

Veterans' Survey 2022 to Census 2021 linkage report

Methods used to link the Veterans' Survey 2022 to Census 2021 deterministically.

Contact:
Rhiannon Brook, Bradley
Salisbury-Finch, Sarah Cummins
linkage.hub@ons.gov.uk

Release date:
15 December 2023

Next release:
To be announced

Table of contents

1. [Main points](#)
2. [Purpose of the linkage](#)
3. [Methods](#)
4. [Quality information](#)
5. [Summary, recommendations and limitations](#)
6. [Related links](#)
7. [Cite this methodology](#)

1 . Main points

- The Veterans' Survey 2022 was linked deterministically to Census 2021 to enable estimation of the representativeness and coverage of the Veterans' Survey 2022 in England and Wales.
- The linkage rate was 81.08% of the Veterans' Survey 2022.
- Potential reasons for this lower linkage rate include respondents leaving the armed forces after Census Day, low coverage of the prison population on the census, ineligible respondents filling in the survey and unidentified duplicates in the Veterans' Survey 2022.
- The estimated precision for the linkage was 97.55% and estimated recall was 99.96%, indicating that the linkage was of high quality.
- Bias analysis was performed on this linkage, showing that the most underrepresented characteristics in the linked data compared with the full Veterans' Survey 2022 data include those younger than 30 years of age, prisoners and all ethnic groups except White.

2 . Purpose of the linkage

This report summarises the linkage between the Veterans' Survey 2022 and Census 2021. The linkage was commissioned by the Adhoc Surveys Development and Delivery (ASDD) and Veterans Topic Analysis teams at the Office for National Statistics (ONS). The linkage was carried out to enable estimation of the representativeness and coverage of the Veterans' Survey 2022 in England and Wales. This allows analysts to weight data if necessary and inform them where geographically to focus any survey boosts.

3 . Methods

Census 2021

The census, administered by ONS, happens every 10 years and provides a report of all the people and households in England and Wales. The most recent census day for England and Wales was on Sunday 21 March 2021.

The dataset used was a subset of the data and contained personal identifiable information, it was created for the purpose of linkage. The Census data were used for linkage before any editing, imputation, estimation and statistical disclosure control were applied. As part of the census, individuals were able to report if they had previously served in the UK armed forces, however this variable was not available in the census dataset used for linkage.

The variables used for linkage included:

- first name
- surname
- sex
- date of birth
- postcode
- alternative postcode

Veterans' Survey 2022

The Veterans' Survey 2022 was a collaboration between the ONS, The Office for Veterans' Affairs (who funded the survey) and the Devolved Administrations. The survey covered UK armed forces veterans aged 18 years and over that lived in the UK and was conducted via a self-select sample. The aim of this survey was to learn more about the lives of the UK armed forces veterans and their families. Responses from veterans in England and Wales only were included in the linkage dataset.

The variables used for linkage included:

- first name
- surname
- sex
- date of birth
- postcode

Pre-processing

Both the Veterans' Survey 2022 and Census 2021 underwent cleaning, steps included:

- case standardising string variables
- standardisation of nulls, missingness and blanks
- standardising date formatting
- removing non-alphabetical characters, and spaces where appropriate
- removing titles from name variables
- splitting forename and surname components into separate variables, for example if a forename was "Joseph James" the resulting forename one (first component) variable would be "Joseph" and the resulting forename two (second component) variable would be "James"
- splitting postcode components into separate variables (sector, district and area)
- adding a unique biography variable (to indicate whether full name and date of birth is unique within the dataset)
- adding a prison flag variable (using address and postcode on both datasets, and current prisoner status on Veterans' Survey 2022)

Following cleaning, the Veterans' Survey 2022 was deduplicated to remove records where the same person had appeared to submit the survey multiple times (and therefore each record had a different survey ID).

Duplicate pairs were identified deterministically using two strict matchkeys. A matchkey is a set of rules or criteria that must be met to make a link. The first matchkey required concordant full name, sex, date of birth and postcode. The second matchkey required concordant first components of both first name and surname, sex, date of birth and postcode.

Duplicate pairs were then scored based on completeness of a selection of both variables used for linkage and variables used for bias analysis only, with a greater weight assigned to linkage variables than bias analysis variables. In cases where one record of a pair had a higher score than the other this record was kept and the record with a lower score was dropped. In cases where the scores were equal within a record pair then one record was dropped arbitrarily.

Before deduplication, the Veterans' Survey 2022 contained 26,038 records. Deduplication removed 88 records; therefore, 25,950 records were remaining for linkage.

Deterministic linkage

The deduplicated Veterans' Survey 2022 data were deterministically linked with Census 2021 in two stages, using 37 matchkeys. Each matchkey consists of a set of rules or criteria that must be met to make a link. To account for expected errors in the data, the criteria are loosened on different linkage variables. Matchkeys are applied hierarchically, starting at the strictest matching criteria, and gradually become looser.

The initial stage involved linking both datasets using the first 30 matchkeys. Where multiple links were found for one record, the link on the strictest (lowest numbered) matchkey was kept. If multiple links for one record formed on the same matchkey then all of these links were discarded as it was not possible to know which one was correct. This cluster resolution process removed 209 links. Following cluster resolution 20,989 one-to-one links remained.

Further exploration at this stage revealed that only a small proportion of the survey responses from individuals in prison had linked. Therefore, a second stage of the linkage was implemented to focus specifically on the prison records. The residual records from the first stage of the linkage were filtered to only include records thought to be for people in prison (based on postcode in each dataset and the prison address and current prisoner variables in Veterans' Survey 2022). These records were then linked with seven new matchkeys, the majority of which used name, sex and date of birth variables only. This allowed for linkage to occur in cases where an individual may have moved prisons in the time period between Census Day and the Veterans' Survey 2022. For the matchkeys that did not include geography variables, records had to possess unique biography (name and date of birth) within each dataset to be linked. This linkage produced 50 one-to-one links.

The links from each stage were then joined together, giving a total of 21,039 one-to-one links and a match rate of 81.08% of the Veterans' Survey 2022. Table 1 shows the list of matchkeys used, which stage of the linkage they were used in, and the number of links from each one in the final linkage (following resolution of multiple links). Table 2 shows how many records were linked and their duplication status.

