

The Changing Meaning Of “No” In Canadian Sex Work

Advertising: Peace of Mind, Safety and Race

Lynn Kennedy^{1*}

¹Sex Work Population Project <https://populationproject.ca/>

*corresponding author

Email address: lynn.kennedy@populationproject.ca

Abstract

Most previous research into the health and safety practices of sex workers is based on questionnaire and interview data from relatively small samples. Large corpora of online advertising can help identify general trends that might be missed in studies with more limited inputs. However, extracting usable information from a large corpus can be challenging. This study addresses these issues by looking at how the word "no" is used in Canadian sex work advertising. Ads were collected during three periods: April 2007 to April 2009 (385729 ads), November 2014 to December 2016 (2951642 ads), and September 2021 to September 2022 (887698 ads). Common terms surrounding the word "no" were identified for each period. Five major themes, associated with 4-25% of workers, were identified: *peace of mind*, *client behavior*, *communication*, *client ethnicity*, and *service restrictions*. Collectives, those advertising in multiple regions, and cis-females were more likely to use "no". Over time *client behavior*, *service restrictions*, and *client ethnicity* became more prominent. Restrictions based on *client ethnicity* were not significant in 2007-2009 but were associated with 12% of advertisers in 2021-2022. More research is needed to understand the experience of sex workers and clients of color.

Introduction

A substantial body of research now exists documenting the adoption of the internet by contact sex workers in industrialized democracies (Argento et al. 2018; Minichiello, Scott, and Callander 2013; Sanders et al. 2018; Sanders, Connelly, and King 2016) and strategies used by

contact sex workers who advertise online to enhance occupational health and safety (Bernier et al. 2021; Bungay and Guta 2018; Cunningham and Kendall 2010; Jiao, Bungay, and Jenkins 2021; Kille et al. 2017; Kolar, Atchison, and Bungay 2014; Machat et al. 2022; Strohmayer, Clamen, and Laing 2019). However, most of this research is based on questionnaire and interview data from relatively small sample populations or small collections of websites or online ads. These samples can be further constrained by geographic location, demographics, long data collection periods, and convenience sampling strategies potentially reducing the representativeness of the research (Kennedy 2022).

There are exceptions to this rule. (Boecking et al. 2015, 2018) analyzed 16 million and 37 million ads from the US and Canada respectively from one prominent classified ads site to analyze temporal and geographic trends in advertising and (Kennedy 2022) analyzed 3.6 million ads from six prominent classified ad sites used by Canadian workers to estimate sex worker population dynamics. (Cunningham and Kendall 2010) analyzed metadata from over 90000 worker profiles on one popular sex work review site in the US and followed this up with a survey of 685 workers to better understand the health and safety practices of workers over multiple years. Clearly, there are opportunities to create a more representative picture of the communication practices of sex workers using publicly available online advertising data. However, given the technical challenges of research using document analysis techniques on large document corpora, few studies have used large samples of online data to better understand the health and safety practices of workers in industrialized democracies.

As a safety strategy, advertising online significantly reduces risk (Argento et al. 2018; Machat et al. 2022; Sanders et al. 2018) and in industrialized democracies, web and telecommunications technologies have become the dominant way in which workers interact with

clients. Contact sex workers who advertise online employ multiple strategies to improve occupational health and safety. Workers describe investing considerable effort in getting to know clients before any in-person contact is initiated (Argento et al. 2018; Bungay and Guta 2018; Jiao et al. 2021) and this can include more formal screening processes in some cases. Workers also reduce risk with careful arrangement of the physical environment where in-person contact takes place (Bungay and Guta 2018). Information sharing between workers is another important safety strategy (Strohmayr et al. 2019) and how workers share information has changed as the internet has become the dominant way that workers advertise (Argento et al. 2018; Bungay and Guta 2018; Strohmayr et al. 2019). Communication in advertising plays an important part in this process as a critical first step in the worker-client relationship (Kille et al. 2017).

This study takes a novel approach to understanding health and safety communications in sex work advertising by examining how the word "no" is used in three large Canadian corpora of online classified ads. The impetus for this work was the result of reviewing over ten thousand ads as part of the process of error estimation for the worker population estimates reported in (Kennedy 2022). The usage of the word "no" emerged as a powerful expression of personal limits and preferences used by many workers.

By considering the context in which a single word is present, it is possible to greatly simplify the analysis of a large number of documents. Bigrams (Jurafsky and Martin 2023), combinations of two words extracted from larger texts, are a common way to reduce the complexity of linguistic analysis in large corpora. These small groups of words can be analyzed for statistical trends far more easily than full documents. In this work, frequently used bigrams beginning with “no” were combined with the most common words preceding and following the bigrams to generate a relatively small number of meta-documents that could be analyzed using

standard grounded theory techniques (Corbin and Strauss 1990).

Ads could be authored by the workers themselves or third parties. Similar to (Cunningham and Kendall 2010), this study uses the abstract entity of an *advertiser* (Kennedy 2022) as a meaningful way to group ads. Advertisers are associated with demographic variables based on advertising data which can be incorporated into the analysis. The use of advertisers as a starting point also reduces bias introduced when individuals produce excessively large numbers of ads. Source documents from three distinct periods (2007-2009, 2014-2016, and 2021-2022) were examined using the variables of region, social context (whether advertisers advertise individually or collectively), gender, and ethnicity to better understand who uses "no" and how this usage has changed over time.

The advantage of this approach, compared to reviews of relatively small random samples of documents, is that the discovered themes are more likely to be relevant. The three corpora likely represent more than 100000 advertisers, who in turn represent an even larger number of workers (Kennedy 2022). Furthermore, compared to convenience sampling often used in other qualitative sex work research, self-selection, and social desirability biases are avoided as the source material is generated by advertisers in a naturalistic setting. This combination of statistical document analysis at scale and qualitative techniques yielded insights that were sometimes surprising into how occupational health and safety strategies evolved over the 15 year period.

Materials and Methods

This study finds evidence for common themes in sex work advertising surrounding how the word "no" is used. The variables of time, region, gender, ethnicity, and social context are considered for the frequency and context of the use of the word "no" by advertisers.

Source materials for the analysis were advertising texts from online classified ads. These ads were collected from six prominent online classified advertising sites described in (Kennedy 2022) during three time periods: April 1, 2007 to March 31, 2009 inclusive, November 1, 2014 to December 31, 2016 inclusive, and September 15, 2021 and September 22, 2022 inclusive. The ad collections for 2014-2016 and 2021-2022 were nationally representative. The 2007-2009 collection is representative of one western Canadian province. The term advertisers refers to unscaled advertisers identified either from contact information in 2007-2009 and 2014-2016 or internal chat ids in 2021-2022 using methods described in (Kennedy 2022).

To prepare the ads, advertisers were extracted and ad text was converted to lowercase ASCII and all non-alphanumeric characters were removed except apostrophes as described in (Kennedy 2022). After the cleaning process, ad texts were scanned and single words (unigrams), pairs of words (bigrams), and triplets of words (trigrams) were stored in a MariaDB database (MariaDB and Widenius 2017) along with a count of the number of advertisers and ads using each.

The 100 most frequently used "no" bigrams for each period were used as the basis for thematic analysis. Three groups of 100 files containing the most common bigrams and common context words were generated as inputs, one for each collection period. These files can be found on the Supplemental materials website (<https://osf.io/download/k85pf/>). QualCoder (Curtain 2023) was used to code underlying themes based on the generated files. For each identified code,

advertiser frequencies were tallied and the relative rank of each theme within each period was calculated.

To better understand how advertiser communication around safety has changed, the terms “safe play”, “safe service”, “safe services”, “safe gfe”, “screening” and “deposit required” were also searched and advertiser frequencies were calculated. To better understand attitudes surrounding race, term frequencies of common racial epithets and the terms “black”, “brown”, “thug” and “pimp” were tabulated in the 2014-2016 corpus. These were used to calculate joint probabilities of the terms “black”, “brown” and “thug” or “pimp”.

