

# Investigating the accuracy of Standard Industrial Classification using BRES

Examining differences between Standard Industrial Classification (SIC) on the Inter-Departmental Business Register (IDBR) and in the results of the Business Register and Employment Survey (BRES) in 2021.

Contact:  
Brittany Black  
idbrdas@ons.gov.uk  
+44 1633 458902

Release date:  
4 May 2023

Next release:  
To be announced

## Table of contents

1. [Overview](#)
2. [Methods for comparing SIC on the IDBR and in the BRES results](#)
3. [Mismatch rates](#)
4. [Isolating genuine errors](#)
5. [Cite this methodology](#)

# 1 . Overview

The Office for Statistical Regulation (OSR) has recently raised concerns over the quality of Standard Industrial Classification (SIC). In response, the Office for National Statistics (ONS) has conducted its own analysis on the quality of SIC.

The SIC is used to categorise business establishments and other statistical units by type of economic activity. SIC supplies a framework for the uniform collection and analysis of data, including use in National Statistics publications. It is therefore important to maintain public trust and transparency in the quality of SIC. This methodological article will be focusing on SIC 2007.

Our analysis investigates the quality of the SIC on the Inter-Departmental Business Register (IDBR) by comparing the SIC of reporting units (RUs) with the results from the Business Register and Employment Survey (BRES). The IDBR primarily sources SIC using Value Added Tax (VAT) registrations from HM Revenue and Customs (HMRC). This is considered more accurate than Companies House and is the principal dataset that the ONS uses to report on SIC. The BRES allows responding businesses to confirm or change the SIC for each of their own Local Units (LUs) and is therefore considered the best source for confirming the quality of SIC.

Our findings have shown that:

- of the 2021 responders to the BRES, 7.1% have a mismatched SIC on some level across the IDBR and the BRES; however not all mismatches will be a genuine error
- the mismatch rate for SIC within the BRES universe is estimated as 11.5% at the full five-digit level, or 8.7% at the division level
- some 3.8% of the 2021 responders (estimated 7.0% for the BRES 2021 universe) have a mismatched section across the IDBR and the BRES
- all SIC sections within the responding sample had a mismatch rate of less than 8.2% at the section level
- responding reporting units (RUs) that had higher employment and turnover displayed lower rates of mismatching; that is, mismatched RUs are more likely to be having a minimal impact on the economy
- the estimate for mismatch rates using employment within the universe is 6.2% at the full five-digit level
- RUs with a "Not Elsewhere Classified" SIC showed higher rates of mismatching within the sample; they were 10.9% at the five-digit level, compared with 6.6% for RUs that did not have a "Not Elsewhere Classified" SIC
- the genuine error rate for SIC is estimated at 8.1% at the full five-digit level within the universe, and this becomes 6.2% when we exclude strata with less than five RUs responding to the BRES 2020 and the BRES 2021 from this estimate; however, these estimates may not display the full picture, as many strata had zero overlap between the two years and so could not be used to isolate a genuine error rate

