Population characteristics in justice health research based on PubMed abstracts from 1963 to 2023: text mining study

Wilson Lukmanjaya, Tony Butler, Patricia Taflan, Paul Simpson, Natasha Ginnivan, Iain Buchan, Goran Nenadic, George Karystianis

Submitted to: JMIR Formative Research
on: May 29, 2024

Table of Contents

Population characteristics in justice health research based on PubMed abstracts from 1963 to 2023: text mining study

Wilson Lukmanjaya[1] BSc; Tony Butler[1] BSc, MSc, PhD; Patricia Taflan[1]; Paul Simpson[1] BSc (Psych), PhD; Natasha Ginnivan[1] BSc, PGD (Pyschology), PhD; Iain Buchan[2] MD, PhD; Goran Nenadic[3] BSc, MSc, PhD; George Karystianis[1] BSc, MSc, PhD

[1]School of Population Health, University of New South Wales Sydney AU
[2]Institute of Population Health, Liverpool University Liverpool GB
[3]School of Computer Science, University of Manchester Manchester GB

Corresponding Author:
Wilson Lukmanjaya BSc
School of Population Health, University of New South Wales
Samuels Building, F25, Samuel Terry Ave, Kensington NSW 2052
Sydney
AU

Abstract

Background: The field of epidemiological criminology (or justice health research) has emerged in the past decade studying the intersection between the public health and justice systems. To ensure research efforts are focused and equitable, it is important to reflect on the outputs in this area and address knowledge gaps.

Objective: The aim of this study was to examine populations researched in a large sample of published outputs and identify research gaps and biases.

Methods: A rule-based text mining method was applied to 34,481 PubMed abstracts published from 1963 to 2023 to identify four population characteristics (sex, age, offender type, nationality).

Results: We evaluated our method in a random sample of 100 PubMed abstracts. Micro-precision was 94.3% with micro-recall at 85.9% and micro-F1-Score at 89.9% across the four characteristics. Half of the abstracts did not have any characteristic mentions (49.4%; 17,039) and only 1.3% (443) reported sex, age, offender type and nationality. From 5,170 (14.9%) abstracts that reported age, 7 out of 10 (69.3%) mentioned young people (under 18 years) and 6 out of 10 (58.7%) abstracts reported adults. Since 1990, studies reporting female only populations increased and in 2023 accounted for almost half of research outputs (48.6%) as opposed to 33.1% for male only populations. Nordic countries (Sweden, Norway, Finland, and Denmark) had the highest number of abstracts proportional to their incarcerated population. Mentally ill offenders were the most common group of interest (17.4%) with an increase from 1990 onwards.

Conclusions: Research reporting on female populations increased and surpassed those involving men despite women representing 5% of the incarcerated population suggesting male prisoners are under researched. Although it has been suggested that the justice health area should focus more on juveniles, our results showed a high number of age reporting abstracts mentioning a population age below 18 years old reflecting a rise of youth involvement in the youth justice system. Those convicted of sex offences and crimes relating to children were not as researched as the existing literature suggests with a focus instead on mentally ill populations whose rates rose steadily in the last 30 years. After adjusting for the size of the incarcerated population, Nordic countries have conducted proportionately the most research. Our findings highlight that despite the presence of several research reporting guidelines, justice health abstracts still do not adequately describe the investigated populations. Our study offers new insights in the field of justice health. It has implications for promoting diversity in the selection of research participants. Men in prison appear to be under researched compared with women.

(JMIR Preprints 29/05/2024:60878)
DOI: https://doi.org/10.2196/preprints.60878

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Population characteristics in justice health research based on PubMed abstracts from 1963 to 2023: text mining study

Abstract

Background: The field of epidemiological criminology (or justice health research) has emerged in the past decade studying the intersection between the public health and justice systems. To ensure research efforts are focused and equitable, it is important to reflect on the outputs in this area and address knowledge gaps.
Objective: The aim of this study was to examine populations researched in a large sample of published outputs and identify research gaps and biases.
Methods: A rule-based text mining method was applied to 34,481 PubMed abstracts published from 1963 to 2023 to identify four population characteristics (sex, age, offender type, nationality).
Results: We evaluated our method in a random sample of 100 PubMed abstracts. Micro-precision was 94.3% with micro-recall at 85.9% and micro-F1-Score at 89.9% across the four characteristics. Half of the abstracts did not have any characteristic mentions (49.4%; 17,039) and only 1.3% (443) reported sex, age, offender type and nationality. From 5,170 (14.9%) abstracts that reported age, 7 out of 10 (69.3%) mentioned young people (under 18 years) and 6 out of 10 (58.7%) abstracts reported adults. Since 1990, studies reporting female only populations increased and in 2023 accounted for almost half of research outputs (48.6%) as opposed to 33.1% for male only populations. Nordic countries (Sweden, Norway, Finland, and Denmark) had the highest number of abstracts proportional to their incarcerated population. Mentally ill offenders were the most common group of interest (17.4%) with an increase from 1990 onwards.
Conclusions: Research reporting on female populations increased and surpassed those involving men despite women representing 5% of the incarcerated population suggesting male prisoners are under researched. Although it has been suggested that the justice health area should focus more on juveniles, our results showed a high number of age reporting abstracts mentioning a population age below 18 years old reflecting a rise of youth involvement in the youth justice system. Those convicted of sex offences and crimes relating to children were not as researched as the existing literature suggests with a focus instead on mentally ill populations whose rates rose steadily in the last 30 years. After adjusting for the size of the incarcerated population, Nordic countries have conducted proportionately the most research. Our findings highlight that despite the presence of several research reporting guidelines, justice health abstracts still do not adequately describe the investigated populations. Our study offers new insights in the field of justice health. It has implications for promoting diversity in the selection of research participants. Men in prison appear to be under researched compared with women.

Keywords: epidemiology; PubMed; criminology; text mining; justice health; offending and incarcerated populations; population characteristics; open research

Introduction

Studies investigating the health needs of offender populations represent an emerging discipline called epidemiological criminology (1, 2) and are affected by factors such as funding, complex and multi-layered ethics approvals, access to prisoners or community-based offender populations, data quality, and reporting bias (3-6). Understanding this population's unique needs enables researchers and policymakers to target specific health and wellbeing needs rather than generalising across all groups (7).

When researchers fail to accurately report their research, biases can occur (8). For that reason, health research reporting has evolved with the introduction of STROBE (STrengthening the Reporting of OBservational studies in Epidemiology)(9), CONSORT (Consolidated Standards of Reporting Trials) (10), SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials)(11) and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)(12) statements which provide guidelines and templates for investigators to structurally report their findings in concise yet detailed manners.

Developing effective population prevention and intervention strategies requires evidence-based reporting of the studied population (13). A 2018 synthesis of reviews on global prisoner health concluded that detained adolescents were not the focus of any of the included studies despite evidence of health inequalities within that particular population (14). Furthermore, minority groups exhibit varying morbidity and mortality rates, suggesting distinct health risks and outcomes (15). Establishing conclusions generated from a minority population towards larger ones, and vice versa, has the potential to lead to ineffective interventions (16, 17). Therefore, it is imperative to accurately report the characteristics of populations involved in research to ensure the transparency and reproducibility of related studies.

As more scientific literature becomes available, the task of manually reading, extracting and synthesising knowledge from large numbers of epidemiological studies becomes more time-consuming (18-20). Automated applications offer investigators the opportunity to quickly and efficiently detect, summarise and incorporate key information from relevant literature (21, 22). However, few studies have attempted to determine a whole-of-discipline perspective by examining the scope and quality of peer reviewed outputs over time. Several research efforts have been made to identify key information (e.g., study design, participant type, arm of intervention, confounding factors) from experimental and observational studies with varying degrees of success, particularly from randomised controlled trials (18, 19, 21, 23-31).

