



# Identifying human trafficking indicators in the UK online sex market

Luca Giommoni<sup>1</sup>  · Ruth Ikwu<sup>2</sup>

Accepted: 25 August 2021  
© The Author(s) 2021

## Abstract

This study identifies the presence of human trafficking indicators in a UK-based sample of sex workers who advertise their services online. To this end, we developed a crawling and scraping software that enabled the collection of information from 17,362 advertisements for female sex workers posted on the largest dedicated platform for sex work services in the UK. We then established a set of 10 indicators of human trafficking and a transparent and replicable methodology through which to detect their presence in our sample. Most of the advertisements (58.3%) contained only one indicator, while 3,694 of the advertisements (21.3%) presented 2 indicators of human trafficking. Only 1.7% of the advertisements reported three or more indicators, while there were no advertisements that featured more than four. 3,255 advertisements (19.0%) did not contain any indicators of human trafficking. Based on this analysis, we propose that this approach constitutes an effective screening process for quickly identifying suspicious cases, which can then be examined by more comprehensive and accurate tools to identify if human trafficking is occurring. We conclude by calling for more empirical research into human trafficking indicators.

**Keywords** Human trafficking · Modern slavery · Internet sex trafficking · Policing sex markets

---

Luca Giommoni (<http://orcid.org/0000-0002-3127-654X>) is a senior lecturer in criminology at the Cardiff School of Social Sciences

Cardiff University. He is interested in quantitative analyses of both the factors that influence international trafficking and online illicit markets.

Ruth Ikwu is a research associate at the Social Data Science Lab at Cardiff University. She is interested in developing analytical solutions to crimes in cyberspace

---

✉ Luca Giommoni  
giommonil@cardiff.ac.uk

<sup>1</sup> School of Social Sciences, Cardiff University, Glamorgan Building, King Edward VII Avenue, Cardiff CF10 3WT, UK

<sup>2</sup> Social Data Science Lab, Cardiff University, Cardiff, UK

## Introduction

Over the last 20 years, human trafficking has become a major social issue for policymakers, practitioners, and academics. In 2008, According to the United Nations Office on Drugs and Crime (UNODC), only 98 countries (out of 155) had enacted laws criminalising human trafficking (UNODC, 2008, p. 22). Conversely, as of August 2018, 168 countries, out of a total of 181, had passed laws criminalizing trafficking in persons (UNODC, 2018, p. 45). Academics have also paid increasing attention to this topic (Wen et al. 2020). Indeed, in a systematic review of the sex trafficking-related literature, Wen et al. (2020, p. 372) found that while between 1982 and 2003 less than twenty articles were published on ‘human trafficking’, ‘sex tourism’, ‘sex slavery’, and ‘child prostitution’, a total of 334 papers were published between 2004 and 2018.

The opportunities presented by the internet for selling and buying sex has created a new line of research on human trafficking. From 2010 onwards, several studies have examined sex workers’ online advertisements to discern human trafficking patterns. This type of research is popular among computer scientists due to the digital nature and volume of the data (e.g. Ibanez and Gazan 2016; Ibanez and Suthers 2014; Wang et al. 2012; Alvari et al. 2017) but also among social scientists (Latonero, 2011; Skidmore et al., 2018). US-based studies that used data from backpage.com and craigslist.com dominate extant literature in this field (e.g. Latonero 2012; Alvari et al. 2017; Hultgren et al. 2018; Shahrokh Esfahani et al. 2019). Only one study (Skidmore et al., 2018) used non-US based data, utilising manual data collection methods and coding, which resulted in a small sample size and limited geographical coverage. One must be cognisant of the fact that human trafficking is a heterogeneous phenomenon, whose features vary according to its location. Hence, human trafficking in the US may vary considerably from Europe and the UK, in terms of traffickers themselves, victims, trafficking flows and governmental responses.

This study builds on extant literature on internet-trafficking by exploring the presence of human trafficking indicators on the largest dedicated platform for sex work services in the UK. Specifically, we developed and identified the presence of ten indicators in a sample of 17, 362 advertisements for female sex workers. This paper adds to the emergent body of work on the nexus between human trafficking and digital technology by (1) extending, for the first time, the analysis across the entire breadth of the UK, by developing a web crawling and scraping software to automatically collect and store relevant information from the largest dedicated website for sex worker advertisements; (2) developing a transparent and replicable methodology for detecting the presence of ten indicators of human trafficking within our sample.

The paper is structured as follows. The next section reviews extant literature on the internet-human trafficking nexus. The third section presents the research aim and scope, while section four outlines data and methods. Section five reports the results, which are then discussed in section six. The article concludes by delineating the implications of the study for future research.

## Sex markets, online data, and human trafficking

According to some studies, digital technologies facilitate and support traffickers to exploit sex workers (Latonero, 2011; Skidmore et al., 2018). Traditionally, traffickers were forced to conduct their activities underground to minimise their risks. The advent of internet advertising presents an opportunity for traffickers to reach a wide audience, while, simultaneously, maintaining a low profile (Dubrawski et al., 2015; Skidmore et al., 2018). Hence, online advertising can increase exploiters' profits and reduce their risk of detection.

Given this threat, scholars have developed two alternative approaches for identifying human trafficking patterns by analysing online advertisements. The first approach involves identifying trafficking indicators within online sex workers' advertisements. The rationale for this approach is that the presence of indicators of trafficking can enable the identification of suspicious advertisements or potential cases of human trafficking (Bounds et al., 2020; Ibanez & Gazan, 2016; Ibanez & Suthers, 2014; Latonero, 2011; Skidmore et al., 2018). For instance, Ibanez and Gazan (2014) identified seven indicators of online sexual exploitation, including different ages and aliases reported across multiple advertisements, reference to specific nationalities/ethnicities, frequent movement to different locations, or the fact that the sex worker only provided in-call services. Their results showed that 82% of Hawaii's backpage escort advertisements contained one or more indicators of human trafficking.

The second approach uses machine learning, data scraping and mining, as well as natural language techniques to discern human trafficking patterns by automatically collecting information from websites dedicated to advertising sex workers' services (Alvari et al., 2016, 2017; Burbano & Hernandez-Alvarez, 2017; Dubrawski et al., 2015; Hultgren et al., 2018; Portnoff et al., 2017; Szekely et al., 2015; Tong et al., 2017; Whitney et al., 2020). The machine learning algorithm is trained to discern between suspicious and non-suspicious advertisements based on different information. This can include common indicators of human trafficking, such as specific language patterns – e.g. use of first-person plurals (i.e. 'we') or third-person pronouns (i.e. she, they) – or a high degree of similarity in the wording of advertisements. For instance, Portnoff and colleagues (2017) developed a machine learning classifier that uses stylometry to distinguish whether advertisements are written by the same person or different people. These studies ordinarily refer to a so-called 'ground truth' to assess the accuracy of their classifiers for discerning between possible cases of human trafficking and genuine advertisements for sex workers. Ideally, this would be a database of known exploited sex workers, but given the absence of such a database the authors adopted alternative approaches, such as using so-called hard identifiers (e.g. phone numbers) or human annotators. For instance, Portnoff et al. (2017) used phone numbers and email addresses as ground truth authorship. In this case, the classifier is considered to be 100% accurate if it identifies as 'cases of human trafficking' those advertisements that include the same phone number or email address. Conversely, other authors (Alvari et al., 2016, 2017) relied on human

trafficking experts to label advertisements as either being ‘of interest’ or ‘not of interest’ as ground truths for testing their classifiers.

These studies exploiting information from sex workers’ advertisements introduced a new approach to studies of human trafficking, in turn, producing new and large datasets as well as sophisticated analyses about a hidden and hard to reach population. However, these studies are not without methodological limitations. Firstly, some of these studies overstate their scope, insofar as they claim to be able to identify online victims of human trafficking or cases of exploitation. For instance, one of the aims of Skidmore et al.’s study (2018, p. 212) was to ‘estimate the extent to which seemingly independent sex workers in the off-street sex trade are controlled or managed.’ These misleading claims about being able to detect human trafficking are also reflected in the titles of some of these studies: ‘Detection and Characterization of Human Trafficking Networks Using Unsupervised Scalable Text Template Matching’ (Li et al., 2018, p. 3111), ‘Semi-supervised learning for detecting human trafficking’ (Alvari et al., 2017, p. 1) and ‘Ensemble Sentiment Analysis to Identify Human Trafficking in Web Data’ (Mensikova & Mattmann, 2018, p. 5). Despite their claims to the contrary, none of these studies is capable of identifying human trafficking or estimating how many sex workers are being exploited, managed, or controlled. That is to say, the presence of multiple indicators does not constitute proof of trafficking. An advertisement can use third-person pronouns, be similarly worded to other advertisements, allude to underage sex workers and include a phone number cited in other advertisements, but this does not prove with any degree of certainty that human trafficking is occurring. As some studies underscore (e.g. Volodko et al. 2020; Bounds et al. 2020; Dubrawski et al. 2015), at best, such research can simply flag up suspicious advertisements that deserve closer attention from law enforcement and practitioners. What it cannot do is identify cases of human trafficking.

