

Dadansoddi ar gyfer Polisi



Analysis for Policy



Llywodraeth Cymru
Welsh Government

Social research number: 59/2026

Publication date: 25/03/2026

Children (Abolition of Defence of Reasonable Punishment) (Wales) Act 2020: estimating police recorded cases of physical punishment against children

Mae'r ddogfen yma hefyd ar gael yn Gymraeg.

This document is also available in Welsh.

© Crown Copyright Digital ISBN: 978-1-83745-248-4

Children (Abolition of Defence of Reasonable Punishment)
(Wales) Act 2020: estimating police recorded cases of
physical punishment against children

Authors: Ryan Nicholls, Daniel Boon, Chloe Whiteley,
Hannah Davies

Full Research Report: Nicholls, R., Boon, D., Whiteley, C., Davies, H.
(2025). Children (Abolition of Defence of Reasonable Punishment)
(Wales) Act 2020: estimating police recorded cases of physical
punishment against children. Cardiff: Welsh Government, GSR report
number 59/2026.

Available at: <https://www.gov.wales/estimating-police-recorded-cases-of-physical-punishment-against-children>

Views expressed in this report are those of the researchers and not
necessarily those of the Welsh Government

For further information please contact:

CFC Evidence and Research team

Equality, Poverty and Children's Evidence and Support Division

Welsh Government

Cathays Park

Cardiff

CF10 3NQ

Email: research.childrenandfamilies@gov.wales

Table of contents

List of tables.....	2
List of figures.....	2
Glossary.....	3
1. Introduction and background.....	4
1.1. Background and context.....	4
1.2. About the model developed.....	4
1.3. About this report.....	5
2. Methodology.....	6
2.1. Data.....	6
2.2. Data preparation.....	6
2.3. Model.....	7
2.4. Classification.....	7
2.5. Limitations.....	8
3. Findings.....	10
3.1. Pre-implementation of the Act.....	10
3.2. Post-implementation of the Act.....	11
4. Conclusions.....	13
References.....	14
Annex A.....	15

List of tables

Table 1: Estimated cases of physical punishment and other assault against children, March 2019 to March 2022 [Note 1]..... 10

Table 2: Estimated cases of physical punishment and other assault against children, March 2022 to June 2025 [Note 1]..... 11

List of figures

Figure 1: Estimated cases of physical punishment and other assault against children, 2019 to 2025 [Note 1]..... 12

Glossary

The Act

Refers to the Children (Abolition of Defence of Reasonable Punishment) (Wales) Act 2020, which makes it illegal for anyone to physically punish a child in Wales by removing the legal defence of “reasonable punishment”.

Data Collection and Monitoring Task and Finish Group (DCM)

The group considered methods for collecting data across organisations to monitor the impacts of the Children (Abolition of Defence of Reasonable Punishment) (Wales) Act 2020 on public services and other organisations. Membership was from a range of organisations with a key interest in monitoring the impact on public services and included representatives from:

Crown Prosecution Services (CPS), Social Services, National Health Service (NHS) Wales, Police forces in Wales, Her Majesty’s Courts and Tribunal Service, Regional Safeguarding boards, Welsh Local Government Association, CAFCASS Cymru, Directors of Education and Welsh Government.

UK Secure e-Research Platform (UK SeRP)

Secure eResearch Platform is the complete customisable solution for Trusted Research Environments: data sharing, linkage and analysis in a safe, secure and controlled environment that’s accredited to the highest international standard. It is air-gapped for security purposes, meaning that it does not have an active internet connection and can only be accessed via a virtual desktop.

Modus operandi (MO)

The description provided for each case in the dataset. This is an open text field which is completed by the police officer recording the crime.

Large language models (LLMs)

Advanced AI systems trained on massive text datasets to understand, generate, and reason with human language. Examples of these include ChatGPT and Microsoft Copilot. Transformer neural networks, one of which is used to conduct the below analysis, form the basis of large language models.

Natural language processing (NLP)

A field of artificial intelligence that enables computers to analyse, understand, and extract meaning from human language in text or speech. LLMs are a type of NLP model that use very large datasets and neural networks to understand and generate human language.

1. Introduction and background

1.1. Background and context

On 20 March 2020 the Children (Abolition of Defence of Reasonable Punishment) (Wales) Act 2020 (hereafter referred to as 'the Act') received Royal Assent. The Act came into force on 21 March 2022.

