Office for Product Safety & Standards

Evaluation of data from a pilot study into the feasibility of collecting accident data from the NHS relating to consumer products

Research report

November 2025

Contents

Executive summary	3
Background	4
Data collection	5
Data	11
Analysis overview	14
Data quality	16
Insights	19
Product safety relevance	23
Conclusions	24
Appendix 1: Additional findings from exploratory analysis	25

Executive summary

To support OPSS's aim of preventing people and places from product-related harm, it is important for OPSS to understand injuries caused by products. In 2018, OPSS commissioned the Royal Society for the Prevention of Accidents (RoSPA) to conduct a pilot study collecting accident and emergency (A&E) data relating to consumer products from a small number of hospital sites that volunteered to participate in the study. This report covers the findings from OPSS's analysis of the resulting dataset.

The dataset contains extensive information about injuries that goes beyond the level of detail in other datasets that OPSS has access to. In particular, the dataset includes free text clinical notes that contain details on how injuries occurred, and in some cases contain information on the specific product involved.

Our analysis encountered a number of data quality challenges, including a large amount of missing data. This led to significant time being spent on data cleaning and processing. There were also some inconsistencies between data from different hospital sites and time periods, and the level of detail in the free text clinical notes varied considerably. The data are not representative of the UK population as they were only collected from a small number of hospital sites, and therefore cannot easily be compared with other data sources.

We found it was difficult to determine from the data whether products were the cause of the injury. As the data sets were not designed to collect product information, it would be difficult to overcome this limitation without changes to the method of data collection. Our analysis suggested that less than 10% of the data involved injuries linked to potentially unsafe products.

There are potential uses of the data for OPSS, particularly around using it to explore injuries related to specific product types identified from other intelligence. However, our analysis found that the data is not suitable for addressing evidence gaps around:

- how many product related injuries occur in the UK each year;
- how many injuries linked to product safety issues occur in the UK each year;
- what trends in product related injuries can be identified that might point to new or emerging risk areas.

Further data collection would require substantial investment in pre-processing (before data are received by OPSS) to improve data quality, as well as considerable resource from OPSS analysts to fully utilise the dataset. While the analysis conducted so far has provided interesting insights, it has not yet directly contributed to OPSS's regulatory work. Therefore, we do not recommend further data collection directly from existing A&E datasets at this time, though there remains a continued need for product-related injury data to support OPSS's product regulation work.

Background

The Office for Product Safety and Standards (OPSS) is the UK's product regulator. We are responsible for the regulation of most consumer goods (excluding food, medicines and vehicles) and we are the national regulator for construction products. We hold policy responsibility for product safety, legal metrology (weights and measures), standards and accreditation, hallmarking, and Primary Authority. We enforce regulations across the product lifecycle from design, manufacture and assessment through to supply, end use and safe disposal.

Injury data was identified during the development of OPSS's Data Strategy 2022-2025 as one of four priority data gaps. While OPSS has access to injury data from a range of sources, there is no single data source that provides a wider view of product related injuries in the UK in sufficient detail to allow injuries to reliably be attributed to specific product categories.

Broadly speaking OPSS's data needs fall into three categories:

1. overall high level figures for the number of product-related injuries in the UK each year;
2. information on the number of injuries relating to specific product categories, that allows monitoring of trends over time; and
3. detailed information on the nature of injuries relating to specific product categories.

Ideally data would cover the whole of the UK, or be collected in a way that data could be considered representative of the whole of the UK population.

In 2018, OPSS commissioned the Royal Society for the Prevention of Accidents (RoSPA) to develop a feasibility study on collecting accident data relating to consumer products from the NHS. The aims of the study were to review the product-related injury data already being collected, and to assess how it could be enhanced to make greater use of free-text fields to collect product related data. Five hospitals in the Southeast and the Midlands participated in the study and data was collected from January 2019 – September 2022. This report describes OPSS's evaluation of the data that was collected.

Aims and objectives

The aim of OPSS's evaluation was to evaluate the usefulness of the accident and emergency data to address the data needs set out above. The assessment focused on the following objectives:

- assessing the quality, completeness and usefulness of the obtained dataset;
- gaining insight from the data on affected age-groups, types of injuries and the products causing them; and
- exploring the potential to gain additional insights by applying data science techniques to the dataset.

Data collection

Hospital sites

Hospital sites were identified based on RoSPA's existing network. Initially, the project had six sites, consisting of five hospitals and one GP surgery. However, technical challenges resulted in two sites (one hospital and the GP surgery) dropping out.

The hospital could not participate due to the fact it could not extract the data required due to the technical limitations of the IT systems. The GP surgery was unable to extract data. After discussions with three GP practices, it was concluded that it was not possible to extract consumer product related injury data from GP practices in general due to the systems and the amount of data, most of which would not be applicable to product related injuries. Another site (site 3) that was recruited at the beginning of the project provided limited data and they dropped out due to a lack of resources at the site. This site was replaced with a new hospital (site 5) which was onboarded in 2021, and as a result data was collected from four sites for three years: from 2019 to 2021. Two sites provided limited data for 2022.

Data processing

It was recognised at an early stage that because the project deals with NHS data, to satisfy data protection requirements the most practical way of securing agreements from the participating hospitals to share their data was to keep the initial transfer within the NHS. Therefore, it was agreed to work with an NHS Commissioning Support Unit to conduct the cleaning and initial analysis of any data received.

For the first two years of the project the North East London Commissioning Support Unit (NEL CSU) provided support with onboarding sites in addition to the data collection, anonymisation and analysis. NEL CSU ceased to exist from April 2022, and Arden and Greater East Midlands Commissioning Support Unit (ARDEN GEM) took over for the remaining year of the project.

In total NEL CSU spent approximately 23 days on this project, receiving 162,016 rows of data and retaining 43,994 rows to then subsequently check for personally identifiable data. AGEM CSU spent approximately 125 days, receiving 369,219 rows and retaining 206,526 to check for personally identifiable data.

Data collection approach and specification

The participating hospitals were asked to provide injury data that is related to injuries and consumer products for each year of the programme. The data requested did not include personal identifiable information. Full details of the data requested, and the mechanisms for transmission and processing, were included in the data protocol which was incorporated into the partnership agreements sent to each site. The data protocol includes the method of data collection and the expectations for each hospital. The relevant parts of that protocol are provided below. Any hospital attendances that did not relate to injury were excluded.

Table 1: fields collected

Field	Name of the field in the Emergency Care Dataset
Age[1]	AGE AT CDS ACTIVITY DATE
Gender	PERSON STATED GENDER CODE
Injury date	INJURY DATE
Injury time	INJURY TIME
Activity date	ARRIVAL DATE
Activity time	ARRIVAL TIME
Presenting complaint	CHIEF COMPLAINT
Place of injury	PLACE OF INJURY
Activity	INJURY ACTIVITY TYPE
Activity status	INJURY ACTIVITY STATUS
Intent	INJURY INTENT
Mechanism	INJURY MECHANISM
Primary diagnosis	DIAGNOSIS
Primary treatment	TREATMENT
Outcome	DISCHARGE DESTINATION
Outcome	DISCHARGE STATUS
Available free text field that may further describe the nature of injury	

[1]Two of the sites aggregated age to the following age bands: 00-04, 05-09, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90+.

While every site was asked to provide all of the fields listed in Table 1, not all were able to do so. This was mainly due to factors such as variations in IT systems and differences in how the data was recorded at each location. (For example, not every hospital collected, or was able to retrieve and report on, data from every field).

Data collection considerations

One of the key challenges of the project was concerns around protecting personal data. The project only required anonymised data with no patient identifiers. Stringent measures were put in place to ensure that the data was first cleaned by the hospital and then by the data analysis centre, but concerns about personal data surfaced repeatedly, particularly when new sites were approached, and it took time to resolve these.

To ensure that personal data was not included, each site was expected to use a data cleaning tool that was provided to them by the NHS Commissioning Unit, and was asked to review data files for any personally identifiable data before releasing it. Once this was complete it was then sent to the NHS Commissioning Unit in a password-protected file with the password sent in a separate email. It was found that some free text fields could

contain data such as names, locations or references to family members that might make patient identification possible. Therefore, once the data was received by the NHS Commissioning Unit, additional checks were made to remove any identifiers.

Data processing

Post processing steps undertaken by the NHS Commissioning Unit included:

- de-identifying participating sites;
- re-indexing records; and
- secondary cleaning of free text.

The processing of data files received by the NHS Commissioning Unit was partitioned into two distinct phases of work. The first phase can be described as structural processing, and the second phase as semantic processing.

Structural processing covered the alignment of data to allow effective comparison between data received from various parties. In general, this processing also focused on the data fields that would have been relatively well managed in their source systems, such as clinical code terminology, event dates, and pre-filled category lists.

