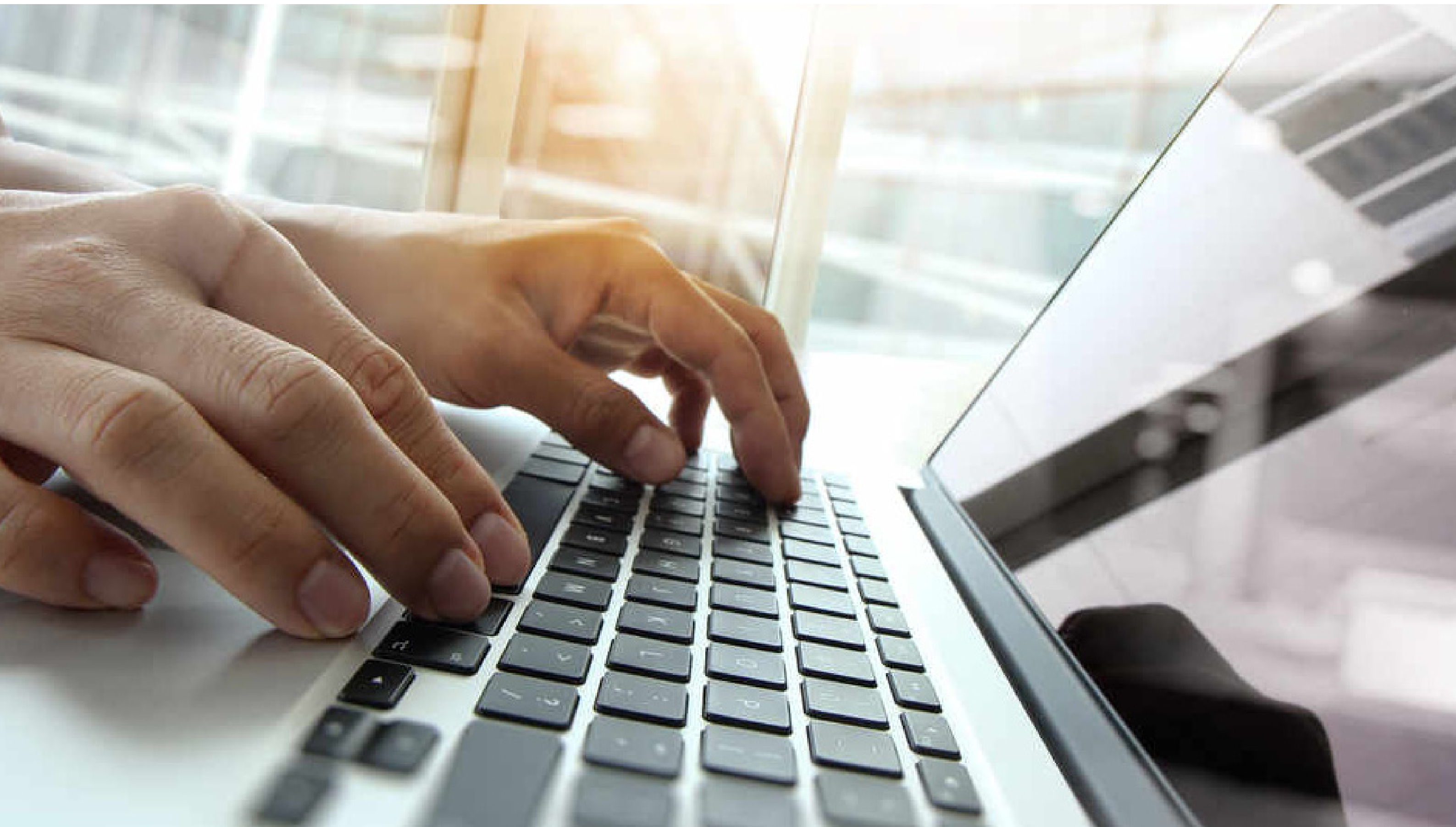
The Military Service Identification Tool:

A utility, feasibility and validation study for identifying veterans from a secondary mental health case register and comparing their sociodemographic, diagnostic and service utilisation characteristics to civilians and Combat Stress veterans

October 2022



SANCTE ET SAPIENTER

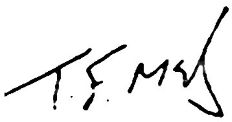
Contents

Figures

Tables

Foreword

The Government's own Veterans Strategy Action Plan recognises the paucity of data on veteran physical and mental health and on their wider well-being. The lack of it hampers targeted approaches to adequately providing for their needs and serves to contribute to a sense of misunderstanding and sometimes neglect within society. In particular, capturing, characterising and mapping the long-term mental health of veterans has always been challenging, with information spread across the spectrum of healthcare services, including those in devolved nations. To date, few studies have been able to examine data that illuminates the mental health needs of UK veterans accessing formal healthcare services. A lack of these data capture has surely contributed to knowledge gaps in the needs profiling of veterans within the UK healthcare system with negative consequences for the effective development of potential treatment pathways. Beginning in 2017, the study started as an exploration of treatment pathways for veterans accessing secondary mental health care services, and unfolded into a feasibility study of using the Military Service Identification Tool as a means of identifying veterans from a secondary mental health care register. This is potentially significant. The first study showed it is feasible to manually identify whether patients who accessed secondary mental health care services have served in the military. The possibility of automating this process could mean it will be quicker and easier to identify veterans for research purposes which should in turn galvanise new initiatives to address the current data and knowledge deficit. This feasibility study is a UK first in attempting to identify military veterans and to explore the characteristics of veterans and matched civilians who have sought help from a secondary mental health care Trust in England. Forces in Mind Trust's mission is that all ex-Service persons and their families make a successful and sustainable transition to civilian life; our Health Programme policy goal is that all veterans and their families are able to access good quality health and social care services when and where they need them. The possibility of realising that goal depends on the development of evidence-based policy and practice, built upon sound data capture. This report indicates that automated means can be developed to help, though more work is needed to make veteran health data capture the norm. But the widely held view of clinical experts that veterans are likely to experience greater comorbidities, seeking help only at crisis point, and exhibiting riskier behaviours linked to their mental health problems places a high priority on fixing those gaps. Therefore, this report should not be the last that seeks to address this problem.



Tom McBarnet
Chief Executive (Acting), Forces in Mind Trust

King’s Centre for Military Health Research, King’s College London

Previously the Gulf War Illness Research Unit, the King’s Centre for Military Health Research (KCMHR) was launched in 2004 as a joint initiative between the Institute of Psychiatry, Psychology and Neuroscience (IoPPN) and the Department of War Studies at King’s College London. KCMHR draws upon the experience of a multi-disciplinary team and is led by Professor Sir Simon Wessely and Professor Nicola T. Fear. The Centre undertakes research investigating military life using quantitative, qualitative and digital methods. Its flagship study is a longitudinal investigation of the health and well-being of the United Kingdom’s (UK) Armed Forces personnel who served during the recent conflicts in Iraq and Afghanistan. This study, funded by the UK Ministry of Defence (MoD), has been running since 2003 and completed its third phase of data collection in 2017. Data from our studies have been used to analyse various military topics and papers have been published in peer reviewed, scientific journals. Our findings are regularly reported in the press and have also been used to inform policies that impact health and well-being of the Armed Forces Community.

Authors

Dr Laura Palmer
Dr Daniel Leightley
Ms Charlotte Williamson
Dr Katharine Mark
Mr David Pernet
Ms Elena Opie
Professor Dominic Murphy
Dr Sharon Stevelink
Professor Nicola T. Fear

Acknowledgements

This research would not have been possible without the veterans and civilians who presented themselves to mental health care services through both the National Health Service (NHS) and the military charity Combat Stress. By seeking professional help, they provided the data upon which this research is built.

The study involved support from the National Institute for Health Research (NIHR) Maudsley Biomedical Research Centre (BRC). This department is a partnership between the South London and Maudsley NHS Foundation Trust and the IoPPN at King's College London. We would particularly like to thank David Chandran (Postdoctoral Informatician), Megan Pritchard (lead in Clinical Record Interactive Search (CRIS) training and development), Debbie Cummings (CRIS administrator), Daisy Kornblum (Clinical Informatician), Karen Birnie (Student helper) and Larisa Maria (Student helper).

We are grateful to the Forces in Mind Trust for funding the research and for their continued support and engagement throughout the project, particularly from Tom McBarnet (Chief Executive (Acting), Kirsteen Waller (Health Programme Manager), Caroline Cooke (Head of Policy), Isabel Summers (Assistant Head of Policy) and former members of the FiMT team including Ray Lock (former Chief Executive) and Harry Palmer (Mental Health Research Programme Coordinator from the Centre for Mental Health).

Publications

To date, the following academic articles have been prepared using data from this study.

- **Identifying veterans using electronic health records in the United Kingdom: A feasibility study.** Katharine M. Mark, Daniel Leightley, David Pernet, Dominic Murphy, Sharon A.M. Stevelink and Nicola T. Fear. *MDPI Healthcare, 2019.*
- **The Development of the Military Service Identification Tool: Identifying Military Veterans in a Clinical Research Database using Natural Language Processing and Machine Learning.** Daniel Leightley, David Pernet, Sumithra Velupillai, Robert J. Stewart, Katharine M. Mark, Elena Opie, Dominic Murphy, Nicola T. Fear, Sharon A.M. Stevelink. *JMIR Medical Informatics, 2020.*

Glossary of terms

Civilian – A person who has not served in the Armed Forces.

Clinical Record Interactive Search (CRIS) – A system which has been developed for use within the NIHR Maudsley Biomedical Research Centre (BRC). It provides authorised researchers with regulated, secure access to pseudonymised information extracted from South London and Maudsley NHS Foundation Trust electronic clinical records system.

Consent for Contact (C4C) – Patients with C4C have joined a register to allow researchers to look at their Trust medical records and to contact them for recruitment into research studies.

Electronic Healthcare Record (EHR) – A digitalised version of a patient's health information that includes medical notes, medication and diagnoses.

Gold standard dataset In statistics a gold standard dataset refers to a diagnostic, test or benchmark that is the best available under reasonable conditions.

Improving Access to Psychological Therapies (IAPT) – A programme which began in 2008 with the objective to improve access for people with anxiety and depression, including obsessive-compulsive disorder, to evidenced based psychological therapies, such as cognitive behavioural therapy.

Interquartile Range – The range of values between the first and third quartiles of a distribution.

Machine learning – Machine learning generates sophisticated statistical models that describe behaviours and patterns in data at an abstract level. There are multiple branches of machine learning; this study used supervised machine learning. Supervised machine learning is capable of automatically learning from data (usually a set of variables, which represent a single dimension of the data), making predictions on what has been observed. This is achieved by using labelled training data (referred as input data) to learn data distributions and patterns, and when presented with testing data (which does not contain a label), the algorithm predicts which label it believes is the outcome.

Military Service Identification Tool (MSIT) – A computer program designed to identify and flag individuals who have, or are, serving in any Armed Forces using medical notes.

Median – The middle number of a sorted list of numbers.

Natural Language Processing (NLP) – The application of a computer program to automatically analyse human text and speech.

Probable [diagnoses] – The term probable is used in the absence of a clinical diagnosis and is based on participants completing a self-report measure such as a questionnaire or taking part in an interview.

Sensitivity – A statistical term reflecting a test's ability to correctly identify positive cases of an outcome.

Specificity – A statistical term reflecting a test's ability to correctly identify negative cases of an outcome.

Statistical significance – Statistical hypothesis testing, using p values, is used to determine whether the relationship between variables is 'significant', i.e., unlikely to be due to chance alone. If the p value falls below 0.05, a statistically significant result has been found. If the p value exceeds 0.05, a non-statistically significant result has been found.

Structured Query Language (SQL) – A commonly used computer language used to query, insert, update or modify data that is stored in a database.

Text mining – Text mining, sometimes referred to as text data mining, is the process of deriving structured information from text created by humans. Information is typically derived using Natural Language Processing and statistical modelling.

Veteran – Anyone who has served for at least one day in Her Majesty's Armed Forces (Regular or Reserve) or Merchant Mariners who have seen duty on legally defined military operations.

Executive Summary

Introduction

A considerable number of military veterans affected by mental health problems fail to seek professional help for their difficulties. Previous research on United Kingdom (UK) veterans who engage in treatment tells us mainly about those accessing primary care services. Although it is useful to investigate the profile of veterans who initiate their mental health treatment through primary care, there is a relative lack of quantitative evidence concerning those who accessed secondary mental health care services – that is, more specialist mental health care, frequently delivered in higher intensity therapeutic specialist clinics or hospitals.

Research objectives

This research used an electronic health record (EHR) Case Register to explore the utility and feasibility of identifying veterans who accessed secondary mental health care services within the UK. The identification of veterans was validated by contacting patients classified as veterans/civilians and confirming their status via self-reported surveys. In addition, the report compared the sociodemographic profiles and the types of mental health diagnoses among veterans who engage in secondary mental health care services compared to their matched civilian counterparts. An additional aim was to explore the service utilisation of both veterans and civilians, however a large volume of missing data made this impracticable. We were, however, able to compare use of the Mental Health Act (1983) in veterans and civilians. To provide comparison to other veteran services, a sub-set of veterans identified through the Case Register were compared to a veteran sample at Combat Stress, a national veterans charity, and analysed.

Method

The South London and Maudsley (SLaM) Biomedical Research Centre (BRC) Case Register, an EHR database, provided data on civilians and veterans who had accessed secondary mental health care services through the SLaM National Health Service (NHS) Foundation Trust. The study procedure involved:

1) developing a manual approach to identifying veterans from the SLaM BRC Case Register;
2) developing a Natural Language Processing (NLP) Military Service Identification Tool (MSIT) to automatically detect veterans from the SLaM BRC Case Register;
3) describing the utility and feasibility of identifying veterans using a manual approach and MSIT;

4) using self-reported surveys to validate whether the veterans/civilians identified by MSIT in the SLaM BRC Case Register had in fact been correctly identified;
5) matching veterans from the SLaM BRC Case Register to a civilian sample on age and gender;
6) matching veterans from the SLaM BRC Case register to a cohort of veterans from Combat Stress on age and gender;
7) extracting data on sociodemographic, diagnostic, clinical and treatment characteristics;
8) backfilling missing data on key sociodemographic variables; and
9) describing the similarities and differences between civilians and veterans in those who had accessed secondary mental health care services.

Utility, feasibility and validation results

This study developed two complementary methods to identify veterans from the SLaM BRC Case Register. The first used a manual approach identifying keywords that are commonly used to describe military service within free-text clinical notes. The second, MSIT, used NLP and machine learning to automatically analyse and classify free-text clinical notes. We found that both approaches were feasible in identifying veterans. However, practical limitations were present: accessing the Case Register was lengthy and involved various administrative hurdles and data security issues.

Manual identification of veterans from the SLaM BRC Case Register was particularly labour intensive. It involved systematically searching the database, using military-related phrases and exclusion criteria, and scrutinising individual records one-by-one. In contrast, MSIT was able to search through large volumes of free-text clinical notes and identify veterans with high precision and, in a few cases, with better accuracy than human coders.

A total of 1,600 individual records were manually evaluated, with the percentage of true veterans identified (i.e. as opposed to the percentage of non-veterans identified) being 43% overall. The study team was cautious regarding who to classify as a veteran by reading through all free-text clinical notes at least twice and only confirmed veteran status when an explicit statement about the patient serving in the Armed Forces was reported by the clinician. The term 'Royal Air Force' correctly identified veterans most often during the searches. Using MSIT, 150,000 individual records were inspected, automatically, with the percentage of true veterans identified being 88% overall.

We further validated MSIT by surveying a subsample of patients within the Clinical Record Interactive Search (CRIS) system (n=146) to determine their self-reported veteran/civilian status and compared this to the MSIT's classifications. We found that 83.6% of the sample were accurately categorised by MSIT. The overall sensitivity of the tool (that is, the ability to detect true civilians) was found to be 0.83, and the specificity was 0.92 (that is, the ability to detect true veterans). An examination of the exceptional cases where MSIT misclassified veteran/civilian status showed that MSIT tended to categorise civilians as veterans (n=23 compared to n=1 veteran inaccurately categorised as a civilian). A manual search of the misclassifications identified keywords to further train the tool to prevent misclassifications of false veterans, e.g. mentions of Salvation Army. Due to the high sensitivity and specificity (>0.80), no substantial changes to the tool were required.

MSIT therefore represents a large saving in human resources, cost, time and manpower required to identify who is a veteran; MSIT is able to run in minutes, whereas manual annotation can take between 6 – 16 minutes on average per individual. Whilst MSIT could be a valuable research tool for future use in SLaM and potentially other trusts, we acknowledge that applying MSIT presents logistical and technical challenges. Examples of these include the unavoidable reliance upon the SLaM NHS Foundation Trust's administration teams to run the tool and extract the data on our behalf; the lengthy and iterative processes required for researchers to obtain ethical approvals, and potentially amendments, from the NHS and relevant Research & Development offices and to obtain access to the patient databases.

Applying the MSIT and the manual approach to the SLaM Case Register identified 2,922 veterans who accessed secondary mental health care services through the SLaM NHS Foundation Trust. Of this sample, 1,288 served in the Armed Forces after National Service conscription was phased out in May 1963. To provide new insights, post National Service era veterans were matched on age and gender to known civilian counterparts for comparison. The final sample size for analyses of sociodemographic, diagnostic, clinical and treatment characteristics was 1,288 civilians and 1,288 veterans.

Extracting, cleaning and analysing data from the SLaM Case Register revealed large amounts of missing data. For example, only 63% of the sociodemographic variables were complete. In order to make optimal use of the notes available in the Case Register, including free-text clinical notes, clinical events and admission notes, variables were systematically backfilled by the research team in a bespoke database, increasing completeness for sociodemographic data to 76% which were then used for analyses.