The overarching objective of the Act is to help protect children's rights by prohibiting the use of physical punishment against children, through the removal of the defence of reasonable punishment. This means that the defence is no longer available within the territory of Wales to parents or those acting in loco parentis (acting with parental responsibility), as a defence to a charge of common assault and battery against a child in their care.

The Welsh Government undertook an extensive multi-media public awareness campaign before the legislation came into force to ensure as many people as possible were aware of the change in the law. This began in June 2021 with digital advertising and television campaigns launched fully in September 2021 and a peak in communications activity between January and March 2022. Welsh Government has continued to raise awareness of the law and provide information, advice and support for parents about positive alternatives to physical punishment.

Under the terms of the Act, Welsh Ministers are required to prepare and publish reports on the effects of the legislative changes both 3 years and 5 years after implementation. The focus of the post implementation review was agreed between Welsh Ministers and the Children, Young People and Education Committee during the passage of the legislation. It centres on the following key areas:

- impact on public services
- awareness of the law
- attitudes to physical punishment
- stakeholder views and experiences

The 3-year post implementation review ([Children \(Abolition of Defence of Reasonable Punishment\) \(Wales\) Act 2020: evidence and data synthesis](#)) was published on 16 December 2025. The synthesis draws on a diverse range of data and research sources, including data from Police Forces in Wales, where available.

1.2. About the model developed

The Welsh Government convened the Data Collection and Monitoring Task and Finish Group (DCM) to develop methods of monitoring the impact of the change in legislation, including on the Welsh Police Forces.

The DCM determined that a standardised approach to identifying cases of physical punishment, such as adding flags to cases of interest, would not be possible because the

Act did not create a new offence. Operating under Home Office counting rules ([Home Office, 2024](#)), prevents the adding of fields to distinguish cases of physical punishment from other forms of assault.

Therefore, it was not possible to identify cases that resulted directly from the Act or to draw on any historical information about when the former “reasonable punishment” defence could have been used. It was therefore, agreed the best way to estimate cases of interest was for Welsh Government analysts to design an algorithm to review case files and estimate which incidents were likely to involve physical punishment.

Due to the limitations of police information systems, any proposed method had to be one that could be applied retrospectively to the data already collected by the police in Wales. This was largely limited to the reports written by officers on police crime management systems, and the contextual information (see Annex A) about the victim and offenders held alongside it. From this information it was determined that some form of text classification would be the most appropriate method of counting cases of interest to the post-implementation reviews of the Act.

In September 2022 the Welsh Government sponsored a member of the Government Statistical Service to undertake a Master’s (MSc) in Data Analytics for Government ([MDataGov](#))^[footnote 1] to design a new automatic approach to classifying the police force data. The conclusion of this work was that a pre-trained large language model (LLM) outperformed a keyword search model used on the same data, achieving an accuracy of 88% when predicting cases compared to 77% for the keyword search model. This demonstrated it would be a better method for identifying cases of physical punishment from police force data, and so could strengthen the approach for monitoring trends and supporting post-implementation reviews of the Act at 3 and 5 years.

This pre-trained language processing model was used by Welsh Government analysts to produce the results detailed in this report.

1.3. About this report

This report supplements the high-level findings published in the 3-year post-implementation review of the Act ([Welsh Government, 2025](#)). The aim of this report is to provide detailed results from the model and provide a high-level description of the methodology used.

Footnote

[1] The Master’s (MSc) in Data Analytics for Government (MDataGov) is a postgraduate programme tailored for public sector organisations. It is a joint initiative between the Office for National Statistics (ONS) Data Science Campus and academic partners across the UK, available to all public sector employees.

2. Methodology

2.1. Data

The data was stored on the UK Secure e-Research Platform (UK SeRP), a secure platform with restricted internet access, due to its sensitive nature. The data consisted of information that police officers record upon receiving reports of a crime, including a description of the crime that was reported, and other contextual information (see Annex A for more information).

As physical punishment of a child is not a distinct classification of crime, the data submitted by the police forces included only cases recorded as common assaults ([CPS, 2025](#)), perpetrated by an adult against a child (person aged under 18), where there was no known victim-offender relationship that would exclude the crime from being considered physical punishment by a parent or those acting in loco parentis.

The main component of the data was an open text field which recorded the modus operandi (MO) of the crime. This was a short description of the crime, inputted by the police officer, and included any details that could be useful for police officers to conduct further investigation. This MO text must always be entered into the text field by the reporting officer for each crime, however there is no minimum or maximum character word limit.