Examples of structural processing tasks included:

- aligning the data fields from each provider – the order of data fields varied between providers, and these had to be brought into the correct sequence;
- quantifying obvious gaps in the data such as a missing month, and providing some summary statistics on the size and shape of the data;
- aligning the column names from each provider – the column names had different labels between providers. These needed to be re-mapped to a standard naming convention. This task included a brief scan of the data field content to ensure the right data was being mapped accurately;
- date and time fields were formatted for consistency and separated into their own fields;
- fields with low cardinality (a low count of distinct values, such as gender) were checked for spelling and consistency. This was done to facilitate the creation of category filters on the data;
- some data providers were only able to share age ranges based on agreed bands, e.g. 0-4, 5-9, etc. Where no age was provided or a value appeared outside of the expected range of 0-120, the value of -1 was substituted;
- the absence of data was standardized, converting fields with content such as ‘na’, ‘N/A’, ‘NULL’, or multiple blank spaces. Leading and trailing space was removed;
- some files arrived damaged, and had to be repaired e.g., where a record was separated over more than one row in the CSV file format.

Data processing of this nature was relatively quick, as much of it could be managed through standard software tools.

Semantic processing covered the meaning of the data and subjective interpretation. Due to the nature of the data being shared, particularly the unstructured clinical narrative describing the injury event, robust data protection steps had to be followed. Following an automated de-identification of the unstructured text, a time intensive process of line-by-line

audit was undertaken. The inclusion of certain Emergency Care Data Set (ECDS) fields helped with data interpretation by having consistent terminology to broadly categorise sets of records by. This was helpful as a cognitive aid to work on blocks of records that had a similar theme, which in turn helped to speed up the sorting and categorisation of those records.

Examples of semantic processing tasks included:

- removal of records that described events unrelated to product injuries. This covered topics such as injuries arising from violence, self-harm, substance abuse, arrest, frailty, medical emergencies, diagnostic reports, and complications arising from prior medical care;
- removal of any patient identifiers not picked up during the automated removal step. The most common causes included words joined or separated in a non-standard grammar, company names, and physical addresses;
- removal of social context to the injury event was intended to make the re-identification of the patient harder to achieve. This included swapping terms such as “father” to “parent”, “sister” to “sibling”, “aunt” to “relative” etc. Determining how far to go with this step was subjective, and largely governed by what time could be spent on each record;
- truncation of long records was undertaken if the cause of injury could be established. In general, it was found that the nature of the injury was well described in the first 300 characters of the text, e.g., in the first three sentences of the clinical narrative or less;
- product categories were applied to each record following an assessment of the cause of injury. This could not easily be automated as interpreting how the patient interacted with their environment affected the category selected. For example, a Bluetooth earpiece has both a battery and magnet, two physical objects that were of interest to the project. The CPSC NEISS product injury coding system was used as a baseline and more general categories evolved that were specific for this project;
- additional consideration was required for describing products, especially if referred to by their colloquial names, or when the spelling and grammar affected the understanding of what happened e.g., “tranpoling” is equivalent to trampoline (as the cause of injury) if the rest of the record is understood in that context;
- in some cases, product brands were referenced in the notes, and these were also transcribed for project analysis.

Data cleaning

There were opportunities to enhance the quality of the data. For example, one site had a minor issue with some duplicate records, which were removed, and another site sent files without free text fields, so the data had to be requested again. In some limited cases, the free text notes were repeated. Spelling and grammar can also make finding relevant information harder. For example, for one site, the word trampoline was spelt (or misspelled) nine different ways and it may have been described using the verb (trampolining) rather than the product. Data determined to be out of scope was removed during data processing. These omissions were made because either no evident mechanism of injury related to a consumer product use could be determined, or the product was used with the deliberate intent to cause harm i.e., not an accident.

The main reasons for exclusion were:

- no data present (nearly 14,000 records removed);
- falls (approximately 13,000 records removed);
- includes terms: RTA, RTC, car accident, head on collision, hit by a car (approximately 7,500 records removed);
- text too short to be descriptive < 20 characters (over 5,000 records removed);
- dyspnoea (difficult or laboured breathing) (approximately 4,500 records removed);
- no injury detected (approximately 3,500 records removed).

These exclusions included records where there was no substantive free text in the record. This ranged from the free text field being empty through to a description that did not describe the event in sufficient detail. For example, one hospital focused on records involving foreign body injuries, however in many of the records the free text was “foreign body”. To group these excluded records, neighbouring fields were also reviewed, such as injury mechanism and presenting complaint, to determine the relevance for the project.

Data categorisation

To categorise and sort the records, where available the NHS Commissioning Unit applied CPSC NEISS categories to each record. This categorisation evolved over time to suit the data that was being received. There were some challenges when categorising and analysing the data, for example:

- difficulty identifying the cause of injury;
- variation in product description; and
- interpreting clinical notes and abbreviations.

Assigning a causation for the injury was particularly challenging when there was more than one product associated with the incident. For example, “81y/o fell on carpet sprained left wrist carrying chair”. This example includes the NEISS categories for chairs and carpets. The carpet may have provided a softer landing for the patient, the patient may have tripped on the carpet, or the chair may have obstructed the patient’s ability to move safely.

Grouping products into categories involved deciding what hierarchies and descriptions link products together. For example, hobs, cookers, and stoves can be used interchangeably for the same freestanding household cooking appliance. The categories used in this analysis are not exhaustive and open to further improvement.

Challenges and limitations

The sites

The time needed to bring hospitals on board varied with each site and presented a number of challenges, including changes in staff and contacts at each site, changes (or updates) to IT systems, approval of contracts for data sharing and working on the project. Two hospital sites approached initially were ultimately not able to participate, although one of these sites did provide some very limited data in the first year. A replacement site entered the programme at a late stage but was able to provide the project with three years of data. The project had to be renewed each year due to the nature of the funding, which led to some errors and delays in extracting data from sites, due to not having long-term data extraction agreements with each site in place.

Data processing

Following NHS reorganisation in April 2022, NEL CSU closed, causing delays and complicating data processing across two sites. Data sharing between these sites was restricted due to missing agreements, requiring received data to be deleted and then resent once new agreements were established.

Covid 19

COVID-19 posed significant challenges to the project in early 2020 around hospital staff availability in light of increased workloads, redeployment, furlough and illness. Some planned activity, such as in-person staff briefings and direct feedback sessions, had to be postponed or modified due to restricted access to hospitals.

NHS Data systems, recording and comparability of data

There are a lot of data points in A&E data and extracting the amount of data, including free text, required for the project was not possible from all the IT systems involved. Different systems used across sites within the NHS made the data extraction more complex for some sites and securing comparable data was challenging. Even if two hospitals used the same IT system, the data and fields collected were different in addition to variation of staff input and staff expertise to extract the data. In addition to the difficulty of achieving comparability between sites, in some cases the data received from individual sites sometimes varied year on year. For example, one site provided free text data without problem in one year but found this problematic in the next year.

Data collection conclusions

The data collection phase of the project resulted in anonymised data on product related injuries being obtained from four hospitals, processed and, with some initial analysis completed, passed to OPSS for further analysis. This involved considerable work in setting up data sharing agreements and close working arrangements with the hospitals and NHS commissioning support units.

The project demonstrated that it is possible to collect product related injury data from NHS hospitals. The willingness of hospitals to participate in the project, especially during a global pandemic and in times of great pressure on the health service was remarkable. There were obstacles in terms of securing data protection agreements, identifying capacity to provide the data and ensuring consistency of the amount and format of data provided, but these were largely overcome, and the data sharing protocols and procedures developed may be useful for future work.

A key learning point was that time needs to be allowed for the development of data sharing procedures, and that each hospital has a different approach to this task. However, in no case did a data protection issue prevent the ultimate release of the data. Collaborating with an NHS organisation, such as the two commissioning support units involved in this project, proved an effective way of overcoming some of the data protection concerns with a project of this nature.

Data

Datasets

In 2022 and 2023, OPSS received two datasets—Dataset A and Dataset B. Before OPSS obtained the data, initial processing was performed by an NHS Commissioning Support Unit. For Dataset A, this task was handled by North East London Commissioning Support Unit (NEL CSU), whereas Dataset B was processed by Arden and Greater East Midlands Commissioning Support Unit (ARDEN GEM). The transition between commissioning support units occurred because NEL CSU ceased operations as of April 2022.

Dataset A

Dataset A contains 43,994 patient records, and 19 variables. Some variables were collected from Accident & Emergency (A&E) hospitals, and additional variables were added during data processing to assist with data merging from different hospital sites and capture information extracted from free text. Dataset A was collected from hospital sites 1, 2, 3 and 4 between 1st January 2019 – 31st December 2020. However, the collection period varied between hospitals, as shown in Table 2.

Table 2: Data collection periods from each hospital in Dataset A

Site	Start date	End date
1	1st January 2019	15th December 2020
2	1st January 2019	31st December 2020
3	2nd February 2019	29th March 2020
4	14th March 2019	30th March 2020

Dataset B

Dataset B contains 206,526 patient records, and 16 variables. Again, some variables were collected from A&E hospitals and additional variables were added during data processing. Datasets A & B are not consistent: age was recorded differently in dataset B (in age bands rather than single years), and the fields added during data processing were different.

Table 3: Data collection periods from each hospital in Dataset B

Site	Start date	End date
1	16th December 2020	5th September 2022
4	1st April 2020	31st January 2022
5	1st January 2019	1st December 2021

An illustrative example of the data is shown in Table 4. Note that this is made-up data and not extracted from the real datasets.