## 2 . Methods for comparing SIC on the IDBR and in the BRES results

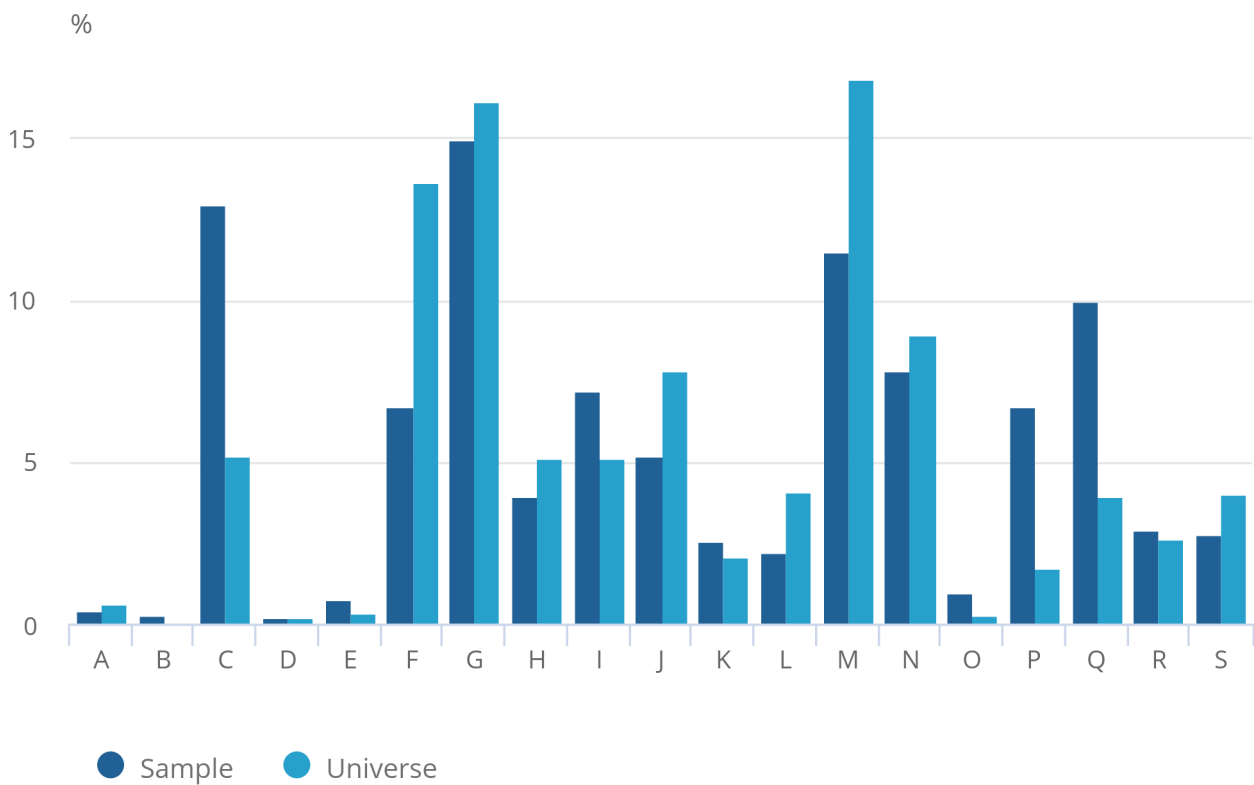
## 2.1 The dataset

The primary dataset for this analysis consists of reporting units (RUs) which responded to the Business Register and Employment Survey (BRES) in 2021 and could also be matched to the 2021 BRES results microdata. This sample consists of 59,230 RUs. The Standard Industrial Classifications (SICs) assigned to this dataset by the Inter-Departmental Business Register (IDBR SICs) cover SIC 2007 Sections A to S.

A bar chart showing the percentage in the dataset of each IDBR section compared with the percentages in the universe for the BRES is shown in Figure 1. Of the 342 sampling strata present in the BRES universe, 333 are covered by the responding sample. IDBR data has been taken from the final selection file for the BRES 2021; BRES data have been taken from the microdata for the BRES 2021 results.

**Figure 1: Comparing percentages of IDBR sections within the BRES universe and the responding 2021 sample**

Figure 1: Comparing percentages of IDBR sections within the BRES universe and the responding 2021 sample



Source: Business Register and Employment Survey from the Office for National Statistics

## 2.2 Determining SIC from BRES data

The results data for the BRES are in a Local-Unit level, so the top-down method was used with employment to find the primary SIC assigned by the BRES to each RU (BRES SIC). However, this method produced two potential primary SICs for 79 RUs (0.13% of the sample). For 73 of the RUs, one of the duplicate SICs matched across the IDBR, so this was assumed to be correct and the mismatched SIC was removed.

For six RUs (0.01% of the sample), both results are equally likely to be the primary SIC; they have both been kept in the data and flagged to avoid being counted twice in one category. This has resulted in some figures not summing up perfectly to their totals.

## 2.3 Defining a mismatch

A match is defined as all five digits of the IDBR SIC and the BRES SIC being identical for an RU. A mismatch is defined as a disparity between the IDBR SIC and the BRES SIC for an RU in the dataset. This can occur on multiple levels, including:

- a Sub-Class Mismatch, which is a difference between the full five-digit SICs
- a Class Mismatch, which is a difference between the first four digits of the SICs
- a Group Mismatch, which is a difference between the first three digits of the SICs
- a Division Mismatch, which is a difference between the first two digits of the SICs
- a Section Mismatch, which is a difference between the SIC 2007 sections of the SICs

For example, SICs 87900 and 87300 would constitute a Group, Class and Sub-Class Mismatch, but would not constitute a Section or Division Mismatch as the first two digits are the same and both SICs belong to section Q. Mismatch rates within this report will be calculated using RU counts unless stated otherwise.

A mismatch does not necessarily indicate an error within the IDBR SIC, as the results of the BRES could also be reflecting a change in the economic activity of an RU, or the SIC on the IDBR is protected to better capture certain types of economic activity. Therefore, there are two types of mismatches occurring. These include:

- "genuine errors", which are instances where the classification stored on the IDBR is incorrect because of an error and has been corrected by the business responding to the BRES
- "benign mismatches", which are instances where the SIC differs across the results of the BRES and the IDBR but not because of a mistake in the IDBR's SIC

### 3 . Mismatch rates

A simple look at the mismatch rates shows that 55,025 (92.9%) reporting units (RUs) in the data completely match across the Inter-Departmental Business Register (IDBR). In other words, the Sub-Class Mismatch rate is 7.1%. The Statistical Classification of Economic Activities in the European Community (NACE), which is the European Union's industrial classification system, only uses the first four digits of the Standard Industrial Classification (SIC). At this Class level, the mismatch rate is 6.7%.

