Methodology of the
Vaccine Allocation Planner for COVID-19

Ariadne Labs and Surgo Foundation

October 26, 2020, v1

Contents

A. Context

Across the United States, people are eagerly awaiting the arrival of safe, effective vaccines against the coronavirus that causes COVID-19. Vaccines will be the surest sustainable way of protecting health, saving lives, and getting the country beyond the pandemic. Health officials and government decision makers at state and local levels must plan carefully to ensure that vaccines are distributed as quickly as possible once stocks are made available.

But there will not be enough stocks to vaccinate everyone immediately. Officials will have to prioritize and allocate them to the groups most in need. The Vaccine Allocation Planner for COVID-19 (VAPC) provides state and county decision makers with the localized data they need to plan vaccine distribution, based on available vaccine doses, priority populations, and vulnerable communities in each state.

Vaccination against a transmittable disease such as COVID-19 is an individual, community, and governmental responsibility that transcends borders. Equitable access to immunization is a core component of the right to health. Strong vaccination allocation systems during extreme resource scarcity, such as the situation we will soon face, are essential to combatting the virus causing the current pandemic. Informed decisions and implementation strategies are critical to ensuring the sustainability of vaccination programs. The full potential of vaccinations can only be realized through learning, continuous improvement and innovation in research and development, as well as quality improvement across all aspects of vaccination. Through the prioritization of vaccination schemes to our frontline workers and the most vulnerable in our population to COVID-19, equitable allocation will have precipitous effects on the remainder of the general public.

B. This document

This document is arranged according to the three main functions in the website:

1. Select groups to vaccinate
2. Count available doses
3. Allocate doses to counties

The reader may want to open the VAPC site to follow along with each section.

Our goal is to provide enough information behind our statistical methods for analysts to understand and potentially recreate each step. If you would like more detail or have other questions please email contact@vaccineallocationplanner.org.

This document does not describe the design or build of the VAPC website itself.

C. General principles about methods

- We strive for transparency at every step.
- We will be updating the VAPC continuously. Results may change as we refine our methods, as recommendations change from various official bodies, and as the qualities of the available vaccines become clear.
- All of the estimates in VAPC reflect our best efforts. We selected the most reliable data sources available, but we did need to make assumptions and imputations at several points, as described in this document. Further, we plan to refine some of our methods going forward, such as providing ranges rather than point estimates. As the statistician George Box wrote, "All models are wrong but some are useful." Recognizing that the VAPC will be inaccurate at times, we strive for it to be useful.

- The VAPC is centered on the county level because this is the smallest geographic unit with reasonably reliable data for all the priority populations. If data becomes available at smaller units, such as the municipality or census tract, we will consider switching.
- We use only publicly available data, and plan to do so moving forward. We will resist using proprietary or commercial data unless the gains to accuracy outweigh the goal of transparency.
- The data science teams used a mix of R, Python, and SAS to implement these methods.

D. The NASEM report

The VAPC closely reflects the [NASEM guidance](#)[1], relying on their careful ethical deliberations regarding vaccine prioritization. We recommend reading the full report, which describes the ethics and rationale behind the settings reflected in the VAPC.

In particular the VAPC centers on the 13 populations arranged in prioritized phases, as presented in the NASEM report's "Table 3-2, Applying the Allocation Criteria to Specific Population Groups." These 13 populations are presented by phase in Table 1 below.

The default values in the VAPC reflect the NASEM recommendations, such as the pre-selection of both populations in phase 1a (high risk health care workers and first responders) and the pre-selected option to take a 10% holdout.

The NASEM report also recommends that "Programs should do everything possible to reach all individuals in one priority group before proceeding to the next one." (page 4-4) At the moment, the VAPC does not reflect this recommendation, but distributes vaccines among all the populations selected by the user, regardless of phase (more information in the section on VAPC function 3, below.)

E. VAPC function 1: Select groups to vaccinate

The first function of the VAPC is the most complex to calculate, requiring estimates of the size of the 13 priority populations and their overlaps in every county.

1. Estimating county populations by group

We estimated population sizes in all US counties for the 13 priority groups. The NASEM report estimates the national total for each group, which we took as a rough benchmark to match with the sum of our county-level estimates. The NASEM report does give sources for its totals, and recognizes that precise estimates are difficult to come by. We followed NASEM's lead in sourcing data, and strove to generally match the NASEM national numbers for each group, unless we had a direct reason for a variance, as described below.

The output of this step is a data frame with one row for each county (n=3,142), one column with the county FIPS code (a standard identifier), one column with the total population of the county from Census estimates, and one column each for the 13 groups, with the number of people (integers) in each group in that county. This section describes how we estimated the 13 groups, and Table 1 summarizes the definitions and data sources for each.

---

[1] National Academies of Sciences, Engineering, and Medicine 2020. Framework for Equitable Allocation of COVID-19 Vaccine. Washington, DC: The National Academies Press. https://doi.org/10.17226/25917.

Table 1: Population group definitions and data sources

Group	Subgroup(s)	VAPC data source
PHASE 1A		
1 High risk workers in health care facilities	Hospitals, physician and other health practitioner offices, outpatient care centers, home healthcare services, pharmacies and drug stores, and nursing and residential care facilities and homes (skilled nursing, mental health, developmental disability, mental and substance abuse, assisted living, retirement communities, other residential care)	Bureau of Labor Statistics 2020 Quarterly Census of Employment and Wages Note: Raw data from BLS QCEW at the county level is highly suppressed (see main text on the imputation method used)
2 First responders	Police	ArcGIS, CA Governor's Office of Emergency Services
	Fire protection services	Bureau of Labor Statistics 2020 Quarterly Census of Employment and Wages
	Other ambulatory health care services	Bureau of Labor Statistics 2020 Quarterly Census of Employment and Wages
PHASE 1B		
3 People with 2+ significant comorbid conditions	Obesity (BMI ≥ 30 kg/m2), diabetes mellitus, COPD, heart disease, chronic kidney disease, and any (1+) condition	Direct estimates of comorbidity rates by county from the CDC (Razzaghi et al. 2020) are adjusted for multimorbidity using Clark et al. 2020 estimates for 1 and 2+ comorbidity populations
4 Older adults in congregate settings	Nursing residents	Centers for Medicare & Medicaid Services - Division of Nursing Homes/Quality, Safety, and Oversight Group/Center for Clinical Standards and Quality
	Residential care residents	Department of Homeland Security - Homeland Infrastructure Foundation-Level Data Note: Includes residents of assisted living facilities for the elderly and continuing care retirement communities
	Crowded households with adults over 65	CDC Social Vulnerability Index - American Community Survey 2014-2018 5-year Estimates Note: Calculated as as a product of crowding (more people than rooms) and persons over 65

Table 1: Population group definitions and data sources (con't)