Veteran group comparison results

Most of the age and gender matched civilians and veterans who accessed secondary mental health care through SLaM NHS Foundation Trust were white, single or separated, with a median age of 41 years.

In terms of sociodemographic variables, many civilians and veterans reported living alone. Veterans were significantly more than likely to live with a partner and/or child than civilians and were significantly less likely to live with their parents.

Veterans were significantly more likely to be given an anxiety, stress, depressive, psychosis or personality disorder diagnosis, whereas civilians were significantly more likely to be given a drug disorder diagnosis. The analysis further indicated that veterans were significantly more likely to have been sectioned under the Mental Health Act (1983) when compared to civilians. Further research is required to ascertain if veterans are at higher risk of being sectioned nationally.

Combat Stress comparison results

A total of 189 veterans identified from the SLaM Case Register were matched on age and gender to 189 veterans from Combat Stress. Most of the age and gender matched veterans who accessed secondary mental health care through both the SLaM NHS Foundation Trust and through Combat Stress were male, which is unsurprising given the mainly male demographic composition of the military. SLaM veterans and Combat Stress veterans had a median age of 40 years.

Analyses indicated that SLaM veterans were significantly more likely to live alone and to be single than Combat Stress veterans. Combat Stress veterans were significantly more likely to live with their partner/children and to be in a relationship than SLaM veterans. Combat Stress veterans were significantly more likely to be of British ethnicity than SLaM veterans.

SLaM veterans were significantly more likely to be given a drug disorder diagnosis, whereas Combat Stress veterans were significantly more likely to be given a depressive, anxiety, stress or alcohol disorder diagnosis.

Discussion

This study is the first in the UK to identify military veterans and to explore veterans and matched civilians who have sought help from a secondary mental health care Trust in England. This research used a Case Register to explore the utility and feasibility of identifying veterans who accessed secondary mental health care services, using manual and automated approaches.

MSIT's predictions of veteran/civilian status were manually checked against electronic healthcare records, and were 97% accurate. When MSIT predictions were compared to participants' disclosed veteran/civilian status, 84% of MSIT's classifications were accurate. MSIT is therefore substantially better than other approaches available, such as using a Structured Query Language search strategy to manually search healthcare records (accuracy= 43%).

Once veterans had been identified, they were matched on age and gender to a civilian cohort extracted from the Case Register. Comparing these two samples presented interesting findings.

More than half of those in secondary mental health care services for both civilians and veterans were White British, and the majority male. This follows a similar profile of the Armed Forces (Fear et al., 2010; Hotopf et al., 2006; Stevelink et al., 2018). Most civilians and veterans reported living alone; previous research has indicated that those who live alone utilise health care services more frequently (Dreyer et al., 2018). Whilst the mechanisms driving higher utilisation are not yet known, it is possible that formal support is sought in the absence of immediate informal support networks.

There is emerging evidence that where people live is an important factor in determining and sustaining inequalities in mental health outcomes (Fone et al., 2007). Just over half of civilians and veterans lived in an area of severe deprivation, which may have impacted negatively on their mental health outcomes (Fone & Dunstan, 2006).

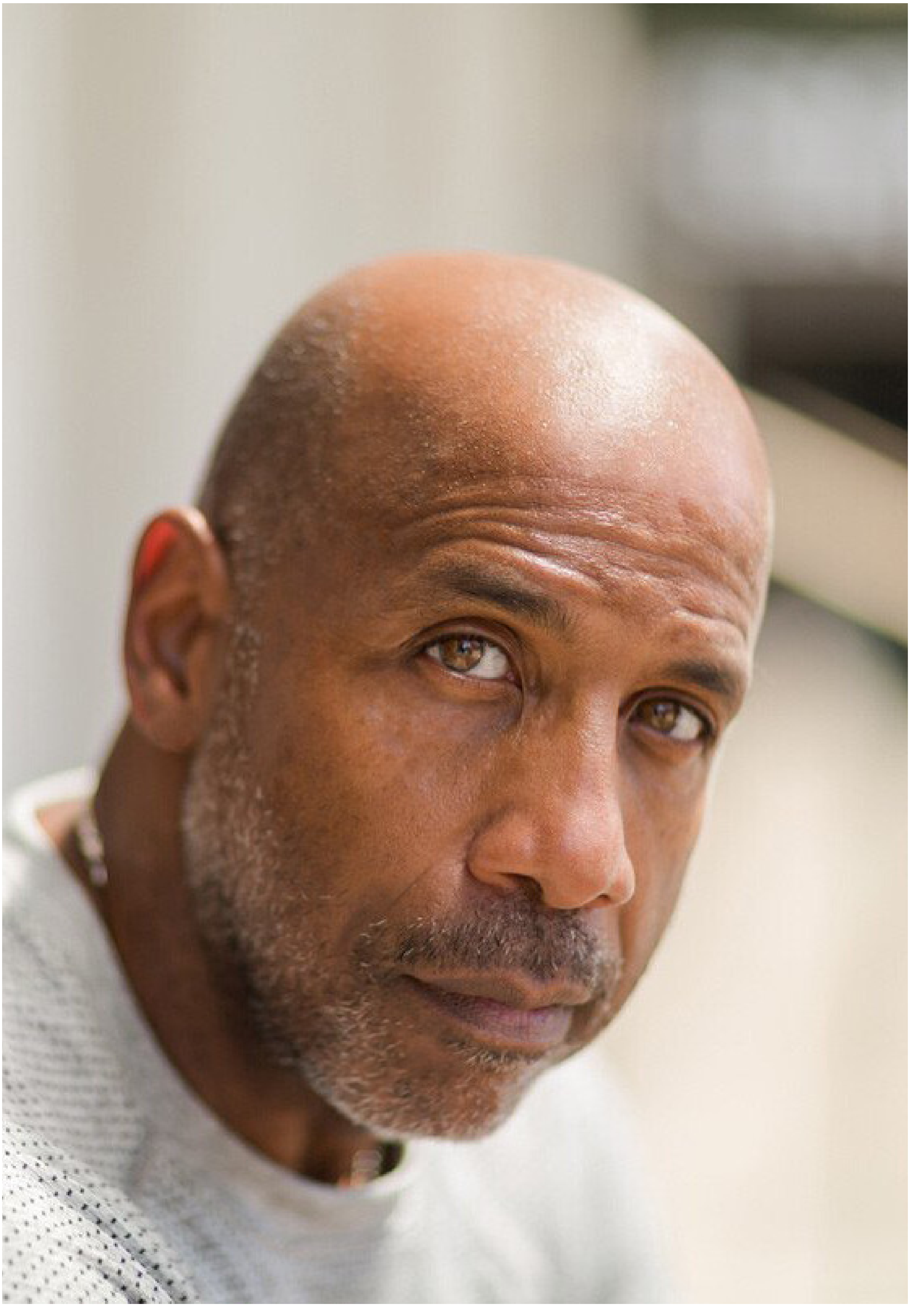
This study found that SLaM veterans were more likely to be given stress, depressive, anxiety, psychosis and personality disorders than civilians. However, we did not find any differences in alcohol use disorder between civilians and veterans despite the literature showing alcohol use is more prevalent in this group (Rhead et al., 2019; Stevelink et al., 2019).

Finally, this study suggested significant differences between civilians and veterans for use of Mental Health Act (1983) sectioning powers, with veterans being significantly more likely to be sectioned than civilians. This could be due to veterans experiencing a higher number of comorbidities, seeking help only at crisis point, having known risk factors for mental health problems (i.e. isolation, living alone, unemployed) and being perceived to be risker in their behaviours (Stevelink et al., 2018; Stevelink et al., 2019; Rhead et al., 2019).

Recommendations

The results of this research have implications for the ways in which veterans receive secondary mental health care services, and in our understanding of how they use these kinds of services. To ensure a broad and realistic discussion of the implications of this research, a stakeholder event was held towards the end of this project. Representatives from secondary mental health care providers, providers of veterans' mental health treatment and support and academics attended. As a result of this event, and the finding of this report, this study suggests the following:

1. We recommend improving the accuracy and efficiency of identifying veterans from the Case Register by ensuring that serving status is asked when a patient is registered;
2. We recommend that the collection of core socio-demographic variables is mandated for all Case Registers;
3. We recommend the implementation of new techniques to minimise missing data, such as mandatory fields in forms and the routine sharing of information between hospital systems;
4. We recommend accelerating the methodology for identifying veterans from the Case Register through further development of the Military Service Identification Tool. Further points for development of the tool are outlined on p.48;
5. We recommend conducting further analysis on the prevalence of mental health problems and how they compare to civilians and further statistical tests on data points available within the Case Register to understand patients' transition between different secondary mental health care services;
6. We recommend conducting further research to assess the generalizability and scalability of our findings at a local, regional and national level;
7. We recommend that future work is conducted to explore the needs of National Service Era veterans;
8. We recommend increasing the number of veterans identified from the Case Register by including a larger number of patient records. This will ensure a large enough sample size for more complex civilian and veteran comparisons;
9. We recommend developing an educational tool for those involved in the care of veterans to highlight their mental health needs.



Chapter One:
Introduction

The UK's veteran population, defined by the British government as those who have been in military service for at least one day (Ministry of Defence, 2016), is estimated to be 2.5 million (Ministry of Defence, 2019). A minority of veterans may experience mental health problems, with estimates ranging from 7% to 22% across psychiatric conditions that include post-traumatic stress disorder (PTSD) and common mental disorders, with some resulting from their experiences in the line of duty (Stevelink et al., 2018). Indeed, when a nationally representative sample of 257 English veterans, between the ages of 16 and 64 years of age, was compared to an age and gender matched sample of 504 non-veterans living in the community in England, the former were found to display more violent behaviour and more suicidal thoughts (Woodhead et al., 2011).

Factors influencing the choice not to access formal treatment can include: a lack of recognition that individuals have mental health problems; their belief that symptoms are not severe enough; a propensity to favour informal over formal sources of help; preferring to deal with problems themselves; fearing adverse occupational consequences; problems accessing services; and concern about the stigma associated with mental ill health (Fear et al., 2010; Giebel et al., 2014; Sharp et al., 2015). Recent research suggests that the latter is only a major issue at treatment initiation when first accessing mental health care services (Sharp et al., 2015).

Of the veterans who do seek formal treatment for mental health issues, around 20% receive medication or counselling (Iversen et al., 2010), typically by entering the health care system at the primary care[1] level within the NHS. The limited existing literature in this field tells us that the profile characteristics for most veterans who present to a primary care settings for mental health treatment are white males who served in the Army at lower ranks (Palmer, 2012; Perera et al., 2009; van Hoorn et al., 2013). However, while it is important to know the profile of those who take the first step to addressing their problem through primary care, research is limited when we consider veterans who access secondary mental health care[2] services.

Moreover there is a lack of quantitative information concerning the profile of veterans which includes whom is receiving secondary mental health care, the type of conditions they present with, the type of diagnoses they receive, the treatment they receive, or how this compares to other participant groups (such as civilians). The work that has been done to date has focused on third sector organisations (Mellotte et al., 2017). This is true for both NHS-based services and military specific services offering treatment, such as veteran charities (i.e. Combat Stress, Help for Heroes).

While healthcare systems are improving to routinely record veteran status, records are lacking at both primary and secondary care levels. MSIT therefore presents a useful solution to identify veterans currently accessing, and who have historically accessed, such services.

[1] *Primary care settings refer to settings like general practitioner surgeries or low-level therapeutic clinics.*
[2] *Secondary mental health care refers to more specialist care, often delivered in higher intensity therapeutic clinics or hospitals.*

Electronic health records

Electronic Health Records (EHR) function as single, complete and integrated electronic versions of traditional paper health records, typically held by hospitals and other care services within the UK (Leightley et al., 2018; Morgan & Jablensky, 2010). These registers have been positioned as a possible 'new generation' for mental health research (Morgan & Jablensky, 2010) and their use, particularly within the NHS, has increased in recent years (Allebeck, 2009). The methodological advantages of Case Registers – including their ability to follow up, in detail, large groups of individuals over time – make them a useful research asset, providing large numbers of participants and measurement points (Stewart, 2014). EHRs in mental health care provide extremely rich material because, in theory, they contain every piece of information recorded in a clinical service about a person's presentation, symptoms and relevant background history (often as free-text), as well as the interventions they have received and their response to treatment (Stewart et al., 2009). Exploration of data within such Case Registers can reveal patterns in health care provisions, patient profiles and their clinical presentation to secondary health care services (Gray et al., 2010; Leightley et al., 2018; Payne et al., 2013; Perera et al., 2016).

With any type of system that requires human entry, EHR systems have some shortcomings which include missing data (Collen, 1990); non-standardised and free-text notes (Wu et al., 2012); and a lack of information regarding undiagnosed mentally ill individuals within the community (Fernandes et al., 2013). However, despite these, and other potential problems with Case Registers, they are hugely advantageous for investigating vulnerable subgroups within the wider population (Perera et al., 2016).

Specifically, for this project, an EHR Case Register provided us with the opportunity to examine treatment pathways, diagnoses and profile characteristics for veterans and matched civilian counterparts accessing mental health care services within the UK on a large scale. However, to be able to perform this comparison we need to know who has served in the UK Armed Forces. This is not possible at present as there is no 'military service marker' in the Case Registers of England or Wales, nor is there a requirement for clinicians to record it in any other way (i.e. free text clinical notes). This makes it difficult to evaluate the unique needs of those who have served in the UK military across Wales and England (Morgan & Jablensky, 2010). In contrast, Scottish Case Registers record veteran status using a 'military service marker' on each individual patient record. This enables the analyses of the veteran profile in Scotland (Bergman et al., 2016). Finally, in Northern Ireland it is more complex, with healthcare managed at a local level and on-going security concerns mean that veteran status cannot be recorded safely.

Natural Language Processing

Natural Language Processing (NLP) provides us with a solution to overcome the challenge of identifying who is a veteran in English and Welsh Case Registers as it has already been used in areas such as retrieval, analysis, transformation and classification of text to great effect (Fernandes et al., 2018; Gundlapalli et al., 2013). Of great importance is its utility in being applied automatically to EHR and free-text clinical notes (Cambria & White, 2014; Fernandes et al., 2018). NLP sub-themes, such as text mining, are represented as a set of programmatic rules or machine learning algorithms (i.e. automated learning from gold standard labelled data) to extract meaning from 'naturally-occurring' text (meaning human generated text) (Fernandes et al., 2018; Leightley et al., 2019). The result is often an output that can be interpreted by humans (Chowdhury, 2005; Manning & Schutze, 1999). For example, suicidal ideation is often not recorded in a structured field; however, by applying text mining it is possible to identify and extract the patients whose free-text clinical notes refer to suicidal ideation in order to undertake further surveillance and analysis (Fernandes et al., 2018). Specifically, for this project, we utilised NLP methodology to develop a tool to identify veterans from free-text clinical notes found within the SLaM NHS Foundation Trust Biomedical Research Centre (BRC) Case Register.

Research objectives

There were four main research objectives for this study[3]:

1. To assess the utility and feasibility of identifying veterans using manual and automated approaches of those who accessed secondary mental health care services using an EHR-based Case Register;
2. To validate whether veterans/civilians identified by MSIT were correctly identified using self-reported surveys;
3. To explore the sociodemographic, diagnostic, clinical and treatment characteristics of gender and age matched veterans and civilians who engaged in secondary mental health care services through the SLaM NHS Foundation Trust;
4. To explore the sociodemographic and diagnostic characteristics of veterans engaged in secondary mental health care services through the SLaM NHS Foundation Trust compared to veterans at Combat Stress.

[3]This report is the culmination of two research projects. The first project (known as SLaM 1) sought to determine the feasibility of identifying and extracting veterans from the same SLaM NHS Foundation Trust. Based on the findings of SLaM 1, the second project (known as SLaM 2) aimed to develop an NLP tool to automatically identify veterans from SLaM NHS Foundation Trust, together with additional veteran profile variables to study, in a considerably larger sample size, ensuring greater robustness of results than the smaller SLAM 1 study. An additional component to SLaM 2 allowed the research team to perform the validation of MSIT using self-reported data on veteran/civilian.

Chapter Two:
Methods

Study materials – South London and Maudsley (SLaM) Biomedical Research (BRC) Centre Case Register

The SLaM BRC Case Register (hereafter Case Register) was set up in 2006 as a novel data resource, derived directly from the routinely collected EHRs of the SLaM NHS Foundation Trust (hereafter SLaM) network (Perera et al., 2016). SLaM is one of Europe's largest mental health providers, serving over 1.2 million residents in four South London boroughs of Croydon, Lambeth, Lewisham and Southwark (Collen, 1990). Specifically, the database holds records for secondary mental health care provisions within SLaM, which include all specialist care (i.e. apart from that provided by general practitioners which is considered primary care) for hospitalisations, outpatient care, community care, psychiatric liaison services to general hospitals and forensic mental health services. It is linked to the SLaM Patient Journal System, which is a bespoke EHR, used across all Trust services within the SLaM network. The Case Register includes patients' demographic details, severity of mental health symptoms, mental health diagnoses, psychometric test scores, medications prescribed and clinical events records (referrals, admissions and discharges). It currently holds over 400,000 cases and sees approximately 53,000 new patient referrals each year. Patient records are updated every 24 hours (Perera et al., 2016).

The process for accessing the Case Register can be found in Appendix 1: Accessing the Case Register (p.52).

Study population

This report consists of civilians and veterans who accessed secondary mental health care services through SLaM. Veterans were identified using a systematic manual approach (see Chapter 3: Utility and feasibility of manual identification of veterans, p. 25) and an automated NLP and machine learning approach (see Section: Chapter 4: Utility and feasibility of the Military Service Identification Tool, p. 28). The study population for the Combat Stress comparison is discussed in Chapter 7: Combat Stress comparison (p. 41). The following criteria were applied for inclusion and exclusion for this study:

Inclusion criteria:

- Civilians and veterans who had accessed SLaM mental health care services within an eleven-year window – between 1st January 2007 and 31st December 2018. The Case Register was implemented in 2007, therefore this was the earliest date that digital records could be accessed; and
- Veterans who had served in the UK Armed Forces – we retained records for those whose country of birth was noted as the UK or was left as blank (as we noticed that this field was often left blank if the individual was a UK national).

