



**Northumbria
University**
NEWCASTLE

Northern Hub for Veterans' & Military Families' Research



MONARCH

Map of Need Aggregated
Research Study

Report to Funders



20th April 2026

Contents

The MONARCH Team	4
Acknowledgements.....	4
Funding and Support.....	4
Participating Organisations	4
Stakeholder Contributions.....	4
Technical and Research Support	4
Veterans Community	4
Foreword.....	5
Map of Need Aggregated Research Study	6
Executive Summary.....	6
Purpose and Context.....	6
Platform Development	6
Platform Capabilities.....	7
Initial Evaluation Results	7
Impact and Future Direction	7
List of Tables and Figures.....	8
Tables	8
Figures.....	8
Introduction.....	9
Background	9
Map of Need Aggregated Research (MONARCH) Project.....	10
Development of the MONARCH Platform.....	11
Phase 1: Focus Groups with Key Policymakers.....	12
Phase 2: Data Sharing and Ingestion	14
Phase 3: Data Cleaning and Anonymisation.....	15
Phase 4: Data Aggregation and Harmonisation.....	16
Phase 5: Data Visualisation	16
Secondary Data Analysis.....	18
Evaluation of the MONARCH Platform.....	19
Quantitative Data.....	20
SUS Findings.....	21
Summary.....	22
Theme 1: Credibility.....	23
Theme 2: Platform Experience and Functionality	23

Theme 3: Key strengths	24
Overview and Reflections	25
Overview	25
Platform Development.....	25
Secondary Data Analyses	25
MONARCH Platform Evaluation	25
Reflections.....	26
Data Coverage and Representativeness	26
Geographic Coverage	26
Platform Accessibility.....	26
Interpretability and Data Literacy.....	27
Scalability and Sustainability	27
Broader Significance	27
Overall Conclusion.....	27
Appendices.....	28

The MONARCH Team

Primary Investigator	Professor Matt Kiernan
Co-Investigator	Professor Marco Tomietto
Co-Investigator	Professor Gavin Oxburgh
Senior Research Assistant	Dr Amy Johnson
Senior Research Assistant	Mr Andrew McGill
Charity Relationship Manager	Meri Mayhew
Visiting Scholar	Dr Michael Rodrigues
Visiting Scholar	Dr Giuseppe Serra M.D.
Visiting Scholar	Dr Federico Turoldo M.D.

Acknowledgements

Funding and Support

The MONARCH project was funded by The Armed Forces Covenant Fund Trust and Forces in Mind Trust. We are grateful for their support in enabling this research into veterans' needs and service provision across the UK.

Participating Organisations

We extend sincere thanks to the ten Armed Forces charities who generously shared their data: SSAFA, Army Benevolent Fund (ABF), Royal Air Forces Association (RAFA), Blind Veterans UK, Combat Stress, Hull 4 Heroes, The Poppy Factory, Royal Marines Association (RMA), The Royal Naval Benevolent Trust (RNBT), and Walking With The Wounded (WWTW). Their collaboration, whilst maintaining the highest standards of data protection and veteran confidentiality, made this research possible.

Stakeholder Contributions

We are grateful to the 29 focus group participants from NHS England, NHS Wales, Armed Forces charities, and local government (October-December 2023) whose insights shaped the platform's design. We are also very grateful to the colleagues who accessed the platform and the participants, across 37 organisations, who participated in the evaluation (July-September 2025). Special recognition goes to the Data Strategy Working Group members whose expertise enabled the harmonisation of complex datasets and development of meaningful need categories aligned with the All-Party Parliamentary Group (APPG) for complex needs and dual diagnosis criteria.

Technical and Research Support

Our appreciation extends to Analytium for their technical support with the SAS Viya platform, and to Northumbria University for their dedication across all eight development phases.

Veterans Community

Most importantly, we acknowledge the UK veterans whose experiences are represented in this data. This platform exists to better understand and meet their needs, ensuring their voices drive improvements in service provision and policy development.

Foreword

The MONARCH Project represents a significant achievement for the UK's Armed Forces community. For the first time, data from across the Armed Forces charity sector has been securely, ethically, and rigorously consolidated into a single national platform - providing a unique insight into the needs, experiences, and circumstances of veterans who have sought charity support. This report sets out not only the scale and technical sophistication of that achievement, but also the significant potential it holds for shaping more proactive and effective support.

The MONARCH platform responds to a long-recognised challenge: while numerous organisations across the UK deliver vital assistance to veterans, the evidence base guiding policy and service provision has historically been fragmented. The absence of linked data, coupled with the complexities of almonisation, meant that policymakers and practitioners lacked a reliable picture of how and where veterans seek support. By bringing together anonymised casework data from ten major Armed Forces charities - representing assistance from more than 170 organisations - the MONARCH project addresses this gap.

What emerges from this work is more than a dataset; it is an enabling infrastructure. MONARCH has created a resource capable of informing early intervention, improving strategic planning, and enhancing the allocation of resources at local, regional, and national levels. The platform has the potential to provide decision makers with the ability to understand need, identify risk factors, observe trends, and support evidence-based service development for years to come.

The commitment shown by the participating charities, the organisations who contributed to the evaluation, and the veterans whose anonymised data has made this work possible, reflecting a shared dedication to improving outcomes for those who have served. The findings of the MONARCH project highlight both achievement and opportunity. They demonstrate what is possible when collaboration, innovation, and evidence-based practice come together. They also remind us that the work of supporting veterans is ongoing, and that we must continue to strengthen data linkage, broaden representation, and refine our understanding of needs across the entire Armed Forces Community.

The Armed Forces Covenant Fund Trust and the Forces in Mind Trust are proud to have supported this important work.

Anna Wright
Chief Executive Officer
The Armed Forces Covenant Fund Trust



Michelle Alston
Chief Executive Officer
Forces in Mind Trust



Map of Need Aggregated Research Study

Executive Summary

Purpose and Context

The MONARCH (Map of Need Aggregated Research) project represents the UK's first consolidated dataset documenting the support needs and service use of military veterans and the Armed Forces Community across the charity sector. Developed by the Northern Hub for Veterans and Military Families' Research at Northumbria University, the MONARCH platform provides a greater understanding of veterans needs which can directly inform tailored service provision and increase understanding of those who may be at risk. In doing so, this could facilitate early intervention for the veteran community at need prior to crisis level.

Platform Development

The MONARCH platform brought together casework data from 10 Armed Forces charities (incl. SSAFA, ABF, RAFA & Combat Stress) through a five-phase structured development process. However, because of *almonisation*, which involves combining and standardising data across organisations, the dataset actually reflects support provided by an additional 173 organisations beyond those original 10.

Phase 1 - Stakeholder Consultation: Focus groups with 29 representatives from NHS England, NHS Wales, Armed Forces charities, and local government (October-December 2023) identified requirements for veteran identification, geospatial analysis¹ capabilities, and an accessible interactive platform with user guidance.

Phase 2 - Data Sharing & Ingestion: Ten charities contributed data through formal data-sharing agreements with GDPR compliance. A secure upload portal was created allowing each charity to submit datasets via unique password-protected URLs directly into the SAS cloud environment².

Phase 3 – Data Cleaning & Anonymisation: A secure hashing algorithm³ anonymised identifying data whilst enabling cross-dataset linkage using service number, date of birth, and gender. An encryption salt⁴ ensured data could not be reverse engineered. Extensive data preparation (4-6 months) addressed inconsistencies in charity data collection protocols, including validation of service numbers, birth dates, and gender information, plus correction of misclassifications and incomplete postcode data in order to enable comparative analysis.

Phase 4 – Data Almonisation, Aggregation and Harmonisation: In military charities, the same individual is often supported by multiple charities for the same issues. To account for this, the process of data aggregation eliminated double counting

¹ A detailed process which involves looking at data on a map to understand where specific things are happening and spot patterns or trends across different locations.

² An online platform that stores, analyses and visualises data using powerful tools without needing additional software on your computer.

³ A method that turns data into a fixed-length code so it can be securely stored and/or checked without revealing the original information.

⁴ A random extra value added to data before it is encrypted or 'hashed' to make it much harder for cyber attackers to guess or crack it.

across 173 organisations. A Data Strategy Working Group refined 280 different recorded needs into 51 distinct categories, mapped to 11 'All Party Parliamentary Group' complex need criteria and the World Health Organisation's (WHO) three health elements (physical health, mental health, social wellbeing).

Phase 5 - Data Visualisation: Built using SAS Viya (a platform already used by MoD and government departments), the interactive platform presents census data, veteran demographics, and needs analysis with dynamic filtering capabilities at local, regional, and national levels. Custom map files were created to address boundary misalignments, with attention to colour-deficiency accessibility and screen-size responsiveness.

Platform Capabilities

The MONARCH platform enables policymakers, HM Government departments, charities, and the NHS (and others) to identify risk factors for early intervention, understand geographic variations in veteran needs and service provision, support evidence-based resource allocation and service planning, in order to prevent veterans reaching crisis point through proactive identification of emerging needs. The platform's aggregated dataset also enables advanced secondary analyses, with research teams already developing machine learning models to identify specific clusters of health and social care needs and AI predictive models for risk factor identification.

Initial Evaluation Results

A three-month prototype evaluation (July-September 2025) with 17 stakeholders revealed strong appreciation for multi-source data integration, geospatial analytical capability, and evidence generation for service planning. However, users identified concerns including missing NHS data contributions, accessibility issues (colour scheme problems for colour vision deficiency, screen reader incompatibility), and unclear terminology affecting non-specialist users.

Impact and Future Direction

MONARCH presents a model for a potential transformative shift from reactive to proactive veteran support, providing the very first unified view of how UK veterans access charity services. The platform supports the NHS 10-year plan's emphasis on early intervention and crisis prevention whilst creating infrastructure for ongoing secondary research to refine understanding of veteran needs and improve service delivery across the entire support ecosystem.

[Access Executive
Summary](#)

[Access Digitised
Interactive Report](#)

List of Tables and Figures

Tables

Table 1. Summary of Thematic Analysis

Table 2. Survey respondent demographics (N = 17)

Table 3. System Usability Scale items with worked example

Table 4. Summary of Thematic Analysis

Figures

Figure 1. Procedure for development of the MONARCH platform.

Figure 2. The opening page of the MONARCH platform

Figure 3. The MONARCH platform summary page

Figure 4. The MONARCH platform census page

Figure 5. The MONARCH platform demographics page

Figure 6. The MONARCH platform needs page

Figure 7. Distribution of System Usability Scale scores

Introduction

Background

There is a substantial body of research on identifying the support needs of UK military veterans (hereinafter referred to as ‘veterans’). These needs can be complex and encompass both physical health conditions (e.g., hypertension⁵, ischaemic heart disease⁶, and type two diabetes⁷); mental health conditions (e.g., post-traumatic stress disorder [PTSD] & depression⁸); and social deprivation (e.g., homelessness & food insecurity^{9,10}). Currently, veterans in the UK receive support through the National Health Service (NHS) and other statutory services in the same way as the general population, supplemented by the Armed Forces charity sector. Both sectors operate a reactive approach to support provision (i.e., intervention is provided for veterans with a self-declared need). However, despite this support infrastructure, research has highlighted the poor help-seeking behaviour from the veteran population, with many only seeking support at the point of crisis^{11,12,13}.

Recent healthcare policy reflects a shift towards a more proactive stance to reduce the instances of crisis, as exemplified in The NHS 10-year plan¹⁴. Examining the veteran population seeking support from the NHS or Armed Forces charity sector can provide insights into key risk factors associated with specific conditions. Previous research has suggested that individuals seeking support from this charity sector are younger/working age and from non-officer ranks who live alone and are recipients of the War Pension Schemes and/or other Armed Forces compensation schemes^{15,16,17}. Identifying such risk factors can enable earlier intervention prior to reaching crisis level and thus, can have significant advantages for both organisational resources and veterans themselves. However, this work is currently in its infancy with considerable scope for improvement.

⁵ Finnegan, A. and Salem, K. (2024). Comparative study of hypertension, diabetes, dementia and smoking in military veterans and non-veterans: a quantitative study using primary healthcare data. *BMJ Mil Health*. <https://doi.org/10.1136/military-2024-002817>

⁶ Bergman, B.P., Mackay, D.F. and Pell, J.P. (2021). Trends in cardiovascular disease in Scottish military veterans: A retrospective cohort study. *BMJ Open*, 11(7). <https://doi.org/10.1136/bmjopen-2020-044465>

⁷ Bergman, B.P., Mackay, D. and Pell, J.P. (2022). Type 2 diabetes in Scottish military veterans: A retrospective cohort study. *BMJ Open*, 12(2). <https://doi.org/10.1136/bmjopen-2021-057431>

⁸ Finnegan, A. and Randles, R. (2023). Prevalence of common mental health disorders in military veterans: using primary healthcare data. *BMJ Mil Health*, 169(6), pp.523-528. <https://doi.org/10.1136/bmjilitary-2021-002045>

⁹ Cusack, M., Montgomery, A.E., Sorrentino, A.E., Dichter, M.E., Chhabra, M. and True, G. (2020). Journey to Home: development of a conceptual model to describe Veterans' experiences with resolving housing instability. *Housing Studies*, 35(2), pp.310-332. <https://doi.org/10.1080/02673037.2019.1598551>

¹⁰ Johnson, A., Serra, G., Tomietto, M. and Kiernan, M.D. (2024). Identifying the risk factors associated with food insecurity in the UK veteran population: a nationwide survey. *Journal of Nutritional Science*, 13, p.63. <https://doi.org/10.1080/02673037.2019.1598551>

¹¹ Hoge, C.W., Castro, C.A., Messer, S.C., McGurk, D., Cotting, D.I. and Koffman, R.L. (2004). Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. *New England Journal of Medicine*, 351(1), pp.13-22. <https://doi.org/10.1056/NEJMoa040603>

¹² Kiernan, M.D., Osbourne, A., McGill, G., Greaves, P. J., Wilson, G. and Hill, M. (2018). Are veterans different? Understanding veterans' help-seeking behaviour for alcohol problems. *Health & Social Care in the Community*, 26(5), pp.725-733. <https://psycnet.apa.org/doi/10.1111/hsc.12585>

¹³ Randles, R. and Finnegan, A. (2022). Veteran help-seeking behaviour for mental health issues: A systematic review. *BMJ Mil Health*, 168(1), pp.99-104. <https://doi.org/10.1136/bmjilitary-2021-001903>

¹⁴ The Department of Health & Social Care (2025). *Fit for the future: 10 Year Health Plan for England – executive summary (accessible version)*. Available at: <https://www.gov.uk/government/publications/10-year-health-plan-for-england-fit-for-the-future/fit-for-the-future-10-year-health-plan-for-england-executive-summary#from-sickness-to-prevention-power-to-make-the-healthy-choice> (Accessed: 16 February 2025).

¹⁵ Kiernan, M.D., Rodrigues, M., Mann, E., Stretesky, P. and Defeyter, M.A. (2022). The Map of Need: Identifying and predicting the spatial distribution of financial hardship in Scotland's veteran community. *BMJ Mil Health*, 168(1), pp.57-63. <https://doi.org/10.1136/bmjilitary-2020-001718>

¹⁶ Serra, G., Tomietto, M., McGill, A. and Kiernan, M. (2025). Developing a nationwide registry of UK veterans seeking help from sector charities: A machine learning approach to stratification. *European Journal of Public Health*, 35(1), pp.5-10. <https://doi.org/10.1093/eurpub/ckae141>

¹⁷ Serra, G., Turolido, F., Tomietto, M., McGill, A. and Kiernan, M.D. (2025). Improving early intervention: identifying risk factors for UK military veterans that access military charities: A case-control study and an AI-powered predictive model. *European Journal of Public Health*, 35(5), pp.867-872. <https://doi.org/10.1093/eurpub/ckaf140>

A wide range of agencies across the UK, including HM Government bodies, health and social care systems, Armed Forces charities, and private-sector stakeholders, collect substantial volumes of data from multiple sources to support the management of operational processes and the delivery of services to veterans. These data points are subsequently used to inform service development and the targeting of specific resources to meet identified needs within individual organisations. However, such a process has been hampered by limited connectivity and linkage between the different systems used by service providers. In addition, the Armed Forces Charity Sector collaborates to support beneficiaries through a process of ‘almonisation’ whereby multiple organisations contribute to the needs of an individual through use of a case management system, such as MOSAIC, sponsored by the Confederation of Service Charities (Cobseo), hosted and managed on behalf of the sector by SSAFA (the Armed Forces charity).

The data challenge caused by almonisation is that one individual, with one need, might be assisted by many organisations through almonisation, who all contribute to that individual’s care. Historically, this has resulted in all of the organisations recording they have helped the individual, resulting in the individual being counted numerous times across different unlinked datasets. Therefore, the absence of effective data linkage across organisations and the process of almonisation presents a significant risk of ‘double counting’ individuals¹⁸ resulting in unreliable datasets and the potential for over-estimation of veteran need on which current and future service provision decisions are based.