Group	Subgroup(s)	VAPC data source
PHASE 2		
5 Critical risk workers (part 1)	Workers in dentist offices, medical and diagnostic laboratories, food and beverage manufacturing facilities and stores, gas stations, cosmetic and beauty supply stores, optical goods stores, other health and personal care stores, transportation industries (air, rail, water, truck, public transit and ground passenger, pipeline, support activities), postal service and other couriers and messengers, general warehousing and storage establishments, and pharmaceutical and medicine manufacturing facilities	Bureau of Labor Statistics 2020 Quarterly Census of Employment and Wages
6 Teachers and school staff	Elementary and secondary school teachers	Bureau of Labor Statistics 2020 Quarterly Census of Employment and Wages
	Child day care service staff	Bureau of Labor Statistics 2020 Quarterly Census of Employment and Wages
7 People with 1 significant comorbid condition	(see above)	Direct estimates of comorbidity rates by county from the CDC (Razzaghi et al. 2020) are adjusted for multimorbidity using Clark et al. 2020 estimates for 1 and 2+ comorbidity populations
8 All older adults	Persons over 65	CDC Social Vulnerability Index - American Community Survey 2014-2018 5-year Estimates
9 People and staff in homeless shelters or group homes	People living in non-institutional group quarters (homeless shelters, group homes for adults, residential rehab treatment centers for adults)	Census Bureau 2010 Decennial Census Note: Will be updated with 2020 census data once available.
	Staff in community food and housing, and emergency and other relief services, and vocational rehabilitation services	Bureau of Labor Statistics 2020 Quarterly Census of Employment and Wages
10 Incarcerated / detained people and staff	Staff in correctional institution establishments	Bureau of Labor Statistics 2020 Quarterly Census of Employment and Wages
	Incarcerated population	Vera Institute of Justice - Incarceration Trends Dataset
PHASE 3		
11 Young adults (18-30)	Persons age 18-30	Census Bureau 2019 American Community Survey
12 Children (3-18)	Persons age 3-18	Census Bureau 2019 American Community Survey
13 Critical risk workers (part 2)	Workers in the following services, establishments, and stores: waste management and remediation, transportation equipment manufacturing, utilities, crop production, specialty trade contractors, oil and gas extraction, animal production and aquaculture, mining (coal, metal ore, nonmetallic mineral), construction of buildings, hardware, clothing and clothing accessories, food services	Bureau of Labor Statistics 2020 Quarterly Census of Employment and Wages

and drinking places, and credit intermediation and
related activities

a. Employment data

For all professions, unless noted otherwise, we relied on industry data from the Bureau of Labor Statistics (BLS) Quarterly Census of Employment and Wages (QCEW), and our numbers are based on employment conditions pre-pandemic (January-March 2020)[2]. The industries were located through their North American Industry Classification System (NAICS) codes. The BLS QCEW affords the finest geographic granularity of employment data available at the county level. While the data is by industry (e.g. education), and not by occupation (e.g. 2nd grade teacher), QCEW data includes all pertinent staff that work alongside these critical workers and would therefore need vaccination as well. Employers in the United States can fall into 4 major ownership types including private, federal, state, and local government. The employment numbers are dispersed among these ownership types and need to be summed to get the total number of employees per industry per county. The average number of employees per industry is then taken across the first 3 months of 2020.

b. Imputation of suppressed data

Privacy laws, based upon stipulations from a Federal Registry Notice[3], introduce suppression issues when accessing employment data from the BLS QCEW at the county level. Approximately 60 percent of the most detailed level data are suppressed for confidentiality reasons.These issues can arise when an industry has few employers within a respective county or when an industry is dominated by state and local government (e.g. education). There are various levels of suppression that can lead to large underestimations at the county level. These levels of suppression include *primary* (dubbed the 80/3 rule) and *secondary*. *Primary* suppression occurs in a county when either a single establishment employs over 80% of the employees or there are less than three establishments total. *Secondary* suppression occurs when the value of the primary suppressed data can be back-calculated with simple arithmetic from the data that is not suppressed in that county. Another undisclosed level of suppression ensures the integrity of the hidden data.

The regulations for suppression can differ depending on the ownership type. For example, federal data is always disclosable, but state governments have much stricter guidelines on what data can be made available. However, all levels of suppression could lead to large differences between the summation of employed individuals per industry from the county level data compared to the national estimates. For instance, the education industry has employment data in 99% of the 3,142 counties. However at least 1 ownership type (i.e. state, local, or private) has suppressed data in 89% of those counties. This amount of suppression leads to ~5.9 million employees being unaccounted for within the education industry.

The magnitude of the suppression in almost every industry made it impossible to simply ignore, and we sought a systematic approach to imputing the suppressed data. Since we know several rules are followed to establish which data gets undisclosed, a multiple imputation method is not warranted since the data would need to be randomly suppressed[4]. Also, imputing data based on

---

[2] Quarterly Census of Employment and Wages. Retrieved from:
https://www.bls.gov/cew/downloadable-data-files.htm

[3] Federal Register, 69, Department of Labor - Bureau of Labor Statistics 19452 (2004 April 13, 2004).
https://www.doleta.gov/grants/sga/04-103sga_correction.pdf

[4] Sterne, J., White, I. R., Carlin, J., Spratt, M., Royston, P., Kenward, M., Carpenter, J. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ, 338(b2393). doi:https://doi.org/10.1136/bmj.b2393

correlations with covariates to industry data (e.g. socioeconomic status, household income, etc.) would require a targeted approach to each individual industry, and we did not have the time in the project.

Instead, we used data from other sources for industries where employment information had been compiled and completely disclosed (e.g. law enforcement employees at the county level). For the majority of industries, other sources did not exist. Therefore, we implemented a simplistic approach given what we knew about the suppressed data. We distributed the difference between national and county level totals by ownership type to the counties with suppressed data weighted by population. We know exactly which counties have suppressed data per industry as well as the ownership type (i.e. state, local, private) in which the data was suppressed. QCEW also lists the national totals by ownership type per industry. Therefore, we can distribute the difference between the national and suppressed estimates among those counties that actually employ that industry ensuring that no employee is assigned to a county where that industry does not exist. Since we can ensure the validity of that industry existing in the county, it is a reasonable assumption to distribute based on population since more employees tend to work in more highly populated areas[5] on average.

c. Volunteer firefighters

We could not identify a county-level count of volunteer firefighters, which are more numerous than their paid counterparts (which are included in the QCEW data above). There are 1,823 counties (containing 81m people) that had no paid firefighters listed in the BLS data nor had their count suppressed by BLS. We distributed the 800,000 volunteer firefighters across these counties weighted by population, meaning 1% of the population of these counties was designated as being a volunteer firefighter. We considered an alternative approach of distributing the 800,000 across all non-urban counties (as most urban counties have paid firefighters[6]) however this would have left a substantial number of urban counties with no firefighters, whereas 2+% of the population in non-urban counties would be counted as a volunteer firefighter.

