

Health worker density

Input data

Our estimates of health worker densities leveraged nationally representative cross-sectional surveys and censuses conducted between 1990 and 2019 to identify members of the general working-age population who self-reported current active employment in a health-related occupation. All such published sources were identified and obtained either through the Global Health Data Exchange (GHDx) using occupation-related keyword searches (eg, “Occupations,” “Occupational risk factors”) or through the International Labour Organization’s (ILO) tabulations and database of employment-related surveys and censuses. Working age was defined as ages 15 to 69, since this was the standard age range for most labour force surveys. The two types of indicators extracted from these sources were employment ratios and occupation distributions.

The employment ratio indicator reflects the proportion of the working-age population that self-reports currently being employed. The corresponding dataset consisted of all downloadable data tabulations of employment-to-population ratios by age and sex available from the ILO, as well as employment levels that we extracted directly from individual-level survey microdata obtained through the GHDx.

Typical survey questions for identifying active employment asked individuals if they had worked for at least one hour in the previous seven days in any of the following capacities: for a wage, as an apprentice, in self-employment, or for a family business. Those reporting only a temporary absence from such a position in the preceding seven days were also considered to be employed.

Among employed individuals, occupation distributions were determined from a respondent’s self-reported description of their main job in the previous seven days, or their typical main job if they were temporarily absent from work. The corresponding extracted indicator was the proportion of the employed population working in different occupational categories. Occupation descriptions were coded according to the source’s occupational coding system, often a version of the International Standard Classification of Occupations (ISCO) or a closely related system. The distribution of our data across these coding systems is summarized in table 2. Such systems apply standard criteria for classifying distinct occupations. Details on the classification criteria for each occupation within ISCO can be found on the ILO website. Coding systems also range dramatically in granularity, typically reflected by the length of the codes themselves, and can thus classify respondents into broadly defined categories of work or provide very specific descriptions of their occupations.

Table 2. Input sources by coding system and level of granularity

Coding system	Number of country-years
Country-specific coding system	78
ISCO 08 4-digit	3038
ISCO 08 3-digit	181
ISCO 88 4-digit	322
ISCO 88 3-digit	588

Total	4207
-------	------

To meet this study's inclusion criteria for occupation data, a survey needed to code respondent occupations to a level of granularity that could identify health workers directly or identify a small aggregate group from which health worker cadres could be accurately split out. This level of granularity corresponds to ISCO three-digit and four-digit codes, as well as those alternate coding systems that could be mapped to ISCO at such a level of detail based on available documentation. This requirement restricted many of the eligible sources, since most standard international survey series and censuses do not code occupations to the level of detail required to identify health workers, let alone specific cadres of health workers. Due to the large number of distinct occupations at this level of granularity, few identified sources released tabulated data that met our inclusion criteria. Consequently, included population-based sources were those for which individual-level microdata could be accessed to allow direct extraction of occupational codes (employment ratio data were directly extracted from these sources as well). In addition to microdata files already accessible through the GHDx, we searched the ILO database of sources containing occupation data coded to at least two digits of ISCO granularity to identify publicly available microdata that also met this study's inclusion criteria. All such sources were added to the GHDx and subsequently extracted. In total, 1469 microdata sources with employment and occupation data were identified and extracted through the GHDx with 1097 country-years of data, 66 of which were censuses and 1031 of which were surveys. Surveys included labour force surveys and household surveys with sufficient labour-related questions. Given the similar nature of both survey types, distinctions between the two are not made in this paper. Both are included in any reference to "labour force participation surveys" in the main text and appendix.

In addition to the survey and census data, we included data from the World Health Organization (WHO) Global Health Observatory database¹⁷. The WHO Global Health Observatory database reports health worker data with sufficient granularity (four digit ISCO-08) to meet our inclusion criteria. By incorporating this data source, we added 2,950 country-years of data for the four main cadres in our study: physicians, nurses and midwives, pharmacists and pharmacist technicians, and dentists and dental assistants. Figures 2a – 2d summarize the total number of country-years comprised in all data sources employed in this study.

Figure 2a. Country-years of occupation data for physicians, by country and territory, 1990-2019

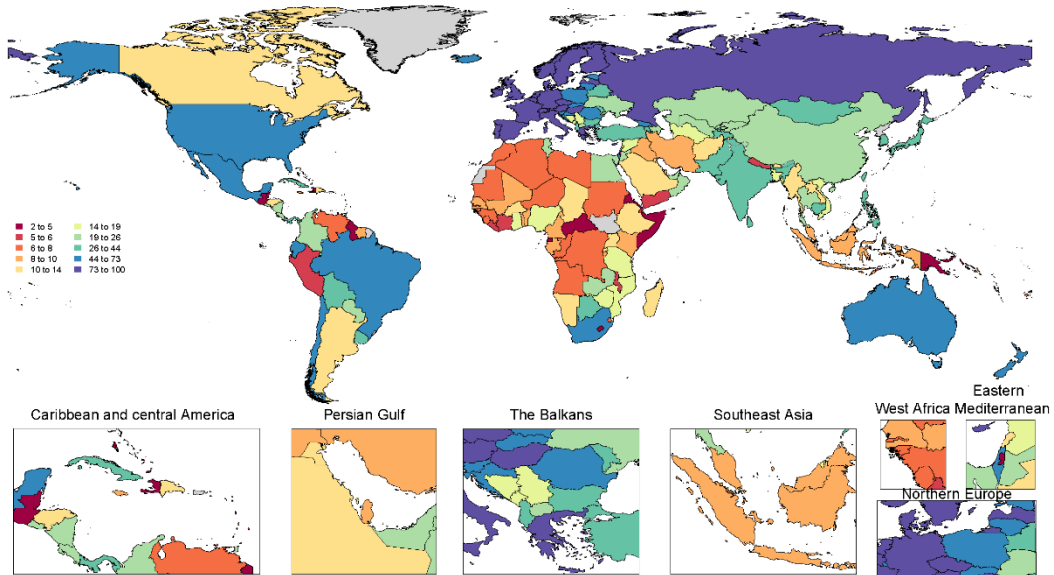


Figure 2b. Country-years of occupation data for nurses and midwives, by country and territory, 1990-2019

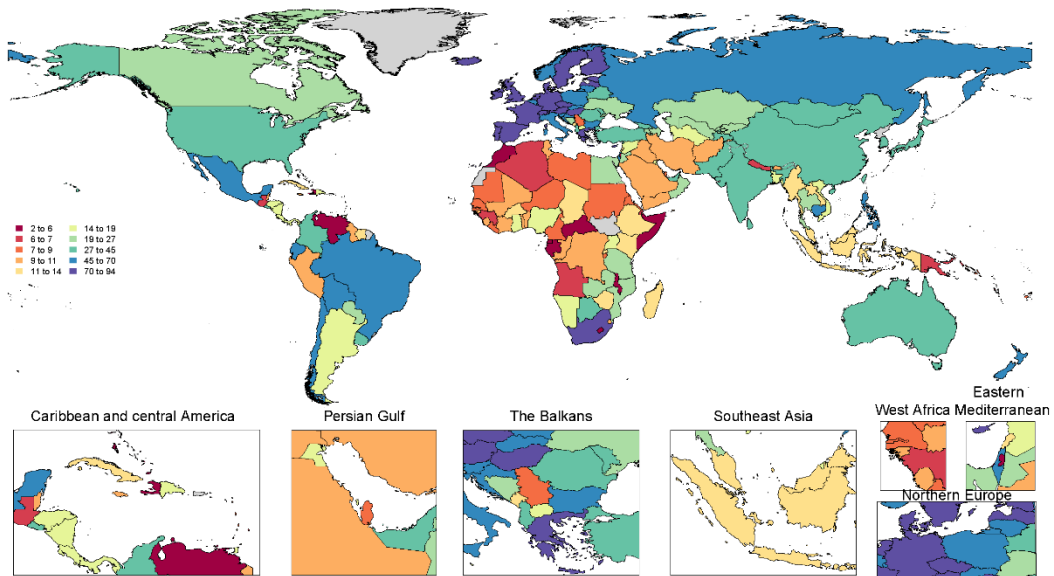


Figure 2c. Country-years of occupation data for dentists, by country and territory, 1990-2019