Exclusion criteria:

- Veterans and civilians aged under 18 years of age, as our focus was on adults; and
- Veterans aged over 64 years (or born before 1943).[4]

[4]This ensured that those who carried out National Service were not included, because this subgroup does not reflect those who voluntarily enter the Armed Forces and biases the sample towards older veterans. National Service was abolished in May 1963. The MSIT was able to identify veterans from the National Service Era, although these were excluded from the analysis in this report.

The validation study, asking patients for their self-reported veteran/civilian status, originally drew from the same sample of veterans and civilians. In this case, our inclusion and exclusion criteria included:

1. Listed as 'Alive' in the SLaM BRC Clinical Record Interactive Search database;
2. Were aged 18 years and old;
3. Had given Consent for Contact (C4C): This mechanism was introduced in January 2017 for patients to join a register giving their consent for researchers to look at their Trust medical records and to contact them for recruitment into further studies.
4. Participant has a valid email address or mobile telephone number;
5. Participants have capacity to consent and deemed fit to participate in research by a clinician. This was corroborated by contacting the clinician or care co-ordinators of a patient, giving a period of 10 working days for a response.
6. Participants will be able to speak English: Participants requiring a translator were therefore excluded from the study. If participants were not fluent in spoken English, there would be substantial barriers to them understanding the participant information and consent form and therefore providing informed consent.

When applying these eligibility criteria, we had very small sample to recruit from. Determining how many of the original cohort had C4C was not possible at the outset. The validation study therefore underwent a series of amendments to obtain the appropriate permissions to expand our sampling pool.

Once obtaining these permissions, we ran MSIT ~500,000 records within the SLaM BRC Case Register who had contact with these services from January 2017 to April 2022. January 2017 represented the inception of the C4C mechanism, thus maximising our selection of patients who would be allowed to be contacted.

After applying the inclusion and exclusion criteria to the new batch, 902 participants (n=723 civilians, n=179 veterans) were eligible to be contacted to take part in the survey. A total of n=149 unique respondents participated in the survey. Three of these responses were ineligible because of duplicated responses that contained contradictory information. Overall, data from 146 participants was used to determine the validation of MSIT. Of these, 112 were categorised by MSIT civilians and 34 were categorised as veterans. The total response rate was 15.5%.

Overall, the sample used in the validation study was sufficient based on our target of n=100-150 - an adequate size for assessing the validity of the MSIT tool. Civilians were oversampled as it was deemed a priority that civilians were not erroneously classified as veterans. This would ensure that any veteran samples determined by MSIT in the future are true veterans, based on self-disclosure.

Whilst we maximised the opportunities to recruit veterans who were able to participate, this did not yield a large veteran sample. The number of veterans compared to civilians will inevitably be smaller as civilians outnumber veterans in the general population. It is possible that the complexity of mental health issues in the veteran group accessing secondary mental health services may mean that they had been ineligible to participate (e.g. due to psychosis, being a current inpatient or refusal from the named care-coordinator or clinician to contact the patient). We also note the lower response rate of veterans, potentially indicating additional barriers to participating in research, such as the aforementioned complexity of their mental health problems.

Study procedure

There were five steps for this study's research procedure – these are described below and in Figure 1.

Identifying veterans using manual and automatic approaches

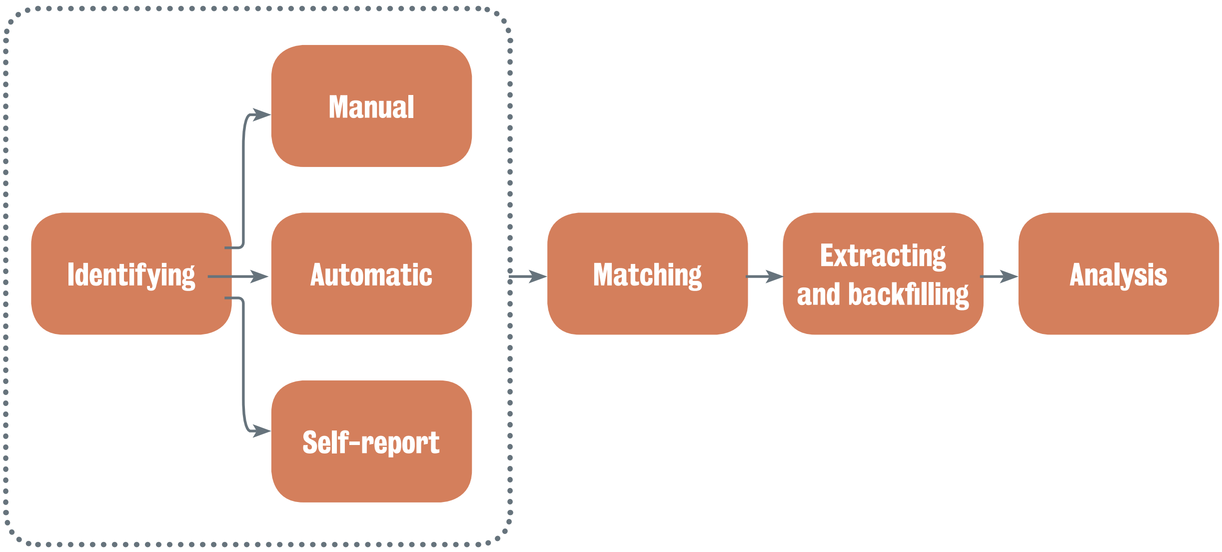
As there is no structured field for identifying veterans within the Case Register, two complementary methods to identify veterans from free-text clinical notes were developed. These approaches were developed using input and guidance from experts in database management and linguistics. The first approach, a systematic manual search approach, is described in detail in Chapter 3: Utility and feasibility of manual identification of veterans (p. 25). The second, an automatic NLP and machine learning approach, is described in detail in Chapter 4: Utility and feasibility of the Military Service Identification Tool (p. 28). Each approach resulted in a list of possible civilians and veterans, with each patient being manually checked by the research team to verify their status.

The research team was careful regarding who to classify as a veteran – each clinical note was read through at least twice and only confirmed as a veteran when an explicit statement about whether the patient had served in the Armed Forces was given. If in doubt, following a conservative approach, the patient was categorised as a civilian.

The process used relied on clinicians correctly recording their status and/or patients' self-reporting (self-disclosure). It is not possible to identify the source of veteran status; for example, if they were directly asked if they had served, or if they volunteered this information freely from electronic healthcare records.

We were, however, able to determine whether categorisation, and therefore the veracity of healthcare records themselves, were accurate via the validation study (full details can be found in Chapter 5: Validating the Military Service Identification Tool, p. 34). In any case, the accuracy of self-report may be compromised among individuals suffering from severe and complex mental health problems which could a) result in delusional thoughts around military service, or b) affect their ability to

Figure 1: The five steps for the study's research procedure



accurately recollect their service/ falsely claiming they have served. Even so, it is likely the number of such falsehoods or inaccuracies would be small (Perera et al., 2016; Stewart et al., 2009) and we were able to partly mitigate against this as patients experiencing psychosis were screened out of the study.

Unfortunately, there was no way to confirm whether some individuals classified as civilians had served in the Armed Forces. In this respect, some veterans may not have volunteered the fact that they had served in the military to their health care provider and thus there would be no mention of their veteran status in their clinical notes. MSIT would subsequently have classified them as civilians. It is worth noting, however, that it is protocol for clinicians to talk through a patient's history/background, including their previous occupations, when they first enter mental health care services.

Matching

Once the veterans who accessed secondary mental health care services through SLaM had been identified, the research team matched the veterans to civilian counterparts by age and gender. Matching on age and gender meant that the overall sample was representative of both civilians and veterans who accessed SLaM secondary mental health care services.

This study is the first in the UK that explores veterans and matched civilians who have sought help from a secondary mental health care Trust in England.

Extracting and backfilling

Once veterans and matched civilians were identified from the Case Register, a set of variables were identified from their medical records. Variables were exported from the Case Register and imported into a Microsoft Excel file (and later into a statistical programme for analysis). Structured fields relating to the veterans' sociodemographic, diagnostic, clinical and treatment characteristics were extracted and transferred to a bespoke study database. In total, 20 specific variables were extracted from the Case Register (Table 1).

Table 1: Variables defined for the study and extracted from the Case Register

1	Age (in years)[5]
2	Gender[6]
3	Ethnicity
4	Marital status
5	Living arrangements
6	Employment status
7	Benefit status
8	Service branch
9	Time in mental health treatment (days) from first diagnosis to last diagnosis
10	Types of mental disorder diagnoses
11	Number of unique mental disorder diagnoses
12	Age(s) at first mental disorder diagnoses
13	Number of outpatient secondary mental health care appointments booked
14	Number of outpatient secondary mental health care appointments attended
15	Number of inpatient secondary mental health care stays
16	Duration of inpatient secondary mental health care stays (in days)
17	Medication received during treatment
18	Number of times talking therapy received based on IAPT appointments
19	Number of times the patient has been sectioned under the Mental Health Act (1983)
20	Deprivation status

[5] *Age was already extracted at step 2 (matching).*
[6] *Gender was already extracted at step 2 (matching).*

The data available for extraction within EHR Case Registers depends upon clinicians to enter information gathered from patients who have accessed care (Leightley et al., 2018). Often, such Case Registers contain large amounts of missing data. To address this, the research team sought to systematically fill variables based on data stored elsewhere where possible (such as correspondence between the hospital and general practitioner). Backfilling of data was only undertaken where there was confidence that the information provided was accurate and did not conflict with other data sources.

Despite the team's attempts to backfill the data, some variables had a high degree of missingness and could not be included in the analysis. These included employment status (variable 6), benefit status (variable 7), service branch (variable 8) and all variables relating to treatment (variables 9-18) and deprivation status (variable 20). Information relating to employment or benefit status may not always be relevant to the clinical consultation and thus will not be recorded. Points of contact with the service may also be variably completed as not all fields will be mandatory.

Protocol of the validation study

To obtain our sample, MSIT was run over subgroups of patients within the SLaM BRC Case Register who had contact with these services from January 2017 to April 2022. This date represented the implementation of the C4C mechanism in CRIS and was therefore the earliest possible date from which we could sample participants from. A BRC administrator applied our eligibility criteria and set up access to patient records for us to obtain their contact details and screen further for their eligibility (including up to date C4C). Where necessary, we contacted participants' named clinicians or care coordinators to determine their capacity to consent or appropriateness to take part in the research. Participants were then contacted via email and/or text messages with a link to the online survey. This method was deemed preferable to postal addresses considering the survey was online.

The survey was hosted on Qualtrics and included an information sheet about the research and consent statements. Participants proceeding to the rest of the survey was considered evidence of their informed consent. Each survey was linked to a unique identifier linked to, but stored separately from, their BRC IDs in order for us to check responses against the MSIT classifications.

The survey itself employed an online questionnaire, which asked a set of (up to 5) questions on participants' previous military service. The first question asked participants whether they had served in the Armed Forces. If yes, additional questions on their military characteristics followed (see Appendix 4, p. 56). The survey ran from February to June 2022. The additional military specific questions were implemented to help elicit further descriptive information to help us refine the MSIT in the event it has made an incorrect prediction and to improve its accuracy.

Participants who took part in the study were invited to voluntarily supply an email address at the end of the questionnaire to enter a prize draw consisting of 1 x £50 Amazon gift voucher, 5 x £25 Amazon gift voucher and 20 x £10 Amazon gift voucher.

Ethics

Ethical approval was given by the East of Scotland Research Ethics Service (EoSRES) within the NHS Research Ethics Service (Ref: 20/ES/0060). Approvals were obtained by the SLaM Research and Development Office at King's College London (Ref: R&D2020/029).

Participants were informed they could withdraw from the study at any point up until data analysis (30th June 2022) by notifying the study team. Details were available on the opening page of the online survey. Where participants asked not to be contacted for research, we updated their C4C status on the CRIS system. If a participant withdrew from the study, no further contact was made. Further details on the ethical process are included in Appendix 4, p. 56.

Data handling

Identifiable data, including patient contact details, was stored on a secure network used by SLaM, only obtainable via CRIS approvals. Survey responses were kept securely on Qualtrics during data collection. Upon completion of data collection, only pseudonymised data was downloaded and stored on the secure server for analysis. The pseudonymised dataset will be kept on secure servers at KCMHR for up to five years and will include no identifiable data. Contact details were kept separately to participants' veteran/civilian status as assigned by MSIT and survey responses.

Analysis

The analyses for this report were split into four sections in line with our research objectives:

1. Utility and feasibility of identifying veterans who accessed secondary mental health care services using manual (see Chapter 3: Utility and feasibility of manual identification of veterans, p. 25) and automated approaches (see Chapter 4: Utility and feasibility of the Military Service Identification Tool, p.28)

   The utility and feasibility of extracting data on veterans presenting to secondary mental health care services was assessed by examining the practicality of accessing the required veteran data, of identifying veterans using manual and automated approaches, of matching our sample to a civilian cohort, and of extracting, cleaning and analysing veteran data. Practical issues the research team encountered included firewall constraints surrounding the Case Register; time and manpower requirements when detecting veteran records; limitations in matching the samples; and amounts of missing data present.

2. Validating MSIT by comparing MSIT classifications to patients' self-reported veteran/ civilian status obtained by an online survey (see Chapter 5: Validating the Military Service Identification Tool, p. 34).

   Once we had surveyed a subsample of participants from the Case Register, we compared their self-reported veteran/ civilian status with their MSIT classification. This was achieved by creating a 2x2 table to cross-tabulate civilian and veteran status. Sensitivity and specificity analyses were further performed to determine the ability of MSIT to correctly identify patients' veteran/civilian status. In the present analysis, sensitivity related to the true classification of a civilian (a positive case) and specificity was the true classification of a veteran (a negative case). Civilians were treated as positive (rather than negative) cases as it was deemed most important that the MSIT would not falsely classify civilians as veterans. This would safeguard a true veteran sample for future analyses. Veterans may be classed as civilians if this was not raised in the consultation. Any inaccurate cases were examined manually to determine reasons for misclassification, with results informing further development of the precision of the tool.

3. Sociodemographic, diagnostic and treatment characteristics of civilians and veterans who accessed secondary mental health care services (see Chapter 6: Veteran and civilian comparison, p. 37)

   Quantitative analyses for this part of the report were carried out using the statistical software STATA (StataCorp, 2015). To address this research objective, a variety of statistics were calculated, such as frequency rates and chi-square tests, to determine whether statistically significant differences existed in sociodemographic profiles and types of mental health diagnoses between civilians and veterans who accessed secondary mental health care services. The data obtained to examine treatment characteristics had a high degree of missingness, thus we were not able to compare the service utilisation of veterans and civilians.

4. Sociodemographic, diagnostic and treatment characteristics of veterans who accessed secondary mental health care services through either SLaM and Combat Stress (see Chapter 7: Combat Stress comparison, p. 41).

Quantitative analyses for this part of the report were carried out using the statistical software STATA (StataCorp, 2015). To address this research objective, a variety of statistics were calculated, such as frequency rates and chi-square tests, to determine whether statistically significant differences existed in sociodemographic profiles and types of mental health diagnoses between veterans who received care through SLaM and Combat Stress. We were unable to investigate differences in the treatment characteristics of SLaM and Combat Stress veterans due to a large amount of missing data in the SLaM dataset on these variables.

SUMMARY

- The Case Register was used to identify veterans who had accessed secondary mental health care services within the South London and Maudsley NHS Foundation Trust
- The Case Register provided data on two population groups:
  1. Veterans who accessed secondary mental health care services through the Trust
  2. Civilians who accessed secondary mental health care services through the Trust
- The study procedure included:
  1. Identifying veterans from the Case Register using manual and automated approaches
  2. Matching veterans from the Case Register to civilians on age and gender
  3. Extracting and backfilling data on sociodemographic, diagnostic, clinical and treatment characteristics for veterans and civilians
  4. Analysing the utility and feasibility of identifying veterans, and exploring similarities and differences between veterans and civilians
  5. Validating whether veterans/civilians identified by MSIT were correctly identified using self-reported surveys.
- Sociodemographic and diagnostic characteristics of SLaM veterans were compared with a) SLaM civilians and b) Combat Stress veterans.
- Due to the large amount of missing data on SLaM treatment variables, a comparison of service utilisation was not possible, with exception to use of the Mental Health Act (MHA) 1983.

Chapter Three:

Utility and feasibility of manual identification of veterans

This chapter presents the utility and feasibility of manually searching the Case Register using Structured Query Language, human verification and experts in database management. This chapter is divided into three sections: 1) we describe the search strategy; 2) we explain how the Structured Query Language search strategy was used and 3) we discuss the utility and feasibility of this approach.

Search strategy

The identification process was iterative in nature and took place in two stages. Firstly, keywords were derived during an initial search[7] of a random selection of patient records, by noting down frequently used military-related terms that appeared in veterans' records. Secondly, the research team compiled their own list of military terms, using their expertise in the field. Together, these two methodologies resulted in a large list of military-related words, of which 19 words proved to be useful, including 'Royal Navy', 'Army', 'Royal Air Force' and 'armed forces' (see Appendix 2: Inclusion and exclusion for the full list, p. 53).

Using the same processes as described above, a list of over 30 exclusion terms were developed to prevent a substantial number of civilians being identified as veterans (often called false positives); these terms included 'navy blue' (see Appendix 2: Inclusion and exclusion for the full list, p. 53). The terms were then combined into a Structured Query Language search strategy, to systematically search the Case Register.