Health research, including that related to the justice system, is indexed in bibliographical databases which publish the abstracts of such studies. Abstracts are written in a relatively structured format within each journal's reporting style and aims to improve communication, making them ideal for the application of a rule-based text mining method (19, 31, 32). They are publicly available in digital form, and not behind a pay wall enabling easy large scale research. The largest database is PubMed, developed by the National Institutes of Health's (NIH), National Library of Medicine, provides access to millions of citations from biomedical journals (33). For example, PubMed has more than 34,000 published articles in the justice health area alone (34). Since epidemiology is a field with its own dictionary with related studies following a semi-structured reporting style (35), we hypothesized that a simple text mining approach (i.e., rules that can identify characteristics of interest) could provide an effective means to extract key information from the whole discipline of justice health.

In this study, we applied a rule-based method on 34,481 PubMed justice health abstracts from 1963 to 2023 to automatically extract a set of population characteristics (age, sex, nationality, offender type) and highlight whether there are biases or gaps in this area from a participant perspective.

Methods

Data

We conducted a literature search in PubMed using an expanded version of an existing query (20) containing search terms related to offenders and prisons. These were combined with either the Medical Subject Headings (MeSH) term "epidemiology" to capture all types of epidemiological

studies or with all the available (in PubMed) publication types (e.g., meta-analysis, clinical trial) to ensure the results will return clinical trials and secondary research (e.g., review). We also added terms related to randomization/natural experiments and synthetic control. These choices prevented articles that made only passing reference to prisoner and offender studies from entering the dataset resulting in a high-quality corpus for our analysis. The search was restricted to English language articles that have an abstract and involved only human participants (i.e., veterinary research was excluded).
The full query was run on the 20th of July 2023:

“prison OR borstal OR jail OR jails OR gaol OR gaols OR penitentiary OR custody OR custodial OR (corrective AND (service or services)) OR ((correctional or detention) AND (centre or centres OR center OR centers OR complex OR complexes or facility or facilities)) OR (closed AND (setting)) OR prisoner OR prisoners OR incarcerated OR criminals OR criminal OR felon OR felons OR remandee OR remandees OR delinquent OR delinquents OR detainee OR detainees OR convict OR convicts OR cellmate OR cellmates OR offenders OR offender OR ((young OR adolescent) AND (offender OR offenders)) OR ((delinquent OR incarcerated) AND youth) OR (juvenile AND (delinquents OR delinquent OR delinquency OR detainee OR detainees OR offender OR offenders)) OR ((young) AND (people) AND (in) AND (custody)) OR ((justice) AND (involved) AND (youth)) OR ((incarcerated) AND (young) AND (people OR person OR persons)) OR ((juvenile OR juveniles) AND (in) AND (custody)) AND english[lang] AND ("epidemiology"[Subheading] OR "epidemiology"[MeSH Terms] OR epidemiology[Text Word] OR clinical study[publication type] OR case reports[publication type] OR clinical trial[publication type] OR clinical trial, phase i[publication type] OR clinical trial, phase ii[publication type] OR clinical trial, phase iii[publication type] OR clinical trial, phase iv[publication type] OR comparative study[publication type] OR controlled clinical trial[publication type] OR evaluation study[publication type] OR meta-analysis[publication type] OR multicenter study[publication type] OR observational study[publication type] OR pragmatic clinical trial[publication type] OR randomized controlled trial[publication type] OR review[publication type] OR systematic review[publication type] OR twin study[publication type] OR validation study[publication type] OR non randomized trial[text word] OR non randomised trial[text word] OR randomization experiment OR randomisation experiment OR natural experiment OR synthetic control)”

Text mining

Dictionaries

To identify the reported sexes, we used various indicators (e.g., boys, girls, men, women, males, females, transgender, trans). A total of 26 terms were used (Multimedia appendix 1). We also compiled a list of offences (36) including their synonyms (e.g., “sex crime”, “sex offending”, “sexual offending”), acronyms (e.g., “ADVO [aggravated domestic violence order]”) and descriptive sentences (e.g., “breach of parole”, “assault with intent to commit rape”). We included grammatical variations of these offences to expand the scope of our dictionary. A total of 1,036 terms were used.

To capture nationalities, we used three dictionaries that indicate a place of origin: one for overall nationalities (1,575), one for country names (363) and one for the largest cities of the world (317). We included nationalities (e.g., Czechoslovakian) and countries (e.g., Yugoslavia) that no longer exist as well as variations of the same nationality and country/region (e.g., Dominicans, Dominicanes, United Kingdom, Great Britain, Britain, UK)(37). Considering how our previous research on examining first author affiliations from justice health PubMed abstracts demonstrated that the United States is the number one country in sheer publication outputs in this area (20), we added three more dictionaries for the US states (50), counties (3,135) and the largest US cities (200).

We also used a dictionary of 259 commonly used terms to describe offending and incarcerated populations (e.g., criminals, incarcerated, reoffending, juvenile, delinquent) (Multimedia Appendix 2).

Rule based text mining approach

From our query results, we randomly selected 100 abstracts as our training set. The training set was annotated by two authors with epidemiological and public health background (GK, TB) for the four population characteristics (i.e., nationality, age, sex, offender type). The inter-annotator agreement calculated as the absolute agreement rate (38) was between 85.0% to 100.0% suggesting reliable annotations.

We developed rules based on common lexical patterns observed in the training set which suggest the presence of any of the four characteristics. The lexical patterns employ frozen syntactical expressions as anchors for certain elements built through verbs, noun phrases, and prepositions, and semantic place holders which can be identified by the dictionary application that indicates a characteristic. For example, the following sentence ("characteristics of sex offenders in") has a mention of "sex offenders" [offender type]. To identify this, the semi frozen lexical expressions "characteristics of" and "in" are matched via two regular expressions; and "sex offenders" gets a match through the offender type dictionary. More than one lexical pattern may be matched in an abstract referring to one or more mentions of a characteristic (which can be duplicates).

An additional sample of 100 randomly selected abstracts was used to serve as the development set aiming to optimize the performance of our method. A total of 140 rules were crafted: 11 for nationality, 47 for age, 4 for sex and 78 for offender type (Multimedia Appendix 3 shows rule examples for each characteristic). To convert the observed lexical patterns into rules, we used the General Architecture for Text Engineering (GATE) and its Java Annotations Pattern Engine (JAPE), a pattern matching language for GATE (39).

Data standardization

To enable descriptive and temporal analysis of the extracted results, mentions of age, sex, nationality and offender type were standardized by using a simple Python script. Unique values from each characteristic were manually inspected by three authors (GK, PS, TB) to identify synonyms (e.g., sex offenders, sex offending, sexual offenders) which then were assigned a respective term (e.g., sex offender). For age, two types of mentions were identified: numeric (e.g., 18 to 24 years old) and textual (e.g., adolescents, adults). We categorized the numerical values according to the Australian Bureau of Statistics' seven age groups: under 18 (minors), 18-24, 25-34, 35-44, 45-54, 55-64 and 65+ years old (40-44). We also assigned a numeric range for the textual mentions (Table 1). If "adults" mentions were stated, this was placed into the "unknown adult" category.

Table 1. Classification of standardized age textual mentions from PubMed abstracts.

Term	Age Range	Classification
Children, minors, juveniles, delinquents, school children, boys, girls	<18 years old	Minors
Juvenile, delinquent	10-17 years old	Minors
Adolescent, teen	13-19 years old	Minors
		18-24 years old
Young, youths, young offenders	15-24 years old	Minors

		18-24 years old
Adults	18+ years old	Unknown adults

To standardize sex mentions, five categories were used: male, female, transgender individuals (i.e., mentions of transgender individuals without specification), transgender men, and transgender women. For nationality mentions, those that belonged to US counties, cities and states were standardized as "American" whereas nationalities that are not in use anymore (e.g., "Czechoslovakian") were assigned a miscellaneous status.

For the offender type, due to the different level of information that each mention might bear (e.g., serial rapist), we used a more generic (when possible) grouping. For example, for populations involved with rape, we maintain the specific offence as rape and assign a higher offence node as "sex offence". A total of six categories were created (child crime related (including child sex abuse) offender, sex offender, violent offender, non-violent offender, mentally ill offender, drug related offender). We also created an additional category called "miscellaneous" to include other non-specific descriptions (e.g., "high risk offenders", "ex-offenders", "juveniles", "delinquents") that could not be mapped to any of the other categories (Multimedia appendix 4). To obtain results at the abstract level for each abstract, we eliminated any duplicate mentions of the standardized characteristics. Table 2 shows some examples of standardizing extracted population mentions according to the four defined characteristics.