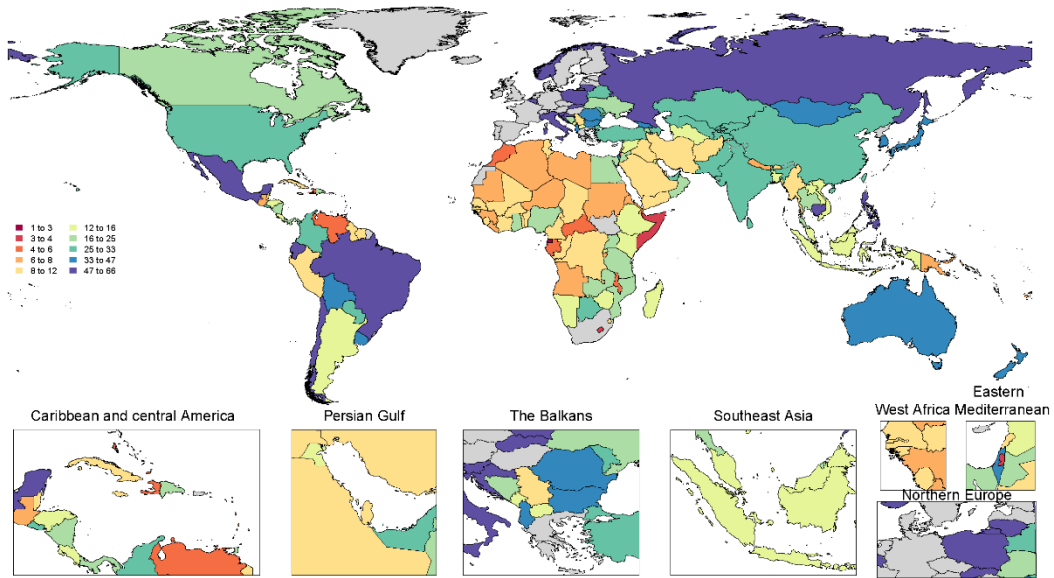
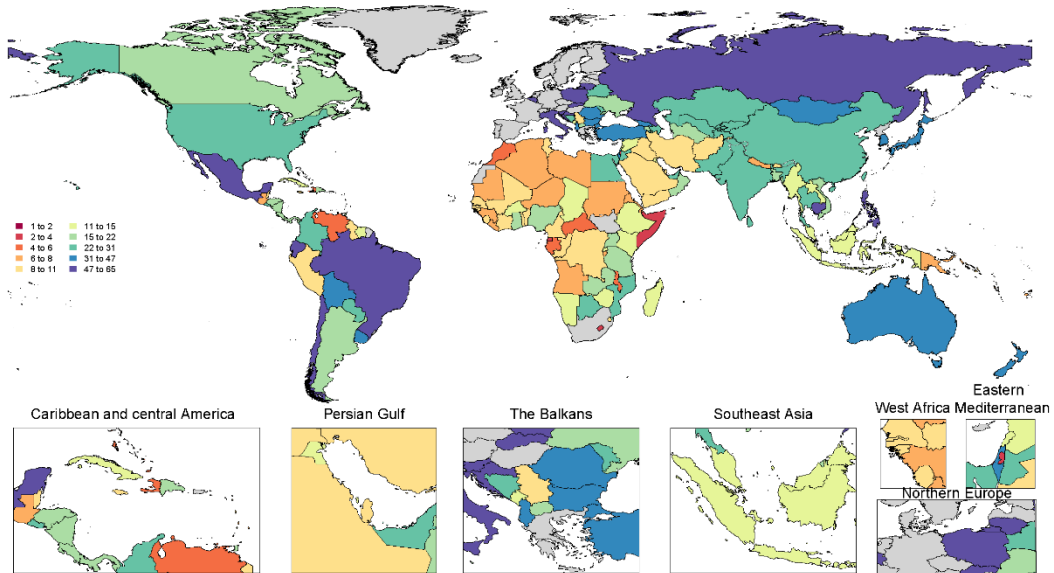


Figure 2d. Country-years of occupation data for pharmacists, by country and territory, 1990-2019



This study considers surveys and censuses to be the gold standard for measuring human resources for health because they are both population-based sources. Both sources rely on population-based sampling and tend to employ nearly identical occupational questionnaires and coding structures. While surveys and censuses do differ in their sample sizes and corresponding sampling errors, these are random errors that do not constitute systematic biases. Therefore, we refer to surveys and censuses as comparable source types, even though they may differ in their level of precision. Labor force surveys and censuses avoid many of the potential biases affecting other data sources, such as concerns related to the quality, coverage, and maintenance of administrative records, payrolls, and registries, which can result in both over-reporting and under-reporting of HRH levels as well as double-counting. For instance, WHO reports that the data in the Global Health Observatory database are reported by countries themselves through the National Health Workforce Accounts reporting portal, but limited information is publicly available about the source of data, any assumptions or modelling used to produce the data, whether data only captured the public sector, whether the data could have double-counting, and any other features. We provide a detailed discussion of the adjustment process we use for the WHO data in section 1.4 below.

Defining health worker cadres

We referred to the WHO Handbook on Monitoring and Evaluation of Human Resources for Health to create a list of relevant health worker cadres identified by four-digit ISCO 88 codes, the highest level of granularity in the coding system.⁵ While some of these cadres are themselves an aggregation of multiple types of health workers that we would have wished to identify individually, we were constrained by the preponderance of data using coding systems that did not provide such levels of detail in their coding structures. The included cadres and their corresponding occupation codes are listed in table 3.

Table 3. Cadres and corresponding occupation codes

Health worker cadre	4-digit ISCO 88 code
Physicians*	2221
Nurses and midwives *	2230, 3231, 3232
Pharmacists*	2224
Pharmaceutical assistants*	3228
Dentists*	2222
Dental assistants*	3225
Physiotherapists and prosthetic technicians	3226
Medical imaging and therapeutic equipment technicians	3133
Medical laboratory technicians	3211
Clinical officers, medical assistants, and community health workers	3221
Health-care aides and ambulance workers	5132, 5139

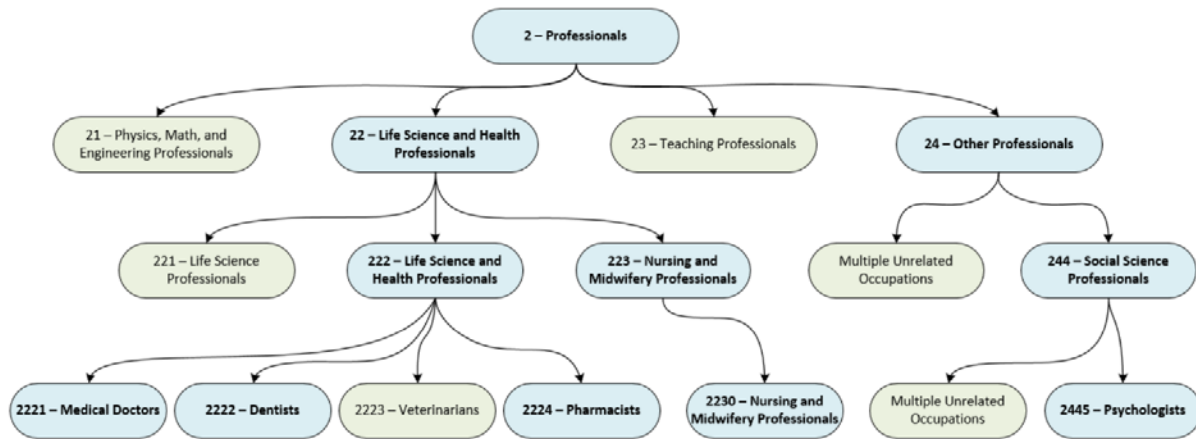
Environmental health workers	3222
Optometrists and opticians	3224
Dieticians and nutritionists	3223
Audiologists, speech therapists, and counsellors	3229
Psychologists	2445
Home-based personal care workers	5133
Traditional and complementary practitioners	3241

*

Included in the definition of SDG indicator 3.c.1 and in this study’s minimum threshold analyses. Similar cadres were grouped together in this study, such that the 20 occupations listed above were consolidated into 16 cadres for the purposes of the analysis. See the threshold analysis section for additional details on cadre groupings.

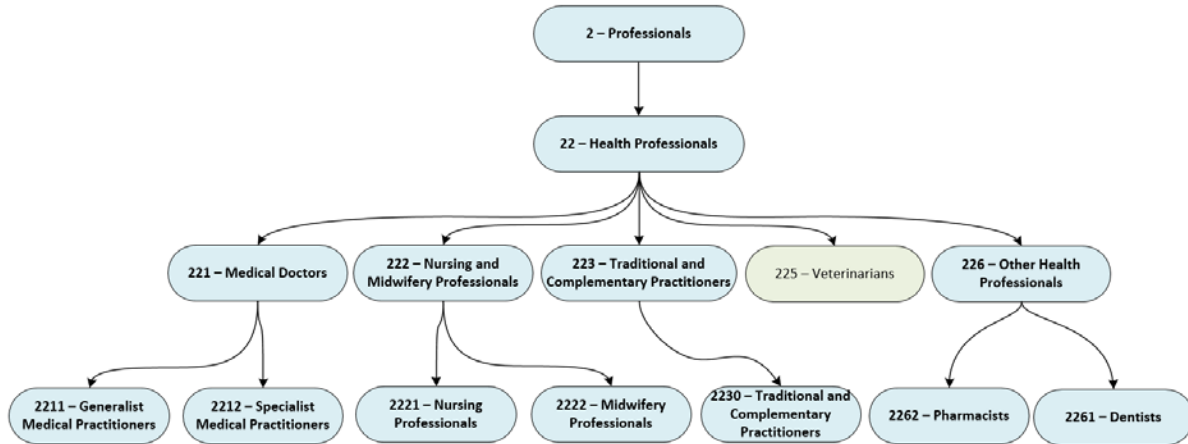
The ISCO has standardised occupation codes, with documentation describing each in detail, and has organised them hierarchically. Codes for the most general category of occupations are one digit in length. Occupational sub-categories are represented by two-digit codes, and further sub-specialties are represented with three-digit and four-digit codes. ISCO also consists of multiple versions, which were developed over time to update the system to reflect modern labour markets. The two versions included in this analysis are ISCO 88 and ISCO 08, released in 1988 and 2008, respectively, which thus cover the majority of the study’s time period of interest. Although the two versions of ISCO are generally similar in their distinctions between occupations, they differ in their hierarchical structuring of the coding system, with ISCO 08 consolidating health workers at higher levels of the hierarchy and facilitating the identification of more detailed occupations at its most granular level.⁶ However, ISCO 88 was substantially more common among identified sources. Mapping between versions inherently results in some loss of precision, and utilising the greater granularity of ISCO 08 for a few cadres would require using available ISCO 08 data to split out the less detailed ISCO 88 categories into smaller occupations. Given the paucity of ISCO 08 sources and the limited temporal overlap in the use of both versions, mapping all data to the ISCO 08 system and splitting less granular codes as necessary was not tenable for this analysis. Consequently, we used ISCO 88 as the gold standard according to which the HRH cadres in this study are defined. At a point when a much greater proportion of surveys rely on ISCO 08, future iterations of this analysis might be able to use ISCO 08 and thus generate more granular cadre-specific estimates. Figures 2 and 3 below present examples of these two coding frameworks.