d. Critical risk workers

The critical risk workers were defined according to the national guidelines set out by the Department of Homeland Security[7]. The list of industries included under critical risk workers is extensive and includes: healthcare and public health, law enforcement and first responders, education, food and agriculture, energy, water and wastewater, transportation, public works, communications, critical manufacturing, hazardous materials, financial services, chemical, defense, real estate, and hygiene services. Some workers in these "essential" industries do not have a high risk of exposure to COVID-19. For instance, 37 percent of jobs in the US[8] can be

---

[5] BLS (2020). County Employment and Wages – First Quarter 2020 [Press release]. https://www.bls.gov/news.release/pdf/cewqtr.pdf

[6] Evarts, Ben; Stein, Gary P. (February 2020). "U.S. Fire Department Profile through 2020". National Fire Protection Association Fire Analysis and Research Division. Retrieved May 6, 2020.

[7]Krebs, C. C. (2020). Advisory memorandum on ensuring essential critical infrastructure workers ability to work during the COVID-19 response. Retrieved from Statistics, U.S. BLS (2020). https://www.cisa.gov/sites/default/files/publications/Version_4.0_CISA_Guidance_on_Essential_Critical_Infrastructure_Workers_FINAL%20AUG%2018v3.pdf

[8] Dingel, J., and B. Neiman. 2020. White Paper: How many jobs can be done at home? Chicago, IL: Becker Friedman Institute for Economics at the University of Chicago.

worked remotely and some of the employees are white collar and can avoid direct interactions with others. The NASEM guidelines acknowledge that there is no complete list of all workers who are critical. The critical worker designation involves both working in vital industries to the functioning of society and in occupations where they cannot avoid exposure risk.

The critical risk workers are separated into 2 groups within the NASEM guidelines: critical workers in high-risk settings (Phase 2) and workers at moderately higher risk of exposure (Phase 3). The high-risk setting designation (Phase 2) is defined according to occupations that have potential exposure both from colleagues and the public as well as the inability to social distance or wear protective equipment (PPE) (e.g. cashiers and food store workers, public transit, etc.). The moderately higher risk worker (Phase 3) are those occupations where the potential exposure comes from their colleagues or the use of PPE and social distancing measures are easily implemented (e.g. factory workers in production, bank tellers, agriculture, etc.).

e. Comorbidity estimates

NASEM guidelines split people with comorbidities into two groups. The severe comorbidities group (Phase 1b) includes people with two or more of the underlying conditions outlined by CDC[9] as leading to an increased risk of severe illness if infected with COVID-19. The high risk group (Phase 2) has exactly one of these conditions. Compiling subsets of a county that have 1 and 2 or more of these comorbidities quickly becomes inaccurate due to double counting that could lead to unrealistic numbers if the overlap between conditions is left unaddressed. For example, an obese person may also have type 2 diabetes mellitus and cancer. If only the total numbers for each underlying condition are counted, this would lead to a drastic overestimation.

Clark et al. (2020)[10] estimated the number of individuals at increased risk of severe disease (defined by the CDC, WHO, and UK public health agencies at the time of publication in June) by age (5-year age groups), sex, and country for 188 countries using prevalence data from the Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) and UN population estimates for 2020. They also analyzed data from Scottish and Chinese multimorbidity studies to calculate the multimorbidity fraction for each country.

However, this approach is biased toward individuals with higher socioeconomic status. It is known that income and education correlate with the presence of comorbidities. Lower socioeconomic populations suffer from greater rates of comorbidities. Therefore, projecting national estimates to the county level will overestimate the true population in wealthier counties, effectively creating an inequitable recommendation for vaccine allocation. We have developed a correction for these estimates based on county-level prevalence of chronic conditions susceptible to COVID-19 calculated by the CDC[11] from the Behavioral Risk Factor Surveillance System (BRFSS) state surveys.

---

[9] CDC (2020, October 16, 2020). People with Certain Medical Conditions. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html

[10] Clark, A., Jit, M., Warren-Gash, C., Guthrie, B., Wang, H. H. X., & Mercer, S. W. (2020). Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study. The Lancet Global Health, 8(8), E1003-E1017. doi:https://doi.org/10.1016/S2214-109X(20)30264-3

[11] Razzaghi H, Wang Y, Lu H, et al. Estimated County-Level Prevalence of Selected Underlying Medical Conditions Associated with Increased Risk for Severe COVID-19 Illness — United States, 2018. MMWR Morb Mortal Wkly Rep 2020;69:945–950. DOI: http://dx.doi.org/10.15585/mmwr.mm6929a1

The CDC used BRFSS data and small area estimation to generate estimates of county-level populations with any of five conditions (obesity (BMI ≥30), heart disease, diabetes mellitus, COPD, and CKD), taking overlap into account. Because these are direct measures of conditions, they already reflect any differences in morbidity by age structure, SES, or any other covariate. This is a strength. Another strength is that this measure reflects obesity, an important predictor for the US, whereas Clark does not. But these estimates pool people with exactly 1 condition and those with 2+ conditions. We need a way to divide them into two subgroups.

According to Clark, 94.3 m people in the US have 1+ comorbidities, comprising 64.7m (69%) with exactly 1 comorbidity, and 29.5m (31%) with 2+ comorbidities. We divided the BRFSS estimates into two groups based on these proportions. So if a county has 1,000 people with 1+ comorbidities according to BRFSS, we assigned 690 to the group with exactly 1 comorbidity and 310 to the 2+ comorbidities group. Applying this ratio across all counties is clearly a large assumption that may be inaccurate in some counties. We plan to refine this methodology in particular moving forward.

f. Older adults in congregate settings

Alongside residents in nursing homes and residential care facilities, NASEM includes all adults over the age of 65 that live below the poverty line in this category as a proxy for older adults living in overcrowded settings. Their contention is that older adults living in overcrowded settings may live in multigenerational households[12] that can typically be found often in lower-income communities.

We found a more direct method of identifying older adults living in overcrowded settings outside of community assisted living facilities such as nursing homes. The ACS collects data on people living in crowded settings, which is defined as more people to a household than there are rooms. This variable is available at the county level so we can get accurate variability. The variable does not disaggregate by age, so we multiplied the percentage of people living in crowded settings by the number of people over 65 to get this subset within the older adults in congregate settings group.

g. Limitations to marginal estimates

Whereas some groups are clearly defined in the NASEM recommendation, such as healthcare workers, teachers, or those with severe comorbidities, other groups are not as clearly defined, such as critical workers. Individual states may have their own definition of which workers are critical, which would change the allocation of doses.

