For all following methods, we use an additional dataset for daily used Janssen vaccines at the state level to differentiate them from other vaccines Koch-Institut (2022b). Because of insufficient data, we do not distinguish the other vaccines. Instead, we take the average of nationwide used shots as an approximation Koch-Institut (2022b). Therefore, we rely on the following assumptions: The ratio of daily Janssen shots at the state level is roughly equal to each county's ratio of Janssen shots. Furthermore, the nationwide ratio of each of the non-Janssen vaccines is also constant for every county.

Using the nationwide vaccine distribution and vaccination effectiveness for given time periods, we can estimate how effective an average vaccination is for each period. We will use this effectiveness in combination with the artificial vaccination registers from the following two sections to determine how high the immunity from vaccination is in a county.

### A.1.3 BOUND CALCULATION OF VACCINE EFFECTIVENESS

Formal description of the linear optimization model for calculating the lower bound of effectiveness:

$$\begin{aligned} \min_{x} \quad & equation\ 3 \\ \text{s.t.} \quad & equation\ 6, equation\ 7, equation\ 8, equation\ 9, equation\ 4 \\ & x \geq 0, \end{aligned} \tag{10}$$

and for the upper bound:

$$\begin{aligned} \max_{x} \quad & equation\ 3 \\ \text{s.t.} \quad & equation\ 6, equation\ 7, equation\ 8, equation\ 9, equation\ 4 \\ & x \geq 0. \end{aligned} \tag{11}$$

## B RESULTS

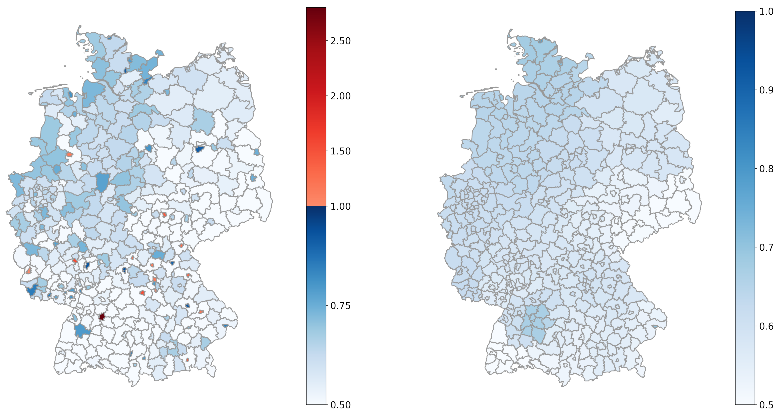
### B.1 COUNTY-WISE REBALANCING

### B.2 VACCINE EFFECTIVENESS

This section presents the result of the collected data from section A.1.2. Table 1 lists how effective a specific vaccine is within a certain number of days for one to three doses. The sources Nordström et al. (2022); Cohn et al. (2021); Pormohammad et al. (2021) reported that the vaccine effectiveness wanes over time, and we can observe this behavior in the effectiveness of each vaccine. In contrast to the other vaccines, this decline is more significant for the Comirnaty and Janssen vaccines.

In table 3, we can see the vaccination distribution for the different vaccines in Germany. Especially Comirnaty shows high shares and is, therefore, the most common vaccine in Germany. At the same time, people in Germany receive Moderna more often as the third dose in relation to the first and second doses.

### B.3 BOUND CALCULATION




(a) Protection level 3 before any rebalancing

(b) Protection level 3 after LP-based rebalancing

Figure 2: Maps of Germany showing vaccination rates per county of people with at least a certain protection level on 2022-05-16. The first row of maps shows vaccination rates before any rebalancing, while the second row shows rates after applying our rebalancing method.

