

Data Science to analyse responses to the National Assembly for Wales Children, Young People and Education Committee's consultation on the Children (Abolition of Defence of Reasonable Punishment) (Wales) Bill

June 2019

Dr Chaitanya Joshi

Data Science Campus (ONS)

Chaitanya.Joshi@ons.gov.uk

<https://datasciencecampus.ons.gov.uk/>

Contents

Data Science to analyse responses to the National Assembly for Wales Children, Young People and Education Committee’s consultation on the Children (Abolition of Defence of Reasonable Punishment) (Wales) Bill	1
Executive Summary:	3
A. Overall responses, in support or otherwise for the Bill:	7
B. Responses categorised in support or otherwise for the Bill:	9
C. Responses categorised by the category of respondent (UserType):	11
D. Individual responses categorised according to region (outside/inside Wales):.....	13
E: Most frequent co-occurring phrases where respondents replied that they do not support the Bill (390 responses):	15
F: Most frequent co-occurring phrases where respondents replied that they do support the Bill (238 responses):	17
G: Most important phrases where respondents replied that they do not support the Bill (390 responses):	19
H: Most important phrases where respondents replied that they do support the Bill (238 responses):.....	21
Annex A:	23
Structure of the dataset: cleaning and anonymisation.....	23
Dataset: initial grouping in various categories:	24
Annex B:.....	25
Dataset: most frequent (co-occurring) and important phrases	25
Most frequent (co-occurring) phrases:	25
Most important phrases:	26

Executive Summary:

- The ONS Data Science Campus used Data Science techniques to analyse free text responses to a consultation gathering opinions on a recent Welsh Government Bill - the Children (Abolition of Defence of Reasonable Punishment) (Wales) Bill (“the Bill”) - introduced by [Julie Morgan AM](#), Deputy Minister for Health and Social Services. The Assembly’s [Business Committee](#) had remitted the Bill to the [Children, Young People and Education Committee](#).
- The Welsh Government’s stated purpose for the Bill is to abolish the common law defence of reasonable punishment so it is no longer available in Wales to parents or those acting in loco parentis as a defence to assault or battery against a child. The Welsh Government states that the Bill is intended to support children’s rights by prohibiting the use of physical punishment, through removal of this defence.
- The [Children, Young People and Education Committee](#) ran an [online consultation](#) between 2 April 2019 and Tuesday, 14 May 2019 to gather opinions on the general principles of the Bill.
- Over the course of this six-week consultation, **650** responses were received. Of the 650 responses, nearly two thirds (**60.0%-390**) do not support the Bill while over one third (**36.6%-238**) of the respondents support the Bill. The remaining respondents either indicate partial support for the Bill (**3.1%; 20**) or state that they do not have a view (**0.3%; 2**).
- The respondents were able to respond in one of three different capacities (described in this report as “UserTypes”):
 - individual responding in a personal capacity (“Individual”)
 - individual responding in a professional capacity (“Professional”)
 - organisation (“Organisation”).
- There were big differences in responses from the different UserTypes. Individuals were more likely not to support the Bill than any other option. Professionals and Organisations were more likely to support the Bill.

- More than four fifths of the total responses were from individuals (**562, 86.5%**), and more than two thirds (**381, 67.8%**) of these do not support the Bill.
 - Of the **59** responses from Organisations (**9.1%**), most (**52, 88.1%**) support the Bill.
 - The remaining **29** responses were from Professionals (**4.4%**) and most of these (**25,86.2%**) also support the Bill.
- **92.7 %** of the Individual responses (521 out of 562) were received from within Wales. The remaining Individual responses (41) were either from outside Wales or the respondents did not specify their postcode.
 - Respondents to the consultation were asked seven questions in connection with the Bill. These were based on the Committee's terms of reference for the Bill's scrutiny.
 - The **most frequent (co-occurring) phrases** (see annex B for more information on how we found co-occurring phrases) used by respondents who do not support the Bill include: *child abuse, love parent, family life, police social service/workers, physical punishment, reasonable chastisement, already protect children, Welsh government, criminalise love parent, unintended consequences.*
 - The **most frequent (co-occurring) phrases** used by respondents who support the Bill include: *physical punishment, corporal punishment, Welsh government, human/ children's right, children young people, positive parent, defence reasonable punishment, committee right child, clear message.*
 - The **most important phrases** (see annex B for more information about how importance is calculated statistically) used by respondents who do not support the Bill include: *child abuse, love parent, social service, protect children, family life, discipline children, reasonable chastisement, criminalise parents.*
 - The **most important phrases** used by respondents who support the Bill include: *corporal punishment, human/children's right, reasonable*

punishment, protect children, positive parent, physical abuse, mental health.

- There is a good degree of overlap between the most frequent phrases and most important phrases emerging from the responses who do not support the bill.
- Likewise, there is a good degree of overlap between the most frequent phrases and most important phrases emerging from the responses who do support the bill.
- In detail, of the **650** responses:
 - **562** were from individuals (representing 86.5% of the overall number of responses)
 - **28.6%** (161 of 562) support the Bill (representing 24.8% of the overall number of responses)
 - **67.8%** (381 of 562) do not support the Bill (representing 58.6% of the overall number of responses); and
 - **3.6%** (20 of 562) partly support the Bill (representing 3.1% of the overall number of responses);
 - **59** were received from Organisations (representing 9.1% of the overall number of responses)
 - **88.1%** (52 of 59) support the Bill (representing 8.0% of the overall number of responses);
 - **8.5%** (5 of 59) do not support the Bill (representing 0.8% of the overall number of responses); and
 - **3.4%** (2 of 59) do not have a view Bill (representing 0.3% of the overall number of responses);
 - **29** were received from Professionals (representing 4.4% of the overall number of responses)
 - **86.2%** (25 of 29) support the Bill (representing 3.8% of the overall number of responses); and

- **13.8%** (4 of 29) do not support the Bill (representing 0.6% of the overall number of responses)
- The methodology is presented in Annex A.

A. Overall responses, in support or otherwise for the Bill:

We first separate the responses according to the type of response in support or otherwise for the Bill (ignoring the UserType category) and express them as a percentage of the total number of the responses as shown in Figure 1 below.