Structured Query Language search strategy

The Structured Query Language strategy, using the derived search terms described above, inspected each patients free-text medical records, analysing each sentence. If the sentence included a military term, the research team was notified. All clinical records that were identified as being that of a veteran were scrutinised individually by the research team to ensure that the patient had served in the military. Each patient's medical records were read through at least twice and confirmed as a veteran when an explicit statement about the patient serving status was reported.

Utility and feasibility

Despite there being no military service marker within the Case Register, we found it was feasible to identify veterans. When considering the individual word searches used to identify veterans, the term 'Army' returned the highest number of potential veteran records, followed by 'Royal Navy'

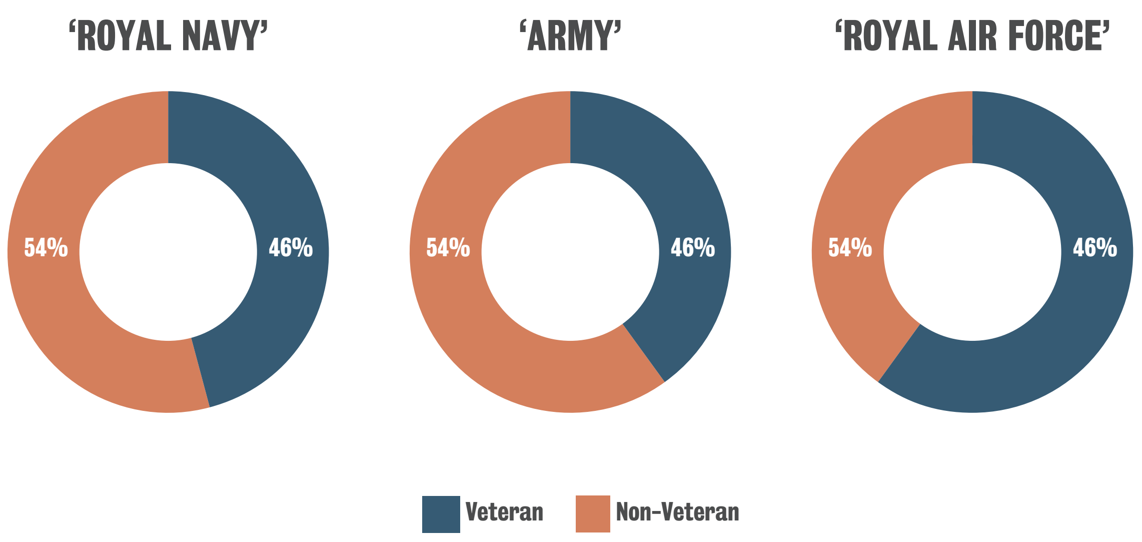
[7] *We used specific search terms such as "served in the Army", "joined the Navy" to identify electronic medical records for further analysis.*

and then ‘Royal Air Force’. However, numbers returned are not reflective of the hit rate of each term (see Figure 2).

A random batch of 457 out of 6039[8] of these potential veteran records were tested, to determine the individual identification rates. These were calculated as the percentage of true veterans identified versus the percentage of non-veterans identified, from the overall number of potential veteran records returned. As shown in Figure 3, ‘Royal Air Force’ had the highest sensitivity rate out of the key words retained that correctly identified a veteran, followed by ‘Royal Navy’ and then ‘Army’. ‘Royal Air Force’ arguably returned the cleanest veteran records when the Case Register was searched – that is, the research team had to implement the fewest exclusion criteria for records returned using this term (see Appendix 2: Inclusion and exclusion for the full list, p. 53).

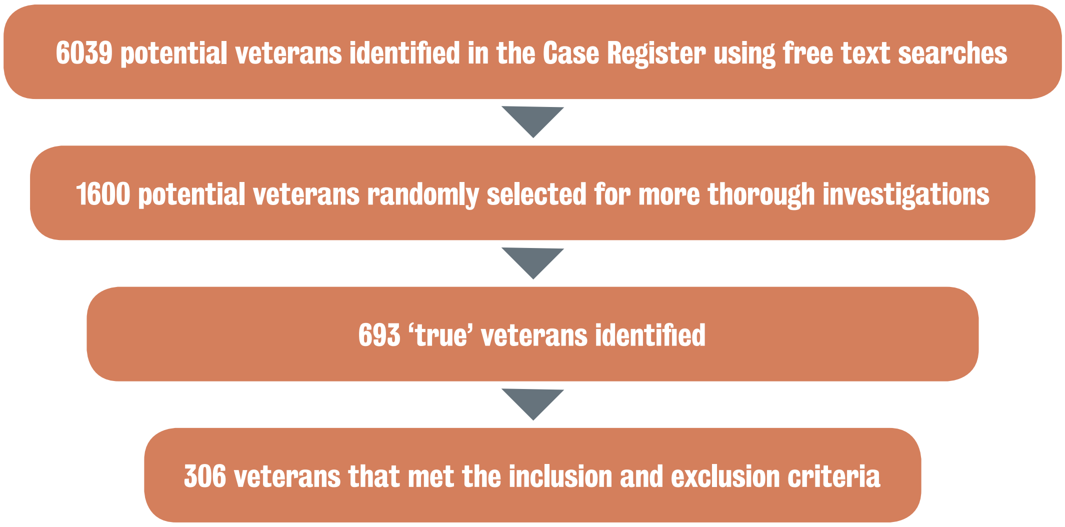
The Structured Query Language search strategy of the Case Register (which is part of the SLaM BRC CRIS system) returned 6,039 potential veterans who fitted these criteria. These records were randomly organised then used for selecting a smaller sample of potential veterans for checking. The research team selected 1,600 records to scrutinise in more detail. Of these 1,600 possible veterans, 693 were identified as being true veterans by the research team. Following implementation of our inclusion and exclusion criteria (see Chapter 2: Methods, p. 17), 306 of the identified veterans were included in the final sample. Figure 3 shows a flow chart of the number of SLaM veteran identified using the manual approach described above. Identifying veterans through the Case Register in this way was labour intensive and time consuming. Manually verifying each potential veteran record involved reading through what was often a vast

Figure 2: Hit rates for the three primary military search terms used in the Case Register – ‘Royal Navy’, ‘Army’ and ‘Royal Air Force’



[8] *This number was selected due to resource and manpower limitations of the study.*

Figure 3: Flow chart showing the numbers of veterans identified through the Case Register using the manual approach



number of free-text notes written by the clinician who had seen the patient. The research team found that reading through each patient's notes took, on average, 15 minutes. The research team manually worked through 1,600 potential veteran records for this project, which equated to 400 hours of reading time or approximately 11 weeks' worth of work/effort.

As mentioned above, of the 1,600 possible veteran records checked, 693 were identified as being true veterans by the research team – this represents an identification rate of 43% of possible veterans being identified as a true veteran. While this percentage reflects a first step in creating a methodology to identify veterans from large clinical databases, it is low compared to automated identification processes that make use of NLP (Chowdhury, 2005).

SUMMARY

- 19 military-related categories and phrases were identified for use in the manual approach.
- 30 terms and phrases that may cause misclassification were identified.
- A Structured Query Language search strategy was developed by experts in database management.
- 6,039 potential veterans were identified using the search strategy.
- Due to time considerations, only 1,600 were randomly selected for manual evaluation.
- 693 were identified as being a true veteran; that is, a clear statement that they served in the Armed Forces was present in their medical record, with 306 veterans meeting the study inclusion criteria and used for analysis.
- The utility and feasibility of a Structured Query Language search strategy combined with manual evaluation was impractical and time consuming.

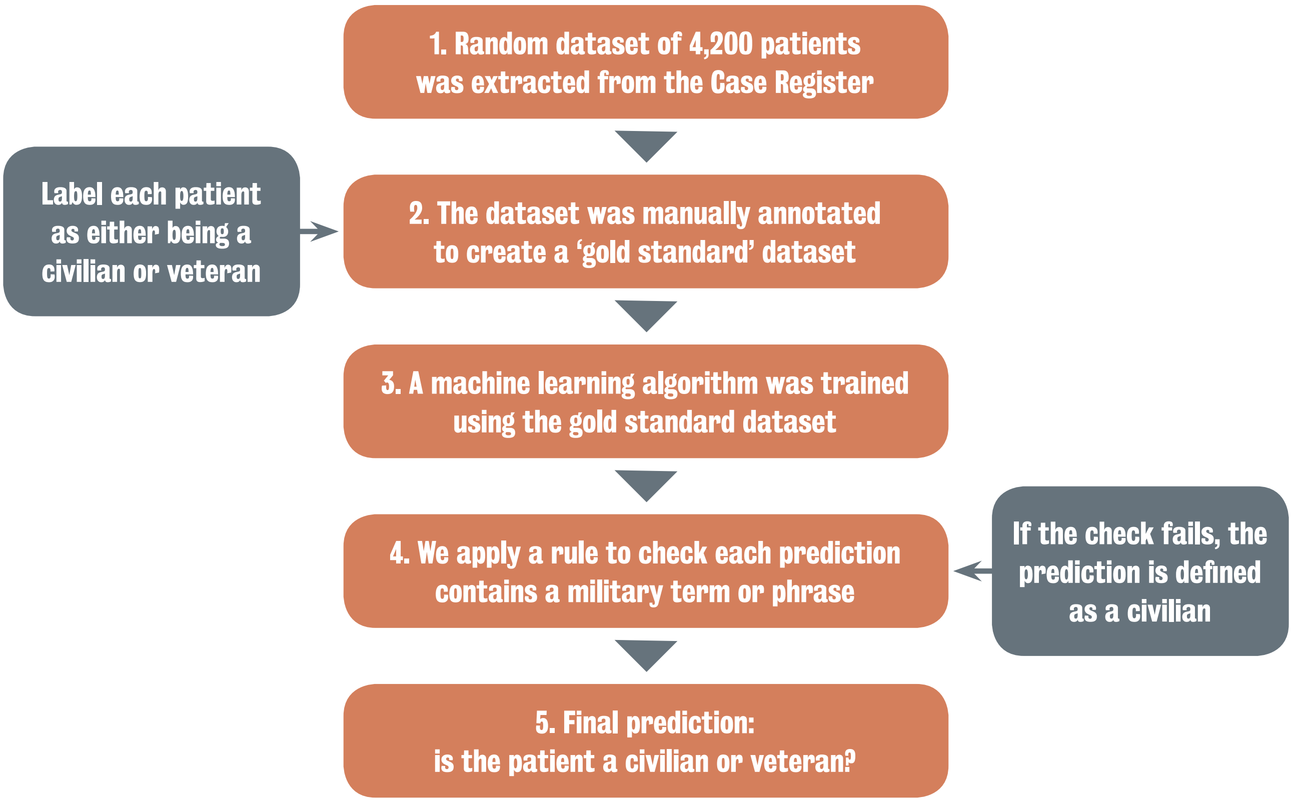
Chapter Four:

Utility and feasibility of the Military Service Identification Tool (MSIT)

This chapter presents the utility and feasibility of the Military Service Identification Tool (hereafter denoted as MSIT) in automatically searching the Case Register to identify and extract veterans.

This chapter is divided into three sections: 1) we describe the MSIT; 2) we discuss utility and feasibility of the MSIT; and 3) we present information on how to access the MSIT.

Figure 4: Military Service Identification Tool flow diagram



Development of the Military Service Identification Tool

The MSIT was created in Python using the Natural Language Processing Toolkit and Scikit-learn – all of which are open source and accessible for free (see Appendix 3: Technical summary of the Military Service Identification Tool, p. 55, for a detailed description). MSIT was developed using the following stages (see Figure 4 above for an overview):

1. Stage 1 –4,200[9] patient medical records were identified and extracted from the Case Register, which forms the basis of the gold standard dataset;
2. Stage 2 – The medical records within the gold standard dataset were manually inspected, annotated and coded by the research team to identify if the record of the patient belonged to a civilian or veteran. Any disagreements on the coding were discussed and resolved within the research team[10];
3. Stage 3 – A machine learning algorithm was 'trained' on a sub-set of the gold standard dataset, with the remainder retained for testing[11];
4. Stage 4 – For each prediction made by the machine learning algorithm a 'rule check' was automatically performed to validate the prediction. The objective was to ensure that the medical record being analysed had at least one mention of a known military term (see Appendix 2: Inclusion and exclusion terms, p. 53); and
5. Stage 5 – Based on the outcome of stage (3) and (4), a prediction is made whether the medical record belongs to a civilian or veteran.

Utility and feasibility

Using free-text clinical notes, the MSIT was able to identify veterans with high precision, accuracy and recall[12] when compared to the manual identification of veterans. To achieve this, 6,672 free-text clinical notes which are part of the patient medical record were manually annotated to identify common military words and phrases. It is important to acknowledge that manual annotation of the gold standard dataset is time consuming, however this only happened once to create the dataset, and did not need to be repeated.

[9] *This comprised 3,300 civilian and 869 veteran patients from the Case Register. In total 6,672 free-text clinical notes from the patient medical records were extracted, annotated and were used for the MSIT model development. Further details of the MSIT development at a document level can be found in Appendix 3: Technical summary of the Military Service Identification Tool, p. 55.*

[10] *The default assumption was that the patient was a civilian.*

[11] *It is common practice to 'hold back' a proportion of the dataset to allow it to be used for testing and evaluation. For example, if we include clinical notes of a patient in both the training and testing datasets the algorithms may be more likely to predict the correct label (i.e. veteran). In this study, 66% of the gold standard dataset was used for training, and the reminder 34% used for testing.*

[12] *We compared the results of the MSIT to manual annotations (the gold standard) allowing for computation of precision (positive predictive value), recall (sensitivity) and accuracy (proportion of true results). Precision was defined as the proportion of correctly identified true veterans over the total number of true veterans identified by the classifier. Recall was defined as the proportion of true veterans identified by the classifier over the total number of actual veterans (identified by manual annotation). Accuracy was defined as the proportion of true results - true positive and true negative veterans – over the total number of actual veterans (identified by manual annotation).*

Table 2: Top five frequently occurring military words and phrases identified during manual [human] annotation of the gold standard dataset

Military words (n=2,611)		Military phrases (n=2,016)	
Word	**Frequency n (%)**	**Phrase**	**Frequency n (%)**
Army	553 (21.20)	Joined the army	167 (8.33)
National Service	445 (17.08)	Left the army	122 (6.07)
RAF	225 (8.65)	Demobbed from the army	101 (5.01)
Navy	166 (6.36)	National Service in the army	65 (3.24)
Royal Navy	124 (4.76)	Two years in the Army	64 (3.19)

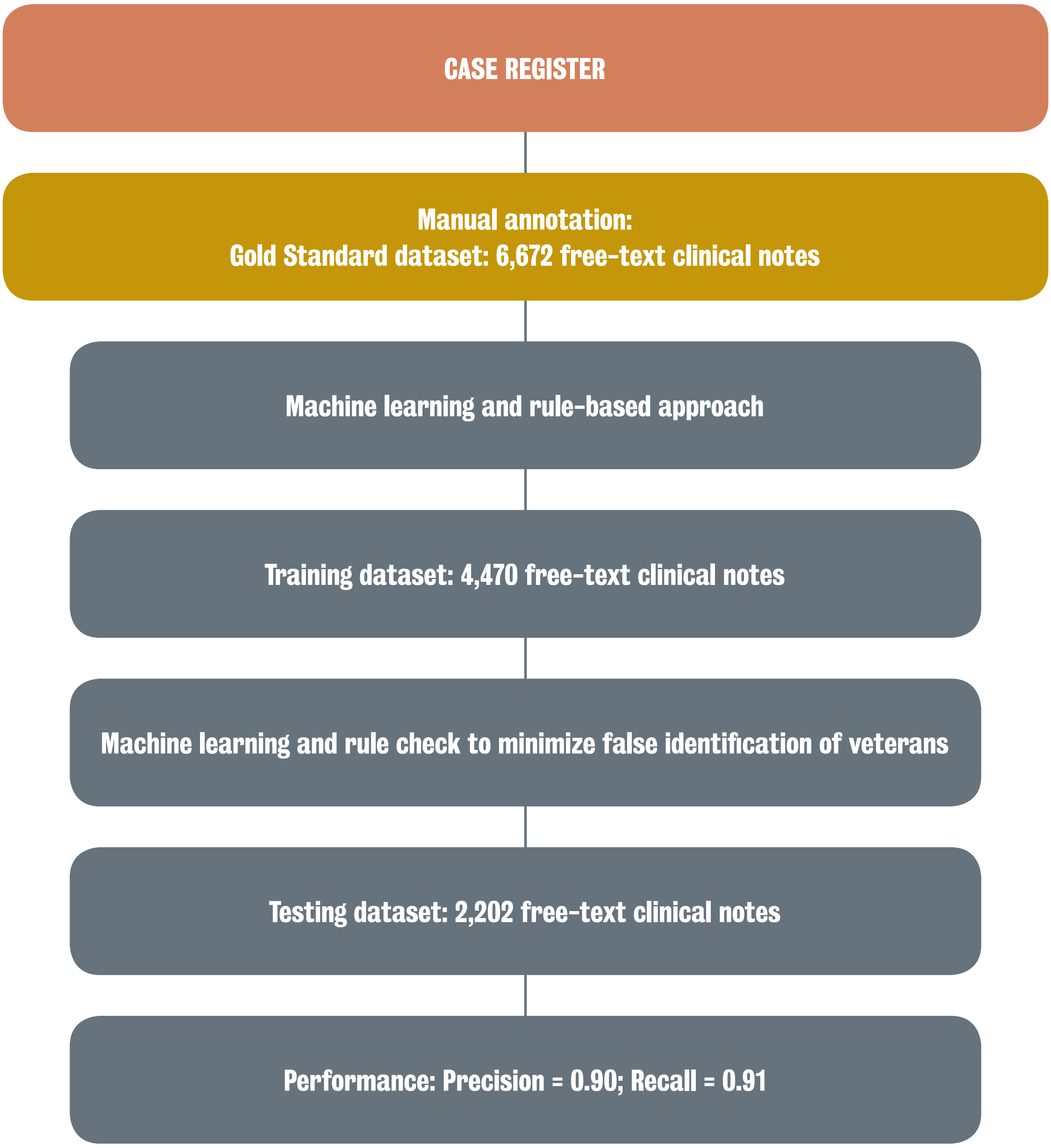
Within the clinical notes, the most common words used related to service branch: 'Army', 'National Service', 'RAF', 'Navy' and 'Royal Navy', whereas the most common military phrases related to either joining the Armed Forces or leaving: 'joined the army', 'left the army', 'demobbed from the army', 'national service in the army' and 'two years in the army' (see Table 2).