The second problem concerns the selection of human trafficking indicators. In several cases, the selection of indicators appears arbitrary and based on authors’ assumptions, rather than on empirical evidence or established indicators. For instance, Alvari and colleagues (2016, 2017) cited the presence of a link to an external website as being an indicator of human trafficking, insofar as it was suggestive of ‘a more elaborate organization’. To support their rationale, the authors also explicitly referenced Kennedy’s study (2012), who claimed to have developed this indicator based on unspecified guidance from law enforcement experts. Similarly, Ibanez and Gazan (2014) consider ‘references to ethnicity or nationality’ within advertisements as a ‘virtual sex trafficking indicator’ without any explanation of how explicit references to nationality, irrespective of the country of origin, can aid the identification of human trafficking victims.

The third problem concerns the operationalisation of some of these indicators, which in some cases appears to be discretionary and arbitrary. For instance, Skidmore et al. (2018) classified a low-level of English proficiency as an indicator of human trafficking. While we agree that this indicator can be expedient for identifying suspicious advertisements, and, indeed, several human anti-trafficking organisations deem it to be an indicator of human trafficking (Home Office, 2016; Polaris Project, 2012; UNODC, 2020), its practical operationalisation is problematic to say the least. Indeed, it is unclear precisely what constitutes either ‘good enough’ or a poor level

of English proficiency for the authors, who cite no quantifiable or objective standard, but rather base their determination solely on their own subjective judgement, which is scarcely reproducible.

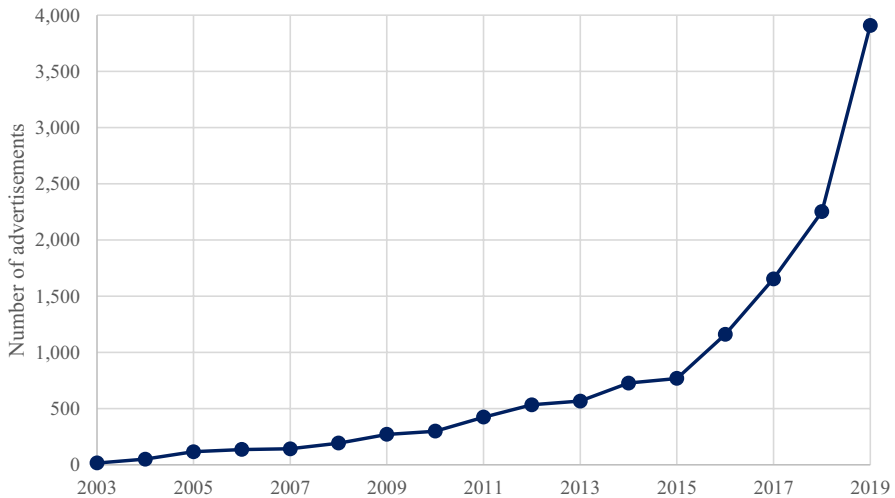
Alongside these three limitations in extant literature, it is also worth noting that there is only one non-US study (Skidmore et al. (2018)), on the use of online sex advertisements to identify indicators of human trafficking. This study identified the presence of certain human trafficking indicators in a sample of 795 profiles advertising sexual services in the south-west of England between July and August 2015. They found that 73 of the profiles (9.2%) cited a phone number that was identical to at least one other sex worker. Via these phone numbers, the authors were able to identify 31 discrete groups comprising between two to five sex workers. This is hitherto the only study to use alternative data sources to the traditional backpage.com and Craigslist.com.

However, Skidmore et al.'s (2018) analysis is limited by their reliance on a manual approach to extracting data from advertisements, which restricted their focus to a limited geographical area due to the intense labour involved. Moreover, replicating their study is either impossible, given the subjective operationalisation of certain indicators, or simply too labour intensive to be extended across an entire country. Consequently, there are no country-wide studies examining the presence of human trafficking indicators outside of the US via the use of automated-data collection methods. Automating data collection, in conjunction with conducting a more critical analysis of human trafficking indicators, is expedient for producing both large datasets and a better understanding of the complexity of human trafficking.

## **The current study**

This study builds on extent literature on internet-trafficking by examining the presence of human trafficking indicators in a sample of UK-based sex workers. More specifically, we developed crawling and scraping software that enabled us to collect information from 17, 362 advertaintments of female sex workers posted on the largest dedicated platform for sex work services in the UK. Compared to previous studies, we established a set of ten indicators deriving from the UNODC indicators of human trafficking (UNODC, 2020). Although empirical research in this area is in its infancy, the UNODC's indicators – which are mostly based on expert opinions – are similar to those of other organisations, such as the Home Office (2016) and the Polaris Project (2012), in part because developed starting from the UNODC's list of human trafficking indicators.

Our approach assumes that the presence of multiple indicators of human trafficking is helpful for flagging-up suspicious advertisements that warrant further attention. The higher the number of indicators in an advertisement, the more suspicious it is deemed to be. Although it cannot categorically identify human trafficking cases, it can guide, and prompt the use of more comprehensive and accurate tools to assess whether human trafficking is occurring. As such, professionals – whether they are healthcare, social work, public health, or law enforcement professionals – can use this approach to determine if further screening for potential human trafficking is



**Fig. 1** Number of new advertisements each year

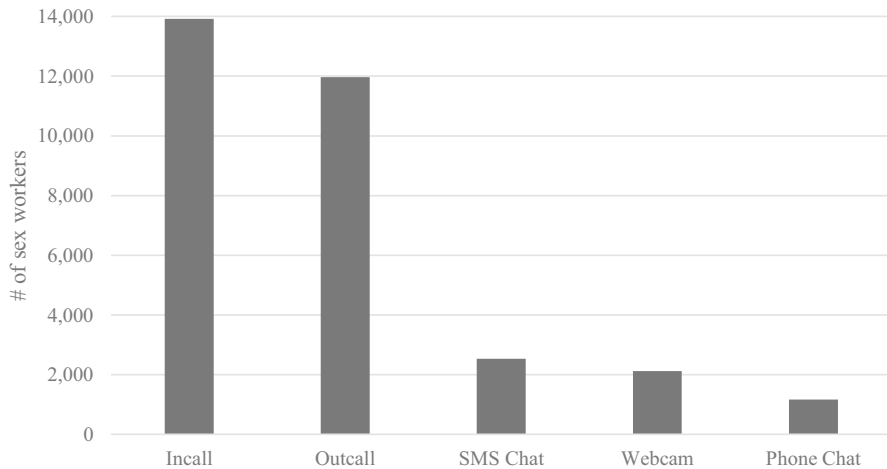
required, based on the number of indicators found in an advertisement. This would help professionals to prioritise their limited resources and protect legitimate sex workers from unwanted police attention (Teela Sanders et al., 2020).

## Methodology

### Data and procedure

We sourced data from a popular multi-service adult entertainment platform in the UK. The platform offers several services, including webcam shows, instant messages and clients' reviews, but the provision of advertising space for sex workers is its principal function. Since launching in 2003, there has been a steady increase in the number of sex workers' advertisements. As Fig. 1 shows, between 2003 and 2011 there were no more than 500 new advertisements each year. This number drastically increased after 2010, culminating in 3,908 new sex workers advertising their services in 2019. This website is now considered a market leader in the UK, given its popularity and capacity to offer a range of services (Stewart Cunningham et al., 2018; Teela Sanders et al., 2018).

While the advertisements are heterogeneous in terms of structure and content, most provide: demographic information about the sex worker (e.g. town where they are active, nationality, age, etc.), information on their physical appearance (e.g. height, hair and eye colour, etc.), sexual orientation (e.g. bi-sexual, heterosexual, etc.), sexual services provided, and pricing. Moreover, most advertisements have free text spaces that the sex workers use to introduce themselves, a public and private gallery to post pictures and videos, and an 'interview' section where sex workers provide more details about themselves and their services.



**Fig. 2** Services advertised by sex workers

We developed a crawling and scraping software to collect this information from the platform. The software automatically accesses, crawls, fetches, and stores information from these advertisements in a database. The data analysed in this study were collected between the 1<sup>st</sup> and 7<sup>th</sup> of February 2020.