Figure 3a. Example subsection of the ISCO 88 hierarchy



Note: only relevant occupations shown.

Figure 3b. Example subsection of the ISCO 08 hierarchy



Note: only relevant occupations shown.

In figures 3a and 3b, occupational categories and their corresponding codes are shown in a tree structure, reflecting the respective hierarchies of the ISCO system. At the top, one-digit codes identify very general categories of occupation. At the bottom, longer codes use additional digits to differentiate occupations at greater levels of detail. Marked in blue are codes for the HRH cadres included in this

analysis, or less-granular codes which contain an HRH cadre within them. More details on the criteria used to differentiate occupational classifications are available on the ILO's website pertaining to ISCO.⁶

For some sources that claimed to use an ISCO version, microdata sometimes exhibited a few codes that did not adhere to standard ISCO frameworks. In all such cases, the few non-ISCO codes were slight deviations from the standard framework, with one or two of the last digits altered either to identify particular occupations not normally highlighted in ISCO, or to express ambiguity when an occupation description was not quite detailed enough for ISCO classification at the desired digit level of detail. Descriptions of such deviations were usually missing from source documentation. In the absence of explanatory documentation, we truncated individual ambiguous codes to the longest length that was valid in the corresponding ISCO system. For example, we truncated invalid four-digit codes to valid two-digit or three-digit codes. When the truncated code was known to be unrelated to the HRH cadre in question (eg, the code pertained to a non-specific type of farmer), which was usually the case, the truncation did not have an impact on the extracted data. In some cases, however, it was unclear whether the truncated code was related to an HRH cadre of interest (eg, a truncated code for social science professionals might have pertained to psychologists or to another non-HRH social scientist). If a source contained a large number of invalid codes with potential relevance to HRH, then the source's entire set of occupation codes was truncated by one digit, and the source was dropped entirely if truncation resulted in occupation codes of less than three digits. If a survey contained few invalid codes with potential relevance to HRH, then all codes were retained and any HRH-ambiguous codes were excluded from both numerator and denominator of the extracted data for only those HRH cadres for which they were ambiguous. For instance, a truncated code potentially pertaining to psychologists was excluded from both numerator and denominator for the psychologist data due to ambiguity, but was still included in the denominator for every other HRH cadre to avoid biasing their respective estimates. Consequently, some sources exhibited minor deviations in extracted sample sizes across the various cadres of interest.

Mapping and splitting

All usable data were mapped to ISCO 88 four-digit codes for this analysis, or were split to such codes if data corresponded to the less detailed three-digit level of granularity. When surveys used country-specific coding systems that were based on ISCO, mapping to ISCO codes could be accomplished easily. When surveys used coding not clearly based on ISCO, only occupations where associated descriptions sufficiently matched those of ISCO were mapped.

In order to map ISCO 08 codes to ISCO 88 codes, we referenced the ILO's ISCO concordance documentation. Frequently, precise matches between four-digit codes existed between versions, allowing exact mapping from one version to the other.⁶ At other times, a four-digit ISCO 88 code corresponded to an aggregation of multiple ISCO 08 codes, which also allowed exact mapping to ISCO 88. When occupation categories between versions did not exactly match or aggregate to one another, we created an approximate mapping, whereby the most common ISCO 88 code corresponding to an ISCO 08 code in the concordance documentation was considered to be the sole match and was mapped accordingly. Table 4 provides examples of health worker cadre mapping across ISCO versions.

Table 4. Examples of health worker cadre mapping across ISCO versions

Occupation titles	ISCO 08 code	ISCO 88 code	ISCO 88-defined health worker cadre	Type of mapping
Dentists	2261	2222	Dentists	Exact match
Clinical officers	2240	3221	Clinical officers, medical assistants, and community health workers	Exact aggregation
Medical assistants	3256	3221	Clinical officers, medical assistants, and community health workers	Exact aggregation
Community health workers	3253	3221	Clinical officers, medical assistants, and community health workers	Exact aggregation
Nursing Professionals	2221	2230	Nurses and midwives	Exact aggregation
Midwifery Professionals	2222	2230	Nurses and midwives	Exact aggregation
Nursing Associate Professionals	3221	2230	Nurses and midwives	Exact aggregation
Midwifery Associate Professionals	3222	2230	Nurses and midwives	Exact aggregation
Environmental health workers	2263	3222	Environmental health workers	Approximate match
Food inspectors	3257	3222	Not HRH (ISCO 08) / Environmental health workers (ISCO 88)	Approximate match
Occupational hygienists	2263	3152	Environmental health workers (ISCO 08) / Not HRH (ISCO 88)	Approximate match
Quality controllers, Electrical product inspector, etc.	3257	3152	Not HRH	Approximate match

Our objective in mapping across ISCO versions was to create consistency and retain as much information as possible, but some inconsistencies and information loss were unavoidable. The table above draws on ISCO concordance documentation to illustrate how we mapped four-digit health worker codes from ISCO 08 to ISCO 88.⁶ For example, the table shows that for dentists there is an exact match between ISCO 08 and ISCO 88, which made it possible to map dentists across versions with no information loss. For other occupation titles, matching to comparable occupation categories was possible, but it resulted in a loss of granularity. For example, while ISCO 08 placed clinical officers, medical assistants, and community health workers into separate occupational categories, each with a different code, ISCO 88 grouped these into one category with a single code. Similar aggregations of seemingly divergent occupations are seen in other ISCO-88 HRH codes, such as audiologists, speech therapists, and counsellors (for which counsellors encompass a broad range of specialties, including family planning and HIV).

In other cases, the grouping of occupations between versions could not be made entirely consistent without including large numbers of non-health-related occupations, as was the case with environmental health workers. ISCO 88 groups environmental health workers with food inspectors, whereas ISCO 08 groups them with occupational hygienists. Finding an exact match between versions would require including codes that also pertain to quality controllers, electrical product inspectors, and a wide range of

other non-health occupations. Instead of including a large number of unrelated professions, we settled on an approximate mapping between versions, such that the environmental health workers cadre identified from ISCO 88 sources would also include food inspectors, whereas the cadre identified from ISCO 08 sources would include occupational hygienists. This resulted in some inconsistency but was preferable to dropping the cadre entirely or diluting it with many unrelated occupations.

Three-digit ISCO occupation codes were common, particularly in the censuses included in this study. While three-digit ISCO codes generally did not allow us to identify specific HRH cadres, information from three-digit occupation codes was useful in creating envelopes from which individual health worker cadres could be split. To do this, we first used data extracted from surveys with four-digit codes to inform preliminary models of each health worker cadre. We used these preliminary models to split three-digit code data into cadre-specific estimates for each GBD location and year. This required modelling not only each HRH cadre, but also the residual categories made up of four-digit codes associated with each three-digit code of interest that did not correspond to any health worker cadre. The advantage of using four-digit data to split three-digit codes rather than enacting one global split is that the proportional makeup of health worker cadre estimates exhibited variation across space and time. Yet, because splits were informed exclusively by four-digit inputs, data prepared from three-digit surveys were dependent upon the quality and coverage of four-digit data.

For country-specific coding systems not derived from ISCO, there was considerable variation in the level of detail available for different health worker cadres, and the digit length of an occupation code was generally not a good predictor of granularity. A two-digit country-specific occupation coding system, for instance, might distinguish between multiple categories of physicians, while using only one code for all nurses and midwives. When it was possible to match or aggregate some country-specific codes to the four-digit ISCO 88 framework, we mapped those codes accordingly. Where codes could be split into multiple estimated cadres without residual groups, we used the method described above for three-digit ISCO codes. Other types of country-specific codes were not usable, as we did not have enough information to map them to ISCO 88. Consequently, sources using country-specific coding systems typically only provided data for a subset of the HRH cadres included in this analysis.

We also generated a dataset for estimating the aggregate of all health workers combined. Data for this group comprised the sum of density data for all cadres – after mapping and splitting – that were obtained from sources whose coding system allowed identification of every cadre in this study (including, for example, all sources that used ISCO 88 or 08 coding).

[Input data adjustment strategy](#)

We consider the survey and census data to be the “gold standard” for measuring human resources for health because they are population-based and nationally representative. In an effort to use all available sources of data, we also included the WHO Global Health Observatory database in our analysis. Besides their origin with countries, there is little information about how these data came about, including any potential biases. Thus, we needed to evaluate the WHO data and determine whether the data were systematically biased from the population-based data. To perform this analysis, we matched the WHO data with survey or census data by country and year, resulting in 2636 matched pairs. Where we had both a survey and a census data point, we took the average to compare to the WHO data point. We depict scatter plots of the matched pairs of WHO data and population-based data for the cadres with

more than 300 data points from the WHO, namely physicians, nurses and midwives, dentists and pharmacists (figure 4a – figure 4d).