Table 4: Illustrative example of collected data

Group	Variables collected	Example data
Site	Site ID	4
Demographics	Age, Age group, Gender	16, 15-19, Male
Clinical	Date, Time, Complaint, Intent, Diagnosis, Disposition	20/02/23, 09:30, Injury to head, Non-intentional injury, Elbow fracture, Admitted
Injury	Mechanism, Activity type, place	Tripping, Indoor recreation, Home
Free text	Free text	Tripped over his slippers and hit head on a metal strip between rooms
Product	Category, Brand	Shoes – slippers, N/A

Hospitals

The data was collected from 5 hospitals. The age range, region, and volume of data per each hospital site is shown in Table 5. Two were paediatric hospitals, and the rest were general hospitals.

Table 5: Age range, region, and data volume per hospital site

Site	Age range	Region	No. of cases
1	0-90+	South East	68,957
2	0-90+	East Midlands	12,744
3	0-16	South East	178
4	0-19	East Midlands	31,079
5	0-90+	East Midlands	88,067

Data processing

Dataset merging

Once OPSS had received the data, some processing and merging was necessary to create a single dataset for analysis. The following steps were applied.

1. Making the data uniform across two datasets, by applying consistent formats.
2. Removing cases where no product was identified, or product categories involved were outside OPSS's remit.
3. Grouping similar product categories.
4. Removing columns not required for further analysis.
5. Splitting the product category variable into category and sub-category.
6. Removing duplicate records.

This resulted in a dataset with 201,025 records, and 15 fields.

Data grouping

Variables such as injury mechanism, activity type, injury place, and disposition contained a high number of unique values. In order to perform analysis and extract meaningful insights, data grouping was performed on those variables to identify similar data and group them into categories.

Analysis overview

After processing, we analysed the data to investigate three key areas.

1. Data quality.
2. What insights could be drawn from the data.
3. To what extent it was possible to identify, from the data, product-related injuries that might indicate a potential product safety concern.

Data quality

The data quality assessment was based on 5 data quality dimensions as defined by the Office for National Statistics. The evaluation measures the state of completeness, validity, consistency, timeliness and accuracy that makes data appropriate for a specific use.

Insights

We carried out exploratory data analysis to summarise the main characteristics of the dataset.

Secondly, topic modelling, specifically Latent Dirichlet Allocation (LDA), was applied to the same data. The aim was to use this technique to identify whether this data highlighted (clustered) products or parts of products (mechanisms) causing injuries to people. Using high level topic modelling, a broad overview showed the key themes across all the documents/client notes. Topics are represented by several words that occur together e.g. ‘burn’ and ‘kettle’. Each topic was further investigated to see if there was any notable clustering within it. Coherence scores were used to help evaluate the significance of topics, but mostly human judgement was the reliable evaluation method. Human judgement includes understanding OPSS’s remit and relevant areas of focus, so that when clusters of words appear, we could assess whether OPSS could utilise this information to inform our regulatory activity.

Finally, we investigated techniques that could be used to automatically extract product information from free text information. Over 70% of the records were missing the product category variable, as pre-processing of Dataset B was not completed before data were provided to OPSS. Extracting this information manually is very time intensive, so we wanted to explore whether we could build an automated tool to do this using a large language model (LLM). We investigated 2 approaches for product extraction:

1. the product category was predicted from free text using the RoBERTa model on existing product categories
2. the product was directly extracted from the free text using the T5 model.

RoBERTa was developed by Facebook. It is a transformers model pretrained on a large corpus of English data in a self-supervised fashion (Hugging Face, 2024). Product category prediction was performed using RoBERTa where free text and possible product categories were given as mode inputs. Zero-shot learning was used where no prior samples were observed during training. As the datasets contain some missing data, it was not possible to perform listwise deletion, instead the term ‘unknown’ was used to fill missing data. 5-fold cross-validation with stratified split by product category was used to assess model performance. Product categories which appear in less than 5 cases were removed, to comply with the cross-validation strategy.

Secondly, instead of performing classification, the products were extracted directly from free text using the Google T5 model. T5 is an encoder-decoder model pre-trained on a multi-task mixture of unsupervised and supervised tasks, and for which each task is converted into a text-to-text format. Specifically, Flan-T5 was used in this study. Flan is a pretraining method that is based on prompting (asking a question to the LLM to get a desired output). A prompt ‘Identify products that caused injury' was used in the model to extract products. We then checked whether the extracted information could be found in the original text to avoid errors.

Due to a high number of unique values in the data sets, product groups were used instead of raw data. The experiment assumes that the grouping performed in the preprocessing stage is correct. There were 48,901 cases with complete free text and product category. 15 product categories with less than 5 cases were removed, resulting in a total of 57 product categories.

Product safety relevance

While it is broadly useful for OPSS to understand all product related injuries, OPSS’s main interest is in injuries resulting from an unsafe product, so we were interested in exploring to what extent we could identify such incidents from the data. We explored two approaches for identifying incidents more likely to be associated with an unsafe product. The first was to develop a classification model to classify whether cases potentially involved an unsafe product, and the second was to use topic modelling to identify whether the data highlighted (clustered) products and injuries.

The first approach aimed to design a supervised classification model that could classify and label whether cases potentially involved an unsafe product. This could be used to filter the dataset and triage the information to relevant OPSS teams. For supervised modelling, labelled data is needed. We asked OPSS colleagues to label 2,000 records. Multiple opinions were necessary because the client notes were sometimes short and vague in description, so it was not always clear whether there was a potential link to an unsafe product. Augmentation techniques were utilised to increase the number of the medical records labelled.

Data quality

Completeness

Tables 6 and 7 show the proportion of missing data for each variable in Datasets A and B and in the merged dataset. Variables with a high proportion of missing data include product category, injury mechanism, activity type, and injury place, which are all important variables for carrying out further analysis of the data. Completeness of variables varied across sites, with sites 4 & 5 in Dataset B having entirely missing data for injury mechanism, activity type, and injury place. The brand variable had very low levels of completeness for all sites.

In general Dataset B had more missing data than Dataset A – in particular product category was 100% populated in Dataset A for all sites, but had much lower levels of completeness in Dataset B. This is because pre-processing of Dataset B was not completed, only around a quarter of the free text was analysed and product information extracted.

In addition, we found that some clinical notes were truncated at 300 characters, so the free text field was not always complete even where it did contain data.

Table 6: Data completeness by site in Datasets A and B

	Dataset A				Dataset B		
site	**1**	**2**	**3**	**4**	**1**	**4**	**5**
age	100%	100%	100%	100%	100%	100%	100%
gender	53%	100%	100%	100%	100%	0%	0%
date	100%	100%	100%	100%	100%	100%	100%
time	0%	100%	100%	0%	0%	100%	0%
complaint	100%	39%	7%	100%	100%	100%	100%
diagnosis	99%	98%	100%	0%	99%	100%	100%
mechanism	90%	100%	91%	63%	57%	0%	2%
activity type	77%	96%	2%	18%	55%	0%	3%
injury place	62%	100%	2%	15%	29%	0%	4%
disposition	100%	100%	100%	100%	100%	100%	100%
free text	100%	96%	100%	100%	75%	94%	83%
category	100%	100%	100%	100%	39%	11%	27%
brand	1%	0%	1%	2%	0%	0%	0%
Number of cases	21,117	20,290	197	2,390	69,335	31,179	106,012

Table 7: Missing data in merged dataset

Column	Number of rows with missing data	Percentage of data missing
Gender	123,587	61.5
Age bracket	36	0
Complaint	8,245	4.1
Diagnosis	2,804	1.4
Mechanism	141,242	70.3
Activity type	147,617	73.4
Injury place	159,443	79.3
Disposition	97	0
Free text	5,572	2.8
Product category	146,552	72.9
Intent	77,127	38.4

The missing data analysis found that injury mechanism and activity type, and injury place and activity type were likely to be missing together.

Accuracy

Because OPSS could not verify if the entered values accurately matched real-world outcomes, accuracy checks were limited to confirming that the data was in the expected format.

We found that in the Diagnosis column, 10.4% of unique variables were codes represented by a mixture of numbers and text (e.g. 01003L), which did not comply with any codes OPSS had access to. All of those cases were from site 2.

The injury place and free text variables sometimes contained misspellings, which made it harder to interpret the data. It was difficult to assess the accuracy of the injury place variable as some of the data were significantly redacted due to the need to remove Personal Identifiable Information (PII) before data were provided to OPSS. Some of the inputs did not appear to be injury places, e.g. 'Assault'.

Consistency and Validity

The datasets have a number of inconsistencies including different coding systems between different sites. Missing data was recorded differently across the datasets. For example, missing data in the Gender variable of Dataset A was recorded as ‘U’, but this code was not used elsewhere.

Dataset A recorded age, while Dataset B recorded age group. Patients’ age range was different across sites as some of the data was collected from paediatric hospitals.

Different hospitals also used different terminologies when recording data. There were a number of terminologies which potentially refer to the same definitions. For example,

‘Foreign body left in wound’ and ‘Foreign body : in wound’; ‘Play – playground’, and ‘Play - playground/soft play’ and ‘Play area – slide’; ‘head injury’ and ‘injury of head’.

In addition, the level of detail in clinical notes varied considerably between records. The text length ranged from 1 to over 700 words.

Timeliness

Timeliness indicates whether the data is available when expected and needed. We looked at the date and time the data was collected and the time it took for the data to be received by OPSS.