Table 1: Matchkey descriptions and number of links per matchkey for the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)

Matchkey	Stage	Description	Number of links
MK1	1	Concordant full name, sex, date of birth and postcode.	3,885
MK2	1	Concordant first components of first name and surname. Concordant sex, date of birth and postcode.	14,992
MK3	1	Concordant full name, date of birth and postcode.	16
MK4	1	First components of resident first name and resident surname (census) concordant with first components of first name and surname (Veterans' Survey 2022). Concordant sex, date of birth and postcode.	75
MK5	1	Concordant full name, sex, year of birth and postcode. Transposed day and month of birth.	1
MK6	1	Concordant full name, sex and date of birth. Census alternative postcode concordant with Veterans' Survey 2022 postcode.	8
MK7	1	Concordant first components of first name and surname. Concordant sex and date of birth. Census alternative postcode concordant with Veterans' Survey 2022 postcode.	40
MK8	1	Jaro-Winkler greater than 0.8 for first components of first name. Concordant first components of surname. Concordant sex and date of birth. Veterans' Survey 2022 postcode concordant with census postcode or alternative postcode.	647
MK9	1	Jaro-Winkler greater than 0.8 for first components of surname. Concordant first components of first name. Concordant sex and date of birth. Veterans' Survey 2022 postcode concordant with census postcode or alternative postcode.	53
MK10	1	Jaro-Winkler greater than 0.8 for first components of first name and surname. Concordant sex and date of birth. Veterans' Survey 2022 postcode concordant with census postcode or alternative postcode.	5
MK11	1	Jaro-Winkler greater than 0.8 for full name. Concordant sex and date of birth. Veterans' Survey 2022 postcode concordant with census postcode or alternative postcode.	142
MK12	1	Soundex full name. Concordant sex and date of birth. Veterans' Survey 2022 postcode concordant with census postcode or alternative postcode.	9
MK13	1	Soundex resident forename (census) and first name (Veterans' Survey 2022). Soundex resident surname (census) and surname (Veterans' Survey 2022). Concordant sex and date of birth. Veterans' Survey 2022 postcode concordant with census postcode or alternative postcode.	2
MK14	1	Concordant full name, sex and date of birth. Veterans' Survey 2022 postcode sector concordant with census postcode sector or alternative postcode sector.	67
MK15	1	Concordant first components of first name and surname. Concordant sex and date of birth. Veterans' Survey 2022 postcode sector concordant with census postcode sector or alternative postcode sector.	221
MK16	1	Jaro-Winkler greater than 0.8 for first components of first name. Concordant first components of surname. Concordant sex and date of birth. Veterans' Survey 2022 postcode sector concordant with census postcode sector or alternative postcode sector.	10
MK17	1	Jaro-Winkler greater than 0.6 for full name. Concordant sex and year of birth. Transposed day and month of birth. Veterans' Survey 2022 postcode concordant with census postcode or alternative postcode.	13
MK18	1	Jaro-Winkler greater than 0.8 for transposed first components of first name and surname. Concordant sex and date of birth. Veterans' Survey 2022 postcode concordant with census postcode or alternative postcode.	2
MK19	1	Concordant full name, sex, year of birth and day of birth. Veterans' Survey 2022 postcode concordant with census postcode or alternative postcode.	125

MK20	1	Concordant full name, sex, year of birth and day of birth. Veterans' Survey 2022 postcode sector concordant with census postcode sector or alternative postcode sector.	3
MK21	1	First components of first name (census) are contained within first components of first name (Veterans' Survey 2022) or vice versa. First components of surname (census) are contained within first components of surname (Veterans' Survey 2022) or vice versa. Concordant sex, year of birth and day of birth. Veterans' Survey 2022 postcode concordant with census postcode or alternative postcode.	457
MK22	1	First components of first name (census) are contained within first components of first name (Veterans' Survey 2022) or vice versa. First components of surname (census) are contained within first components of surname (Veterans' Survey 2022) or vice versa. Concordant sex, year of birth and day of birth. Veterans' Survey 2022 postcode sector concordant with census postcode sector or alternative postcode sector.	4
MK23	1	Jaro-Winkler greater than 0.8 for full name. Concordant sex and date of birth. Veterans' Survey 2022 postcode concordant with census postcode or alternative postcode.	17
MK24	1	Standardised Levenshtein distance greater than 0.7 for full name. Concordant sex and date of birth. Veterans' Survey 2022 postcode concordant with census postcode or alternative postcode.	12
MK25	1	Full name with no spaces (census) contained within full name with no spaces (Veterans' Survey 2022) or vice versa. Concordant sex and date of birth. Veterans' Survey 2022 postcode concordant with census postcode or alternative postcode.	42
MK26	1	Jaro-Winkler greater than 0.7 for full name with no spaces. Concordant sex, year of birth and day of birth. Veterans' Survey 2022 postcode concordant with census postcode or alternative postcode.	58
MK27	1	Standardised Levenshtein distance greater than 0.8 for full name with no spaces. Concordant year of birth and day of birth. Veterans' Survey 2022 postcode concordant with census postcode or alternative postcode.	10
MK28	1	Standardised Levenshtein distance greater than 0.8 for full name with no spaces. Concordant sex, year of birth and day of birth. Veterans' Survey 2022 postcode sector concordant with census postcode sector or alternative postcode sector.	1
MK29	1	First components of first name (census) are contained within first components of first name (Veterans' Survey 2022) or vice versa. First components of surname (census) are contained within first components of surname (Veterans' Survey 2022) or vice versa. Concordant year of birth and day of birth. Veterans' Survey 2022 postcode sector concordant with census postcode sector or alternative postcode sector.	67
MK30	1	Jaro-Winkler greater than 0.8 for full name with no spaces. Concordant date of birth. Veterans' Survey 2022 postcode concordant with census postcode or alternative postcode.	5
MK31	2	Within the datasets filtered to only include prison records. Concordant first components of first name and surname. Concordant year of birth. Concordant day and/or month of birth. Prison postcode on Veterans' Survey 2022 concordant with census postcode or alternative postcode, or census address concordant with prison address on Veterans' Survey 2022.	1
MK32	2	Within the datasets filtered to only include prison records. Concordant full name, sex and date of birth. Unique biography for both Veterans' Survey 2022 and Census 2021.	8
MK33	2	Within the datasets filtered to only include prison records. Concordant first components of first name and surname. Concordant sex and date of birth. Unique biography for both Veterans' Survey 2022 and Census 2021.	33
MK34	2	Within the datasets filtered to only include prison records. Concordant first components of first name and surname. Concordant date of birth. Unique biography for both Veterans' Survey 2022 and Census 2021.	2