Figure 4a. Matched pairs of IHME population-based points and WHO data points for physicians, 1990-2019

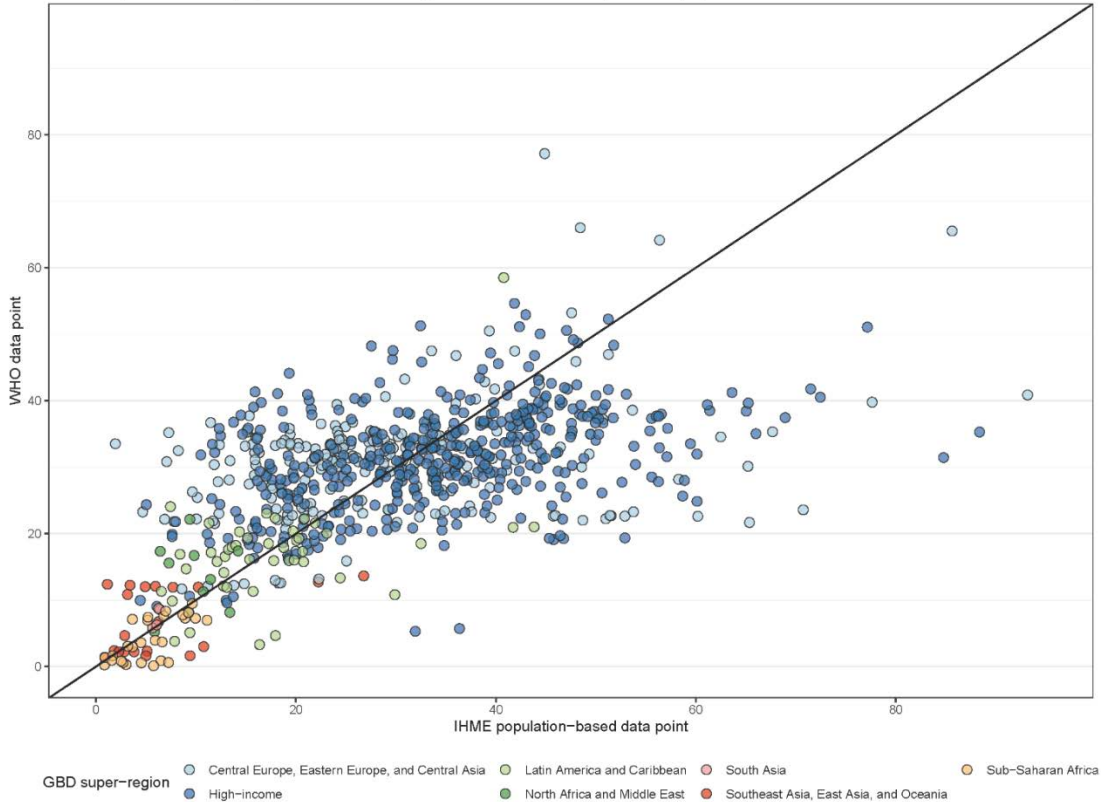


Figure 4b. Matched pairs of IHME population-based points and WHO data points for nurses and midwives

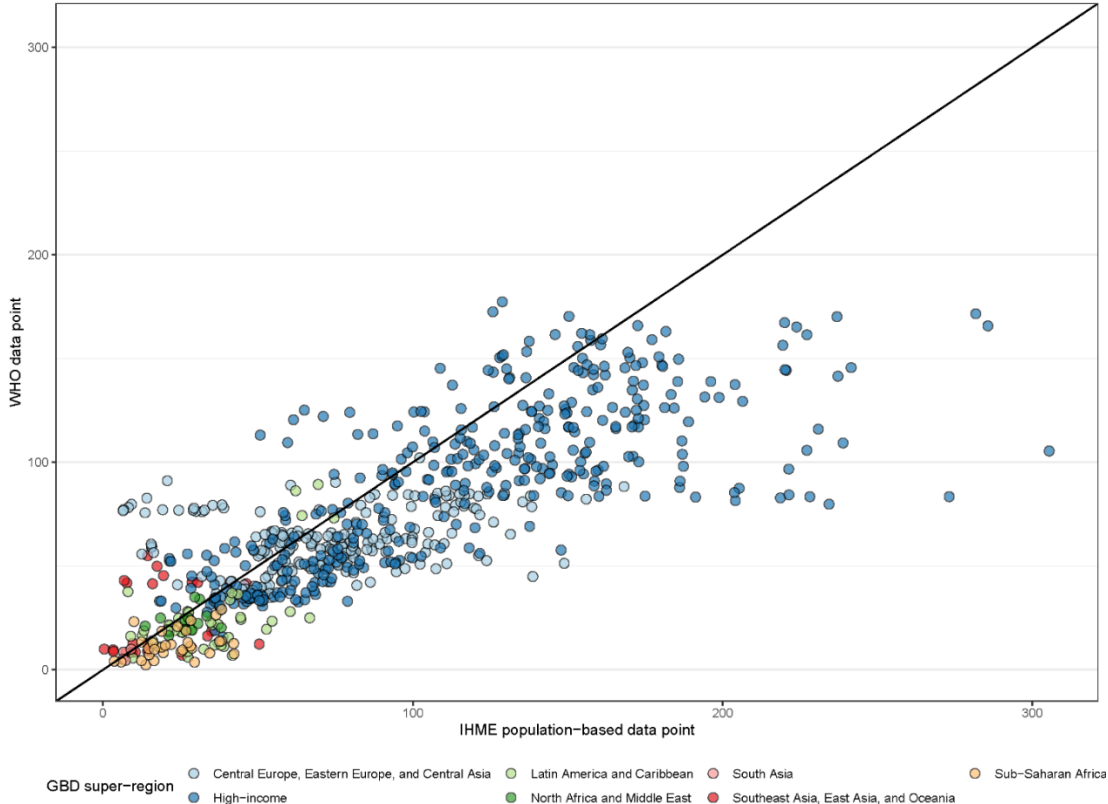


Figure 4c. Matched pairs of IHME population-based points and WHO data points for dentists

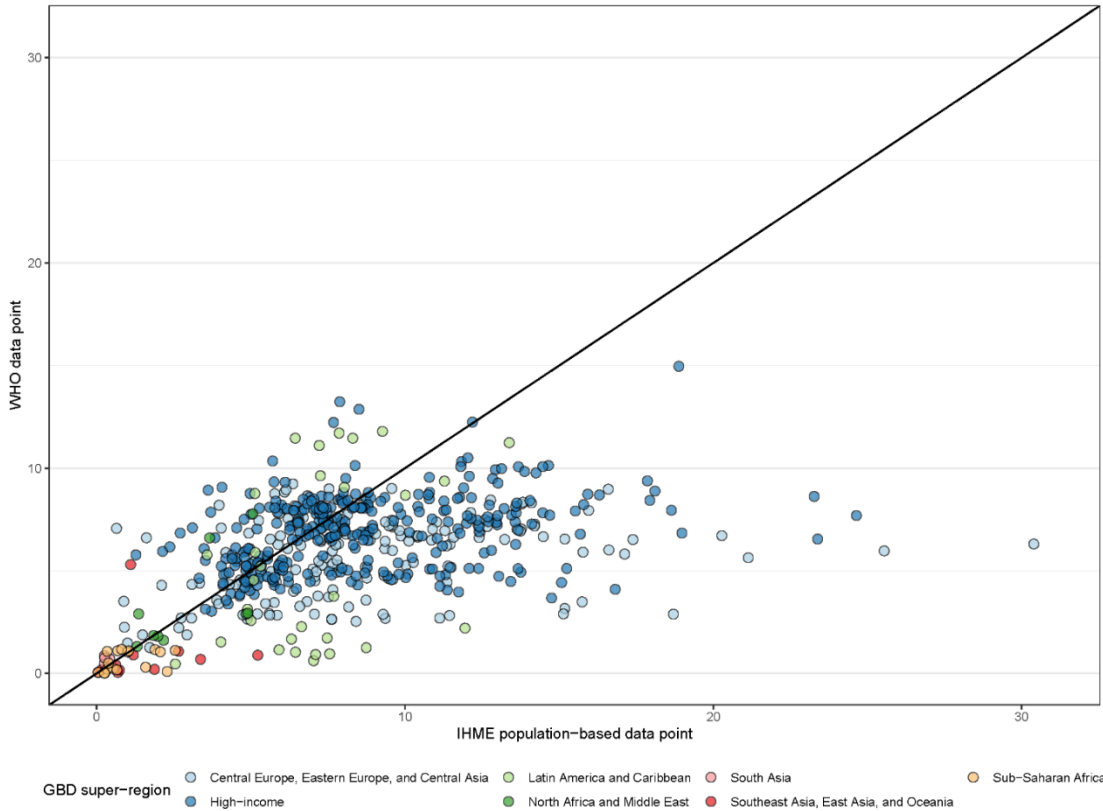
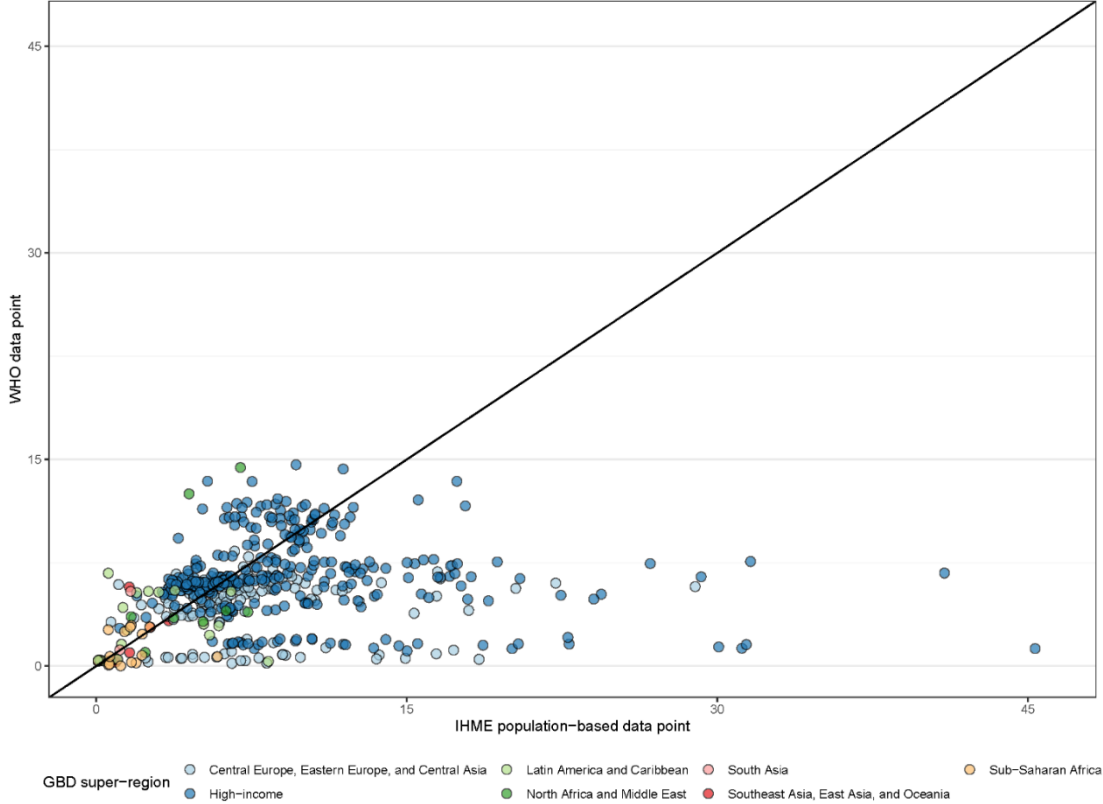