A random batch of 1,428 patients (a total of 2,202 free-text clinical notes) were extracted. This batch represented 66% of the gold standard dataset. MSIT was evaluated by comparing the results from MSIT against the results from the gold standard dataset. This allowed for calculation of precision (positive predictive value) and recall (sensitivity).

As shown in Figure 5, MSIT had a high precision and recall with an overall accuracy of 97%, demonstrating the suitability of MSIT in identifying veterans from the Case Register (see Appendix 3: Technical summary of the Military Service Identification Tool, p. 55, for detailed evaluation results). However, it is important to acknowledge that the results did indicate that the tool was not perfect, and it does struggle to identify, or mis-identify, a very small proportion of civilians as veterans in the Case Register. This is due to difficulty in distinguishing if the record is describing the patient or another individual (i.e. father served in the military and not the patient).

To further validate the MSIT tool's performance, we surveyed patients directly to ascertain their veteran/civilian status (Chapter 5 presents the results of this part of the study, p. 34).

Figure 5: Flow diagram of the Military Service Identification Tool and its performance when evaluated against the gold standard dataset



Applying the Military Service Identification Tool to the Case Register

The MSIT was applied to a sample of 150,000 patient medical records extracted from the Case Register[13] in a staged approach. The sample was divided into blocks of 25,000 records and passed to the MSIT, it took an average of 3 minutes to make a prediction for each block, with each veteran prediction manually verified for use in analysis[14]. As shown in Figure 6, MSIT and the manual search strategy identified 2,922 veterans, with 1,634 veterans excluded as they were part of the National Service Era cohort (pre-May 1963). The remaining 1,288 veterans were matched to a civilian cohort and used for the analyses presented in this report.

Accessing the Military Service Identification Tool

To support the use of the MSIT[15], the Tool is publicly available under a GNU General Public License (v3)[16]. It is hoped that the research community will further develop, refine and contribute to the tool and apply it to different datasets across the UK.

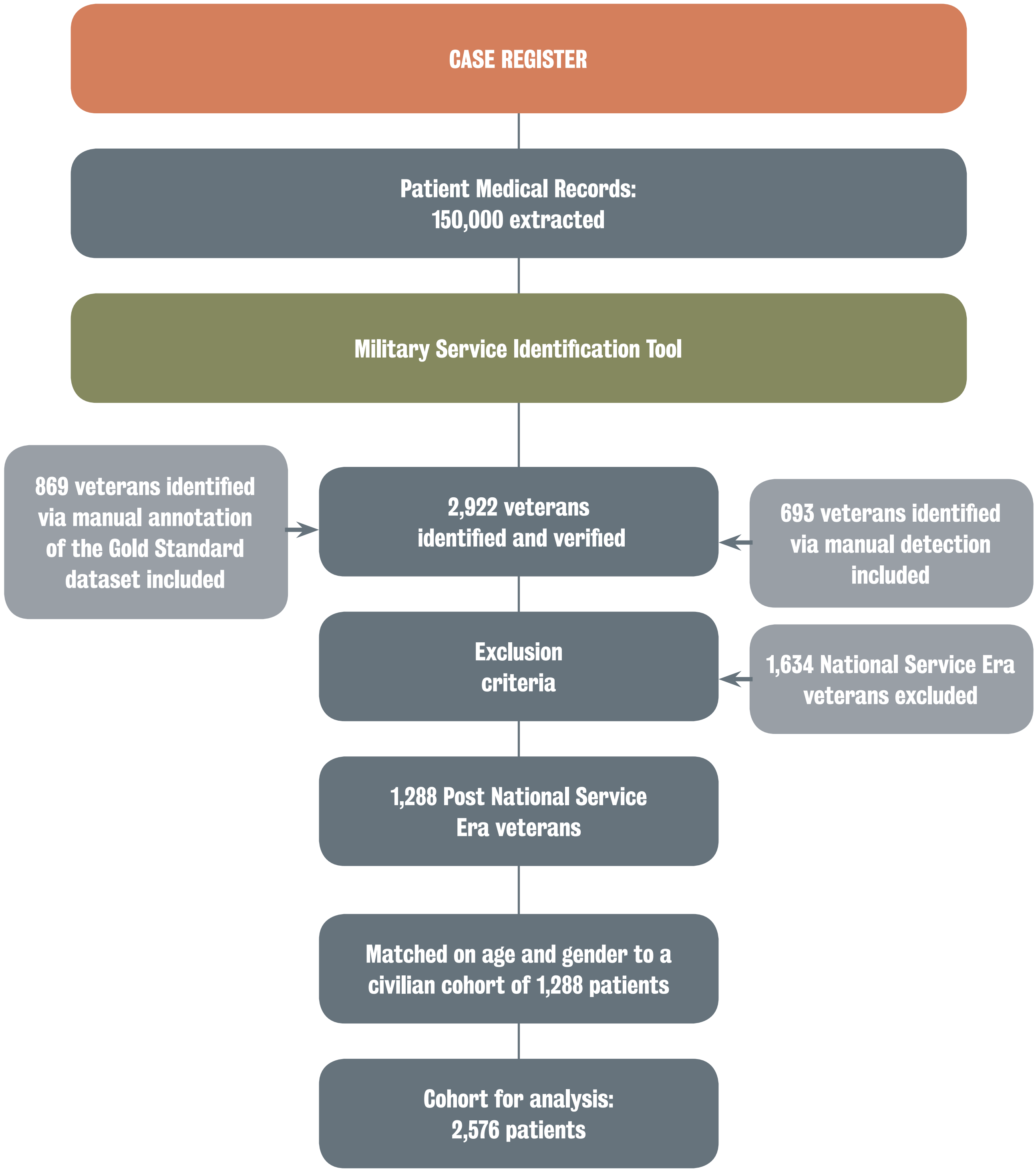
A Jupyter Notebook demonstrating the tool with artificial data can be found **here**.

SUMMARY

- The Military Service Identification Tool was created using data from 4,200 patients extracted from the Case Register.
- The Military Service Identification Tool obtained an accuracy of 97% in identifying civilians and veterans using a manually annotated gold standard dataset.
- The Military Service Identification Tool was applied to 150,000 patient records extracted from the Case Register.
- 2,922 true veterans were identified and verified to ensure they had a clear statement that they had served in the Armed Forces.
- There was evidence that MSIT misidentifies a small proportion of civilians as veterans. These cases appeared to relate to records on patients' family members, for example.
- Overall, the Military Service Identification Tool was found to be feasible in identifying veterans accurately and quickly.
- 1,288 veterans were matched to a civilian dataset on age and gender after applying study exclusion criteria to create a sample to be used for analyses.

[13]Considering firewall constraints surrounding the Case Register, time, and manpower requirements to verify veteran records, we restricted the number of documents analysed by the MSIT for each patient to three.
[14]We manually checked a random sample of records labelled as being a civilian, with misclassification in line with our training and testing results described earlier.
[15]The datasets used and generated during the development of MSIT are based on pseudonymised patient data from the Case Register which is not publicly available. Access to this data requires a formal application to the SLaM BRC Patient Data Oversight Committee of the National Institute of Health Research Maudsley Biomedical Research Centre. On request and after suitable arrangements are put in place, data and models employed in this study can be viewed within the secure firewall.
[16]This license has been selected as it requires those who use the Military Service Identification Tool and make any improvements to make those improvements accessible to the public.

Figure 6: Flow chart to show the numbers of veterans identified through the Case Register via the Military Service Identification Tool



Chapter Five:

Validating the Military Service Identification Tool

Whilst the MSIT has high precision, it is still unclear whether the veterans identified by the MSIT are true veterans. MSIT identifies veterans based on the notes provided by the clinician in the electronic healthcare record, however self-disclosure and recording may be prone to error. To overcome this limitation and to demonstrate whether the MSIT is correctly identifying military service (serving personnel, veterans or civilian), it was deemed important to survey those who were classified as veterans and civilians to ask about potential military service.
A further validation was undertaken to ascertain whether individuals identified by the MSIT had or had not served in the Armed Forces. This was achieved by sending an online survey to a subsample of patients in the SLaM BRC Case Register to determine their self-reported status and compare to their classification according to MSIT. Responses allowed us to validate the MSIT algorithm predictive ability and contribute to more precise categorisations. Incorrect classifications would provide further information with which to refine the algorithm and improve its accuracy. The Methods of this section are outlined in Chapter 2: Methods, p. 17.

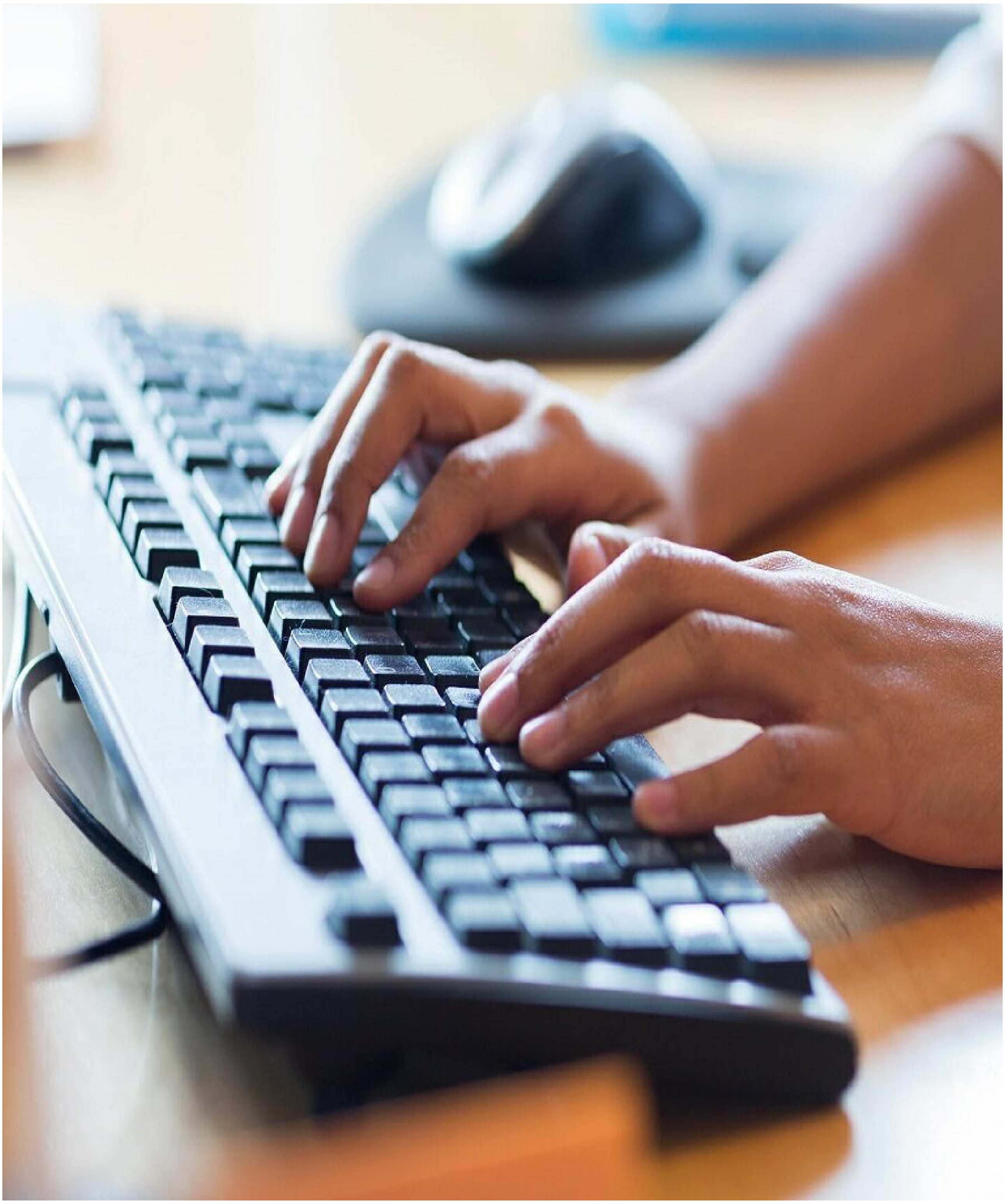
Results

The final sample included 146 participants (n= 112 MSIT civilians; n= 34 MSIT veterans). When corroborating survey responses and MSIT classifications, we found:

- 83.6% of the total sample were accurately categorised by MSIT (n=122/146).

A sensitivity and specificity analysis was further conducted to determine how many veterans and civilians were misclassified. Overall, 23 true civilians were inaccurately categorised by MSIT as veterans, and 1 veteran was inaccurately categorised by MSIT as a civilian. Therefore:

- The sensitivity of the tool, i.e. the ability to correctly classify civilians, was 0.83
- The specificity of the tool, i.e. the ability to correctly classify veterans, was 0.92



Manual investigation of inaccurate classifications

According to this validation study, MSIT classifications had a very high degree of accuracy with some exceptions. In the minor examples of misclassification, MSIT may be more prone to assign civilians with veteran status. It is not possible to determine if this represents a significant difference due to the small sample size of veterans on which this was based.

A manual search was conducted to examine the data MSIT based misclassifications on. The most common reasons for misclassification were the mentioning of military family members and support received from the Salvation Army. This information can be used to apply minor changes to the tool, however no substantial changes to the tool were needed.

In the future, minor revisions to MSIT can be considered to reflect changes in military terminology and works used amongst the general population. However, in its current form, MSIT is functional and further testing in other datasets should be undertaken to further validate the findings.

SUMMARY

- The Military Service Identification Tool was found to have utility and feasibility in the previous chapter, obtaining an accuracy of 97% in identifying civilians and veterans using a manually annotated gold standard dataset.
- According to the validation study, MSIT was able to correctly identify 83.6% of participants surveyed.
- MSIT’s tendency to attribute veteran status to civilians would be remedied by including further keywords and making minor amendments; overall, no substantial changes to MSIT were necessary.
- MSIT has been released publicly.

Chapter Six:

Veteran and civilian comparison

This chapter describes the sociodemographic profiles and types of mental disorders across two groups – civilians and veterans - who accessed secondary mental health care services through SLaM. This chapter is divided into two sections: 1) sociodemographic and military characteristics; and 2) diagnostic characteristics. Another objective was to examine the treatment characteristics of veterans and civilians. Whilst treatment characteristics could not be explored due to a large amount of missing data on these variables, we were able to analyse usage of the MHA (1983) in the veteran and civilian samples.

Sociodemographic and military characteristics

Across the sample, civilians and veterans were matched on age and gender (see Table 3); the median age was 40.9 years (interquartile range (IQR) = 31.0 – 51.0 years), 88.4% of each sample were male (total n=2,276) and 11.6% female (total n=300) (Table 3). Most were 'single, separated, divorced or widowed' (civilian: 76.0%; veteran: 69.1%). Veterans were significantly more likely to be 'married, or 'in a partnership or relationship' compared to civilians who were significantly more likely to be 'single, separated, divorced or widowed'. Over half of the total sample (53.8%) lived in an area of low deprivation[17] though no differences were found between civilians and veterans.

Most of the sample endorsed having British (white) ethnicity (74.5% of civilians and 75.3% of veterans). However, detailed analysis indicated that veterans were significantly more likely to report 'any other ethnicity', which included mixed or multiple ethnicities, with civilians significantly more likely to report being 'Asian or Asian British'.

A large proportion of civilians and veterans reported living 'alone' (42.9% vs. 43.6% respectively). Veterans were significantly more likely to live with a 'partner and/or children' than civilians and were significantly less likely to live with their 'parents'.

[17]Describes an area which has the potential for health risk from an ecological concentration of poverty, unemployment, economic disinvestment, and social disorganisation.

Table 3: Sociodemographic and military variables across SLaM civilians and veterans identified using the Military Service Identification Tool[18]

	Overall (n=2,576)	Civilian (n=1,288)	Veteran (n=1,288)	p value
Age (years), mean (SD)	40.9 (12.7)	40.8 (12.7)	40.9 (12.7)	-
Gender, n (%)				
Male	2,276 (88.4)	1,138 (88.4)	1,138 (88.4)	-
Female	300 (11.6)	150 (11.6)	150 (11.6)	
Deceased, n (%)				
Yes	320 (12.4)	154 (12.0)	166 (12.9)	0.473
No	2,256 (87.6)	1,134 (88.0)	1,122 (87.1)	
Ethnicity, n(%)				
Asian/Asian British	79 (3.5)	51 (5.0)	28 (2.3)	0.002*
African/Caribbean/Black British	305 (13.8)	144 (14.0)	161 (13.6)	
White British	1,659 (74.9)	764 (74.5)	895 (75.3)	
Any other ethnicity	172 (7.8)	67 (6.5)	105 (8.8)	
Marital status, n (%)				
Married/Relationship	576 (27.7)	227 (24.0)	349 (30.9)	< 0.001*
Single/Separated/Divorced/Widowed	1,503 (72.3)	721 (76.0)	782 (69.1)	
Living arrangements, n (%)				
Alone	708 (43.3)	316 (42.9)	392 (43.6)	0.002*
Parents	120 (7.3)	71 (9.6)	49 (5.4)	
Partner and/or children	522 (31.9)	209 (28.4)	313 (34.9)	
Relatives	47 (2.9)	23 (3.1)	24 (2.7)	
Other[19]	238 (14.6)	118 (16.0)	120 (13.4)	
Deprivation status[20], n (%)				
Most deprived (high)	1,240 (53.8)	566 (51.8)	674 (55.7)	0.172
Middle	833 (36.2)	411 (37.6)	422 (34.8)	
Least deprived (low)	231 (10.0)	116 (10.6)	115 (9.5)	
Service status, n (%)				
Overseas	-	-	60 (4.7)	-
UK			1,228 (95.3)	

*[18]Note. n = number of civilians or veterans; IQR = interquartile range (lower quartile – upper quartile); * = statistically significant p value. For age, the p value refers to difference between means. Missing data was backfilled where possible using free-text clinical notes. Percentages refer to the proportion that populated each specific field and do not include missing data.*
[19]This includes staying with foster parents and friends.
[20]The Index of Multiple Deprivation is the official measure of relative deprivation for small areas (or neighbourhoods) in England. The Index of Multiple Deprivation ranks every small area in England from 1 (most deprived area) to 10 (least deprived area) based on a range of factors.