Dataset A was collected between 1st January 2019 – 31st December 2020. The data was processed and sent to OPSS in 2022. Dataset B was collected between 1st January 2019 – 5th September 2022. The data was processed and sent to OPSS in May 2023. This indicates at least a 7 month delay in receiving the data. This was mainly due to the significant time it took to perform data pre-processing.

Data quality conclusions

The results from the data quality evaluation showed that the data sets suffer a number of data quality issues. There were high percentages (> 70%) of missing data in variables particularly of interest to OPSS, such as product category. Inconsistencies between the data collected from different sites and the pre-processing that was carried out on the two datasets also present significant data quality challenges.

Spending more time on pre-processing could address some of these issues, but that would impact the timeliness of the data and would also be very resource intensive.

Insights

Exploratory data analysis

We completed univariate (each variable is explored separately) and bivariate (two variables are explored together) analysis on the data sets. As the data sets are from selected hospitals, the results cannot be generalised and may not be representative of injuries in England. We did not explore any comparisons between the demographics of the data and the demographics of the local populations around the hospitals. Note that as previously reported, we encountered significant data quality issues, which limits the insights that can be drawn from the data.

Table 8: A breakdown of top 10 product categories

Product category	Number of cases	Percentage of cases
Missing	146,552	72.9
Buildings	13,006	6.5
Furniture	8,853	4.4
Vehicles	6,643	3.3
DIY & Trade	4,493	2.2
Play	3,203	1.6
Kitchen Utensils	2,171	1.1
Trampolines	2,141	1.1
Cycling	1,975	1.0
Household Items	1,132	0.6
Others	10,856	5.4

Table 8 shows a breakdown of product categories in the data set. 72.9% of the product categories were missing. Falls were the leading cause of injury in reported incidents where the product category was identified, with blunt and sharp injuries following. Notably, just because an injury involves a certain product does not mean that the product was unsafe or failed to meet standards; for instance, someone could be injured by tripping over a fully compliant product.

We found that more than half of the injuries happened at home (56%), followed by public area (21%), and outdoor (13%). Over half of the injuries occurred either during indoor recreation (22%), social play/recreation (15%), descending stairs (8%), or walking indoors (8%).

Appendix 1 shows further breakdowns of the data produced during the exploratory analysis.

We looked into what the data could tell us about some specific products or categories of product that other OPSS work has given us insights into.

- E-bikes and e-scooters: most of the injuries related to micro-mobility vehicles in the data were due to falls, whereas OPSS's main interest in these products relates to lithium-ion battery fires.
- Toys: the main injury types in the play and toy categories were falls and foreign body object ingestion. The main products associated with these injuries were trampolines, blocks, lego, play dough, and magnetic balls.
- Musical instruments: in Wave 6 of OPSS's Consumer Research survey, one of the findings was that the prevalence of safety incidents with musical instruments was high (31% in wave 6), with reported incidence having risen since wave 1 (6%). We investigated injuries linked to musical instruments and the analysis suggested that these injuries were largely due to accidents unrelated to product safety concerns (e.g. someone walking into a piano).

The findings from the exploratory data analysis indicated that a high proportion of the injuries appeared unlikely to be caused by unsafe products. While it isn't possible to draw any conclusions from the hospital data about whether the product concerned was compliant with relevant standards, an accident such as someone tripping over a toy that has been left on the floor is unlikely to provide much useful information from a product safety perspective. Some information on injuries relating to specific product types could be extracted from the data, particularly via reading the free text notes for each relevant entry, however this was of mixed usefulness. The fact that a high proportion of the records are missing at least one of product category or injury mechanism is a significant limitation.

Topic modelling

Topic modelling was applied to clinical notes to uncover key insights. This method detects patterns of words and phrases in text data, allowing documents to be grouped by shared language and themes. Initially, we used topic modelling to gain a broad understanding of the content within the clinical notes, which helped us identify themes.

Topic modelling identified six distinct topics, which can be visualised as separate bubbles on a graph. These bubbles did not overlap, indicating clear topic boundaries. Increasing the number of topics would result in the bubbles beginning to overlap, making the topic representations less distinct. The six identified topics were burns, personal vehicles, eye-related accidents, laceration accidents, falls, and playground related injuries.

Topic 1: Burn injuries

This topic consisted of 4,043 household accidents involving burns (2% of total accidents). We found two notable sub-categories: burns from hot liquids (3,024 accidents) and children swallowing small, non-food items (1,019 accidents).

Topic 2: Personal vehicles

This topic consisted of 6,640 incidents (3.3% of total incidents) involving falls from scooters and bikes, leading to various injuries.

Topic 3: Eye injuries

This topic consisted of 1,056 eye injuries (0.5% of all injuries), from a range of causes including chemical exposures and impact injuries from everyday objects.

Topic 4: Laceration accidents

This topic included cuts from kitchen knives (2,597 incidents), broken glass (1,231 incidents), and injuries from tools like saws and chainsaws (1,237 accidents).

Topic 5: Falling and tripping

This topic included around 80,000 incidents of falls, primarily within home settings. These incidents involve stairs, staircases, steps, doorsteps, and related structures such as stair gates.

Topic 6: Playground related injuries

This topic consisted of 6,761 incidents (3.4% of total incidents) related to injuries from playground and construction equipment, such as swings, climbing frames, monkey bars, slides, and scaffolding. These incidents occurred in diverse settings, including schools, parks, gardens, and construction sites.

The results of the topic modelling provide limited value from a product safety perspective, as none of the topics identified specifically address categories of injuries that are likely attributable to unsafe products. The burns injury category does include burns injuries from children swallowing small, non-food items, but this is only a subset within a broader category identified from the topic modelling. So, while the topic modelling successfully clustered different types of injuries (choking, burns, falls etc.), the general focus of the data on all types of injuries constrains its utility in informing targeted product safety activity.

Product extraction

Our findings showed that RoBERTa had a longer inference time compared to the T5 model. As a result we limited the evaluation of RoBERTa to 100 randomly selected cases on selected products with a high number of cases. The product categories assessed included play, kitchen utensils, trampolines, household items, small items, personal care, home appliances, and baby & infant items. Overall the model's performance was poor. We observed better results when the product category appeared in the free text, even if it was misspelled. We also noticed that in some cases the product categories were not presented in free text but were derived from the activity type variable, which would make it difficult for the model to produce accurate predictions.

The T5 model was more successful: using this method, we were able to extract products from 81% (of around 158,000 rows) of free text. It is important to note that subsequent verification to ascertain whether the extracted products were responsible for the reported injuries was not conducted.

Overall, the T5-flan model showed some potential for being able to automatically extract the products from free text. If more data were to be collected, this would be useful to explore further as it has potential to save time on the pre-processing of the data. However, it would be essential to have strong model monitoring and evaluation mechanisms in place.

Table 9: F1 scores from RoBERTa on 100 randomly selected cases

Product category	Precision	Recall	F1
Baby & infant	0.00	0.00	0.00
Home appliances	1.00	0.12	0.22
Household items	1.00	0.18	0.31
Kitchen utensils	0.33	0.23	0.27
Personal care	0.31	0.50	0.38

Play	1.00	0.04	0.07
Small items	0.12	0.86	0.21
Trampolines	0.68	0.85	0.76

Note: Precision measures the proportion of positive predictions that are correct. Recall measures the proportion of actual positive cases that the model gets right. The F1 score combines precision and recall into a single measure of how well the model is performing.

Product safety relevance

Of the 2,000 medical cases labelled, only approximately 6% were labelled as relevant to product safety, making them a minority class. To address this, data augmentation was employed to generate additional examples of relevant cases, aiming to reduce the dataset's imbalance. However, it encountered challenges with very short sentences in clinical notes; altering or removing even a single word risked obscuring critical details such as the mechanism or location of injury, resulting in incoherent examples with limited value for training. Despite this, a classification model based on the Random Forest method was developed using the newly augmented training dataset. The model classified 8% of relevant product safety medical records correctly, indicating that 92% of relevant cases were misclassified as not relevant, a high false negative rate. Consequently, despite the implementation of augmentation methods, it wasn’t possible to develop a reliable automated tool for classifying cases relevant to product safety.

A key issue here is that the free text is not typically product focused. The focus of the clinical notes tended to be on the injury, and the level of detail recorded about the product was inconsistent.

Conclusions

Our analysis drew out some interesting findings from the data, and suggested some potential use cases. The data contained some injury information that was not available in other data sets OPSS have access to, particularly regarding the detail of what activity led to the injury, and the location of the incident. The free text clinical notes provide additional information which in some cases gives a detailed description of how injuries occurred, though the level of detail in the clinical notes varied widely. We found that the data complements other data sets OPSS has access to – for example, we were able to gain insights into musical instrument related injuries to help explain the findings from wave 6 of OPSS's consumer research survey.

Nonetheless, our analysis of the data proved difficult and required significant resources. Much of the data was of poor quality and there were many gaps. There were a high number of unique values, which had to be grouped together before we could draw meaningful conclusions. OPSS spent considerable time cleaning and processing the information. Further complications arose from inconsistencies across different sites and varying detail in free-text clinical notes, making analysis difficult. Additionally, since this data originated from just five hospital sites, including several paediatric hospitals, it does not accurately reflect the UK population and cannot be reliably compared with other sources.