MK35	2	Within the datasets filtered to only include prison records. Jaro-Winkler greater than 0.8 for first components of first name. Concordant first components of surname. Concordant sex and date of birth. Unique biography for both Veterans' Survey 2022 and Census 2021.	3
MK36	2	Within the datasets filtered to only include prison records. Jaro-Winkler greater than 0.8 for first components of surname. Concordant first components of first name. Concordant sex and date of birth. Unique biography for both Veterans' Survey 2022 and Census 2021.	2
MK37	2	Within the datasets filtered to only include prison records. Standardised Levenshtein distance greater than 0.7 for first components of surname. Concordant first components of first name. Concordant sex and date of birth. Unique biography for both Veterans' Survey 2022 and Census 2021.	1

Source: The Veterans' Survey 2022 to Census 2021 linked data from the Office for National Statistics

Table 2: Linkage and duplication status counts for the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)

	No duplicates identified	Duplicates identified
Linked	20,957	82
Not linked	4,905	6
Dropped	0	88

Source: The Veterans' Survey 2022 to Census 2021 linked data from the Office for National Statistics

4 . Quality information

Clerical review

The standard approach to estimate error in the linked data is to perform clerical review (manual checking) on a sample of links and rejected record pairs, to estimate the number of true positives (correct links), false positives (incorrect links) and false negatives (missed matches). In linkage, there is a trade-off between two types of error - precision and recall.

Precision is a measure of the accuracy of the matches that have been made:

$$\textit{precision} = \frac{\textit{number of true positives}}{\textit{number of true positives} + \textit{number of false positives}}$$

Recall is a measure of the proportion of matches that have been made from all the possible matches:

$$\textit{recall} = \frac{\textit{number of true positives}}{\textit{number of true positives} + \textit{number of false negatives}}$$

Clerical review was carried out in two stages. Firstly, false positive analysis was carried out on a sample of the deterministically linked records to estimate precision. Secondly, false negative analysis was carried out on a sample taken from unlinked (rejected) record pairs to estimate recall. To do this, a basic probabilistic (score-based) linkage of the residuals was implemented using [Splink](#) for the purposes of creating samples for clerical review only. Record pairs for both the false positive analysis and false negative analysis were run through the Data Linkage Hub's Clerical Review Online Widget (CROW) tool for review on a pair-wise basis.

For most groups, sample sizes for clerical review were derived with [Statulator](#) using a confidence level of 95.00%, an expected proportion (of false positives) of 0.05, and a relative precision of 0.4. For sample groups consisting of a small number of record pairs, all record pairs were clerically reviewed.

Precision and recall for the entire population were derived using total estimated errors. This is the sum of multiplying the error rate with the number of record pairs for each bucket and then aggregating up to the entire population.

Confidence intervals (CIs) estimated for the population were derived using the Agresti-Coull method with a confidence level of 95.00% for each bucket, and then aggregated up to the population level. Upon aggregation of the overall CI of precision or recall, a value of zero was applied for both the lower and upper bounds for each bucket where no false positives or false negatives were found and where the sample size was small. For large sample sizes with no errors, the lower bound was adjusted to zero and the upper bound was calculated by dividing 3 by the sample size. For buckets where all record pairs were reviewed the precise error rates were used.

For the false positive analysis of deterministic links, 2,115 record pairs were sampled for clerical review, stratified by matchkey and 1,792 positives were identified (shown in Table 3). Precision was estimated to be 97.55% CI [97.34%, 97.70%] (shown in Table 4).

Table 3: Results of false positive analysis of the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)

Sample group	Matchkeys in group	Count of record pairs in group	Record pairs sampled	True positives found	False positives found	True positive estimate	False positive estimate
1	1	3,885	409	409	0	3,885.00	0.00
2	2	14,992	443	443	0	14,992.00	0.00
3	3 to 7	140	140	140	0	140.00	0.00
4	8 to 11	847	297	296	1	844.15	2.85
5	12 to 15	299	181	180	1	297.35	1.65
6	16 to 20	153	153	45	108	45.00	108.00
7	21 to 22	461	230	41	189	82.18	378.82
8	23 to 20	212	212	188	24	188.00	24.00
9	31 to 37	50	50	50	0	50.00	0.00

Source: The Veterans' Survey 2022 to Census 2021 linked data from the Office for National Statistics

Table 4: Precision results from the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)

Total true positive estimate 20,523.67

Total false positive estimate 515.33

Precision estimate 97.55%

Source: The Veterans' Survey 2022 to Census 2021 linked data from the Office for National Statistics

For the false negative analysis buckets were created based on match weight, giving a total of 5 buckets. A sample of 1,346 pairs were clerically reviewed to detect false negatives and 6 false negatives were identified in the sample (shown in Table 5). Recall was estimated to be 99.96% CI [99.74%, 99.98%] (shown in Table 6).

Table 5: Results of false negative analysis of the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)

Sample group	Match weight	Count of record pairs in group	Record pairs sampled	False negatives found	True negatives found	False negative estimate	True negative estimate
1	Greater than 19.9	62	62	5	57	5.00	57.00
2	15 to 19.9	1,138	326	1	325	3.49	1,134.51
3	10 to 14.9	616	263	0	263	0.00	616.00
4	5 to 9.9	1,076	321	0	321	0.00	1,076.00
5	Less than 5	2,069	374	0	374	0.00	2,069.00

Source: The Veterans' Survey 2022 to Census 2021 linked data from the Office for National Statistics

Table 6: Recall results from the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)

Total false negative estimate 8.49

Total true positive estimate 20,523.67

Recall estimate 99.96%

Source: The Veterans' Survey 2022 to Census 2021 linked data from the Office for National Statistics

Note that clerical review is used as a proxy for the true match status of a record pair. Therefore, the estimates for precision and recall do not factor in the uncertainty resulting from subjective human decision making. A review of the false positives from group 6 in Table 3 identified that clerical reviewers had been quite conservative in their judgement for a pair to be a match in these cases. Therefore, it is expected that the estimated precision is a slight under-estimate of the true precision.