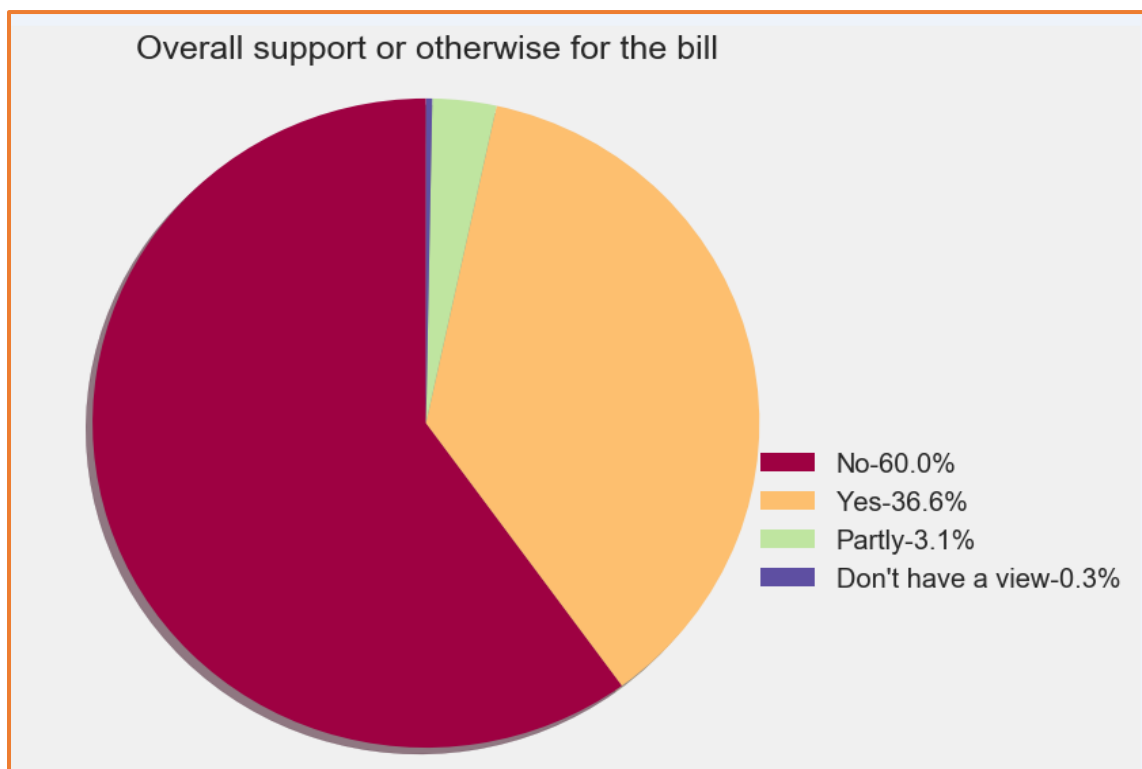
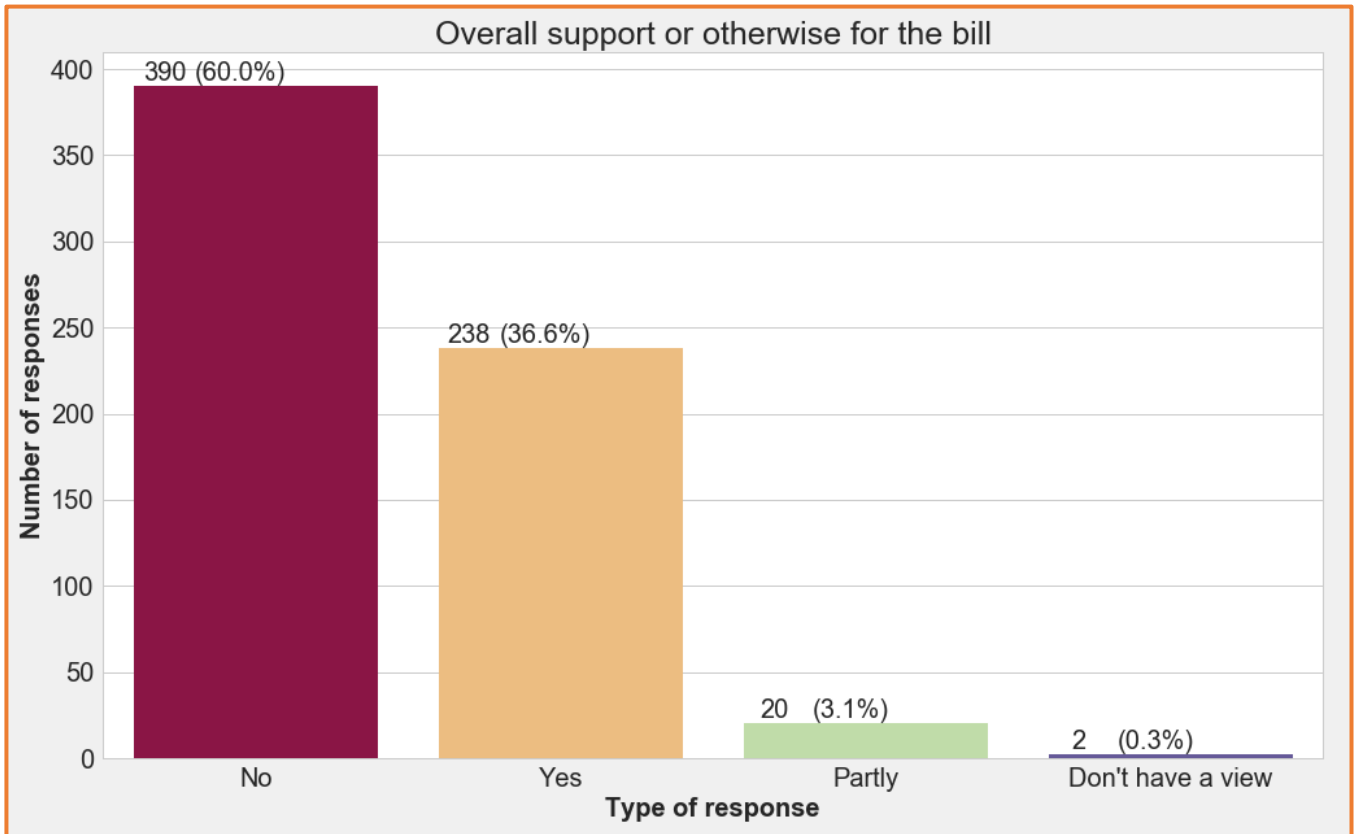
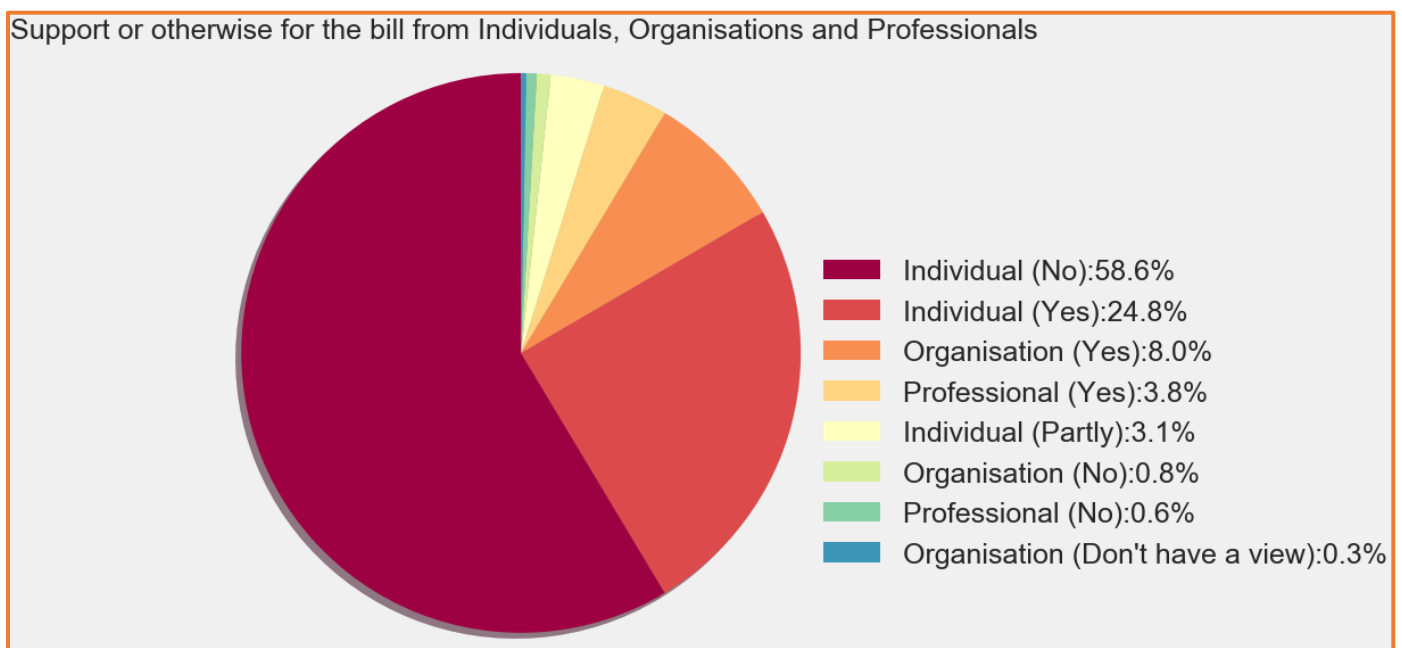
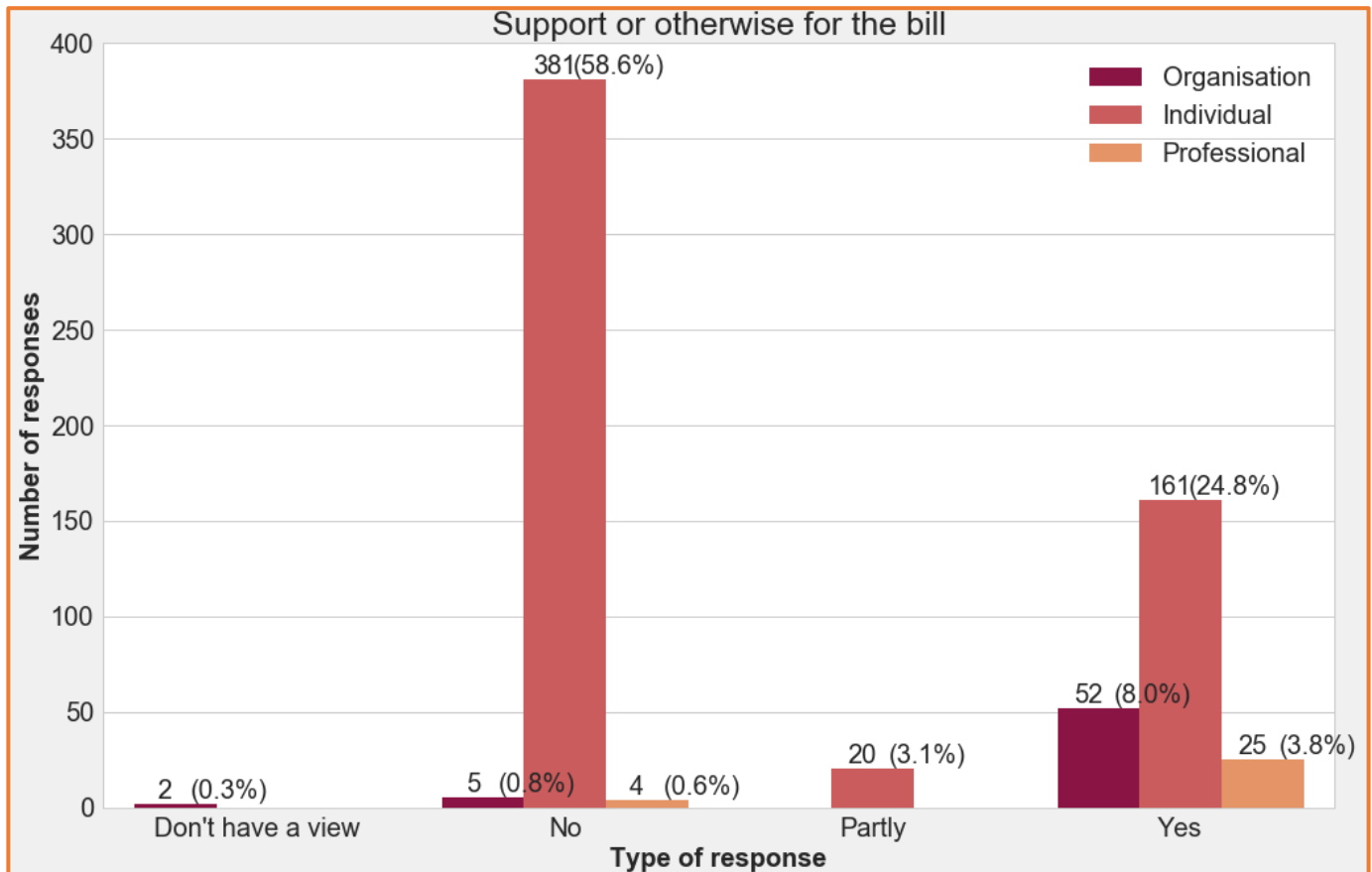