Data was submitted by police forces in Wales to Welsh Government via a secure data file transfer system to a limited number of analysts working on the model, and immediately transferred to the UK SeRP. Three of the four police forces in Wales submitted data pre- and post-implementation of the Act, covering the period April 2019 to June 2025.

A total of 13,860 cases were included in the full dataset submitted to the Welsh Government, of which 6,420 were recorded by North Wales Police, 3,875 by Dyfed Powys Police, and 3,565 by Gwent Police. Data are rounded to the nearest 5. Totals and percentages are calculated with the raw numbers and may appear not to match. This data is not publicly available.

2.2. Data preparation

The dataset was originally created for the purposes of crime management and was not intended to be used for text classification. Much of the recorded text was inputted informally and in many cases spelling, grammar, punctuation, and use of upper and lower case were used either incorrectly or inconsistently. In addition, some personal information featured in the data entries, and best efforts were made to remove this for data protection purposes.

Upon receipt of the data to Welsh Government, the dataset passed through Microsoft Presidio ([Microsoft, 2022](#)), a customisable data anonymisation algorithm used to remove details such as names, addresses, phone numbers and anything else which could enable individuals to be identified. This was done before Welsh Government analysts working on the data undertook any further actions with the data.

A combination of built-in and custom personal information identifiers were used to remove non-standard types of personal identifiable information. Instances of identifiable information were replaced by placeholders; for example, a name would be replaced by “<PERSON>”. Manual spot checks were then completed by Welsh Government analysts to ensure any remaining personal identifiable information was removed.

The data was further transformed in the following ways to maintain consistency, for ease of interpreting analysis, and to make it easier to train the model:

- All text was transformed to lower case. This removed issues where cases were recorded entirely in capitals or where capitalisation was used incorrectly
- The most common acronyms were replaced and written out in full. Cases often contained acronyms rather than whole terms. For example, the police reported that ‘oel’ and ‘ssd’ were commonly used, instead of ‘officer enquiry log’ and ‘social services department’ respectively
- Punctuation, including special characters, were removed from the text. Cases were often inputted informally, and use of punctuation was inconsistent and often incorrect

2.3. Model

Due to where the data was stored, importing and setting up various models to test was difficult. As such, the type of model was chosen according to research literature at the time ([Vaswani et al. 2023](#)), which indicated that transformer neural networks were consistently the highest performing text classification tools. Due to limited resources, and the resource-intensive nature of transformers, the choice of available model was somewhat limited.

In consideration of the available resources, the model chosen for this analysis was ‘distilbert-base-uncased’. DistilBERT is a smaller and quicker approximation of a larger, more commonly used model, called BERT. The model is an early version of part of the base model used in applications such as ChatGPT and Copilot, trained in 2019. The standard version of DistilBERT was used, which is a pre-trained model that can interpret text data, and does not take account of whether characters are upper or lower case. This was useful in this context as the data often used upper and lower case in non-standard ways, such as cases being recorded entirely in capital letters. Other available models were too large to save or run on the instance of the UK SeRP that was used.

Two cycles of the model have been run to date. The first, as outlined in the introduction, was run as part of the data science master’s project sponsored by Welsh Government, and served as a proof-of-concept, then established that the DistilBERT model was suitable for classifying police-recorded cases of physical punishment. The final model, used to produce the findings in this report, was refined based on the outcome of the first.

2.4. Classification

A definition for physical punishment of a child was required to train the model and undertake the analysis. The definitions below were developed by analytical officials for the purposes of this model only, and are not a formal Welsh Government policy definition of what physical punishment of a child constitutes in Wales. Cases are inherently difficult to compare and

categorise due to the unique circumstances of each incident, so best judgement of the analysts involved was used when classifying cases.