Note that all Section Mismatches are also Sub-Class Mismatches, but not all Sub-Class Mismatches will be Class Mismatches or Section Mismatches. Over half of the mismatches that occur in the sample happen on a sectional level, that is, in the most severe category of change.

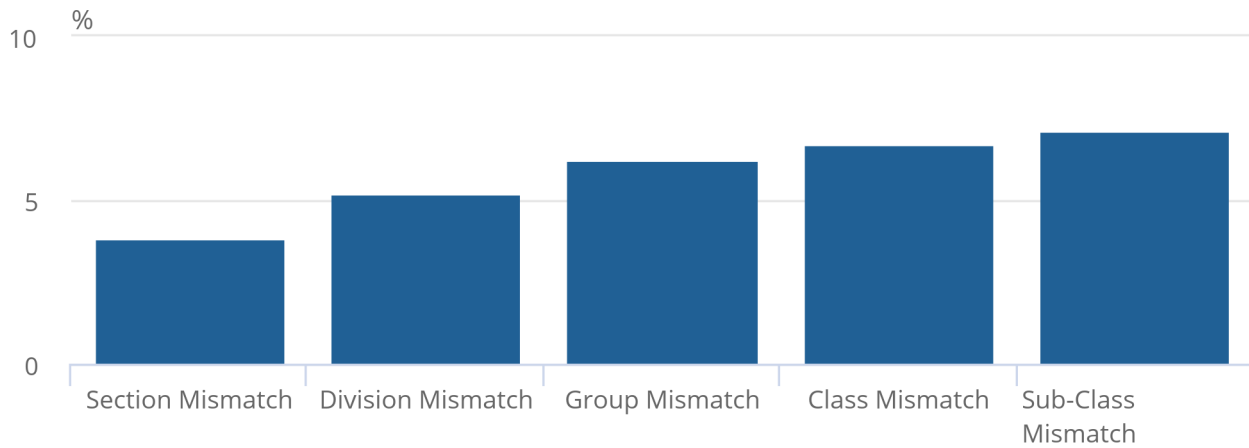
Table 1: Counts and percentages of SIC mismatch types in the responding sample

<b>SIC Match Type</b>	<b>Count</b>	<b>% of Sample</b>
<b>Complete Match</b>	55,025	92.9%
<b>Sub-Class Mismatch</b>	4,205	7.1%
<b>Class Mismatch</b>	3,956	6.7%
<b>Group Mismatch</b>	3,680	6.2%
<b>Division Mismatch</b>	3,073	5.2%
<b>Section Mismatch</b>	2,281	3.8%

Source: Business Register and Employment Survey from the Office for National Statistics

**Figure 2: Percentage of mismatch types within the responding sample**

Figure 2: Percentage of mismatch types within the responding sample



**Source: Business Register and Employment Survey from the Office for National Statistics**

The mismatch rate within the sample is not necessarily the same as the mismatch rate within the universe. This is because not all sections, employment bands, and so on represented in the universe will be sampled proportionately, and not all reporting units sampled will respond.

We can estimate the mismatch rates within the Business Register and Employment Survey (BRES) 2021 universe by isolating the mismatch rates in each sampling stratum of the data and using the universe stratum populations to create a weighted average. This is shown in Table 2. The estimated Sub-Class mismatch rate for the universe is 11.5%. As for the responding sample, over half of the mismatches that occur in the universe happen on a sectional level.

Table 2: Weighted average for mismatch rates in the BRES 2021 universe

Section Mismatch Rate	Division Mismatch Rate	Group Mismatch Rate	Class Mismatch Rate	Sub-Class Mismatch Rate
7.0%	8.7%	10.0%	10.7%	11.5%

Source: Business Register and Employment Survey from the Office for National Statistics

In the following subsections, we examine how the properties of RUs correlate with mismatch rates in the sample.

### 3.1 Sections

There are multiple metrics which we can use to assess the impacts of mismatched SIC on the different sections of SIC. Firstly, the Section mismatch rate shows how often the IDBR and the results of the BRES are misaligned on the section of an RU's SIC. Table 3 shows Sections A to S ordered by descending Section mismatch rate. Section B is the first row as it has the most severe Section mismatch rate. This is because, for all RUs with a SIC in Section B on the IDBR, 8.2% are not assigned to Section B in the results of the BRES.

Table 3: SIC sections' mismatch rates in the IDBR from the results of BRES

<b>Section</b>	<b>Section Mismatch %</b>
<b>B: Mining and Quarrying</b>	8.20%
<b>A: Agriculture, Forestry and Fishing</b>	7.90%
<b>S: Other Service Activities</b>	7.40%
<b>N: Administrative and Support Service Activities</b>	6.70%
<b>R: Arts, Entertainment and Recreation</b>	6.10%
...	...
<b>I: Accommodation and Food Service Activities</b>	1.60%

Source: Business Register and Employment Survey from the Office for National Statistics

#### Notes

1. Ordered by section mismatch rate (descending).

We also look at the count changes of the RUs assigned to sections on the IDBR and in the results of the BRES. This is shown in Table 4, where sections are ordered by descending absolute count percentage change. Once again, section B has the most severe change. This is because the count of all RUs with a SIC in Section B on the IDBR is 11 counts fewer or 6.4% lower than the count of all RUs with a SIC in Section B from the results of the BRES. It is worth noting that a count change of 0 does not mean that all reporting units assigned to section D in the IDBR are assigned to section D by the BRES.