Moreover, accurate and reliable data is not always available, nor does it consistently provide sufficient detail to support policy makers in identifying key risk factors. Consequently, these limitations hinder the development of a comprehensive and accurate understanding of the true scale of veteran need and patterns of service access at both regional and national levels. A key challenge identified by the MoN project¹⁹ was the need to develop a process to identify individuals anonymously in order to facilitate secure information-sharing and database linkage, whilst simultaneously safeguarding individual privacy and ensuring that all personal data remained confidential. Moreover, such a system had to support the planning of future interventions and the efficient allocation of resources where needed. In essence, all collected data must offer actionable insights, be continually updated and readily accessible in a dynamic and responsive manner.

Map of Need Aggregated Research (MONARCH) Project

The MONARCH project was developed to deliver all the above aspects and map the Armed Forces charity provision across the UK at a local, regional and national level²⁰, something that had never previously been achieved. Having a resilient and future-proof system is of paramount importance, thus, the overall aim of the MONARCH project was to inform early intervention service provision and reduce instance of health and social care crisis by UK veterans. In doing so, it had two key objectives:

1. The creation of a national, aggregated and dynamic visual data platform of Armed Forces charity data that will provide appropriate information to service providers;

¹⁸ The process where the same veteran receives support from multiple organisations but appears in counts multiple times causing inflated numbers.

¹⁹ The Armed Forces Covenant Fund Trust (no date) *The Map of Need*. Available at: <https://covenantfund.org.uk/the-map-of-need/> (Accessed 16 February 2026).

²⁰ *Ibid* Kiernan et al., (n 12).

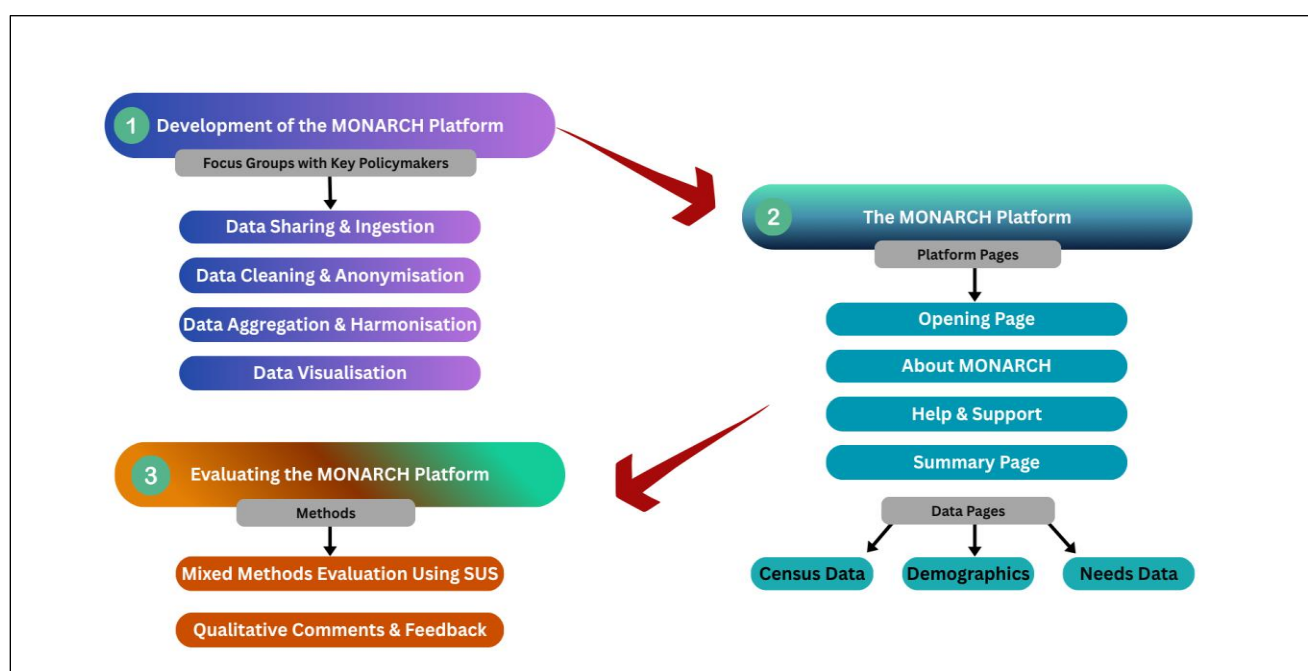
2. The creation of a methodologically robust dataset for scientific investigation, encompassing data design and different coding schemas.

A unified dataset showing how veterans use available services was subsequently created. The collected data was transformed into an interactive dashboard designed to help policymakers, HM Government departments, charities and the NHS better understand veterans' needs and improve service delivery, with the intent of preventing crisis situations arising.

Development of the MONARCH Platform

The MONARCH project was the first UK initiative to consolidate casework data from multiple organisations within the Armed Forces Charity Sector, and present it via a user-friendly, online platform. Full ethical clearance from Health and Life Sciences Ethics Department, Northumbria University was obtained before the project commenced (reference number: 2055). As outlined in Figure 1, the procedure for its initial development was separated into five distinct phases split within three broad categories: (i) development; (ii) the platform itself, and; (iii) evaluation.

Figure 1: Procedure for development of the MONARCH platform.



Due to some overlap of the processes, we have presented these as five separate sections (or phases):

1. Stakeholder consultation
2. Data sharing and ingestion
3. Data cleaning and anonymisation
4. Data aggregation and harmonisation
5. Data visualisation

Each of the five phases are summarised below, however, please refer to Appendices A and B (published papers) for more comprehensive information and explanations.

Phase 1: Focus Groups with Key Policymakers

Four focus groups were conducted between October and December 2023 and included representatives (N=29) from the NHS, various Armed Forces charities, and local government across England and Wales. Participants were selected based on their role in making policy and strategic decisions regarding veteran support. The research team aimed to recruit individuals across diverse occupations and geographic regions, however, despite sustained efforts, focus groups with policymakers from Scotland and UK Central Government could not be organised due to scheduling constraints.

The focus groups aimed to establish: (i) what data were required to inform key policy decisions; and (ii) how this information could be most effectively presented through a data platform. All sessions were audio-recorded and lasted for a mean of 45 minutes (range = 38 – 49 minutes) and were facilitated by a MONARCH team member, with another team member recording key points.

Braun and Clarke's²¹ Thematic Analysis (TA) framework was used to inductively analyse the focus group data. This approach was selected for its flexibility whilst providing a structured and well-established analytical method. The six-stage process comprised of: (i) familiarisation with the data; (ii) generating initial codes from transcripts; (iii) searching for themes; (iv) reviewing themes; (v) defining and naming themes; and (vi) producing the report. The MONARCH team possess extensive experience working with, and researching, the ex-Armed Forces community with several members being veterans themselves. These perspectives informed our interpretation whilst maintaining analytical rigour acknowledging their reflexivity and positioning. Subsequent TA of the focus groups identified three main themes which are outlined in Table 1 and explained in more detail below.

Table 1. Summary of Thematic Analysis

Themes	Sub-themes	Summary
Overview of services and needs	Identifying and evidencing needs Resource planning & service collaboration	Participants emphasised the need to identify veteran populations, evidence complex needs, and support resource planning whilst also preventing crisis
Locations	Geospatial analysis	Geospatial analysis at multiple levels (i.e., local, regional, national) was deemed essential to understand geographic variation in needs and service provision
Dataset	Pre-existing datasets Autonomy and Accessibility Concerns	Participants requested an interactive online platform compatible with existing datasets, providing user autonomy for analysis while including guidance to prevent data misinterpretation and ensure GDPR compliance. Concerns were raised about partial data coverage (e.g. veterans not self-identifying), misinterpretation risks, and the need for controlled access with user guidance.

²¹ Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), pp.77-101.
<https://doi.org/10.1191/1478088706qp063oa>

Theme 1: Overview of Services and Needs

Focus group participants outlined the key challenges of identifying veterans due to organisations outside of the Armed Forces not routinely collecting this data and veterans themselves not identifying with the term. Participants outlined the need to develop a complete picture of the geospatial spread of veterans and their support needs, including understanding the complexity and possible risk factors that may not be apparent at initial contact with services. Such information would provide an evidence base for the allocation of targeted resources at the macro level (i.e., county and national). In particular, the opportunity to predict future need and identify risk factors could be valuable for identifying veterans who have unmet needs and who were not engaging with services. This could enable early intervention (facilitating a preventative approach) which could prevent veterans reaching crisis. However, some participants were concerned about the impact this approach could have on already strained services.

Information on veteran support use (e.g., the number of services available in geographical regions) was seen as valuable for identifying gaps, evidencing resource provision and impact, whilst also avoiding duplication and enabling collaboration with other organisations. Providing this information was seen as enabling services with limited funding to deliver effective services cost-efficiently.

Theme 2: Locations

It was clear that veterans' needs varied across the UK, both nationally and locally, and understanding these core differences could inform tailored resource provision. There was a requirement for the MONARCH platform to include data from multiple organisations across different regions of the UK which could highlight key challenges, available resources and needs unique to specific locations. However, it was noted that any data provided would be valuable. Participants requested different levels of geographical focus depending on their occupation and information needs. For some, the ability to analyse data at a county level, NHS Integrated Care Board (ICB), or by local authority would be useful, whereas others requested only national level data.

Theme 3: Dataset

Participants described using pre-existing databases and tools to help them within their role. The ability to link the MONARCH platform to these databases could enable them to develop a complete picture of veteran needs. As such, an online platform would be most applicable format and included numerous benefits in terms of accessibility and providing up-to-date data.

There were varying opinions about how accessible the platform should be to users. For some participants, the provision of figures and percentages would fulfil their information needs and organisational requirements. Others spoke of needing direct access to the data to analyse independently. Ultimately, participants advised that the platform needed to be interactive and able to be tailored to the user's needs. However, there were notable concerns about the risk of data misinterpretation by those with limited experience or malicious intent which could directly impact service provision and relationships with other organisations. Based on this, participants advised that there should be tight controls over who would have access to the platform, or at the very least, guidance on using and interpreting the data.

Additional challenges included gaining access to focused and informative data, whilst still adhering to data protection requirements and ensuring there was no risk of identification. Finally, there were concerns about the platform providing skewed data as only individuals who identified as veterans were included in the platform.

Summary

In summary, the focus groups highlighted the following points that should be considered when developing the MONARCH platform:

- To outline the number of veterans across the UK and provide a comprehensive overview of their needs (including unmet needs & identifying possible risk factors) at different levels (e.g., nationally, ICB, regionally).
- To predict future need, enabling services to operate a proactive as opposed to a reactive approach to support provision.
- To outline the available services for veterans across the UK to determine whether need has been met and facilitate collaborations.
- The ability to merge, or use in conjunction with, other data platforms to provide an overall picture of veteran need.
- To produce an interactive online platform that can be interrogated based on the user's information requirements.
- To be mindful of potential concerns, such as misinterpretation and biased analysis.

Many of the above points were achieved, or were in progress, at the end of the project, except for providing information on the number of services within geographic locations as this was not within the remit for the project. This objective was achieved by the initial MoN project²² through the Veteran Gateway App, however, those specific outcomes were not further developed under MONARCH.

Phase 2: Data Sharing and Ingestion

Ten Armed Forces charities (SSAFA, Army Benevolent Fund [ABF], Royal Air Forces Association [RAFA], Blind Veterans UK, Combat Stress, Hull 4 Heroes, The Poppy Factory, Royal Marines Association [RMA], the Royal Naval Benevolent Trust [RNBT], and Walking With The Wounded [WWTW]) agreed to contribute data. However, it must be noted that through the process of almonisation, the data from these charities accounted for the assistance given by a further 173 different organisations.

Data was shared through formal data sharing agreements stipulating specific variables, GDPR²³ and UK Data Protection Act compliance, together with secure storage protocols. Legally approved agreements were signed between Northumbria University and each charity prior to moving forward. The aggregated dataset is managed by the Northern Hub for Veterans and Military Families' Research via an SAS Viya interactive portal, with each charity retaining access to their own data and the research team only retained anonymised information.

²² *Ibid* The Armed Forces Covenant Fund Trust (n19)

²³ Data Protection Act (2018), c.12. Available at: <https://www.legislation.gov.uk/ukpga/2018/12/contents> (Accessed: 16 February 2026).

Phase 3: Data Cleaning and Anonymisation

Datasets were initially shared to the project by e-mailing password-protected Microsoft Excel files which were then uploaded into the SAS environment. A portal was later created that allowed each charity to upload their own datasets via a secure, password protected, webpage. Each charity had their own URL and password combination which prevented unwanted or malicious files being imported. Files submitted were deposited directly into the SAS cloud environment to be processed via data cleaning.

To address the issue of ‘double counting’ caused by almonisation, a Secure Hashing Algorithm (SHA)²⁴ was used and applied to each individual case. This process is a cryptographic technique that uses minimal identifiable data to generate a unique 64-item ID for each individual. In the case of MONARCH, gender, date of birth and the military service number were used. Regardless of which dataset the data originates from, if the gender, date of birth and military service number remain constant, the algorithm will always return the same 64-digit ID for that individual, providing an anonymised identifier. This process addressed the challenges of almonisation (and double counting) and allowed us to link data across an infinite number of datasets.

Members of the MONARCH team describe a UK-wide initiative to address fragmented data on veterans’ support services by creating a unified, anonymised dataset from over 1,800 military charities²⁵. The authors explain this process in more detail²⁶ and report how the SHA was tested for efficiency which, initially, was completed with only two demographics (gender & military service number), then later refined to include date of birth as the third demographic to improve accuracy. As outlined in the study at Appendix B²⁷, their results showed that the secure hashing algorithm, and subsequently used in the MONARCH project, identified a significantly greater number of unique cases, avoiding false matches, compared to previous approaches. The secure hashing algorithm successfully linked records across the different years and charity datasets and, through the process of encrypting personal information, securely protected confidential data. For the purposes of platform development and to further enhance security, a ‘salt’ string²⁸ was added to the hashing process to ensure the data could not be reverse engineered at any point.

Data preparation took approximately six months due to variation in charity data collection practices. The process involved removing duplicates and validating the three demographics outlined above. Key challenges included auto-filled birth dates defaulting to 1900, placeholder service numbers (i.e. 123456), inconsistencies in gender reporting across charities, misclassification of family members as female veterans, and incomplete postcode data. Solutions included age-range validation (18-107 years), aggregating genders (male, female, other), IF formulas²⁹ to identify invalid service numbers, communication with charities for variable reclassification, and creation of ‘family member’ and ‘unknown’ categories where appropriate. Logic-based error detection and spot-check quality control were applied throughout.

²⁴ Tomietto, M., McGill, A., & Kiernan, M.D. (2023). Implementing an electronic public health record for policy planning in the UK military sector: Validation of a secure hashing algorithm. *Heliyon*, 9(6). <https://doi.org/10.1016/j.heliyon.2023.e16116>

²⁵ *Ibid* Serra et al., (n 16).

²⁶ See also Appendix B to this report.

²⁷ *Ibid* Serra et al., (n 16).

²⁸ The process by which random data is added to the initial hashing algorithm which acts as a unique security ‘phrase’. Without this ‘phrase’ it is impossible to recreate the hashing, thus ensuring complete encryption and anonymisation at all times.

²⁹ A logical function used to test a condition and then return one value if the condition is true, and another if it is false.

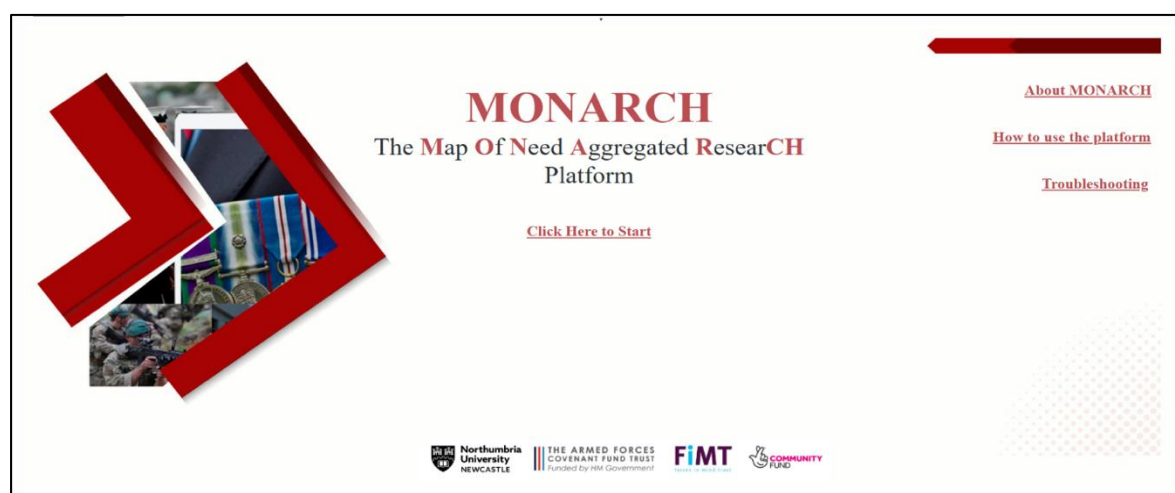
Phase 4: Data Aggregation and Harmonisation

Data aggregation and harmonisation involved linking charity datasets under common elements, despite independent record systems recording variables differently. The classification of needs was completed via 3 levels of aggregation³⁰ (see Appendix B for full published paper). First, a Data Strategy Working Group comprising charity sector stakeholders developed an operational definition of 'need'³¹. Individuals were included only if they received a funded service (excluding signposting) for an identified need verified by the charity. Second, the initial 280 needs recorded across charities were refined through DSWG discussions to 51 distinct needs, mapped to 11 of the 14 All Party Parliamentary Group complex need criteria (e.g., poverty, homelessness, physical/mental health issues, employment problems, learning disability, social isolation, relationship difficulties, substance misuse, domestic violence) and, third, further categorised using WHO's three health elements: (i) physical health; (ii) mental health, and; (iii) social wellbeing.