The following statistical measures were generated. For each code, the proportion of advertisers associated with the code was calculated and segmented by five variables: social context (collective versus individual advertisers), period, region, self-identified gender, and self-identified ethnicity (2021-2022 only). These proportions represent the probability that an advertiser fitting a specific demographic category was associated with a given code (e.g. $p(\text{communication} = \text{True} \mid \text{gender} = \text{male})$). For each code, the differences between advertiser proportions for each of the demographic variables and the proportions for all advertisers were tested for significance using the R *prop.test* function (R Core Team 2021). The R *lm* function was used to create univariate and multivariate models of the proportion of advertisers by year and the proportion of advertisers by year and code respectively (R Core Team 2021). Two measures of advertising behavior were compared for advertisers using “no” and those who did not: days from the first ad to the last ad (duration) and ad frequency. Advertised hourly prices were compared for advertisers who restrict clients based on ethnicity and the general population. In both cases significance was tested using Welch’s modified two sample t-test, *tsum.test* from the R BSDA package (Arnholt and Evans 2021). R version 4.2.2 was used to perform all tests

using a 95% confidence level.

Ethics statement

All source data used in this study consisted of publicly available data at the time it was collected and was collected in accordance with the policies of the sites in effect at the time. The methods used are conformant with the ethical standards of the Canadian Sociology Association (section 4.10 II) and the American Sociology Association (section 10.5 c) (American Sociological Association 2018; CSA-SCS Policy, Ethics, and Professional Concerns Subcommittee 2018). As the replicability of the main results of this paper is important, a data set is provided as part of the supporting information along with the code used to process it. However, to protect the safety and privacy of advertisers and third parties, all identifying information has been removed including the names of the source websites.

Results

Table 1 summarizes the advertising data in three corpora covering a 15 year period from April 2007 to September 2022 inclusive. Overall 4225069 ads and associated metadata were used (2007-2009: 385729 ads, 2014-2016: 2951642 ads, 2021-2022: 887698 ads). Of these, 39% (N=1628698) were ads that contained the word “no”. Most advertisers (54% or 114842 out of 214458) used the word “no” in at least one ad.

Themes and codes

A total of 168 unique "no" bigrams were represented in the top 100 bigrams from every period. A total of 14 codes were identified from the analysis of these bigrams and context words using Qualcoder (Curtain 2023). Table 2 shows the number of advertisers, the most common trigrams, and the affect associated with each identified code. The vast majority of "no" advertisers are associated with both the top 100 bigrams for each period (92%, N=105838) and the top 5 codes (88%, N=100688). Mean bigrams per advertiser were 2.3 (SD 2.1) and mean codes were 1.8 (SD 1.1). For advertisers using the word "no", on average 49% (SD 33%) of the associated ads contained the word "no". This varied somewhat between corpora with 2007-2009 averaging 45% (SD 32%), 2014-2016 averaging 46% (SD 33%), and 2021-2022 averaging 56% (SD 34%).

Table 2: Codes relating to the usage of the word “no” based on the top 100 bigrams for each period. Advertiser counts represent all periods. Affect refers to whether the code refers to restrictions (-), inducements (+), or neutral statements (=).

codes	advertisers	top trigrams	affect
<i>peace of mind</i>	54084 (25%)	look no further,no rush no	+
<i>communication</i>	47130 (22%)	no blocked calls,_NNNN_ no blocked	=
<i>client ethnicity</i>	32612 (15%)	no blacks no,_NNNN_ no black	-
<i>client behavior</i>	23863 (11%)	calls no explicit,no explicit talk	-
<i>service restrictions</i>	8545 (4%)	no greek no,only no bare	-
<i>service location</i>	3532 (2%)	no out calls,only no car	=
<i>services offered</i>	2786 (1%)	no restrictions _NN_,service no restrictions	+
<i>no pictures</i>	2713 (1%)	accurate no face,no pic messaging	=

<i>pimps or law enforcement</i>	2679 (1%)	no thugs no,only no thugs	-
<i>service time</i>	1426 (1%)	no hh no,_N_h no half	=
<i>employment</i>	988 (<1%)	no experience necessary,_NN_ no experience	+
<i>client age</i>	917 (<1%)	no no young,only no young	-
<i>appearance</i>	294 (<1%)	have no tattoos,no no tattoos	+
<i>payment</i>	169 (<1%)	no e transfers	=

The top 5 themes were coded as *communication*, *peace of mind*, *client behavior*, *client ethnicity*, and *service restrictions*. Less common themes used by 2% or fewer advertisers were coded as *services offered*, *no pictures*, *pimps or law enforcement*, *service time*, *employment*, *client age*, *appearance*, and *payment*. The *communication* code related to restrictions on how the advertiser wished to be contacted including requiring caller id for phone calls. *Peace of mind* refers to “no” statements that were intended to reassure prospective clients. *Client behavior* refers to restrictions related to etiquette. *Client ethnicity* refers to restrictions based on race or ethnicity. *Service restrictions* refers to statements regarding disallowed types of services - the inverse of *services offered*. Other service related themes were *payment*, *service location*, and *service time* which described restrictions on methods of payment, venues, and service duration respectively. In addition to *client ethnicity*, *client age*, and *pimps or law enforcement* are restrictions on who can contact the advertiser and could be said to overlap with *client behavior*. The *no pictures* and *appearance* themes relate to restrictions on presentation in the ad and presentation in person respectively. Finally, *employment* is unique to advertisers looking for prospective employees rather than clients.

Codes could have a negative, positive, or neutral affect. Affect refers to how a reader might potentially react to a given “no” bigram. Positive affect refers to codes that attempt to

reassure or induce prospective clients in some way, negative affect refers to codes that discuss risky or offensive client behavior or limits who can respond to the ad and neutral affect refers to codes that are simple declarative statements. Distributions of advertisers using “no” by affect were similar. Positive affect was associated with 26% (N=56629) of these advertisers, negative affect was associated with 25% (N=52946) and neutral affect was associated with 25% (N=54068).

Differences in advertising behavior were found between advertisers who used the word “no” and others. In all cases, advertisers who used the word "no" advertised more frequently and were found to advertise longer than advertisers who did not. For all periods the mean duration between the first and last ad and the number of ads for advertisers using "no" was 118 days (SD 172) and 30 ads (SD 264) respectively. In contrast, advertisers who did not use “no” were less active; they advertised for 75 days (SD 131) and posted 8 ads (SD 67) on average. Welch’s modified two sample t-test indicated that the differences between the “no” group and other advertisers were significant for both duration ($t = -65.731$, $df = 211139$, $p < 0.001$) and ads posted ($t = -27.317$, $df = 131888$, $p < 0.001$).

The following sections describe how the variables of period, social context (advertising individually versus collectively), region, self-identified gender, and self-identified ethnicity are associated with the 5 most prevalent codes. In all cases the R *prop.test* function showed that the overall proportions of subgroups of advertisers using “no” were significantly different from the proportions of all advertisers using “no” ($p < 0.001$). Except where noted, the proportions of advertiser subgroups associated with individual codes were also significantly different from the proportion of all advertisers associated with that code ($p < 0.001$).

Time

The proportion of advertisers using "no" was 54% (N=114842) and ranged from 51% (N=19876) in 2021-2022 to 54% (N=90820) in 2014-2016. A univariate model of the proportion of advertisers using "no" and year was not significant ($p = 0.7529$). Table 3 shows the distribution of advertisers by period.