Table 4: Changes to SIC sections' counts in the IDBR from the results of BRES

<b>Section</b>	<b>Count change</b>	<b>Count % change</b>
<b>B: Mining and Quarrying</b>	-11	-6.4%
<b>L: Real Estate Activities</b>	52	4.0%
<b>N: Administrative and Support Service Activities</b>	-149	-3.2%
<b>E: Water Supply; Sewerage, Waste Management and Remediation Activities</b>	12	2.8%
<b>K: Financial and Insurance Activities</b>	37	2.4%
...	...	...
<b>D: Electricity, Gas, Steam and Air Conditioning Supply</b>	0	0.0%

Source: Business Register and Employment Survey from the Office for National Statistics

#### Notes

1. Ordered by absolute count percentage change (descending).

Finally, we examine the changes to employment. Table 5 shows the changes in employment for a section between the IDBR and the results of the BRES ordered by absolute percentage change. Section B still has the largest change by this metric. This is because the total employment of all RUs with a SIC in Section B on the IDBR is 3,510 counts fewer or 9.7% lower than the summation of employment of all RUs with a SIC in Section B from the results of the BRES.

Table 5: Changes to SIC sections' employment in the IDBR from the results of BRES

<b>Section</b>	<b>Employment change</b>	<b>Employment % change</b>
<b>B: Mining and Quarrying</b>	-3,510	-9.7%
<b>E: Water Supply; Sewerage, Waste Management and Remediation Activities</b>	-5,641	-7.2%
<b>O: Public Administration and Defence; Compulsory Social Security</b>	-54,426	-5.2%
<b>F: Construction</b>	-14,740	-4.3%
<b>C: Manufacturing</b>	36,014	3.2%
...	...	...
<b>K: Financial and Insurance Activities</b>	338	0.1%

Source: Business Register and Employment Survey from the Office for National Statistics

#### Notes

1. Ordered by absolute employment percentage change (descending).

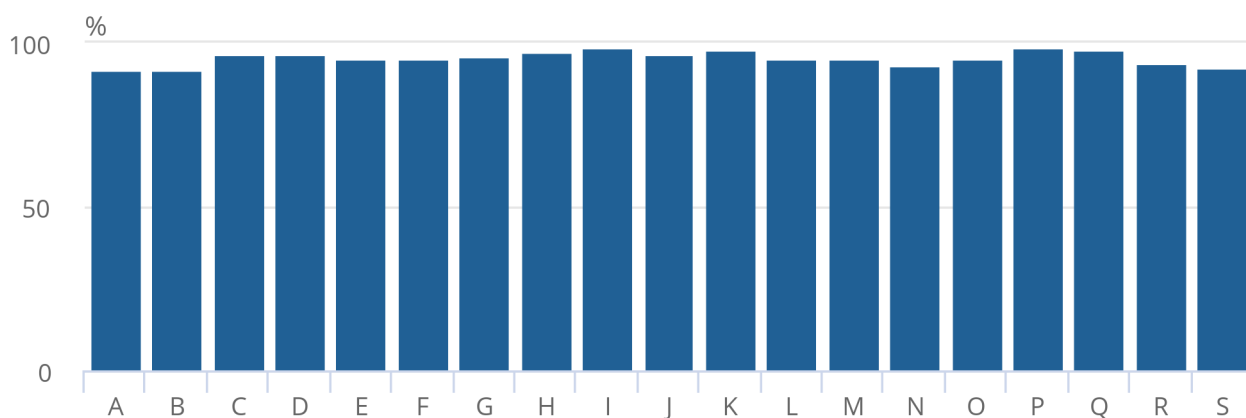


Section B is the most affected section on all metrics, however it is worth noting that it is one of the least represented sections in the sample. Of the RUs listed as Section B in the IDBR, 14 were listed as different sections in the results of the BRES. Three RUs not listed as Section B in the IDBR were listed in this section in the results of the BRES. The most common section (5 of 14) that RUs were assigned to by the BRES is Section M: professional, scientific, and technical activities. Of these RUs, three were specifically reassigned to SIC 71122: Engineering-related scientific and technical consulting activities.

Figure 3 shows at least 91.8% of RUs within an IDBR section are assigned to the same section in the results of the BRES. Figure 4 shows at least 88.4% of employment within an IDBR section is assigned to the same section in the results of BRES.