Data on employment from the BLS is currently available for the first quarter of 2020 only (as of November 2020) and may not reflect the current levels given the massive unemployment rates spurred by the pandemic.

Older adults in congregate settings may be an underestimate since we are not accounting for multi-generational households where there are the same number of people as rooms or fewer.

---

[12] Miller, R. B., and C. A. Nebeker-Adams. 2017. Multigenerational households. In Encyclopedia of couple and family therapy, edited by J. Lebow, A. Chambers, and D. C. Breunlin. Cham, Switzerland: Springer International Publishing. Pp. 1–3.

2. Estimation of overlapping populations

After getting these marginal estimates of the 13 groups for each county, our next task was to estimate the amount of overlap between these groups. Understanding the degree of overlap is important because it has a large effect on the number of vaccine courses needed by a population. Many groups clearly overlap strongly, such as elderly in congregate settings and those with 2+ comorbidities, and other groups do not overlap at all, such as children and elderly. For this reason, we did not feel we could assume independence.

We took a Monte Carlo approach to estimate the overlap between all 13 groups. Given the expected correlations between groups and the marginal populations, we would randomly generate a simulated population of 1 million for each county, with 13 binary variables indicating group membership.

Our goal was to estimate the proportion of people in every possible combination of belonging to 13 groups. For each county, the desired dataset would have 13 columns of binary variables indicating membership in each group, and one column with the proportion of people in that county who were members of exactly those groups. This would imply 8,192 (2^{13}) rows for each of 3,142 counties, for a total dataset of 14 columns and about 26 million rows. Luckily, memberships in more than a few groups is rare, and the final dataset is much smaller and more manageable.

a. Conditional probabilities for each pair of groups

To generate the simulated population, we needed a correlation matrix for the 13 groups. Our first step was to consider the conditional probabilities of group membership. Within one county, there are 78 pairs of group memberships. Each pair of membership in group X and group Y relates to a standard 2x2 table of probabilities as shown in Table 2.

Table 2: Sample 2x2 table

		Member of group Y?		
		Yes	No	
Member of group X?	Yes	A	B	P(X)
	No	C	D	1 – P(X)
		P(Y)	1 – P(Y)	1

We have P(X) and P(Y) from the estimates of population size in each county. Our goal was to estimate cell A for each pair, which would be the amount of overlap between the two groups. Under independence, cell A would equal P(X)*P(Y), but we wanted to reflect more accurate conditional probabilities. In the cases with zero overlap, cell A is zero, as are the conditional probabilities. In cases with larger overlap, the conditional probabilities could range widely. Further, there are two "directions" in which to resolve cell A: the probability of being in group X given membership in group Y [P(X|Y)] and the probability of being in group Y given being in group X [P(X|Y)].

We assembled estimates of every P(X|Y) and P(Y|X) (see appendix A). We began by populating all the conditional probabilities that are logically zero:

- Children, young adults, and older adults are mutually exclusive groups (also affects older adults in congregate settings)

- Severe comorbidity (2+ conditions) and moderate comorbidity (1 condition only) are mutually exclusive (groups 3 and 7)

We then set several conditional probabilities to zero according to some simplifying assumptions:

- All occupations are mutually exclusive (occupations: 1,2,5,6,13)
- Homeless/incarcerated/older in congregate settings are mutually exclusive groups
- Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
- Critical risk workers of type 1 and 2 are mutually exclusive (groups 5 and 13)
- Children do not have jobs, and those with jobs are not children

There may be exceptions to these cases, but we assume they are negligible. We will investigate these assumptions further in future iterations.

Finally, we drew estimates of the remaining conditional probabilities from a variety of analyses using several data sources, including:

- US Census
- National Health Interview Survey
- Global Burden of Disease database
- Bureau of Justice Statistics
- Bureau of Labor Statistics
- Dept of Housing and Urban Development

Details on how we estimated each are in this spreadsheet, and appendix A presents all the conditional probability estimates used in this analysis. We also assessed the quality of the data behind each conditional probability as:

- Based on direct measurement (strongest)
- Based on indirect estimate, such as age standardization
- Educated guess, with reference to related data (weakest)
- Assumed to be zero overlap
- No reasonable or reliable estimate

Note that we used the same set of conditional probabilities for every county. We recognize that these probabilities likely vary from county to county; we leave that complexity for a future version of the VAPC.

Finally, we only estimated conditional probabilities for the bivariate case, not for higher dimensional relationships (membership in a group given membership in several other groups.) To facilitate computation, we assumed independence for all dimensions greater than two.

b. Resolving each pair of conditional probabilities

For each county, we cycled through each 2x2 table, pairing each condition with every other condition. We substituted the county-specific marginal values for $P(X)$ and $P(Y)$, and then considered both ways of solving for cell A: $P(X|Y)$ and $P(Y|X)$.

In order to choose which "direction" to use, we first considered cases where the conditional probability resulted in an impossible table, for example where $P(X|Y)*P(Y) > P(X)$. In that case, we used the conditional in the other direction to complete cell A. If both directions were possible, then we selected the option with better data quality. If the data quality was the same, then we selected the direction resulting in the smaller overlap. We tested using the smaller versus the larger overlap in these cases, and found that the smaller overlaps generally resulted in more

accurate estimates. Finally, in the rare case that neither direction worked (one single case in the first 1500 counties), we assumed independence between the margins.

c. Generating a covariance matrix for each county

Once cell A was established for the relationship between two groups in a county, then we calculated the tetrachoric correlation (ρ), a measure of correlation between two binary variables. Based on a review of methods for calculating ρ[13], we used:

$$\rho = \cos\left(\frac{\pi}{1 + \sqrt{AD/BC}}\right)$$

where A, B, C, and D are the probabilities in the 2x2 table. We then arranged the ρ's into a standard symmetric covariance matrix with 1's on the diagonal.

If the expected overlap is zero (A=0), then ρ works out to -1, which we changed to -.999 to avoid computational problems.

d. Monte Carlo simulation for each county

The Monte Carlo simulation drew random values from a 13-dimensional space of normally distributed continuous variables. For each county, the inputs included a vector of 13 means, all set to zero, and the covariance matrix between the 13 variables. In most cases, the covariance matrix was found to be non-positive definite, which we resolved by adding a quantity λ to the diagonal elements. λ was equal to the absolute value of the smallest (negative) eigenvalue of the matrix.