Overall, we collected 25, 056 sex workers' advertisements, i.e., all the profiles that were active at the at the time of the data collection. In accordance with previous research (Alvari et al., 2016; Dubrawski et al., 2015; Skidmore et al., 2018), this study solely focused on female sex workers selling services to male clients, and removed advertisements that solely concerned online sex practices. We excluded male sex and online exploitation due to the relative dearth of literature in both these areas.

As Fig. 2 shows, most of the sex workers offered some form of offline face-to-face services: 44% of the advertisements offered in-call services (i.e. the client visits the sex worker), while 38% offered out-call services (the sex worker visits the client). Conversely, some advertisements offered both offline and online services (SMS chat, webcam and phone chat), while a small number of sex workers solely offered online services. Overall, there were 2, 532 advertisements offering SMS chat (8%), 2, 116 offering webcam services (7%) and 1, 164 offering phone chat services (4%). Our final sample included 17, 362 advertisements by female sex workers providing offline sex services to male clients.

### Indicators of human trafficking

This study aims to identify indicators of human trafficking for the purposes of sexual exploitation in our sample of sex workers' advertisements. These indicators were developed through recourse to the UNODC indicators, which include 36 general indicators of human trafficking, and 19 specific indicators of sexual exploitation. (UNODC, 2020).

In total, we developed a list of 10 different operational indicators to apply to our data, two of which were designed to identify the sexual exploitation of minors. We began by attempting to associate the UNODC indicators with the information in the advertisements. For instance, one of the UNODC indicators of sexual exploitation is that ‘There is evidence that groups of women are under the control of others’ (UNODC, 2020, p. 2). We operationalised this specific indicator by searching for information that indicated the presence of a third-party who was controlling sex workers. For instance, if more than one advertisement cited the same contact number, we considered this to be indicator of human trafficking, insofar as it may signal that the sex worker was controlled by others.

Unfortunately, in several cases we were not able to operationalise certain UNODC indicators of human trafficking with the data in the advertisements, such as ‘Show fear or anxiety’, ‘Feel that they cannot leave’, ‘Have limited contact with their families or with people outside of their immediate environment’. (UNODC, 2020, p. 1). The information in the advertisements did not allow for inference about sex workers’ social or family interactions, their level of trust in authorities, or anything regarding their feelings. Therefore, from an initial list of 59 indicators, we narrowed this down to ten operational indicators Table 1 summarises these indicators. As the tables shows, indicators I1-I8 are based on the UNODC indicators of human trafficking (2020). Indicators I9 and I10, instead, were specifically designed to identify the sexual exploitation of minors and were based on previous literature (Alvari et al., 2016; Ibanez & Gazan, 2016). It is also worth to note, that indicators I1, I2, and I3 refer to the same dimension of exploitation, although they are operationalised in a different way.

Below, we present the indicators used in this study. We coded each indicator as ‘1’ if present and ‘0’ if absent. Finally, we coded a lack of information as ‘0’. Appendix 1 describes in more depth how we operationalised them, while all data used in the analysis are available on [Figshare](#).

*I1: Use of third- or first-person plural pronouns.* We classified those advertisements that used first-person plural pronouns (i.e. we) or third-person pronouns (i.e. she, they) as signalling a potential risk of exploitation. The presence of these pronouns can indicate that a third-party has written the advertisement, thus putting the sex worker at a higher risk of being controlled by others (UNODC, 2020, p. 2). To this end, we used computational linguistics (i.e. grammatical tagging) to identify verbs, pronouns and adjectives used in the advertisements, based on their definition and context (see Appendix 1 for more details).

*I2: Same phone number used in more than one advertisement.* We flagged-up advertisements that cited the same phone number, as this can signal that a sex worker is being controlled by others (UNODC, 2020, p. 2), and is one of the most commonly used indicators of human trafficking in extant literature (Latonero, 2012; Portnoff et al., 2017; Skidmore et al., 2018). To reduce the collection of sensitive information and ensure the anonymity of the sex workers, we developed a web-scraping software that automatically transformed phone numbers into a hash (i.e. an alphanumeric code) via an algorithm. We also manually inspected those advertisements citing identical phone numbers to remove cases where the same sex workers had multiple accounts.



**Table 1** Indicators of human trafficking \*

#	Indicator	Rationale
11	The advertisement uses third- or first-person plural pronouns	'There is evidence that groups of women are under the control of others.'
12	Same phone number cited in more than one advertisement	'There is evidence that groups of women are under the control of others'
13	High degree of similarity between sex workers' advertisements	'There is evidence that groups of women are under the control of others'
14	Sex workers offer risky or violent sexual services	'There is evidence that suspected victims cannot refuse unprotected and/or violent sex.'
15	Advertisements promoting inexpensive sex services	'Receive little or no payment'
16	Sex workers moving frequently across several locations	'Move from one brothel to the next or work in various locations'
17	Sex workers moving to different locations along with other sex workers	'Live or travel in a group, sometimes with other women who do not speak the same language'
18	Sex workers offering in-call services only	'Show signs that their movements are being controlled'
19	Advertisements using words that allude to the youthful characteristics of the sex workers	This indicator aims to uncover possible cases of underage sex workers
110	Stating a dress size that is typical of underage women	This indicator aims to uncover possible cases of underage sex workers

\*Indicators 11-18 are based on the UNODC indicators of human trafficking (2020). Indicators 19 and 110 were specifically designed to identify the sexual exploitation of minors and were based on previous literature (Alvari, Shakarian, and Snyder 2016; Ibanez and Gazan 2016).

*13: High degree of similarity between sex workers' advertisements.* We used Natural Language processing (McMahon & Hansen, 2018) to identify the similarity between advertisements. If two advertisements were highly similar, then it is possible that the same person wrote them, which may signal that two sex workers could be under the control of others (UNODC, 2020, p. 2). Appendix 1 provides a more detailed discussion of this method.

*14: Sex workers offering risky or violent sexual services.* According to UNODC indicators of human trafficking, those sex workers who have unprotected or violent sex, or who cannot refuse unprotected or violent sex, are at a higher risk of exploitation (UNODC, 2020, p. 2). Based on this, we drafted a list of all the services offered by the sex workers in our sample and identified those advertisements offering unprotected or violent sex practices. These practices encompassed oral sex without protection, cim (abbreviation for oral ejaculation), receiving domination, receiving BDSM (i.e. bondage, discipline, dominance, submission, and sadomasochism), swallowing, receiving humiliation, receiving fisting, barebacking (i.e. sex performed without a condom) and unprotected sex. All these sex practices involve some form of violence or unprotected sex, increasing the risk of the sex workers contracting sexually transmitted infections.

*15: Advertisements promoting inexpensive sex services.* The fifth indicator pertains to the pricing of sex services. Specifically, we identified those advertisements where the advertised price was lower than £50 for outcall services and lower than £25 for in-call services as signalling risk of exploitation. This operational indicator was based on the UNODC indicator that victims of human trafficking 'Receive little or no payment'. We identified low outliers by drawing upon an interquartile range analysis. Appendix 1 provides more details about how we identified low outliers (Table 2). It is worth to note that the price requested in the advertisement may not correspond to the earnings of victims of human trafficking. In fact, victims of human trafficking can request an amount of money above the threshold indicated, but still receive little or no payments. Despite this, the economic analysis is still able to flag up those cases where a traffickers can force sex workers to keep their prices low while increasing the pool of potential clients thanks to the more competitive prices.

*16: Sex workers moving frequently between several locations.* Based on UNODC indicators, we also flagged-up sex workers that moved frequently between places based on what reported in their advertisements (UNODC, 2020, p. 2). To achieve this, we identified sex workers that reported travelling to different geographical areas and staying there for less than five days, e.g. if they are in Location A at  $t_1$ , and then in Location B at  $t_1 + 3$  days. The 5 days threshold was calculated according to the average number of days spent in each location by sex workers that were on tour.

*17: Sex workers moving to different locations along with other sex workers.* Like the previous indicator, this also tracked sex workers' movements, namely whether

**Table 2** Interquartile Analysis

	MIN	MAX	Q1	Q3	IQR	$e$	$Q3+e$	$Q1-e$
INCALL	£30	£1500	£100	£150	£50	£75	£225	£25
OUTCALL	£20	£4500	£140	£200	£60	£90	£290	£50

multiple sex workers displayed the same movement patterns, i.e. if they moved from Location A to Location B in the same day. According to the UNODC, living or travelling in groups is a sign of possible exploitation.

*I8: Sex workers offering in-call services only.* This indicator concerns those advertisements that offered in-call services only, i.e. when the client visits a sex worker. This may indicate that sex workers are unable to leave their premises, thus highlighting cases where sex workers' movements are restricted (UNODC, 2020).