Phase 5: Data Visualisation

The MONARCH platform was developed in SAS Viya, a cloud-based data visualisation platform chosen for its existing use by MoD and other HM Government departments to facilitate policy-maker adoption. The MONARCH platform presents the user with different pages, including: (i) opening page; (ii) summary page; (iii) the English, Welsh and Scottish census; (iv) demographics (showing age at contact, receipt of benefits, gender, service branch/trade, rank, dependents, marital status, and service affiliation); and (v) needs (which involved mapping data to APPG complex need criteria and WHO health determinants). Users could interactively filter data by selecting map areas and/or specific figures to refine analysis. It was also possible to identify trends and key changes in variables by observing the data across different years, in this sense, it was a live and interpretable platform which could be adapted or filtered to meet the specific user's informational needs. The images of the five main pages can be viewed below:

Figure 2. The opening page of the MONARCH platform



³⁰ *Ibid* Serra et al., (n 16).

³¹ Defined as a verified request for assistance by an individual, for a specific reason, within a given year, where support was provided by the charity – see *ibid* Serra et al., (n 17).

Figure 3. The MONARCH platform summary page

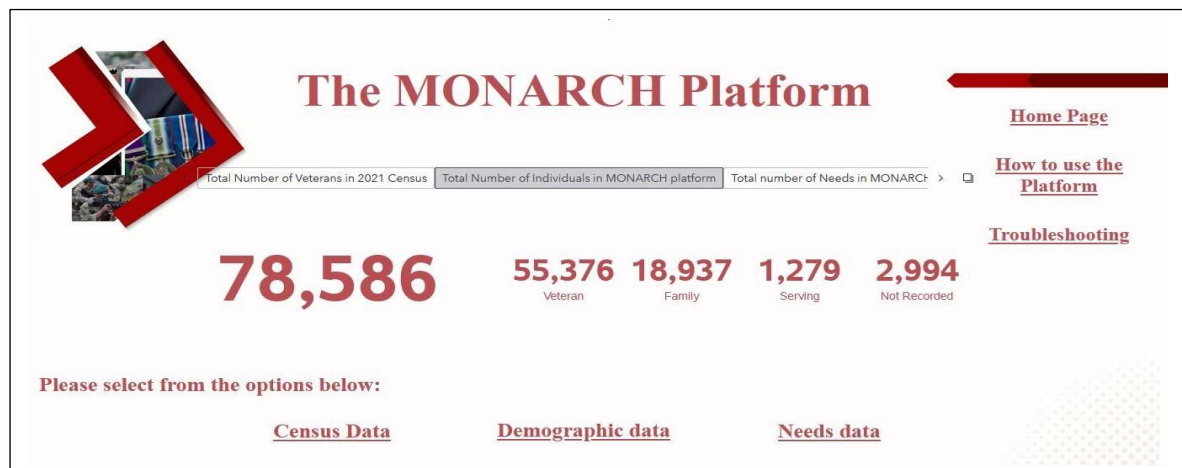


Figure 4. The MONARCH platform census page

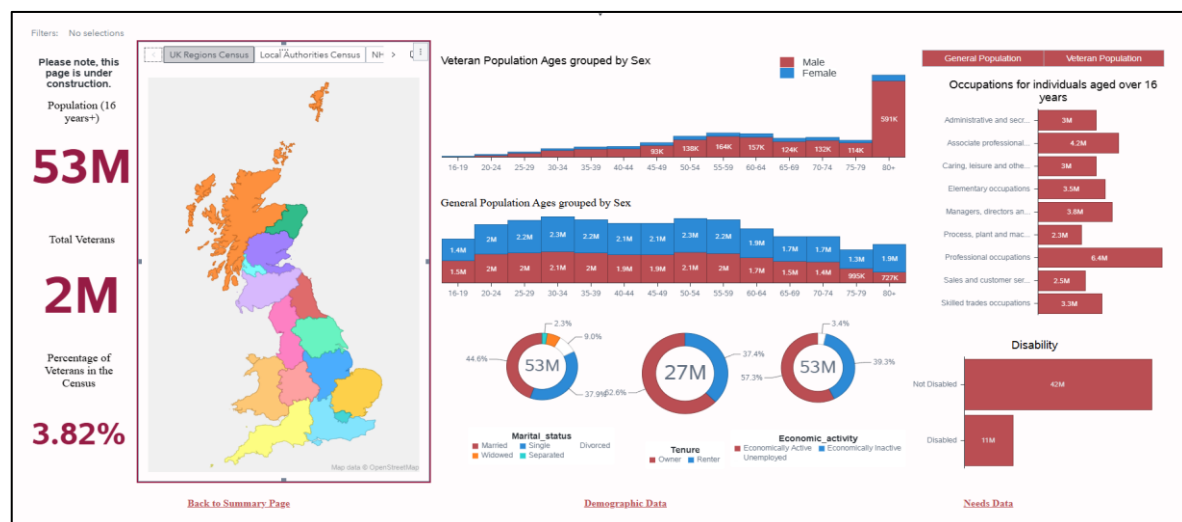


Figure 5. The MONARCH platform demographics page

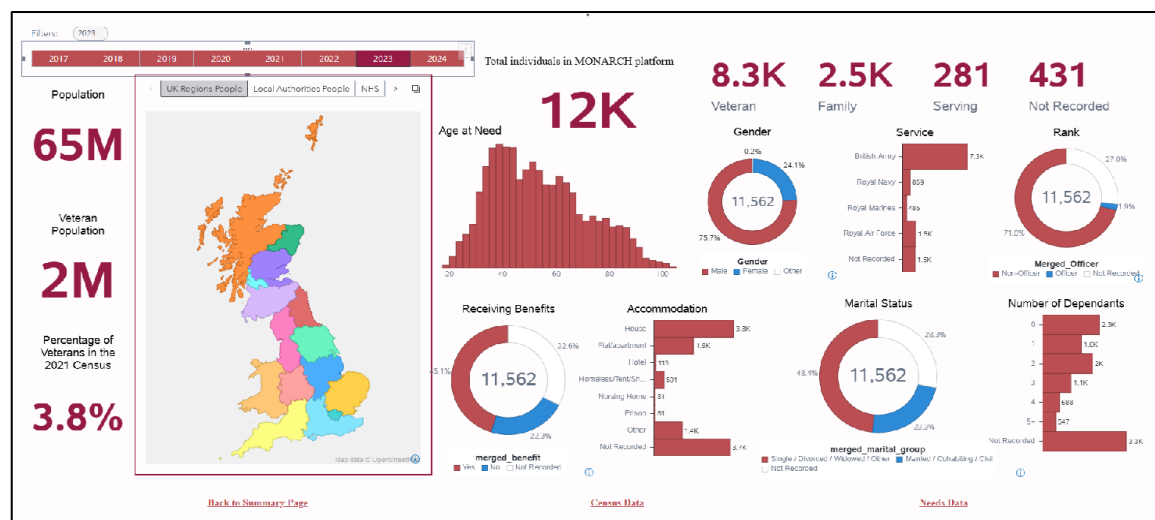
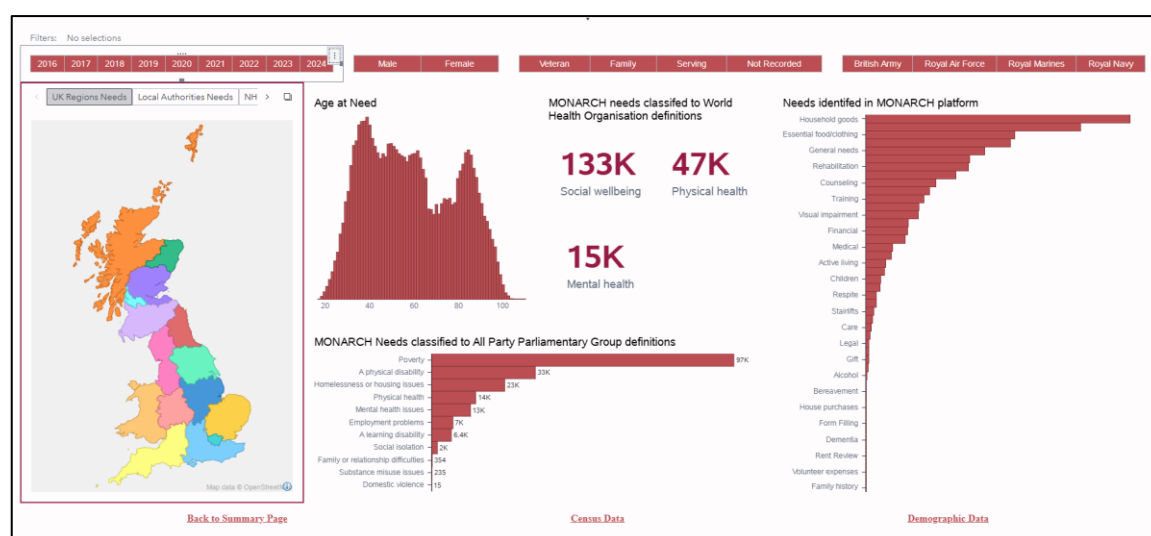


Figure 6. The MONARCH platform needs page



A key consideration was ensuring that the platform was accessible to individuals with health conditions, such as colour vision deficiency or neurodiversity. To this end, members of the research team consulted individuals with expertise in Equality Diversity and Inclusion (EDI) to ensure the platform was as inclusive as possible. Whilst some changes were easily implemented (e.g., colour choice), others were restricted by the platform, such as the inability to use a screen reader with the platform. To ensure ongoing consideration of accessibility, the evaluation of the platform included some focus on this to guide future development.

The operating system used

The MONARCH platform used SAS Viya which operates as a tile-based platform (i.e., figures and content automatically snap to a grid-based layout). Consequently, there were challenges in visualising the data due to the vast amount of information provided on the MONARCH platform, which often required the user to scroll down to view all the figures. Not only could this affect usability, but it could also result in key data being missed. This challenge was overcome using a ‘precision box’ which enabled more flexibility in placing figures, however, this was reliant on manually positioning figures. Similarly, challenges arose in optimising screen-size responsiveness, with the platform layout dynamically adjusting according to display resolution. For example, users accessing the platform on lower-resolution screens (e.g., standard laptops) occasionally encountered overlapping figures or were required to scroll to view visualisations in full. Guidance regarding recommended screen resolutions, together with practical workarounds, was provided to users during tutorial sessions and through a dedicated troubleshooting page embedded within the platform.

Finally, further development challenges included the production of bespoke mapping files to address boundary differences between local authority areas and NHS Integrated Care Boards. This was in addition to the integration of Scottish census data which was released one year later than the corresponding datasets for England and Wales.

Secondary Data Analysis

The MONARCH platform has the capacity to scale beyond previous data management and visualisation functions. The aggregated dataset provides a foundation for

further secondary analyses and the integration of additional analytical features to support policy decision-making. The research team has already created machine learning models for the identification of specific clusters of health and social care needs among veterans (outlined previously)³² and AI predictive models for identifying the risk factors leading veterans to access military charities³³. Please refer to Appendix B for the full published paper.

Evaluation of the MONARCH Platform

The MONARCH platform prototype was disseminated to 66 stakeholders across 38 organisations (incl. Armed Forces charities, local government, NHS, central/devolved governments) for three months (July-September 2025). Each user was invited to attend a brief 30-minute tutorial to introduce them and their colleagues to the platform. In total, five tutorial sessions were held. Everyone who received a link to the MONARCH platform were invited to participate in the evaluation.

A convergent mixed-methods evaluation using the System Usability Scale (SUS)³⁴ and qualitative online survey, hosted on JISC surveys, assessed platform usability. In total, 17 respondents took part (see Table 2 for demographics). A key consideration during the visualisation stage of the platform development was ensuring this was accessible to all users, therefore, all were asked to state any health conditions that may have affected their use of the platform. It is worthy of note that number of respondents we could distribute the link to the dashboard was restricted by our Research and Development license with SAS.

³² *Ibid* Serra et al., (n 16)

³³ *Ibid* Serra et al., (n 17)

³⁴ Brooke, J. (1996). In, Jordan, P.W., Thomas, B., McClelland, I.L., and Weerdmeester, B. (eds.). *Usability evaluation in industry* SUS-A quick and dirty usability scale, pp.4-7. <https://doi.org/10.1201/9781498710411>

Table 2. Survey respondent demographics (N= 17)

Demographic	Response	Respondents
Age	20-30 years	2 (11.76%)
	31-40 years	2 (11.76%)
	41-50 years	4 (23.53%)
	51+ years	7 (41.18%)
	Prefer not to say/did not answer	2 (11.76%)
Gender	Male	9 (52.94%)
	Female	6 (35.29%)
	Prefer not to say/did not answer	2 (11.76%)
Job role	Executive/Director	5 (29.41%)
	Civil Servant	3 (17.65%)
	Research	1 (5.88%)
	Charity Sector	1 (5.88%)
	Management	5 (29.41%)
	Prefer not to say/did not answer	2 (11.76%)
Length of time in role	Below 5 years	7 (41.18%)
	6-10 years	6 (35.29%)
	10+ years	3 (17.65%)
	Prefer not to say/did not answer	1 (5.88%)
Health conditions*	Dyslexia	2 (66.67%)
	Dyspraxia	1 (33.33%)
	ADHD	1 (33.33%)
	Autism	1 (33.33%)
	Colour Vision Deficiency	1 (33.33%)

* Some participants (N=3) reported multiple health conditions.

Quantitative Data

The System Useability Scale (SUS) is a widely used scale to assess the usability of online resources and has high levels of reliability and validity³⁵. Participants rate 10 statements (see table 3) via a Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). To calculate the overall usability score, the data is converted to adjusted scores as follows:

- For each odd-numbered question, 1 is subtracted from the given value.
- The answers from the even numbered questions are subtracted from 5.
- Each of these adjusted scores are added together and then multiplied by 2.5.
- The new values are the usability score which range between 0 and 100 with higher scores indicating better usability.

³⁵ Lewis, J.R. (2018). The system usability scale: past, present, and future. *International Journal of Human-Computer Interaction*, 34(7), pp.577-590. <https://doi.org/10.1080/10447318.2018.1455307>

Table 3. System Usability Scale items with worked example

Statement Number	Statement	Worked Example 1 (Strongly Disagree) to 5 (Strongly Agree)
1	I think I that I would like to use this system frequently	$5 - 1 = 4$
2	I found the system unnecessarily complex.	$2 (5 - 2) = 3$
3	I thought the system was easy to use	$4 - 1 = 3$
4	I think that I would need the support of a technical person to be able to use this system.	$3 (5 - 3) = 2$
5	I found the various functions in this system were well integrated.	$5 - 1 = 4$
6	I thought there was too much inconsistency in this system.	$2 (5 - 2) = 3$
7	I would imagine that most people would learn to use this system very quickly.	$4 - 1 = 3$
8	I found the system very cumbersome to use.	$2 (5 - 2) = 3$
9	I felt very confident using the system.	$4 - 1 = 3$
10	I needed to learn a lot of things before I could get going with the system.	$3 (5 - 3) = 2$



= Odd numbered question from which
1 is subtracted from the total

Sub-total = 30 (x 2.5) = **Total useability score of 75**
which equates to **Acceptable Useability**

The SUS score is categorised using a university grade analogue³⁶ outlined below:

- Acceptable useability - Scores of 70 and above
- High marginal usability - Scores between 62 and 70
- Low marginal usability - Scores between 50 and 62
- Significant usability concerns - Scores below 50

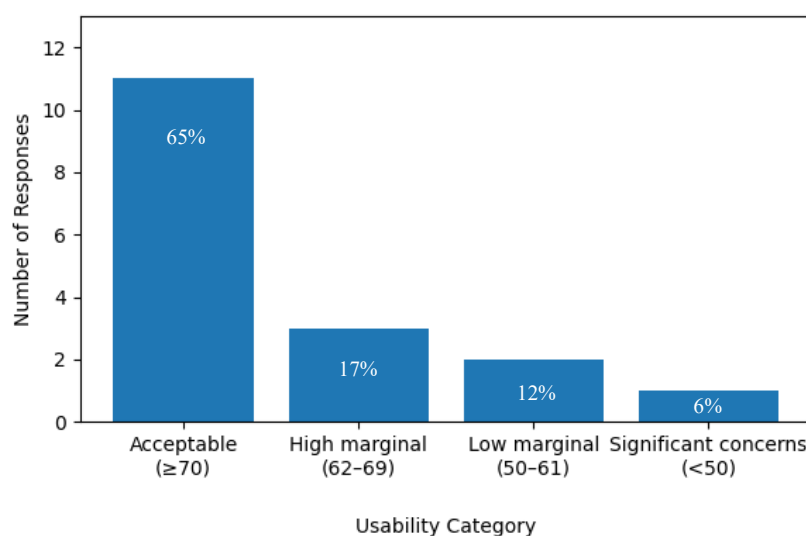
SUS Findings

The SUS scores in the MONARCH project (see Figure 3) ranged from 32.5 - 97.5 (Mean = 74.12, SD = 17.52). Of these scores:

- 11 indicated acceptable usability (65%)
- 3 indicated high marginal usability (17%)
- 2 indicated low marginal usability (12%)
- 1 indicated unacceptable usability (6%)

³⁶ Bangor, A., Kortum, P.T. and Miller, J.T. (2008). An empirical evaluation of the system usability scale. *Intl. Journal of Human-Computer Interaction*, 24(6), pp.574-594. <https://doi.org/10.1080/10447310802205776>

Figure 7: Distribution of System Usability Scale scores



The overwhelming majority of users viewed the MONARCH platform as having acceptable or high marginal usability. Of those who reported low marginal or unacceptable levels of usability, one reported having only seen a short demonstration of the platform, with another highlighting the need for more definitions and making the selections on the filtering system clearer. Both suggested the need for further training and, due to the survey being anonymous, it was not possible to identify whether these users had attended the tutorial sessions offered. It is possible that these specific users were unaware of the definitions (although they were included within hover boxes). One reported being neurodiverse and, despite best efforts to ensure the platform was fully accessible, there were some key challenges encountered. These included presenting a vast amount of information on one page and ensuring it was user friendly for all. Had the project continued, additional focus would have been placed on enhancing accessibility.