**Figure 3: Percentage of RUs reassigned to their IDBR Section in BRES**

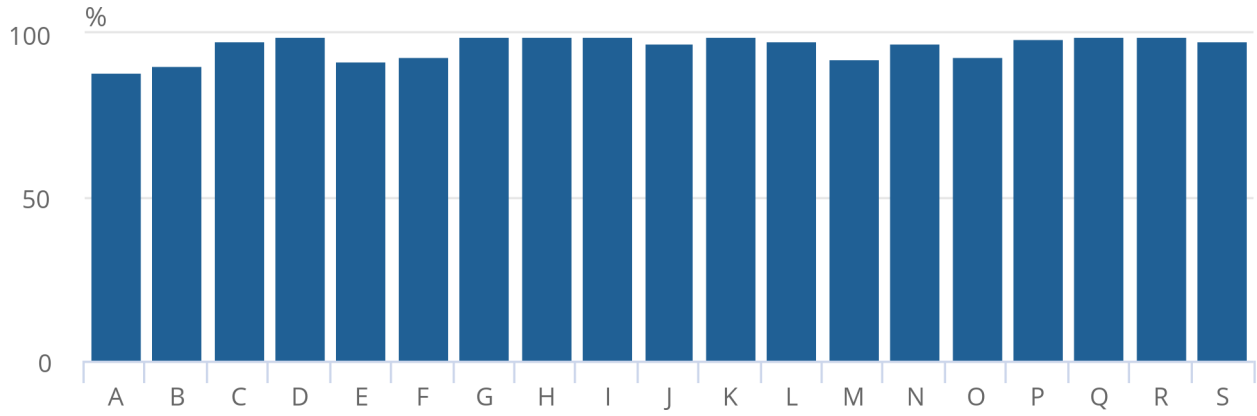
Figure 3: Percentage of RUs reassigned to their IDBR Section in BRES



Source: Business Register and Employment Survey from the Office for National Statistics

**Figure 4: Percentage of employment reassigned to its IDBR section in BRES**

Figure 4: Percentage of employment reassigned to its IDBR section in BRES



**Source: Business Register and Employment Survey from the Office for National Statistics**

## 3.2 Employment and turnover

Employment and turnover can be used to estimate the extent of a business's impact on the economy. Note that these values are not independent, as companies with higher turnover will tend to have more employees.

Table 6 shows that RUs grouped within the lower bands of employment display much higher mismatch rates than RUs in higher employment bands. Similarly, grouping by turnover shows that RUs within the lower bands of turnover were more likely to be mismatched across the IDBR and the BRES than those with higher turnover (see Table 7). RUs with low employment and low turnover have higher mismatch rates, that is, RUs that are mismatched are more likely to be having a minimal impact on the economy.

Table 6: Mismatch rates by grouped employment

<b>Grouped Employment</b>	<b>RU Count</b>	<b>Section Mismatch Rate</b>	<b>Division Mismatch Rate</b>	<b>Group Mismatch Rate</b>	<b>Class Mismatch Rate</b>	<b>Sub-Class Mismatch Rate</b>
<b>Less than 20 Employment</b>	15,080	7.3%	9.2%	10.5%	11.2%	12.0%
<b>20 - 99 Employment</b>	27,952	3.3%	4.6%	5.6%	6.0%	6.4%
<b>100 or More Employment</b>	16,198	1.6%	2.4%	3.3%	3.6%	3.8%

Source: Business Register and Employment Survey from the Office for National Statistics

Table 7: Mismatch rates by grouped turnover

<b>Grouped Turnover</b>	<b>RU Count</b>	<b>Section Mismatch Rate</b>	<b>Division Mismatch Rate</b>	<b>Group Mismatch Rate</b>	<b>Class Mismatch Rate</b>	<b>Sub-Class Mismatch Rate</b>
<b>Less than £250,000 turnover</b>	11,350	6.80%	8.50%	9.60%	10.20%	11.00%
<b>£250,000 - £5 million turnover</b>	24,439	3.90%	5.40%	6.70%	7.20%	7.70%
<b>£5-15 million turnover</b>	11,602	2.60%	3.90%	4.80%	5.20%	5.40%
<b>Turnover more than £15 million</b>	11,839	2.10%	2.80%	3.40%	3.70%	3.90%

Source: Business Register and Employment Survey from the Office for National Statistics

By calculating mismatch rates using employment rather than count, we can better approximate the impact of mismatched SICs when reporting on the economy. We can use a weighted average to estimate these figures for the universe by isolating the employment mismatch rates in each sampling stratum and using the total stratum employments within the universe as weights. This is shown in Table 8. As in Table 2, the estimated mismatch rates for the universe are higher than those in the sample, and over half of the mismatches that occur happen on a sectional level. In line with the results of Table 6, mismatch rates using employment are lower impact than those using RU count.

Table 8: Mismatch rates using employment rather than countshowing actual rates within the sample and a weighted average estimate for the rates within the universe.