Diagnostic characteristics

Across the sample (both groups), the median age for receiving a mental health diagnosis within SLaM was 45.0 years (IQR = 35.0 – 55.0 years). For civilians and veterans, the most common mental health diagnoses were 'depressive disorder' (15.5% and 26.2% respectively, see Table 4). Veterans were more likely to have diagnoses of 'anxiety disorder', 'depressive disorder', 'personality disorder', 'psychosis disorder', 'stress disorder' and 'other mental disorders' than civilians, whilst civilians were more likely to experience 'drug disorders' than veterans.

No significant differences were found between 'alcohol use disorder' (13.7% and 14.1% respectively) or 'other mental disorders' (12.6% and 14.8% respectively) between civilians and veterans.

There is a huge amount of public interest in the rates of PTSD among veterans (Otis et al., 2003; Stevelink et al., 2018; van Hoorn et al., 2013). Further analyses of the data revealed that 1.4% of civilians, and 3.6% of veterans were assigned a diagnosis of PTSD (p = 0.048).

We attempted to examine the treatment characteristics of both veteran and civilian samples, however there was a high degree of missingness for number of inpatient/outpatient appointments and duration of inpatient stays, and other such variables. We were, however, able to examine the levels of Mental Health Act (1983) usage in both samples.

Overall, of the 272 patients sectioned[23] under the Mental Health Act (1983), a total of 960 section/re-section notices were issued, with veterans (15.5%) being significantly more likely to be sectioned than civilians (5.6%) (Table 5).

Table 4: Mental health diagnoses of civilians and veterans in SLaM

Diagnosis[21]	Overall (n=2,576)	Civilian (n=1,288)	Veteran (n=1,288)	
	n (%)	n (%)	n (%)	p value
Alcohol use disorder	358 (13.9)	176 (13.7)	182 (14.1)	0.733
Anxiety disorder	186 (7.2)	68 (5.3)	118 (9.2)	< 0.001*
Depressive disorder	536 (20.8)	199 (15.5)	337 (26.2)	< 0.001*
Drug disorder	267 (10.4)	161 (12.5)	106 (8.2)	< 0.001*
Other mental disorders[22]	352 (13.7)	162 (12.6)	190 (14.8)	0.108
Psychosis disorder	285 (11.1)	106 (8.2)	179 (13.9)	< 0.001*
Personality disorder	123 (4.8)	42 (3.4)	81 (6.3)	< 0.001*
Stress disorder	234 (9.1)	57 (4.4)	177 (13.7)	< 0.001*

** denotes a significant difference between civilians and veterans.*

[21]Each diagnosis is a group of International Classification of Diseases (version 10) coding. A full breakdown of the coding can be found in Appendix 5: Diagnosis categories (p. 57).
[22]Note: The 'other mental disorders' category represents civilians and veterans who received a mental diagnosis which fell outside the scope of this report (i.e. unspecified mental disorder, conduct disorders, hyperkinetic disorders).
[22]Being sectioned means that you are kept in hospital under the Mental Health Act 1983. You can be sectioned if your own health or safety are at risk, or to protect other people (Wickersham et al., 2019).

Table 5: The number of patients sectioned, and re-sectioned, under the Mental Health Act (1983)

	Civilian (n=1,288)	Veteran (n=1,288)	
	n (%)	**n (%)**	**p value**
Number of patients being issued with a section notice under the Mental Health Act (1983)	72 (5.6)	200 (15.5)	< 0.001 *
	Median (IQR)	**Median (IQR)**	**p value**
Number of times the Mental Health Act (1983) 'section' or 're-section' has been issued	2 (1-3)	2 (1-4)	0.097

SUMMARY

- A total of 2,576 patients were included in the analysis, comprising of 1,288 civilians and 1,288 veterans.
- Most civilians and veterans endorsed having a British (white) ethnicity.
- Most civilians and veterans were single, separated, divorced or widowed.
- Veterans were significantly more likely to be given a stress, depressive, anxiety, psychosis or personality disorder diagnosis than civilians.
- No significance difference between civilians and veterans for alcohol misuse were observed.
- Veterans in the present sample were more likely to be sectioned under the Mental Health Act (1983) than civilians.

Chapter Seven:
Combat Stress comparison

Unlike within the Case Register where veteran status is not known, individuals within the Combat Stress Case Register are all veterans who have served in the Armed Forces. To understand how the results in this report compare, we undertook a small sub-study which compared veterans identified from the Case Register to Combat Stress.

Combat Stress

The Combat Stress electronic Case Register was set up in 2013, and is a digital version of the charity's paper records system (Weijers & Busuttil, 2015). Combat Stress was established in 1919, after the end of the First World War, to support veterans experiencing mental health difficulties. Today it is the UK's largest military charity in terms of the number of individual's treated, providing both inpatient and outpatient secondary mental health services to veterans and specialising in PTSD. Since 2011, Combat Stress has been funded by the NHS to provide a national specialist PTSD clinical service for ex-serving personnel (Weijers & Busuttil, 2015).

Approximately 2,000 new veteran patients present to the charity's services each year and they treat individuals from across the UK(Murphy et al., 2015). Similarly, to the Case Register, the Combat Stress Case Register holds records for veterans who have accessed secondary mental health care services through the charity. The database includes patients' demographic details, severity of mental health symptoms, mental health diagnoses, scores on mental health questionnaire measures, medications prescribed and clinical events records. While the data held is broadly similar across the Case Register and Combat Stress Case Register, the EHR systems and their associated outputs are completely distinct, in terms of both the search process and the structure used.

The Combat Stress Case Register

We gained access to 1,136 pseudo anonymised veteran records from the Combat Stress Case Register, all of which fitted our inclusion and exclusion criteria (see Chapter 2: Methods, p. 17). After matching took place, 189 were included in the final sample.

BRC Case Register and Combat Stress Case Register Comparison

Sociodemographic and military variables

The SLaM veterans and Combat Stress veterans were matched on age bands and gender (see Table 6). Post matching, the median age of SLaM veterans was 40.0 (interquartile range (IQR) 31.0 – 49.0) and the median age of Combat Stress veterans was 40.8 (IQR = 31.4 – 50.5). Most veterans were male (95.8%).

Many of the SLaM veterans reported living alone (48.9%), whereas most Combat Stress veterans reported living with their 'partner or children' (76.6%). Most veterans endorsed having

Table 6: Sociodemographic and military characteristics of SLaM veterans and Combat Stress veterans[24]

	SLaM Veterans (n=189)	Combat Stress Veterans (n=189)	
	Median (IQR)	Median (IQR)	p value
Age (in years)	40.0 (31.0-49.0)	40.8 (31.4-50.5)	0.88
	n (%)	n (%)	
Gender:			
Male	181 (95.8)	181 (95.8)	1.00
Female	8 (4.2)	8 (4.2)	
Who the veteran lives with:			
Alone	46 (48.9)	20 (18.0)	< 0.001*
Friends/family/other	22 (23.4)	6 (5.4)	
Partner/children	26 (27.7)	85 (76.6)	
Ethnicity:			
British	132 (80.5)	158 (98.1)	< 0.001*
Other	32 (19.5)	3 (1.9)	
Marital status:			
In a relationship	40 (25.8)	118 (62.4)	< 0.001*
Separated/divorced/widowed	31 (20.0)	39 (20.6)	
Single/presently separated	84 (54.2)	32 (16.9)	
Service branch:			
Royal Navy	13 (18.3)	16 (8.5)	0.12
Army	52 (73.2)	154 (81.5)	
Royal Air Force	6 (8.5)	19 (10.1)	

*[24]Note. n = number of veterans; IQR = interquartile range (lower quartile – upper quartile); * = statistically significant p value differences. For age, the p value refers to differences between means. Missing data for service branch was were backfilled for SLaM veterans where possible, using free-text clinical notes. This allowed us to assign a service branch for 71 SLaM veterans who previously had this field missing. Percentages refer to numbers that had each specific field populated and do not include missing data.*

a ‘British ethnicity’ – 80.5% of SLaM veterans and 98.1% of Combat Stress veterans, though this was significantly higher for Combat Stress veterans than for SLaM veterans. 54.2% of SLaM veterans reported being ‘single or other’, whereas 62.4% of Combat Stress veterans reported being ‘in a relationship’. Most veterans reported having served in the ‘Army’ – 73.2% of SLaM veterans and 81.5% of Combat Stress veterans.

The sociodemographic and military variables with the highest levels of missing data were who the veteran lives with (45.8% missing) and service branch (31.2% missing, after backfilling), however, this was only the case for SLaM identified veterans.

Mental health variables

Across the total sample, the most common mental health diagnoses given were ‘stress disorders’ (43.7%), ‘depressive disorders’ (41.8%) and ‘anxiety disorders’ (41.8%; see Table 7). However, the diagnoses given varied when looking at the SLaM veterans and the Combat Stress veterans separately. A greater diversity of diagnoses was given across the SLaM veterans than the Combat Stress veterans.[25]

For SLaM veterans, the most common mental health diagnoses given were ‘other’ mental disorders[26] (15.3%), depressive disorders (14.3%) and alcohol use disorder (11.6%). For Combat Stress veterans, the most common mental health diagnoses given were ‘anxiety disorders’ (78.3%), ‘stress disorders’ (76.7%) and ‘depressive disorders’ (69.3%).

SLaM veterans were significantly more likely to be given a ‘drug misuse’ (4.8%) or an ‘other mental disorder’ diagnosis (15.3%) or to be assigned ‘psychosis disorder’ (2.1%) than Combat Stress veterans. Combat Stress veterans were significantly more likely to be given a ‘depressive disorder’ (69.3%), ‘anxiety disorder’ (78.3%) or ‘stress disorder’ (76.7%) diagnosis than SLaM veterans.

Table 7: Number of SLaM veterans and Combat Stress veterans diagnosed with a mental health disorder.

Diagnosis	Overall (n=2,576)	SLaM Veterans (n=189)	Combat Stress Veterans (n= 189)	
	n (%)	n (%)	n (%)	p value
Alcohol use disorder	58 (15.34)	22 (11.64)	36 (19.05)	0.046 *
Anxiety disorder	158 (41.80)	10 (5.29)	148 (78.31)	< 0.001 *
Depressive disorder	158 (41.80)	27 (14.29)	131 (69.31)	< 0.001 *
Drug disorder	9 (2.38)	9 (4.76)	0 (0)	0.002 *
Other mental disorders[26]	32 (8.47)	29 (15.34)	3 (1.59)	< 0.001 *
Psychosis disorder	4 (1.06)	4 (2.12)	0 (0)	0.044 *
Stress disorder	165 (43.65)	20 (10.58)	145 (76.72)	< 0.001 *
Overseas	-	-	60 (4.7)	-
UK			1,228 (95.3)	

[25] *‘This is not surprising considering that Combat Stress mainly focuses on veterans who have been diagnosed with Post-Traumatic Stress Disorder.*

[26] *‘Other’ disorders include dementia, delirium, dissociative disorders, somatoform disorders, eating disorders, sexual disorders, developmental disorders, hyperkinetic disorders, mood disorders (not including depression), bipolar disorder, personality disorders, neurotic disorders and self-harm and self-poisoning.*

SUMMARY

- 189 SLaM veterans and 189 Combat Stress veterans were included in this analysis.
- The majority were male, with a median age of 40.0 years for SLaM veterans and 40.8 years for Combat Stress veterans.
- SLaM veterans were most likely to live alone, whereas Combat Stress veterans were most likely to live with their partner/children.
- Combat Stress veterans were more likely to have British ethnicity than SLaM veterans.
- The most common mental health diagnoses for SLaM veterans were depressive disorders.
- The most common mental health diagnoses for Combat Stress veterans were anxiety disorders.
- Combat Stress veterans were significantly more likely to have diagnoses of anxiety, depressive, stress and alcohol disorders than SLaM veterans, whilst SLaM veterans were more likely to have drug, psychosis disorder and other mental disorders.

Chapter Eight: Discussion

This research used a Case Register to explore the utility and feasibility of identifying veterans who accessed secondary mental health care services within the UK using manual and automated approaches. Once veterans had been identified using these approaches, they were matched on age and gender to a civilian cohort extracted from the same Case Register. This allows us to compare the sociodemographic profiles and the types of mental health diagnoses of civilian and veterans who engage in secondary mental health care services within the UK. Further, an additional comparison was performed on a subset of SLaM veterans compared to veterans extracted from Combat Stress.

Key findings

Utility and feasibility for manual identification of veterans

There is currently no marker for identifying veterans in England, Wales or Northern Ireland, unlike Scotland (Leightley et al., 2018; Morgan & Jablensky, 2010), therefore identifying this specific sub-population is a challenge. To the best of our knowledge, only one study exists in relation to secondary health care and the Armed Forces in England and Wales (Leightley et al., 2018). However, this study does not explore secondary mental health care services, it required access to a military cohort, and only explores a limited number of physical health conditions. In the current study, in conjunction with experts in database management, we developed a Structured Query Language search strategy based on 19 military-related terms and phrases to identify veterans from the Case Register. Due to time, manpower and cost restraints we were only able to run the approach on a small sub-sample of the Case Register, with each patient taking approximately 11-16 minutes to review. Nevertheless, we were able to identify 693 veterans, with each being manually verified as being a veteran. However, we found the utility and feasibility of this approach was impractical, time consuming and not suitable for wide-scale use.

Utility, feasibility and validation of the Military Service Identification Tool

Building upon our knowledge of the Structured Query Language approach, we developed the MSIT, a combined NLP and machine learning approach which automatically inspects a patient's medical record to determine whether they are a civilian or veteran. Testing of the MSIT showed that it can achieve accuracy rates of 97%, with performance comparable to that of a human coder. When applied to the Case Register, MSIT was able to identify 2,922 veterans in under 15 minutes when executed across 150,000 patient records. This far outperformed the Structured Query Language search strategy.

We conducted a validation study which compared MSIT classifications to patients' self-reported veteran/civilian statuses. We obtained this information via an online self-reported survey. Overall, 83.6% of the MSIT classifications were accurate. We calculated a 0.83 sensitivity and 0.92 specificity for MSIT. Among the small number misclassified, most were civilians miscategorised as veterans; keywords that contributed to misclassification have been used to further train the tool, however no substantial changes to the tool were needed.

Whilst a self-reported validation supported the MSIT's accuracy overall, this method relied upon the truthful disclosure of participants. This is also an issue for medical records which often collect patient characteristics (e.g. sociodemographic details) based on what patients tell clinicians and what they believe is relevant to raise during points of contact. The reliance upon self-report is therefore a ubiquitous issue that will affect the correct recording of veteran status and other key demographic or occupational indicators beyond the scope of this study.

Overall, the ability to identify veterans could, as shown in other findings of this report, provides insights into the unique mental health needs of veterans and their pathways and use of secondary mental health care services. EHR-based case registers, such as the one used in this study, function as single, complete and integrated electronic versions of traditional paper health records (Leightley et al., 2018). These registers have been positioned as a 'new generation' for health research and are now mandatory in the UK. The methodological advantages of these registers – including their longitudinal nature, largely structured fields and their detailed coverage of defined populations – make them an ideal research and surveillance tool (Stewart, 2014). To our knowledge, this is the only study applying NLP to the problem of identifying veterans, developed and tested against a large Case Register.

Veteran and civilian comparison

In this study, we wanted to explore Post National Service era veterans who accessed secondary mental healthcare services. The present analysis found the majority of those in secondary mental health care services were White British, and male. This follows a similar profile of the Armed Forces (Fear et al., 2010; Hotopf et al., 2006; Stevelink et al., 2018). Civilians were significantly more likely to be Asian or British Asian, reflecting the catchment area of SLaM which is highly ethnically diverse (Perera et al., 2016; Stewart, 2014).

Most civilians and veterans reported living alone; previous research has indicated that those who live alone utilise health care services more frequently (Dreyer et al., 2018). Research conducted using data extracted from the Case Register showed that living alone resulted in an increase of health care service utilisation for older patients (Dreyer et al., 2018). Future work should seek to explore the role of social isolation and secondary mental health care utilisation in relation to complex mental health needs among veterans.

There is emerging evidence that the places where people live are an important factor in determining and sustaining inequalities in mental health outcomes (Fone et al., 2007). The majority of the sample included in this study lived in a deprived area, which may have impacted negatively on their presentation of symptoms, adherence to care and mental health outcomes (Fone & Dunstan, 2006).

This study found that veterans were more likely to be given a 'stress disorder', 'depressive disorder', 'anxiety disorder', 'psychosis disorder 'or 'personality disorder' than civilians. These diagnoses in veterans are expected and mirrors previous research (Fear et al., 2010; Iversen et al., 2010; Stevelink et al., 2019), with exception to personality and psychosis disorders where there is a lack of research into these conditions in military populations. This is likely to be because they can be deemed incompatible with military service. Further research into the characteristics of veterans with these diagnoses would be enlightening.