*I9: Advertisements using words that allude to the youthful characteristics of the sex workers.* This indicator relates to the sexual exploitation of minors. While advertisements cannot report that sex workers are younger than 18 years-old, they can use words alluding to them being underage. Hence, based on prior research, we identified advertisements that used the following words 'young', 'teen', 'petite' and 'tiny' (Alvari et al., 2016; Ibanez & Gazan, 2016).

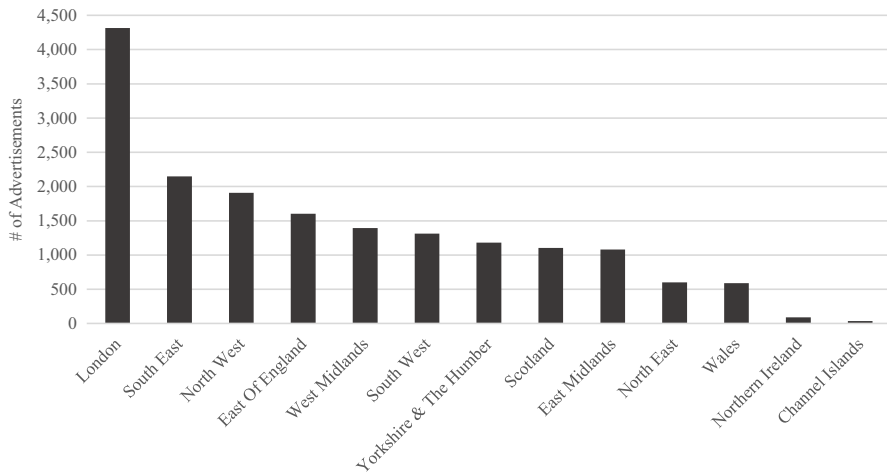
*I10: Stating a dress size typical of underage women.* This indicator also aims to flag-up cases of underage sexual exploitation, by identifying those advertisements that stated that the sex worker was a size 4 (the smallest dress size reported), which is a particularly small size that is more typical of very young women.

It is worth noting that this analysis is only able to identify indicators based on what is reported in the advertisements. However, we are cognisant that there might be a disparity between these advertisements and actual reality. For instance, sex workers might state one price in their advertisements, but charge a different price to clients; or they may promise unprotected sex, but later decline. Moreover, sex workers can provide false information about their age, ethnicity, and so on. Hence, it would be more accurate to say that we can derive the presence of these indicators, but not 'observe' their presence.

Finally, the operationalisation of some of these indicators required setting a threshold beyond which we would consider the indicator to be present. This means that we had to dichotomize certain continuous variables, such as in the case of I5, i.e. price. We did this based on specific statistical procedures (I5) or in line with previous literature (I3). We also considered the option of using alternative scales (e.g. low, medium, high), but given the presence of some strictly binary indicators (I2), we rejected this option. Finally, given the exploratory nature of this study, it was decided that a binary scale would be better at showing the prevalence of these indicators.

## Ethics

The study was reviewed and approved by the Social Science Research Ethics Committee at Cardiff University (SREC/3197). Ethical considerations were paramount in this project, and thus three principles governed the study (1) data collection was acceptable insofar as the data was public; (2) data collection and analysis did not raise the possibility of harm to any party; (3) no individual identifying information would be collected or revealed at any stage of the research process.



**Fig. 3** Number of sex workers' advertisements in each region in the UK

We collected the data from an openly accessible website dedicated to sex work services. Advertisements are easily searchable and retrievable for anyone, as users do not need to create an account to access them.

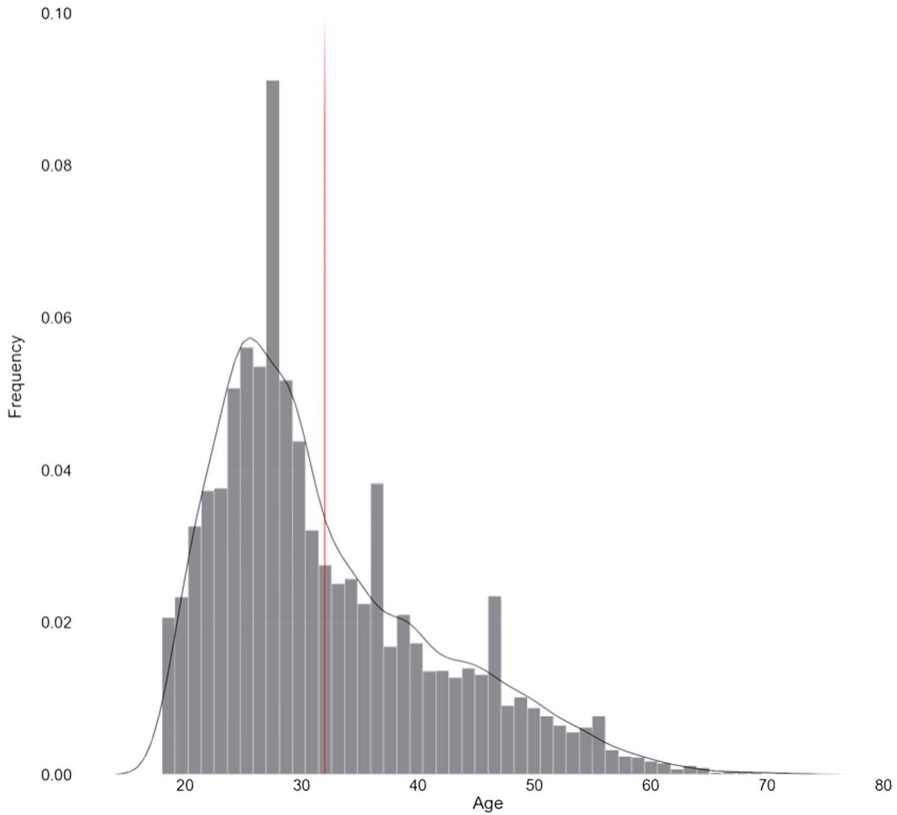
To collect data from these advertisements, we developed a web crawling and scraping software. Contrary to manual data collection, the crawling and scraping software allows for a higher level of privacy by anonymising specific parts of the website. For instance, although the website included identifiable information such as sex workers' phone numbers,<sup>1</sup> this was not collected. Phone numbers were transformed into an alphanumerical code which made it impossible to relate the information to an identifiable natural person. Other personal information that could identify sex workers, such as profile pictures, were also not collected.

## Results

### Overview of the online sex market in the UK

Our sample of 17,362 advertisements comprises sex workers from over 130 countries, including from every European country and every continent. Despite the heterogeneity of the sample, this variable follows a Pareto distribution, in the sense that a handful of nationalities account for most of the sex workers, while the remaining countries account for a much smaller number. In fact, five countries account for 85% of the sample – UK (62%), Romania (10%), Brazil (7%), Poland (4%) and Hungary

<sup>1</sup> We define identifiable information as 'any information relating to an identified or identifiable natural person'.



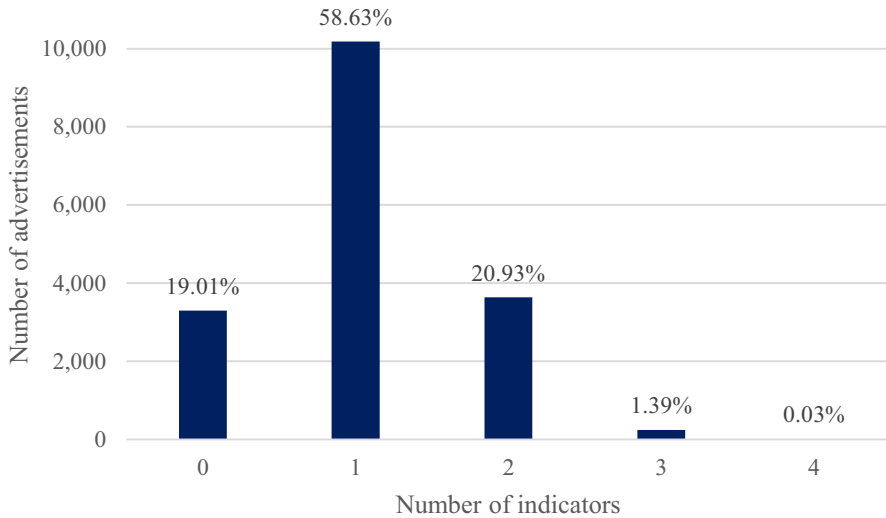
**Fig. 4** Age distribution of sex workers in the sample

(3%) – while the remaining 50 other nationalities combined do not make up 1% of the sample.