Showing actual rates within the sample and a weighted average estimate for the rates within the universe

<b>Mismatch Rates using Employment for...</b>	<b>Section Mismatch Rate</b>	<b>Division Mismatch Rate</b>	<b>Group Mismatch Rate</b>	<b>Class Mismatch Rate</b>	<b>Sub-Class Mismatch Rate</b>
<b>Sample</b>	2.0%	2.6%	3.6%	3.8%	3.9%
<b>Universe</b>	3.3%	4.3%	5.4%	5.8%	6.2%

Source: Business Register and Employment Survey from the Office for National Statistics

### 3.3 "Not Elsewhere Classified" SICs

By definition, "Not Elsewhere Classified" or n.e.c. SICs are businesses that do not easily fit into a pre-existing SIC. Therefore, we would expect higher rates of mismatching for RUs with these SICs. Table 9 supports this, as RUs with n.e.c. SICs have more than double the section mismatch rate and consistently higher mismatch rates of each type.

Table 9: Mismatch rates by N.E.C. SICs

<b>SIC Type</b>	<b>RU Count</b>	<b>Section Mismatch Rate</b>	<b>Division Mismatch Rate</b>	<b>Group Mismatch Rate</b>	<b>Class Mismatch Rate</b>	<b>Sub-Class Mismatch Rate</b>
<b>N.E.C. IDBR SIC</b>	7,121	7.00%	8.90%	9.90%	10.50%	10.90%
<b>Specific IDBR SIC</b>	52,109	3.40%	4.70%	5.70%	6.20%	6.60%

Source: Business Register and Employment Survey from the Office for National Statistics

## 4 . Isolating genuine errors

When only considering the 2021 responding sample of the Business Register and Employment Survey (BRES), it is not possible to figure out whether a mismatch is benign or whether it is a genuine error. However, by looking at the reporting units (RUs) which responded to the BRES 2020 and the BRES 2021, we can isolate some of the mismatches as benign misclassifications. This is because we can assume that any business that has previously responded to the BRES will have been able to correct genuine errors on the Inter-Departmental Business Register (IDBR) then.

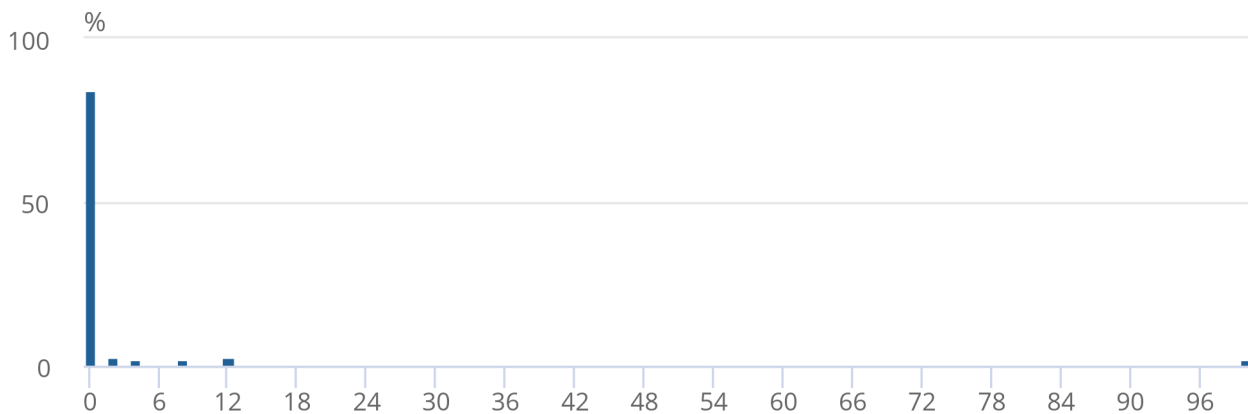
We can assume that the mismatch rate in the responders to the BRES in 2020 and 2021 consists only of benign mismatches. The mismatch rate for RUs who only responded to the BRES 2021 consists of the genuine error rate plus the benign mismatch rate. Therefore, by subtracting the 2020 and 2021 responders' mismatch rate from the 2021 only responders' mismatch rate, we can estimate the genuine error rate. This will be done on a sampling stratum level so that each part of the mismatch rate will be calculated for similar RUs.

## 4.1 Using all strata with 2020 and 2021 responders

Of the 342 sampling strata, 101 have at least one RU that has responded to the BRES in 2020 and 2021. Histograms showing the distributions of the estimated benign mismatch and genuine error rates for each stratum are shown in Figure 5 and Figure 6. Some estimates shown in Figure 6 for the genuine error rate are less than 0%, which is clearly impossible. This is because of sampling error and small samples sizes.