The output was a dataset with 1 million rows and 13 continuous variables Y_1 to Y_{13}, each with mean zero and standard deviation λ. We first standardized all 13 to have a standard deviation of 1 by dividing by λ. Then we generated binary group membership variables G_1 to G_{13} as:

$$G_g = \begin{cases} 1 \text{ if } Y_g < \text{probit}(M_g) \\ 0 \text{ if } Y_g \geq \text{probit}(M_g) \end{cases}$$

where M_g is the marginal proportion of group g in the county and probit(x) returns the percentile of a standard normal distribution such that x% of the standard normal distribution lies below it.

e. Final analytic dataset

In order to create our desired analytic dataset for each county, we used the simulated data with 1 million cases to generate frequencies of every possible combination of the 13 group variables. Some combinations of group membership did occur in the simulated data but were rare, resulting in negligible proportions. We removed these combinations based on the size of the county, according to the algorithm in Table 3.

Table 3: Algorithm for removing negligible proportions

If the county population is:	Approx. % of all counties that are this size	Then remove any combination of group membership with proportion less than:
Large (> 500,000)	5%	0.00001

---

[13] El-Hashash, E. F., & El-Absy, K. M. (2018). Methods for Determining the Tetrachoric Correlation Coefficient for Binary Variables. *Asian Journal of Probability and Statistics*, 2(3), 1-12.

Medium (<= 500,000 and >5,000)	85%	0.0001
Small (<= 5,000)	10%	0.001

Although 2^{13} rows per county were theoretically possible, in practice counties had on average about 150 rows each due to low marginal proportions, the many zero overlaps in the conditional probabilities, and the removal of negligible proportions. The final analytic dataset for all US counties and all groups has about 500,000 rows, and includes the variables shown in Table 4.

Table 4: Variables in the final analytic dataset

Variable	Description
State	State name
FIPS	County identifier
Countypop	Total population of the county
G_1 to G_{13}	Binary variables indicating membership in each group
P	Percent of people in this county who are members of the indicated groups
N	Number of people in this county who are members of the indicated groups

To calculate the marginal probabilities for a group G_X in a county, we sum the variable P across all rows where G_X = 1 in that county. To calculate the number of people to vaccinate in G_X and G_Y, we sum the variable N across all rows where G_X = 1 *or* G_Y = 1 in that county. To calculate the percent of people who overlap between groups G_X and G_Y, we sum the variable P across all rows where G_X = 1 *and* G_Y = 1 in that county.

f. Accuracy of marginal and overlap estimates

We checked the accuracy of our final analytic dataset against the expected marginal proportions in a sample of 1,500 counties. Of these 19,500 margins (1,500 * 13) in the simulated data, 99.5% are within 1/10 of 1% of the margins in the actual populations. The maximum difference is ½ of 1%.

Similarly, we checked the overlaps in the simulated data against the cell A estimates from each 2x2 table in each county. With 78 pairs of groups from the sample of 1,500 counties, we had 117,000 overlaps to check. Of those, 1,115 had zero margins, meaning zero overlap, which left 115,885 overlap pairs to check. Table 5 shows statistics on the absolute difference between the actual overlaps drawn from the conditional probabilities and the overlaps from the simulated data.

Table 5: Comparison of actual and simulated overlaps in 1,500 counties

	Absolute value of difference between actual and simulated overlaps between any two groups			
	N	Mean	Median	Maximum
Total	115,885	0.004%	0.001%	16.021%
By how cell A in the 2x2 table was resolved:				
Where only one direction worked:				
P(Y|X) worked, P(X|Y) did not	4,412	0.060%	0.021%	8.791%
P(X|Y) worked, P(Y|X) did not	9,216	0.056%	0.023%	1.619%

Both directions worked:				
P(Y|X) had better data quality	5,185	0.191%	0.083%	2.385%
P(X|Y) had better data quality	2,830	0.802%	0.376%	16.021%
Both directions worked and they had the same data quality:				
P(Y|X) had smaller overlap	14,600	0.545%	0.280%	5.228%
P(X|Y) had smaller overlap	21,141	0.805%	0.325%	5.369%
Neither direction worked, assumed independence	1	0.535%	0.535%	0.535%
Zero overlap	58,500	0.303%	0.011%	7.664%

The mean and median error in all groups are well under 1%, which suggests good matching between target and simulated overlaps. There are some outliers, however, up to fully 16 percentage points difference for one overlap in one county. We are investigating the cause of these and other outliers to address them.

F. VAPC function 2: Count available doses

The second function involves calculating the number of doses and courses that may be available for allocation in each state. The VAPC assumes an initial purchase of 10 million doses by Operation Warp Speed (OWS), and that the doses will be allocated to states proportional to overall population in each state. These assumptions may be changed as information develops, but at the moment these settings are not modifiable by the user.

If the user opts to take a 10% holdout, then the VAPC reduces the number of doses available by that amount.

Finally, if the user selects a 2-dose vaccine then the VAPC halves the number of doses available. At this point, the VAPC refers to "courses" of vaccines, where a course refers to a person-dose of supply. The number of courses equals the number of individual people who can be vaccinated with a given supply of doses.

With time, the vaccine landscape is likely to be more complex than a single vaccine requiring one or two doses. Several vaccines may gain approval, each potentially with different dosing, manufacturing characteristics, safety profiles, duration of protection, and efficacy in various populations, as an extensive modeling effort has revealed. We plan to update the VAPC as these circumstances evolve.

G. VAPC function 3: Allocate doses to counties

1. Proportional allocation

Once the VAPC has estimates of the selected population sizes by county, the overlaps between those groups, and the number of vaccine courses available for the state, it allocates the vaccines to counties proportional to the selected population. Table 6 presents a simple, fictional example in which there are 5,000 courses to allocate in a state with 200,000 people living in 8 counties. The user has selected priority populations 1 and 2 (phase 1a) to vaccinate. Columns A through E show each step in the allocation by distribution. In proportional distribution, the vaccine coverage of the priority populations is always the same across all counties and the state (18% in this example.)

2. Adjustment for SVI or CCVI

In addition to prioritizing the 13 populations in phases, the NASEM guidance recommends prioritizing allocation to geographic areas that are especially socially vulnerable. It suggests

using either the CDC's Social Vulnerability Index (SVI) or the COVID-19 Community Vulnerability Index (CCVI). Both of these measures are rank scores ranging from 0 to 1, where 1 represents more vulnerability, and both are publicly available on the county level. NASEM suggests selecting one of these measures and prioritizing allocation to the quartile of counties with the highest vulnerability.

We implemented this prioritization in the VAPC as follows, illustrated by columns F through M in Table 6. We first distributed 90% of the available vaccine courses proportional to population across all counties as described above. We then identified the quartile of counties with the highest vulnerability according to the measure chosen by the user: SVI or CCVI. We allocated the remaining 10% of courses among the vulnerable counties, proportional to the eligible population within that quartile. The final vaccine coverage in the top quartile (26% in this example) is always higher than in the rest of the counties (16% in this example), but the vaccine coverage *within* these two groups will always be consistent.

Note that the SVI and CCVI do not always select the same counties into the top quartile.