Bias analysis of linked versus unlinked Veterans' Survey 2022 records

Bias analysis is important for telling us about the representativeness of our linked data. It provides a measure of whether linkage failure is random or related to characteristics of those in the data. If there is bias in a linkage process, which causes a certain demographic to be more or less likely to appear in the linked data, then any conclusions drawn from the linked data could be incorrect or misleading.

Linkage failure may be caused by linkage methods more effectively linking western than non-western names and because of the presence of nicknames. Nicknames are commonly used in the military. If a nickname has been used when completing either Census 2021 or the Veterans' Survey, the likelihood of a successful linkage is decreased.

However, it is also important to consider whether failure to link is because of differing coverage of the two data sources (meaning that the link does not exist). Of people on the Veterans' Survey 2022, 18.92% were not found on the census, but the estimated recall of 99.96% suggests that nearly all possible links were found. Therefore, it is likely that most individuals from the Veterans' Survey 2022 who did not link to Census 2021, could not be found on the census. This needs to be accounted for when examining the bias analysis.

This coverage error could be caused by many minor factors, which result in the linkage rate being lower than might be expected. Veterans who left the armed forces after Census Day may not appear on the census if they were serving overseas. No verification of veteran status and address was required so it is possible that some respondents were veterans living outside of the UK or other ineligible people using inaccurate details. Also, it is possible that there are further duplicate records in the Veterans' Survey 2022 as the deduplication methods used were very precise.

For this bias analysis, demographic characteristics for the linked data were compared with the underlying Veterans' Survey 2022 data. The variables from the Veterans' Survey 2022 are self-reported and so accuracy cannot be guaranteed.

Bias was analysed using proportional discrepancy. Positive proportional discrepancies convey overrepresentation of the characteristic evaluated, while negative proportional discrepancy scores convey underrepresentation. For example, a proportional discrepancy score of negative 0.05 suggests the linked data have only matched 95% of the expected number of matches given the overall match rate. When a group is described as being underrepresented, this means that they are underrepresented in the linked data compared with the unlinked Veterans' Survey 2022 data, given the match rate. This does not mean that this group is underrepresented in the survey compared with the veteran population. The ONS publication [The Veterans' Survey 2022, Demographic Overview and Coverage Analysis, UK](#), compared census records of veterans who appeared in the linked data with veterans in the census who did not link. It also compared all Veterans' Survey 2022 respondents from England and Wales to Census 2021. These comparisons enabled use of the census as a population base to indicate the representativeness of the Veterans' Survey 2022 compared with the veteran population. This differs from the bias analysis of the linked data, which informs how representative the linked data are compared with the Veterans' Survey 2022 data.

For most of the variables where proportional discrepancy is examined there are some records with null values, which represents missing data. These null groups often show as being underrepresented in the data. It is possible that records that have null values for some variables are less likely to have linked because they are of poorer quality than the more complete records. In the following analysis, null groups are shown on the graphs.

Biases in age, years served and year of leaving service

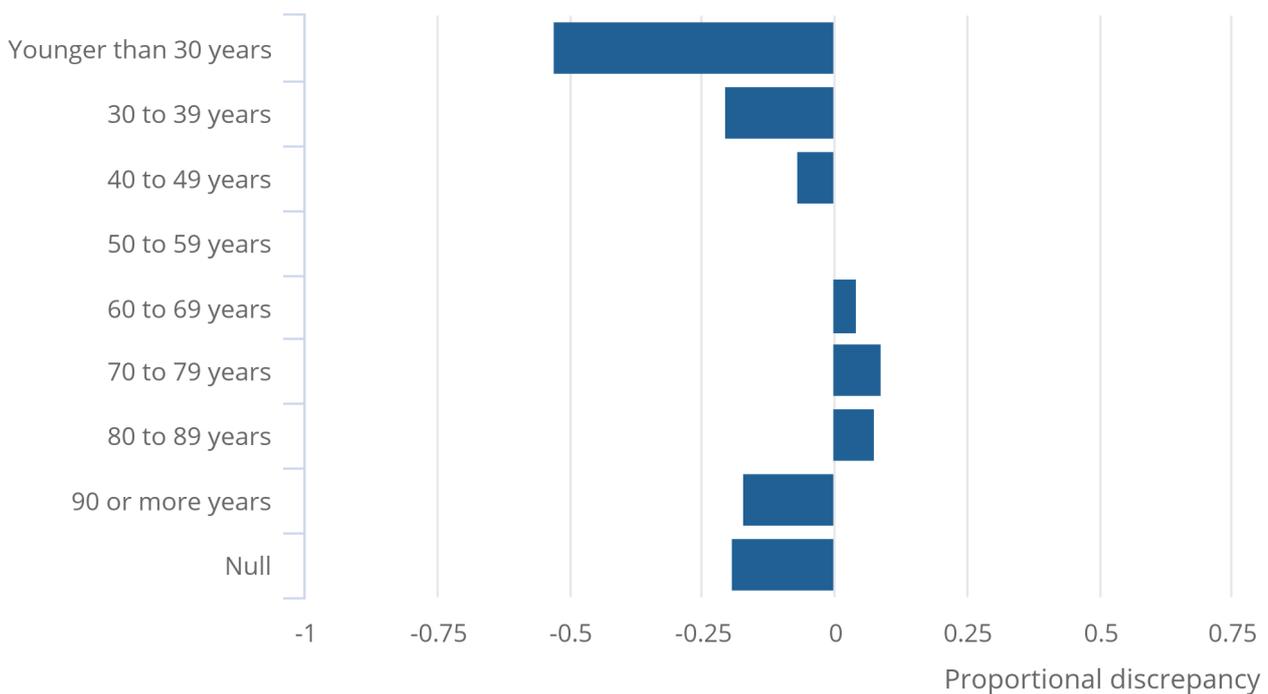
Age on Census Day was calculated from date of birth recorded on the Veterans' Survey 2022. Figure 1 shows the examination of proportional discrepancy for age group, showing that the age groups of younger than 50 years and 90 years or older were underrepresented. Typically, an underrepresentation of younger and working-aged people in linked data could be because they are more likely to move addresses or change name, making them harder to link.

When the residuals from the underrepresented age groups were examined, a higher proportion than would be expected compared with the rest of the data had missing postcode information or were current prisoners. Prisoners were also found to be underrepresented in the linked data, partly because of low coverage in the census. The most underrepresented group (younger than 30 years of age) is small compared with the others, so the impact of this bias should be minimal.