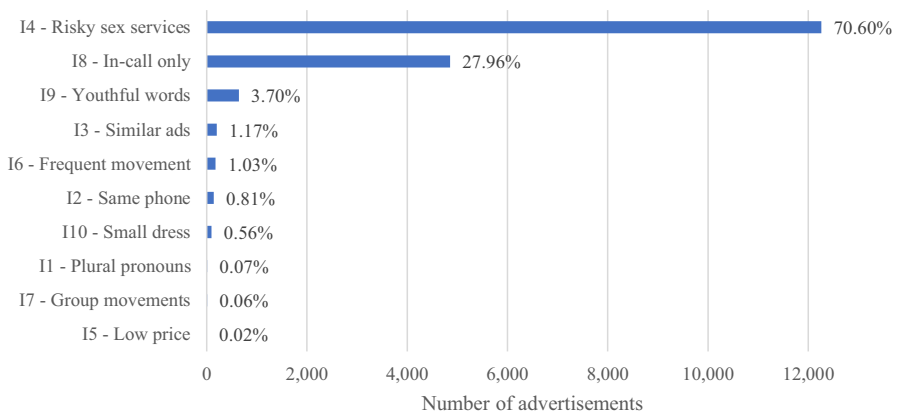
Figure 3 shows the concentration of sex workers' advertisements across the UK. As expected, with 4,313 advertisements (24.8%), London is the area with the highest number of sex workers, followed by the South East (2,149; 12.4%) and North West (1,908; 11.0%) of England. Scotland, Wales and Northern Ireland make up 1,103 (6.4%), 590 (3.4%), and 92 (0.5%) of the advertisements, respectively. Our sample includes sex workers aged from 18 to 72 years-old (see Fig. 4). The average age is 32, while the most reported age is 25 (6.3% of cases). 319 cases reported their age as 18 or 19. One in 5 sex workers in the sample are over 40.

#### Presence of human trafficking indicators.

Figure 5 shows the number of indicators observed within each advertisement. The average number of indicators observed per advertisement is 1.1. Most advertisements (58.63%) present only one indicator. 3,634 advertisements (20.93%) report two indicators of human trafficking, while 3,301 advertisements (19%) do not contain any of the identified indicators. Very few advertisements contain more than two



**Fig. 5** Number of indicators observed in the sample

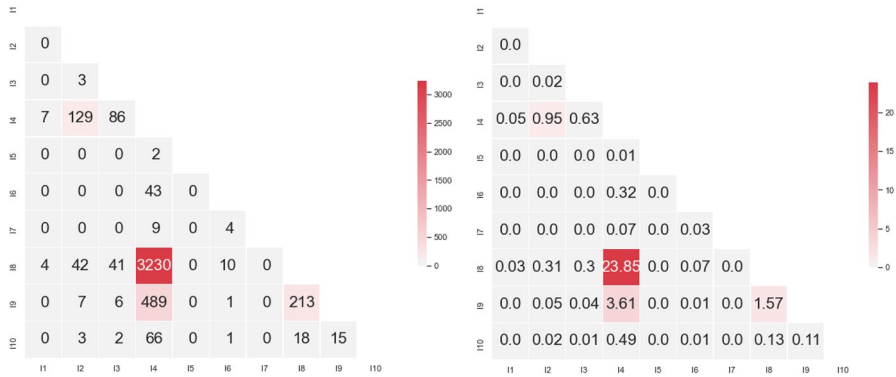


**Fig. 6** Presence of each of the human trafficking indicators within sex workers' advertisements

indicators. 242 advertisements (1.49%) display 3 indicators, while 6 contain four. No advertisements contain more than four indicators of human trafficking.

Figure 6 shows the distribution of the ten human trafficking indicators, ranging from those with the highest prevalence to the lowest. The indicator that occurs most often – I2, 258 (70.6% of the sample)—is number 4, that is, providing risky or violent sexual services, while the second most common indicator is the provision of in-call services from sex workers (I8), which occurred in 4,855 advertisements (27.96%).

After I4 and I8, there are only two indicators that occur in more than 1% of the sample. I10, which is designed to flag-up possible cases of underage sex workers,



**Fig. 7** Indicators co-occurrence matrix. Count (left) and percentage (right)

is found in 643 (3.7%) advertisements. *I3*, which identifies advertisements that display a high degree of similarity, is found in 1.17% of the sample. Overall, 116 sex workers have a profile description that is so similar to other advertisements that it suggests they have been written by the same person. It is worth reminding the readers that this indicator aims to identify if ‘there is evidence that groups of women are under the control of others’ (UNODC, 2020, p. 2).

All the other indicators of human trafficking show an incidence below 1%. *I2*, the use of the same phone number in more than one advertisement, which is one of the most common indicators used in the exploration of the internet-trafficking nexus, is present in 141 advertisements, which amounts to 0.8% of the sample. Only 98 sex workers report having a dress size of 4, which is often associated with underage women.

*I6*, *I1*, *I7* and *I5* have an extremely low incidence, proximate to 0%. During the period of data collection, there were only 59 sex workers (0.3%) that moved across different locations and stayed in each one for less than 5 days (*I6*). There are only 12 advertisements using pronouns suggestive of a third-party, e.g. a trafficker, writing the advertisement on the behalf of the sex worker (*I1*). Similarly, only 11 sex workers display patterns of movement that overlap perfectly with those of other sex workers. This suggests that the number of sex workers travelling in groups, a possible indicator of exploitation according to UNODC (2020), is extremely low. Finally, merely four advertisements are clear outliers, insofar as the sex workers were charging less than £25 per hour (*I5*).

Figure 7 provides a co-occurrence matrix of the indicators included in this analysis, namely the simultaneous presence of two indicators within the same advertisement. Given the low incidence of several indicators in our sample, it is worth noting that many cells report a value of 0. Just over half (24) of the 45 potential co-occurrences have two or more indicators in an advert. The highest co-occurrence is between offering in-call services (*I8*) and the provision of unprotected or violent sex (*I4*): this takes place 3,230 times, i.e. in 23% of the sample. This is followed by the co-occurrence of *I4* and *I9*: 489 advertisements (3.54%) refer both to youthful characteristics and the provision of risky sex services.

Some indicators refer to the same dimension of exploitation. For instance, *I9* (i.e. advertisements using words that allude to the youthful characteristics of the sex workers) and *I10* (dress size typical of underage women), both of which aim to highlight cases of underage sex workers, are both contained in 15 advertisements (0.11%).

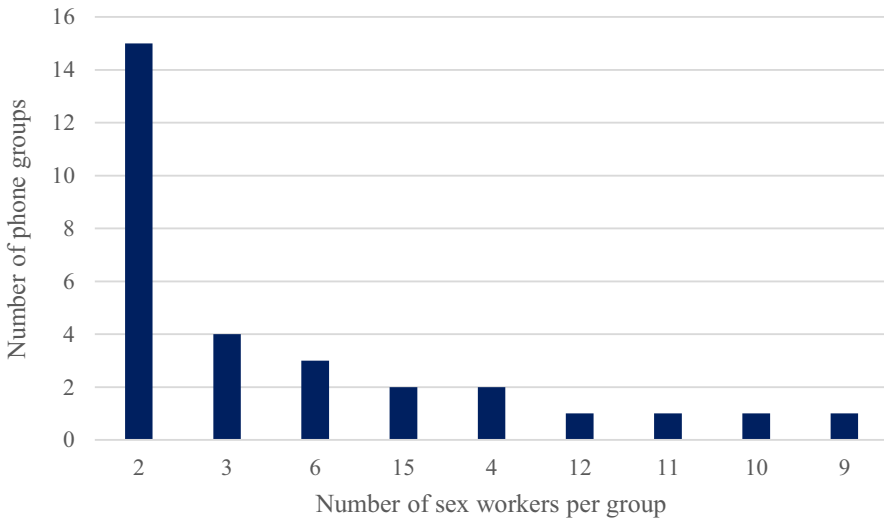
Similarly, only 3 advertisements show a particularly high degree of similarity in terms of their content (*I3*) and cite the same phone number (*I2*), thus suggesting that very few cases meet the UNODC indicator ‘There is evidence that groups of women are under the control of others’. Finally, in only 3 cases is there both *I7* (movement between different locations along with other sex workers) and *I6* (frequent movement between different locations), which also indicates a lack of organised coordination in sex workers’ movements.

## Discussion

This study finds that despite the media hype and attention generated by previous studies in this area (Latonero, 2011; Malo, 2018; Martin & Hill, 2019), there were few indicators of sex trafficking in our dataset. Among a total of 17,362 sex workers’ advertisements, there were only 18,404 indicators identified, which corresponds to only 11% of the indicators that could have been present in the sample. While many advertisements reported at least one indicator, few advertisements contained multiple indicators. No advertisements had 5 or more indicators and only 11 contained 4 indicators. Most sex workers’ advertisements (58%) displayed only 1 out of a total of 10 different indicators of human trafficking. In comparison, (Ibanez & Suthers, 2014) found that 15% of the advertisements (N=216) they examined had four or more indicators. Similarly, Skidmore et al. (2018) reported that all the groups in their analysis, i.e. sex workers connected to a unique phone number, presented at least some indicators. Although these studies are not wholly comparable due to differences in the datasets, methodology and selected indicators, they indicate a higher concentration of human trafficking indicators than we observed. We also want to underscore that our analysis found only two indicators that showed a high incidence, namely: *I4* – offering unprotected and/or violent sex; and *I8* – providing in-call services only. However, interpreting these two indicators is not straightforward.