**Figure 5: Benign mismatch rates for each sampling stratum with data with at least 1 RU responding to 2020 and 2021 BRES**

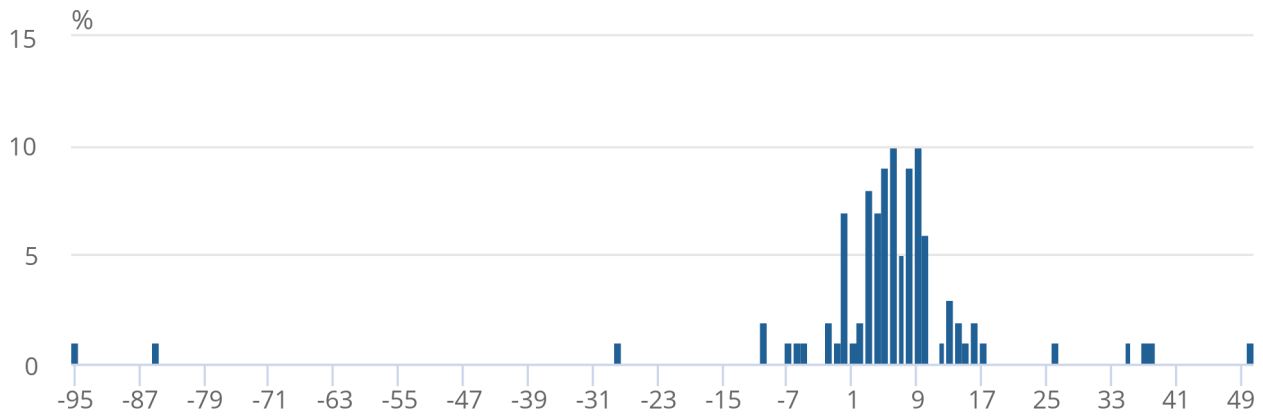
Figure 5: Benign mismatch rates for each sampling stratum with data with at least 1 RU responding to 2020 and 2021 BRES



Source: Business Register and Employment Survey from the Office for National Statistics

**Figure 6: Genuine error rates for each sampling stratum with data with at least 1 RU responding to 2020 and 2021 BRES**

Figure 6: Genuine error rates for each sampling stratum with data with at least 1 RU responding to 2020 and 2021 BRES



Source: Business Register and Employment Survey from the Office for National Statistics

Table 10 shows that the estimate for benign mismatch rates is 0% for at least 75% of strata with data. The quartiles for genuine error rate are more varied, with a median of 6%.

Table 10: Median and quartiles of the benign mismatch and genuine error rates for the BRES sampling strata  
All strata with 2020 and 2021 responders

Metric	Lower (25th Percentile)	Median (50th Percentile)	Upper Quartile (75th Percentile)
<b>Benign Mismatch Rate</b>	0.0%	0.0%	0.0%
<b>Genuine Error Rate</b>	2.7%	6.0%	9.1%

Source: Business Register and Employment Survey from the Office for National Statistics

By calculating a weighted average, we can gain an estimate of each type of mismatch rate for the 2021 sample and universe. This is shown in Table 11. We have estimates of the genuine error rate as 4.0% in the sample and 8.1% in the universe.

Table 11: Weighted averages for the benign mismatch rate and the genuine error rate  
All strata with 2020 and 2021 responders

<b>Weighted Average for...</b>	<b>Using 2021 Sample Weights</b>	<b>Using 2021 Universe Weights</b>
<b>Benign Mismatch Rate</b>	3.7%	2.3%
<b>Genuine Error Rate</b>	4.0%	8.1%

Source: Business Register and Employment Survey from the Office for National Statistics

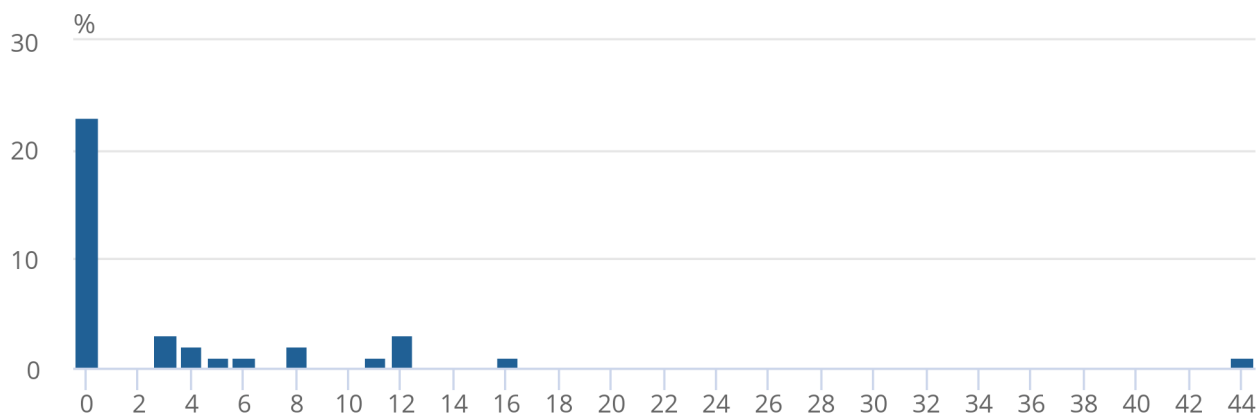
## 4.2 Excluding strata with small sample sizes

Many of the strata have only one RU that responded to the BRES 2020 and the BRES 2021 from which a benign mismatch rate can be estimated. As seen in Figure 5, this has resulted in estimated benign mismatch rates of 100% for some strata. To reduce the risk of our estimated metrics being affected by insufficient sample size, we have recalculated these figures excluding strata for which we do not have at least five RUs that responded to both years of the BRES. Of the 342 sampling strata, 38 have at least 5 RUs that responded to the BRES in 2020 and 2021.