A similar result is seen for year of leaving service, where those who left service after 2020 or before 1950 were underrepresented. Little bias was found in number of years' service with those serving 12 years or less slightly underrepresented.

Figure 1: Proportional discrepancy for age group (years) from the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)

Figure 1: Proportional discrepancy for age group (years) from the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)



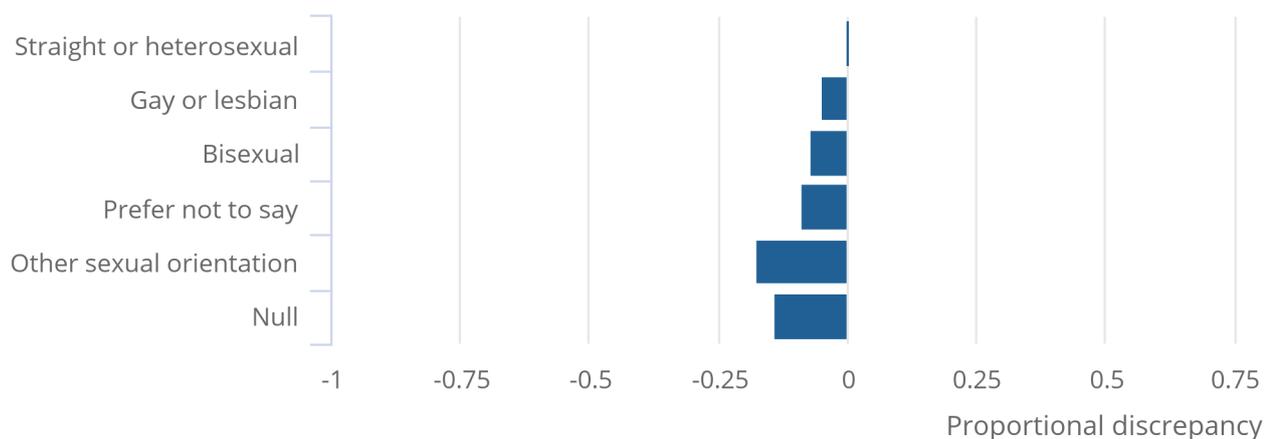
Source: The Veterans' Survey 2022 to Census 2021 linked data from the Office for National Statistics

Biases in sex and sexual orientation

No notable bias was found for sex, although females were slightly underrepresented. Figure 2 shows the examination of proportional discrepancy for sexual orientation, showing that all groups except straight or heterosexual were underrepresented.

Figure 2: Proportional discrepancy for sexual orientation from the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)

Figure 2: Proportional discrepancy for sexual orientation from the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)



Source: The Veterans' Survey 2022 to Census 2021 linked data from the Office for National Statistics

Biases in geography

Overall, no bias was found between countries (England and Wales). Figure 3 shows the examination of proportional discrepancy for English counties. The only underrepresented county was Inner London. Figure 4 shows the examination of proportional discrepancy for Welsh principal areas, which shows a mix of principal areas that are under and overrepresented. The most underrepresented principal areas were Swansea and Torfaen.

Figure 3: Proportional discrepancy for English counties from the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)

Figure 3: Proportional discrepancy for English counties from the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)