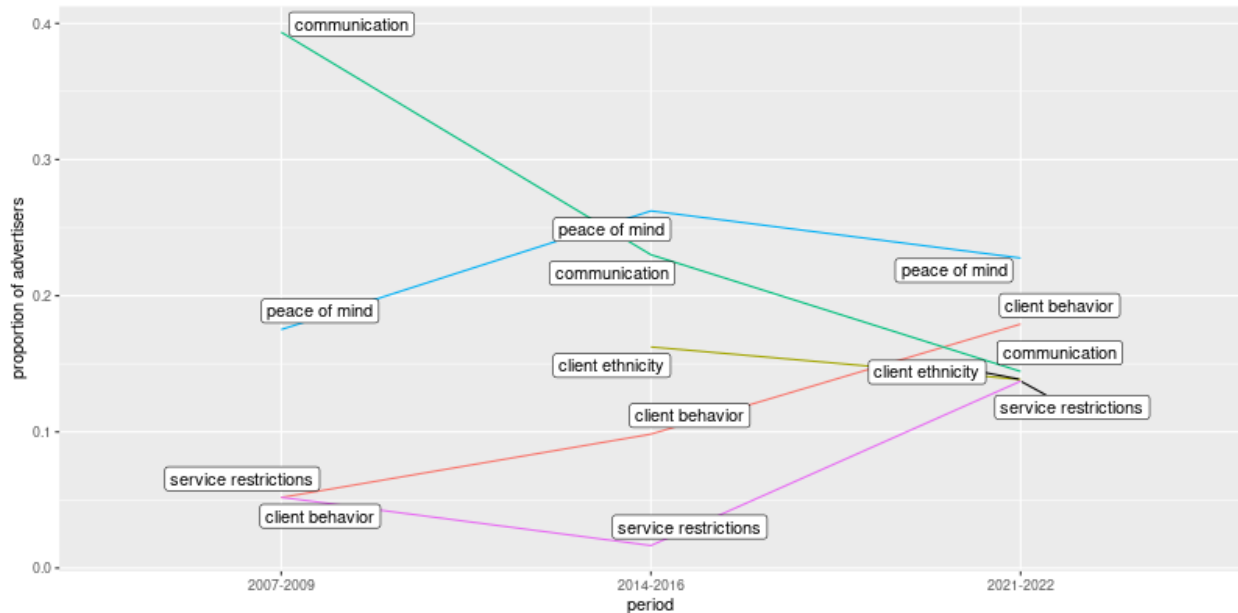
Table 3: Advertisers by time period.

period	advertisers	using "no"
2007-2009	7939	4146 (52%)
2014-2016	167539	90820 (54%)
2021-2022	38980	19876 (51%)
all	214458	114842 (54%)

The relative ranks of the top 5 codes changed significantly between periods. A multivariate model of the proportion of advertisers using "no" to a combination of year and code created with the R `lm` function was significant ($p < 0.001$). Testing using the R `prop.test` function also showed that the proportions of advertisers for each period were significantly different from the overall proportion for all periods for each of the top 5 codes. What is notable is the drop in the rank of *communication* and the increase in the rank of *client behavior*, *service restrictions*, and *client ethnicity*. *Client ethnicity* did not appear in the top 100 bigrams for 2007-2009 but is ranked fourth by 2021-2022. Figure 2 shows that while service restrictions dropped in rank the proportion of advertisers associated is higher in 2021-2022 (14% of all advertisers) than in 2007-2009 (5% of all advertisers). Indeed, except for *communication*, more advertisers were using the top 5 codes in 2021-2022 (40% = 15551/38980) than in 2007-2009 (22% =

1776/7939).

Figure 2: Proportions of advertisers using the top 5 codes by time period.



Safety-related messaging generally became more common between 2007 and 2022. In addition to safety related "no" terms, the usage of the word "safe" itself provided further evidence for this trend. To disambiguate other uses of the word "safe", counts of advertisers using the common bigrams "safe service", "safe services", "safe gfe" and "safe play" were made for each period. Over all periods 6% (N=16127) of advertisers used at least one of these "safe" bigrams. In 2007-2009 these bigrams were not common with only 1.3% (N=105) of advertisers using them. However, in 2014-2016 this increased to 7% of advertisers (N=12166) and 10% (N=3856) in 2021-2022. Most advertisers using these "safe" bigrams also used "no" (84%, N=13522 all periods).

Safety messaging was not limited to health. The terms “deposit required” and “screening” were less commonly used but represent important risk reduction strategies. These became more prevalent between 2007 and 2022. In 2007-2009 0% (N=2) of advertisers used the bigram

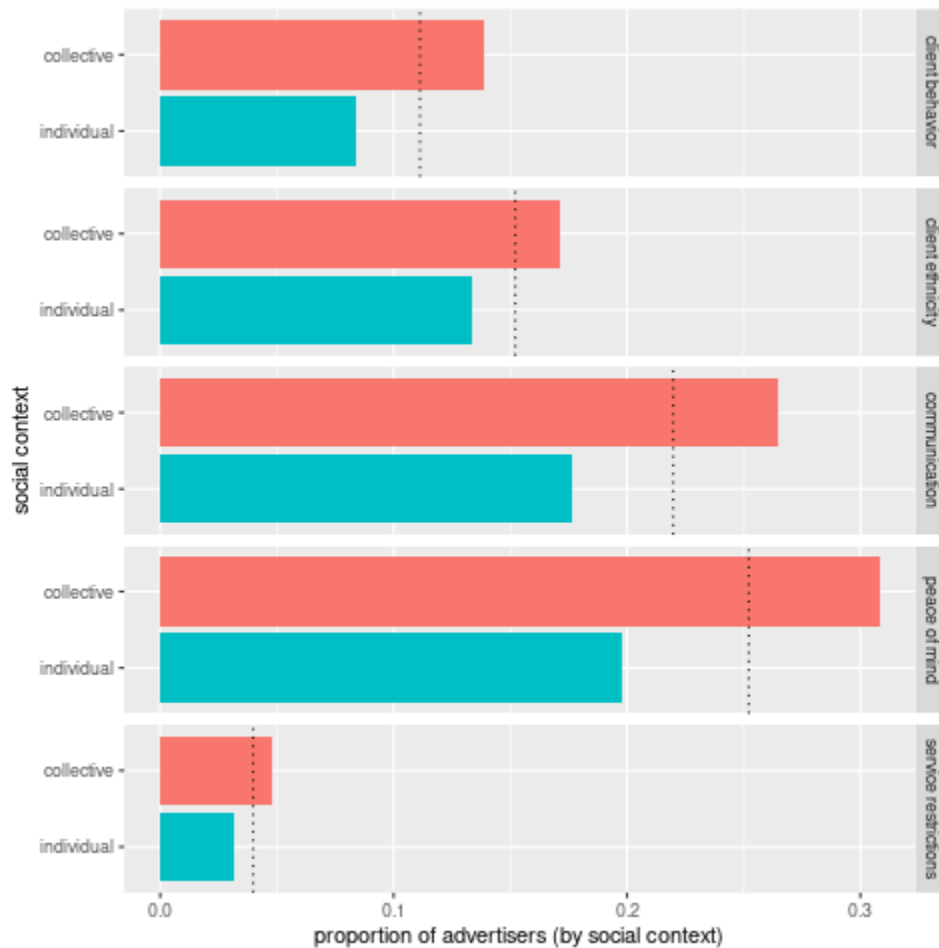
“deposit required” however by 2021-2022 usage increased to 4% (N=1725). Similarly, the term “screening” in 2007-2009 was used by 0% (N=8) of advertisers but in 2021-2022 this grew to 1.3% (N=522).

Social context

Social context, or whether an advertiser advertised as an individual or a collective as defined in (Kennedy 2022), was associated with whether that advertiser used “no” in ads. An estimated 50% of advertisers (N=107207) were advertising as individuals with 44% (N=47305) using “no” in ads. However, 63% (N=67537) of collective advertisers (N=107251) used “no” at least once. Over time, the proportions of individual advertisers decreased from an estimated 49% (N=3861) in 2007-2009 to 41% (N=15847) in 2021-2022.