Summary

Whilst these findings provide valuable insight into the usability of the MONARCH platform as a concept and future development, it is possible that an increased sample size could have provided a different perspective. Additionally, as previously noted, use of the MONARCH platform was largely needs-led (e.g., to inform funding applications) and it is entirely possible that the platform was not accessed during the three-month reference period.

Qualitative Data

Using TA³⁷ to analyse the qualitative responses, three main themes were identified (see Table 4). For confirmation of the six stages of TA and research team reflexivity, please refer to p.10 of this report.

³⁷ *Ibid* Braun and Clarke (n21)

Table 4. Summary of Thematic Analysis

Theme	Sub-themes	Summary
Credibility	a. Data integrity b. Lack of clarity	Users praised multi-source data integration but raised concerns about missing data (particularly NHS contributions) and the use of unclear terminology/definitions affecting confidence and accessibility for non-specialists
Platform Experience and Functionality	a. Usability & accessibility b. Confidence using the platform	There was positive feedback on quick data access, filtering capabilities, and visual presentation, however, issues included red/white colour scheme impacting users with colour vision deficiency, screen reader incompatibility, information density, and inconsistent cross-page integration
Key Strengths	Analytical capability	The key strengths of the MONARCH platform included analytical capability for geospatial analysis, trend identification, and evidence generation for service planning and funding bids, though debate existed about whether the platform provided a complete picture without integrating additional external datasets.

Theme 1: Credibility

Survey respondents praised the inclusion of multiple data sources on the MONARCH platform, which was seen as providing valuable insight and allowing users to analyse data thereby increasing perceptions of credibility. However, concerns were raised regarding data integrity due to missing data and the lack of contribution from some organisations, such as the NHS, which impacted the perceived trustworthiness of the platform. However, it must be noted that many Armed Forces charities still rely on volunteers to undertake casework and data collection, and throughout the project the amount of missing data did reduce considerably as the project team worked with the organisations through the Data Strategy Working Group. It is also important to acknowledge that the inclusion of NHS data was not a primary aim of this project

The data categories and definitions (or perceived lack of) was seen as needing further clarification, particularly to those who were unfamiliar with the Armed Forces Community and military terminology. Similarly, terms such as ‘not recorded’ could be confusing to those with limited experience using data platforms. These factors could reduce the trustworthiness of the platform and impact the user’s ability to interpret the data, consequently influencing use. Despite this, the platform was praised by survey respondents for its consistency and its links to established and reputable conceptualisations of health (i.e., APPG and WHO). There were instances of participants encouraging other colleagues to access the MONARCH platform either to provide a visual overview or to inform their own work.

Theme 2: Platform Experience and Functionality

The MONARCH platform was praised overall by survey respondents (or their work colleagues), for the ability to access information quickly and directly relevant to their needs. Most found that the platform was consistent, well integrated, easy to navigate, and,

considering the breadth of coverage, not too complex. Others valued the visual elements of the platform. However, there were contrasting opinions regarding the consistency between pages, with one participant describing difficulty extrapolating data. Similarly, whilst some users described the MONARCH platform as clear and straightforward, others found this overwhelming and overloaded with information. On reflection there appeared to be a clear divide between users, based on digital literacy and familiarity with data.

The majority of respondents reported confidence in using the platform, attributing this primarily to prior experience working with data together with familiarity with comparable databases. Several indicated that continued engagement with the platform would likely strengthen this confidence further. However, others identified a need for additional support, particularly in relation to data interpretation. Concerns were raised that users without analytical experience might struggle to draw accurate conclusions from the outputs. Perceptions of the platform's relevance to respondents' professional roles were mixed. Whilst some viewed it as a valuable resource, others did not consider it necessary within their current remit. Among those who anticipated using it, engagement was expected to be needs-driven and intermittent, most commonly in support of discrete tasks such as funding applications as opposed to routine operational use.

A key recommendation from respondents was the colour choice on the platform. The use of red and white made it difficult for some to identify which filter was selected or to read text on the figures. The choice of colours could directly impact the accessibility to those with health conditions, such as colour vision deficiency. Despite the research team's attempt to ensure the accessibility of the MONARCH platform was as inclusive as possible, there were areas of improvement noted. Respondents suggested potential improvements that could be made to the platform including:

- Percentages included with total figures
- Clearer drop-down options
- Guidance in interpreting the data
- The ability to look at the national picture
- The ability to assess trends in data
- Ability to download figures or export data

Theme 3: Key strengths

Overall, the MONARCH platform facilitated focused and comparative analysis across different variables and locations. One such example was the ability to, “...*drill down into real granular detail*” (Survey Respondent 6) and geospatial analysis, which was facilitated using filters on the platform. However, some respondents suggested that including more filters could help enhance the analytical capability.

The platform was also praised for providing evidence for amending services, bid writing, policy planning and, of course, identifying the needs of the veteran population. The ability to access needs data across multiple years provided the opportunity to explore possible trends in the population. In this sense, the MONARCH platform could directly impact service development and provision. However, there appeared to be a debate about whether the platform provided a complete picture of veteran needs and the use of the Armed Forces Charity Sector. Reflecting on this observation, there appears to be a poor understanding of the breadth of almonisation, as the bulk of casework is carried out by a small number of charities

(which were included in this project). The only national charities that declined to take part in the project were the Royal British Legion, The Royal Air Force Benevolent Fund and Help for Heroes, however, a significant proportion of their casework was accounted for through almonisation. Again, it is also worth noting that the primary aim of this project was the aggregation of Armed Forces charity data, whilst developing the capacity to absorb other data from organisations such as the NHS or the Department of Work and Pensions. To integrate data from the statutory sector would be a separate project with its own challenges.

Overview and Reflections

Overview

Platform Development

The MONARCH project represents a landmark achievement in the mapping and understanding of Armed Forces charity provision across the UK. Prior to its development, no unified dataset existed that could provide a comprehensive, dynamic, and geospatially granular picture of veteran need and service access at local, regional, and national levels. By consolidating casework data from 10 Armed Forces charities (which included assistance from a further 173 organisations through almonisation), developing a robust anonymisation and data harmonisation methodology, and presenting aggregated findings through an interactive visualisation platform, MONARCH has addressed a longstanding gap in the evidence-base underpinning veteran support services.

The project demonstrated that effective cross-organisational data linkage is both technically feasible and analytically valuable. The secure hashing algorithm developed and validated by the MONARCH team proved to be more accurate and effective to previous approaches, identifying unique cases with greater accuracy whilst also protecting individual confidentiality. The harmonisation of 280 disparate need categories into 51 distinct needs (mapped to established frameworks including the APPG complex need criteria and WHO health determinants), further ensured that the resulting dataset was not only internally consistent but also theoretically grounded and policy relevant.

Secondary Data Analyses

The secondary data analyses produced from the MONARCH dataset have generated important empirical contributions to the field. The machine learning and predictive modelling work identified clear risk profiles for veterans most likely to experience complex and recurrent needs, specifically younger, non-officer veterans living alone. These findings are not only consistent with but extend beyond previous literature in this area. Such insights provide a robust, evidence-based foundation for the development of proactive and targeted intervention strategies, directly addressing the reactive model of support provision that has historically characterised both NHS and Armed Forces charity sector service delivery.

MONARCH Platform Evaluation

Although the platform evaluation was limited, it provided broadly positive findings. The mean SUS score of 74.12 falls within the range denoting acceptable usability, and qualitative feedback affirmed the platform's analytical capability, its capacity to support evidence generation for funding and policy purposes, and the value of its geospatial

functionality. These findings suggest that the MONARCH platform has strong potential as a practical tool for policymakers, service commissioners, and charity sector stakeholders.

Reflections

Whilst the MONARCH project has achieved considerable progress, a number of important reflections and limitations warrant consideration, both to contextualise the current findings but also to inform any future development of the platform.

Data Coverage and Representativeness

The dataset currently captures only those veterans who have self-identified and sought support from participating Armed Forces charities. As outlined in the focus groups, a substantial proportion of the veteran population does not identify with the term 'veteran' and/or does not engage with formal support services, particularly prior to reaching crisis. Consequently, the platform may systematically underrepresent the most marginalised and hard-to-reach segments of the veteran population which are precisely the individuals who stand to benefit most from early intervention.

The absence of NHS data from the platform was raised as a concern by some evaluation participants. However, whilst this is a valid point, it is a more complex problem and one that requires additional funding to establish further information. It must also be borne in mind that whilst its absence somewhat limits the comprehensiveness of the overall outline of needs presented, it was never the aim of the project to include NHS data. Future development of the platform should prioritise the integration of NHS and wider public sector datasets to move closer to a whole-system view of veteran need.

Geographic Coverage

Despite sustained efforts, focus group representation from Scotland, Northern Ireland (NI) and UK Central Government could not be achieved. Integrating Scottish census data was delayed by one year relative to England and Wales and there are unique challenges with presenting NI data (e.g., the NI census did not ask the 'Veteran Question'). The MONARCH team held discussions with NI stakeholders about how it might include NI, however, the project concluded before this could become a reality.

Whilst these gaps were managed pragmatically, they highlight the structural challenges of cross-jurisdictional data initiatives and the importance of building sustained relationships with devolved government stakeholders from the outset of the project. It is important to note the challenges

Platform Accessibility

Despite the research team's proactive efforts to ensure inclusive design, the evaluation identified that the colour scheme presented difficulties for users with colour vision deficiency, and screen reader incompatibility was also noted. Given that accessibility was an explicit design priority, these findings underscore the importance of involving users with relevant health conditions in iterative co-design processes throughout development, rather than as evaluators at the end. These issues are, however, readily addressable and should be prioritised in any future development of the platform.

Interpretability and Data Literacy

Qualitative feedback consistently highlighted concerns about the risk of data misinterpretation by users unfamiliar with data analysis or military terminology. This tension, between providing open, accessible data and ensuring it is used responsibly, is not unique to MONARCH but is particularly relevant given the potential for misinterpretation which may directly (or indirectly) influence policy and resource allocation decisions. The development of embedded guidance, contextual definitions, and tiered access controls would help mitigate this risk and increase stakeholder confidence in the platform.

Scalability and Sustainability

The MONARCH platform was designed to be resilient and future-proof, and the secondary data analysis work already underway demonstrates its capacity for ongoing scientific contribution. However, sustaining data sharing agreements across multiple organisations over time, ensuring data remains current, and managing the resource demands of continued platform development, will require stable funding and appropriate governance structures. The engagement of a Data Strategy Working Group was a valuable model in this respect, and its continued involvement in overseeing platform development and data standards should be maintained.

Broader Significance

As a methodological model, the MONARCH project has demonstrated that it is possible to aggregate sensitive casework data across independent organisations, protect individual privacy, and produce a platform that is both scientifically rigorous and practically useful to policymakers. In doing so, it offers a transferable framework that could be adapted for other complex populations or service contexts where fragmented data systems present similar barriers to effective evidence-based planning.

Overall Conclusion

There is no doubt that the MONARCH platform represents a significant step forward in the evidence infrastructure supporting UK veteran services. The foundations laid by this project (e.g., technically, analytically, & cross-sector collaboration) provide a strong platform on which to develop further. Addressing the limitations identified here, particularly around data coverage, accessibility, and interpretability, will be essential to realising the platform's full potential as a tool for preventing veteran crisis and improving service delivery at scale.

----- End of Report -----

Appendices

Appendix A – Published paper 1: Tomietto, M., McGill, A., & Kiernan, M.D. (2023). Implementing an electronic public health record for policy planning in the UK military sector: Validation of a secure hashing algorithm. *Heliyon*, 9(6).

Appendix B – Published paper 2: Serra, G., Tomietto, M., McGill, A. and Kiernan, M.D. (2025). Developing a nationwide registry of UK veterans seeking help from sector charities: A machine learning approach to stratification. *European Journal of Public Health*, 35(1), pp.5-10.

Appendix C – Published paper 3: Serra, G., Turollo, F., Tomietto, M., McGill, A. and Kiernan, M.D., 2025. Improving early intervention: identifying risk factors for UK military veterans that access military charities: A case-control study and an AI-powered predictive model. *European Journal of Public Health*, 35(5), pp.867-872.

Heliyon 9 (2023) e16116



Research article

Implementing an electronic public health record for policy planning in the UK military sector: Validation of a secure hashing algorithm

Marco Tomietto^{*}, Andrew McGill, Matthew D. Kiernan

Department of Nursing, Midwifery and Health, Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, United Kingdom

ARTICLE INFO

Keywords:
Secure Hashing Algorithm
Military veterans
Military families
Unique identifier
Big data
Data aggregation
Data confidentiality

ABSTRACT

The digitalisation of healthcare services is a major resource to inform policy-makers. However, the availability of data and the establishment of a data flow present new issues to address, such as data anonymisation, records' reliability, and data linkage. The veterans' population in the UK presents complex needs and many organisations provide social and healthcare support, but their databases are not linked or aggregated to provide a comprehensive overview of service planning. This study aims to test the sensitivity and specificity of a Secure Hashing Algorithm to generate a unique anonymous identifier for data linkage across different organisations in the veterans' population. A Secure Hashing Algorithm was performed by considering two input variables from two different datasets. The uniqueness of the identifier was compared against the single personal key adopted as a current standard identifier. Chi-square, sensitivity, and specificity were calculated. The results demonstrated that the unique identifier generated by the Secure Hashing Algorithm detected more unique records when compared to the current gold standard. The identifier demonstrated optimal sensitivity and specificity and it allowed an enhanced data linkage between different datasets. The adoption of a Secure Hashing Algorithm improved the uniqueness of records. Moreover, it ensured data anonymity by transforming personal information into an encrypted identifier. This approach is beneficial for big data management and for creating an aggregated system for linking different organisations and, in this way, for providing a more comprehensive overview of healthcare provision and the foundation for precision public health strategies.

1. Introduction

The digitalisation of healthcare services is a major resource to inform policy-makers. However, the availability of data and the establishment of a data flow present new issues to address, such as big data management, in order to properly transform data into useful information for healthcare- and social-services planning [1].

The implementation of Electronic Health Records (EHRs) is key to providing quality and cost-effective healthcare provision.

Abbreviations: SHA, Secure Hashing Algorithm; EHR, Electronic Health Record; RNBT, Royal Naval Benevolent Trust; ABF, Army Benevolent Fund.

^{*} Corresponding author.

E-mail addresses: marco.tomietto@northumbria.ac.uk (M. Tomietto), andrew.mcgill@northumbria.ac.uk (A. McGill), matt.kiernan@northumbria.ac.uk (M.D. Kiernan).

<https://doi.org/10.1016/j.heliyon.2023.e16116>

Received 11 April 2023; Received in revised form 7 May 2023; Accepted 7 May 2023

Available online 19 May 2023

2405-8440/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Moreover, it is essential for integrated care and to provide a comprehensive health record system, which follows individuals along their lifespan and along their contacts with different healthcare services and professionals [2,3].

The benefits of an EHR include administrative efficiency, improved access to information by healthcare professionals, patients, researchers, decision-makers, increased patient safety, and costs reduction [3]. Along with these benefits, there are also three major barriers to consider for the implementation of an EHR. At the organisational level (1), the cost of implementation is an initial bottleneck, which affects the healthcare systems worldwide in adopting an EHR. While the benefits over time are evident, the investment for implementation is not always affordable. Also, at the human resource management level (3), there is a need for trained professionals to develop and maintain the EHR [4]. Moreover, it is also necessary to plan the healthcare professionals' education to properly adopt the EHR and to overcome the resistance to change [5]. A third major barrier is related to privacy and security concerns (3). This aspect is strictly related to national and international regulations, and it needs to consider specific solutions to protect individual data [6]. This last barrier is a key issue to consider in future investments in EHRs and in human resources education and training.