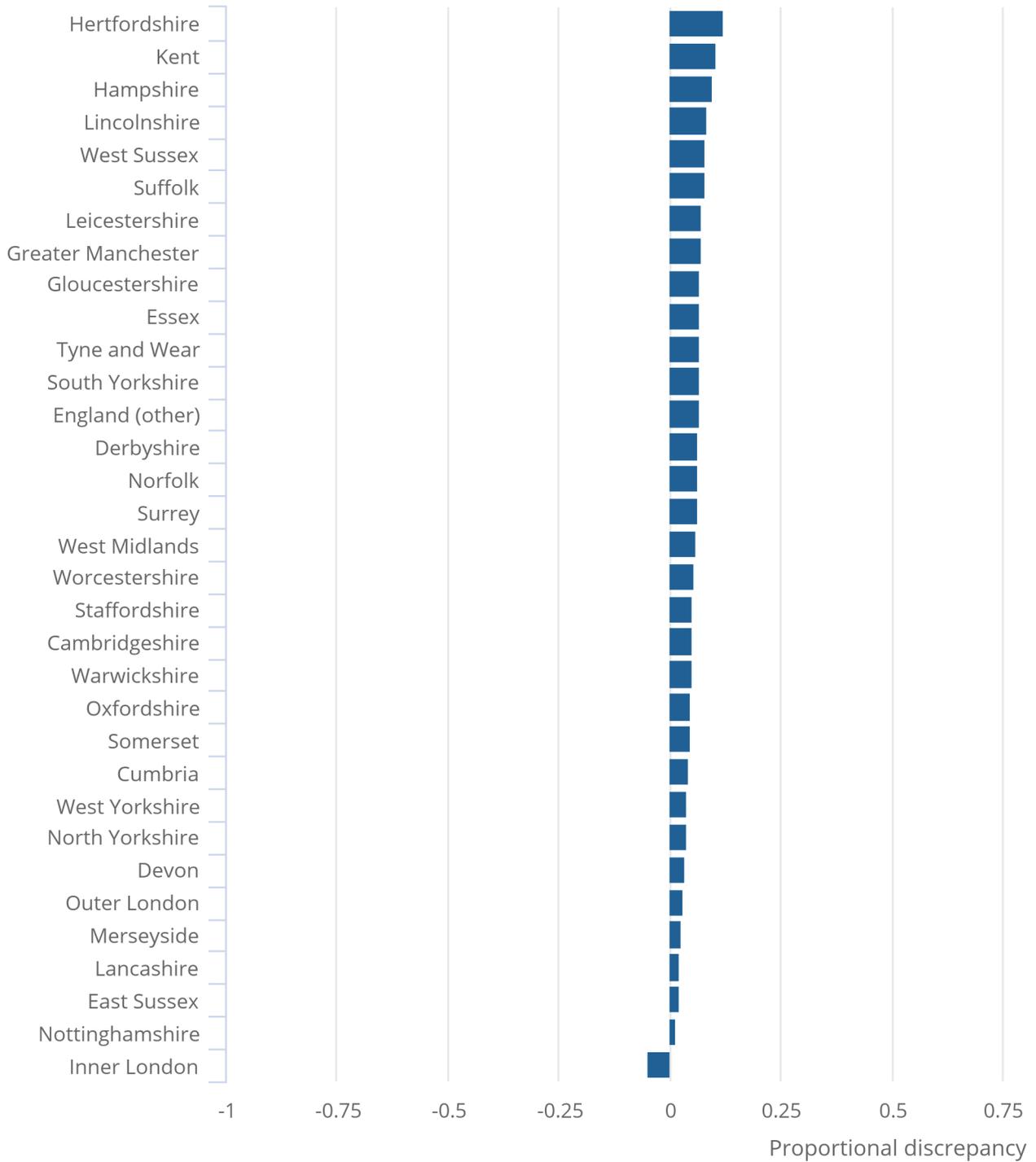
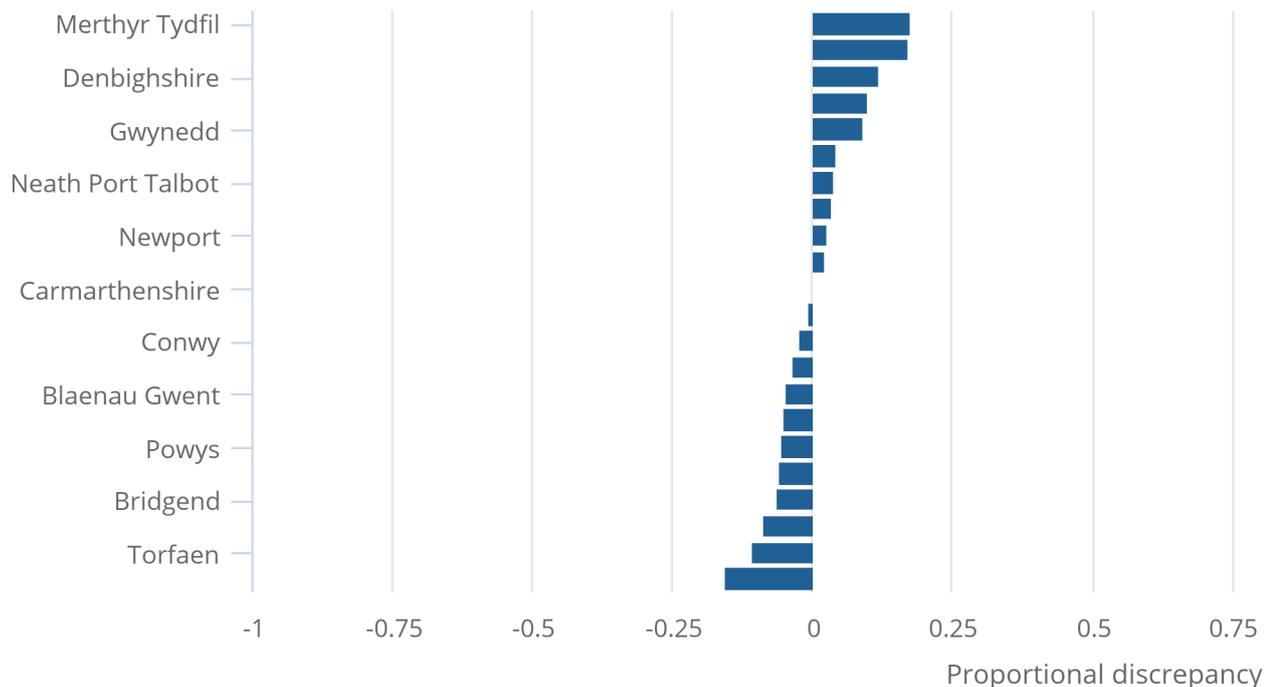


Figure 4: Proportional discrepancy for Welsh principal area from the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)

Figure 4: Proportional discrepancy for Welsh principal area from the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)



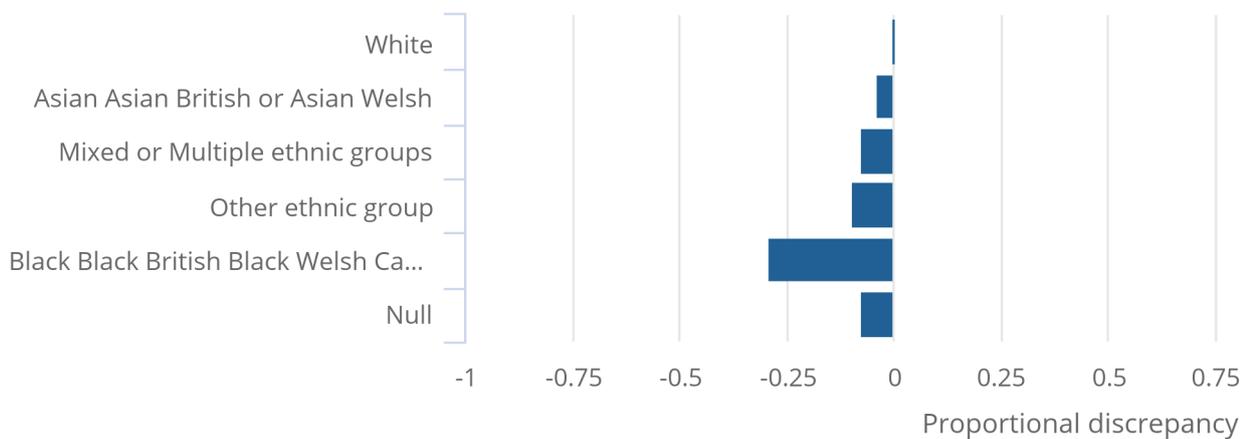
Biases in ethnic group and UK citizenship during service

Figure 5 shows the examination of proportional discrepancy for ethnic group, showing that all groups except White are underrepresented. The Black, Black British, Black Welsh, Caribbean or African group was the most underrepresented group.

Figure 6 shows the examination of proportional discrepancy for UK citizenship status during service, showing that all groups were underrepresented except those who were a UK citizen throughout their service. The linkage process is likely to be better at matching Western names than non-Western names. This may explain some of the biases seen here.