Figure 4d. Matched pairs of IHME population-based points and WHO data points for pharmacists, 1990 - 2019



In figures 5a-5d, we provide maps of the average percent difference between the WHO points and the population-based data by country and territory for each of these cadres.

Figure 5a. Average percent difference between IHME population-based data and WHO data for physicians

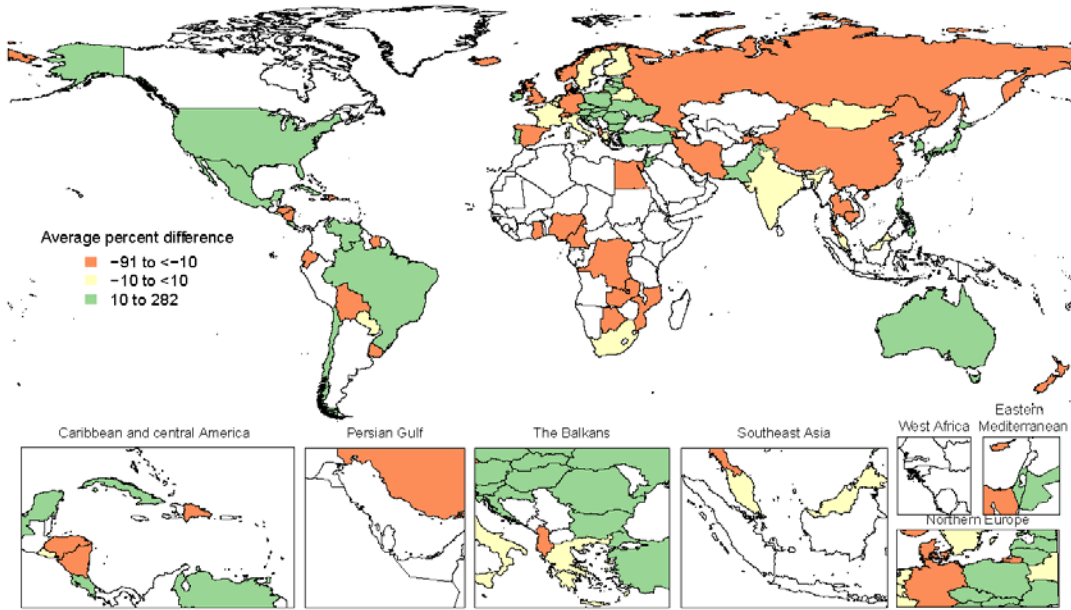


Figure 5b. Average percent difference between IHME population-based data and WHO data for nurses and midwives

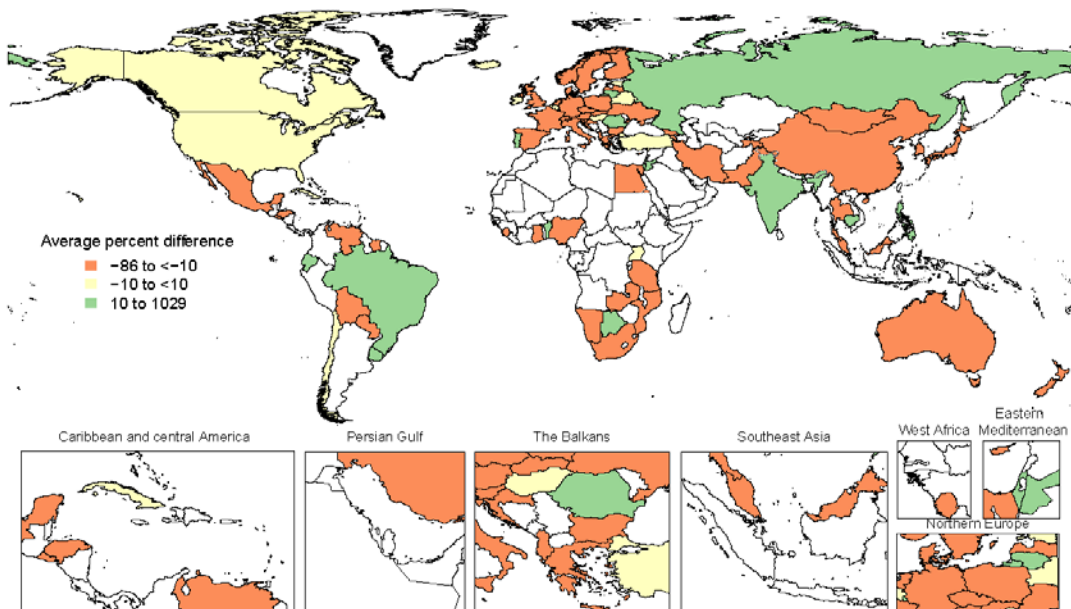


Figure 5c. Average percent difference between IHME population-based data and WHO data for dentists

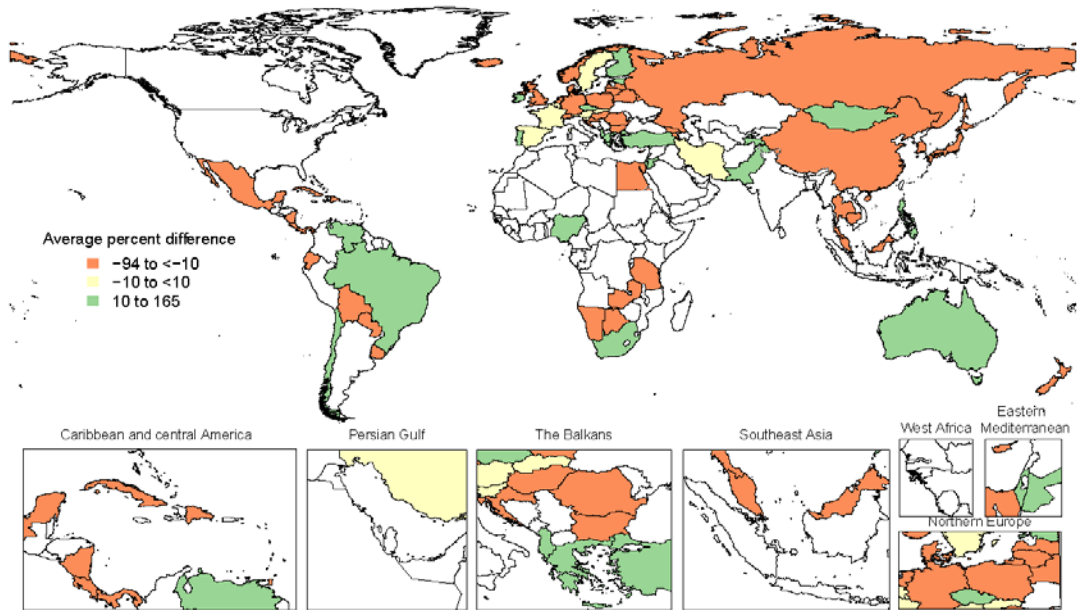


Figure 5d. Average percent difference between IHME population-based data and WHO data for pharmacists