Table 6: Example allocations proportional to population and by SVI or CCVI

	Setup			Proportional to population		Prioritized by SVI or CCVI			Top quartile only				
	Total pop.	Eligible pop.	% dist. of eligible pop.	Courses allocated	Vaccine coverage	90% of courses allocated	SVI or CCVI rank	Highest quartile?	Eligible pop.	% dist of eligible pop.	10% of courses allocated	Total courses allocated	Vaccine coverage
	From margins data	From analysis data	Col B / Total col B	Col C * 5,000 courses	Col D / Col B	Col C * 90% of 5,000 courses	From SVI/CCVI data		Col B	Col I / Total col I	Col J * 10% of 5,000 courses	Col F + Col K	Col L / Col B
	A	B	C	D	E	F	G	H	I	J	K	L	M
County A	10,000	6,000	22%	1,091	18%	982	0.450	No				982	16%
County B	20,000	2,000	7%	364	18%	327	0.521	No				327	16%
County C	40,000	2,000	7%	364	18%	327	0.645	No				327	16%
County D	30,000	10,000	36%	1,818	18%	1,636	0.694	No				1,636	16%
County E	10,000	2,000	7%	364	18%	327	0.703	No				327	16%
County F	20,000	500	2%	91	18%	82	0.769	No				82	16%
County G	40,000	2,000	7%	364	18%	327	0.811	Yes	2,000	40%	200	527	26%
County H	30,000	3,000	11%	545	18%	491	0.945	Yes	3,000	60%	300	791	26%
Totals	200,000	27,500	100%	5,000	18%	4,500			5,000	100%	500	5,000	18%

H. Refining the methods

The VAPC methods are not static. We intend to carry out additional iterations, tests, and refinements to the method and the scope of the tool. Here are some future developments we are considering:

- Include Puerto Rico and US territories
- Incorporate recommendations and populations from other governing bodies, such as ACIP, when they become available. These would be options for the user to select.
- Give ranges rather than point estimates.
- Perform sensitivity analyses on our assumptions regarding marginal and conditional probabilities.
- Improve our county-level estimation of the populations with exactly 1 versus 2+ severe comorbidities
- Resolve the outliers on overlap estimates, including improving our estimation of zero overlaps.
- Use state-specific conditional probability tables.
- Allow users to change the size of the Operation Warp Speed purchase or their state's share of doses.
- Allow users to upload their own estimates of the county-level populations of the 13 groups.

.

Appendix A: Conditional probability estimates

Estimated conditional probabilities for all pairs of group memberships

		Est. conditional probability		Data quality*	Comments
Probability of being in group **1 Health care workers**, given membership in:					
1 Health care workers	P(1|1)	100%		1	Logical relationship.
2 First responders	P(1|2)	0%		2	Assumption: All occupations are mutually exclusive (occupations: 1,2,5,6,13)
3 Two+ comorbidities	P(1|3)	2.31%		4	See tab "Occupation given comorbidities" in [calculations].
4 Older in congregate	P(1|4)	0%		2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/ incarcerated/ older in congregate settings
5 Critical risk 1	P(1|5)	0%		2	Assumption: All occupations are mutually exclusive (occupations: 1,2,5,6,13)
6 Teachers	P(1|6)	0%		2	Assumption: All occupations are mutually exclusive (occupations: 1,2,5,6,13)
7 One comorbidity	P(1|7)	3.30%		4	See tab "Occupation given comorbidities" in [calculations].
8 All older	P(1|8)	1.34%		3	See tab "Occupation age structure" in [calculations].
9 Homeless	P(1|9)	0%		2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
10 Incarcerated	P(1|10)	0%		2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
11 Young adults	P(1|11)	4.31%		3	See tab "Occupation age structure" in [calculations].
12 Children	P(1|12)	0%		2	Assumption: Children do not have jobs, and those with jobs are not children
13 Critical risk 2	P(1|13)	0%		2	Assumption: All occupations are mutually exclusive (occupations: 1,2,5,6,13)
Probability of being in group **2 First responders**, given membership in:					
1 Health care workers	P(2|1)	0%		2	Assumption: All occupations are mutually exclusive (occupations: 1,2,5,6,13)
2 First responders	P(2|2)	100%		1	Logical relationship.
3 Two+ comorbidities	P(2|3)	0.27%		4	See tab "Occupation given comorbidities" in [calculations].
4 Older in congregate	P(2|4)	0%		2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings

5 Critical risk 1	P(2|5)	0%	2	Assumption: All occupations are mutually exclusive (occupations: 1,2,5,6,13)
6 Teachers	P(2|6)	0%	2	Assumption: All occupations are mutually exclusive (occupations: 1,2,5,6,13)
7 One comorbidity	P(2|7)	0.39%	4	See tab "Occupation given comorbidities" in [calculations]().
8 All older	P(2|8)	0.07%	6	Prefer P(8|2). See tab "Occupation age structure" in [calculations]().
9 Homeless	P(2|9)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
10 Incarcerated	P(2|10)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
11 Young adults	P(2|11)	0.45%	6	Prefer P(11|2). See tab "Occupation age structure" in [calculations]().
12 Children	P(2|12)	0%	2	Assumption: Children do not have jobs, and those with jobs are not children
13 Critical risk 2	P(2|13)	0%	2	Assumption: All occupations are mutually exclusive (occupations: 1,2,5,6,13)

Probability of being in group **3 Two+ comorbidities**, given membership in:

1 Health care workers	P(3|1)	6.98%	4	See tab "Comorbidities given occupation" in [calculations]().
2 First responders	P(3|2)	5.25%	4	See tab "Comorbidities given occupation" in [calculations]().
3 Two+ comorbidities	P(3|3)	100%	1	Logical relationship.
4 Older in congregate	P(3|4)	38.72%	5	See tab "Congregate settings" in [calculations]().
5 Critical risk 1	P(3|5)	7.45%	4	See tab "Comorbidities given occupation" in [calculations]().
6 Teachers	P(3|6)	7.13%	4	See tab "Comorbidities given occupation" in [calculations]().
7 One comorbidity	P(3|7)	0%	1	Severe comorbidity (2+ conditions) and moderate comorbidity (1 condition only) are mutually exclusive (groups 3 and 7)
8 All older	P(3|8)	33.67%	3	See tab "Comorbidities and age" in [calculations]().
9 Homeless	P(3|9)	10.70%	5	See tab "Homeless and comorbidities" in [calculations]().
10 Incarcerated	P(3|10)	16.45%	3	See tab "Incarcerated and comorbidities" in [calculations]().
11 Young adults	P(3|11)	0.55%	3	See tab "Comorbidities and age" in [calculations]().
12 Children	P(3|12)	0.19%	3	See tab "Comorbidities and age" in [calculations]().
13 Critical risk 2	P(3|13)	5.77%	4	See tab "Comorbidities given occupation" in [calculations]().