Table 2. Examples of standardized extracted mentions of the four population characteristics (i.e., age, sex, nationality, offender type) including attributes that describe offender types.

Extracted mention	**Characteristic**	**Standardized version**		**Offender type**
Women	sex	female		-
Boys and girls	sex	male	female	
Women with borderline personality disorder	sex	female		-
	offender type	mentally ill offender		borderline personality disorder
Child molester	offender type	child crime related offender		child sex abuse
Male sex offenders	offender type	sex offender		-
Serial rapists	offender type	sex offender		rape
Psychotic inmates	offender type	mentally ill offender		psychosis
Ex-offenders	offender type	miscellaneous		ex-offender
Age 18-25 years	age	18-24	25-34	-
Iowa	nationality	American		-
Norway	nationality	Norwegian		-

Results

Text mining evaluation

The system's performance was evaluated at the abstract level and used the standard definitions of the precision, recall and F1-Score metrics (45). True positive was defined as the identification of either all the correct mentions of a population characteristic or the extraction of a number of mentions for one population characteristic, even if the system failed to pick up some mentions in an abstract. For example, if an abstract had two mentions of female sex (e.g., "females", "women"), then the detection of either one or both mentions would be considered a true positive at the abstract level with "female" being the standardized sex in this abstract. The same process was applied in cases where there can be more than one different mention of a population characteristic (e.g., mentions of two

different nationalities for the investigated population). A false positive (FP) is an identification of an incorrect mention for a population characteristic while a false negative (FN) is an incorrectly ignored mention of a population characteristic.

Overall, at the abstract level, the mean precision and recall were 95.2% and 90.9%, respectively, whereas the F1-score was 93.0%. However, since the number of mentions between the four characteristics varied drastically in the evaluation set, we reported on the micro values of precision, recall and F1-Score to offer a more weighted approach to the system's performance. Micro precision was 94.3%, micro recall at 85.9% and micro F1-Score at 89.9%. The largest recall drop was observed in age (6.2%), and it was the only recall with a value below 80.0% (78.8%) while nationality had the highest recall (95.2%). Sex had the highest precision (100.0%) followed by age (97.6%) (Table 3). The highest F1-Score was observed for sex (96.4%) followed by nationality (93.0%). Age and offence type had similar F1-Scores with 87.1% and 87.3% respectively (Table 3).

Table 3. Precision, recall and F1-Score for the training, development and evaluation sets including the number at the abstract level of True Positives (TP), False Positives (FP) and False Negatives (FN) for age, sex, offender type and nationality.

Age	**TP**	**FP**	**FN**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Training set	51	6	10	89.4	83.6	86.4
Development	57	1	10	98.2	85.0	91.1
Evaluation set	47	1	11	97.6	78.8	87.1
Sex						
Training set	52	5	5	91.2	91.2	91.2
Development	54	6	5	90.0	91.5	90.7
Evaluation set	54	0	4	100.0	93.1	96.4
Offender type						
Training set	78	15	15	83.8	83.8	83.8
Development	94	11	8	89.5	92.1	90.7
Evaluation set	98	8	19	92.4	83.7	87.3
Nationality						
Training set	36	3	3	92.3	92.3	92.3
Development	32	5	1	86.4	96.9	92.3
Evaluation set	40	4	2	90.9	95.2	93.0

Query results

Our query returned a total of 34,481 justice health study abstracts with the earliest recorded in 1963 (Multimedia appendix 5). Half of the abstracts (49.4%; 17,039) did not have any characteristic mentions. Most abstracts either mentioned only age and nationality (4.9%; 1,676) or age and offender type (3.1%; 1,082). Only 1.3% (443) of abstracts reported all four characteristics (sex, age, offender type, nationality).

Age

A total of 5,170 abstracts (14.9%) reported the population's age with 7 out of 10 abstracts (69.3%) mentioning minors, 6 out of 10 (58.7%) reporting adult populations and 3.5% not specifying the age.

The largest adult group was that of 18-24 years old (33.4%) followed by 35-44 years old (10.4%). Studies involving 55-64 years old had the lowest number of mentions (7.4%) (Table 4).

Table 4. Number of justice health abstracts (n=5,170) in PubMed from 1963 to 2023 reporting age. Note that one abstract can include more than one age group.

Initial age group	Number of abstracts (%)	Age group	Number of abstracts (%)
Minors (<18)	3,581 (69.3)	Minors (<18)	3,581 (69.3)
Adults (>18)	3,037 (58.7)	18-24	1,728 (33.4)
		25-34	519 (10.0)
		35-44	538 (10.4)
		45-54	398 (7.7)
		55-64	382 (7.4)
		65+	473 (9.1)
		Unknown adult	183 (3.5)

Sex

A total of 8,169 (23.6%) abstracts reported the sex of the investigated population in the abstract. 39.7% (3,241) of the abstracts reported only males, 42.9% (3,501) reported only females and 17.4% (1,418) reported both males and females. Less than 1% (21) of abstracts reported transgender populations. Although there has been a gradual increase in study rates involving only female populations since 1990 (see trend line, Figure 1), from 2014 onwards, a decrease was noted with regards to abstracts reporting only male populations, surpassed by abstracts reporting only female populations. In 2023, 48.6% of abstracts reported only female populations versus 33.1% reporting only male populations (Figure 1) (Multimedia Appendix 6 shows the rates per year in detail).

Nationality

A total of 9,525 (27.6%) abstracts reported the nationality of the investigated population. The most common nationality was United States (i.e., American) (31.4%; 2,992) followed by United Kingdom (9.2%; 786) and Australia (7.6%; 730) (Table 5). However, to account for the size of the country population, which we assumed to be broadly linked to the size of its prisoner population (Pearson r=0.73), and this in turn being a potential driver of the volume of research reflected by the number of publications, we derived a publication rate based on the average prisoner population size over the period of 2000 to 2020 (46) and calculated a rate per 1000 prisoner population. In this case, the Nordic countries were in the top 4 in terms of publication rate followed by Australia (Table 5). Only four countries from Asia (China, India, Japan, South Korea) and one country from Africa (South Africa) were in the top twenty of both crude and rate ranks.

Table 5. Top 20 most common nationalities reported in 9,525 justice health articles in PubMed from 1963 to 2023 along with the respective continent, number of articles, prisoner population (average 2000-2020), article rate per 1000 prisoners, and rate rank.

Crude rank	Country	Continent	Articles (%)	Prisoner population[a]	Article rate per 1000 prisoners[b]	Rate rank
1	United States	North America	2,922 (31.41%)	2,120,277	1.4	16
2	United Kingdom	Europe	993 (10.43%)	88,274	11.2	9
3	Australia	Oceania	730 (7.66%)	30,685	23.8	5
4	Canada	North America	498 (5.23%)	38,321	13.0	8
5	China	Asia	371 (3.90%)	1,627,290	0.2	20
6	Germany	Europe	340 (3.57%)	68,437	5.0	10

7	France	Europe	281 (2.95%)	62,158	4.5	11
8	Sweden	Europe	261 (2.74%)	6,510	40.1	3
9	India	Asia	249 (2.61%)	385,832	0.6	18
10	Netherlands	Europe	249 (2.61%)	14,470	17.2	7
11	Italy	Europe	227 (2.38%)	56,090	4.0	12
12	Japan	Asia	202 (2.12%)	65,348	3.1	13/14
13	Spain	Europe	193 (2.03%)	61,751	3.1	
14	Brazil	South America	162 (1.70%)	509,602	0.3	19
15	South Africa	Africa	146 (1.53%)	164,629	0.9	17
16	Norway	Europe	138 (1.45%)	3,289	42.0	1
17	Switzerland	Europe	135 (1.42%)	6,257	21.6	6
18	Finland	Europe	135 (1.42%)	3,238	41.7	2
19	Denmark	Europe	125 (1.31%)	3,729	33.5	4
20	South Korea	Asia	94 (0.99%)	52,989	1.8	15

a Average prisoner population 2000 to 2020 (46).
b Rate per 1000 prisoners

Offender type

4,814 (13.9%) abstracts mentioned the offender type. Mentally ill offenders were reported in 17.4% (840) of the PubMed abstracts followed by sex offenders (12.9%; 620). Child crime related offenders (e.g., child abusers) had the lowest number of mentions with 1.7% (84) (Table 6).