Figure 1: The total number of responses is 650. 60.0% of the responses were not in support for the bill, followed by 36.6% of the responses in support for the bill. 3.1% of the responses were in partial support for the bill and 0.3% of the responses expressed no views on the bill. The responses above show the cumulative figure for different UserTypes (Individuals, Organisations and Professionals).

B. Responses categorised in support or otherwise for the Bill:

The respondents were able to respond under three different UserTypes (Individual/Professional/Organisation). We thus separate the responses according to the type of response in support for the Bill for each UserType category and express them as a percentage of the total number of the responses as shown in Figure 2.



UserType	Number of responses	% of overall number of responses
Individual	562	86.5
Organisation	59	9.1
Professional	29	4.4

UserType	BillSupport	Number of responses	% of overall number of responses
Individual	No	381	58.6
	Partly	20	3.1
	Yes	161	24.8
Organisation	Don't have a view	2	0.3
	No	5	0.8
	Yes	52	8
Professional	No	4	0.6
	Yes	25	3.8

Figure 2: Responses categorised by different UserType/categories in support or otherwise for the Bill. Most of the responses received were submitted by Individuals (86.5%), followed by Organisations (9.1%) and Professionals (4.4%). Individuals who do not support the Bill constitute 58.6% of all responses. 24.8% of all responses were from Individuals who support the Bill and 3.1% of all responses were from Individuals who partially support the Bill. 8% of all responses were received from Organisations who support the Bill. Of the remaining 1.1% of responses received from Organisations, 0.8% do not support the Bill and 0.3% do not have a view. Lastly, 3.8% of the overall responses came from Professionals supporting the bill and 0.6% of the responses from Professionals were not in support for the bill.

C. Responses categorised by the category of respondent (UserType):

Since most responses were received from Individuals, in Figure 2 the proportion of responses received from Organisation and Professional was not very visible. It is therefore worthwhile to separate the responses based on the respondent type and visualise the type of responses received under three different categories separately. We categorised three separate sets of information based on UserType and type of response in support for the Bill and the resulting information is displayed in Figure 3.