There were 3 definitions agreed upon and cases in the dataset were assigned into one of the 3 categories below:

- Lower level physical punishment – included cases of low-level physical punishment, committed against a child by someone who could reasonably be acting in loco parentis. This is similar to [New Zealand's \(2007\)](#) definition of "minor acts of physical discipline", and includes lighter forms of punishment, such as a slap on the wrist or hitting on the arm which would only leave a temporary reddening of the skin or no mark at all at the time the crime was reported
- Higher level physical punishment – included cases of high-level or more severe physical punishment, committed against a child by someone who could reasonably be acting in loco parentis. This is similar to [New Zealand's \(2007\)](#) definition of "smacking" and includes more severe cases of punishment, such as punching, hands on neck, or hair pulling, which would either leave a temporary reddening of the skin or no mark at all at the time the crime was reported
- Other assault – included any cases not of interest with regards to the Act. This included any cases where the perpetrator was not someone acting in loco parentis, or there was not enough evidence in the MO to determine whether a case was related to physical punishment or not

Of the full dataset, 15% of cases (2,080) were randomly selected and classified manually, using the above definitions, by Welsh Government analysts. This data was then split randomly into a training dataset, validation dataset and test dataset, each containing 70%, 15% and 15% respectively. The training dataset was used to help the model determine the types of cases that fit into each of the 3 categories. The accuracy of the model was tested against the validation dataset, by measuring how many of the manually classified cases it accurately predicted.

The model was fine-tuned by changing its parameters, which included how fast it learned (learning rate), how much data it processed each time (batch size), how many rounds of training it performed (number of epochs), and a parameter to reduce the risk of overfitting by limiting how strongly the model can respond to individual details in the training data (weight decay). Each combination of parameters gave slightly different results against the validation dataset. Several combinations of settings were applied until the final model was chosen.

Once the final model was selected, it was run on the test dataset. This dataset contained the remaining 15% of the labelled cases that the model had not seen during its training. It correctly labelled 78% of cases in the test dataset, and therefore it is assumed that it has the same level of accuracy on the unseen, unlabelled data.

2.5. Limitations

The main limitations relate to the raw data provided, and by the platform to which the research was confined.

The model was not pre-trained to be accustomed to text in the original form that it was received. Although the data was transformed to mitigate this risk, not all inconsistencies would have been accounted for.

The manually classified cases were selected at random, which reduces the risk of systematic bias, but does not guarantee an even distribution of cases across time periods or police forces. It is assumed that the way in which the MO text is recorded has remained consistent over time and across forces; however, cases are inherently difficult to compare due to the unique circumstances of each incident.

The model used in this research was introduced in 2019 and represents an earlier generation of transformer-based language models. The field of natural language processing has advanced rapidly in recent years ([Naveed et al., 2025](#)), with newer large language models offering significantly improved reasoning ability and contextual understanding. A newer model would likely have performed better. The instance of UK SeRP used here, whilst vital in keeping data secure, limited the choice of model and available packages, due to the processing and accessibility limits of the version of the platform which the Welsh Government could access. A larger, more powerful model may have performed even better than the model implemented in this analysis. The ability to install software packages and download models without limitation would have enabled a deeper testing of different models and implementations. Alternative models and access to more computing power on UK SeRP is currently being explored by Welsh Government analysts.

Finally, data was not received from all police forces in Wales. A more complete dataset would enable the model to be trained to understand more deeply the context of physical punishment cases across the entirety of Wales.

3. Findings

All the figures below are estimates only. No models are 100% accurate, and some degree of misclassification is inevitable even in well-tested models.

A total of 13,860 police case files were received from 3 police forces in Wales. Data are rounded to the nearest 5. Totals and percentages are calculated with the raw numbers and may appear not to match or sum to 100%.

In the 3-year post-implementation review of the Act ([Welsh Government, 2025](#)), figures were published as a total of cases of physical punishment. The below findings provide a full breakdown of those figures by classification and year.

3.1. Pre-implementation of the Act

The dataset contained 5,395 cases of common assault for the period 21 March 2019 to 20 March 2022. Of these, 960 (18%) were estimated to be higher level physical punishment, 2,910 (54%) as lower level physical punishment, and 1,525 (28%) as other assault.

The model estimated that cases of higher and lower level physical punishment decreased slightly during the second year of this period (April 2020 to March 2021), but then increased during the period April 2021 to March 2022, especially for cases of lower level physical punishment.

Table 1: Estimated cases of physical punishment and other assault against children, March 2019 to March 2022 [Note 1]

Classification of case	March 2019 to March 2020	April 2020 to March 2021	April 2021 to March 2022	Total
Lower level physical punishment	885	865	1,160	2,910
Higher level physical punishment	350	230	375	960
Other assault	450	450	625	1,525
All cases	1,685	1,545	2,160	5,395

Source: 5,395 police case files received from 3 police forces in Wales for the period March 2019 to March 2022

[Note 1] Data are rounded to the nearest 5

3.2. Post-implementation of the Act

The dataset contained 8,465 cases of common assault for the period 21 March 2022 to 8 June 2025, an increase of 3,070 (57%) from the period relating to pre-implementation of the Act. Of these, 1,210 (14%) were estimated to be higher level physical punishment, 4,970 (59%) as lower level physical punishment, and 2,285 (27%) as other assault.