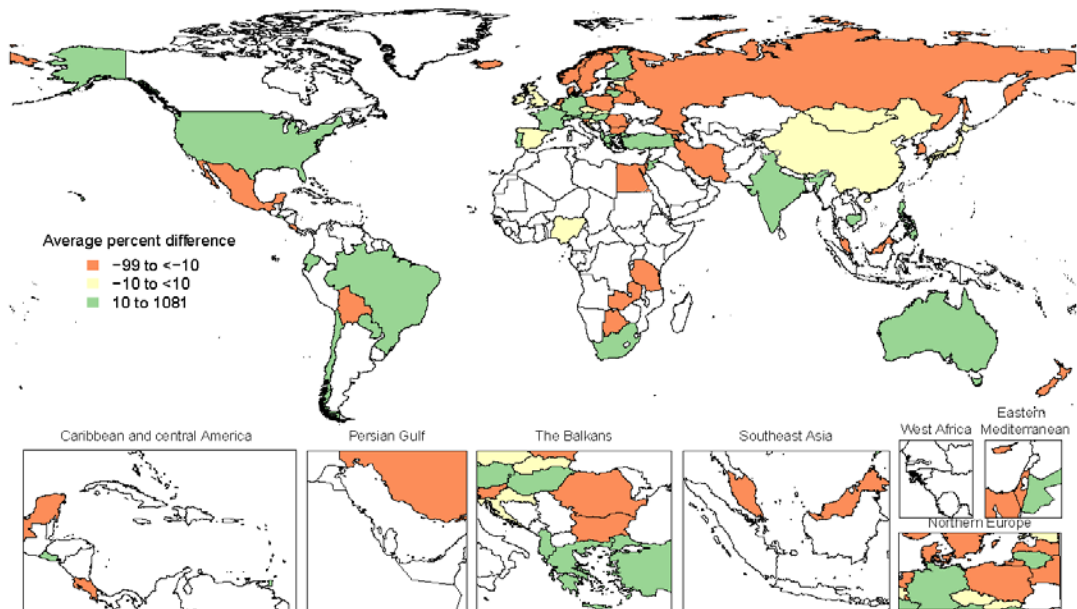


Figure 5a-5d highlight that the WHO data were not consistently biased in one direction or another, although regional patterns are apparent. For example, WHO data tends to be lower than population-based data in sub-Saharan Africa across all cadres, suggesting the WHO data for the region potentially

only captures the public sector. Another regional trend depicted figure 5a is that the WHO physician data appears to be systematically higher than population-based sources in eastern Europe. Based on these types of regional patterns, we performed bias adjustments based on GBD region and GBD super-region, where no matched pairs existed for a given location. We computed location-specific adjustments when matched pairs were available for a given location.

For each geographical unit, we tested whether there was sufficient evidence to implement an adjustment first. There were some locations where the WHO data were comparable to the population-based data we did not want to make an adjustment. We initially tested whether a location needed adjustment using a lasso regression, with cadre per population regressed on indicators for region and super-region, or alternatively, location. Where the indicators were zero in the lasso regression, we did not conduct an adjustment. The penalty (lambda) in the lasso regressions was based on the 1st standard deviation from the mean minimum lambda selected from minimum root mean square error from cross-validation, iterated 1000 times to ensure fold selection did make our results unstable. Table 5 shows the lambdas selected and the RMSE from each lasso model. Figures 6a-6d show the countries in which a location adjustment was used, a super-region or region adjustment was used or no adjustment was applied.

Table 5. Lambdas and RMSE from the lasso covariate selection regressions, by cadre

Cadre	Location adjustment		Super-region/region adjustments	
	Lambda	RSME	Lambda	RSME
Physicians	0.010507	0.494645	0.011254	0.554383
Nurses and midwives	0.004176	0.409446	0.007937	0.484956
Dentists	0.009028	0.572764	0.014393	0.602905
Pharmacists	0.009017	0.589479	0.046735	0.989457