Another key limitation of the data for OPSS is that it was difficult to determine whether injuries were caused by unsafe products. Since clinical notes were not designed to record details about the products involved, the information available on products varied considerably. Addressing this issue would likely require changing how data is collected. Our analysis suggests that less than 10% of the data involved injuries linked to potential unsafe products. Topic modelling of the clinical notes revealed that, for most of the topics identified there was insufficient detail to draw useful conclusions about any particular product's role in injuries, and the topics mainly captured groups of accidents unlikely to be consistently related to product safety issues (e.g. falls).

There are potential uses of the data for OPSS, particularly around using it to explore injuries related to specific product types identified from other intelligence. However, our analysis found that the data is not suitable for addressing evidence gaps around:

- how many product related injuries occur in the UK each year;
- how many injuries linked to product safety issues occur in the UK each year;
- what trends in product related injuries can we identify that might point to new or emerging risk areas.

Collecting additional data would require significant additional investment in pre-processing (before data are received by OPSS) to address some of the data quality issues. Even after pre-processing, OPSS analysts would still require significant resource to make effective use of the resulting dataset. So far, our analysis has been interesting, but has not yielded any findings that have been used to inform OPSS's regulatory activity. Therefore, we do not recommend further data collection directly from existing A&E datasets for OPSS product regulation work at this time. There does however remain an evidence gap around product-related injury data to support OPSS's product regulation work.

Appendix 1: Additional findings from exploratory analysis

This section shows a range of charts which were produced during the exploratory analysis of the dataset. When reviewing these, it is important to bear in mind the data quality issues and other limitations of the dataset discussed in the main body of this report - particularly that the data cannot be taken to be representative of the whole UK population, and that there was a great deal of missing data for some key fields. The charts are shown here to give an indication of what was covered by the exploratory analysis of the dataset, and the types of insights that could be drawn from the dataset.

Time series trend and patients' characteristics

Figure 1: Number of cases per month

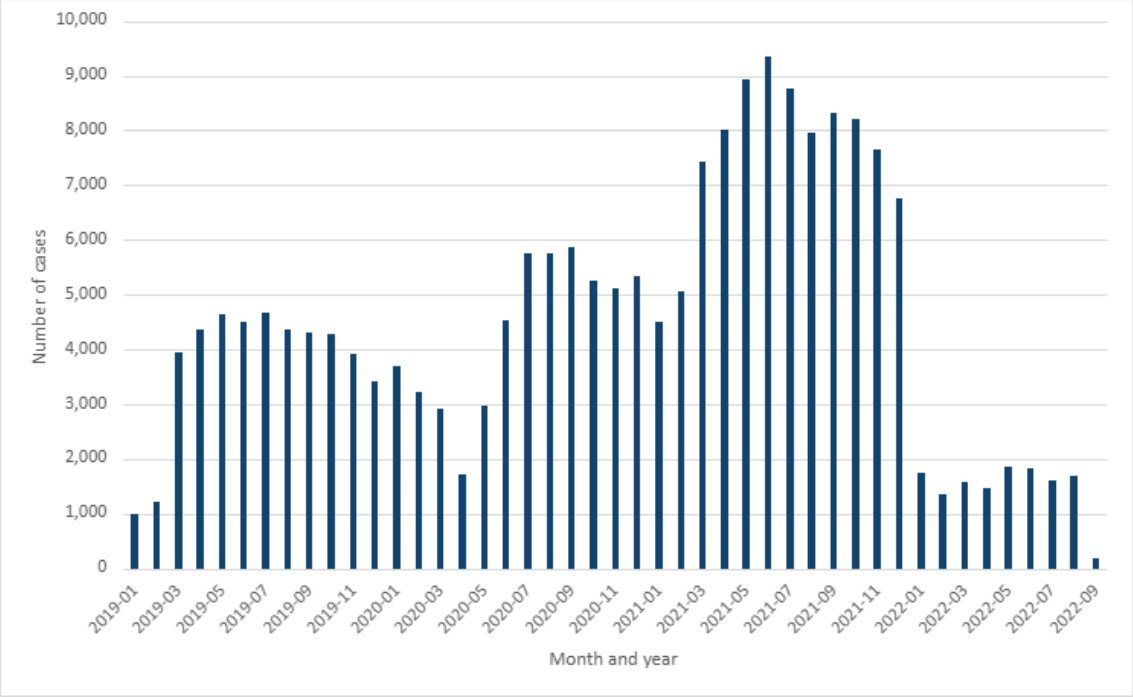
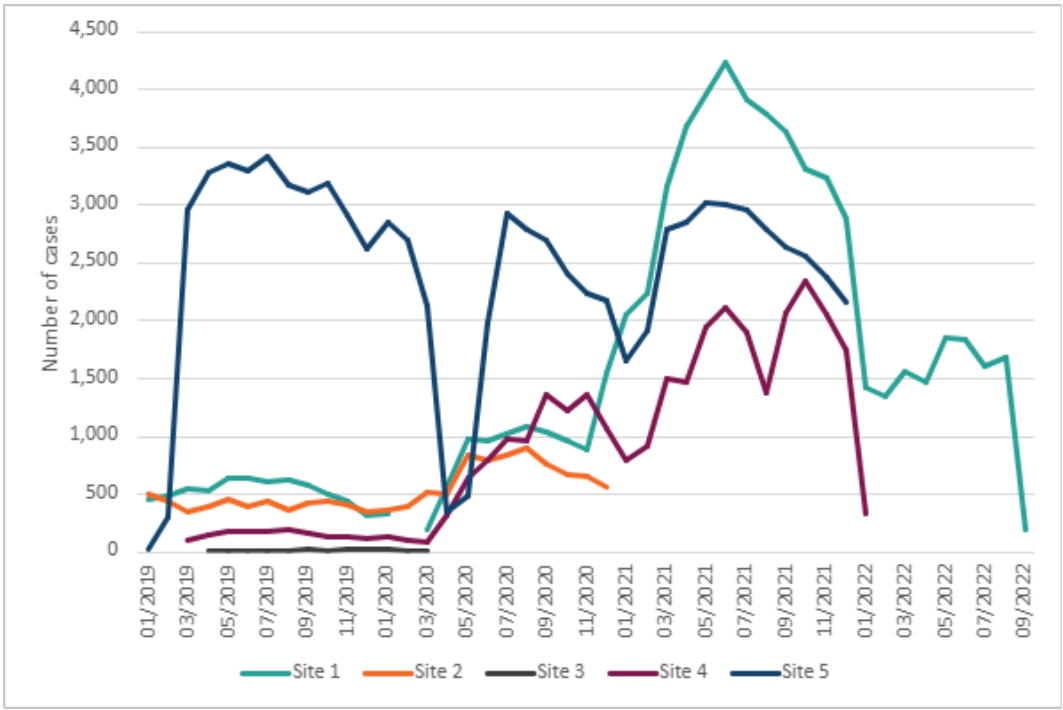


Figure 2: Number of complaints per month by hospital site



Figures 1 and 2 show the number of cases included in the data by month and across different sites. The overall number of cases each month is primarily driven by the sites that data was collected from each month.