Histograms showing the distributions of the estimated benign mismatch and genuine error rates for each stratum are shown in Figure 7 and Figure 8. Nonetheless, some estimates in Figure 8 show the histogram of genuine error rates to be less than 0%, though there are fewer impossible values than shown in Figure 6.

**Figure 7: Benign mismatch rates for each sampling stratum with data with at least 5 RUs responding to 2020 and 2021 BRES**

Figure 7: Benign mismatch rates for each sampling stratum with data with at least 5 RUs responding to 2020 and 2021 BRES

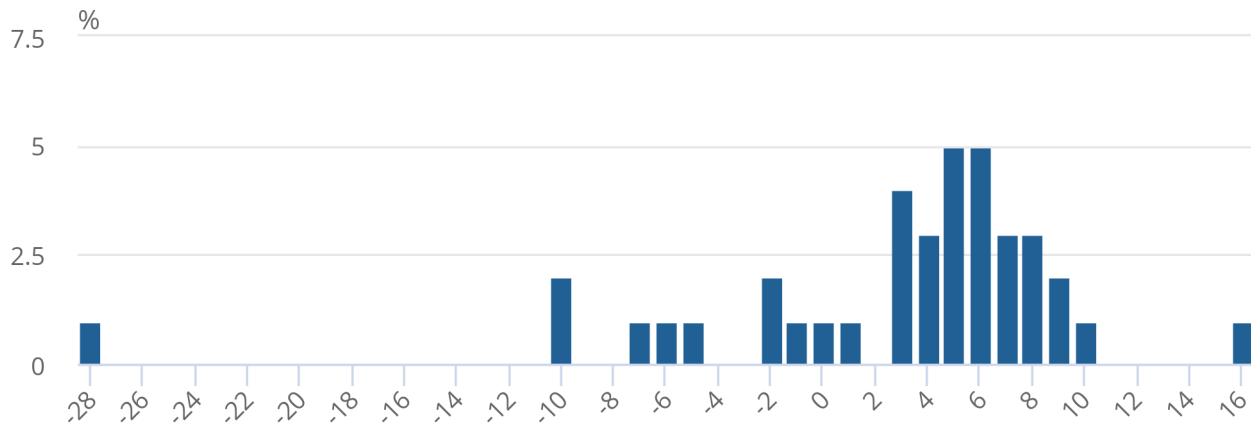


Source: Business Register and Employment Survey from the Office for National Statistics



**Figure 8: Genuine error rates for each sampling stratum with data with at least 5 RUs responding to 2020 and 2021 BRES**

Figure 8: Genuine error rates for each sampling stratum with data with at least 5 RUs responding to 2020 and 2021 BRES



Source: Business Register and Employment Survey from the Office for National Statistics

Table 12 shows that the estimate for benign mismatch rates is less skewed towards zero when excluding small sample sizes from the data. However, it is still 0% for at least 50% of RUs. The interquartile range for the genuine error rate (6.3%) is the same as in Table 10, but the median is slightly lower at 5.2%.

Table 12: Median and quartiles of the benign mismatch and genuine error rates for the BRES sampling strata  
All strata with five or more 2020 and 2021 responders

Metric	Lower (25th Percentile)	Median (50th Percentile)	Upper Quartile (75th Percentile)
<b>Benign Mismatch Rate</b>	0.0%	0.0%	5.2%
<b>Genuine Error Rate</b>	0.7%	5.2%	7.0%

Source: Business Register and Employment Survey from the Office for National Statistics

Table 13 shows the weighted averages excluding the strata with fewer than five RUs responding to both the 2020 and 2021 BRES. The estimates of the genuine error rate are lower: 3.2% in the sample and 6.2% in the universe.

Table 13: Weighted averages for the benign mismatch rate and the genuine error rate  
All strata with five or more 2020 and 2021 responders

<b>Weighted Average for...</b>	<b>Using 2021 Sample Weights</b>	<b>Using 2021 Universe Weights</b>
<b>Benign Mismatch Rate</b>	4.1%	2.0%
<b>Genuine Error Rate</b>	3.2%	6.2%

Source: Business Register and Employment Survey from the Office for National Statistics

## 5 . Cite this methodology

Office for National Statistics (ONS), released 4 May 2023, ONS website, methodology, [Investigating the accuracy of Standard Industrial Classification using BRES](#)