Table 6. Number of justice health abstracts (n=4,814) in PubMed with an offender type across female and male populations. Note that one abstract might have more than one offender type and might include both male and female populations.

Offender type	Frequency	%	Male (%)	Female (%)	Unknown (%)
Miscellaneous	3,389	70.4	1,162 (34.3%)	942 (27.8%)	1,741 (51.4%)
Mentally ill offender	840	17.4	193 (23.0%)	103 (12.3%)	596 (71.0%)
Sex offender	620	12.9	211 (34.0%)	81 (13.1%)	386 (62.3%)
Drug related offender	521	10.8	111 (21.3%)	84 (16.1%)	356 (68.3%)
Violent offender	364	7.6	134 (36.8%)	77 (21.2%)	201 (55.2%)
Non-violent offender	96	2.0	25 (26.0%)	16 (16.7%)	67 (69.8%)
Child crime related offender	84	1.7	26 (31.0%)	16 (19.0%)	54 (64.3%)

From 1990 to 2023, the overall number of PubMed abstracts with an offender type increased (Multimedia appendix 7). However, rate wise (i.e., the number of PubMed abstracts with a specific offender type divided by the total number of PubMed abstracts that had a mention of an offender type) revealed a general increase for mentally ill offenders. Mentions for sex, drug related, non-violent and violent offenders had an overall decrease with the biggest noted for sex offenders (10.0%) (Figure 2).

When the sex offenders are broken down by male and female, the rates of study abstracts through time remain roughly the same since 1990 with an average rate of 33.0% for male offenders and 15.5% for female offenders. A minimal increase towards male sex offenders and a minimal decrease on female sex offenders were observed (Figure 3).

Discussion

Principal Results

This text mining study demonstrated that key population characteristics (age, sex, offending type, nationality) can be derived by applying text mining to a large corpus of study abstracts available in PubMed. Our findings enable researchers to investigate the presence of potential research and knowledge gaps over time that arise from examining certain offending groups within an entire discipline. Half of abstracts (49.4%) did not report any characteristics while the number of abstracts that mentioned one characteristic ranged from 13.9% (offender type) to 27.6% (nationality). This highlights a larger problem towards the reporting of necessary information in abstracts for the description of populations within the justice health area as only 1.3% of our sample abstracts reported sex, age, offender type and nationality.

Previous research has showcased that despite myriad reporting guidelines covering observational, experimental and secondary study reporting (e.g., STROBE, CONSORT, SPIRIT, PRISMA) (9-12), justice health abstracts do not appear to adequately detail their study designs and examined variables (32, 34) and based on our results, nor do they adequately describe the population under investigation. The importance of population description in research is needed to not only understand predictors for recidivism but also to enable the conduction of meta-analyses and other future studies (47).

Age

We initially standardized the extracted age into two groups (below 18 years old and over) to examine age-related trends in offending populations. However, age was mentioned in 14.9% of our PubMed abstract sample, thus this finding should be taken with caution. Although it has been suggested that the justice health area should focus more on juveniles (14), our results showed that 7 out of 10 (69.3%) abstracts mentioned a population age below 18 years old. This finding reflects the rise in youth involvement in the youth justice system (48); the US saw a 30% increase in juvenile cases between 1985 and 2009 (49) and Australia noted a recent increase of 6% from 2021-22 to 2022-2023 (50).

The high number of abstracts reporting minors in the justice health area could also be explained through reporting practices. It is possible that the majority of researchers who investigate minor populations are more likely to speficy the age. Most abstracts referred to the investigated population in generic terms such as “offenders” or “incarcerated individuals” which could imply adult populations. This would separate them from younger people who are described by more specific terms such as “adolescent”, “juvenile” and “delinquent”. This, however, was not taken into account for this research. Considering this and along with the inspection of full text studies which might describe in accuracy the age of the participant population, our finding could be different.

Sex

Although males make up the overwhelming majority of incarcerated populations (10.9 million worldwide versus <1 million for women), there was an over-representation of studies involving female offenders suggesting that men in prison are an under-researched group (51). Increased research into women since 2000 aligns with an increase in the number of incarcerated women worldwide since 2000 (52) with female offender studies rising from 29.2% in 1990 to 48.6% in 2023. Conversely, male focussed research decreased from 50.8% in 1990 to 33.3% in 2023. This disparity evokes a consideration of equity in justice health research. Equity and not equality should be prioritized in health (53, 54). Therefore, it is possible that although inequality is shown through rates of research between male and female offending populations, an equity approach can contribute

to our understanding on why there is disproportionately more research on female populations.

However, since only 23.7% of our data sample reported sex, it is possible that the remaining studies which did not detail the population's sex in the abstract focused on men. Given that most prisons hold male prisoners only, investigators focusing on female populations in the justice health area might be better at reporting the female sex. Nevertheless, this highlights the need for more detailed reporting in PubMed abstracts, to allow other researchers to accurately synthesize information more effectively and accurately.

Nationality

Nationality was the population characteristic with the highest prevalence in our sample (27.6%). Using the crude rank, the US was the most common nationality. However, using the publication rate the US dropped to nineteenth with the Nordic countries (Norway, Finland, Sweden, Denmark) occupying the top four spots and Australia the fifth one. Previous research analysing PubMed justice health abstracts showed similar rankings for the Nordic countries in terms of their total published outputs (20). Nordic countries are regarded as having a progressive approach to offender rehabilitation with proportionally lower numbers of incarcerated individuals and recidivism rates compared to many other countries (55, 56). However, these results are based only on 29.7% of PubMed abstracts with nationality reported so it is possible that in full text studies, the actual nationality of the examined population is described which could reveal a different picture in the rankings.

Offender type

Mentally ill offenders were the most common group identified from the abstracts (17.4%). In the United States, it has been estimated that 24% of the inmate population have a mental illness (57) with approximately 50% to 75% of the 2 million juveniles meeting criteria for a mental health disorder (58, 59). In the last ten years, reliance on the juvenile justice system to meet its population's mental health needs has increased and so has the research to examine the effectiveness of intervention and treatment programs (60). Mentally ill offenders have higher rates of recidivism exhibiting rehabilitation needs and prison adjustment difficulties that differentiate them from the general offender population (61, 62). Our results highlight the depth of this problem with researchers examining a total of 58 unique mental illness concepts (Multimedia appendix 8) in the last 70 years ranging from behavioural disorders (e.g., attention deficit hyperactive disorder) to mood disorders (e.g., depression, bipolar) and anxiety disorders (e.g., post-traumatic stress disorder) with substance use disorders and intellectual disability receiving the most focus.

Most research on the longitudinal pattern of criminal careers has focused on generally violent offenders (63) which could explain the relatively low number of abstracts mentioning non-violent offenders (2.0%) involved for example, with theft and shoplifting. This indicates the need to investigate a more diverse range of offender groups (64, 65). Crimes such as theft, stalking, and driving under the influence may cause significant harm towards others and yet, there is a lack of related work focusing on cases of, for example, fraud and sextortion which can have significant effects on victims (66, 67).

Considering that sex offenders are regarded as one of the serious offender groups (68, 69), it is not surprising they were the second most commonly researched group in our sample (12.9%). European surveys have suggested that up to 10% of male offenders commit sexual violence against adult women (70, 71) with Australia noting an average of 36.4% of all offences recorded been related to sexual assault in the last 15 years (72). US statistics also put the prevalence of sexual assault at half a

million incidents per year (73). Despite males comprising the majority of the sex offender population (e.g., in Australia 97% of sex offenders are men)(74), research suggests that the proportion of female sex offenders are higher than thought (75). A recent meta-analysis with data from 12 countries reported that victimization surveys indicated prevalence rates of female sex offenders were six times higher than official data (11.6%). This disparity is similar to our findings that saw female sex offenders comprising 13.1% vs male sex offenders with 34.0% from 620 abstract studies.