While it is likely that victims of sexual exploitation cannot refuse to perform certain sex services, it is also true that many independent sex workers offer risky sex services. Unprotected sex is in high-demand among clients and thus offers an economic premium to those sex workers willing to provide it (Eccles & Clarke, 2014; Gertler et al., 2005; Quaipe et al., 2019). As aforesaid however, this indicator is solely based on advertisements, and sex workers may refuse to perform unprotected or violent sex when meeting clients. Similarly, the provision of in-call services only, which can signal that sex workers’ movements are being controlled or limited, calls for a critical analysis of how precise it is in terms of discriminating between real cases of human trafficking and independent sex workers. One could argue that while exploiters can force sex workers to only offer in-call services to limit their independence, sex workers can also independently make this decision, insofar as not meeting





**Fig. 8** Number of phone groups in the sample

clients in unfamiliar locations limits their exposure to risk. Hence, while these two indicators can highlight suspicious cases, they can also overestimate potential cases of human trafficking.

Interestingly, the three indicators used to proxy the presence of possible traffickers showed a low prevalence and were not correlated. There were 12, 140 and 178 advertisements that reported *I1*, *I2* and *I3*, respectively. *I2* (same phone number cited in more than one advertisement) and *I3* (high degree of similarity between sex workers' advertisements) co-occurred in only 2 advertisements. Conversely, we were able to identify that 140 sex workers were associated with 30 phone numbers. As Fig. 8 shows, the distribution of sex workers connected to the same phone number varied from a minimum of 2 to a maximum of 15. Half of the phone numbers (i.e., 15) had only 2 sex workers associated with it. However, it is interesting to note that there were groups with up to 15 (2 groups), 12 (1 group), 11 (1 group) and 10 sex workers (1 group) linked to a unique phone number. While we agree that multiple advertisements citing the same phone number raises suspicions and warrants further attention, we do not believe this constitutes a 'hard identifier' of sexual exploitation. Indeed, it is possible that independent sex workers freely decide to work 'together' and share a phone number, or that escort agencies and massage parlours also use this platform to advertise sex workers' services. While the same sex worker could have multiple advertisements, we checked for duplicates and can thus exclude this hypothesis. Once again, although potentially useful for detecting suspicious cases, this indicator also potentially encompasses cases involving genuine and independent sex workers.

The two indicators concerning sex workers' movements, *I6* and *I7*, also showed a low prevalence. There were 178 sex workers moving frequently between places, and only 11 sex workers moving in groups to the same location; they both co-occurred in

3 advertisements. While these indicators can be expedient for tracking some interesting human trafficking patterns, their presence or absence, deserves further attention. While movement-related indicators can correctly flag-up suspicious advertisements, they can also identify independent sex workers attempting to maximise their profits by moving to different locations (Teela Sanders et al., 2018). Conversely, exploited sex workers may remain in the same location for extended periods.

Finally, only a few advertisements used words alluding to youthful characteristics (19) or a dress size associated with underage girls, 643 (3.7%) and 98 (0.56%), respectively. Although these indicators may flag-up cases of underage sex workers, these advertisements could also merely be a genuine or deceitful response to the desire for youthfulness within the commercial sex industry. In their study of the UK sex market, Cunningham et al. (2018) showed that several advertising platforms now use stringent verification procedures to verify sex workers' identities, such as requiring a personal ID or proof of residency. However, this primarily targets migrant sex workers to ensure they are eligible to work (and to mitigate against the risk of police raids in response to Home Office pressures on illegal migration). It remains unclear if this approach is used to prevent underage people from joining these platforms.

There are two important findings emerging from this study. Firstly, that the examined platform does not contain many suspicious advertisements. While there are many advertisements presenting at least one indicator, a small percentage of the advertisements (1.70%) report three or more indicators, while none present more than 4 indicators. In recent years, several sources have noted how criminal organisations make use of digital platforms like advertising websites (Ibanez & Suthers, 2014; Latonero, 2011; Skidmore et al., 2018). Although traffickers use these technologies, many independent sex workers also use them for legitimate purposes. There is little to no evidence that these platforms can facilitate sexual exploitation, while it remains unproven that without such platforms, victims will no longer be exploited. Rather, there is mounting evidence that these platforms have positively impacted upon the sex market, by helping sex workers reach large numbers of clients, build their reputation, avoid risks associated with street sex work (e.g. arrests and violence), and aid them in performing preliminary checks on clients (Scott Cunningham & Kendall, 2011; Stewart Cunningham et al., 2018). Most studies on online sex work agree that the internet has had an overall positive impact on the sex market (Scott Cunningham and Kendall 2011; Campbell et al. 2018; Sanders and Platt 2017; Sanders et al. 2019). Shutting down websites advertising sex workers' services due to concern over traffickers' use of them has two potential negative consequences. First, traffickers can easily move to alternative platforms and continue to advertise trafficked sex workers (Heil & Nichols, 2014; Volodko et al., 2020). Second, it can further marginalise legitimate sex workers who lose out on the benefits of these platforms, and push them into more risky settings.

Perhaps more importantly, one of this study's main findings concerns the elusive nature of human trafficking indicators. While Volodko et al (2020) posit that some indicators are more useful than others, we find that it is hard to conclude which of these indicators are sufficiently reliable for identifying human trafficking given the dearth of empirical research in this area. This introduces the problem of 'false positives' and 'false negatives'. False positives are advertisements identified by an

indicator as signalling human trafficking, but which in fact are not. False negatives comprise actual cases of human trafficking that are not flagged-up by an indicator. Our approach is likely to suffer from both false positives and false negatives. For instance, *I4* identified around 70% of sex workers' advertisements in our sample, insofar as most of them reported performing some form of violent or risky sex practices. It is likely that many of these were false positive cases, which does not mean that *I4* is not a useful indicator of online human trafficking. While not of all those identified are victims, arguably all victims of human trafficking are forced to perform risky sex practices. Conversely, there are some indicators that can overlook instances of human trafficking. For instance, *I6—Sex workers move to different locations along with other sex workers*, overlooks all cases of exploitation in which victims are not trafficked between different places, and thus probably under-identifies cases of human trafficking.

False positives and false negatives raise problems in terms of developing interventions based on this approach. If we decided to administer more comprehensive screening tools targeting those advertisements that presented at least one indicator, then this would likely be overkill and ineffective. Here, as with other studies, most sex workers' advertisements contained at least one indicator of human trafficking, and, hence, administering more precise tools to all of them would be unfeasibly expensive. Conversely, targeting those with multiple indicators would only increase the number of false negatives. In fact, the higher the number of indicators chosen, the higher the number of false negatives there would be. This means that practitioners should set a threshold on the number of indicators that are required before engaging in more targeted interventions. Basing this threshold on sound empirical evidence would require further research into the precision and validity of human trafficking indicators, which is currently lacking.

## Conclusions

This study contributes to the study of the internet-trafficking nexus by identifying the presence of human trafficking indicators in a sample of 17, 362 UK-based sex workers' advertisements. It used an automated data-collection method and an objective and reproducible approach for identifying the presence of ten human trafficking indicators. In this respect, this study can help professionals flag-up suspicious advertisements, in conjunction with more precise tools, but that it cannot readily identify cases of human trafficking. Hence, the identification of these online indicators serves as a guide for quickly recognising and prioritising those cases that warrant further attention. Practitioners could then administer more comprehensive and precise tools to confirm if human trafficking is indeed occurring.

We believe that more empirical knowledge is required on the indicators of human trafficking to further progress research on the internet-trafficking nexus. We suggest two possible directions for future research. First, researchers should compile a list of empirically-tested human trafficking indicators. These variables need not be predictive or causal, but merely correlated, so that they can be identified via cross-sectional studies. In the UK, the National Referral Mechanism (NRM) – a framework

for identifying and referring potential victims of modern slavery and ensuring they receive the appropriate support – could provide the initial sample necessary for performing the analyses and identify significantly correlated variables. Once a list of empirically-tested correlates is compiled, subsequent analyses could aim to identify their presence within sex workers' online advertisements. Second, it would be instructive to compile a dataset of advertisements involving exploited sex workers. This would afford the possibility to identify differences between the advertisements of genuine sex workers and those including victims of human trafficking. Moreover, the availability of these data would also enable us to train machine learning algorithms to automatically identify advertisements that include common traits of those known to involve exploited sex workers and classify these as suspicious. We believe that either of these two approaches constitute a necessary step in the development of a methodology based on systematic research evidence, rather than the prevailing approach that is based on unclear professional judgment or arbitrariness.