Complaints and injury mechanisms

Figure 3: Top 15 complaints

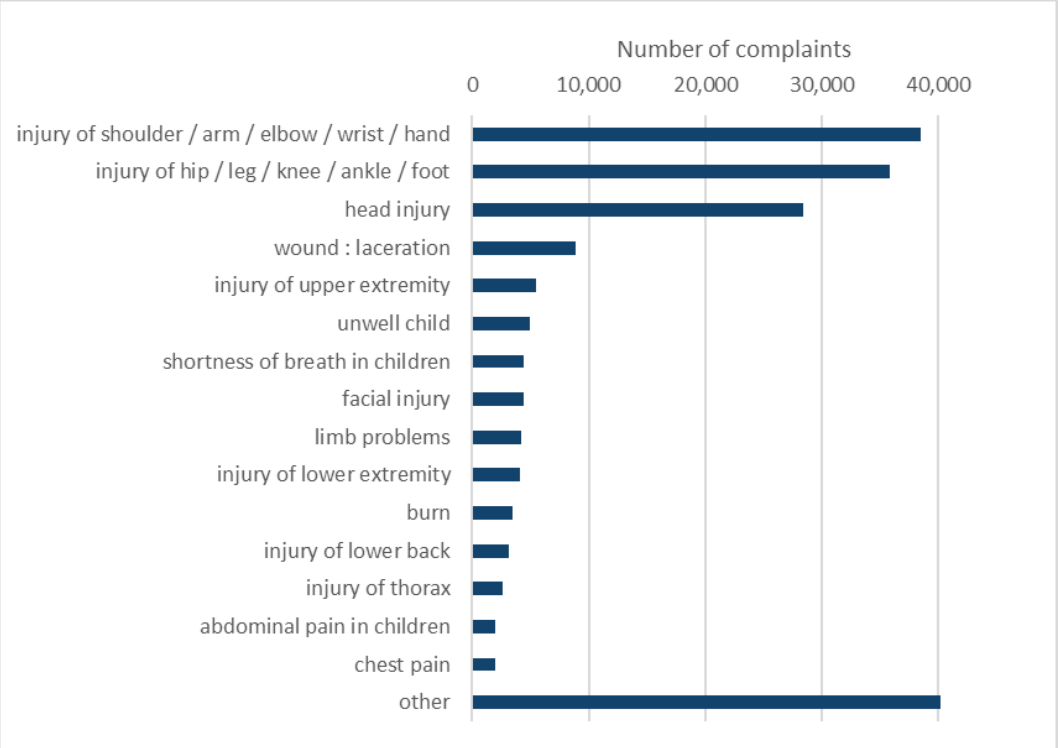
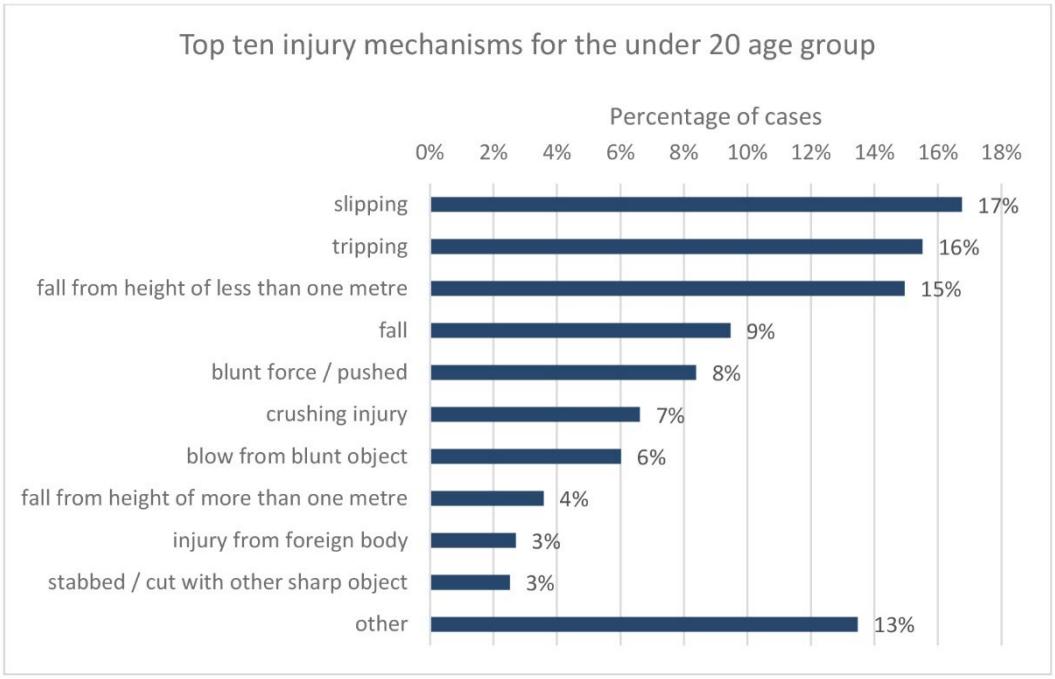
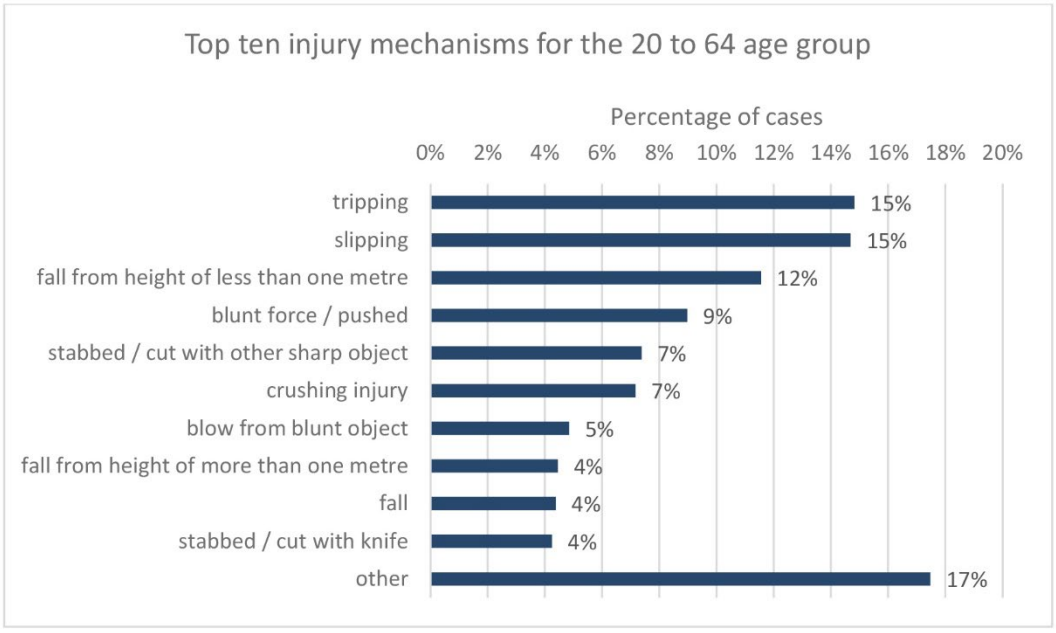
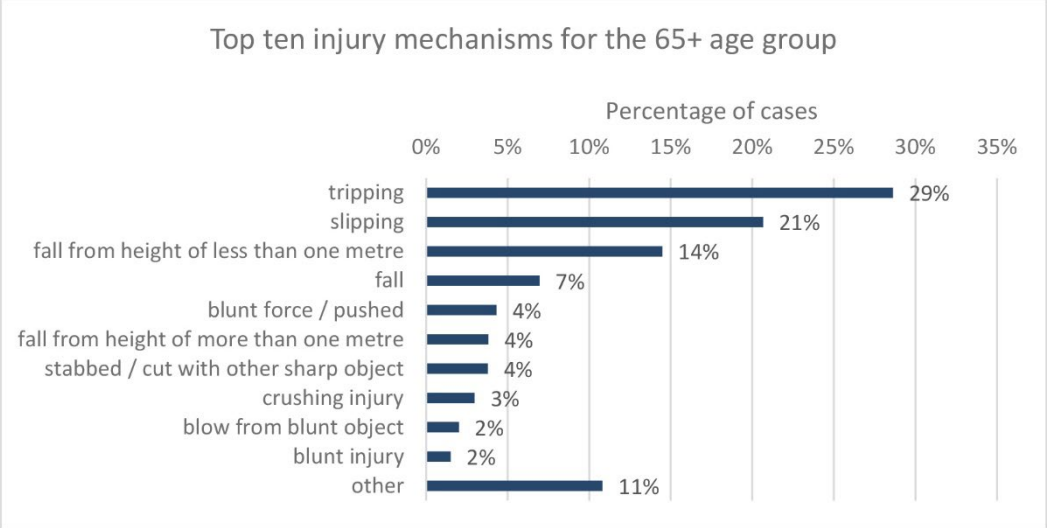


Figure 3 shows the top fifteen complaints, which together make up around 80% of all available values. Figure 4 shows the top ten injury mechanisms for the different age groups. Slipping, tripping and falling were the top injury mechanisms for all age groups.

Figure 4: Injury mechanism across age groups



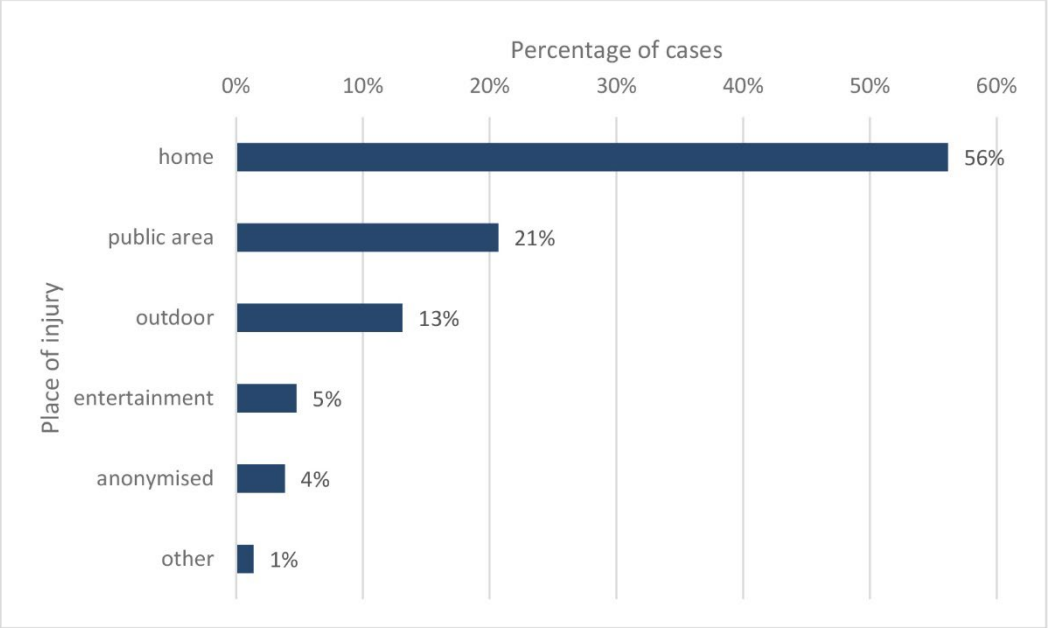




Note: the charts in Figure 4 show percentages based on the total number of cases where the injury mechanism was complete for each age group. It excludes cases where injury mechanism was blank.