Figure 5: Proportional discrepancy for ethnic group from the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)

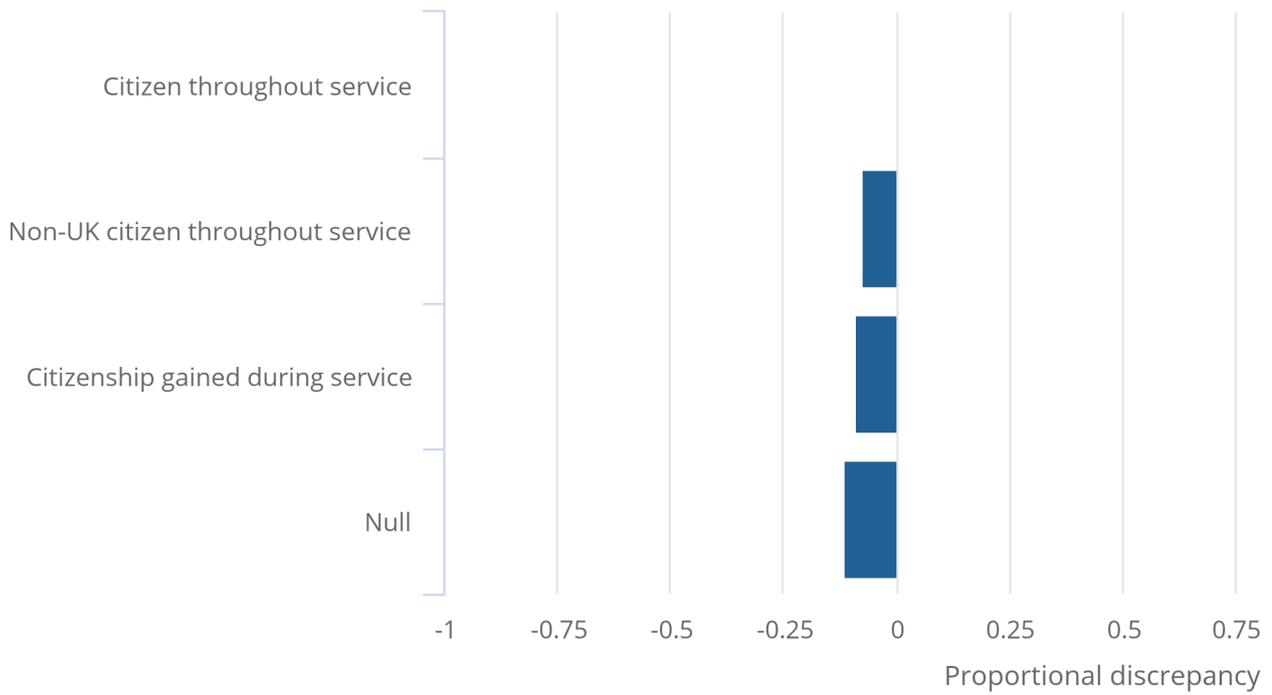
Figure 5: Proportional discrepancy for ethnic group from the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)



Source: The Veterans' Survey 2022 to Census 2021 linked data from the Office for National Statistics

Figure 6: Proportional discrepancy for UK citizenship status during service from the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)

Figure 6: Proportional discrepancy for UK citizenship status during service from the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)



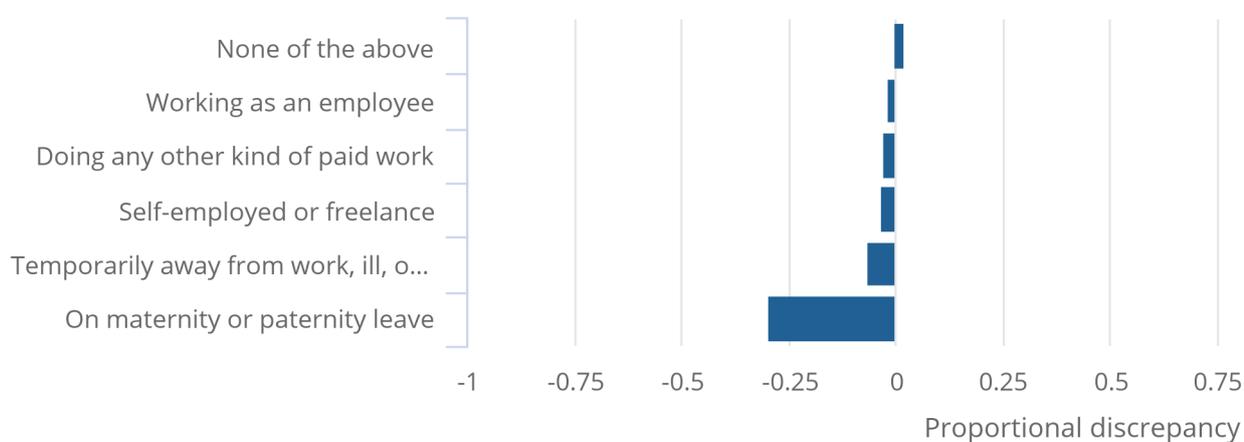
Source: The Veterans' Survey 2022 to Census 2021 linked data from the Office for National Statistics

Biases in employment status

For the employment status question, respondents were able to select multiple options, or “none of the above” options. Figure 7 shows the examination of proportional discrepancy for employment status, showing that all groups are underrepresented except for those who selected the option for “none of the above”. As the age groups 60 to 69 years, 70 to 79 years and 80 to 89 years are overrepresented in the linked data, it is not surprising that veterans who are in this group are overrepresented, as they are likely to be retired.

Figure 7: proportional discrepancy for employment status from the Veterans’ Survey 2022 to Census 2021 linkage (England and Wales)

Figure 7: proportional discrepancy for employment status from the Veterans’ Survey 2022 to Census 2021 linkage (England and Wales)



Source: The Veterans’ Survey 2022 to Census 2021 linked data from the Office for National Statistics