Despite an estimate of 1 billion children aged 12-17 years experiencing child abuse and maltreatment (76), our findings suggest that individuals responsible for committing such offences are under researched with only 1.7% of the abstracts reporting child related crimes. Since such offences are hard to detect due to the involvement of minors and adolescents, only 1.7% of our sample over seventy years involved those convicted of child sex offences highlights a research gap in justice health. To design and implement effective prevention and intervention programs for child related crimes, it is necessary to have more evidence-based research on individuals committing this type of offence.

Text mining error analysis

Utilizing a rule-based method returned encouraging results (mean F1-Score 93.0% across four characteristics) although the micro-F1-Score was at 89.9% which can offer a more weighted performance due to each class's different number of mentions in the evaluation set.

False positives

Some nationality terms were either part of a population's ethnicity (e.g., Mexican-American) or referred to nationalities irrelevant to the current study s (e.g., "As in the earlier (British) [False Positive] study", "Despite being shown on alcohol-related harm as well as with young [False Positive] people in the USA") leading to the generation of FPs. It is safe to assume that such cases can be present in our larger study sample despite their low prevalence as FPs (4 in total). Using generic terms (e.g., "delinquent", "criminal", "adolescent) to capture the age and the offender type of populations led to the generation of one (i.e., "Maternal depression is a risk factor for adolescent [False Positive] depression") and eight FPs (e.g., "that drug use was more strongly related to disruptive and delinquent [False Positive] behavior, for both sexes", "drinking problems and criminal [False Positive] arrests were interrelated") respectively. This indicates that perhaps more specific terms in our dictionaries could potentially limit the generation of FPs on that front.

False negatives

The lack of implemented rules due to not previously seen syntactical patterns was as a source of FNs particularly in age (e.g., "majority of these incarcerated youth [False Negative] have one", "considerations for minors [False Negative] facing delinquency"), sex (e.g., "Males [True Positive]greatly outnumbered females [False Negative]", "males [True Positive] were less likely than females [False Negative] to have") and offender type (e.g., "which girls committed aggressive offenses [False Negative], "residential location of a serial offender [False Negative]"). Another source of errors was the presence of misspellings in the published text that did not trigger the respective rules for age (e.g., "compared between 12and14-year-old [False Negative] boys who attended a delinquency") and the lack of terms from our nationality (e.g., "offenders committed to Iowa Department of Corrections [False Negative]", "records of three private Minnesota [False Negative] adoption agencies") and offender type dictionaries (e.g., "individuals who are subject to a restraining order or have been convicted of a domestic violence misdemeanour [False Negative]", "serial commercial robber").

Limitations

Our study comes with several limitations. First, PubMed articles might not be sufficient to portray a complete picture of offending and incarcerated populations since government articles and reports can remain unpublished and so fall outside the scope of this study. Second, research with a sociological and criminological focus is unlikely to appear in journals covered by PubMed. Thus, our data set could potentially underestimate the total number of research outputs in this area. Third, we focused only on English abstracts which carried a risk of "English-language" bias. However, the incorporation of non-English abstracts in our PubMed sample could ensure greater research transparency and findings and reducing bias.

Fourth, using only abstract text almost certainly does not give a full picture with regards to the investigated population. As noted in our findings, only a fraction of abstracts reported any of the four characteristics we examined – offender type (13.9%), nationality (27.6%), age (14.9%), and sex (23.6%). It is likely that the use of full text articles, especially those that might adhere to official reporting guidelines (e.g., PRISMA, STROBE) detail their population reporting in the body text of the article which would elicit different findings than presented here. However, this was not feasible and would have involved permissions from numerous publishing houses and be extremely costly.

Finally, despite a reliable performance from our methodology, the number of identified characteristics could be underrepresented (especially for age and offence type). Using specific rules might not have been enough to identify mentions of age while more descriptive cases for offender types could have resulted in FNs.

Conclusions

Our study demonstrated that it is feasible and efficient to extract key information from populations within a large sample of justice health study abstracts over time. Our findings align with existing research that has highlighted a focus on female offender studies and have revealed an emphasis on mentally ill offenders and minors with rising rates the last thirty years. Interestingly, research involving child related offenders were not common. Despite the US having the largest incarcerated population in the world, adjusting its publication rate by the prisoner population demonstrates that Nordic countries with progressive approaches to offender rehabilitation have published proportionately more research. Our findings offer new insights in the whole area of justice health with clear implications to promote diversity in cohort selection and limitation of bias and research gaps.

Data Availability

The data sets used in this study are publicly available in PubMed by implementing our query described in the Data Sample subsection under the Methods section.

Authors' Contributions

WL contributed to study initialization, literature review, method implementation, result analysis, result interpretation, manuscript initialization and revision. TB was responsible for result interpretation, manuscript revision and study supervision. PT, PS, NG, IB, GN contributed to the result interpretation and manuscript revision. GK was responsible for study conception and initialization, method implementation, result interpretation, manuscript initialization and revision, study supervision. All authors contributed to the manuscript and approved the submitted version. No generative artificial intelligence was used in any part of the manuscript.

Conflicts of Interest

None declared.

Abbreviations

ADVO: Aggravated Domestic Violence Order
CONSORT: Consolidated Standards of Reporting Trials
GATE: General Architecture for Text Engineering
FN: False Negative
FP: False Positive
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials
STROBE: STrengthening the Reporting of OBservational studies in Epidemiology
TP: True Positive

References

1. Waltermaurer E, Akers T. Epidemiological criminology: Theory to practice: Routledge; 2014.
2. Akers TA, Lanier MM. "Epidemiological criminology": Coming full circle. American journal of public health. 2009;99(3):397-402.
3. National Institute of Justice. Challenges conducting research in prisons 2024. https://nij.ojp.gov/topics/articles/challenges-conducting-research-prisons.
4. Thelwall M, Simrick S, Viney I, Van den Besselaar P. What is research funding, how does it influence research, and how is it recorded? Key dimensions of variation. Scientometrics. 2023;128(11):6085-106.
5. Binswanger IA, Maruschak LM, Mueller SR, Stern MF, Kinner SA. Principles to guide national data collection on the health of persons in the criminal justice system. Public Health Reports. 2019;134(1_suppl):34S-45S.
6. Simpson PL, Guthrie J, Butler T. Prison health service directors' views on research priorities and organizational issues in conducting research in prison: Outcomes of a national deliberative roundtable. International Journal of Prisoner Health. 2017;13(2):113-23.
7. Jones RM, Manetsch M, Gerritsen C, Simpson AI. Patterns and predictors of reincarceration among prisoners with serious mental illness: a cohort study: modèles et prédicteurs de réincarcération chez les prisonniers souffrant de maladie mentale grave: une étude de cohorte. The Canadian Journal of Psychiatry. 2021;66(6):560-8.
8. Simundic A-M. Bias in research. Biochemia medica. 2013;23(1):12-5.
9. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. The lancet. 2007;370(9596):1453-7.
10. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Journal of Pharmacology and pharmacotherapeutics. 2010;1(2):100-7.
11. Chan A-W, Tetzlaff JM, Altman DG, Dickersin K, Moher D. SPIRIT 2013: new guidance for content of clinical trial protocols. The Lancet. 2013;381(9861):91-2.
12. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Bmj. 2021;372.
13. Langlois ÉV, Daniels K, Akl EA, Organization WH. Evidence synthesis for health policy and systems: a methods guide: World Health Organization; 2018.
14. Kinner SA, Young JT. Understanding and improving the health of people who experience incarceration: an overview and synthesis. Epidemiologic reviews. 2018;40(1):4-11.
15. Fazel S, Baillargeon J. The health of prisoners. The Lancet. 2011;377(9769):956-65.
16. Balzer LB. "All generalizations are dangerous, even this one."—Alexandre Dumas. Epidemiology. 2017;28(4):562-6.
17. Fornaro M, Anastasia A, Valchera A, Carano A, Orsolini L, Vellante F, et al. The FDA "black box" warning on