During this post-implementation period, the model estimated that cases of higher level physical punishment steadily decreased after the first year, while lower level physical punishment cases increased slightly during the second year of the post-implementation period, but experienced a decrease in the most recent year of data.

Table 2: Estimated cases of physical punishment and other assault against children, March 2022 to June 2025 [Note 1]

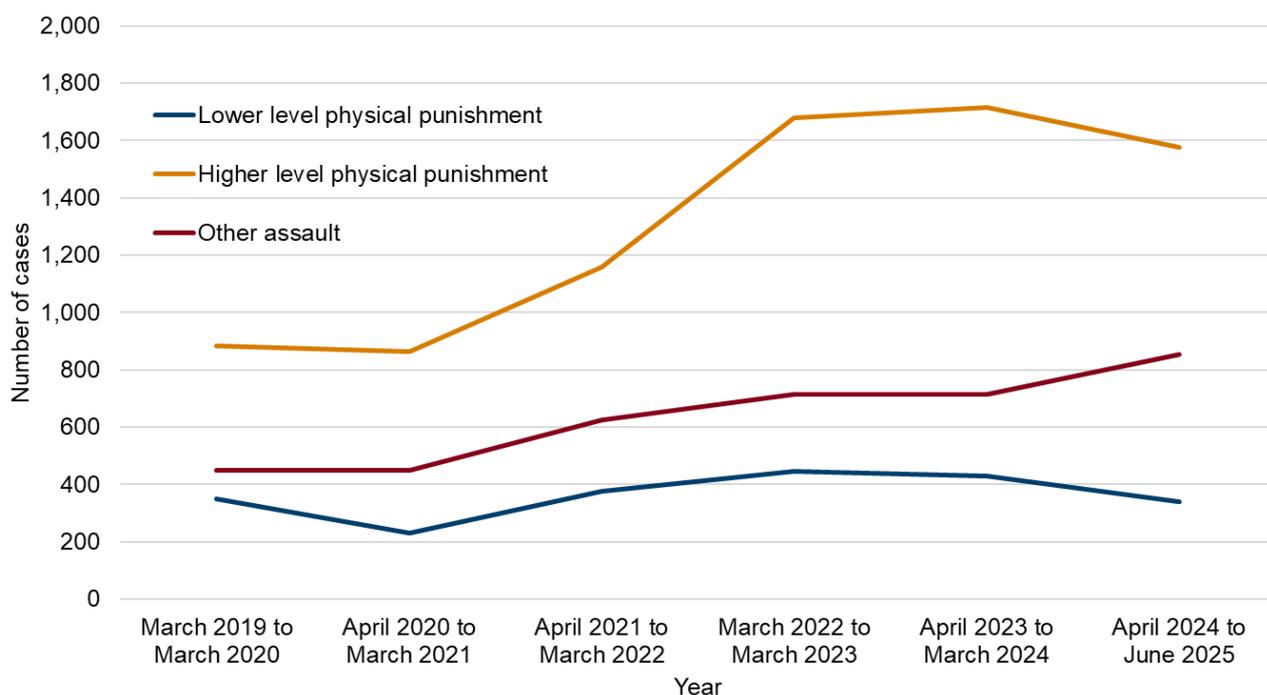
Classification of case	March 2022 to March 2023	April 2023 to March 2024	April 2024 to June 2025	Total
Lower level physical punishment	1,680	1,715	1,575	4,970
Higher level physical punishment	445	430	340	1,211
Other assault	715	715	855	2,285
All cases	2,840	2,860	2,765	8,465

Source: 8,465 police case files received from 3 police forces in Wales for the period March 2022 to June 2025

[Note 1] Data are rounded to the nearest 5

This is an increase of 250 (26%) cases of higher level physical punishment and 2,060 (71%) cases of lower level physical punishment compared to the estimated cases relating to the period pre-implementation of the Act. Cases of lower level physical punishment accounted for 67% (3,070 cases) of the increase in total common assault cases before and after the implementation of the Act.