Analyses also indicated no differences in 'alcohol use disorder' between both samples despite the current literature that indicates greater alcohol use in veterans (Rhead et al., 2019; Stevelink et al., 2019). This could be due to veteran seeking help infrequently; being highly functional despite misusing alcohol or geographical differences, with around 53% accessing formal medical support for alcohol misuse (Hines et al., 2014; Irizar et al., 2020; Stevelink et al., 2019).

Finally, this study found that there were significant differences between civilians and veterans for use of Mental Health Act (1983) sectioning powers, with veterans being significantly more likely to be sectioned than civilians. This could be due to veterans experiencing a higher number of comorbidities, seeking help only at crisis point, having known risk factors for mental health disorders (i.e. isolation, living alone, unemployed), being recruited from deprived areas and potentially being considered riskier in their behaviours.

As part of the study design, 1,634 National Service Era veterans were excluded. This subgroup represented 56% of the sample, which may indicate that older veterans access secondary mental health care services more frequently than younger veterans. Future work should seek to explore this further. It is important to acknowledge that those included in this study have sought help for a mental health problem, and consideration should be given to those who are not seeking any formal help (Stevelink et al., 2019).

Combat Stress comparison

SLaM veterans were more likely to be single and to live alone, whereas Combat Stress veterans were more likely to be in a relationship and to live with their partner and/or children. These sociodemographic differences suggest that SLaM veterans are more socially isolated and perhaps that the veterans of Combat Stress receive more support from their families than veterans accessing other services. Additionally, Combat Stress veterans were more likely to have British ethnicity than SLaM veterans. This highlights that SLaM veterans are a more ethnically diverse group of individuals.

The more frequent diagnoses of depressive, anxiety, alcohol and stress disorders in the Combat Stress sample and the more frequent diagnoses of drug and psychosis disorders in the SLaM sample are unsurprising considering Combat Stress' specialism and the broad array of services offered by SLaM.

Strengths and limitations

In this report, we have demonstrated that it is possible to identify veterans who accessed secondary mental health care services in the UK, by using a Case Register; and assessed their sociodemographic and diagnostic characteristics in comparison to civilians. This report represents the first study of its kind to compare veterans with a matched civilian cohort in a help-seeking secondary mental health care population.

To support the identification of military veterans, we developed the MSIT, to systematically search patient records and flag any that are a potential military veteran. The tool was able to identify potential veterans with a high precision. We were able to validate the MSIT classifications by comparing to patients' self-reported veteran/civilian status via a validation study. Results suggested a high degree of accuracy with 83.6% of cases correctly identified when corroborating with patients' self-reported status. We did find, however, that veterans may be misclassified as civilians if they did not disclose their veteran status in consultations or if this was not recorded by clinicians. There was some evidence of civilians being misclassified as veterans due to keywords arising in other contexts, e.g. in relation to family members, as metaphors, or other services. This allowed us to make minor revisions to the keywords used by the tool, however our findings suggest that MSIT does not require any substantial changes.

Overall, a key strength of the MSIT was the exploitation of NLP, which is advantageous for automating the process of identification and reducing the possibility of human error and bias, and overcomes challenges found when using military cohorts linked to case registers(Leightley et al., 2018; Rhead et al., 2019). MSIT does not rely on any codes (clinical or otherwise) or structured fields, which broadens its application to types of uses, such as diagnosis, occupation and ethnicity detection. Further work is required to refine the tool to function on other datasets, including data collected locally, regionally and nationally. To that end, we have released the source-code of the tool (accessible via this **link**) and hope that other researchers and organisations will contribute to its development.

Overall, this project encountered numerous logistical, bureaucratic and technical challenges. For example, CRIS is managed by an internal administration team and therefore KCMHR researchers did not have direct access to all information or processes. As a result, the researchers were unable to investigate which exclusion criteria reduced the original eligible sample of ~500,000 extracts to 902 eligible contacts. We therefore could not determine whether veterans in SLaM were more likely to lack C4C, to be a current inpatient, or to be experiencing psychosis or dementia than civilians. It is, however, possible that the comparatively small veteran sample of the present study is attributable to the complexity of mental health problems compared to civilians.

Lastly, we were unable to compare the treatment characteristics of veterans and civilians due to a high degree of missing data on SLaM service utilisation variables. Despite this, we were able to analyse usage of the Mental Health Act (MHA) 1983. Further research will be required to ascertain the characteristics of those sectioned under the MHA, and to compare veterans and civilians in other trusts nationally.

Next Steps

The results of this research have implications for the ways in which veterans receive secondary mental health care services, and in our understanding of how they use these kinds of services. To ensure a broad and realistic discussion of the implications of this research, a stakeholder event was held towards the end of this project. Representatives from secondary mental health care providers, providers of veterans' mental health treatment and support and academics attended including:

- King's College London;
- Combat Stress;
- Centre for Mental Health;
- Ministry of Defence;
- Help for Heroes;
- NHS;
- Department of Health and Social Care;
- NHS Digital;
- The Royal British Legion;
- Chester University;
- Northumbria University; and
- Veterans Office (Northern Ireland).

After hearing the research findings presented in this report, stakeholders worked together in three groups to discuss the results of the research, the potential impact on practice and policy, and future research directions. The session culminated in in a set of key recommendation for the academic, public and Government sectors to address.

Recommendations

1. We recommend improving the accuracy and efficiency of identifying veterans from the Case Register by ensuring that serving status is asked when a patient is registered:
   - We were able to check the accuracy of MSIT by corroborating classifications with medical records and patients' self-reported veteran/civilian status.
   - The survey and medical records rely upon participants accurately disclosing their veteran/civilian status. For medical records, this also relies upon clinicians asking about, and accurately recording, this information.
   - To overcome this problem, we recommend the implementation of a military service

marker in the Case Registers and similar NHS databases, and if possible, that these then be verified with Ministry of Defence records, and/or alternatively that approval be sought from SLaM to contact patients to validate their military service. This would indicate which patients had previously served in the Armed Forces, however this could be time consuming and will not enable retrospective analyses.

- We recommend refining the military-related terms employed to search the Case Register by using more specific and detailed phrases, such as individual regiment or operation names, to increase the potential of improving hit rates for detecting veterans in large, non-military specific Case Registers.

2. We recommend that the collection of core socio-demographic variables is mandated for all Case Registers:
   - We endorse backfilling missing data for outcome variables by using clinical written notes to manually work through patient records one-by-one. Details left out of the database's structured fields are often included within these free-text fields and would allow us to improve data quality. We believe that future work should seek to develop NLP tools to aid backfilling data automatically.

3. We recommend the implementation of new techniques, to ensure that missing data within the Case Registers is kept to a minimum:
   - Data held by case registers is not collected primarily for research purposes and therefore often has large amounts of missing values23. Indeed, this was the case for the current report. It would be helpful if data entry for some fields was made mandatory in Case Registers to ensure that the most important information was available for all patients.

4. We recommend accelerating the methodology for identifying veterans from the Case Register through further development of the Military Service Identification Tool:
   - The validation exercise achieved in the present study has provided further information with which to refine MSIT and will contribute to more precise classifications.
   - A further step would involve us teaching the computer to identify veterans by automatically annotating military words without human involvement. Developing the MSIT in this way enables future projects using the Case Register (or other register(s)) and the same data inputs to implement a consistent, reliable and efficient approach to identifying veteran medical records.
   - It is important to acknowledge that veteran status is a highly sensitive personal attribute therefore further research should explore how MSIT can be deployed in a safe manner.
   - This report has demonstrated that the MSIT identified probable veterans with a high degree of accuracy. Future research needs to be conducted to rigorously evaluate the tool to ensure it is suitable for large-scale deployment.

5. We recommend conducting further analysis on the prevalence of mental health problems and how they compare to civilians. In addition, we recommend carrying out further statistical tests on socio-demographic data points which are available within the Case Register to provide a more detailed account of the patient's experiences in mental health care services:
   - We have reported on a range of socio-demographic and diagnostic characteristics already. We suggest exploring how to address issues with incomplete data for key variables of interest, such as benefit and employment status, service utilisation, treatment pathways and treatment outcomes.

6. We recommend conducting further research to assess the generalizability and scalability of our findings at a local, regional and national level:
   - This study identified significant differences between civilian and veteran diagnoses of mental health disorders, specifically for alcohol misuse. Findings also highlighted that veterans were more likely to be sectioned under the MHA than civilians. These should be considered further, especially in light of potential geographical difference differences across the UK.
   - This study indicated that veterans who seek help have more complex mental health needs. It is important that we explore how secondary mental health care services are coping with these complex needs, and what else can be done, future work should seek to answer these points.
   - The findings in this report represent a single secondary mental health care provider in London. We recommend that future research seek to compare these rates to national data sources, and to compare with other providers of veteran secondary mental health care such as charities.

7. We recommend that future work be conducted to explore the needs of National Service Era veterans, and how these needs might be different from Post National Service Era veterans:
   - As a result of the MSIT, we were able to identify National Service Era veterans and overseas veterans who may present with unique needs and a specific burden upon healthcare services. This was a key discussion point at the stakeholder event, with the recommendation that future work should explore this further.

8. We recommend increasing the number of veterans identified from the Case Register by including a larger number of patient records:
   - As a result of the current method for identifying veterans, we were only able to include 1,288 civilians and veterans in each of our two groups. While this sample is sufficient for this study, it is lacking in statistical power for more complex analyses.

9. We recommend the development of an educational tool for those involved in the care of veterans to highlight their mental health needs and to support recognising a veteran to enable accurate recording of their status.

Conclusions

This study is the first in the UK to identify military veterans and to explore veterans and matched civilians who sought help from a secondary mental health care trust in England. This research used a Case Register to explore the utility and feasibility of identifying veterans who accessed secondary mental health care services, using manual and automated approaches. Based on a validation of patients' self-reported status, MSIT was 83.6% accurate in determining veteran and civilian status. This shows MSIT presents a more accurate and less labour-intensive solution than manual identification, where accuracy was estimated to be 43%. Once veterans had been identified, they were matched on age and gender to a civilian cohort extracted from the Case Register.

Comparing these two samples presented interesting findings. Our analyses showed that veterans experience more complex and severe mental health problems than civilians and that those accessing NHS secondary mental health treatment and Combat Stress have different demographic and diagnostic profiles. This report has also shed new insights into the high levels of sectioning in veterans compared to civilians.

When comparing SLaM veterans and civilians, there were no significant differences between civilians and veterans for alcohol use disorder, which is a surprise considering research consistently indicates that veterans misuse alcohol more than the UK general population. However, we did find that veterans were more likely to be diagnosed with anxiety, depressive, psychosis, personality and stress disorder. These findings suggest that veterans were more likely to be sectioned under the MHA than civilians. We must, however, caution that further research is required to explore why this might be the case.

To complement the analysis, we performed a comparison of help-seeking veterans at Combat Stress. We found that Combat Stress veterans had higher frequency rates of anxiety, depressive, stress and alcohol disorders than SLaM veterans, indicating the complexity of patients Combat Stress treats. This demonstrates that veterans accessing services vary across the board, and a more holistic, individualised approach is needed to support them through treatment.

Appendices

Appendix 1: Accessing the Case Register

There were several steps involved in gaining full access to the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (BRC) Case Register. In total, these processes took approximately six months to gain the required approvals for the study team.

Project approval

The research team applied to the BRC to request approval of the study. This detailed a summary of the research, objectives, rationale, the types of variables required and the expected outputs. A copy of this application is available upon request. The application was then discussed at a BRC committee meetings (which meets monthly) and was approved.

Research passports

Following study approval, the research team was required to obtain research passports to access the Case Register. Research passports are the equivalent of an honorary contract with the SLaM NHS Foundation Trust and ensure that researchers are contractually obliged to adhere to relevant Trust policies regarding confidentiality and data protection.

Firewall constraints

The Case Register (and all data needed for the study) could only be accessed using the SLaM firewall. This meant that the research team had to source computers at the BRC (located at the King's College London Hospital) to access the Case Register. However, we were also able to apply for access for a remote connection to the SLaM network, allowing us to carry out data cleaning and analyses from our King's College London campus computers.

Appendix 2: Inclusion and exclusion terms

Table 8: Inclusion and exclusion criteria used in the identification of veterans from SLaM.

Included key words	Exclusion	Notes
Army	“who was/is in (the) army”	Majority of times this refers to someone other than the patient
	“Salvation Army”	
	“army knife”	
	“army gear”	
	“army style”	
	“army cadet”	
	“army cadette”	
	“army themed”	
	“child army”	
	“army family”	
	“rebel army”	
	“refugee army”	
	“army service”	
	“private army”	
	“army green”	
	“army <item of clothing>”	Clothing
	“army type”	
	Foreign armies: Eritrea, Sri Lanka	Reference to service in non-UK army, or experiences relating to non-UK army
Navy	“navy blue”	Clothing
	“dark navy”	Clothing
	“navy colour”	Clothing
	“wearing (a) navy”	Clothing
	“dressed in navy”	Clothing
	“navy <item of clothing>”	Clothing
	“Merchant Navy”	
	“Army and Navy Store”	
	“worked for Navy, Army, Air Force Institute”	NAAFI
	“<family member> was/is in (the) navy”	Family member in Navy
	“due to join the Navy”	(Thinking of) joining Navy
	“accepted into Navy”	(Thinking of) joining Navy
	“potential careers, including Navy”	(Thinking of) joining Navy
	Foreign navies: Italian, US, Israeli, Portuguese, Burmese, Eritrea	Reference to service in non-UK navy, or experiences relating to non-UK navy

Continued overleaf

Table 8: *Continued*

Included key words	Exclusion	Notes
RAF / air force	"<family member> was/is in (the) RAF"	Family member in RAF
Ex-service		
Veteran		
Armed forces		
Afghan		Deployment location
Iraq		Deployment location
Bosnia		Deployment location
Kosovo		Deployment location
Falklands		Deployment location
N Ireland		Deployment location
Cyprus		Deployment location
Germany		Deployment location
Enlisted		
National service		
Combat Stress		Military charity
SSAFA		Military charity
Help for Heroes		Military charity

Appendix 3: Technical summary of the Military Service Identification Tool

This Appendix presents the technical details for the development of the Military Service Identification Tool. For more information: kcmhr@kcl.ac.uk

Data approach for the Military Service Identification Tool

There are approximately 500,000 correspondence documents within the Case Register, due to the large volumes of data a sub-set was extracted for the development of the Military Service Identification Tool (MSIT). This subset was extracted using the Personal History Detection tool which has been developed by the Biomedical Research Centre based at King's College London(NIHR Biomedical Research Centre (BRC) - King's College London, 2019).

Each personal history record contains a forensic analysis of each patient's life events since birth; these include education attainment, childhood adversity, employment and relationship information. Each record is written by a clinician. The personal history dataset contains 98,395 documents, after an informal scoping exercise and discussions with Natural Language Processing (NLP) experts whom have experience of using the Case Register the decision was made to retain only 6,672 documents (called the Gold Standard dataset in this report), which represented 4,200 patients. For the machine learning and rule-based combined tool, a decision was made to retain 66% of the dataset for training, and the remainder 34% used for testing and evaluation.

The gold standard dataset was independently annotated by Dr Daniel Leightley and Elena Opie with acceptable inter-rater agreement as indicted by a Cohen's kappa of 0.83 for veterans and 0.89 for civilians.

Developing the Military Service Identification Tool

The MSIT utilised the gold standard dataset which was refined upon first test-run of the classification algorithms as new cases were discovered. The classification framework was trained to identify veterans compared to civilians based on the use of military terms and phases. A training set of 4,470 annotated documents was used to select a machine learning classifier. There is sparse literature on which machine learning algorithms are best suited for a specific task, not only in the field of NLP but also for areas such as healthcare, agricultural and security(Ahad et al., 2008; Cunningham et al., 2018; Leightley et al., 2013, 2017).

To ensure the appropriate selection the classifier for the tool, a comparison was made based on ten-fold cross validation performance of the following machine learning classifiers: Random Forest, Decision Tree, Linear Support Vector Classifier, Support Vector Classifier, Multinomial Naïve Bayes, k-Nearest Neighbour, Logistic Regression and Multi-layered Perception. Linear Support Vector Classifier obtained the highest accuracy (see Table 9) and was used as the machine learning classifier for MSIT.

Table 9: Machine learning classifier accuracy based on gold standard document dataset

Classifier	Accuracy
Random Forest	0.84
Decision Tree	0.91
Linear Support Vector Classifier	0.95
Support Vector Classifier	0.84
Multinomial Naïve Bayes	0.90
k-Nearest Neighbour	0.89
Logistic Regression	0.88
Multi-layered Perception	0.94

To improve the true positive rate of the MSIT, and to reduce the potential for false positives, a post-processing of the machine learning outcome was applied based on known military terms and phrases described earlier in this report. For each document that was predicted as being that of a veteran, a check was performed to ensure the document used a military term of phrase (i.e. 'served in the military', 'national service', 'served in the forces'). The MSIT was then applied to the Case Register to identify veterans required for the study.

Appendix 4: Further information for the Validation study

Data collection

Four additional questions were asked of patients who had a self-reported status of 'veteran'. These were as follows:

Question 2:
Which part of the Armed Forces did you serve in?

- Royal Navy
- Royal Marines
- Army
- Royal Air Force

Question 3:
What was the highest rank when you left the military?

- Senior Commissioned Officer (Cdr/Lt Col/Wg Cdr and above)
- Commissioned Officer (to Lt Cdr/Maj/Sqn Ldr)
- Senior Non-Commissioned Officer
- Junior Non-Commissioned Officer
- Other ranks (AB/Pte/AC/JT or equivalent)
- Other (please specify your highest rank only)*

**Please note: Free text responses are not reviewed during data collection.*

Question 4:
How long did you serve in the Armed Forces?