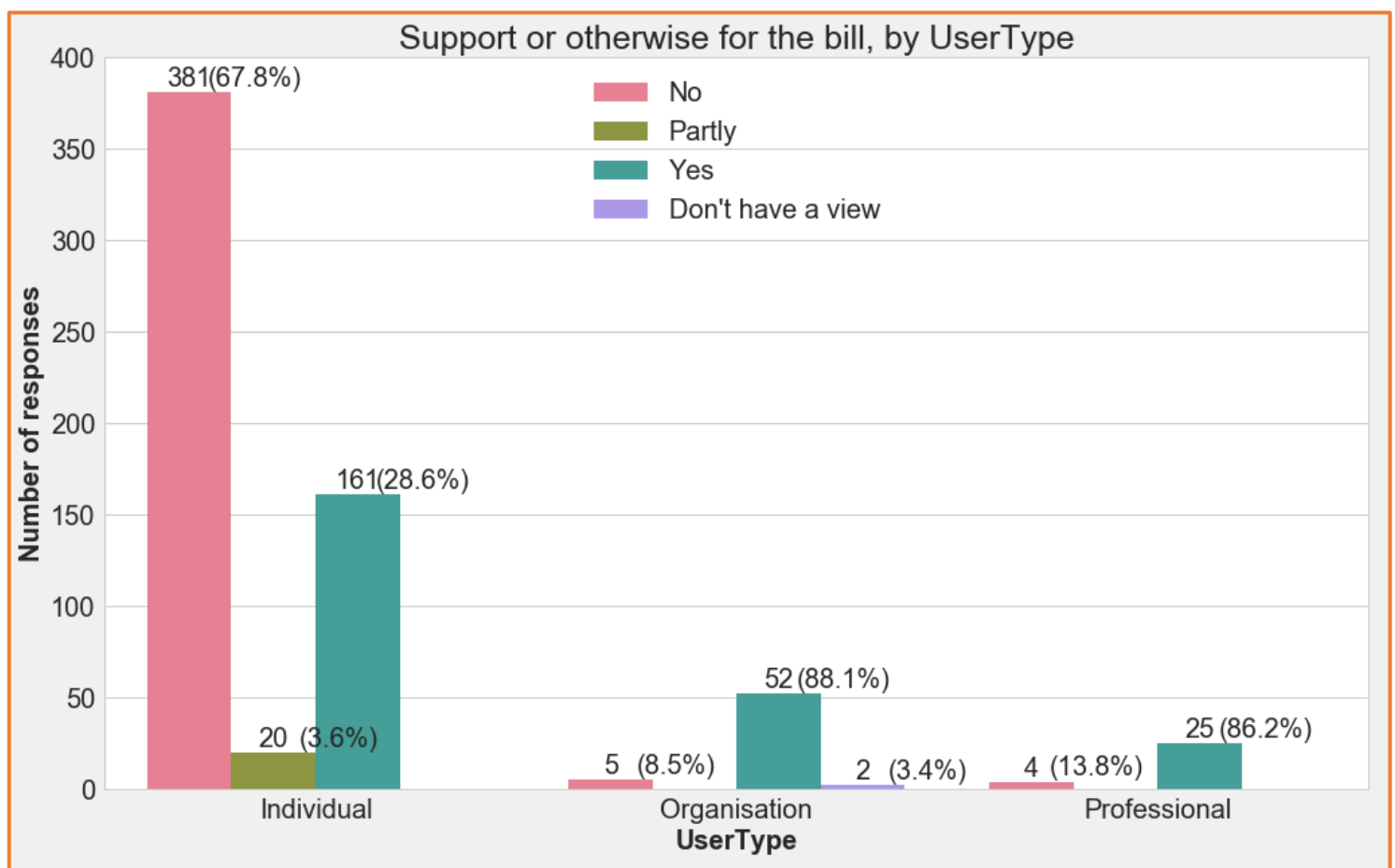
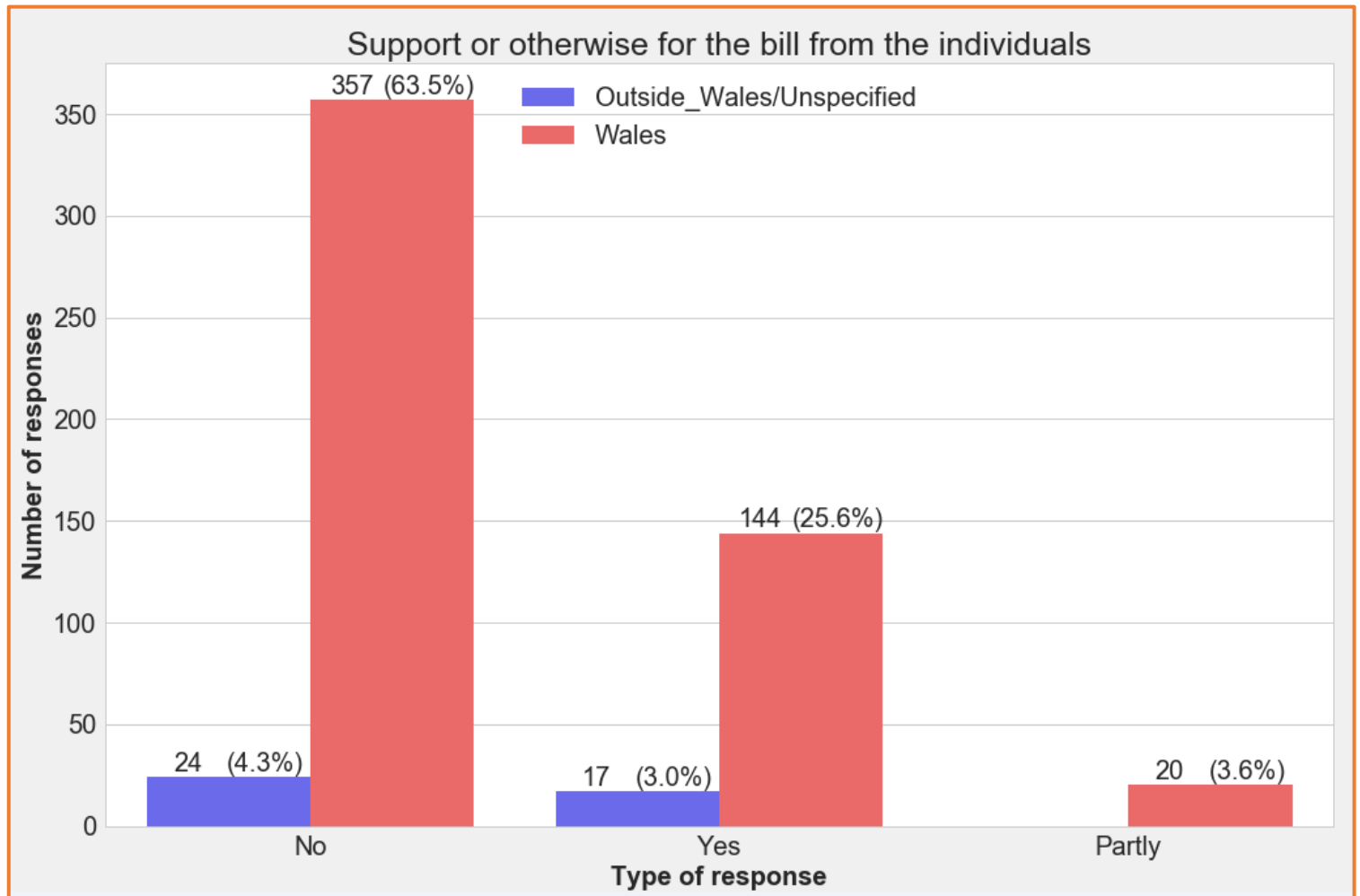


Figure 3: Responses categorised according to their support or otherwise for the Bill for each UserType. As shown above, most of the Individual responses do not support the Bill (67.8%). The reverse was true for responses received from both Organisations (88.1% in support for the Bill) and Professionals (86.2% in support for the Bill). However, it should be noted that overall the sample size of the responses in the Individual category is the highest (562 responses) followed by

Organisation (59 responses) and Professional (29 responses) and, when all user types are taken together, most of the responses are not in support for the bill (390 of 650 responses – 60%).