## Appendix 1 – Human trafficking indicators

### Use of third- or first-person plural pronouns

We used grammatical tagging to identify verbs, pronouns and adjectives in advertisement descriptions. Grammatical tagging is used in linguistic studies to match words in a sentence to their corresponding parts-of-speech (Santorini, 1990). Tokenization (breaking up a sentence into words or tokens) and parts of speech tagging (attributing each token to parts of speech (POS) in the context of the sentence) are Natural Language Processing steps for attributing lexical and semantic meaning to a sentence. While tokenization splits a sentence into words (tokens), POS taggers assign each word to one of the 34 POS delineated by (Santorini, 1990), which are based on the function the word plays in the sentence. We applied the following steps to identify third- and first-person pronouns in each advertisement:

- We created a list of text documents (corpus) from the 17, 362 advertisement summary descriptions. Each text in our corpus corresponded to a description of a sex worker.
- We removed punctuation, emojis and special characters from each advertisement description.
- Further, we created a sequence of word tokens from each cleaned document and applied our POS tagger to each token. Therefore, each word, in each sex workers' advertisement was attributed to a POS.
- We removed all words from each advertisement that were not tagged as third- or first-person pronouns.

To identify third- and first-person pronouns, we use the compendium of Penn Treebank as a standard for speech tagging terms in our corpus (Santorini 1990). In this format of parts of speech tagging, pronouns are tagged as PRP and WP

(Personal Pronouns), PRP\$ and WP\$ (Possessive Pronouns). Therefore, we eliminate all tokens where the corresponding part of speech tag is not one of PRP, WP, PRP\$ or WP\$ retaining only third- and first-person pronouns.

### High degree of similarity between sex workers advertisements

To identify if multiple profile descriptions were written by the same person, we used Natural Language processing to estimate the cosine similarity between profile descriptions (McMahon & Hansen, 2018). The cosine similarity score measures the cosine of two arrays containing the word counts of two text documents (McMahon & Hansen, 2018). Given two documents  $A$  and  $B$ , we created word-frequency vectors  $x$  and  $y$  for  $A$  and  $B$ , respectively. We further estimated the cosine similarity between  $A$  and  $B$  as:

$$(\text{sim}(x, y) = \frac{x \cdot y}{\|x\| \|y\|})$$

where  $\|x\|$  and  $\|y\|$  refers to the Euclidean norm (Oghbaie & Mohammadi Zanjireh, 2018) of the vectors  $x$  and  $y$ . The score is different to Euclidean distance measures (McMahon & Hansen, 2015), insofar as it is not sensitive to the size of the documents. Therefore, the degree of linguistic similarity between two advertisements can be determined irrespective of the length of the advertisement. This score is on a range of 0 (No similarity between two texts) through 1 (Exactly the same text).

We created an indicator for this by thresholding the cosine similarity score for ‘significant similarity’ at 0.90. To create binary values for each sex worker, we assigned a value of 1 to a sex worker who had a similarity score above 0.90.

### Advertisements promoting inexpensive sex services

Outliers are observations that are different or vary significantly from most of the data. The interquartile range shows how the data is spread around the median and can be used to detect outliers i.e. points that do not fall within this range. The interquartile range was estimated by subtracting the first quartile from the third quartile  $\text{IQR} = Q_3 - Q_1$ . Since the interquartile range indicates how spread out the middle half of the data is, certain data points that fall considerably outside this range are classified as outliers. Tukey (1977) defines the 1.5 IQR rule technique for detecting outliers. The rule states that a data point is an extreme outlier if it is either more than  $1.5 * \text{IQR}$  above the third quartile or below the first quartile, given that the centre of the interval is  $(Q1 + Q3)/2$  and its radius is  $3.5 * \text{IQR}$ . We defined extreme outliers below the first quartile by applying the following steps:

- Estimated the IQR for in-call and outcall service charges.
- Estimated the outlier determining constant  $e$  by multiplying the IQR by 1.5.
- Estimated  $Q1 - e$  for extreme lower bound outliers

Our results are shown in the table above, with £25 standing for the lower bound outlier cut-off for in-call prices and £50 standing for the lower bound outlier cut-off for outcall prices. Therefore, all sex workers with in-call prices below £25 were flagged-up as extreme outliers, while sex workers with outcall prices below £50 were flagged-up as extreme outliers

Of all the sex workers who advertised in-call or outcall per hour prices, 44% of the sex workers' advertisements did not offer in-call services while 46% of the advertisements did not offer outcall prices. None of the 3, 088 sex workers' advertisements that offered in-call services met the criteria for lower extreme outliers, while there were only four ads with outcall prices below £50

**Acknowledgements** The authors would like to thank Mike Levi, Nicolas Trajtenberg and Giulia Berlusconi for their comments on previous drafts of this article.

**Funding** This work was supported by the Economic and Social Research Council under Grant ES/S008853/1

## Declarations

**Disclosure statement** The authors declare that they have no conflicts of interest.

**Data deposition** The data that support the findings of this study are available on Figshare: <https://figshare.com/projects/CyberTNOC/94890>

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Alvari, H., Shakarian, P., & Snyder, J. E. K. (2016). A Non-Parametric Learning Approach to Identify Online Human Trafficking. ArXiv:1607.08691
- Alvari H, Shakarian P, Snyder JEK (2017) Semi-supervised learning for detecting human trafficking. Security Informatics 6(1):1. <https://doi.org/10.1186/s13388-017-0029-8>
- Bounds D, Delaney KR, Julion W, Breitenstein S (2020) Uncovering Indicators of Commercial Sexual Exploitation. J Interpers Violence 35(23–24):5607–5623. <https://doi.org/10.1177/0886260517723141>
- Burbano D, Hernandez-Alvarez M (2017) Identifying human trafficking patterns online. IEEE Second Ecuador Technical Chapters Meeting (ETCM) 2017:1–6. <https://doi.org/10.1109/ETCM.2017.8247461>
- Campbell, R., Aydin, Y., Cunningham, S., Hamer, R., Hill, K., Melissa, C., Pitcher, J., Scoular, J., Sanders, T., & Valentine-Chase, M. (2018). Technology mediated sex work: Fluidity, networking & regulation in the UK. In S. Dewey, I. Crowhurst, & C. Izugbara (Eds.), *Routledge International*