Probability of being in group **4 Older in congregate**, given membership in:

1 Health care workers	P(4|1)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
2 First responders	P(4|2)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
3 Two+ comorbidities	P(4|3)	6.91%	4	See tab "Comorbidities and age" in calculations.
4 Older in congregate	P(4|4)	100%	1	Logical relationship.
5 Critical risk 1	P(4|5)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
6 Teachers	P(4|6)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
7 One comorbidity	P(4|7)	3.96%	4	See tab "Comorbidities and age" in calculations.
8 All older	P(4|8)	n/a	8	Calculated separately for each county as number of older adults in congregate settings / all older adults
9 Homeless	P(4|9)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings are mutually exclusive groups
10 Incarcerated	P(4|10)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings are mutually exclusive groups
11 Young adults	P(4|11)	0%	1	Children, young adults, and older adults are mutually exclusive groups (also affects older adults in congregate settings)
12 Children	P(4|12)	0%	1	Children, young adults, and older adults are mutually exclusive groups (also affects older adults in congregate settings)
13 Critical risk 2	P(4|13)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
Probability of being in group **5 Critical risk workers 1**, given membership in:				
1 Health care workers	P(5|1)	0%	2	Assumption: All occupations are mutually exclusive (occupations: 1,2,5,6,13)
2 First responders	P(5|2)	0%	2	Assumption: All occupations are mutually exclusive (occupations: 1,2,5,6,13)
3 Two+ comorbidities	P(5|3)	4.76%	4	See tab "Occupation given comorbidities" in calculations.
4 Older in congregate	P(5|4)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
5 Critical risk 1	P(5|5)	100%	1	Logical relationship.

6 Teachers	P(5|6)	0%	2	Assumption: All occupations are mutually exclusive (occupations: 1,2,5,6,13)
7 One comorbidity	P(5|7)	6.80%	4	See tab "Occupation given comorbidities" in [calculations].
8 All older	P(5|8)	3.10%	3	See tab "Occupation age structure" in [calculations].
9 Homeless	P(5|9)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
10 Incarcerated	P(5|10)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
11 Young adults	P(5|11)	8.77%	3	See tab "Occupation age structure" in [calculations].
12 Children	P(5|12)	0%	2	Assumption: Children do not have jobs, and those with jobs are not children
13 Critical risk 2	P(5|13)	0%	2	Assumption: Critical risk workers of type 1 and 2 are mutually exclusive (groups 5 and 13)
Probability of being in group **6 Teachers**, given membership in:				
1 Health care workers	P(6|1)	0%	2	Assumption: All occupations are mutually exclusive (occupations: 1,2,5,6,13)
2 First responders	P(6|2)	0%	2	Assumption: All occupations are mutually exclusive (occupations: 1,2,5,6,13)
3 Two+ comorbidities	P(6|3)	2.37%	4	See tab "Occupation given comorbidities" in [calculations].
4 Older in congregate	P(6|4)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
5 Critical risk 1	P(6|5)	0%	2	Assumption: All occupations are mutually exclusive (occupations: 1,2,5,6,13)
6 Teachers	P(6|6)	100%	1	Logical relationship.
7 One comorbidity	P(6|7)	3.38%	4	See tab "Occupation given comorbidities" in [calculations].
8 All older	P(6|8)	1.47%	3	See tab "Occupation age structure" in [calculations].
9 Homeless	P(6|9)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
10 Incarcerated	P(6|10)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
11 Young adults	P(6|11)	4.17%	3	See tab "Occupation age structure" in [calculations].
12 Children	P(6|12)	0%	2	Assumption: Children do not have jobs, and those with jobs are not children

13 Critical risk 2	P(6|13)	0%	2	Assumption: All occupations are mutually exclusive (occupations: 1,2,5,6,13)

Probability of being in group 7 One comorbidity, given membership in:

1 Health care workers	P(7|1)	20.87%	4	See tab "Comorbidities given occupation" in calculations.
2 First responders	P(7|2)	19.25%	4	See tab "Comorbidities given occupation" in calculations.
3 Two+ comorbidities	P(7|3)	0%	1	Severe comorbidity (2+ conditions) and mdoerate comorbidity (1 condition only) are mutually exclusive (groups 3 and 7)
4 Older in congregate	P(7|4)	48.73%	5	See tab "Congregate settings" in calculations.
5 Critical risk 1	P(7|5)	21.51%	4	See tab "Comorbidities given occupation" in calculations.
6 Teachers	P(7|6)	21.13%	4	See tab "Comorbidities given occupation" in calculations.
7 One comorbidity	P(7|7)	100%	1	Logical relationship.
8 All older	P(7|8)	42.38%	3	See tab "Comorbidities and age" in calculations.
9 Homeless	P(7|9)	23.47%	5	See tab "Homeless and comorbidities" in calculations.
10 Incarcerated	P(7|10)	24.75%	3	See tab "Incarcerated and comorbidities" in calculations.
11 Young adults	P(7|11)	6.12%	3	See tab "Comorbidities and age" in calculations.
12 Children	P(7|12)	2.94%	3	See tab "Comorbidities and age" in calculations.
13 Critical risk 2	P(7|13)	18.46%	4	See tab "Comorbidities given occupation" in calculations.

Probability of being in group 8 All older, given membership in:

1 Health care workers	P(8|1)	6.42%	3	See tab "Occupation age structure" in calculations.
2 First responders	P(8|2)	2.68%	3	See tab "Occupation age structure" in calculations.
3 Two+ comorbidities	P(8|3)	62.78%	3	See tab "Comorbidities and age" in calculations.
4 Older in congregate	P(8|4)	100%	1	Logical relationship.
5 Critical risk 1	P(8|5)	7.18%	3	See tab "Occupation age structure" in calculations.
6 Teachers	P(8|6)	6.85%	3	See tab "Occupation age structure" in calculations.
7 One comorbidity	P(8|7)	36.03%	3	See tab "Comorbidities and age" in calculations.
8 All older	P(8|8)	100%	1	Logical relationship.
9 Homeless	P(8|9)	6.00%	4	See tab "Homeless age structure" in calculations.
10 Incarcerated	P(8|10)	2.94%	3	See tab "Incarcerated age structure" in calculations.
11 Young adults	P(8|11)	0%	1	Children, young adults, and older adults are mutually exclusive groups (also affects older adults in congregate settings)
12 Children	P(8|12)	0%	1	Children, young adults, and older adults are mutually exclusive groups (also affects older adults in congregate settings)
13 Critical risk 2	P(8|13)	4.59%	3	See tab "Occupation age structure" in calculations.