Figure 6a. Adjustment type for physicians

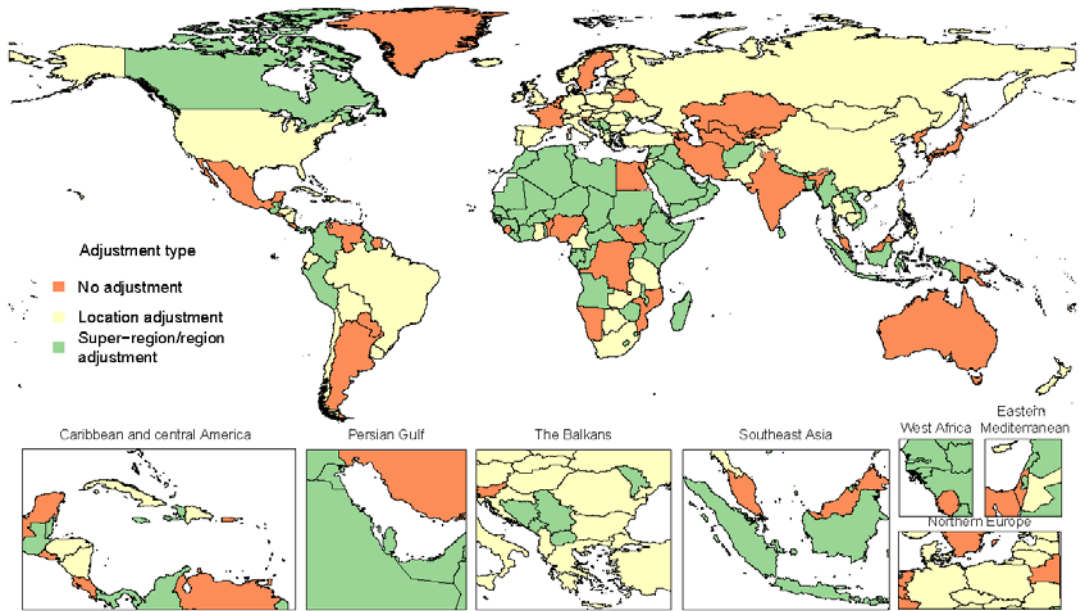


Figure 6b. Adjustment type for nurses and midwives

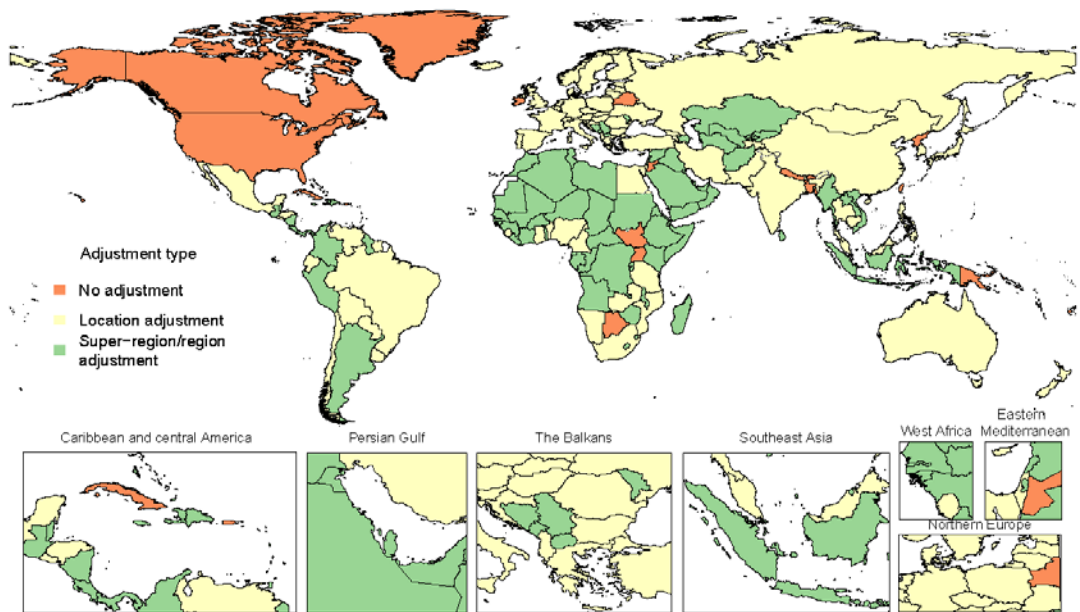


Figure 6c. Adjustment type for dentists

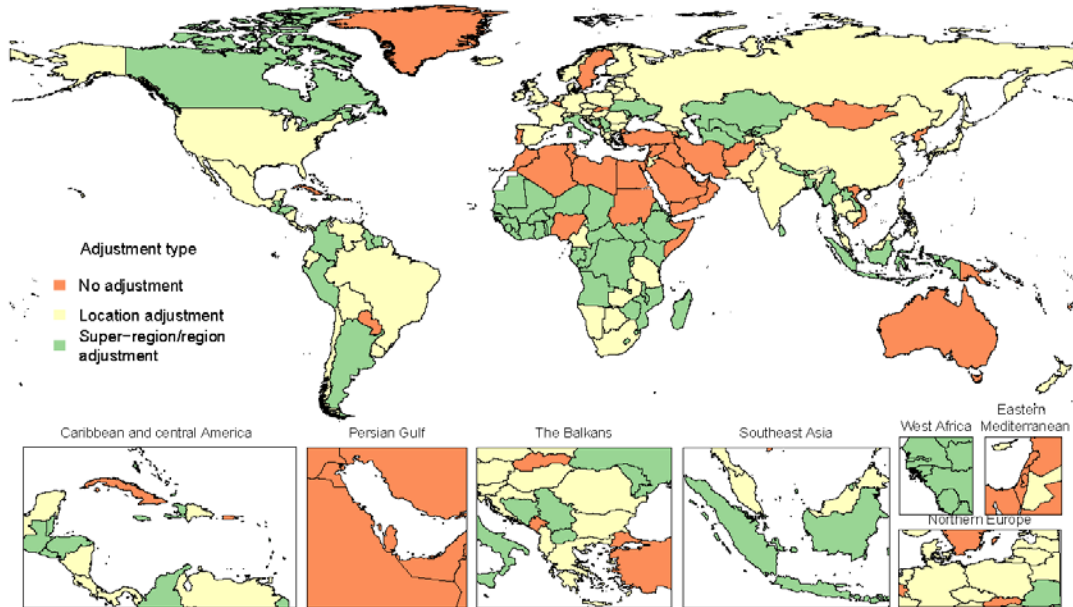
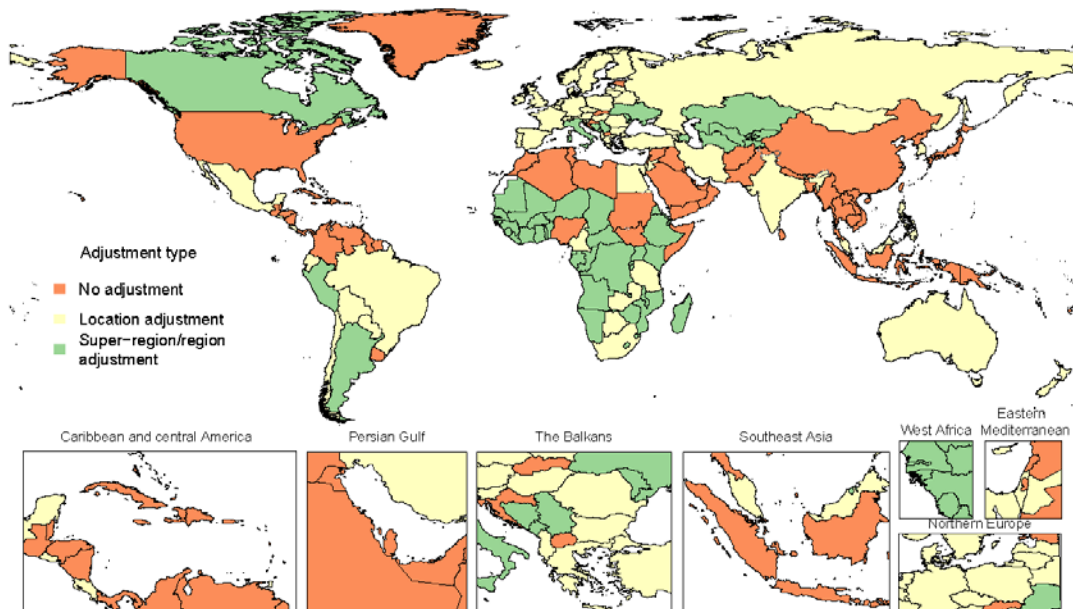


Figure 6d. Adjustment type for pharmacists



To implement the adjustment, we performed two crosswalks using the MR-BRT tool developed for GBD, estimating adjustments based on the ratio of population-based sources to WHO data.^{18, 19} MR-BRT propagates model uncertainty and data uncertainty through to the adjusted estimates (of the WHO data), important for subsequent modeling steps. The WHO data do not have reported variance since they are sourced from reports from countries. We imputed variance of the WHO data based on the

mean regional variance of the labor force surveys and censuses where population-based sources were available; we imputed the mean variance of these sources by super-region in locations without population-based sources. Because in some locations, regions and super-regions, data were sparse and we were concerned about over-fitting the adjustment model, we used a Gaussian prior in the crosswalk. The first crosswalk was for region-super-region adjustments, based on the indicators selected in the lasso regression with geographic indicators for region and super-region. The second crosswalk was for location-specific adjustments, with adjustments estimated only for locations selected in the lasso regression. We opted to perform a single super-region adjustment in sub-Saharan Africa instead of individual regional adjustments due to a limited number of matched pairs in the super-region (only 27 matched pairs for physicians for the super-region and just 1 matched pair for central sub-Saharan Africa in particular). Furthermore, we declined to implement super-region-region adjustments where the direction of the adjustment was inconsistent with population-based sources – if the adjustment would lower the WHO point but a population-based source was higher than the WHO point, we concluded we did not have sufficient evidence to implement the adjustment.

$$\ln\left(\frac{HRH\ cadre_i}{10,000\ population}\right) = \beta * I(location) + \varepsilon_i$$

$$\ln\left(\frac{HRH\ cadre_i}{10,000\ population}\right) = \beta * I(region) + \beta * I(super\ region) + \varepsilon_i$$

Where:

i denotes a given matched pair of population-based sources to WHO data

$I(location)$ denotes an indicator for each location selected in the lasso regression

$I(region)$ denotes an indicator for each region selected in the lasso regression

$I(super\ region)$ denotes an indicator for each super-region selected in the lasso regression

β is the adjustment factor for locations, regions or super-regions selected in the lasso regression, modeled with a Gaussian prior in MR-BRT

ε_i is the residual

Maps of the percentage adjustment relative of the WHO data are found in figures 7a-7d. The MR-BRT crosswalk package computes uncertainty for each point, based on both data and model uncertainty.

Figure 7a. Average percent adjustment to the WHO data input data, physicians

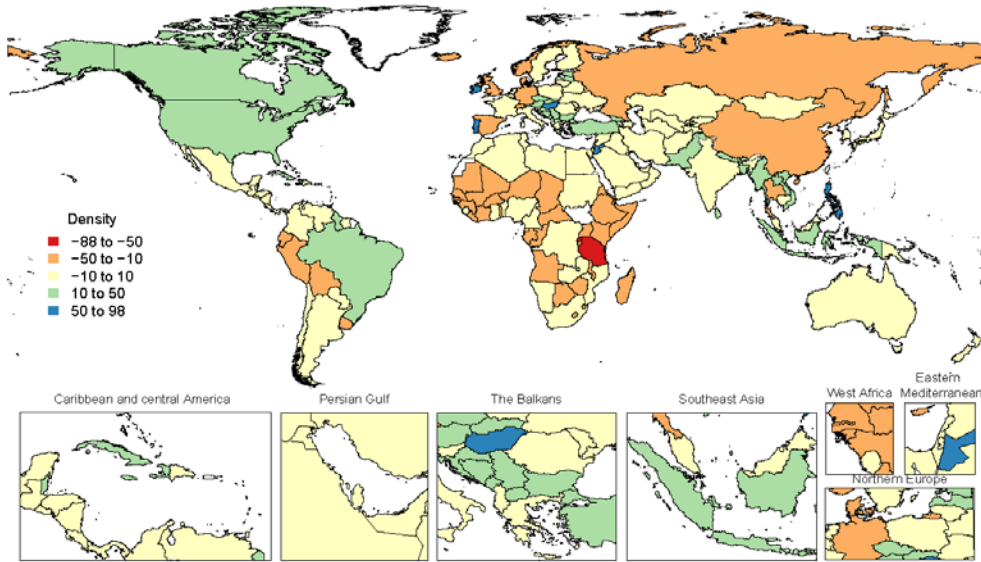


Figure 7b. Average percent adjustment to the WHO data input data, nurses and midwives