Figure 3 shows the proportions of advertisers using the top 5 codes based on social context. These represent the probability, given that the advertiser was found in the given context, that the advertiser is associated with the given code. All proportions of advertisers were significantly different from the overall proportion for each code.

Figure 3: Prevalence of top 5 codes relative to social context. The dotted lines indicate the proportion of all advertisers associated with the code.



Region

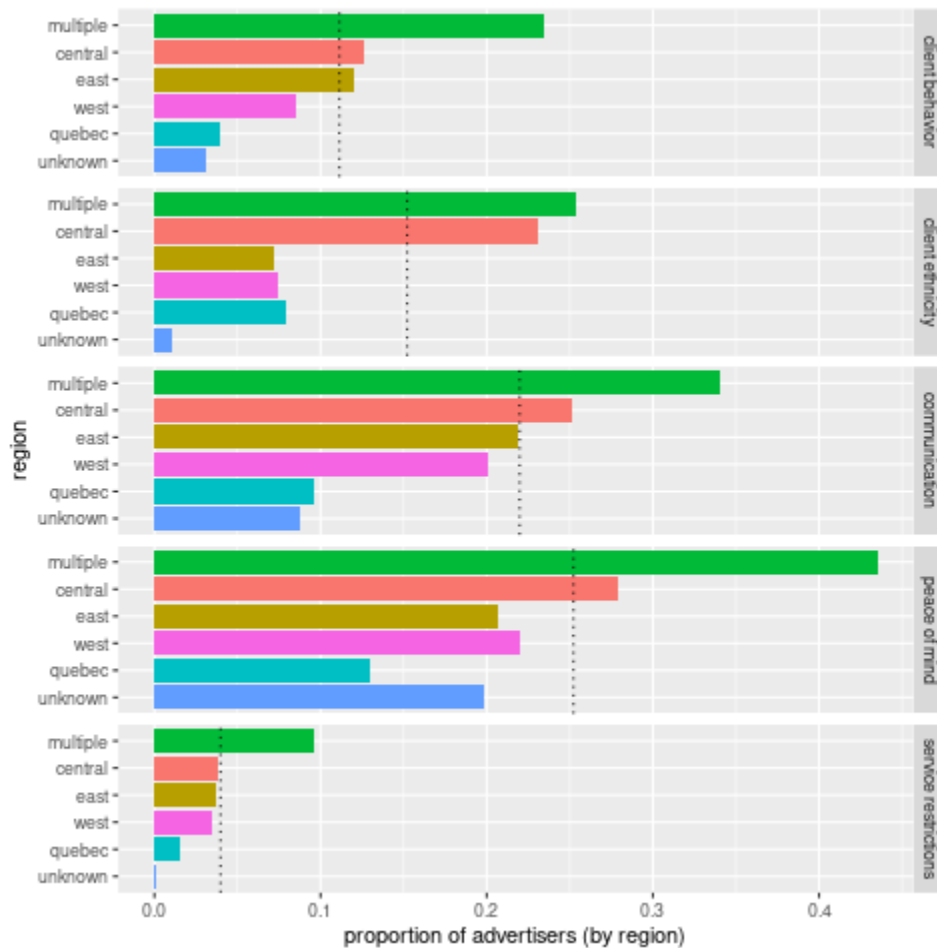
Table 4 shows the distribution of advertisers by region. More advertisers (75%, N=17641) associated with multiple provinces, such as those who tour, used the word “no” compared to advertisers who only advertised in a single province (51%, N=97201). Apart from those advertising in multiple provinces, the central region (61%, N=48491) and advertisers with no geographic information (56%, N=1734) were more likely than average to use “no”.

Table 4: Advertisers by region.

region	provincial codes	advertisers	using “no”
<i>multiple</i>	multi-province and the north (nw,nv,yk)	23676	17641 (75%)
<i>central</i>	on	79621	48491 (61%)
<i>unknown</i>	no geographic information	1734	970 (56%)
<i>west</i>	ab,bc,mb,sk	73369	36061 (49%)
<i>east</i>	nb,nf,nl,ns,pei	6927	3319 (48%)
<i>quebec</i>	qc	29131	8360 (29%)

Regional differences in code usage are illustrated in Figure 4. All proportions of advertisers were significantly different from the overall proportion for each code except for the *service restrictions* in the *central* region ($p = 0.09$) and *service restrictions* ($p = 0.35$), *communication* ($p = 0.81$), and *client behavior* ($p = 0.015$) in the *east* region.

Figure 4: Proportions of advertisers by region associated with the top 5 codes. The dotted lines indicate the proportion of all advertisers associated with the code.



For the advertisers associated with *multiple* regions, the most common provinces were Ontario (67%, N=11074), Alberta (52%, N=8652), Quebec (43%, N=7067), British Columbia (34%, N=5668), Saskatchewan (22%, N=3688), Manitoba (17%, N=2852), New Brunswick (11%, N=1789). These top 7 provinces were associated with 99.8% of *multiple* region advertisers and the top two provinces, Ontario and Alberta, were associated with 91% of advertisers.

Worker gender

Table 5 shows the number of advertisers associated with self-identified worker gender.

Advertisers who advertised with multiple genders were the most likely to use “no” (62%, N=6289) followed by cis-female identified advertisers (57%, N=100865). Advertisers using “no” associated with multiple genders represented 1.8% of advertisers overall. These advertisers were mostly associated with cis-female and cis-male identity (62%, N=2425) or cis-female and female-identified transgender identity (25%, N=984).

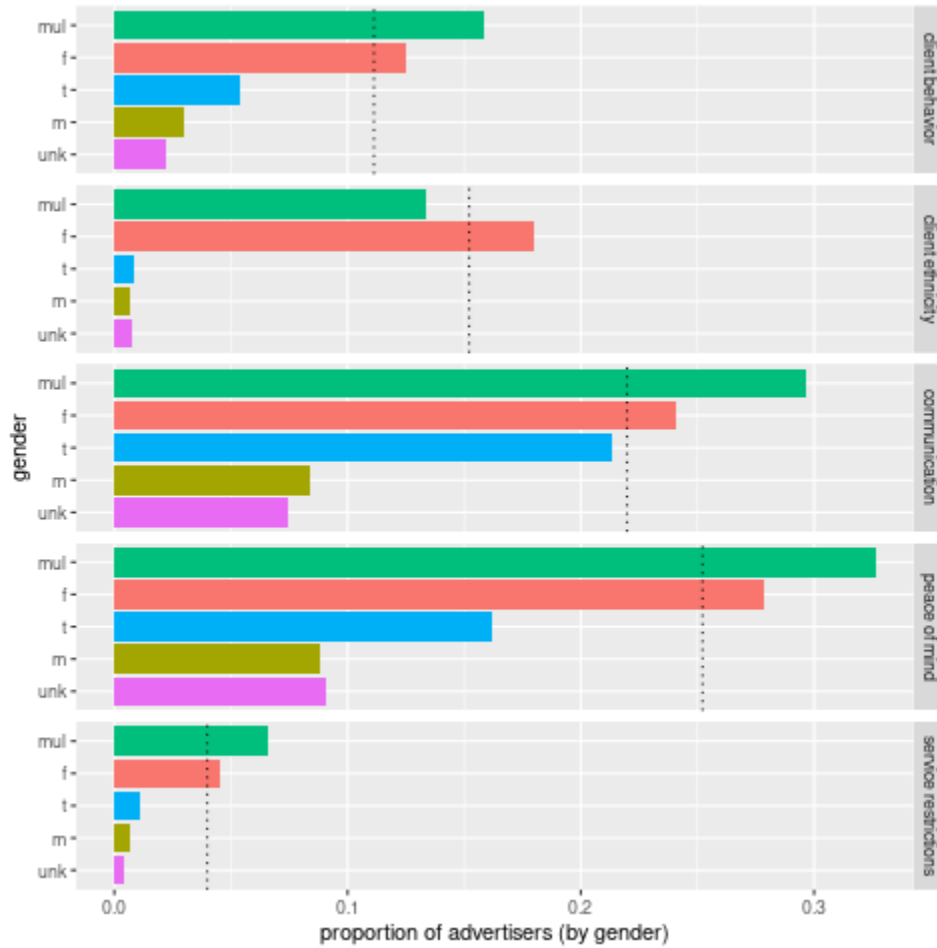
Table 5: Advertisers by gender. Some advertisers are associated with multiple genders.