- Handbook of Sex Industry Research* (pp. 533–543). Routledge. <http://eprints.whiterose.ac.uk/140432/>
- Cunningham, Scott, & Kendall, T. D. (2011). Prostitution 2.0: The changing face of sex work. *Journal of Urban Economics*, 69(3), 273–287. <https://doi.org/10.1016/j.jue.2010.12.001>
- Cunningham S, Sanders T, Scoular J, Campbell R, Pitcher J, Hill K, Valentine-Chase M, Melissa C, Aydin Y, Hamer R (2018) Behind the screen: Commercial sex, digital spaces and working online. *Technol Soc* 53:47–54. <https://doi.org/10.1016/j.techsoc.2017.11.004>
- Dubrawski A, Miller K, Barnes M, Boecking B, Kennedy E (2015) Leveraging Publicly Available Data to Discern Patterns of Human-Trafficking Activity. *Journal of Human Trafficking* 1(1):65–85. <https://doi.org/10.1080/23322705.2015.1015342>
- Eccles C, Clarke J (2014) Levels of advertised unprotected vaginal and oral sex by independent indoor female sex workers in West Yorkshire. *UK Sexually Transmitted Infections* 90(1):36–37. <https://doi.org/10.1136/sextrans-2013-051150>
- Gertler P, Shah M, Bertozzi SM (2005) Risky Business: The Market for Unprotected Commercial Sex. *J Polit Econ* 113(3):518–550. <https://doi.org/10.1086/429700>
- Heil E, Nichols A (2014) Hot spot trafficking: A theoretical discussion of the potential problems associated with targeted policing and the eradication of sex trafficking in the United States. *Contemporary Justice Review* 17(4):421–433. <https://doi.org/10.1080/10282580.2014.980966>
- Home Office. (2016). *Modern Slavery Act 2015 – Statutory Guidance for England and Wales*. Home Office. <https://www.gov.uk/government/publications/modern-slavery-how-to-identify-and-support-victims>
- Hultgren, M., Whitney, J., Jennex, M., & Elkins, A. (2018). A Knowledge Management Approach to Identify Victims of Human Sex Trafficking. *Communications of the Association for Information Systems*, 42(1). <https://doi.org/10.17705/1CAIS.04223>
- Ibanez, M., & Gazan, R. (2016). *Virtual indicators of sex trafficking to identify potential victims in online advertisements*. 818–824. <https://doi.org/10.1109/ASONAM.2016.7752332>
- Ibanez, M., & Suthers, D. D. (2014). Detection of Domestic Human Trafficking Indicators and Movement Trends Using Content Available on Open Internet Sources. *2014 47th Hawaii International Conference on System Sciences*, 1556–1565. <https://doi.org/10.1109/HICSS.2014.200>
- Kennedy, E. (2012). *Predictive Patterns of Sex Trafficking Online* [Thesis, Carnegie Mellon University]. <https://doi.org/10.1184/R1/6686309.v1>
- Latonero, M. (2011). *Human Trafficking Online: The Role of Social Networking Sites and Online Classifieds* (SSRN Scholarly Paper ID 2045851). Social Science Research Network. <https://papers.ssrn.com/abstract=2045851>
- Latonero M (2012) The Rise of Mobile and the Diffusion of Technology-Facilitated Trafficking. *SSRN Electron J*. <https://doi.org/10.2139/ssrn.2177556>
- Li L, Simek O, Lai A, Daggett M, Dagli CK, Jones C (2018) Detection and Characterization of Human Trafficking Networks Using Unsupervised Scalable Text Template Matching. *IEEE International Conference on Big Data (big Data) 2018*:3111–3120. <https://doi.org/10.1109/BigData.2018.8622189>
- Malo, S. (2018, February 1). Is the Super Bowl really the U.S.'s biggest sex trafficking magnet? *Reuters*. <https://uk.reuters.com/article/us-football-nfl-superbowl-trafficking-an-idUSKBN1FL6A1>
- Martin, L., & Hill, A. (2019). Debunking the Myth of ‘Super Bowl Sex Trafficking’: Media hype or evidenced-based coverage. *Anti-Trafficking Review*, 13, 13–29. <https://doi.org/10.14197/atr.201219132>
- McMahon, M., & Hansen, S. (2015). A Primer on Text Mining for Economists Introduction Traditional focus in ( monetary ) economics on quantitative information. *Presentation*.
- McMahon, M., & Hansen, S. (2018). *A Primer on Text Mining for Economists*. Economic and Social Research Council, University of Warwick. [https://nanopdf.com/download/a-primer-on-text-mining-for-economists-preliminaries-pre-processing\\_pdf](https://nanopdf.com/download/a-primer-on-text-mining-for-economists-preliminaries-pre-processing_pdf)
- Menskivova, A., & Mattmann, C. A. (2018). Ensemble sentiment analysis to identify human trafficking in web data. *Workshop on Graph Techniques for Adversarial Activity Analytics (GTA 2018)*, Marina Del Rey, CA, USA, 5–9.
- Oghbaie, M., & Mohammadi Zanjireh, M. (2018). Pairwise document similarity measure based on present term set. *Journal of Big Data*, 5(1). <https://doi.org/10.1186/s40537-018-0163-2>
- Polaris Project. (2012). *Domestic Sex Trafficking: The Criminal Operations of the American Pimp*. Polaris Project. <https://www.dcs.virginia.gov/sites/dcs.virginia.gov/files/publications/victims/domestic-sex-trafficking-criminal-operations-american-pimp.pdf>

- Portnoff RS, Huang DY, Doerfler P, Afroz S, McCoy D (2017) Backpage and Bitcoin: Uncovering Human Traffickers. *KDD* DOI 10(1145/3097983):3098082
- Quaife M, Lépine A, Deering K, Terris-Prestholt F, Beattie T, Isac S, Paranjape RS, Vickerman P (2019) The cost of safe sex: Estimating the price premium for unprotected sex during the Avahan HIV prevention programme in India. *Health Policy Plan* 34(10):784–791. <https://doi.org/10.1093/heapol/czz100>
- Sanders, T., & Platt, L. (2017). *Is sex work still the most dangerous profession? The data suggests so.* <https://researchonline.lshmt.ac.uk/4647656/>
- Sanders, T., Scoular, J., Pitcher, J., Campbell, R., & Cunningham, S. (2019). *Beyond the Gaze and Well Beyond Wolfenden and: The practices and rationalities of regulating/ policing sex work in the digital age.* <https://ira.le.ac.uk/handle/2381/43762>
- Sanders, Teela, Scoular, J., Campbell, R., Pitcher, J., & Cunningham, S. (2018). *Internet Sex Work.* <https://www.palgrave.com/gp/book/9783319656298>
- Sanders, Teela, Vajzovic, D., Brooks-Gordon, B., & Mulvihill, N. (2020). Policing vulnerability in sex work: The harm reduction compass model. *Policing and Society*, 0(0), 1–17. <https://doi.org/10.1080/10439463.2020.1837825>
- Santorini, B. (1990). Part-of-Speech Tagging Guidelines for the Penn Treebank Project (3rd Revision). *University of Pennsylvania 3rd Revision 2nd Printing*, 53(MS-CIS-90–47), 33. <https://doi.org/10.1017/CBO9781107415324.004>
- Shahrokh Esfahani, S., Cafarella, M. J., Baran Pouyan, M., DeAngelo, G., Eneva, E., & Fano, A. E. (2019). Context-specific Language Modeling for Human Trafficking Detection from Online Advertisements. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 1180–1184. <https://doi.org/10.18653/v1/P19-1114>
- Skidmore, M., Garner, S., Desroches, C., & Saggiu, N. (2018). The Threat of Exploitation in the Adult Sex Market: A Pilot Study of Online Sex Worker Advertisements. *Policing: A Journal of Policy and Practice*, 12(2), 210–218. <https://doi.org/10.1093/police/pax007>
- Szekely, P., Knoblock, C. A., Slepicka, J., Philpot, A., Singh, A., Yin, C., Kapoor, D., Natarajan, P., Marcu, D., Knight, K., Stallard, D., Karunamoorthy, S. S., Bojanapalli, R., Minton, S., Amanatullah, B., Hughes, T., Tamayo, M., Flynt, D., Artiss, R., ... Ferreira, L. (2015). Building and Using a Knowledge Graph to Combat Human Trafficking. In M. Arenas, O. Corcho, E. Simperl, M. Strohmaier, M. d'Aquin, K. Srinivas, P. Groth, M. Dumontier, J. Heflin, K. Thirunarayan, & S. Staab (Eds.), *The Semantic Web—ISWC 2015* (pp. 205–221). Springer International Publishing. [https://doi.org/10.1007/978-3-319-25010-6\\_12](https://doi.org/10.1007/978-3-319-25010-6_12)
- Tong, E., Zadeh, A., Jones, C., & Morency, L.-P. (2017). Combating Human Trafficking with Deep Multimodal Models. *ArXiv:1705.02735*
- Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2). Reading, MA.
- UNODC. (2008). *Global report on trafficking in persons*. United Nation Office on Drugs and Crime. [https://www.unodc.org/documents/Global\\_Report\\_on\\_TIP.pdf](https://www.unodc.org/documents/Global_Report_on_TIP.pdf)
- UNODC. (2018). *Global report on trafficking in persons*. United Nation Office on Drugs and Crime. <https://www.unodc.org/unodc/data-and-analysis/glotip.html>
- UNODC. (2020). *Human Trafficking Indicators*. United Nation Office on Drugs and Crime. [https://www.unodc.org/pdf/HT\\_indicators\\_E\\_LOWRES.pdf](https://www.unodc.org/pdf/HT_indicators_E_LOWRES.pdf)
- Volodko A, Cockbain E, Kleinberg B (2020) “Spotting the signs” of trafficking recruitment online: Exploring the characteristics of advertisements targeted at migrant job-seekers. *Trends in Organized Crime* 23(1):7–35. <https://doi.org/10.1007/s12117-019-09376-5>
- Wang H, Cai C, Philpot A, Latonero M, Hovy E, Metzler D (2012). Data Integration from Open Internet Sources to Combat Sex Trafficking of Minors. <https://doi.org/10.1145/2307729.2307769>
- Wen J, Klarin A, Goh E, Aston J (2020) A systematic review of the sex trafficking-related literature: Lessons for tourism and hospitality research. *J Hosp Tour Manag* 45:370–376. <https://doi.org/10.1016/j.jhtm.2020.06.001>
- Whitney, J., Hultgren, M., Jennex, M., Elkins, A., & Frost, E. (2020). *Using Knowledge Management and Machine Learning to Identify Victims of Human Sex Trafficking* (pp. 360–389). <https://doi.org/10.4018/978-1-7998-2355-1.ch014>

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.