1.1. Background and significance

Within the UK, it is reported that there are 2.4 million veterans [7]. Each year in the UK, approximately 14,000 men and women leave the armed forces and move back to civilian life [8] and a proportion of those will experience social and healthcare needs such as mental and physical health issues, homelessness and substances misuse [9].

When considering the veterans' services, data exists in three distinctive areas: government data (1) (such as data within the Department of Work and Pensions, Veterans UK or the Ministry of Justice) National Health Service Data (2) and data held by the Military Charities (3). Military Charities play a significant role in the provision of health and welfare services for former military personnel [10,11]. In 2017, it was reported that there were 278 military charities, 283 service funds (charities that provide financial funds to veterans), 82 military associations, and 65 mixed types of charities that provide support to the armed forces community [12]. Each of these charities within this sector collects data on how veterans use those services, and many veterans who need assistance will engage with one or more of those services over time. This multitude of different data flows that exist within a sector that provides significant care and welfare resources to this population creates an enormously complex, unintelligible picture of how those services are used by the veterans' population.

Despite the large population and the wide network of services and charities addressing the needs of veterans, there is a lack of summary indicators of complexity, and a lack of coordination between different organisations. Each service and charity have their own record system where different data flows currently exist. Harmonising and aggregating these data flows is crucial to develop a clear overview of the complexity of needs and to improving the social and healthcare services at the National level [12].

There is a significant need to harmonize these data flows at a population level to understanding the characteristics of those using services, where those services are being accessed and what needs are being met by those services. A harmonised or aggregated dataset such as this would create a national resource that would provide an evidence base for policy, resource allocation and service planning.

This situation is similar to the current status of epidemiological research in which researchers need to merge different datasets and develop different linkage techniques to create a comprehensive dataset for research and for informing policy-makers.

In recent years, the development of advanced cryptography techniques such as Secure Hashing Algorithms (SHAs) allowed researchers to generate a unique identifier starting from a set of variables, which should identify a unique pattern. In this way, by including a reasonable number of variables to identify a unique pattern, it is possible to generate an anonymous unique identifier across different datasets and merge them [13].

The main point to address and verify is the uniqueness of the identifier generated by the SHA and to determine if the set of variables to generate it are wide and specific enough to ensure this uniqueness.

This approach has the advantage to overcome the privacy and security concerns of an EHR and the potential to ensure a valid and reliable way to merge different datasets or dataflows from different sources over time.

1.2. Aims of this study

This study aims to test a SHA to generate an anonymous unique identifier for merging multiple datasets from the Military Charities Sector in the UK. Specifically, this study aims to:

1. Test the uniqueness, sensitivity and specificity of a unique identifier generated by a SHA;
2. Test data linkage by adopting the SHA-generated unique identifier.

The SHA unique identifier will be tested against the current standard identifier (military service number).

2. Materials and methods

2.1. Study design

This study adopted a retrospective design by considering two datasets from two charities in the UK. One charity provided data from two years (2021 and 2022). In this way, this study also tested both the uniqueness of the SHA-generated unique identifier in each

charity and the data linkage between two different years in the same charity.

2.2. Secure Hashing Algorithm

The SHA is a cryptography hash function ensuring both data integrity and authentication through a digital certificate. A hash function considers an arbitrary input (e.g. a set of variables) and generates a fixed “fingerprint” string. The uniqueness of this fingerprint depends on the input. SHAs are a family of hash functions designed for the abovementioned purpose [14].

In this study, an encryption tool has been adopted and customized to the specific variables and datasets of the Military Charities Sector in the UK. The package for generating the SHA unique identifier is an open-source package available at: <https://github.com/ClinicalLaboratory/Generating-Unique-IDs-from-Patient-identification-Data-Using-Security-Models> implemented in R Studio [15].

2.3. Data sources

Two data sets were used from two large UK Military Charities. In detail, two years of data from the Army Benevolent Fund-The Soldiers Charity (ABF 2021 and ABF 2022) and one data set from the Royal Naval Benevolent Trust (RNBT). The ABF offers support in independent living, elderly care, education and employability, mental fitness, families and housing to those that have served in the British Army. The RNBT provides financial assistance and support to those that have served in the Royal Navy, and their families. As ABF only provides support to those that have served in the British Army and RNBT support to those that have served in the Royal Navy, the SHA should not identify an individual receiving assistance from both charities, as it is expected that veterans will be supported by their single service charity only.

2.4. Statistical analyses

Data preparation was performed in Excel and datasets were converted in .csv format to be imported into R Studio. The SHA unique identifiers were generated by adopting R Studio and the package mentioned above [15].

Datasets were prepared by recording gender as “M” or “F”. Military service numbers have been checked for consistency of the format (e.g. upper or lower cases, sequence of letters and numbers expected). Missing data in one or both variables considered for the SHA generation (gender and military service number) have been deleted. The unique identifiers were generated by using the same set of input variables (gender and military service number) for the SHA.

In step 1, descriptive statistics have been performed to compare the frequency of unique records between the SHA-generated unique identifiers and the current gold standard. Chi-Square test was performed to compare those frequencies ($p < 0.05$). Sensitivity and specificity of the SHA unique identifiers were calculated against the current gold standard (military service number) to detect the SHA ability to correctly identify a unique record (sensitivity) and to correctly reject a duplicate record (specificity). The ROC (Receiver Operator Characteristics) curve was plotted and the area under the curve was calculated. The SHA unique identifiers, sensitivity, specificity, chi-square and the ROC curve were analyzed using SPSS v28 for Windows [16].

In step 2, data linkage between the datasets and their records was compared by considering the two identifiers and by adopting a one-to-one procedure in Stata v13 [17].

Fig. 1 presents the data sources, the steps for data analysis and the main outcomes of this study.

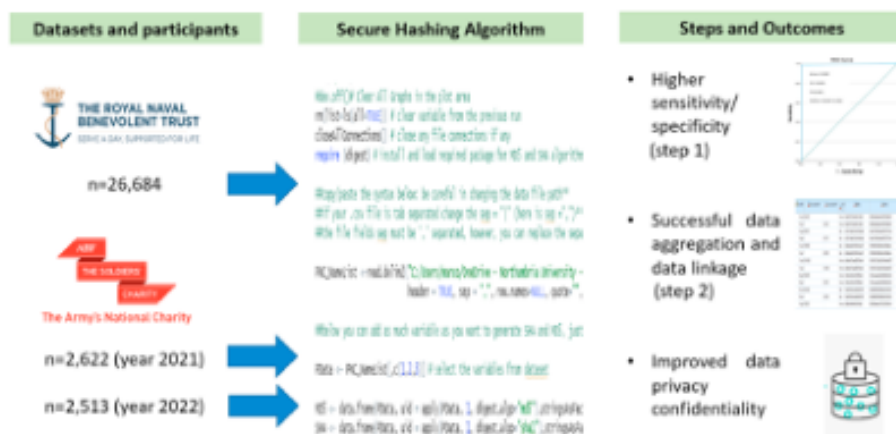


Fig. 1. Data sources, steps for data analysis and main outcomes.

2.5. Ethical considerations

This research complies with the Data Protection Act (2018), which includes the General Data Protection Regulations [18]. Ethical approval has been provided by Northumbria University (reference number: 2055).

3. Results

The dataset from ABF 2021 had 2622 records suitable for SHA generation, the ABF from 2022 dataset 2513 records, and the RNBT dataset 26,684.

In each dataset primary (unique) and duplicate cases were identified by military service numbers and SHA unique identifiers for comparing the rate of unique records detected by the SHA and for calculating sensitivity and specificity.

In the ABF 2021 dataset, 2296/2622 primary cases (87.6%) were identified by military service number and 2306/2622 (87.9%) by SHA unique identifier. The rate of primary cases identified by the SHA algorithm, compared to those identified by the military service number, was statistically significant (chi-square = 2530.549, $p < 0.001$) with a sensitivity of 100% (95%CI = 99.8%–100%) and a specificity of 96.9% (95%CI = 94.4%–98.5%) (Table 1). The area under the ROC curve was 99.8% (95%CI = 99.6–99.9, $p < 0.001$) (Fig. 2).

In the ABF 2022 dataset, 2344/2513 primary cases (93.3%) were identified by military service number and 2353/2513 (93.6%) by SHA unique identifier. The rate of primary cases identified by the SHA algorithm was significantly associated with the standard approach (chi-square = 2370.071, $p < 0.001$) with a sensitivity of 100% (95%CI = 99.8%–100%) and a specificity of 94.7% (95%CI = 90.1%–97.5%) (Table 2). The area under the ROC curve was 99.8% (95%CI = 99.7–100, $p < 0.001$) (Fig. 3).

In the RNBT dataset, 6601/26,684 primary cases (24.7%) were identified by military service number and 6743/26,684 (25.3%) by SHA unique identifier. In total 142 primary cases were additionally identified by the SHA (chi-square = 25947.662, $p < 0.001$), with a sensitivity of 100% (95%CI = 99.9%–100%) and a specificity of 99.3% (95%CI = 99.2%–99.4%) (Table 3). The ROC curve supports the optimal sensitivity and specificity of the SHA-generated unique identifiers compared to the gold standard, the area under the curve was 98.9% (95%CI = 98.7–99.1, $p < 0.001$) (Fig. 4).

A screening of the primary records detected by the SHA compared to those ones detected by the military service number across the two datasets revealed that the same military service number was associated with different individuals (meaning a different pattern of variables). In this way, it is possible to confirm that the identifier generated by the SHA was able to identify more accurately the uniqueness of each individual.

In step 2, data linkage was tested. The RNBT dataset was matched with the ABF 2021 and the ABF 2022 datasets by the SHA unique identifiers: as expected, no match was found between the datasets. Similarly, the ABF 2021 dataset and the ABF 2022 dataset were matched by the SHA unique identifiers, and 1771 records were matched by the SHA unique identifier which would indicate that 1771 individuals were receiving care from the charity over two years. Due to the lower sensitivity of the military service number, it was not possible to match the records from the two ABF datasets by military service number.

4. Discussion

Ensuring the privacy and security of an EHR is a complex issue that includes solutions at different levels: physical, administrative, and technical [19]. At the physical level, it is necessary to focus on network security from the workstations and servers to WiFi radio frequencies and physical access to the devices in the network. From an administrative perspective, it is necessary to establish a risk management system, identify a chain of responsibilities for the system security evaluation, and plan a recovery strategy. The technical level considers users' authentication, virus checking, firewall protection, and data encryption [19]. This paper focused on this last aspect to link different datasets and dataflows in a valid and reliable way and, at the same time, to protect data confidentiality by ensuring data anonymisation. This paper described a procedure to ensure the uniqueness of records' identification by adopting a SHA. This approach demonstrated a better performance in improving the identifiers' uniqueness, compared to the current standard. Moreover, the SHA generated unique identifier allowed the anonymisation of data by deleting the individuals' personal information once the unique identifier is generated. The SHA unique identifier also demonstrated an improved data linkage.

Previous approaches for matching records have shown controversial levels of accuracy [20]. A sufficient number of input variables is crucial to ensure the uniqueness of an identifier and accurate records' matching. The SHA tested in this paper demonstrated an optimal sensitivity and specificity and it also performed better than the previous approach when matching different datasets. Moreover, the sensitivity of the SHA unique identifiers is optimal in the 3 datasets and the specificity improved with the sample size. This

Table 1
ABF 2021 dataset. Cross-tabulation military service number and SHA identifier.

		Military Service Number		Total	
		Primary cases	Duplicate cases		
SHA Identifier	Primary cases	2296 (100%*)	10	2306	$\chi^2 = 2530.549$ $p < 0.001$
	Duplicate cases	0	316 (96.9%**)	316	
	Total	2296	326	2622	

*sensitivity; **specificity.

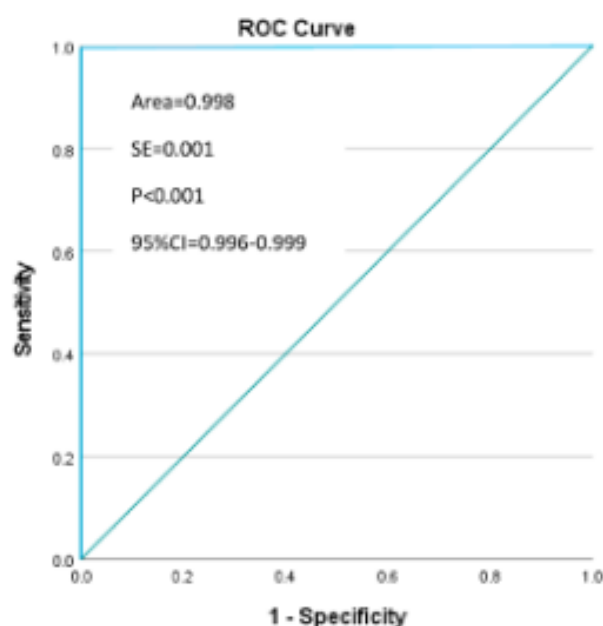


Fig. 2. ABF 2021 dataset. ROC curve of the SHA identifier against the gold standard. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 2

ABF 2022 dataset. Cross-tabulation military service number and SHA identifier.

		Military Service Number			
		Primary cases	Duplicate cases	Total	
SHA identifier	Primary cases	2344 (100%*)	9	2353	$\chi^2 = 2370.071$ $p < 0.001$
	Duplicate cases	0	160 (94.7%**)	160	
	Total	2344	169	2513	

*sensitivity; **specificity.

supports the adoption of the unique identifier generated by the SHA, as it is able to improve the records' uniqueness, data anonymisation and data linkage, especially when managing large datasets and big data [13].

Furthermore, this approach allows data to be aggregated from different organisations. This enables the interactions between veterans and military charities to be observed at the population level, irrespective of which arm of the British military Forces they served in, and in turn, representing their social and health needs. This is a fundamental step to start planning healthcare provision at the National level and addressing the complex needs of this population by preventing poor health and social conditions, promoting healthy behaviours, and early detection of conditions or situations which might impact severely on the service users. The role of big data management in public health and the military veterans space is crucial to allocate the resources to promote healthy populations [21] and the implementation of unique and anonymous identifiers through a SHA is a promising solution to develop reliable data aggregation from different sources. Furthermore, this approach will help develop the implementation of health and social care strategies for veterans, and it will contribute to a more precise analysis of individuals and groups at the population level, leading to improved public health interventions for veterans [22]. While the benefit of public health is clear, it is necessary to consider also the risks related to this approach, as this level of detail is a potential harm for the privacy and security of individuals. However, it is argued that the major advantage of the SHA is that it enables both data aggregation and precision public health, but also ensures data anonymity and confidentiality [19]. The SHA and the generation of a reliable and anonymous unique identifier is the gateway for data aggregation, and, over time, implementing big data in the military veterans' population. This then has the potential to then develop machine learning and artificial intelligence solution for precision public health [23,24].

5. Future implementation

The SHA approach is recognised as the gold standard for generating a unique identifier. It requires good-enough computing

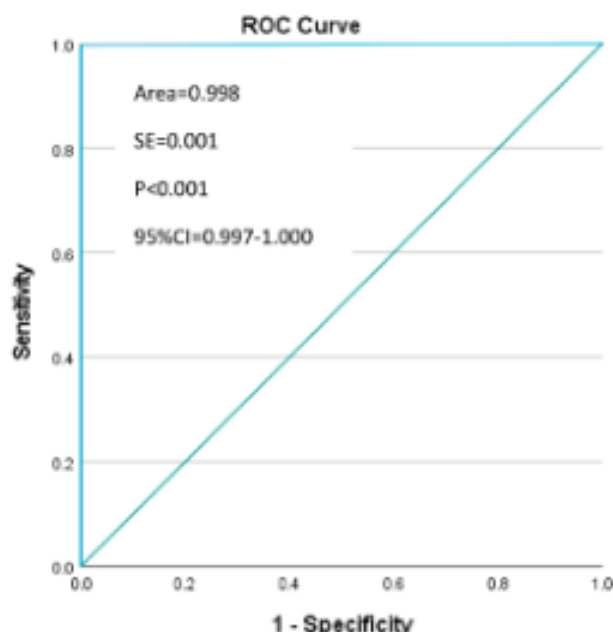


Fig. 3. ABF 2022 dataset. ROC curve of the SHA identifier against the gold standard. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 3

RNBT dataset. Cross-tabulation military service number and SHA identifier.

		Military Service Number		
		Primary cases	Duplicate cases	Total
SHA identifier	Primary cases	6601 (100%*)	142	6743
	Duplicate cases	0	19,941 (99.3%**)	19,941
	Total	6601	20,083	26,684

*sensitivity; **specificity.

capability to be performed, especially for large datasets [13]. If the computing capability is not a reason of concern, this approach is recommended.

In some cases, the SHA can generate the same identifier for different records due to data entry flaws, which could lead to the same input variables for different records. For this reason, data checking and preparation before performing the SHA is essential to ensure the unique identifier's accuracy [25].

In this study, the largest dataset included 26,684 records. While this sample size was adequate for methodologically testing the algorithm, further tests on big data are recommended. Moreover, the SHA has been performed by considering two input variables to generate the unique identifier. While the results are satisfactory, it is recommended to use more variables to better ensure the identifiers' uniqueness.