Figure 1: Estimated cases of physical punishment and other assault against children, March 2019 to June 2025 [Note 1]



Description of figure 1: this line chart shows that, since the period March 2019 to March 2020, the number of lower level physical punishment cases is estimated to have increase from 885 to 1,715 by the period April 2023 to March 2024, before decreasing slightly for the period April 2024 to June 2025. Cases of other assault are estimated to have increased steadily over the same period from 450 to 855, while higher level physical punishment cases are estimated to have fluctuated between 200 and 450.

Source: 13,860 police case files received from 3 police forces in Wales for the period March 2019 to June 2025

[Note 1] Data are rounded to the nearest 5

4. Conclusions

The methods used to identify and categorise cases of physical punishment against children in this report (LLMs) have been shown to improve upon previous keyword search methods. This is because LLMs can use the full narrative of each record to identify relevant cases, apply consistent classification over time, and distinguish physical punishment from other forms of assault. This means that LLMs can produce more robust estimates and therefore provide a stronger and more reliable analysis.

The model estimated that more cases of physical punishment against children were reported in the 3 years post implementation of the Act than in the 3 years before the Act came into force. During this time there was an extensive multi-media public awareness campaign undertaken by the Welsh Government. This was especially noticeable for estimated cases of lower level physical punishment, increasing by over 2,000 post implementation of the Act compared to the 3 years prior. This increase in estimated cases of lower level physical punishment was the main reason for the total increase in common assault cases between the 2 periods, while estimated cases of higher level physical punishment remained similar over time.

To support the Welsh Government's 5-year post-implementation review of the Children (Abolition of Defence of Reasonable Punishment) (Wales) Act 2020, it is anticipated that this model will be developed further and the analysis undertaken again. Future plans for estimating police recorded cases of physical punishment against children are focused on 2 areas: firstly, coverage can be improved by obtaining a full dataset from all police forces in Wales; a larger and more varied training set is expected to improve generalisability and reduce uncertainty in the estimates. Accuracy in identifying and assigning a categorisation of cases of physical punishment against children for the purposes of producing the estimates will also be further improved. This focus will be supported by exploring access to instances of the UK SeRP with increased computational power, as this would allow more powerful and comprehensive, and therefore accurate, models to be tested.

References

Children (Abolition of Defence of Reasonable Punishment) (Wales) Act 2020, '[Children \(Abolition of Defence of Reasonable Punishment\) \(Wales\) Act 2020](#)', (Accessed: 25/02/2026)

Crown Prosecution Service (CPS) (2025), '[Offences against the Person, incorporating the Charging Standard | The Crown Prosecution Service](#)', (Accessed: 25/02/2026)

Home Office (2024), '[Crime Recording Rules for Front line Officers and Staff](#)', UK Home Office

Microsoft (2022), '[Presidio: Data Protection and de-identification SDK, Microsoft Presidio](#)', (Accessed 25/02/2026)

Naveed H, Khan A, Qiu S, Saqib M, Anwar S, Usman M, Akhtar N, Barnes N, Mian A (2025), '[A Comprehensive Overview of Large Language Models](#)', ACM Trans. Intell. Syst. Technol, Volume 16, Issue 5, Article 106 (October 2025)

New Zealand Police (2007), '[Three Month Review: Background and Methodology | New Zealand Police](#)', (Accessed: 25/02/2026)

Stripe N (2021), '[Police recorded offences involving knives or sharp instruments: methodology changes - Office for National Statistics](#)', (Accessed: 25/02/2026)

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, Kaiser L, Polosukhin I (2017), '[Attention Is All You Need](#)', (Accessed: 25/02/2026)

Welsh Government (2020), '[Data Collection and Monitoring Task and Finish Group meeting 26 February 2020](#)', Welsh Government

Welsh Government (2025), '[Children \(Abolition of Defence of Reasonable Punishment\) \(Wales\) Act 2020: synthesis report | GOV.WALES](#)', (Accessed: 25/02/2026)

Annex A

Contextual information provided in the dataset about the victim and offenders

The data consisted of the following variables:

- MO_text – the description of the crime
- RecordDate - the date the crime was recorded
- CommitDate - the date the crime was committed
- VOR - the relationship between the victim and offender
- OutcomeCode - the outcome of the investigation as per the Home Office Counting Rules crime classification (Home Office, 2024)
- CA - A flag indicating child abuse
- CC - A flag indicating child concern
- CP - A flag indicating a child protection order