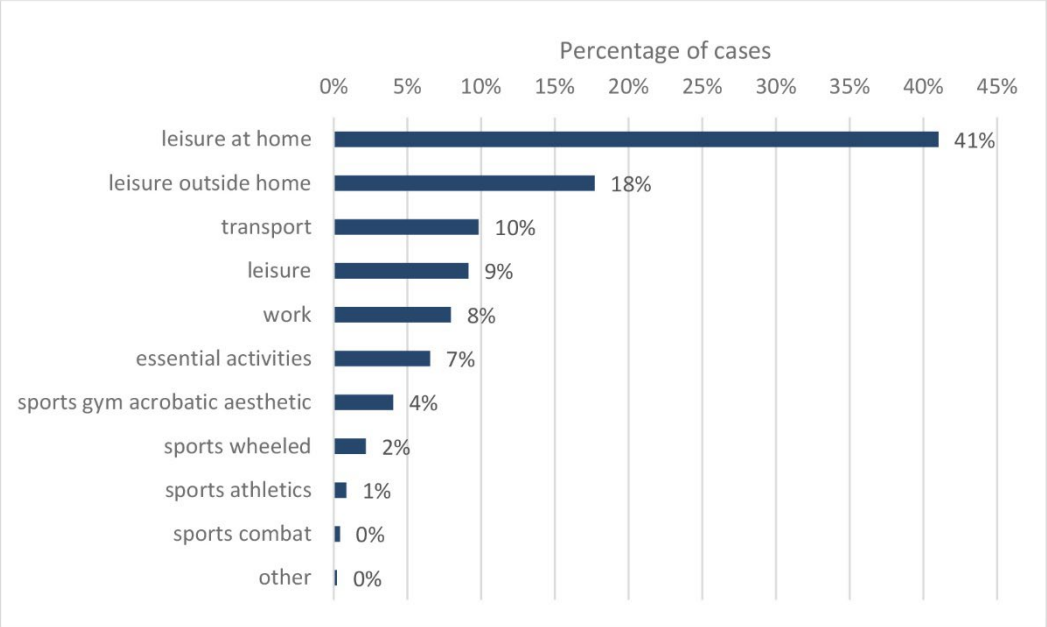
Figures 5 and 6 show that the most common place of injury was the home, and 'leisure at home' was the most common activity during which the injury occurred.

Figure 5: Top five injury places in the data set



Note: Figure 5 shows percentages based on the total number of cases where the place of injury was complete for each age group. It excludes cases where place of injury was blank.

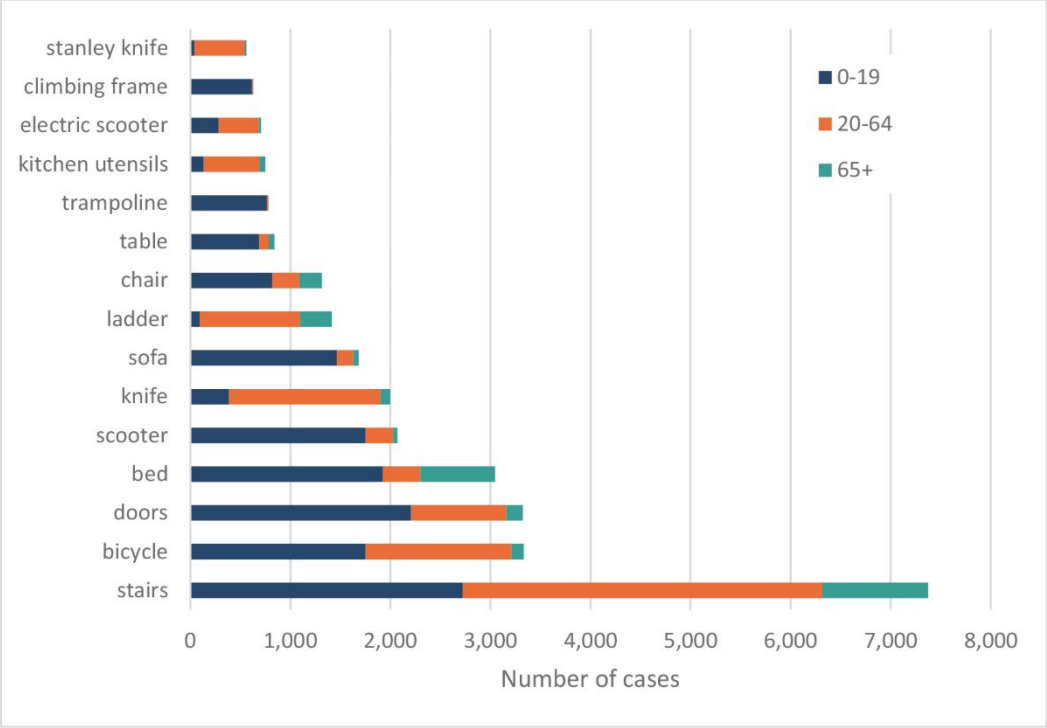
Figure 6: Top 10 activities involved during injuries



Note: Figure 6 shows percentages based on the total number of cases where the activity data was complete for each age group. It excludes cases where activity was blank.

Products and injuries

Figure 7: Age group distribution across the top 15 product categories

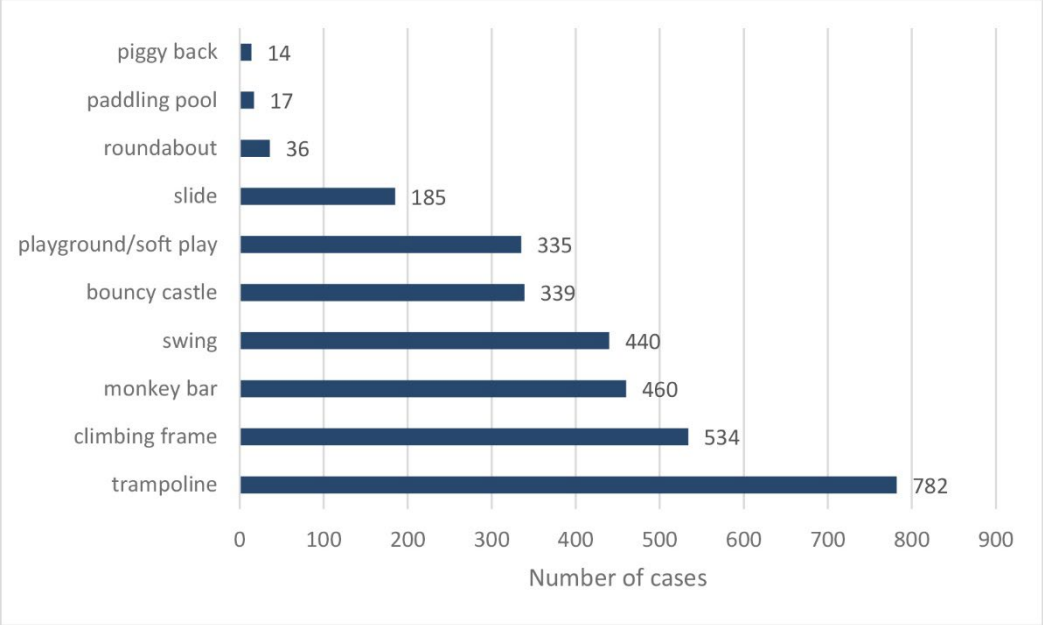


More than 75% of values were missing for product category. For those records for which product category was populated, there were 832 unique values. The top 15 categories shown in Figure 7 represent 65% of all available values. Over 7,000 cases involved stairs, and are associated with injuries such as falling and tripping.

The figures below show data on some specific product types we investigated, to illustrate the sorts of insights that could be obtained from the dataset. These are intended to be illustrative, and do not cover the entirety of what was explored during OPSS's analysis of the dataset.

Play

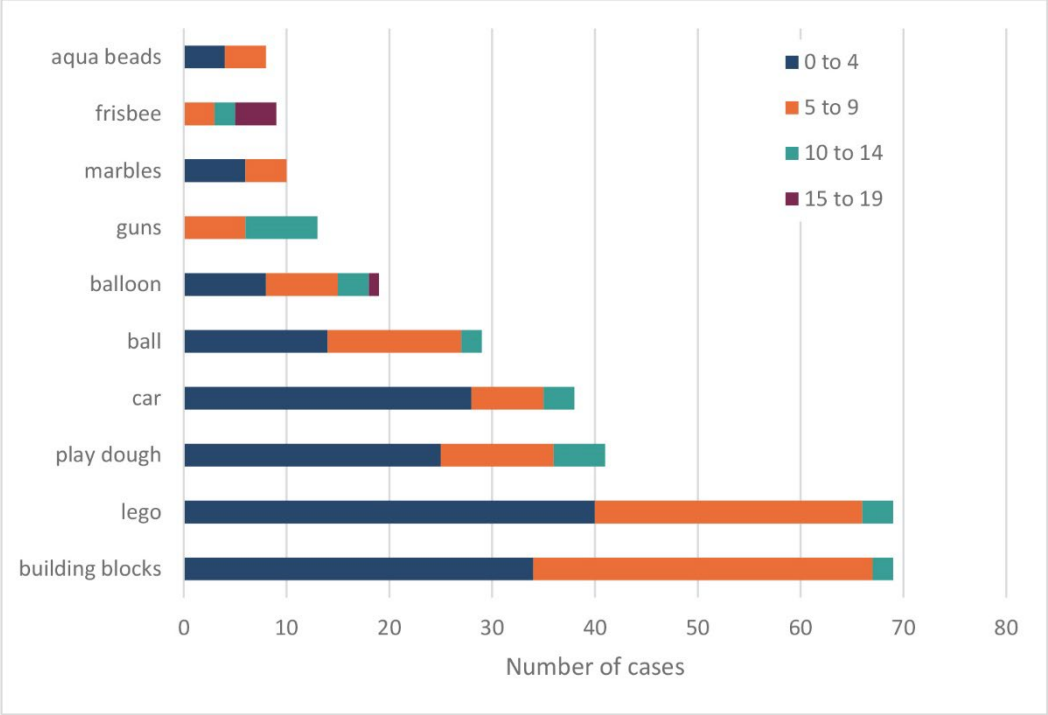
Figure 8: Top 10 product categories under the 'play' grouping



There were around 3,200 cases in the 'Play' product category grouping. The specific product category is missing in 40% of cases. The largest product category in "Play" is trampoline. Other cases largely refer to playground equipment such as climbing frames, monkey bars, swings, slides etc.