D. Individual responses categorised according to region (outside/inside Wales):

We identified the respondents' region from the postcode information available from the dataset. This information was only applicable for Individual responses (562), and the split according to the regions is shown below in Figure 4. Out of 562 individual responses, we see 92.7% (521 of 562) were received from within Wales.



BillSupport	Region	Number of Individual responses	% of overall Individual responses
No	Outside_Wales/Unspecified	24	4.3
	Wales	357	63.5
Partly	Wales	20	3.6
Yes	Outside_Wales/Unspecified	17	3
	Wales	144	25.6

Figure 4: Individual responses classified according to the region of the respondents. 92.7 % of the Individual responses were received from within Wales (63.5% who do not support the Bill and 25.6% who fully support the Bill and 3.6% in partial support for the Bill). The rest of the Individual responses (7.3%) were from either outside Wales or the respondents did not specify their postcode.

E: Most frequent co-occurring phrases where respondents replied that they do not support the Bill (390 responses):

We focussed on detecting most prevalent themes in the responses using word collocation, which can be understood as expressions of multiple words which commonly co-occur e.g. 'New York', 'Curly hair', 'Blue sky'. Please refer to Annex A and B for further technical details.

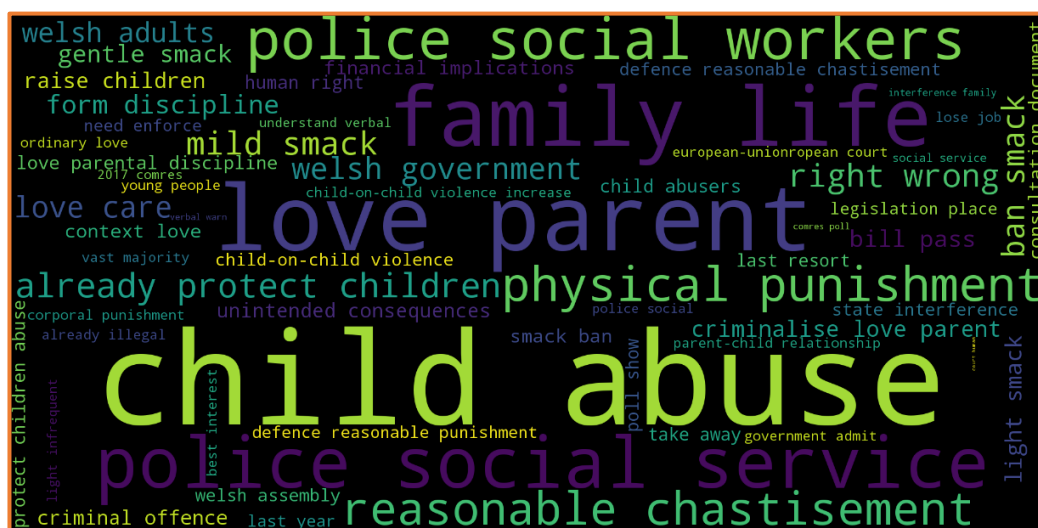
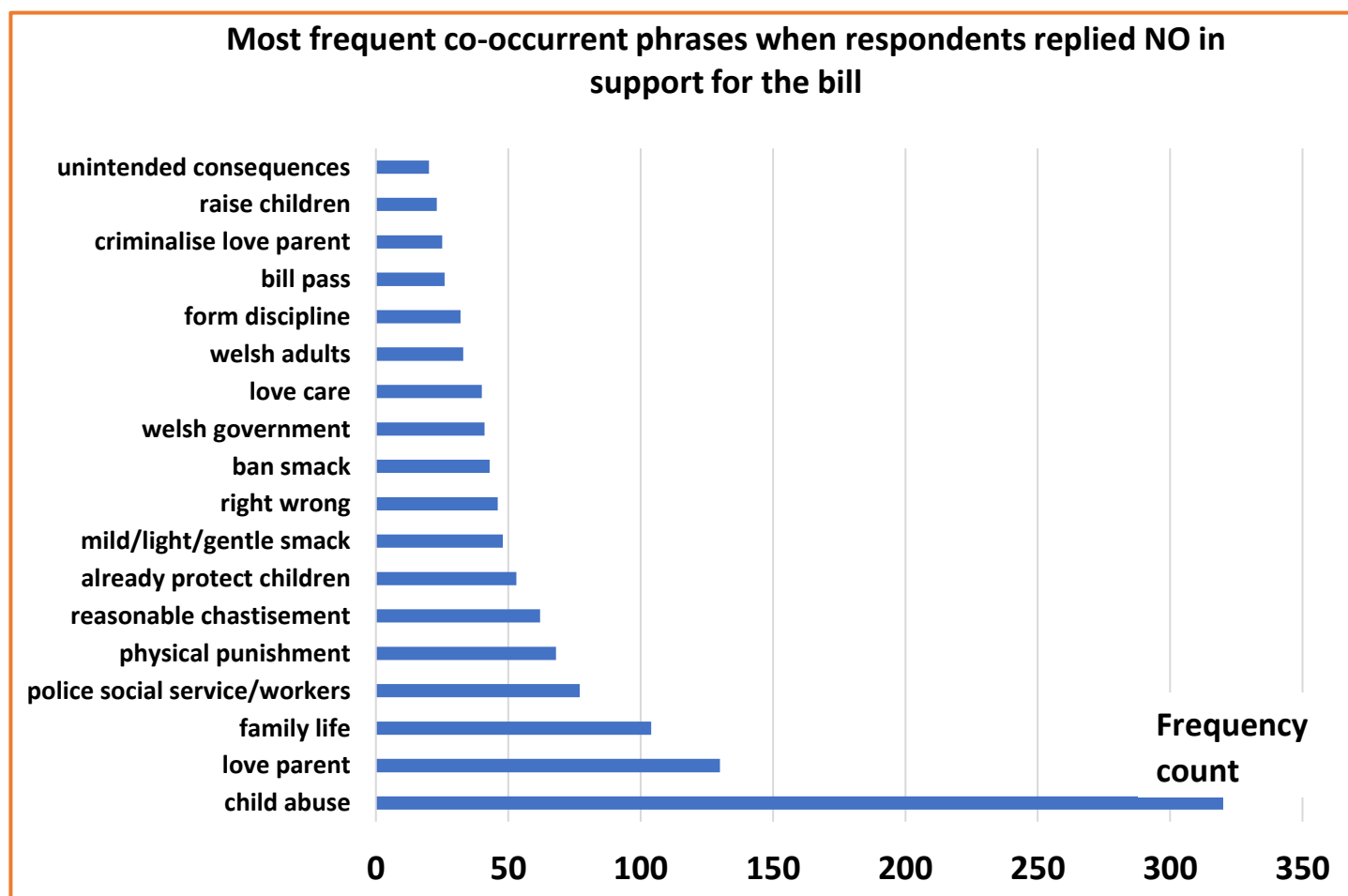


Figure 5: (Top) Bar chart showing most frequent (co-occurring) phrases and their frequency counts when respondents replied **NO** in support for the Bill where the y-axis shows the frequency of the most frequent (co-occurring) phrases. Larger text size in the word cloud (bottom) represents phrases with higher frequencies. Please refer to Annex A and B on the evaluation of co-occurrent phrases.