6. Conclusion

Veterans are a considerable population in the UK and they present with complex needs at the social and public health level. Military Charities are key in addressing this complexity and in providing support to this population. However, they act as separate organisations and different datasets, and information flows exist without being aggregated or providing a comprehensive overview of how those services are used by the veterans' population. This study implemented a SHA for generating a unique identifier to facilitate data aggregation and linkage, so as to aggregate these datasets and support a better healthcare provision at the public health level. This technical solution also addresses data confidentiality and anonymisation by protecting personal data while ensuring accurate data aggregation.

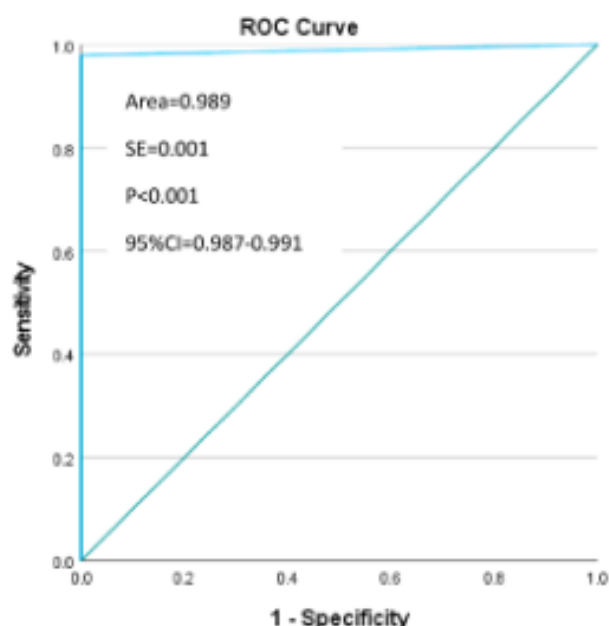


Fig. 4. RNBT dataset, ROC curve of the SHA identifier against the gold standard. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Authors contributions

Marco Tomietto: Wrote the paper, designed the study, analyzed and interpreted data; drafted the article, and critically revised its intellectual content; approved the version submitted. Andrew McGill: Wrote the paper, acquired data, analyzed and interpreted data; drafted the article, and critically revised its intellectual content; approved the version submitted. Matthew D. Kiernan: Wrote the paper, designed the study, interpreted data; drafted the article, and critically revised its intellectual content; approved the version submitted.

Data availability statement

The data that has been used is confidential.

Funding

The study was funded by the Armed Forces Covenant Fund - Map of Need - Ref 3590.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Royal Navy Benevolent Trust and Army Benevolent Fund – The Soldiers Charity.

References

- [1] World Health Organization, Global diffusion of eHealth: making universal health coverage achievable: report of the third global survey on eHealth. <https://www.who.int/publications/i/item/9789241511780>, 2017 accessed on January 2023.
- [2] E. Vayena, J. Duenowagis, J.S. Brownstein, A. Sheikh, Policy implications of big data in the health sector, *Bull. World Health Organ.* 96 (1) (2018) 66–73.
- [3] World Health Organization, Electronic Health Records: Manual for Developing Countries, WHO Regional Office for the Western Pacific, 2006. <https://apps.who.int/iris/handle/10665/207504>, accessed on January 2023.

- [4] L.L. Fragidis, P.D. Chatzoglou, Implementation of a nationwide electronic health record (EHR): the international experience in 13 countries, *Int. J. Health Care Qual. Assur.* 31 (2) (2018) 116–130.
- [5] G. Jimenez, P. Spinazzze, D. Matchar, G.K.C. Huat, R.M. van der Kleij, N.H. Chavannes, J. Car, Digital health competencies for primary healthcare professionals: a scoping review, *Int. J. Med. Inf.* 143 (2020), 104260.
- [6] J.L. Fernández-Alemán, L.C. Señor, P.A.O. Loaysa, A. Tosal, Security and privacy in electronic health records: a systematic literature review, *J. Biomed. Inf.* 46 (3) (2013) 541–562.
- [7] Ministry of Defence, UK Armed Forces Veterans residing in Great Britain, in: Ministry of Defence (Ed.), 2017: Annual Population Survey, 2019. <https://www.gov.uk/government/publications/mod-national-and-official-statistics-by-topic/mod-her-majesty-s-stationary-office>, accessed on January 2023.
- [8] L. Scullion, K. Jones, P. Dwyer, C. Hynes, P. Martin, Military veterans and welfare reform: bridging two policy worlds through qualitative longitudinal research, *Soc. Pol. Soc.* 20 (4) (2021) 670–683.
- [9] K. Gordon, K. Burnell, C. Wilson, Outside the military “bubble”: life after service for UK ex-armed forces personnel, *Front. Public Health* 8 (2020) 50–62.
- [10] R. Gribble, S. Wessely, S. Klein, D.A. Alexander, C. Dandeker, N.T. Fear, Public awareness of UK veterans’ charities, *RUSI J.* 159 (1) (2014) 50–57.
- [11] A. Herman, R. Yarwood, From warfare to welfare: veterans, military charities and the blurred spatiality of post-service welfare in the United Kingdom, *Env. Plan. Econ. Space* 47 (12) (2015) 2628–2644.
- [12] M. Mayhew, The UK Armed Forces Charity Sector: a Summary of Provision, 2018. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/759046/6.5046_MOD_Military_Charity_Report_v3_webNH.PDF, accessed on January 2023.
- [13] E.A. Mohammed, J.C. Slack, C.T. Naugler, Generating unique IDs from patient identification data using security models, *J. Pathol. Inf.* 7 (1) (2016) 55.
- [14] W. Penzer, T. van Werkhoven, On the secure hash algorithm family, *Cryptogr. Cont.* (2008) 1–18.
- [15] RStudio Team, RStudio: Integrated Development for R, RStudio, PBC, Boston, MA, 2020. URL, <http://www.rstudio.com/>.
- [16] IBM Corp, IBM SPSS Statistics for Windows, Version 28.0, IBM Corp, Armonk, NY, 2021.
- [17] StataCorp, Stata Statistical Software: Release 13, StataCorp LP, College Station, TX, 2013.
- [18] General Data Protection Regulation (GDPR), available at: <https://gdpr-info.eu/>, 2018. accessed on November 2022.
- [19] C.S. Kruse, B. Smith, H. Vanderlinden, A. Newland, Security techniques for the electronic health records, *J. Med. Syst.* 41 (8) (2017) 1–9.
- [20] S.B. Johnson, G. Whitney, M. McAuliffe, H. Wang, E. McCreedy, L. Rozenblit, C.C. Evans, Using global unique identifiers to link autism collections, *J. Am. Med. Inf. Assoc.* 17 (6) (2010) 689–695.
- [21] S.J. Mooney, V. Pejver, Big data in public health: terminology, machine learning, and privacy, *Annu. Rev. Publ. Health* 39 (2018) 95–112.
- [22] S. Dolej, Big data’s role in precision public health, *Front. Public Health* 68 (6) (2018) 1–12.
- [23] K. Benke, G. Benke, Artificial intelligence and big data in public health, *Int. J. Environ. Res. Publ. Health* 15 (12) (2018) 2796–2807.
- [24] O. Ota, K. Sedig, The challenge of big data in public health: an opportunity for visual analytics, *Orl. J. Pub. Health Inform.* 5 (3) (2014) 223–254.
- [25] F. Chen, X. Jiang, S. Wang, L.M. Schilling, D. Menker, T. Ong, M.E. Matheny, J.N. Doctor, L. Ohno-Machado, J. Vaidya, Perfectly secure and efficient two-party electronic-health-record linkage, *IEEE Inter. Comput.* 22 (2) (2018) 32–41.

European Journal of Public Health, Vol. 35, No. 1, 5–10
© The Author(s) 2024. Published by Oxford University Press on behalf of the European Public Health Association.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>),
which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
<https://doi.org/10.1093/ejpub/ckae141> Advance Access published on 9 September 2024

Developing a nationwide registry of UK veterans seeking help from sector charities—a machine learning approach to stratification

Giuseppe Serra ^{1,2}, Marco Tomietto^{1,*}, Andrew McGill¹, Matthew Kiernan¹

¹Department of Nursery, Midwifery and Health, Northumbria University, Newcastle upon Tyne, United Kingdom
²Department of Medicine (DMED), University of Udine, Udine, Italy

*Corresponding author. Department Nursing, Midwifery and Health, Faculty of Health and Life Sciences, Coach Lane Campus East Building, Northumbria University, Newcastle upon Tyne, NE7 7TR, United Kingdom. E-mail: marco.tomietto@northumbria.ac.uk.

Abstract

The assistance to veterans in the UK is provided by the National Health Service and over 1800 military charities. These charities count services using different definitions and reporting systems, so to date a national registry of service usage does not exist. The aim of the Map Of Need Aggregation Research study is to build a standardized registry of service usage data for the military charity sector. Data are completely anonymized by adopting a Secure Hashing Algorithm. A unique anonymous identifier is generated allowing both privacy protection and avoiding double counts. Data are standardized and linked with an automated process to create an aggregated dataset. The dataset describes the population, using both *a priori* and machine learning approaches. To date a total of 42 509 veterans with 128 423 needs are included. The mean age was 60.1 years, and 90% were male. 65% were receiving other benefits, 5% were homeless and 1% were in prison. 65% of the needs recorded concerned social wellbeing. 40% of veterans received assistance in at least two different years. The k-means clustering approach returned 4 subgroups of use that were identical to those created using *a priori* knowledge. The dataset is the most comprehensive source of veteran charity usage data in the UK to date. Service usage is generally homogenous among subgroups, but some differences were highlighted indicating that younger, non-officer veterans may be more at risk of presenting with more complex needs. These first useful insights can help allocate resources to build an effective preventive strategy for more complex cases.

Introduction

Veterans as a vulnerable population

There is little information on the prevalence of complex needs in veterans. Those that do experience multifactorial difficulties when they leave the military should be considered as vulnerable, as they generally have interconnected mental, physical, and social health problems [1]. Besides the risks directly connected to service (physical [2], chemical [3], biological [4]) that may cause health consequences over a long period of time, some veterans face numerous challenges upon their return to civilian life. They can suffer from mental health conditions such as PTSD, depression, and anxiety [5, 6], and face many social challenges after discharge, for instance economic hardship, unemployment, homelessness, and food insecurity [7].

These physical, mental, and social health issues often present simultaneously in a single veteran, interacting and exacerbating one another [8]. Some examples of these interactions are unemployment and financial hardship that increase the risk of many mental health conditions [9]; mental health conditions such as depression can increase the risk of hypertension, dyslipidaemia, and diabetes [10, 11]. However, survival studies have demonstrated a longer life expectancy for veterans [12] compared to general population. This effect is well-known as soldiers are pre-selected amongst physically performing individuals and this leads to an initial advantage known as healthy soldier effect (HSE). Interestingly, there is increasing evidence that this effect is eroding in recent years, bringing veteran mortality closer to that of general population [13]. Moreover, while younger veterans have a higher mortality, older veterans keep

showing a clear HSE, and this may be due to a higher prevalence in younger generations of obesity [14] and unhealthy behaviour such as smoking, drinking, and drug use [15].

Assistance for veterans in the UK

Despite the complexity of veterans' needs, it is hard to find comprehensive data in this sector in the United Kingdom (UK) [16]. For health issues, veterans are cared for by the National Health Service (NHS), but there is no reliable national marker to identify their military veteran status [17].

In addition, Military charities in the UK provide a parallel form of assistance, ranging from income support to private doctor appointments or direct mental health support. It is, therefore, clear that studying the health of veterans while relying solely on NHS data provides an incomplete picture for this population and its complexity.

Usage data from the UK Military charities sector is fragmented and recorded in a broad range of ways that make it difficult to use. An additional challenge is that as of 2020, there were 1843 military charities in the UK, ranging from major national charities to small local ones [18]. Many of these charities provide different services as well as supporting veterans vicariously through other charities, creating a risk of over-counting service delivery.

An initial study that investigated the potential to use charity data to explore the wellbeing of veterans was undertaken by the Northern Hub for Veterans and Family Research at Northumbria University as part of the Map of Need project [7]. While it provided some insights about veterans' health in the UK, it covered one major charity and relied on a static dataset [19]. The Map of Need project

also mapped veterans' charity engagement across the UK, but data was limited to users interacting with an app-based signposting platform.

To provide a more comprehensive and self-updating source of data, the Map Of Need Aggregation Research (MONARCH) project was established. The aim of this study was to build an aggregated self-updating dataset, collecting, aggregating, and making accessible charity usage data of veterans nationwide.

The aim of this article is the exploration and description of the MONARCH project data, providing the first comprehensive picture of charity service usage by veterans in the UK.

A secondary objective is to stratify the veterans according to their characteristics, in order to identify possible differences in usage patterns among subgroups.

Methods

Study design

Cross-sectional study.

Study population

The MONARCH project data consists of the veterans who have accessed military charity services between 1 January 2018 and 25 June 2023 (date of last extraction). The dataset included five of the major national charities in the UK, which covers support for all branches of the British Armed Forces. The largest of those organizations undertakes case management for over 128 smaller charities or military charity funding bodies.

Anonymizing data and assigning a unique identifier

The first step of linking the datasets was to assign a unique identifier for the people assisted throughout all organizations, and, for this purpose, a Secure Hash Algorithm (SHA) was used as implemented by Tomietto *et al.* [20]. This approach uses three variables to generate a 64-digit code, namely service number, date of birth, and gender. Hashing algorithms are deterministic, so they always return the same output for a given input, and this makes them vulnerable to dictionary attacks [21]. To further secure anonymization, a 'salt' variable (a password only known to data providers) was added as a fourth input making dictionary decrypting impossible. The identifying variables are then removed from the dataset before it is made available for analysis, leaving only the completely anonymous unique hash identifier. This approach is used by the UK Office for National Statistics for anonymizing personally identifiable information [22].

Measures

Definition of need and aggregation

The statistic unit of this registry was each 'need' captured by the charities, defined as a request, by a given person, for a given reason, in a given year, for which the veteran received assistance.

The types of needs were then classified using 3 layers of aggregation:

- (1) A category of need as provided by the charities
- (2) An aggregation based on the classification of needs used by the reports of the Dual Diagnosis National programme, specifically: family or relationship difficulties, mental health issues, social isolation, substance misuse issues, physical health, a history of offending behaviour, a learning disability, a physical disability, domestic violence, employment problems, homelessness or housing issues, poverty [23]
- (3) A further aggregation based on the three pillars of health (physical, social, and mental) according to the definition of the WHO [24]

Outcome variables

The main outcome variables are the number of needs met, the number of accesses and the repeated request for the same form of assistance. The number of accesses was computed, for each veteran, counting the number of years in which the same subject was helped. The repetition of need was assessed, for each subsequent access, checking if a need was met at least once in previous years.

Exposures

The characteristics of the veterans recorded by the charities were linked and harmonized:

- Age was calculated as the difference between the date of the first need recorded and the date of birth. Age was then classified in 'working age' (≥ 66 years) and 'non-working age' (< 66 years).
- For the variables 'gender', 'rank', 'accommodation', 'marital status', 'service' and 'in receipt of other benefits', the last value to be declared was taken into consideration.
- People recorded both as a veteran and as a family member (i.e. spouse of a veteran which is also a veteran) were classified as a veteran.

Statistical method

The statistical analysis consisted of a description of the population, of service usage and a stratification according to different usage patterns. Categorical variables were reported as absolute frequencies and percentages.

Stratification by outcome occurrence was performed in two subsequent steps: first it was done by *a priori* knowledge, dividing the population in four subgroups based on the ones provided by charities. Then, a K-means technique was used to create clusters. The results of the two approaches were then compared. In this way, it was possible to test the performance of the clusters created with the machine learning approach (K-means) against the stratification made by *a priori* assumptions. In other words, the K-means algorithm was tested to determine its effectiveness in uncovering a previously hypothesized pattern.

For the classification '*a priori*', veterans assisted only in one year were further divided by the number of needs met (single vs multiple). Veterans assisted in two or more years were further divided in a subgroup always presenting with different needs (called 'all new needs') and a subgroup that came back at least once with the same request made in previous years (called 'repeated need'). This led to the formation of four subgroups:

- **Single access, single need.** People that accessed a charity with a single request and did not come back. In this case, it is possible that the problem was simple, and it was solved by the access.
- **Single access, multiple needs.** People that accessed a charity with multiple requests and did not come back. In this case, it is possible that the problem was more complex than the previous scenario, but it was probably solved as the veteran did not come back.
- **Multiple access, repeated needs.** People that accessed a charity with one or more problems and came back at least in one different year with the same issue. In this case, it is possible that the effectiveness of the intervention was not optimal, as the problem kept representing over time.
- **Multiple access, all new needs.** People that accessed a charity with one or more problems and kept coming back in different years with always different requests. In this case, it is possible that the requests of services are a symptom of a deeper problem that may not be properly addressed by service providers.

For the classification based on machine learning, a K-means technique was used to create clusters of use based on: the total number of accesses, the total number of needs and the number of years in

which each single need was met. K-means cluster analysis is an unsupervised machine learning technique that is useful for identifying latent patterns within a sample distribution and providing descriptions of these patterns and subgroups. To determine the ideal number of subgroups, a two-step cluster analysis was conducted, utilizing silhouette measures of cohesion, and employing the Euclidean method [25].