Toys

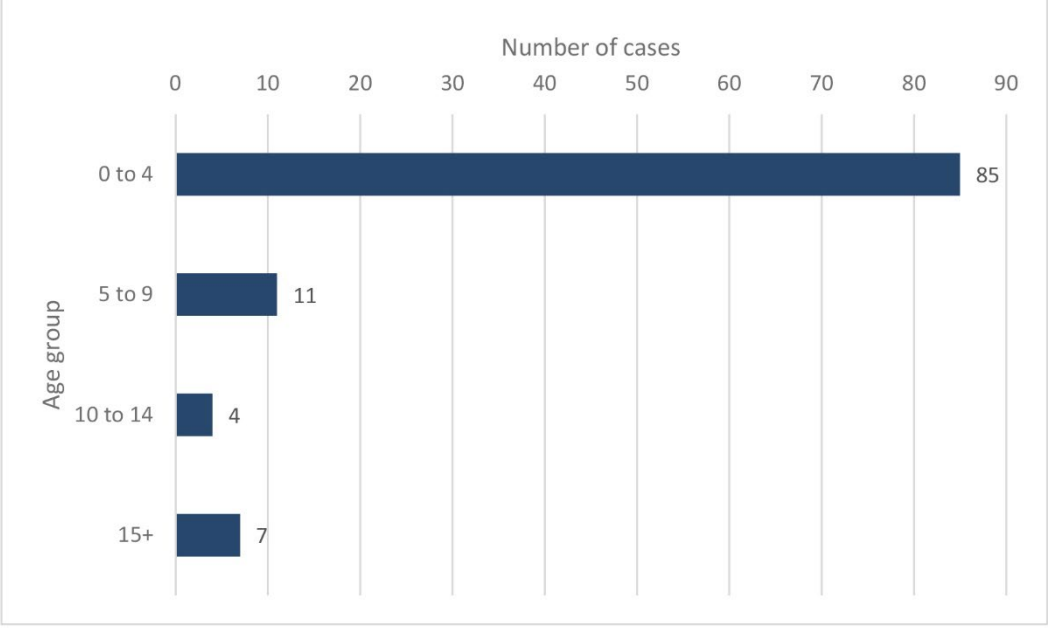
Figure 9: Age group distribution for the top 10 product categories under the toys grouping



There are 616 cases in the 'Toys' grouping, with 55 different specific product categories. Out of those, about a third do not have a product category recorded.

Button batteries

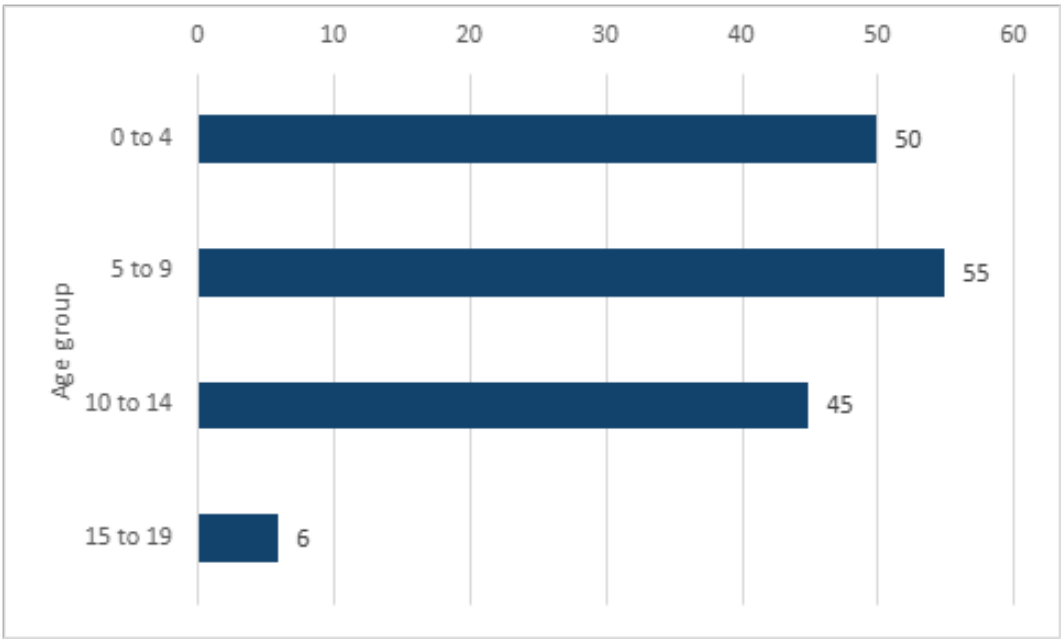
Figure 10: Age group distribution for the button batteries category



There were 107 records related to button batteries. The majority of cases were related to ingestion or inhalation of foreign bodies. Under 5s were the main age group affected. Data on mechanism, activity type and injury place in this category was largely missing.

Magnets

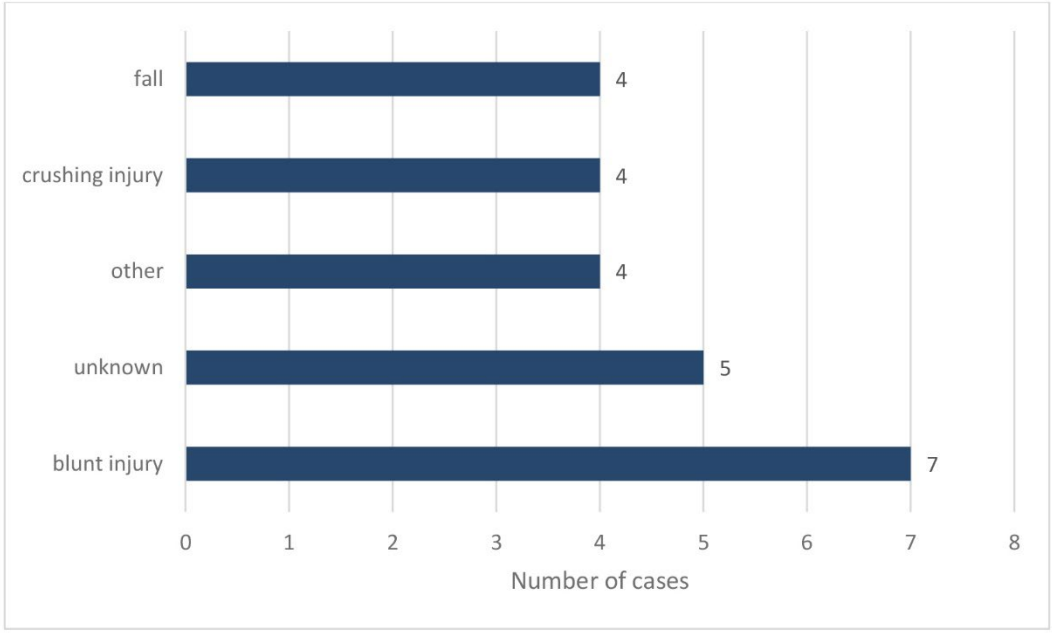
Figure 11: Age groups distribution and injury type under the magnets category



There were 156 cases involving magnets. The majority of cases were related to ingestion of foreign bodies, with a more varied age distribution than was seen for button batteries. Data on mechanism, activity type and injury place in this category was largely missing.

Musical instruments

Figure 12: Injury mechanism under the musical instruments category



Wave 6 of OPSS's Consumer Research survey showed that of respondents who had bought a product in the last 6 months, those who had bought a musical instrument were most likely to have experienced a safety issue. This prompted us to explore musical instrument injuries in this dataset. We identified 24 injuries from musical instruments. Manual review of the free text clinical notes suggested that these were largely accidental injuries with no particular indication of product safety concerns (e.g. someone walking into a piano).

© Crown copyright 2026

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated.

To view this licence, visit www.nationalarchives.gov.uk/doc/open-governmentlicence/version/3/ or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gsi.gov.uk. Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned.

Contact us if you have any enquiries about this publication, including requests for alternative formats, at: OPSS.enquiries@businessandtrade.gov.uk

Office for Product Safety and Standards

Department for Business and Trade,
4th Floor, Multistory, 18 The Priory Queensway, Birmingham B4 6BS
https://www.gov.uk/government/organisations/office-for-product-safety-and-standards