antidepressant suicide risk in young adults: more harm than benefits? Frontiers in psychiatry. 2019;10:294.
18. Glasziou P, Vandenbroucke J, Chalmers I. Assessing the quality of research. Bmj. 2004;328(7430):39-41.
19. Karystianis G, Buchan I, Nenadic G. Mining characteristics of epidemiological studies from Medline: a case study in obesity. Journal of biomedical semantics. 2014;5(1):22.
20. Karystianis G, Lukmanjaya W, Simpson P, Schofield P, Ginnivan N, Nenadic G, et al. An analysis of PubMed abstracts from 1946 to 2021 to identify organizational affiliations in epidemiological criminology: descriptive study. Interactive journal of medical research. 2022;11(2):e42891.
21. Jonnalagadda SR, Goyal P, Huffman MD. Automating data extraction in systematic reviews: a systematic review. Systematic reviews. 2015;4:1-16.
22. Tsafnat G, Glasziou P, Karystianis G, Coiera E. Automated screening of research studies for systematic reviews using study characteristics. Systematic reviews. 2018;7(1):64.
23. Hara K, Matsumoto Y. Extracting clinical trial design information from MEDLINE abstracts. New Generation Computing. 2007;25(3):263-75.
24. Chung GY. Sentence retrieval for abstracts of randomized controlled trials. BMC medical informatics and decision making. 2009;9:1-13.
25. Kiritchenko S, De Bruijn B, Carini S, Martin J, Sim I. ExaCT: automatic extraction of clinical trial characteristics from journal publications. BMC medical informatics and decision making. 2010;10:1-17.
26. Tooth L, Ware R, Bain C, Purdie DM, Dobson A. Quality of reporting of observational longitudinal research. American journal of epidemiology. 2005;161(3):280-8.
27. Xu R, Garten Y, Supekar KS, Das AK, Altman RB, Garber AM. Extracting subject demographic information from abstracts of randomized clinical trial reports. Medinfo. 2007;129:550-4.
28. Hansen MJ, Rasmussen NØ, Chung G. A method of extracting the number of trial participants from abstracts describing randomized controlled trials. Journal of Telemedicine and Telecare. 2008;14(7):354-8.
29. Luo Z, Johnson SB, Lai AM, Weng C, editors. Extracting temporal constraints from clinical research eligibility criteria using conditional random fields. AMIA annual symposium proceedings; 2011: American Medical Informatics Association.
30. Luo Z, Miotto R, Weng C. A human–computer collaborative approach to identifying common data elements in clinical trial eligibility criteria. Journal of biomedical informatics. 2013;46(1):33-9.
31. Karystianis G, Thayer K, Wolfe M, Tsafnat G. Evaluation of a rule-based method for epidemiological document classification towards the automation of systematic reviews. Journal of biomedical informatics. 2017;70:27-34.
32. Karystianis G, Simpson P, Lukmanjaya W, Ginnivan N, Nenadic G, Buchan I, et al. Automatic Extraction of Research Themes in Epidemiological Criminology From PubMed Abstracts From 1946 to 2020: Text Mining Study. JMIR formative research. 2023;7:e49721.
33. Canese K, Weis S. PubMed: the bibliographic database. The NCBI handbook. 2013;2(1).
34. Karystianis G, Lukmanjaya W, Buchan I, Simpson P, Ginnivan N, Nenadic G, et al. An analysis of published study designs in PubMed prisoner health abstracts from 1963 to 2023: a text mining study. BMC Medical Research Methodology. 2024;24(1):68.
35. Last JM. Dictionary of epidemiology. CMAJ: Canadian Medical Association Journal. 1993;149(4):400.
36. Criminal Lawyers Australia. List of offences 2024. https://www.criminal-lawyers.com.au/offences.
37. Wikipedia. List of existing nationalities 2024. https://en.wikipedia.org/wiki/Lists_of_people_by_nationality.
38. Kim JD TJ. Corpora and their annotations. In: Ananiadou S MJ, editor. Text Mining for Biology and Biomedicine: Artect House; 2006.
39. Cunningham H, Tablan V, Roberts A, Bontcheva K. Getting more out of biomedical documents with GATE's full lifecycle open source text analytics. PLoS computational biology. 2013;9(2):e1002854.
40. Australian Institute of Health and Welfare. Australia's children—in brief. : Cat. no. CWS 72. Canberra: AIHW.; 2019. https://www.aihw.gov.au/reports/children-youth/australias-children-in-brief/summary.
41. Judicial Commission of New South Wales. Children's Court of NSW Resource Handbook - What makes juvenile offenders different from adult offenders 2023.