Transgender refers to female-identified transgender.

gender	advertisers	using “no”
multiple (<i>mul</i>)	6289	3895 (62%)
female (<i>f</i>)	175513	100865 (57%)
transgender (<i>t</i>)	3392	1511 (45%)
unknown (<i>unk</i>)	13359	4253 (32%)
male (<i>m</i>)	15905	4318 (27%)

Figure 5 shows the relative proportions of advertisers associated with the top 5 codes. All proportions of advertisers were significantly different from the overall proportion for each code except for female-identified transgender advertisers (*t*) associated with the *communication* code ($p = 0.38$). Compared to other variables, worker gender with *client ethnicity* showed the largest difference in code usage when comparing cis-female and multiple (18% or 32382 out of 181802 advertisers) and the other gender categories (7% or 230 out of 32656 advertisers).

Figure 5: Proportions of advertisers by gender using the top 5 codes. Gender codes: *f* cis-female, *m* cis-male, *t* female-identified transgender, *mul* more than one gender was associated with the advertiser and *unk* means that no gender was associated with the advertiser. The dotted lines indicate the proportion of all advertisers associated with that code.



Worker ethnicity

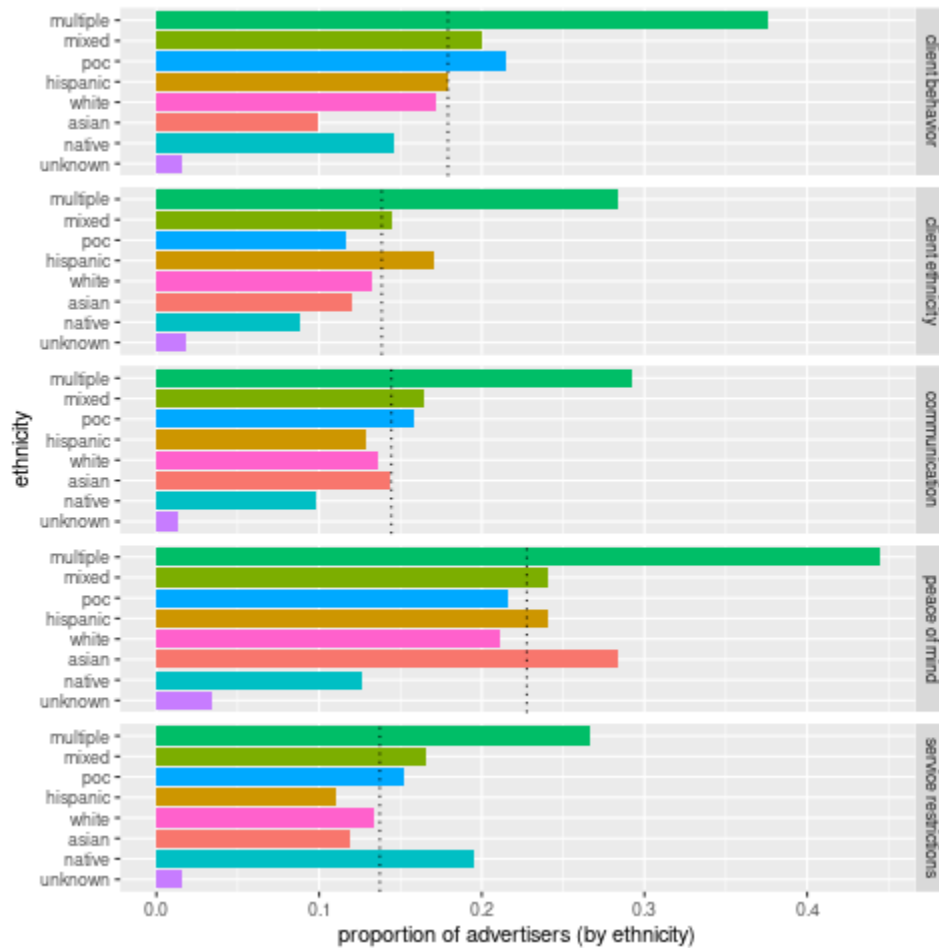
Consistent data on self-identified ethnicity was only available for the 2021-2022 period. The total number of advertisers in 2021-2022 was 38980 with 19876 (51%) using "no" at least once in an ad. Site 3, the source site for this analysis, had an "ethnicity" field which could be optionally filled with one of *Latino/Hispanic, Asian, Mixed, Caucasian/White, Black, Indo Canadian, Middle Eastern, Native, or Canadian Born Chinese*. These ethnic categories were grouped as shown in Table 6 as *Multiple* (used more than one ethnicity option), *Mixed, Hispanic, PoC* (People of Color), *White, Asian, Native*, and *Unknown* to simplify the analysis. Advertisers associated with *Multiple* ethnic identities were the most likely to use “no” (75%, N=4766) followed by *Mixed* (60%, N=1912) and *Hispanic* (57%, N=1029). Advertisers who provided no ethnic identity information were much less likely to use “no” in ads (9%, N=609). The most prevalent ethnic group was *White* (40%, N=15551) where 55% (N=8560) used the word “no”.

Table 6: Advertisers by self-identified ethnicity in 2021-2022.

ethnicities	group	advertisers	using “no”
Advertiser uses multiple ethnicities	<i>Multiple</i>	6380	4766 (75%)
“Mixed” ancestry	<i>Mixed</i>	3165	1912 (60%)
Hispanic	<i>Hispanic</i>	1811	1029 (57%)
Black,Indo Canadian,Middle Eastern	<i>PoC</i>	2668	1509 (57%)
Non-Hispanic White	<i>White</i>	15551	8560 (55%)
Asian,Asian Canadian	<i>Asian</i>	2559	1384 (54%)
First Nations	<i>Native</i>	205	107 (52%)
No ethnicity provided	<i>Unknown</i>	6641	609 (9%)

Figure 6 shows the proportions of advertisers associated with each of the top 5 codes in 2021-2022. All proportions of advertisers were significantly different from the overall proportion for each code. While the magnitude of the differences between advertisers using *Multiple* ethnic identifiers and others is notable, the *Native* advertisers are notable for being the least likely to discriminate based on *client ethnicity* (9%, N=18) and the second most likely to communicate *service restrictions* (20%, N=40). The *Hispanic* advertisers showed the opposite trend: they were the second most likely to discriminate based on *client ethnicity* (17%, N=309) and the least likely to communicate *service restrictions* (11%, N=200). The *Asian* advertisers were the least likely to be associated with *client behavior* restrictions (10%, N=254), the second most likely to be associated with *peace of mind* (28%, N=725), and the third least likely to communicate *service restrictions* (12%, N=306). After *Multiple*, the *PoC* (22%, N=574) and *Mixed* (20%, N=634) advertisers were the most likely to communicate *client behavior* restrictions. The *PoC* (15%, N=406) and *Mixed* (17%, N=523) groups were the most likely after *Multiple* and *Native* to use service restrictions.

Figure 6: 2021-2022 advertisers associated with the top 5 codes by ethnicity. *Mixed* in this context is a single ethnic option but *Multiple* refers to advertisers who indicated more than one ethnicity in their ads such as collective advertisers with ads for multiple individual workers. The dotted lines indicate the proportion of all 2021-2022 advertisers associated with that code.