In addition to finding frequently used phrases in the responses, we can also find out the context around key phrases. This is particularly important when a similar key phrase appears in responses supporting/opposing the bill. A few of the most frequent (co-occurring) phrases along with some of the respondents' comments not in support for the Bill are shown below.

Child abuse: 'there is currently sufficient legislation in place to deal with child abuse', 'interpreting smacking as child abuse is erroneous', 'banning smacking would criminalise innocent parents whilst real child abusers could be overlooked', 'this will waste precious public resources that should be focused on child abuse'...

Love parent: 'a loving parent will know far better than anyone else what their own child needs', 'it is quite wrong to criminalise loving parents who want to bring their children up to be responsible citizens', 'loving parents demonstrate their love by teaching their children right from wrong'...

Family life: 'the state is already interfering too much in family life', 'family life will be disrupted as parents will be afraid to reasonably discipline their children', 'do you really want to see children put into the care system when their parents are those best placed to bring them up even if their views differ from yours and when their of family life demonstrates loving care'...

Police social service/workers: 'our police and social workers are already overstretched', 'the police and social services have better things to be doing than interfering in healthy stable families', 'police and social services will be overwhelmed and distracted from serious cases that really do need their attention'...

Physical punishment: 'the law already bans abusive and excessive physical punishment so there is no need for this bill', 'it can be much more damaging for parents to resort to shouting swearing etc rather than keeping control and giving reasonable and appropriate physical punishment', 'there is no evidence that mild physical punishment harms children and the government even upheld this in a consultation last year', 'whilst I accept there are circumstances in which it could be harmful to use physical punishment there are also all sorts of non physical punishments that are equally damaging.'...

Welsh government: 'the Welsh Government has the responsibility to listen to the voice of the Welsh people and not to ignore it', 'this bill should be thrown out immediately and the Welsh Government should use their resources to address real and pressing issues regarding children such as educational standards.'...

We focussed on detecting most prevalent themes in the responses using word collocation, which can be understood as expressions of multiple words which commonly co-occur e.g. ‘*New York*’, ‘*Curly hair*’, ‘*Blue sky*’. Please refer to Annex A and B for further technical details.

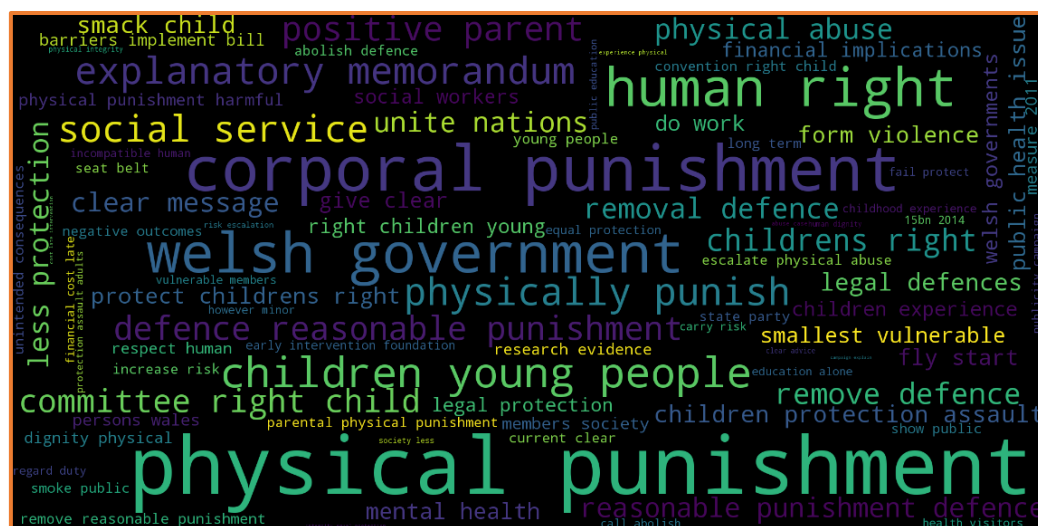


Figure 6: (Top) Bar chart showing most frequent (co-occurring) phrases and their frequency counts when respondents replied **YES** in support for the Bill where the y-axis shows the frequency of the most frequent (co-occurring) phrases. Larger text size in the word cloud (bottom) represents phrases with higher frequencies. Please refer to Annex A and B on the evaluation of co-occurrent phrases.

In addition to finding frequently used phrases in the responses, we can also find out the context around key phrases. This is particularly important when a similar key phrase appears in responses supporting/opposing the bill. A few of the most frequent (co-occurring) phrases along with some of the respondents' comments in support for the Bill are shown below.

Physical punishment: 'I believe that physical punishment of children is the abuse of the most vulnerable people who need protecting and nurturing not hurting', 'physical punishment does not work and because it does not work there is a danger of it escalating into physical abuse and in some cases lead to the death of a child', 'there is extensive research evidence that physical punishment is ineffective and it can cause considerable harm', 'removing the reasonable punishment defence acknowledges children's status in society as holders of human rights and is consistent with the due regard duty in the rights of children and young persons wales measure 2011'...

Welsh government: 'extremely pleased the Welsh Government is leading the way in the UK and taking this forward', 'the proposal put forward by the Welsh Government will ensure that Welsh law is consistent with the UK's international human rights obligations and will provide children with equal protection against criminal assault as that presently enjoyed by adults in Wales', 'I believe removing the defence of reasonable punishment in law in Wales will help to safeguard children from assault and is in line with the Welsh Government's commitment to children's rights'...