Question 5:
When you left the Armed Forces, were you:

- Regular
- Reserve
- Both

Ethical information

The validation study had minimal risks as we only asked for minimal information about participants' occupational status. It was possible that those most affected by answering questions on their occupational status did not take part in the study or fill in certain questions. However, some participants might have negative perceptions of military service due to certain life experiences or opinions. To manage and mitigate potential risks, we adopted the following process:

1. We provided contact details for any concerned participants to contact the research team directly;
2. Those working on the study have extensive experience recruiting, and liaising with, vulnerable groups and were therefore equipped to manage any concerns. Where participants reported distress (n=1) or required us to contact their care-coordinator (n=1), we responded by ensuring the first participant was not re-contacted, was removed from C4C, and ensured that any named clinician/ care co-ordinator who gave us permission to contact was recorded;
3. A risk protocol was implemented whereby the research team would support the participant and signpost to relevant services if there were low risk signs of distress. This was not necessary in the current study;
4. If distress was more serious, Dr. Dominic Murphy, Clinical Psychologist at Combat Stress, was available if participants consented to a call back. This was not necessary in the current study;
5. Where risk to the participant or others is serious and/or imminent, confidentiality would have been breached and relevant clinical teams and authorities would have been contacted. Participants were made aware of the possibility of a breach in the Participant Information Sheet. This was not necessary in the current study.

Appendix 5: Diagnosis categories

Table 10: Diagnostic grouping of ICD-10 diagnostic coding

Diagnosis Group	ICD-10 Coding
Alcohol Use Disorder	F10 - Mental and behavioural disorders due to use of alcohol
Drug Disorder	F11 - Mental and behavioural disorders due to use of opioids
	F12 - Mental and behavioural disorders due to use of cannabinoids
	F13 - Mental and behavioural disorders due to use of sedatives or hypnotics
	F14 - Mental and behavioural disorders due to use of cocaine
	F15 - Mental and behavioural disorders due to use of other stimulants, including caffeine
	F16 - Mental and behavioural disorders due to use of hallucinogens
	F17 - Mental and behavioural disorders due to use of tobacco
	F18 - Mental and behavioural disorders due to use of volatile solvents
	F19 - Mental and behavioural disorders due to multiple drug use and use of other psychoactive substances
Stress Disorder	F43 - Reaction to severe stress, and adjustment disorders
Depressive Disorder	F30/F39 - Mood [affective] disorders
Anxiety Disorder	F40 - Phobic anxiety disorders
	F41 - Other anxiety disorders
	F42 - Obsessive-compulsive disorder
	F46 - Other neurotic disorders
	F48 - Other neurotic disorders
Psychosis Disorder	F20/F29 - Schizophrenia, schizotypal and delusional disorders
Personality Disorder	F60 - Specific personality disorders
	F61 - Mixed and other personality disorders
Other Mental Disorders	All other 'F' codes not denoted above.

References

Ahad, Md. A. R., Tan, J. K., Kim, H. S., & Ishikawa, S. (2008). Human activity recognition: Various paradigms. 2008 International Conference on Control, Automation and Systems, 1896–1901. https://doi.org/10.1109/ICCAS.2008.4694407

Allebeck, P. (2009). The use of population based registers in psychiatric research. Acta Psychiatrica Scandinavica, 120(5), 386–391. https://doi.org/10.1111/j.1600-0447.2009.01474.x

Bergman, B. P., Mackay, D. F., Smith, D. J., & Pell, J. P. (2016). Long-Term Mental Health Outcomes of Military Service: National Linkage Study of 57,000 Veterans and 173,000 Matched Nonveterans. The Journal of Clinical Psychiatry, 77(6), 793–798. https://doi.org/10.4088/JCP.15m09837

Cambria, E., & White, B. (2014). Jumping NLP Curves: A Review of Natural Language Processing Research. IEEE Computational Intelligence Magazine, 9(2), 48–57.

Chowdhury, G. G. (2005). Natural language processing. Annual Review of Information Science and Technology, 37(1), 51–89. https://doi.org/10.1002/aris.1440370103

Collen, M. F. (1990). Clinical research databases? A historical review. Journal of Medical Systems, 14(6), 323–344. https://doi.org/10.1007/BF00996713

Cunningham, R., Sánchez, M., May, G., & Loram, I. (2018). Estimating Full Regional Skeletal Muscle Fibre Orientation from B-Mode Ultrasound Images Using Convolutional, Residual, and Deconvolutional Neural Networks. Journal of Imaging, 4(2), 29. https://doi.org/10.3390/jimaging4020029

Dreyer, K., Steventon, A., Fisher, R., & Deeny, S. R. (2018). The association between living alone and health care utilisation in older adults: a retrospective cohort study of electronic health records from a London general practice. BMC Geriatrics, 18(1), 269. https://doi.org/10.1186/s12877-018-0939-4

Fear, N. T., Jones, M., Murphy, D., Hull, L., Iversen, A. C., Coker, B., Machell, L., Sundin, J., Woodhead, C., Jones, N., Greenberg, N., Landau, S., Dandeker, C., Rona, R. J., Hotopf, M., & Wessely, S. (2010). What are the consequences of deployment to Iraq and Afghanistan on the mental health of the UK armed forces? A cohort study. The Lancet, 375(9728), 1783–1797. https://doi.org/10.1016/S0140-6736(10)60672-1

Fernandes, A. C., Cloete, D., Broadbent, M. T., Hayes, R. D., Chang, C.-K., Jackson, R. G., Roberts, A., Tsang, J., Soncul, M., Liebscher, J., Stewart, R., & Callard, F. (2013). Development and evaluation of a de-identification procedure for a case register sourced from mental health electronic records. BMC Medical Informatics and Decision Making, 13(1), 71. https://doi.org/10.1186/1472-6947-13-71

Fernandes, A. C., Dutta, R., Velupillai, S., Sanyal, J., Stewart, R., & Chandran, D. (2018). Identifying Suicide Ideation and Suicidal Attempts in a Psychiatric Clinical Research Database using Natural Language Processing. Scientific Reports, 8(1), 7426. https://doi.org/10.1038/s41598-018-25773-2

Fone, D., Dunstan, F., Lloyd, K., Williams, G., Watkins, J., & Palmer, S. (2007). Does social cohesion modify the association between area income deprivation and mental health? A multilevel analysis. International Journal of Epidemiology, 36(2), 338–345. https://doi.org/10.1093/ije/dym004

Fone, D. L., & Dunstan, F. (2006). Mental health, places and people: A multilevel analysis of economic inactivity and social deprivation. Health & Place, 12(3), 332–344. https://doi.org/10.1016/j.healthplace.2005.02.002

Giebel, C. M., Clarkson, P., & Challis, D. (2014). Demographic and clinical characteristics of UK military veterans attending a psychological therapies service. The Psychiatric Bulletin, 38(6), 270–275. https://doi.org/10.1192/pb.bp.113.046474

Gray, L., Batty, G. D., Craig, P., Stewart, C., Whyte, B., Finlayson, A., & Leyland, A. H. (2010). Cohort Profile: The Scottish Health Surveys Cohort: linkage of study participants to routinely collected records for mortality, hospital discharge, cancer and offspring birth characteristics in three nationwide studies. International Journal of Epidemiology, 39(2), 345–350. https://doi.org/10.1093/ije/dyp155

Gundlapalli, A. V, Carter, M. E., Palmer, M., Ginter, T., Redd, A., Pickard, S., Shen, S., South, B., Divita, G., Duvall, S., Nguyen, T. M., D'Avolio, L. W., & Samore, M. (2013). Using natural language processing on the free text of clinical documents to screen for evidence of homelessness among US veterans. AMIA Annu Symp Proc, 2013, 537–546.

Hines, L. A., Goodwin, L., Jones, M., Hull, L., Wessely, S., Fear, N. T., & Rona, R. J. (2014). Factors Affecting Help Seeking for Mental Health Problems After Deployment to Iraq and Afghanistan. Psychiatric Services, 65(1), 98–105. https://doi.org/10.1176/appi.ps.004972012

Hotopf, M., Hull, L., Fear, N. T., Browne, T., Horn, O., Iversen, A., Jones, M., Murphy, D., Bland, D., Earnshaw, M., Greenberg, N., Hacker Hughes, J., Tate, A. R., Dandeker, C., Rona, R., & Wessely, S. (2006). The health of UK military personnel who deployed to the 2003 Iraq war: a cohort study. The Lancet, 367(9524), 1731–1741. https://doi.org/10.1016/S0140-6736(06)68662-5

Irizar, P., Leightley, D., Stevelink, S., Rona, R., Jones, N., Gouni, K., Puddephatt, J.-A., Fear, N., Wessely, S., & Goodwin, L. (2020). Drinking motivations in UK serving and ex-serving military personnel. Occupational Medicine. https://doi.org/10.1093/occmed/kqaa003

Iversen, A. C., van Staden, L., Hughes, J. H., Browne, T., Greenberg, N., Hotopf, M., Rona, R. J., Wessely, S., Thornicroft, G., & Fear, N. T. (2010). Help-seeking and receipt of treatment among UK service personnel. British Journal of Psychiatry, 197(02), 149–155. https://doi.org/10.1192/bjp.bp.109.075762

Leightley, D., Chui, Z., Jones, M., Landau, S., McCrone, P., Hayes, R. D., Wessely, S., Fear, N. T., & Goodwin, L. (2018). Integrating electronic healthcare records of armed forces personnel: Developing a framework for evaluating health outcomes in England, Scotland and Wales. International Journal of Medical Informatics, 113, 17–25. https://doi.org/10.1016/j.ijmedinf.2018.02.012

Leightley, D., Darby, J., Baihua Li, McPhee, J. S., & Moi Hoon Yap. (2013). Human Activity Recognition for Physical Rehabilitation. 2013 IEEE International Conference on Systems, Man, and Cybernetics, 261–266. https://doi.org/10.1109/SMC.2013.51

Leightley, D., McPhee, J. S., & Yap, M. H. (2017). Automated Analysis and Quantification of Human Mobility Using a Depth Sensor. IEEE Journal of Biomedical and Health Informatics, 21(4), 939–948. https://doi.org/10.1109/JBHI.2016.2558540

Leightley, D., Pernet, D., Velupillai, S., Stewart, R. J., Mark, K. M., Opie, E., Murphy, D., Fear, N. T., & Stevelink, S. A. M. (2019). The Development of the Military Service Identification Tool: Identifying Military Veterans in a Clinical Research Database using Natural Language Processing and Machine Learning (Preprint). JMIR Medical Informatics. https://doi.org/10.2196/15852

Leightley, D., Williamson, V., Darby, J., & Fear, N. T. (2019). Identifying probable post-traumatic stress disorder: applying supervised machine learning to data from a UK military cohort. Journal of Mental Health, 28(1), 34–41. https://doi.org/10.1080/09638237.2018.1521946

Manning, C., & Schutze, H. (1999). Foundations of Statistical Natural Language Processing. The MIT Press.

Mellotte, H., Murphy, D., Rafferty, L., & Greenberg, N. (2017). Pathways into mental health care for UK veterans: a qualitative study. European Journal of Psychotraumatology, 8(1), 1389207. https://doi.org/10.1080/20008198.2017.1389207

Ministry of Defence (2016). Veterans: Key facts. https://www.armedforcescovenant.gov.uk/wp-content/uploads/2016/02/Veterans-Key-Facts.pdf

Ministry of Defence (2019) Population Projections: UK Armed Forces Veterans residing in Great Britain, 2016 to 2028. Accessed via https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/775151/20190107_Enclosure_1_Population_Projections_-_UK_Armed_Forces_Veterans_residing_in_Great_Britain_-_2016_to_2028.pdf

Morgan, V. A., & Jablensky, A. V. (2010). From inventory to benchmark: quality of psychiatric case registers in research. British Journal of Psychiatry, 197(01), 8–10. https://doi.org/10.1192/bjp.bp.109.076588

Murphy, D., Hodgman, G., Carson, C., Spencer-Harper, L., Hinton, M., Wessely, S., & Busuttil, W. (2015). Mental health and functional impairment outcomes following a 6-week intensive treatment programme for UK military veterans with post-traumatic stress disorder (PTSD): a naturalistic study to explore dropout and health outcomes at follow-up. BMJ Open, 5(3).

NIHR Biomedical Research Centre (BRC) - King's College London. (2019). https://www.kcl.ac.uk/lsm/research/divisions/hscr/research/biomedical-research-centre-(brc)

Otis, J. D., Keane, T. M., & Kerns, R. D. (2003). An examination of the relationship between chronic pain and post-traumatic stress disorder. Journal of Rehabilitation Research and Development, 40(5), 397–405.

Palmer, I. P. (2012). UK extended Medical Assessment Programme for ex-service personnel: the first 150 individuals seen. The Psychiatrist, 36(07), 263–270. https://doi.org/10.1192/pb.bp.110.033266

Payne, R. A., Abel, G. A., Guthrie, B., & Mercer, S. W. (2013). The effect of physical multimorbidity, mental health conditions and socioeconomic deprivation on unplanned admissions to hospital: a retrospective cohort study. Canadian Medical Association Journal, 185(5), E221–E228. https://doi.org/10.1503/cmaj.121349

Perera, G., Broadbent, M., Callard, F., Chang, C.-K., Downs, J., Dutta, R., Fernandes, A., Hayes, R. D., Henderson, M., Jackson, R., Jewell, A., Kadra, G., Little, R., Pritchard, M., Shetty, H., Tulloch, A., & Stewart, R. (2016). Cohort profile of the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLaM BRC) Case Register: current status and recent enhancement of an Electronic Mental Health Record-derived data resource. BMJ Open, 6(3), e008721. https://doi.org/10.1136/bmjopen-2015-008721

Perera, G., Soremekun, M., Breen, G., & Stewart, R. (2009). The psychiatric case register: noble past, challenging present, but exciting future. The British Journal of Psychiatry : The Journal of Mental Science, 195(3), 191–193. https://doi.org/10.1192/bjp.bp.109.068452

Rhead, R., Deirdre, M., Margaret, J., Greenberg, N., Fear, N. T., & Goodwin, L. (2019). Mental health disorders and alcohol misuse among UK military veterans and the general population: A comparison study. Lancet Psych.

Sharp, M. L., Fear, N. T., Rona, R. J., Wessely, S., Greenberg, N., Jones, N., & Goodwin, L. (2015). Stigma as a barrier to seeking health care among military personnel with mental health problems. Epidemiologic Reviews, 37(1), 144–162. https://doi.org/10.1093/epirev/mxu012

StataCorp. (2015). Stata: Release 15. Statistical Software.

Stevelink, S. A. M., Jones, M., Hull, L., Pernet, D., MacCrimmon, S., Goodwin, L., MacManus, D., Murphy, D., Jones, N., Greenberg, N., Rona, R. J., Fear, N. T., & Wessely, S. (2018). Mental health outcomes at the end of the British involvement in the Iraq and Afghanistan conflicts: a cohort study. The British Journal of Psychiatry, 213(6), 1–8. https://doi.org/10.1192/bjp.2018.175

Stevelink, S. A. M., Jones, N., Jones, M., Dyball, D., Khera, C. K., Pernet, D., MacCrimmon, S., Murphy, D., Hull, L., Greenberg, N., MacManus, D., Goodwin, L., Sharp, M.-L., Wessely, S., Rona, R. J., & Fear, N. T. (2019). Do serving and ex-serving personnel of the UK armed forces seek help for perceived stress, emotional or mental health problems? European Journal of Psychotraumatology, 10(1), 1556552. https://doi.org/10.1080/20008198.2018.1556552

Stewart, R. (2014). The big case register. Acta Psychiatrica Scandinavica. https://doi.org/10.1111/acps.12279

Stewart, R., Soremekun, M., Perera, G., Broadbent, M., Callard, F., Denis, M., Hotopf, M., Thornicroft, G., & Lovestone, S. (2009). The South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLAM BRC) case register: development and descriptive data. BMC Psychiatry, 9(1), 51. https://doi.org/10.1186/1471-244X-9-51

van Hoorn, L. A., Jones, N., Busuttil, W., Fear, N. T., Wessely, S., Hunt, E., & Greenberg, N. (2013). Iraq and Afghanistan veteran presentations to Combat Stress, since 2003. Occupational Medicine, 63(3), 238–241. https://doi.org/10.1093/occmed/kqt017

Weijers, B. & Busuttil, W. (2015). Exploring Patterns in Referrals to Combat Stress for Uk Veterans with Mental Health Difficulties between 1994 and 2014. International Journal of Emergency Mental Health and Human Resilience, 17(3). https://doi.org/10.4172/1522-4821.1000250

Woodhead, C., Rona, R. J., Iversen, A., MacManus, D., Hotopf, M., Dean, K., McManus, S., Meltzer, H., Brugha, T., Jenkins, R., Wessely, S., & Fear, N. T. (2011). Mental health and health service use among post-national service veterans: results from the 2007 Adult Psychiatric Morbidity Survey of England. Psychological Medicine, 41(2), 363–372. https://doi.org/10.1017/S0033291710000759

Wu, C.-Y., Chang, C.-K., Hayes, R. D., Broadbent, M., Hotopf, M., & Stewart, R. (2012). Clinical risk assessment rating and all-cause mortality in secondary mental healthcare: the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLAM BRC) Case Register. Psychological Medicine, 42(8), 1581–1590. https://doi.org/10.1017/S0033291711002698

Wickersham A, Nairi S, Jones R, Lloyd-Evans B. The Mental Health Act Assessment Process and Risk Factors for Compulsory Admission to Psychiatric Hospital: A Mixed Methods Study. Br J Soc Work. April 2019. doi:10.1093/bjsw/bcz037