In 2021-2022 13% (N=6380) of advertisers were associated with *Multiple* ethnic identifiers. Table 7 provides a breakdown of ethnicity usage by these advertisers. The Caucasian/White and Mixed identifiers are associated with 95% (N=4463) of *Multiple* advertisers. Of the 4766 advertisers in the Multiple group that used “no”, the majority (94%, N=4490) also advertised collectively.

Table 7: Ethnic identifiers used by advertisers in the *Multiple* group.

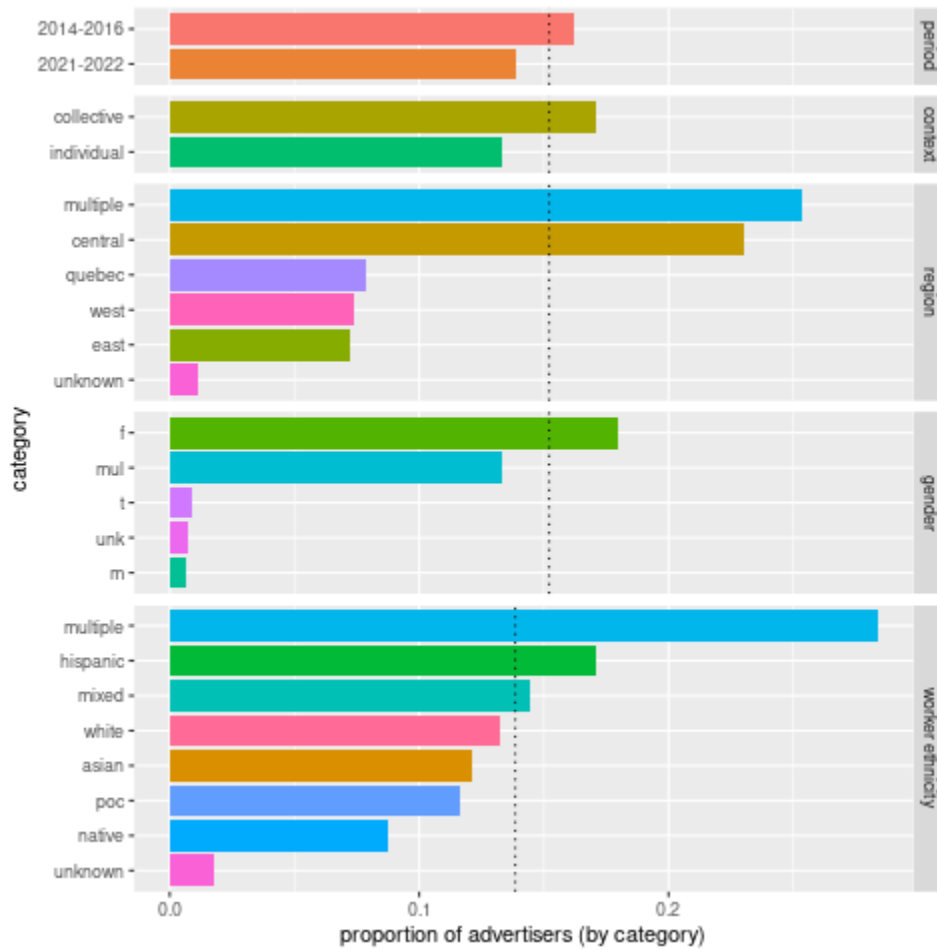
ethnicity	frequency
Caucasian/White	3273 (69%)
Mixed	2979 (62%)
Latino/Hispanic	1920 (40%)
Black	1358 (28%)
Asian	1179 (25%)
Middle Eastern	644 (14%)
Indo Canadian	465 (10%)
Native	216 (5%)
Canadian Born Chinese	201 (4%)

Client ethnicity

In the later corpora, significant minorities of advertisers restricted clients based on ethnicity as indicated by the proportions of advertisers associated with the *client ethnicity* code. In 2014-2016 at least 16% (N=27215) of advertisers restricted clients based on ethnic background. This proportion decreased in 2021-2022 but was still substantial (12%, N=5397). Figure 7 summarizes the demographic characteristics of the advertisers who used this type of restriction. Cis-female identified advertisers (18%, N=32382) (the *f* and *mul* categories described in the *Worker gender* section) and advertisers who traveled (23%, N=5994) or only advertised in Ontario (25%, N=18353) were the most likely to state this type of restriction. In 2014-2016 28% (19998 out of 72264) of these advertisers were associated with *client ethnicity*. In 2021-2022 this

proportion decreased to 25% (4208 out of 17168).

Figure 7: Proportion of advertisers by category who restrict clients based on *client ethnicity*. The dotted lines are the proportion of all advertisers with this type of restriction. Note that 2021-2022 was the only period with reliable data for worker ethnicity.



Within regions, there could be variation in the use of *client ethnicity* restrictions. In the east, advertisers from Newfoundland (4% = 38/930) were the least likely to use these restrictions followed by New Brunswick (7% = 134/2027), Prince Edward Island (8% = 284/3602), and Nova Scotia (12% = 43/357). In the west, advertisers from British Columbia were the least likely to use client ethnicity restrictions (3% = 883/31987) followed by Manitoba (8% = 509/6316),

Saskatchewan (11% = 578/5336), and Alberta (12% = 3475/29730).

Ethnic self-identification affected the probability of being associated with the *client ethnicity* code. The proportions of advertisers, based on worker ethnicity, associated with the client ethnicity code were significantly different from the overall population proportion (14% = 5397/38980) for this code in 2021-2022 ($p < 0.001$). Advertisers associated with the *Multiple* ethnicities group, 94% of whom are collective, were the most likely to be associated with client ethnicity restrictions (28% = 1812/6380) followed by the *Hispanic* (17% = 309/1811), *Mixed* (14% = 457/3165), *White* (13% = 2063/15551), *Asian* (12% = 309/2559), *PoC* (12% = 310/2668), *Native* (9% = 18/205) and *Unknown* (2% = 119/6641). The *PoC* group consisted of three subgroups: Middle Eastern (14% = 60/425), Black (11% = 209/1851), and Indo Canadian (10% = 41/392).

In 2021-2022, workers who are associated with the *client ethnicity* code charged significantly more: mean CAD \$281 per hour (SD CAD \$83, N=3000) versus mean CAD \$250 per hour (SD CAD \$109, N=16984) (Welch's two sample t-sum test: $t = -17.971$, $df = 5013.4$, $p < 0.001$).

Evidence for racial stereotyping of Persons of Color

Given the prominence of restrictions based on client ethnicity, is there evidence of stereotyping language in advertising data? All of the most commonly used bigrams relating to client ethnicity appear to refer to people of African ancestry ("no black", "no african", "no aa", "no blacks", "no blackgents", "no blk"). To see if other offensive language appeared in advertising, a list of related slurs, gathered from (Wikipedia contributors 2023), was used as a basis for searching ad text. The S1 file contains the regular expression (Perl 5 Porters 2021) used to search for the terms. Occurrences were rare with only 2.4 documents in 1000 matching any of

the terms.

The words “pimp”, “thug”, “brown” and “black” were more common. Term counts of the terms “pimp” or “thug” coinciding with “black” and “brown” were tabulated for the 2014-2016 corpus which had the highest number of advertisers restricting clients based on client ethnicity. Searching for joint occurrences of these terms avoids some of the issues of ambiguous usage of the common words “black” and “brown”. Using document frequencies, the word “black” appeared in 46% of the ads containing the words “pimp” or “thug”:

$$p((pimp \text{ or } thug) \text{ and } black) = 6986/2995315 = 0.0023 \text{ (1)}$$

$$p(pimp \text{ or } thug) = 15305/2995315 = 0.0051 \text{ (2)}$$

$$p(black | pimp \text{ or } thug) = p((pimp \text{ or } thug) \text{ and } black) / p(pimp \text{ or } thug) = 0.46 \text{ (3)}$$

The joint probability $p(brown \text{ and } (thug \text{ or } pimp))$ was 0 indicating that “pimp” or “thug” is likely not associated with those of Indo Canadian or Middle Eastern ancestry. The joint probability of $p(racial \text{ epithet and } (thug \text{ or } pimp))$ was also 0 where *racial epithet* refers to strings matching the S1 file regular expression.