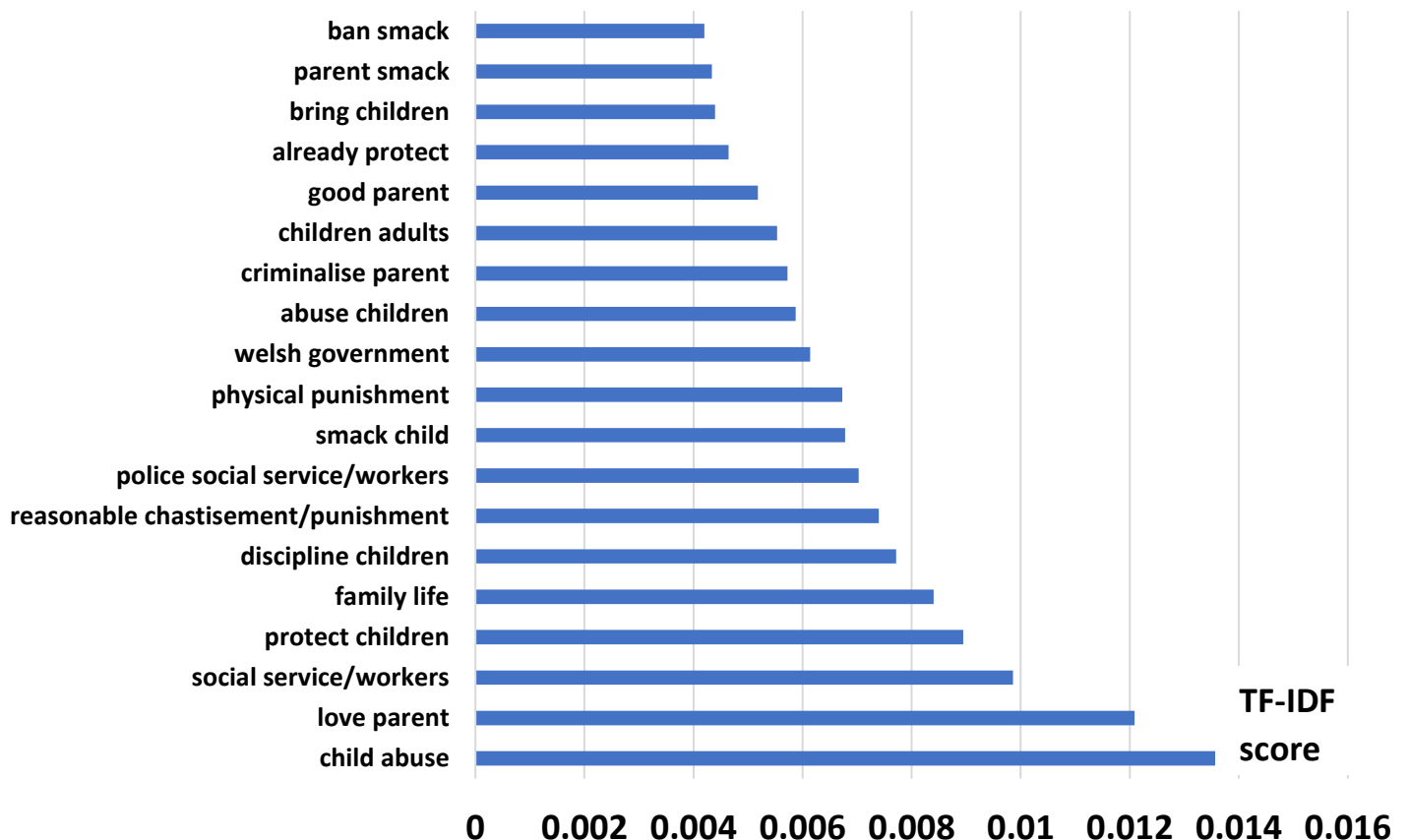
Committee right child/United Nations: 'the United Nations Committee on the Rights of the Child recognises that any physical punishment of children however minor is incompatible with the human rights of children and has called for it to be abolished', 'international treaty bodies have repeatedly called for the UK to legislate to remove the reasonable punishment defence including recommendations from the UN Committee on the Rights of the Child which repeated this call for the fourth time in June 2016. International human rights bodies are unequivocal in stating that children have the same right to legal protection from assault as adults.'...

Clear message: 'a total ban on all physical punishment of children is clear and unambiguous whereas the current defence of reasonable chastisement is open to interpretation and can create confusion uncertainty and gives a clear message that children the most vulnerable members of our society do not enjoy equal rights to adults', 'this clear message removes any legal loop holes regarding parental physical chastisement of children in Wales which will be no longer be defensible under the legislation'...

G: Most important phrases where respondents replied that they do not support the Bill (390 responses):

In a diverse set of responses, it is not enough to only focus on frequently (co) occurring phrases. It is equally desirable to pull out salient/important key phrases across all the responses- we achieve this by evaluating the term frequency inverse document frequency (tf-idf) score for phrases in the responses. The higher the tf-idf score, the more important/salient the phrase. Please refer to Annex A and B for further technical details (including definition of tf-idf score) on the evaluation of important phrases.

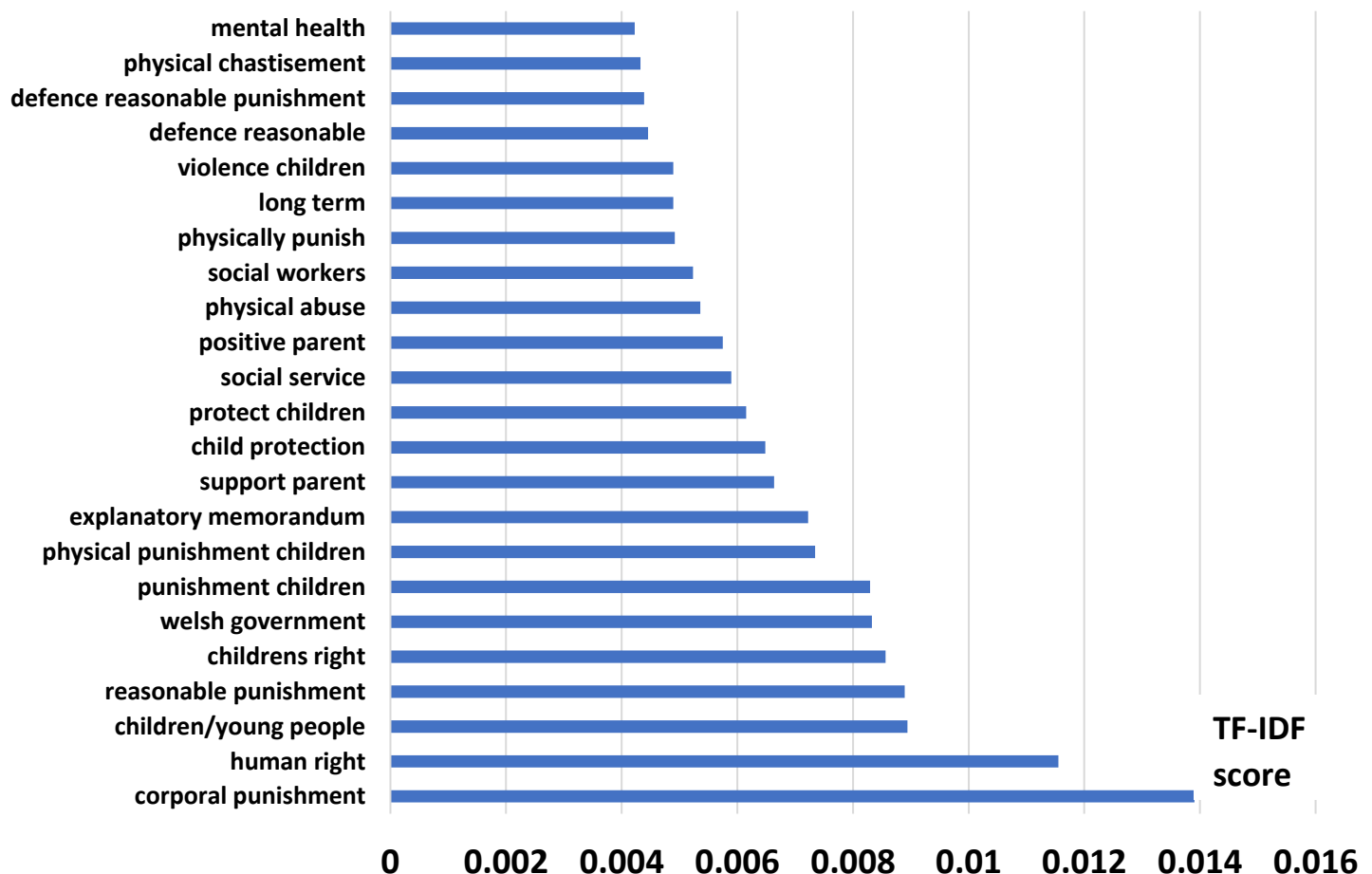
Most important phrases when respondents replied NO in support for the bill



H: Most important phrases where respondents replied that they do support the Bill (238 responses):

In a diverse set of responses, it is not enough to only focus on frequently (co) occurring phrases. It is equally desirable to pull out salient/important key phrases across all the responses- we achieve this by evaluating the term frequency inverse document frequency (tf-idf) score for phrases in the responses. The higher the tf-idf score, the more important/salient the phrase. Please refer to Annex A and B for further technical details (including definition of tf-idf score) on the evaluation of important phrases.