The Chi-Square test was adopted to assess the statistical significance of the differences among subgroups. A P values $< .05$ stated an adequate statistical significance.

The analyses were conducted using SAS Enterprise Guide version 8.2 [26] and SPSS version 28 [27].

Ethical and data management considerations

Data management was structured according to the General Data Protection Regulation (2018) and the UK Data Protection Act (2018). Electronic data were securely stored in a protected folder accessible to the research team. Ethical approval was obtained (ref: 2055, date: 12/01/23). A data agreement was signed and legally approved between the research centre and each charity. Each dataset was then aggregated into an anonymized dataset. The aggregated registry is managed by the research centre and is accessible to each charity. Access to the aggregated data will be possible through an interactive portal in SAS Visual Analytics.

Results

The database to date consists of 42 264 veterans and 113 521 needs. The needs concerned: 74 037 (65.2%) social health, 664 (0.6%) mental health, and 38 820 (34.2%) physical health.

Table 1 shows the characteristics of veterans who accessed charities for various needs. The majority of the veterans were male (88.6%), working age (60.9%), non-officer rank (97.7%), single/divorced/widowed (61.8%), and served in the army (64.4%). The most common type of accommodation that beneficiaries lived in was a house (49.0%), followed by flat or an apartment (27.3%) and 371 veterans (1.0%) reported living in prison. The majority of the veterans were in receipt of other benefits (64.2%) and had one or more dependants (74.0%). Most of the veterans were helped for social well-being (79.2%), roughly one third for a physical health need (35.0%), and some for a mental health need (1.2%). 60.1% did not return for further assistance, while the remaining 39.9% came back in more than one year, suggesting ongoing need. Around half of the veterans with a single access (37% of total) only reported one need, while the rest (23.1% of total) reported multiple needs. Of the veterans that were helped in more than 1 year, a small number presented with all new needs (6.8%) and a third returned for further assistance for needs that they were already helped with (33.1%). Overall, 29.4% of the users only had one need, 27.2% two needs, 15.9% three needs, 9.4% four needs, and 18.2% five or more needs. In summary, around 40% of veterans came back in two or more years, and around 70% presented with more than one need in the study period.

Table 2 compares the characteristics of veterans who accessed charities for various needs by usage pattern. The results show that the following characteristics were significantly associated with the usage pattern: age, rank, marital status, service, accommodation, and receipt of other benefits; the variables gender, number of dependants, type of need, and needs per veteran were not.

In particular, veterans in working age tend to present multiple needs and both single and multiple accesses. Those being married tend to have single access and single or multiple needs. On the contrary, veterans declaring being not married/not co-habiting tend to have multiple accesses and present multiple needs. Moreover, veterans with multiple accesses with multiple or repeated needs have the highest frequency of living in precarious forms of housing (Table 2).

Table 1. Characteristics of veterans accessing charities

	Frequency	Percentage
Gender		
Female	4802	11.4
Male	37 453	88.6
Other	9	0.0
Age		
Non-Working Age	16 540	39.1
Working age	25 724	60.9
Rank		
Non-Officer	29 550	97.7
Officer	709	2.3
Marital status		
Married/Cohabiting/Civil	16 159	38.2
Single/Divorced/Widowed	26 105	61.8
Service		
Army	27 280	64.4
Other	1361	3.2
Royal Air Force	8602	20.4
Royal Marines	860	2.0
Royal Navy	4195	9.9
Accommodation		
Flat/apartment	10 398	27.3
Homeless/Tent/Shelter	2178	5.7
Hotel	578	1.5
House	18 673	49.0
Nursing Home	654	1.7
Other	5233	13.7
Prison	371	1.0
In receipt of other benefits		
No	12 553	35.8
Yes	22 529	64.2
N. of dependants		
No dependants	9194	26.0
1 dependant	6552	18.5
2 dependants	9189	26.0
3+ dependants	10 449	29.5
Type of need		
Helped for physical health need	14 781	35.0
Helped for mental health need	495	1.2
Helped for social well-being	33 467	79.2
Accesses per veteran		
Did not come back (1 need)	12 409	37.0
Did not come back (multiple needs)	13 018	23.1
Came back (All new needs)	2880	6.8
Came back (Repeated needs)	13 957	33.1
Needs per veteran		
1	12 409	29.4
2	11 493	27.2
3	6714	15.9
4	3968	9.4
5 or more	7680	18.2

Table 3 compares the clusters made 'a priori' and the clusters computed with the k-means technique. By applying the silhouette measures of cohesion, the ideal number of clusters was found to be 4, and the clusters computed by the unsupervised approach were exactly correspondent to the ones created 'a priori'. This means that each subject was grouped in the same way by the unsupervised algorithm and by the researchers, with a complete correspondence.

Discussion

Our findings show that veterans accessing charities tend to be younger than the general veteran population (61% in working age vs 47%). Most notably, the gender distribution of military charity usage was similar to the general veteran population (11% female vs 13%) [28] which counters the current narrative in the UK, that women veterans face barriers to accessing services from military charities [29, 30]. This may indicate that younger veterans, irrespective of gender, are more likely to seek help from military charities, and this is consistent with previous research [7].

Table 2. Characteristics of veterans accessing charities by n. of accesses

	Multiple access, all new needs		Multiple access, repeated needs		Single access, single need.		Single access, multiple needs.		P
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage	
Gender									.1343
Female	310	10.8	1662	11.9	1372	11.1	1458	11.3	
Male	2570	89.2	12 294	88.1	11 033	88.9	11 556	88.6	
Age									< .0001
Non-working	734	25.5	6485	46.5	4887	39.4	4434	34.1	
Working age	2146	74.5	7472	53.5	7522	60.6	8584	65.1	
Mean age									
Rank									.041
Non-officer	1930	98.3	9860	97.4	8733	97.6	9042	97.9	
Officer	33	1.7	259	2.6	214	2.4	194	2.1	
Marital status									< .0001
Married/Cohabiting/Civil	1029	35.7	5290	37.9	4952	39.9	4888	37.6	
Single/Divorced/Widowed	1851	64.3	8667	62.1	7457	60.1	8130	62.5	
Service									< .0001
Army	1986	69.0	8460	60.7	8319	67.2	8443	64.9	
Other	84	2.9	453	3.3	406	3.3	418	3.2	
Royal Air Force	484	16.8	3214	23.0	2277	18.4	2627	20.2	
Royal Marines	49	1.7	243	1.7	270	2.2	298	2.3	
Royal Navy	275	9.6	1579	11.3	1116	9.0	1225	9.4	
Accommodation									< .0001
Flat/apartment	743	28.7	3277	26.9	2974	25.8	3404	29.0	
Homeless/Tent/Shelter	177	6.9	612	5.0	624	5.4	765	6.5	
Hotel	48	1.9	192	1.6	141	1.2	197	1.7	
House	1256	48.5	5821	47.7	5896	51.1	5700	48.5	
Nursing Home	24	0.9	294	2.4	198	1.7	138	1.2	
Other	309	11.9	1890	15.5	1593	13.8	1441	12.3	
Prison	34	1.3	108	0.9	114	1.0	115	1.0	
In receipt of other benefits									< .0001
No	774	31.4	3330	33.3	4258	38.9	4191	35.9	
Yes	1689	68.6	6670	66.7	6677	61.1	7493	64.1	
N. of dependants									.9044
No dependants	644	25.9	2599	25.7	2951	27.0	3000	25.3	
1 dependant	459	18.5	1882	18.6	1970	18.0	2241	18.9	
2 dependants	643	25.9	2671	26.4	2795	25.6	3080	26.0	
3+ dependants	738	29.7	2975	29.3	3220	29.4	3516	29.8	
Type of need									
Helped for physical health need	1006	34.9	5527	39.6	3924	31.6	4324	33.2	< .0001
Helped for mental health need	35	1.2	163	1.2	126	1.0	171	1.3	.1769
Helped for social well-being	2412	83.8	10 873	77.9	9657	77.8	10 525	80.9	< .0001
Needs per veteran									< .0001
1	0	0.0	0	0.0	12 409	100.0	0	0.0	
2	1034	35.9	2301	16.5	0	0.0	8158	62.7	
3	914	31.7	2572	18.4	0	0.0	3228	24.8	
4	532	18.5	2303	16.5	0	0.0	1133	8.7	
5 or more	400	13.9	6781	48.6	0	0.0	499	3.8	

The most common type of need presented by veterans was social wellbeing, followed by physical health. Only a small percentage presented with a mental health need, while other studies report much higher prevalence of these conditions [5, 6].

Several characteristics that may indicate that veterans are a vulnerable population were identified. In of detail, the repetition of access, the receipt of other benefits, and the living conditions of some veterans. In particular, the fact that 39.9% of veterans accessed charity services more than once suggests that their needs are not fully met in the first access. Furthermore, 67% of them received other forms of benefits, suggesting that the need for support is often not limited to a simple need, but probably part of a more complex problem [23]. This suggests that the veteran population may have a high level of complexity [31, 32], and this is consistent with previous studies [5], where veterans were found to be resistant to ask for help until the problems become multiple and unmanageable.

In addition, a significant percentage of the veterans who accessed charity services were living in prison or were homeless, indicating a severe form of difficulty. The fact that 25% of people who accessed had dependants (children or spouses) may indicate that the number

of those in need might be even higher than what reported from sector charities.

The stratification 'a priori' based on usage pattern showed statistically significant differences among subgroups. Specifically, the subgroup with a single access and a single need had a higher frequency of being married and a low percentage of other benefits' claim. This is consistent with previous literature [33] and could be attributable to the fact that those with a stronger family support net tend to go less frequently in crisis. It is therefore possible that this subgroup is the one with the lowest complexity, having stronger family support and less need for charitable services.

On the other hand, the subgroup with multiple accesses presenting with all different needs was the one with the lowest age and lowest prevalence of officers. In addition, this group had the highest rates of people living in prison or being homeless and the highest percentage of people asking for other kind of benefits. These characteristics corroborate the assumption that this subgroup is the one with the highest complexity, probably suffering from issues that cannot be solved by a charity service but probably need a more holistic evaluation. In fact, it is possible that younger veterans

Table 3. Clusters by pattern of access

Clusters	1	2	3	4
Came back (All new needs)	0	0	0	2880
Came back (Repeated needs)	13 957	0	0	0
Did not come back (1 need)	0	12 409	0	0
Did not come back (multiple needs)	0	0	13 018	0

with lower ranks have a lower education level [34], which is a recognized social determinant of health [9].

However, the differences among subgroups are modest in absolute terms, indicating that the veteran population who accessed charity services is relatively homogenous. This may suggest that the factors that influence the usage of military charity services could be related to other variables that were not included in our study. Future research should explore other possible determinants and their associations with the usage of charity services by veterans.

Strengths and limitations

One of the main strengths of this study is that it provides the first comprehensive registry of veteran charity usage data in the UK. This is the first study that combines data from multiple charities that offer different types of support to veterans, applying a set of definitions and harmonizing heterogeneous datasets. This approach allowed us to gain a deeper understanding of the characteristics of the veterans who access charity services, and to identify patterns that may not be evident from a single dataset.

Another strength is that for the first time in the UK it includes a high number of veterans who accessed military charity services. This number can be considered reliable, as double counting people who were helped by multiple charities, (a process called 'almonisation', where multiple charities financially contribute to the resolution of a single problem) was eliminated. Prior to this study, each charity involved in an 'almonised' case would record that case as an individual helped, therefore, a person was counted multiple times for the same problem. In the military charity sector, this created a significant inflation of usage rates, since the datasets were not linked with proper caution. This process was made possible by the use of SHA as pilot tested by Tomietto *et al.*, ensuring data anonymity and improving the unique identification of each veteran across different data sources [20]. The aggregated dataset has also the potential to be linked to other large datasets (e.g. from NHS registries) to provide a wider understanding of the health and social care needs of this population.

One of the main limitations is the coverage of the charity sector in the UK. Our study is not yet fully inclusive of the whole military charity sector. Therefore, our results may not reflect the full spectrum of services that are used by veterans in the UK and may be biased towards the types of services that are offered by the charities in our study. In the future, the MONARCH project will expand the coverage of the charity sector and include data from more charities that provide different types of support to veterans.

Another limitation of this study is the data incompleteness. Not all the charities in our study provided data on all the variables that were analysed, and this may constitute a source of bias. To address this issue, feedback was provided to charities sharing data, and a working group on data strategy was set up. It is, therefore, expected that in following years, the quality of data will improve constantly.

A further limitation of this study is the absence of a control group, as we did not compare the veterans who accessed charity services with those who did not. This limits our ability to identify the factors that determine the usage of charity services by veterans. Future research should focus on recruiting an appropriate a control group to conduct a comparative analysis.

Finally, the machine learning approach should be both considered as a strength and a limitation. The 'unsupervised' automated approach returned the same results as the 'a priori' approach, strengthening the validity of the hypothesized pattern. However, the K-means algorithm needs a careful variable selection to provide useful insights, and, if this condition is met, the results are similar to those obtained using an approach 'a priori'.

Conclusion

This study provides a novel and comprehensive analysis of the usage of charity services by veterans in the UK. This work established, for the first time in the sector, a registry of veterans and a methodology for data aggregation and harmonization based on a shared definition of health and social needs. This study provides useful insights to decision-makers and may be used to guide future planning and health and social care provision. Future developments will expand the registry by including a wider consortium of charities and involving the NHS. This will strengthen the study of the determinants of service usage and to uncover the causal pathways leading veterans to a crisis.

Conflict of interest: None declared.

Funding

The MONARCH project was funded by the Armed Forces Covenant Fund Trust (grant reference number 3590) and the Forces in Mind Trust (project reference number FiMT22/1024NUN).

Data availability

The data supporting this study's findings are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Key points

- The MONARCH dataset is the first comprehensive nationwide registry in the UK of military charity data
- Previous issues with this data such as double counting were resolved using an innovative method of anonymization
- The understanding of usage patterns can lead to the design of targeted preventive strategies

References

- 1 Oster C, Morello A, Venning A *et al.* The health and wellbeing needs of veterans: a rapid review. *BMC Psychiatry* 2017;17:414.
- 2 Veterans Benefits Administration. *Compensation: Service-Connected Disability or Death Benefits*. https://www.benefits.va.gov/REPORTS/about/docs/2022-compensation.pdf?trk=public_post_comment-text (May 2024, date last accessed).
- 3 Amiri H, Solaymani-Dodaran M, Mousavi B *et al.* Long-term health outcomes among survivors exposed to sulfur mustard in Iran. *JAMA Netw Open* 2020;3:e2028894.
- 4 Fausde MK, Hoffmann R, Fausde KM *et al.* Malaria re-emergence in northern Afghanistan. *Emerg Infect Dis* 2007;13:1402-4.
- 5 Hoge CW, Castro CA, Messer SC *et al.* Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. *N Engl J Med* 2004;351:13-22.
- 6 Rhead R, MacManus D, Jones M *et al.* Mental health disorders and alcohol misuse among UK military veterans and the general population: a comparison study. *Psychol Med* 2022;52:292-302.
- 7 Kiernan MD, Rodrigues M, Mann E *et al.* The map of need: identifying and predicting the spatial distribution of financial hardship in Scotland's veteran community. *BMJ Mil Health* 2022;168:57-63.