Discussion

Over the 15 year period represented by the corpora in this study, roughly similar proportions of advertisers (overall proportion 54%, range by period 51-54%) and posts (overall 39%, range 36-47%) used the word “no”. However, how advertisers used “no” in the earliest corpus is substantially different from how “no” was used in later corpora. In the 2007-2009 corpus, *communication* preferences were the most important consideration followed by statements reassuring prospective clients (*peace of mind*). However, in later corpora the word “no” was more likely to refer to restrictions (*client behavior, service restrictions, or client*

ethnicity). Advertisers using the word "no" were found to post ads more frequently and advertise longer than other advertisers. Overall, female identified advertisers (cis or trans), advertisers who travel, and advertisers who were likely collectives (based on social context, use of multiple genders, or use of multiple ethnic identities) tended to use the word "no" more often than others. People of Color and Mixed ethnicity advertisers tended to be somewhat more likely to communicate restrictions on services and client behavior but this was far less than advertisers who used multiple ethnic identifiers.

The differences in source websites between corpora may have affected the relative rank of the most prevalent codes. The geographic distribution of the 2007-2009 corpus was only one western Canadian province and this may have affected the relative prevalence of codes seen in this period. Indeed, except for eastern Canada, advertisers located in western Canada were less likely to be associated with the *client ethnicity* code in 2014-2016 than other advertisers. However, for the source site of the 2021-2022 corpus, no differences were found in the overall relative rank of codes in 2014-2016 and the subset of advertisers using this site.

Changes in the use of “deposit required” and “screening” might be accounted for by an increase in the number of independent providers. However, as can be seen in the *Social context* section in Results, this was not the case between 2007 and 2022.

This study poses several questions that should be followed up by future work. There seems to be a general lack of research on the effectiveness of occupational health and safety techniques used by sex workers. This begs the question, is there a reliable way to find out how effective specific safety measures are? The analyses presented in this work show that workers who travel may have a heightened sense of risk. To what degree are touring workers more at risk than sedentary workers? A striking feature of the usage of the word “no” was the extent to which

advertisers attempted to reassure clients. Although this can be interpreted as a safety strategy (Bungay and Guta 2018), to what degree is the prospective client's perception of risk reflected in advertising data? Lastly, why was health messaging relatively less prominent than other types of messaging?

Restricting clients based on race or ethnicity

“No black brown middle east east indians too many bad experience [SIC]”

A significant minority of advertisers mostly located in central Canada restricted clients based on ethnicity. The overwhelming majority of these advertisers identified as cis-female. Furthermore, significant minorities of advertisers who identified as People of Color used this restriction.

The effect of location might be explained by the uneven distribution of Men of Color in Canada. Based on the Canadian 2016 census 55% (N=755430) of all Black, South Asian, and Arab men live in Ontario (Statistics Canada 2019). However, 17% (N=228705) and 27% (N=377325) of Men of Color live in Quebec and western Canada respectively where only 7% of advertisers had ethnic restrictions.

As this type of restriction has not been explored extensively in the literature, why mainly cis-female advertisers and why People of Color are associated with it are unanswered questions. In 2014-2016, 72 ads mention "bad experiences" when discussing this type of restriction. However, advertisers did not elaborate further on what this meant.

Overtly offensive language was rare. Furthermore, the frequency estimates of offensive language may be overestimating its use as many of the terms searched had additional meanings. Informal reviews of ads using offensive language showed that these ads could be authored by

disgruntled clients who had used contact information to identify the worker.

Very little research has been done specifically on People of Color who are clients. In *Paying for Sex in a Digital Age* (Sanders and Brents 2020), a comprehensive exploration of the experience of sex buyers in the US and UK, only 30 out of 1893 respondents were People of Color and no specific analysis of the experience of these respondents was undertaken. Earlier research in the US suggested that White clients were proportionally more prevalent online (85%) compared to the number identified in arrest records (57%) (Milrod and Monto 2012; Monto and Milrod 2014). Could it be the case that the increased risk of arrest for People of Color is partly explained by reduced access to online workers?

People of Color have been extensively studied in the context of online dating. The research shows that Men of Color experience disadvantages in both gay and straight online dating contexts (Bany, Robnett, and Feliciano 2014; Feliciano, Robnett, and Komaie 2009; Lin and Lundquist 2013; Mendelsohn et al. 2014; Robnett and Feliciano 2011; Smith 2014; Wade et al. 2022; Wade and Harper 2021; Wade and Pear 2022). This type of discrimination whether subtle or overt, has a significant negative effect on those experiencing it (Bedi 2015).

While the majority of workers appear to not restrict clients based on ethnicity or race, seeing these statements in ads must impact the health and well-being of this client population. However, more work is needed to understand what the effects are and how clients mitigate them.

Limitations

The advertiser proportion measure used in this study identifies the relative prominence of each theme for the authors of the ads. These may not necessarily be the attitudes of the workers represented in the ads although in many cases these would be identical. This research should be

followed up by further qualitative research which grounds its samples in advertiser data to directly compare worker attitudes with how they are portrayed in advertising.

Unscaled advertisers could represent one or more workers and in some cases, advertisers using different contact details or multiple chat ids might be counted more than once. Advertisers identified by contact information, the case in 2007-2009 and 2014-2016, can change contacts over long periods, and small proportions, typically around 5%, are not sex workers (Kennedy 2022). For example, the 172767 advertisers initially identified by contact data in 2014-2016 were estimated to represent 75600 (95% CI 74087–77219) effective advertisers and 169473 (95% CI 166870-172226) workers. The fact that advertisers who use "no" advertised longer than those who did not mitigates this problem somewhat as these advertisers are less likely to change contact information and may be more likely to be relevant advertisers. Similarly, advertisers identified by online profiles, the case in 2021-2022, can have multiple profiles. To mitigate this, the 2021-2022 dataset was extensively screened for non-sex work related advertisers and this appears to have significantly reduced the detected number of duplicate profiles. However, even for this group, ~5% of profiles appear to be duplicated.

While many advertisers in Quebec used “no”, the way similar types of statements are represented in French language ads is different from how these are represented in English resulting in undercounting of Francophone advertisers. More work is needed to understand the effect of linguistic identity and occupational health and safety communication for these advertisers. Similarly, even for Anglophone advertisers, “no” statements could be expressed in other ways. The proportions shown here should be considered a lower bound.

Sex work in Canada exists in an evolving environment where structural variables can have a significant impact on workers and clients. Source websites used in this study changed

between corpora at least in part because of the hostile legal environment engendered by the Protection of Communities and Exploited Persons Act (PCEPA) (Government of Canada 2014). It is conceivable that these external structural changes could affect how workers communicate with clients.

Conclusion

The word "no" is a powerful tool to communicate personal preferences used by majorities of online sex workers. Online sex work advertising is a balancing act between presenting an appealing, welcoming image and setting limits. Messages relating to "no" appeared to be evenly balanced between negative, neutral, and positive affect. However, "no" usage could be surprising. It is disturbing that significant subpopulations of workers restrict clients on ethnicity: the perspectives of both workers and clients of color are needed to understand why this is happening.

Supplemental Materials

Supplemental materials including all source data can be found at <https://osf.io/hwzsn/>

S1 file. Racial epithet regular expression.

Bibliography

- American Sociological Association. 2018. "ASA Code of Ethics."
- Argento, Elena, Matthew Taylor, Jody Jollimore, Chrissy Taylor, James Jennex, Andrea Krusi, and Kate Shannon. 2018. "The Loss of Boystown and Transition to Online Sex Work: Strategies and Barriers to Increase Safety Among Men Sex Workers and Clients of Men." *American Journal of Men's Health* 12(6):1994–2005. doi: 10.1177/1557988316655785.
- Arnholt, Alan T., and Ben Evans. 2021. *BSDA: Basic Statistics and Data Analysis*.