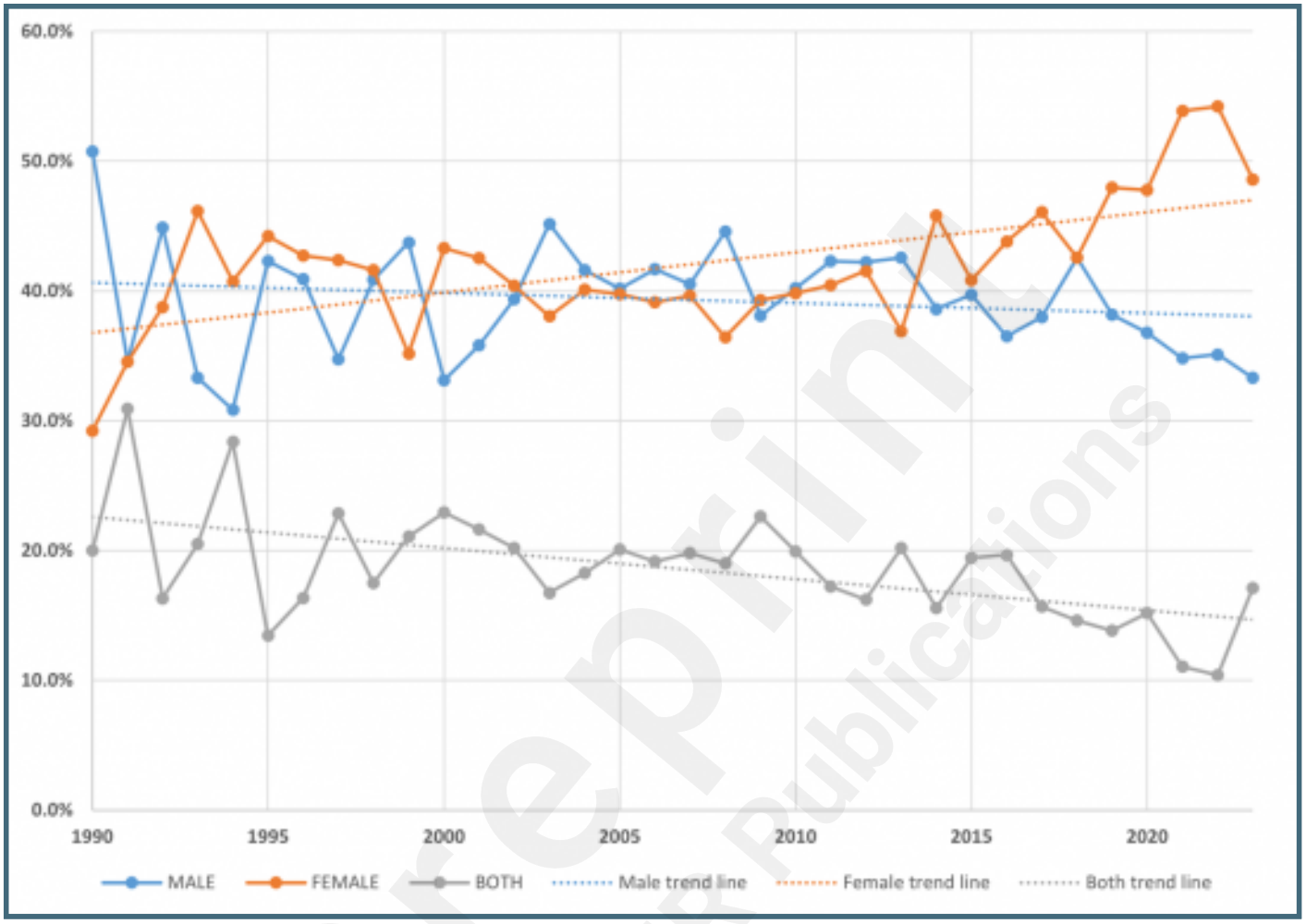
https://www.judcom.nsw.gov.au/publications/benchbks/children/cm_juvenile_offenders_different.html#ftn.d5e25761.
42. Australian Institute of Health and Welfare. Youth detention population in Australia 2023 2023. https://www.aihw.gov.au/reports/youth-justice/youth-detention-population-in-australia-2023/report-editions.
43. Australian Institute of Health and Welfare. Australia's Youth 2021. https://www.aihw.gov.au/reports/children-youth/australias-youth/contents/demographics.
44. Australian Institute of Health and Welfare. Making progress: the health, development and wellbeing of Australia's children and young people. Cat. no. PHE 104. Canberra: AIHW; 2008.
45. Ananiadou S, McNaught J. Text mining for biology and biomedicine: Citeseer; 2006.
46. World Prison Brief. 2024. https://www.prisonstudies.org/
47. Gendreau P, Little T, Goggin C. A meta-analysis of the predictors of adult offender recidivism: What works! Criminology. 1996;34(4):575-608.
48. Harms P. Detention in delinquency cases, 1989-1998: US Department of Justice, Office of Justice Programs, Office of Juvenile; 2002.
49. Hockenberry S, Puzzanchera C. Juvenile court statistics 2013. 2015.
50. Australian Bureau of Statistics. Prisoners in Australia. 2023.
51. United Nations Office on Drugs and Crime. Data Matters 2021. https://www.unodc.org/documents/data-and-analysis/statistics/DataMatters1_prison.pdf.
52. Institute for Crime and Justice Policy Research. World Female Prison Population Up By 60% Since 2000 2022. https://www.icpr.org.uk/news-events/2022/world-female-prison-population-60-2000.
53. Anand S, Peter F, Sen A. Public health, ethics, and equity: OUP Oxford; 2004.
54. Iacobucci G. "Levelling up" plan needs more funding and a focus on health inequalities, say experts. British Medical Journal Publishing Group; 2022.
55. Johnsen B, Granheim PK, Helgesen J. Exceptional prison conditions and the quality of prison life: Prison size and prison culture in Norwegian closed prisons. European Journal of criminology. 2011;8(6):515-29.
56. Lappi-Seppälä T, Tonry M. Crime, criminal justice, and criminology in the Nordic countries. Crime and Justice. 2011;40(1):1-32.
57. James DJ, Glaze LE. Mental health problems of prison and jail inmates. 2006.
58. Grande T, Hallman J, Caldwell K, Underwood L. Using the BASC-2 to assess mental health needs of incarcerated juveniles: Implications for treatment and release. Corrections Today. 2011;73(5):100-2.
59. Wald J, Losen DJ. Defining and redirecting a school-to-prison pipeline. New directions for youth development. 2003;2003(99):9-15.
60. Underwood LA, Washington A. Mental illness and juvenile offenders. International journal of environmental research and public health. 2016;13(2):228.
61. O'keefe ML, Schnell MJ. Offenders with mental illness in the correctional system. Journal of Offender Rehabilitation. 2007;45(1-2):81-104.
62. Skeem JL, Winter E, Kennealy PJ, Louden JE, Tatar II JR. Offenders with mental illness have criminogenic needs, too: toward recidivism reduction. Law and human behavior. 2014;38(3):212.
63. Piquero NL, Weisburd D. Developmental trajectories of white-collar crime. The criminology of white-collar crime: Springer; 2009. p. 153-71.
64. Restum ZG. Public health implications of substandard correctional health care. American Journal of Public Health. 2005;95(10):1689-91.
65. Rayirala A, Bunga D, Rani KS, Umashankar M. Psychiatric morbidity in violent and nonviolent offenders: A cross-sectional comparative study. Archives of Mental Health. 2019;20(2):55-60.
66. O'Malley RL. Short-term and long-term impacts of financial sextortion on victim's mental well-being. Journal of interpersonal violence. 2023;38(13-14):8563-92.
67. Button M, Lewis C, Tapley J. Not a victimless crime: The impact of fraud on individual victims and their families. Security Journal. 2014;27:36-54.
68. Zimring FE. An American travesty: Legal responses to adolescent sexual offending: University of Chicago Press; 2009.

69. Rosselli MK, Jeglic EL. Factors impacting upon attitudes toward sex offenders: The role of conservatism and knowledge. Psychiatry, Psychology and Law. 2017;24(4):496-515.
70. Dombert B, Schmidt AF, Banse R, Briken P, Hoyer J, Neutze J, et al. How common is men's self-reported sexual interest in prepubescent children? The Journal of Sex Research. 2016;53(2):214-23.
71. Krahé B, Tomaszewska P, Kuyper L, Vanwesenbeeck I. Prevalence of sexual aggression among young people in Europe: A review of the evidence from 27 EU countries. Aggression and Violent Behavior. 2014;19(5):545-58.
72. Australian Bureau of Statistics. Recorded Crime - Offenders [Internet]. Canberra: ABS; 2022-23.
73. Bierie DM, Davis-Siegel JC. Measurement matters: Comparing old and new definitions of rape in federal statistical reporting. Sexual Abuse. 2015;27(5):443-59.
74. Australian Bureau of Statistics. 97 per cent of sexual assault offenders are male [Internet]. Canberra: ABS. 2022.
75. Cortoni F, Babchishin KM, Rat C. The proportion of sexual offenders who are female is higher than thought: A meta-analysis. Criminal Justice and Behavior. 2017;44(2):145-62.
76. World Health Organization. Violence against children 2022. https://www.who.int/news-room/fact-sheets/detail/violence-against-children#:~:text=Globally%2C%20it%20is%20estimated%20that,lifelong%20health%20and%20well%2Dbeing.