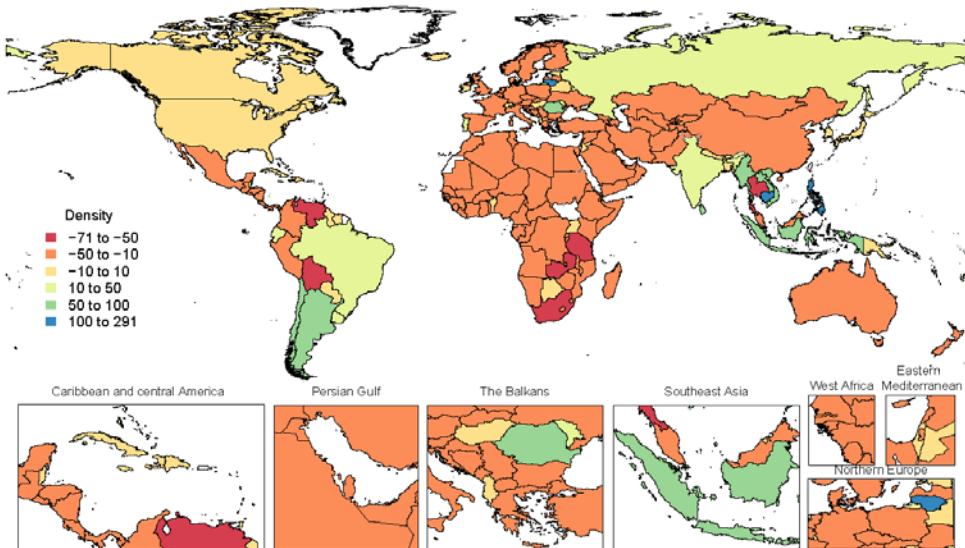


Figure 7c. Average percent adjustment to the WHO data input data, dentists

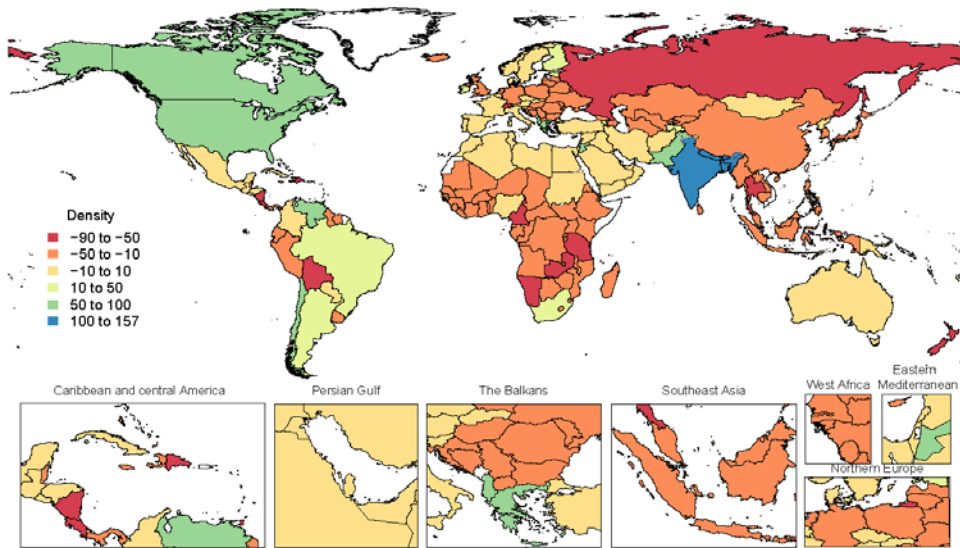
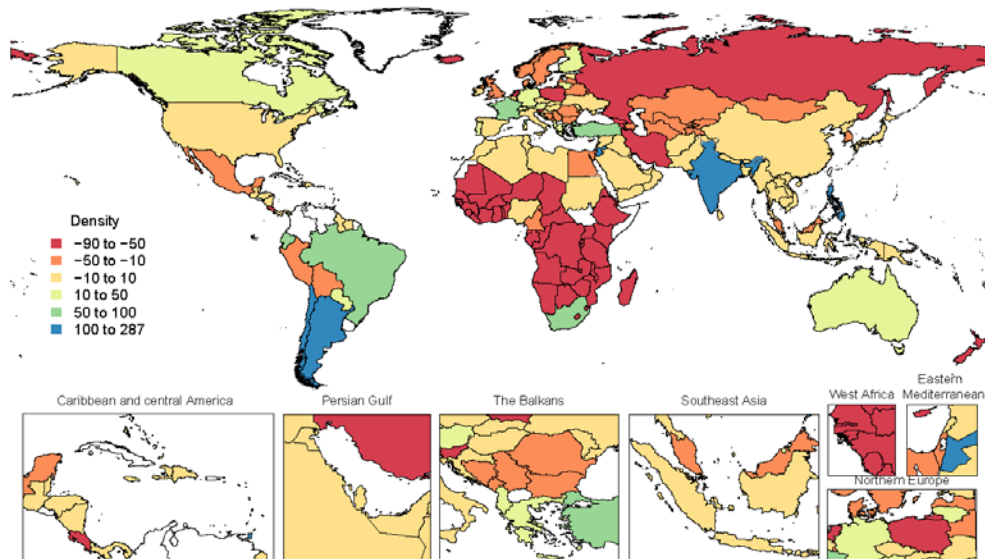


Figure 7d. Average percent adjustment to the WHO data input data, pharmacists



Modelling strategy

To estimate the prevalence of health workers around the world over time, this analysis used a three-stage spatiotemporal Gaussian process regression (ST-GPR). ST-GPR is a flexible modelling strategy that synthesises noisy data by incorporating covariates and borrowing strength across both geography and time to produce comprehensive time series estimates of an indicator with corresponding uncertainty. ST-GPR has been used widely in GBD research and has been described in detail in other recent publications.⁷ As stated in the main text, the first stage of the model fits a linear regression to the data with random effects on specified covariates. The second stage smooths the residuals between the regression fit and the data across time and geography to generate a non-linear trend that

better follows available data in each national location, as well as in the surrounding region and super-region. The third stage uses that trend as a mean function in a Gaussian process regression to account for input data variance and to generate uncertainty in the final estimates. In this study, ST-GPR was used to model 1000 draws of densities for each health worker cadre separately, all health worker cadres combined, and employment ratios, for every location from 1990 to 2019. Health workers were not modelled separately by age or sex. Uncertainty was propagated through all analytical steps, such that final UIs reflect uncertainty from survey sampling as well as from the models themselves.

In addition to handling noisy data and providing estimates of uncertainty, the space-time smoothing component of ST-GPR made it an ideal tool for modelling HRH. Since available input data on this topic are heavily skewed toward high-income geographies like Europe, it was important to select a model that would not extrapolate findings from such data-rich areas to unduly influence estimates for locations lacking in data, which tend to have very different health worker densities and distributions. By incorporating data from surrounding regions and super-regions in the model fit, ST-GPR ensures that estimates for locations lacking in data better reflect patterns observed in inputs from nearby geographies, rather than patterns from the most data-rich locations. ST-GPR therefore makes the default assumption (unless the data indicates otherwise) that locations nearby geographically will follow similar patterns in HRH levels, though this assumption is partially mitigated by the use of additional covariates to inform location-specific differences in anticipated workforce densities. Geographical proximity in ST-GPR is determined from the GBD 2019 location hierarchy, which divides the world into seven super-regions, 21 regions, and 204 countries and territories. GBD regions and super-regions were generated to group countries that are similar in physical geography as well as in epidemiological profiles (for example, patterns in causes of death). Given the wide range of health topics estimated within the GBD framework, these regional distinctions do not always align perfectly with patterns in the disease or indicator being estimated. Nevertheless, these groupings were constructed to capture many potentially unmeasured contextual similarities (related to culture, climate, economics, health burdens, etc.) that produce similar trends across a wide range of health-related indicators. We assume that a country's regional grouping is no less informative to the modelling of health worker densities than it is to the modelling of other prevalent health issues (eg, heart disease, cancer, exposure to lead, etc.). While health workers are certainly needed everywhere, the local disease burden, training capacity, workforce demand (in terms of monetary and workplace incentives for workers), and general health system infrastructure are all incredibly relevant factors to a country's health worker densities that do follow distinct regional patterns.

In order to model health worker cadres – both separately and in the aggregate – in ST-GPR, we first used linear models with fixed effects on combinations of the following GBD 2019 estimated covariates: Socio-demographic Index (SDI), log-transformed total national per capita health expenditure, and estimates of the size of the professional workforce. Estimates of SDI and total national per capita health expenditure were generated by affiliated research groups using methodology described elsewhere.^{8,9} The professional workforce size was calculated as the proportion of the employed population working in ISCO-defined professional occupations. Data for the professional workforce covariate came from the same types of censuses and surveys used in extracting health worker cadre data. However, many more censuses and surveys were available for this covariate due to the fact that professional occupations can be identified from even those sources that only code occupations to the ISCO one-digit level of detail. ST-GPR was used to model the professional workforce across all GBD locations and years. Additional

details on the modelling process for professional occupations are available in the Global Burden of Diseases, Injuries, and Risk Factors Study 2017 comparative risk assessment appendix.¹⁰ To model each health worker cadre, we used the same model settings for intermediate and final estimates.

Intermediate estimates, which were exclusively used in splitting, were run using only four-digit mapped censuses and surveys, while final estimates were run on all available data after three-digit codes had been split into the underlying four-digit cadres. The combinations of covariates chosen for inclusion in each model varied and were based on their empirical predictive ability for a given cadre during model development.

We generated employment ratio estimates in ST-GPR by age and sex, using a linear model with fixed effects on total government expenditure levels, average educational attainment in years, the proportion of the population that is Muslim (a proxy used only in the female model fit to reflect the notably lower levels of employment recorded among women, primarily in North Africa and the Middle East), and five-year age group, and with random effects on GBD location, region, and super-region. We then aggregated results within every location-year to calculate the employed population ages 15-69 as a proportion of the total population, using GBD 2019 estimates of age-specific populations.

Due to the small size of specific health worker cadres relative to the total employed population – and the instability that consequently resulted from data transformations in the modelling process – we removed cadre-specific values of zero and modelled the remaining proportion data in log space as the number of workers per 10 000 employed population. This greatly increased the stability of these models. Since all health workers combined constituted a larger proportion of the total employed population, it was not necessary to adjust these data in any way, and it was modelled directly as a proportion in logit space. To control for unrealistic trends due to stochastic variation in smaller cadre models and to ensure consistency across modelled results, estimates for all health workers in the aggregate were used as a more trustworthy envelope to which cadre-specific estimates were raked at the draw level. Raking here refers to the application of a rescale factor to all cadre-specific estimates to ensure that they summed to the envelope category of all health workers. Finally, we converted raked estimates of health worker cadres from proportions of employed populations ages 15-69 to proportions of total populations, using the output draws from the employment ratio model. 1000 draws of final estimates were summarised using the mean and the 2.5th and 97.5th percentile as the 95% uncertainty interval.