| Protection | Vaccine | 0 days | $\leq$ 14 days | $\leq$ 31 days | $\leq$ 61 days | $\leq$ 121 days | $\leq$ 181 days |
|---|---|---|---|---|---|---|---|
| 1 | AstraZeneca | 0 | 0.44 | 0.48 | 0.39 | 0.17 | 0.17 |
| | Comirnaty | 0 | 0.57 | 0.54 | 0.5 | 0.27 | 0.14 |
| | Moderna | 0 | 0.72 | 0.69 | 0.62 | 0.48 | 0.45 |
| | Janssen | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | AstraZeneca | 0.44 | 0.65 | 0.69 | 0.58 | 0.36 | 0.36 |
| | Comirnaty | 0.57 | 0.92 | 0.89 | 0.85 | 0.52 | 0.39 |
| | Moderna | 0.72 | 0.96 | 0.93 | 0.86 | 0.72 | 0.69 |
| | Janssen | 0 | 0.84 | 0.81 | 0.75 | 0.53 | 0.13 |
| 3 | AstraZeneca | 0.65 | 0.65 | 0.69 | 0.58 | 0.36 | 0.36 |
| | Comirnaty | 0.92 | 0.92 | 0.89 | 0.85 | 0.52 | 0.39 |
| | Moderna | 0.96 | 0.96 | 0.93 | 0.86 | 0.72 | 0.69 |
| | Janssen | 0.84 | 0.84 | 0.81 | 0.75 | 0.53 | 0.13 |

Table 1: Vaccine effectiveness against SARS-CoV-2 infection up to 181 days after the last vaccine dose grouped by protection level and vaccine.

| Vaccine | Protection 1 share | Protection 2 share | Protection 3 share |
|---|---|---|---|
| AstraZeneca | 15.36% | 5.48% | 0.01% |
| Comirnaty | 76.23% | 78.93% | 63.31% |
| Moderna | 8.41% | 9.85% | 36.64% |
| Janssen | 0% | 5.74% | 0.04% |

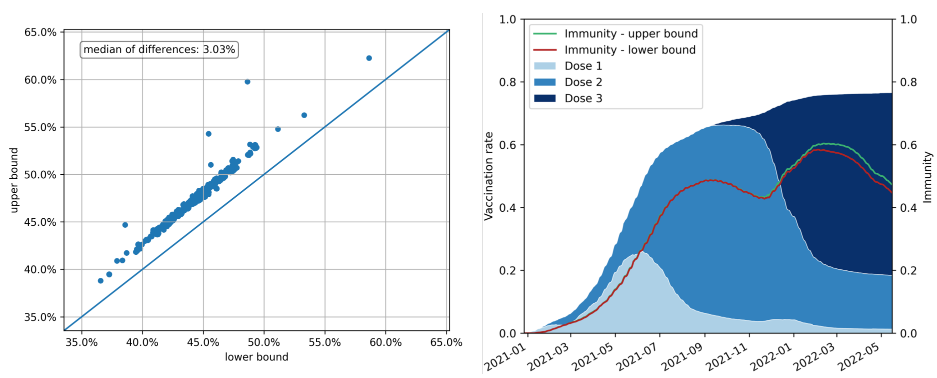
Table 2: Vaccination distribution in Germany by vaccine and protection level until 2022-05-16. For the distribution, we consider only the listed vaccines from Table 1.

| Vaccine | Protection 1 share | Protection 2 share | Protection 3 share |
|---|---|---|---|
| AstraZeneca | 15.36% | 5.48% | 0.01% |
| Comirnaty | 76.23% | 78.93% | 63.31% |
| Moderna | 8.41% | 9.85% | 36.64% |
| Janssen | 0% | 5.74% | 0.04% |

Table 3: Vaccination distribution in Germany by vaccine and protection level until 2022-05-16. For the distribution, we consider only the listed vaccines from Table 1.

| type | min. distance |
|---|---|
| first to second shot | 20 days |
| second to booster shot | 90 days |
| Janssen to booster shot | 90 days |

Table 4: The minimal time that needs to have passed for a person to receive their next vaccination. The numbers used are based on Koch-Institut (2022a).




(a) Each county's immunity bounds in Germany.

(b) Immunity rate and bounds of Berlin.

Figure 3: Using our rebalanced data, the scatter plot of each county's immunity in Germany is shown, while on subplot 3b, we showcase the results of Berlin as an example. Here, immunity represents the total estimated protection against infections.