Most important phrases when respondents replied YES in support for the bill



Annex A:

Structure of the dataset: cleaning and anonymisation

The responses to the online consultation on the general principles of the Bill were provided to the ONS Data Science Campus in a csv file format. The dataset was analysed using a numerical code developed by the [Data Science Campus](#) and written in [Python](#) which is a general-purpose programming language. All the results provided are fully reproducible provided one has access to the data and the algorithm developed by the Campus which can be made freely available by the Campus.

One of the very first steps in applying data science techniques is data cleaning. We cleaned the dataset by removing those columns all rows of which had missing values. The resulting shape of the dataset retrieved from the online portal (Smart Survey) was **650 X 32** i.e. the dataset has **650** responses and each specific response had **32** different attributes. However, we focussed only on the attributes relevant for the analysis, some of which are shown in Figure Annex.1. Since the dataset contained sensitive personal information, the dataset was anonymised before it was analysed (removing names, email addresses and actual postcodes replaced with location inside/outside of Wales).

WhatLangSubmit-personal	Are you representing an Org	UserType	Organisation - Eng	Q6 BillSupport	Q10 12 Please outline your reasons for your answer to question	Q11 13 Do you think there is a need for legislation to deliver	Q12 21 Do you have any comments about any potential barriers t	Q13 22 Do you think the Bill takes account of these potential	Q14 31 Do you think there are any unintended consequence	Q15 41 Financial Implications	Q16 51 Other Comments
2.0	NaN	Individual	-	Yes	I believe that children are among the most vu...	Yes I do, people interpret reasonable force ...	-	-	-	-	-
NaN	2.0	Professional	-	Yes	Children should be afforded no less protection...	Yes.\nWhilst this defence exists within the la...	There will be resistance to the change from so...	Yes	No	Professionals will be able to intervene earlie...	-
2.0	NaN	Individual	-	No	The Bill's proposals, if implemented, would be...	I have heard that this Bill is to some extent ...	No, if the Bill was to fall for lack of time o...	-	-	As has already been pointed out by opponents o...	-
2.0	NaN	Individual	-	Yes	I was a battered child. This type of punishme...	Yes - must be approved by law to stop it	I think many will object! Usual things will b...	No. You need a lot of publicity and re-educat...	Yes / backlash from voters!	-	There is no defence for assaulting a child. I...
2.0	NaN	Individual	-	Yes	Assault is always wrong and the person/s respo...	Yes, assault if an is illegal, the same should...	-	-	-	-	-

Figure Annex.1: A snippet of the dataset obtained showing the relevant attributes used in the analysis.

Dataset: initial grouping in various categories:

- *'Usertype'*: The respondents were asked to submit their opinions under three different capacities: Individuals, Professionals and Organisations. The total number of responses received under these three different categories is **650**.
- *'BillSupport'*: The respondents were asked to express their support or otherwise for the Bill by choosing one of the following answers to the question *'1.1. Do you support the principles of the Children (Abolition of Defence of Reasonable Punishment) (Wales) Bill?'*: NO, YES, PARTLY, DON'T HAVE A VIEW.
- The responses were first split according to the following attributes of the dataset: *'UserType'* and *'BillSupport'*. Thereafter, responses in each category were collated for the following seven questions:

'Please outline your reasons for your answer to question 1.1'

'Do you think there is a need for legislation to deliver what this Bill is trying to achieve?'

'Do you have any comments about any potential barriers to implementing the Bill?'

'Do you think the Bill takes account of these potential barriers?'

'Do you think there are there any unintended consequence arising from the Bill?'

'Do you have any comments on the financial implications of the Bill (as set out in Part 2 of the Explanatory Memorandum)?'

'Do you have any other points you wish to raise about this Bill?'

Annex B:

Dataset: most frequent (co-occurring) and important phrases

- Most data science techniques rely on a large volume of training data to train machine learning models with a hope that these models can then be confidently used for predictive analysis on the (unseen) test data. In the present scenario, the number of free text responses were relatively modest for many [natural language processing](#) techniques including [topic modelling](#) to be effective.
- We thus focussed on those data science techniques which do not critically depend on the volume of training data. We started the text mining analysis with first combining the responses under two main categories- full support or opposition to the bill (ignoring the UserType and Region categories). Splitting the responses under these two categories allowed us to collate a sizeable number of responses to allow data science techniques to extract most frequent (co-occurring) and important/key phrases present in the respondents' responses to the bill.

Most frequent (co-occurring) phrases:

- We performed the analysis by first extracting the most frequent phrases in the responses under the above two categories. We focussed on detecting word collocation, which can be understood as expressions of multiple words which commonly co-occur e.g. 'New York', 'Curly hair'.
- We used [Gensim](#) to automatically detect common phrases and multi-word expressions from a stream of sentences. Gensim is an open-source library for unsupervised topic modelling and natural language processing, using modern statistical machine learning.
- It is possible that certain key phrases can appear in responses both in support/opposition to the bill. This report therefore also accompanies html tables showing the context of these frequently occurring phrases (e.g. 'physical punishment' as a phrase appear in both categories: 'NO' and 'YES').
- **While detecting co-occurring phrases we can make use of two parameters: min_count and threshold to ensure only meaningful key phrases are picked up by our model. Min_count** ignore all phrases with total collected count lower than this **and threshold** represents a threshold for forming the phrases (higher means fewer phrases). We used

Min_count=10 and threshold=5 in detecting most frequent (co-occurring) phrases.

Most important phrases:

- In a diverse set of responses, it is not enough to only focus on frequently (co) occurring phrases. It is equally desirable to pull out salient/important key phrases across all the responses. We apply tf-idf/TFIDF, short for [term frequency-inverse document frequency](#), which is a numerical statistic that is intended to reflect how important a word is to a response in a collection of comments. The tf-idf value increases proportionally to the number of times a word appears in the document and is offset by the number of documents in the corpus that contain the word, which helps to adjust for the fact that some words appear more frequently in general (common words a, the, of, at etc are very frequent but do not convey much meaning on their own).
- For example, consider a response containing 20 words wherein the phrase '*good-discipline*' appears 3 times. The term frequency (i.e., tf) for '*good-discipline*' is then $(3 / 20) = 0.15$. If we assume we have 5000 responses and the phrase '*good-discipline*' appears in 10 of these, the inverse document frequency (i.e., idf) is calculated as $\log(5000 / 10) = 2.69$. Thus, the tf-idf weight is the product of these quantities: $0.15 * 2.69 = 0.4035$. Thus tf-idf allowed us to increase the weight of this key-phrase from 0.15 to 0.40.
- We used [Scikit-learn](#) which is another free software machine learning library for the Python programming language to extract important phrases based on their tf-idf weights.
- In the sections relating to the most frequent and most important words we lay out both bar charts and word cloud outputs for most frequent co-occurring phrases and important phrases. With the larger text size in the word clouds representing terms with higher weights (either co-occurrence frequency or tf-idf weights).