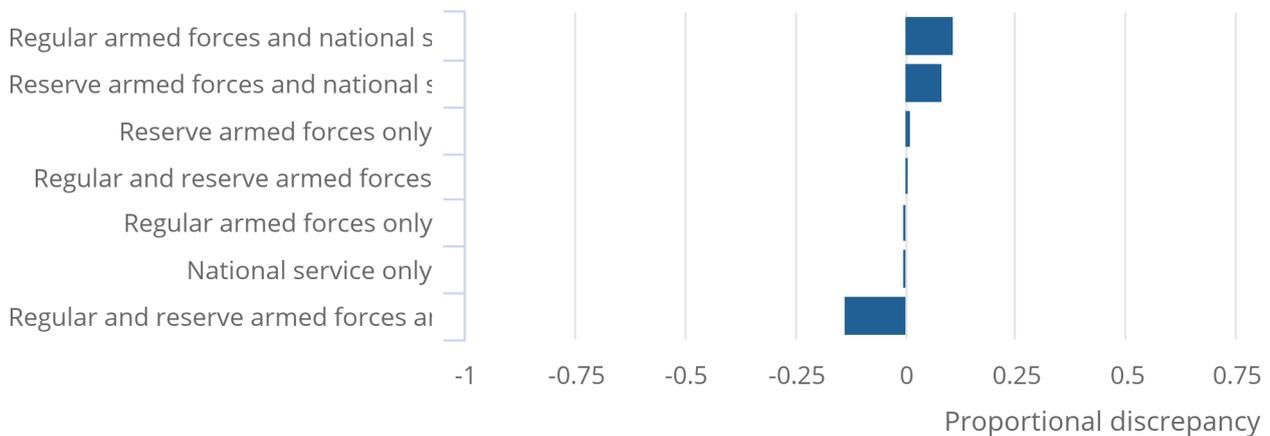
Biases in service, rank and veteran type

No bias was found for service type (Army, Marine, Navy or RAF) or rank (Officer or Other rank), although those with a missing (null) rank were highly underrepresented with a proportional discrepancy of negative 0.47.

In the survey, veterans were asked to state if they were veterans of national service, the reserve armed forces and /or the regular armed forces. All veterans stated that they were members of one, two or all three of these groups. Figure 8 shows that the only group notably underrepresented in the linked data is those who have served in all three of the national service, the reserve and regular armed forces.

Figure 8: Proportional discrepancy for veteran type from the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)

Figure 8: Proportional discrepancy for veteran type from the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)



Source: The Veterans' Survey 2022 to Census 2021 linked data from the Office for National Statistics

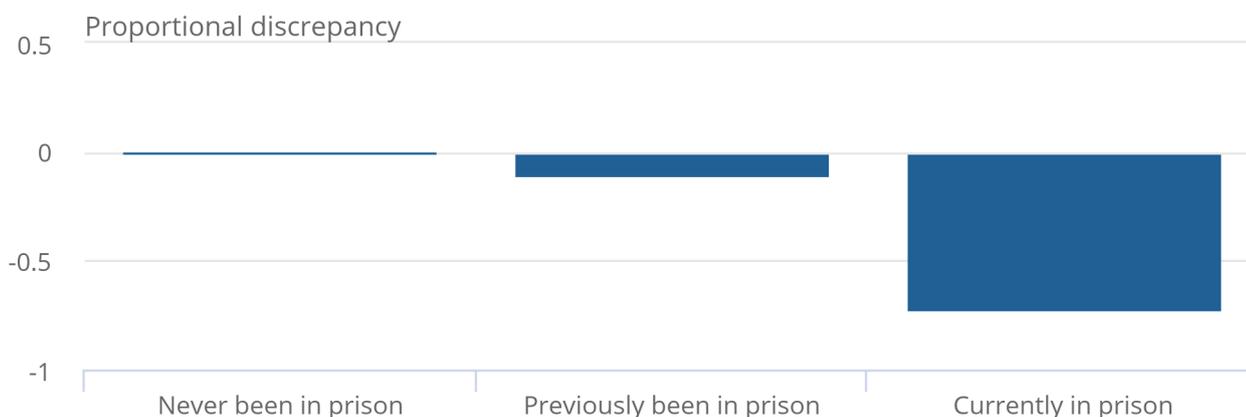
Biases in prisoner status

As part of the survey, respondents were asked if they had ever served a prison sentence and then if so, if they were currently serving a prison sentence. Figure 9 shows the examination of proportional discrepancy for this variable, showing that respondents who reported that they had served a prison sentence are underrepresented in the linked data and that those currently serving a prison sentence are very underrepresented. Those who never served a prison sentence are well represented.

There are two possible explanations for the low linkage rate of serving prisoners: Individuals moving between different prison addresses causing differences in address between the Veterans' Survey 2022 and Census 2021; and low response rates from prisons in the census leading to low coverage of prisoners on the census. You can read more about census response rates from prisons in our [Communal establishment methodology](#).

Figure 9: Proportional discrepancy for current prisoner status from the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)

Figure 9: Proportional discrepancy for current prisoner status from the Veterans' Survey 2022 to Census 2021 linkage (England and Wales)



Source: The Veterans' Survey 2022 to Census 2021 linked data from the Office for National Statistics

5 . Summary, recommendations and limitations

The linkage between the Veterans' Survey 2022 and Census 2021 has been shown to be of high quality, with an estimated precision and recall of 97.55% and 99.96%, respectively. The linkage rate was slightly lower than we would have expected at 81.08%. Possible reasons for the linkage rate not being higher include:

- respondents leaving the armed forces after Census Day (these respondents will either be absent from the census if serving abroad or have a different postcode to the one given on the survey)
- unidentified duplicates in the Veterans' Survey 2022 data
- for the prisoner population, low response rate in the census and possible movement within prisons and between prison and non-prison addresses
- veterans from outside England and Wales completing the survey or other ineligible respondents filling in the survey with inaccurate details

Bias analysis comparing the linked data with the Veterans' Survey 2022 showed underrepresentation of several groups in the linked data compared with the full Veterans' Survey 2022 data. This may be because of a mixture of linkage failure and coverage error between the surveys. The underrepresented groups included those aged under 30 years, all sexual orientations except straight or heterosexual, those living in Inner London, Swansea or Torfaen, all ethnic groups except White, and non-UK citizens and prisoners.

If analysis is carried out using the linked data, caution should be taken when observing patterns for these groups as analysis outcomes could be the result of linkage bias rather than reflecting true trends or patterns.

6 . Related links

[Veterans' Survey methodology](#)

Methodology | released 15 December 2023

An overview of the background, methodology, data collection, processing and weighting of the UK Veterans' Survey 2022.

[The Veterans' Survey 2022, demographic overview and coverage analysis, UK](#)

Article | released 15 December 2023

Coverage and sample bias analysis of the Veterans' Survey 2022, with weighted estimates for veteran responses in the UK by personal characteristics.

7 . Cite this methodology

Office for National Statistics (ONS), released 15 December 2023, ONS website, methodology, [Veterans' Survey 2022 to Census 2021 linkage report](#)