Probability of being in group 9 Homeless, given membership in:

1 Health care workers	P(9|1)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with

				jobs are not homeless/incarcerated/older in congregate settings
2 First responders	P(9|2)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
3 Two+ comorbidities	P(9|3)	n/a	7	Use P(3|9) due to very different population sizes.
4 Older in congregate	P(9|4)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings are mutually exclusive groups
5 Critical risk 1	P(9|5)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
6 Teachers	P(9|6)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
7 One comorbidity	P(9|7)	n/a	7	Use P(7|9) due to very different population sizes.
8 All older	P(9|8)	n/a	7	Use P(8|9) instead. See tab "Homeless age structure" in calculations.
9 Homeless	P(9|9)	100%	1	Logical relationship.
10 Incarcerated	P(9|10)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings are mutually exclusive groups
11 Young adults	P(9|11)	n/a	7	Use P(11|9) instead. See tab "Homeless age structure" in calculations.
12 Children	P(9|12)	n/a	7	Use P(12|9) instead. See tab "Homeless age structure" in calculations.
13 Critical risk 2	P(9|13)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
Probability of being in group **10 Incarcerated**, given membership in:				
1 Health care workers	P(10|1)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
2 First responders	P(10|2)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
3 Two+ comorbidities	P(10|3)	n/a	7	Use P(3|10) due to very different population sizes.
4 Older in congregate	P(10|4)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings are mutually exclusive groups
5 Critical risk 1	P(10|5)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings

6 Teachers	P(10|6)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
7 One comorbidity	P(10|7)	n/a	7	Use P(7|10) due to very different population sizes.
8 All older	P(10|8)	n/a	7	Use P(8|10). See tab "Incarcerated age structure" in calculations.
9 Homeless	P(10|9)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings are mutually exclusive groups
10 Incarcerated	P(10|10)	100%	1	Logical relationship.
11 Young adults	P(10|11)	0.64%	6	Prefer P(11|10). See tab "Incarcerated age structure" in calculations.
12 Children	P(10|12)	n/a	7	Use P(12|10). See tab "Incarcerated age structure" in calculations.
13 Critical risk 2	P(10|13)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
Probability of being in group **11 Young adults**, given membership in:				
1 Health care workers	P(11|1)	21.19%	3	See tab "Occupation age structure" in calculations.
2 First responders	P(11|2)	18.65%	3	See tab "Occupation age structure" in calculations.
3 Two+ comorbidities	P(11|3)	1.06%	3	See tab "Comorbidities and age" in calculations.
4 Older in congregate	P(11|4)	0%	1	Children, young adults, and older adults are mutually exclusive groups (also affects older adults in congregate settings)
5 Critical risk 1	P(11|5)	20.94%	3	See tab "Occupation age structure" in calculations.
6 Teachers	P(11|6)	20%	3	See tab "Occupation age structure" in calculations.
7 One comorbidity	P(11|7)	5.36%	3	See tab "Comorbidities and age" in calculations.
8 All older	P(11|8)	0%	1	Children, young adults, and older adults are mutually exclusive groups (also affects older adults in congregate settings)
9 Homeless	P(11|9)	18.00%	3	See tab "Homeless age structure" in calculations.
10 Incarcerated	P(11|10)	24.65%	3	See tab "Incarcerated age structure" in calculations.
11 Young adults	P(11|11)	100%	1	Logical relationship.
12 Children	P(11|12)	0%	1	Children, young adults, and older adults are mutually exclusive groups (also affects older adults in congregate settings)
13 Critical risk 2	P(11|13)	28.19%	3	See tab "Occupation age structure" in calculations.
Probability of being in group **12 Children**, given membership in:				
1 Health care workers	P(12|1)	0%	2	Assumption: Children do not have jobs, and those with jobs are not children
2 First responders	P(12|2)	0%	2	Assumption: Children do not have jobs, and those with jobs are not children
3 Two+ comorbidities	P(12|3)	0.33%	3	See tab "Comorbidities and age" in calculations.

4 Older in congregate	P(12|4)	0%	1	Children, young adults, and older adults are mutually exclusive groups (also affects older adults in congregate settings)
5 Critical risk 1	P(12|5)	0%	2	Assumption: Children do not have jobs, and those with jobs are not children
6 Teachers	P(12|6)	0%	2	Assumption: Children do not have jobs, and those with jobs are not children
7 One comorbidity	P(12|7)	2.35%	3	See tab "Comorbidities and age" in calculations.
8 All older	P(12|8)	0%	1	Children, young adults, and older adults are mutually exclusive groups (also affects older adults in congregate settings)
9 Homeless	P(12|9)	8.00%	3	See tab "Homeless age structure" in calculations.
10 Incarcerated	P(12|10)	2.56%	3	See tab "Incarcerated age structure" in calculations.
11 Young adults	P(12|11)	0%	1	Children, young adults, and older adults are mutually exclusive groups (also affects older adults in congregate settings)
12 Children	P(12|12)	100%	1	Logical relationship.
13 Critical risk 2	P(12|13)	0%	2	Assumption: Children do not have jobs, and those with jobs are not children

Probability of being in group **13 Critical risk workers 2**, given membership in:

1 Health care workers	P(13|1)	0%	2	Assumption: All occupations are mutually exclusive (occupations: 1,2,5,6,13)
2 First responders	P(13|2)	0%	2	Assumption: All occupations are mutually exclusive (occupations: 1,2,5,6,13)
3 Two+ comorbidities	P(13|3)	8.58%	4	See tab "Occupation given comorbidities" in calculations.
4 Older in congregate	P(13|4)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
5 Critical risk 1	P(13|5)	0%	2	Assumption: Critical risk workers of type 1 and 2 are mutually exclusive (groups 5 and 13)
6 Teachers	P(13|6)	0%	2	Assumption: All occupations are mutually exclusive (occupations: 1,2,5,6,13)
7 One comorbidity	P(13|7)	12.24%	4	See tab "Occupation given comorbidities" in calculations.
8 All older	P(13|8)	3.57%	3	See tab "Occupation age structure" in calculations.
9 Homeless	P(13|9)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
10 Incarcerated	P(13|10)	0%	2	Assumption: Homeless/incarcerated/older in congregate settings do not have jobs, and those with jobs are not homeless/incarcerated/older in congregate settings
11 Young adults	P(13|11)	21.26%	3	See tab "Occupation age structure" in calculations.

12 Children	P(13|12)	0%	2	Assumption: Children do not have jobs, and those with jobs are not children
13 Critical risk 2	P(13|13)	100%	1	Logical relationship.

* Data quality codes

1. Logical relationship
2. Assumed to be zero overlap
3. Based on direct measurement (strongest)
4. Based on indirect estimate, such as age standardization
5. Educated guess, with reference to related data (weakest)
6. Estimate is unreliable; prefer its mirror image
7. Estimate is not available at all; use its mirror image
8. Calculate separately by county (one probability only)