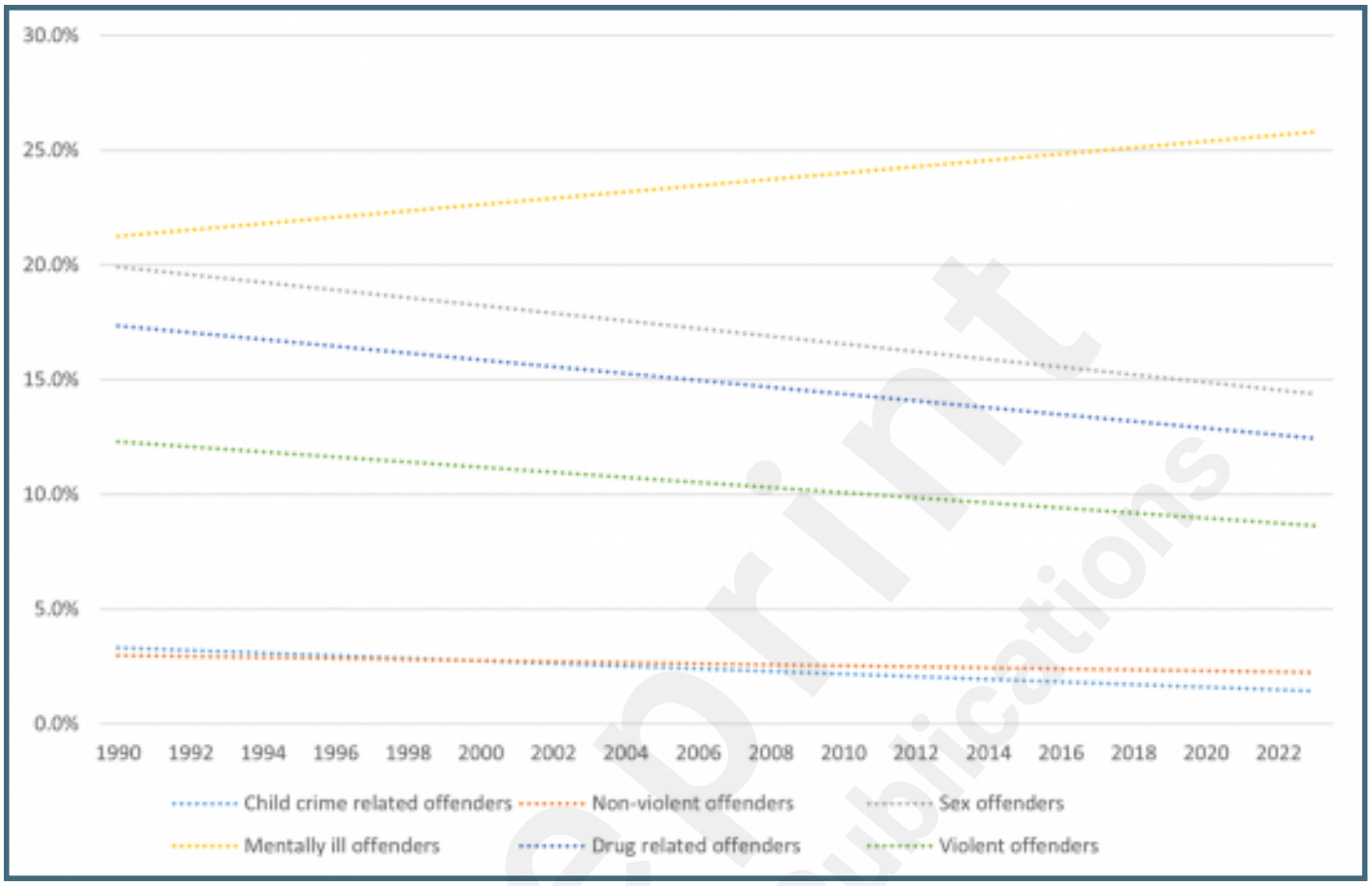
Supplementary Files

Figures

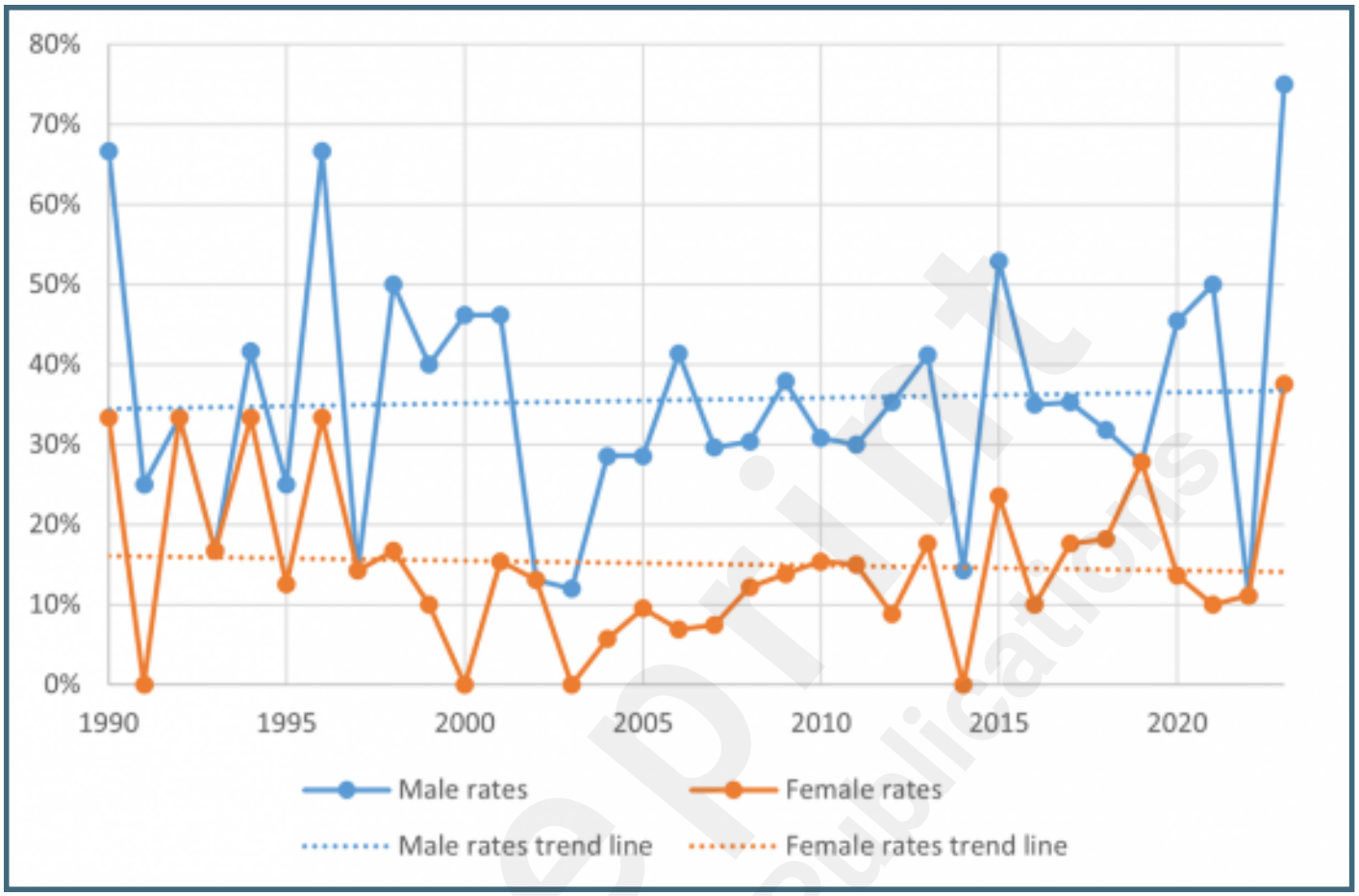
Proportions of PubMed study abstracts that reported female only and male only populations from 1990 to 2023. Note: Due to the very low rates for transgender populations, these were not included in the graph.



Trend lines for the rates of the six offender types in justice health abstracts (n=4,814) in PubMed from 1990 to 2023.



Males and females sex offender rates in justice health abstracts (n=4,814) in PubMed from 1990 to 2023.



Multimedia Appendixes

Terms used to identify the population's sex in PubMed abstracts.
URL: http://asset.jmir.pub/assets/88d5913fb1c89d1a281bcf7e5eb68bdb.docx

Terms used to describe offending and incarcerated populations in PubMed abstracts.
URL: http://asset.jmir.pub/assets/c9ca72c5457cc073555559eb5995eaae.docx

Rule examples for each population (i.e., age, sex, nationality, offender type) characteristic.
URL: http://asset.jmir.pub/assets/6a7d06661fa55297cea4c39894264d5a.docx

Miscellaneous terms used to describe offending and incarcerated populations in PubMed abstracts.
URL: http://asset.jmir.pub/assets/15cb6090744ba009c09a526bf445fcc7.docx

Number of published articles (n = 34,481) in PubMed related to epidemiological criminology from 1963 to 2023.
URL: http://asset.jmir.pub/assets/9a00f526359ce1907b5653fa35559edb.docx

Rates of justice health abstracts (n=8,169) in PubMed that reported male only, female only, trans only and both female and male populations from 1990 to 2023.
URL: http://asset.jmir.pub/assets/48c4620b28d87efa005daf26136d6ed8.docx

Number of justice health abstracts (n=4,814) in PubMed that mention one offender characteristic in one of the six defined group classes.
URL: http://asset.jmir.pub/assets/45cda202c80bbd187f32a14bac97e3b6.docx

Mental health concepts related to offending populations in justice health PubMed abstracts.
URL: http://asset.jmir.pub/assets/561854c62a9d226ac84071afd56b61d7.docx