- Bany, James A., Belinda Robnett, and Cynthia Feliciano. 2014. "Gendered Black Exclusion: The Persistence of Racial Stereotypes Among Daters." *Race and Social Problems* 6(3):201–13.
- Bedi, Sonu. 2015. "Sexual Racism: Intimacy as a Matter of Justice." *The Journal of Politics* 77(4):998–1011. doi: 10.1086/682749.
- Bernier, Thérèse, Amika Shah, Lori E. Ross, Carmen H. Logie, and Emily Seto. 2021. "The Use of Information and Communication Technologies by Sex Workers to Manage Occupational Health and Safety: Scoping Review." *Journal of Medical Internet Research* 23(6):e26085–e26085.
- Boecking, Benedikt, K. Miller, M. Barnes, B. Boecking, and E. Kennedy. 2015. "Leveraging Publicly Available Data to Discern Patterns of Human-Trafficking Activity." *Journal of Human Trafficking* 1(1):65–85.
- Boecking, Benedikt, Kyle Miller, Emily Kennedy, and Artur Dubrawski. 2018. "Quantifying the Relationship between Large Public Events and Escort Advertising Behavior." *Journal of Human Trafficking* 5(3):220–37. doi: 10.1080/23322705.2018.1458488.
- Bungay, Vicky, and Adrian Guta. 2018. "Strategies and Challenges in Preventing Violence against Canadian Indoor Sex Workers." *American Journal of Public Health (1971)* 108(3):393–98.
- Corbin, Juliet M., and Anselm Strauss. 1990. "Grounded Theory Research: Procedures, Canons, and Evaluative Criteria." *Qualitative Sociology* 13(1):3–21. doi: 10.1007/BF00988593.
- CSA-SCS Policy, Ethics, and Professional Concerns Subcommittee. 2018. "Statement of Professional Ethics."
- Cunningham, Scott, and Todd D. Kendall. 2010. "Risk Behaviours among Internet-Facilitated Sex Workers: Evidence from Two New Datasets." *Sexually Transmitted Infections* 86(Suppl 3):iii100–105. doi: 10.1136/sti.2010.044875.
- Curtain, Colin. 2023. "QualCoder."
- Feliciano, Cynthia, Belinda Robnett, and Golnaz Komaie. 2009. "Gendered Racial Exclusion among White Internet Daters." *Social Science Research* 38(1):39–54.
- Government of Canada. 2014. *Protection of Communities and Exploited Persons Act*.
- Jiao, Sunny, Vicky Bungay, and Emily Jenkins. 2021. "Information and Communication Technologies in Commercial Sex Work: A Double-Edged Sword for Occupational Health and Safety." *Social Sciences (Basel)* 10(1):1–13.
- Jurafsky, Dan, and James H. Martin. 2023. "N-Gram Language Models." in *Speech and Language Processing*.
- Kennedy, Lynn. 2022. "The Silent Majority: The Typical Canadian Sex Worker May Not Be Who We Think." *PloS One* 17(11):e0277550–e0277550.
- Kille, Julie, Vicky Bungay, John Oliffe, and Chris Atchison. 2017. "A Content Analysis of Health and Safety Communications Among Internet-Based Sex Work Advertisements: Important Information for Public Health." *Journal of Medical Internet Research* 19(4):e111. doi: 10.2196/jmir.6746.
- Kolar, Kat, Chris Atchison, and Vicky Bungay. 2014. "Sexual Safety Practices of Massage Parlor-Based Sex Workers and Their Clients." *AIDS Care* 26(9):1100–1104. doi: 10.1080/09540121.2014.894611.
- Lin, Ken-Hou, and Jennifer Lundquist. 2013. "Mate Selection in Cyberspace: The Intersection of Race, Gender, and Education1." *The American Journal of Sociology* 119(1):183–215.
- Machat, Sylvia, Tara Lyons, Melissa Braschel, Kate Shannon, and Shira Goldenberg. 2022.

- “Internet Solicitation Linked to Enhanced Occupational Health and Safety Outcomes among Sex Workers in Metro Vancouver, Canada 2010–2019.” *Occupational and Environmental Medicine (London, England)* 79(6):373–79.
- MariaDB, and Michael Widenius. 2017. “MariaDB.”
- Mendelsohn, Gerald A., Lindsay Shaw Taylor, Andrew T. Fiore, and Coye Cheshire. 2014. “Black/White Dating Online: Interracial Courtship in the 21st Century.” *Psychology of Popular Media Culture* 3(1):2–18.
- Milrod, Christine, and Martin A. Monto. 2012. “The Hobbyist and the Girlfriend Experience: Behaviors and Preferences of Male Customers of Internet Sexual Service Providers.” *Deviant Behavior* 33(10):792–810.
- Minichiello, Victor, John Scott, and Denton Callander. 2013. “New Pleasures and Old Dangers: Reinventing Male Sex Work.” *The Journal of Sex Research* 50(3–4):263–75. doi: 10.1080/00224499.2012.760189.
- Monto, Martin A., and Christine Milrod. 2014. “Ordinary or Peculiar Men? Comparing the Customers of Prostitutes With a Nationally Representative Sample of Men.” *International Journal of Offender Therapy and Comparative Criminology* 58(7):802–20. doi: 10.1177/0306624X13480487.
- Perl 5 Porters. 2021. “Perlre - Perl Regular Expressions.”
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Robnett, Belinda, and Cynthia Feliciano. 2011. “Patterns of Racial-Ethnic Exclusion by Internet Daters.” *Social Forces* 89(3):807–28.
- Sanders, T., and B. G. Brents. 2020. *Paying for Sex in a Digital Age: US and UK Perspectives*. Routledge.
- Sanders, Teela, Laura Connelly, and Laura Jarvis King. 2016. “On Our Own Terms: The Working Conditions of Internet-Based Sex Workers in the UK.” *Sociological Research Online* 21(4):133–46.
- Sanders, Teela, Jane Scoular, Rosie Campbell, Jane Pitcher, and Stewart Cunningham. 2018. “Characteristics and Working Practices of Online Sex Workers.” Pp. 55–85 in *Internet Sex Work: Beyond the Gaze*. Cham: Springer International Publishing.
- Smith, Jesus. 2014. “Getting Off Online: Race, Gender, and Sexuality in Cyberspace.” Pp. 109–20 in *Illuminating How Identities, Stereotypes and Inequalities Matter through Gender Studies*. Dordrecht: Springer Netherlands.
- Strohmayr, Angelika, Jenn Clamen, and Mary Laing. 2019. “Technologies for Social Justice: Lessons from Sex Workers on the Front Lines.” Pp. 1–14 in *Conference on Human Factors in Computing Systems - Proceedings, CHI '19*. ACM.
- Wade, Ryan M., Alida M. Bouris, Torsten B. Neilands, and Gary W. Harper. 2022. “Racialized Sexual Discrimination Who Seek Intimate Partners Online.” *Sexuality Research & Social Policy* 19(3):1341.
- Wade, Ryan M., and Gary W. Harper. 2021. “Racialized Sexual Discrimination (RSD) in Online Sexual Networking: Moving from Discourse to Measurement.” *The Journal of Sex Research* 58(6):795–807. doi: 10.1080/00224499.2020.1808945.
- Wade, Ryan M., and Matthew M. Pear. 2022. “A Good App Is Hard to Find: Examining Differences in Racialized Sexual Discrimination across Online Intimate Partner-Seeking Venues.” *International Journal of Environmental Research and Public Health* 19(14):8727.

Wikipedia contributors. 2023. "List of Ethnic Slurs and Epithets by Ethnicity — Wikipedia, The Free Encyclopedia."