- 8 Iversen A, Nikolau V, Greenberg N et al. What happens to British veterans when they leave the armed forces? *Eur J Public Health* 2005;15(2):175–184.
- 9 Social Determinants of Health. https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1. (May 2024, date last accessed).
- 10 Cohen BE, Marmar C, Ren L et al. Association of cardiovascular risk factors with mental health diagnoses in Iraq and Afghanistan war veterans using VA health care. *JAMA* 2009;302:489–92.
- 11 Boos CJ, Schofield S, Cullinan P, et al. Association between combat-related traumatic injury and cardiovascular risk. *Heart* 2022;108:367–74.
- 12 Hartal M, Kreiss Y, Yavni N. Relative longevity among retired military personnel: a historical-cohort study. *Mil Med Res* 2015;2:29.
- 13 Bollinger MJ, Schmidt S, Pugh JA et al. Erosion of the healthy soldier effect in veterans of US military service in Iraq and Afghanistan. *Popul Health Metr* 2015; 13:8.
- 14 Smith TJ, Marriott BP, Dotson L et al. Overweight and obesity in military personnel: sociodemographic predictors. *Obesity (Silver Spring)* 2012;20:1534–8.
- 15 Bray RM, Hourani LL. Substance use trends among active duty military personnel findings from the United States Department of Defense Health Related Behavior Surveys, 1980–2005. *Addiction* 2007;102:1092–101.
- 16 Moore E, Williams K, Jaynes Z. *United Kingdom Veteran Landscape*. Center for a New American Security, 2020. <https://www.cnas.org/publications/reports/united-kingdom-veteran-landscape>. (May 2024, date last accessed).
- 17 Leighton D, Pernet D, Velupillai S et al. The development of the military service identification tool: identifying military veterans in a clinical research database using natural language processing and machine learning. *JMIR Med Inform* 2020; 8:e15852.
- 18 Cole S, Robson A, Doherty R. *Armed Forces Charities*. www.maproom.net (2020). (May 2024, date last accessed).
- 19 *Annual Report of the Usage of the Veterans' Gateway Online and Mobile Directory of Services*. 2020. https://covenantfund.org.uk/wp-content/uploads/2021/02/2020_Veterans-Gateway-Usage-Report.pdf. (May 2024, date last accessed).
- 20 Tomietto M, McGill A, Kiernan MD. Implementing an electronic public health record for policy planning in the UK military sector: validation of a secure hashing algorithm. *Heliyon* 2023;9:e16116.
- 21 Guendon M, Benzamine E, Gadoche K et al. Securing data linkage in french public statistics. *BMC Med Inform Decis Mak* 2016;16:129.
- 22 Office for National Statistics. Linking with anonymised data—how not to make a hash of it. 2021. <https://www.gov.uk/government/publications/joined-up-data-in-government-the-future-of-data-linking-methods/linking-with-anonymised-data-how-not-to-make-a-hash-of-it>. (May 2024, date last accessed).
- 23 Gerry A, Dodd T. *Dual Diagnosis National Programme Update*. www.dualdiagnosis.co.uk. (2009). (May 2024, date last accessed).
- 24 Constitution of the World Health Organization. (United Nations, 1948). <https://www.who.int/about/governance/constitution>. (May 2024, date last accessed).
- 25 Wu J. *Advances in K-Means Clustering*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. 10.1007/978-3-642-29807-3.
- 26 SAS Institute Inc. 2023. *SAS/STAT® 15.3 User's Guide*. Cary, NC: SAS Institute Inc.
- 27 IBM Corp. Released 2023. *IBM SPSS Statistics for Windows, Version 29.0.2.0*. Armonk, NY: IBM Corp.
- 28 Office for National Statistics. [https://www.ons.gov.uk/peoplepopulationandcommunity/armedforcescommunity/articles/characteristicsofukarmedforcesveteransenglandandwalescensus2021/census2021#:~:text=Of%20the%20total%20\(just%20over,20%20%25%20in%20England\)](https://www.ons.gov.uk/peoplepopulationandcommunity/armedforcescommunity/articles/characteristicsofukarmedforcesveteransenglandandwalescensus2021/census2021#:~:text=Of%20the%20total%20(just%20over,20%20%25%20in%20England)). (2023). (May 2024, date last accessed).
- 29 Godier-McBard LR, Gillin N, Fossey MJ. 'Treat everyone like they're a man': stakeholder perspectives on the provision of health and social care support for female veterans in the UK. *Health Soc Care Commun* 2022;30:e3966–76.
- 30 Bailey M, Phillips R, Morrison Z. Improving access to service charities for female veterans. (2023) <https://rgo-repository.worktribe.com/output/2079145>. (May 2024, date last accessed).
- 31 Faddeeva A, Tiwari A, Mann E et al. A protocol for developing a complex needs indicator for veterans (CNIV) in the UK. *Public Health Pract (Oxf)* 2022;4:100281.
- 32 Faddeeva A, Tomietto M, Tiwari A et al. Validation of a complex needs indicator for veterans in the UK using a composite indicators' method. *Public Health Pract (Oxf)* 2024;7:100464.
- 33 Umberson D, Thomeer MB. Family matters: research on family ties and health, 2010–2020. *J Marriage Fam* 2020;82:404–19.
- 34 Ministry of Defense (MOD). *Officer Entry Requirements*. <https://jobs.army.mod.uk/regular-army/entry-options/officer/>. (May 2024, date last accessed).

European Journal of Public Health, Vol. 35, No. 5, 867–872
© The Author(s) 2025. Published by Oxford University Press on behalf of the European Public Health Association.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>),
which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
<https://doi.org/10.1093/ejpub/ckaf140> Advance Access published on 13 August 2025

Improving early intervention: identifying risk factors for UK military veterans that access military charities—a case-control study and an AI-powered predictive model

Giuseppe Serra^{1,2} , Federico Turoldo³, Marco Tomietto^{1,*}, Andrew McGill¹,
Matthew D. Kiernan¹

¹Department of Nursing, Midwifery and Health, Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, United Kingdom

²Department of Medicine (DMED), University of Udine, Udine, Italy

³Department of Medicine, Surgery and Health Sciences, University of Trieste, Trieste, Italy

*Corresponding author: Prof. Marco Tomietto, Department of Nursing, Midwifery and Health, Faculty of Health and Life Sciences, Coach Lane Campus, Northumbria University, Newcastle upon Tyne, NE7 7TR, United Kingdom. E-mail: marco.tomietto@northumbria.ac.uk.

Abstract

Some veterans face unique physical, mental, and social challenges, leading them to seek assistance from military charities. This case-control study uses data from the MONARCH Study and the tri-service food insecurity study, with the aim to identify key risk factors associated with charity usage among UK veterans. Cases (veterans who accessed charities in 2022) were compared to controls (veterans who did not access charities). Logistic regression and a random forest algorithm were used to identify risk factors for charity use. Several risk factors for charity use were identified: younger age, living alone, being a non-officer, and living in rented accommodation. Having dependents was found to be protective but emerged as a risk factor for veterans living alone and protective for veterans living with others. The use of a random forest algorithm confirmed the statistical importance of these variables, offering deeper insights into complex interactions. These results improve our understanding of the risk factors for charity usage by veterans and provide a predictive model that could be implemented in planning service provision in public health. Additionally, it could be used as the basis for the implementation of targeted preventive interventions, allowing for proactive measures to be taken to support veterans before they reach a point of needing charity services in a period of crisis. These predictive models could enable more efficient resource allocation and the development of tailored strategies to address the specific needs of at-risk veteran subgroups.

Introduction

Complexity of the veteran population

Veterans are recognized as a population that can face increased challenges, and there is evidence that in recent years this vulnerability may be increasing [1]. While historically veterans had a longer life expectancy than the general population, a phenomenon known as healthy soldier effect (HSE), in recent years, this advantage is eroding, possibly indicating a worsening of overall health status [1, 2].

In fact, some veterans may suffer the physical health consequences of risks related to military service, such as the chemical risk from exposure to explosives or gases [3], noise-induced hearing loss (explosions, heavy machinery), or biological risk derived from deployment in areas with a high prevalence of infectious disease. Moreover, the stress of being in combat leads some veterans to develop mental health conditions such as depression or Post-Traumatic Stress Disorder (PTSD) [4]. It has been argued that the worsening of health status may be associated with improved survival rates to injuries that in previous wars would have been un-survivable [5] (this may have caused an increase in veterans living with disabilities), or to the higher prevalence in younger veterans of risk factors such as smoking, drinking, and obesity [6, 7]. Beyond physical and mental health issues, veterans can face social challenges, including unemployment and financial difficulties [8].

Since the vulnerabilities of this population are multiple and affect different aspects of health, the needs of these individuals have been defined as complex [9]. An attempt to measure complexity in veterans has been made by Fadeeva *et al.*, who developed an indicator and validated it using military charity data [10, 11]. The results showed that, while the majority of veterans presented a low level of complexity, a significant minority had high levels.

Assistance to veterans provided by charities

In the UK, veterans' healthcare is provided by the NHS along with the rest of the general population. This is supplemented with over 1800 military charities that provide additional support for physical, mental, and social health issues [12]. Although military charity data are highly fragmented, the Map of Needs project [8] demonstrated significant informative potential. This initiative was further developed through the MONARCH Study [13], which ultimately established a comprehensive national registry of charity usage data from veterans in the UK.

The first analysis of the dataset unveiled the characteristics of veterans accessing charities, indicating they differ from the general veteran population as reported in the 2021 census. Specifically, veterans using charities tend to be younger, non-officers, living in rented accommodation, and often living alone [13]. While these

findings provide valuable insights they are limited to charity users, so the conclusions remain purely hypothetical.

The benefits of early intervention

Research indicates that veterans often refrain from seeking assistance until their problems become severe and unmanageable [14]. Therefore, it can be hypothesized that veterans that refer to a charity have struggled for a certain amount of time and with a certain number of issues before seeking help. As issues that affect veterans tend to influence and worsen one another [13], it is plausible that a single issue can escalate into a series of complex problems, making resolution significantly more challenging.

Consequently, early intervention for veterans is more beneficial and may be more cost-effective than allocating resources to assist veterans in crisis. It is possible that a simple intervention made at an early stage on healthy veterans could prevent a situation where a veteran is forced to seek help from a charity for complex needs.

However, implementing large-scale intervention measures is currently impractical for several reasons. First, there are approximately 2 million veterans in the UK [15, 16], necessitating an intervention of substantial scale. Second, early findings from the MONARCH Study demonstrate that only a small percentage of veterans [13] experience issues that lead them to seek help from military charities. Third, there is no validated intervention proven to prevent the specific physical, mental, and social health issues faced by veterans.

In addition, in recent years, the concept of Precision Public Health has emerged as a response to the increasing availability of extensive data, aiming to tailor public health interventions similarly to advancements in precision medicine. This approach seeks to enhance the effectiveness of public health strategies by using population-specific data to deliver targeted interventions at the right time and place [17–19].

Previous evidence on risk factors for adverse outcomes

An alternative method of implementing an intervention would be to identify a subgroup with higher risk, to focus on veterans that could benefit the most. However, the risk factors associated with military charity usage remain unexplored, making it challenging to pinpoint a target population.

Johnson *et al.* [20] attempted to identify risk factors for adverse outcomes through a cross-sectional study assessing the prevalence of food insecurity among UK veterans. Their findings indicated that younger, non-officer veterans residing in rented accommodations were at higher risk compared to older, officer veterans owning their own homes. Nonetheless, the cross-sectional nature of the study did not allow for the determination of a definitive causal relationship.

Objective of the study

The objective of this study is to determine the characteristics of military veterans which are risk factors for military charity usage, by comparing a group of veterans that accessed a charity (cases) and a group of veterans that did not (controls).

Methods

Study design

Case-control study.

Setting and participants

Cases are selected by simple random sampling [21] among veterans who received services from at least one of the military charities included in the MONARCH dataset [13] between 1st of January 2022 and 31st of December 2022. The year 2022 was chosen because it was the most recent year with complete data from all charities

included in the database. The selection was performed using the 'sample' function of basic R [22] (version 4.3.1), without replacement and using the seed 1234 to ensure reproducibility.

Controls are the veterans who were interviewed during the tri-service food insecurity study [20], a nationwide survey for which invitations were sent to all members of multiple veteran associations. Controls were selected only among those that reported never accessing a military charity for assistance.

The dataset of cases (15 007 records of benefits) was much larger than the dataset of controls (881 records). For this reason, only a sample of cases was used to maintain a 1:1 ratio between cases and controls. A flow diagram of cases and controls sampling and inclusion in the study is provided in [Supplementary Fig. S1](#).

Variables and data sources

The outcome of interest is a military veteran receiving at least one service from one of the military charities included in the MONARCH dataset in the study period. The source data come directly from the charities, where a means tested assessment accurately records each service provided for financial reporting purposes.

The exposures of interest are the socio-demographic characteristics and the service history of the veterans. For cases, this information is recorded by the charities before they provide the service. For controls, the information is gathered through the questionnaire used for the tri-service food insecurity study [20].

Quantitative variables and groupings

The variables are defined as follows:

- Age. For cases, it is the difference in days between 1st of January 2022 and the date of birth, divided by 365.25 and rounded down. For controls, it is the rounded down age, in years, declared during of the data collection period (01 February 2023–31 March 2023). Age is categorized into 10-year intervals and further divided into working age (<66 years) and non-working age (≥66 years).
- Gender, rank, marital status, accommodation, and number of dependents are recorded by the charities for cases, while for controls, this information is gathered through a questionnaire.
- Only 'male' and 'female' genders were considered, with other declared genders excluded prior to sampling due to differing definitions in the two datasets.
- Rank was coded according to standard NATO definitions and categorized as 'officer' or 'non-officer'.
- Marital status was classified into 'married', 'cohabiting', 'in a civil relationship', 'single', 'divorced', and 'widowed'. The categories 'married', 'cohabiting', and 'in a civil relationship' were combined and used as a proxy for living with someone, while 'single', 'divorced', and 'widowed' were grouped and used as a proxy for living alone.
- Accommodation was categorized as 'owner' or 'rented', irrespective of the accommodation type (house, apartment, etc.).
- Dependents were classified based on the number of declared dependents. The variable was further grouped into 'with dependents' and 'without dependents'.

Statistical analysis

Data were analysed using R software 4.3.1 [22].

Descriptive statistics were used to describe demographic data, socio-economic status, and other characteristics using the 'gtsummary' and 'tableone' package [23]. Categorical variables were presented as frequency, and continuous variables were presented as mean and standard deviation. Default tests were the Wilcoxon rank sum test for continuous variables, Pearson's Chi-squared test without Yates' correction for categorical variables with all expected cell counts ≥ 5, and Fisher's exact test for categorical variables with any expected cell count < 5. Missing data were

handled by listwise deletion. The statistical significance was set to < 0.05 . Standardized mean difference (SMD) was also used to assess differences between subgroups, given the high numerosity of the sample. A value of $\text{SMD} > 0.1$ was considered significant.

Univariate and multivariate logistic regression were used to calculate odds ratios and 95% confidence intervals.

The multivariate model was built with variables that were significant in the univariate models, and consistent with similar models in available literature [20]. The variable 'accommodation' was removed due to the high number of missing values.

As part of the sensitivity analysis, five additional random samples were generated. Then, the models were compared to the results of the initial extraction to assess consistency.

As a secondary analysis, a random forest classification algorithm was fitted to the dataset to extrapolate the variable importance of each independent variable, and then compare the result with the multivariate logistic regression and explore any differences. The study dataset was used as a training set, while the test set consisted of 6864 complete cases from the MONARCH dataset that were not used in the training process.

The direction of the association was evaluated by comparing the predicted probability of being a case in different sub-categories. The strength of the association was evaluated using the Boruta algorithm [24], and the estimates of variable importance were expressed as mean decrease in accuracy. Multiple R packages were used for this secondary analysis [25–28].

To further explore the possible relationship between the number of dependents and living condition, a post-hoc analysis was performed, calculating a stratified logistic regression model by living condition.

Ethical considerations

Data management was conducted in compliance with the General Data Protection Regulation (GDPR) (2018) and the UK Data Protection Act (2018). Ethical approval was obtained for both data sources [13, 20]. All personal information was removed from the datasets before making them available to the analysts.

Results

Table 1 shows frequency distribution and characteristics of cases and controls.

Of the 838 cases (veterans who accessed a charity in 2022), 75% were male and 60% were working age. The most represented age groups were >80 (23%), 30–39 (17%), and 50–59 (15%). Most veterans were living alone (65%). Seventy-five percent had dependents and 51% lived in rented accommodation. The most represented service was Army (61%) followed by Royal Air Force (21%) and Royal Navy (12%). Ninety-eight percent 98% of cases were non-officers.

Of the 838 controls (veterans who declared not accessing a charity in 2022), 84% were male and 41% were in working age. The most represented age groups were 50–59 (17%), 60–69 (37%), and 70–79 (30%). Most veterans were living with others (76%), 81% had dependents and 85% were homeowners. The most represented service was Army (50%) followed by Royal Navy (34%) and Royal Air Force (16%). Seventy-eight percent of controls were non-officers.

All differences assessed were statistically significant.

Table 2 shows the results of univariate logistic regression predicting being a case. Results show that female gender was a risk factor for charity usage (OR 1.72; 95% CI 1.35–2.20), together with working

Table 1. Frequency distribution and characteristics of cases and controls included in the study

Characteristic	Total <i>N</i> = 1676 ^a	Controls <i>N</i> = 838 ^a	Cases <i>N</i> = 838 ^a	SMD ^b	SMD ^b * > 0.1	<i>P</i> -value ^c
Gender				0.218	*	<0.001
Female	341 (20%)	134 (16%)	207 (25%)			
Male	1335 (80%)	704 (84%)	631 (75%)			
Age (grouped)				1.127	*	<0.001
18–29	42 (3%)	0 (0%)	42 (5%)			
30–39	148 (9%)	8 (1%)	140 (17%)			
40–49	155 (9%)	38 (5%)	117 (14%)			
50–59	270 (16%)	143 (17%)	127 (15%)			
60–69	417 (25%)	300 (37%)	117 (14%)			
70–79	346 (21%)	246 (30%)	100 (12%)			
≥ 80	281 (17%)	86 (10%)	195 (23%)			
Age (binary)				0.376	*	<0.001
< 66	842 (51%)	340 (41%)	502 (60%)			
≥ 66	817 (49%)	481 (59%)	336 (40%)			
Living condition				0.920	*	<0.001
Living with others	881 (57%)	628 (76%)	253 (35%)			
Living alone	676 (43%)	197 (24%)	479 (65%)			
Dependents				0.129	*	<0.001
0	310 (22%)	155 (19%)	155 (25%)			
≥ 1	1126 (78%)	650 (81%)	476 (75%)			
Accommodation				0.777	*	<0.001
Owner	721 (80%)	678 (84%)	43 (49%)			
Rented	177 (20%)	133 (16%)	44 (51%)			
Service				0.637	*	<0.001
Army	890 (56%)	419 (50%)	471 (61%)			
Royal Air Force	300 (19%)	138 (16%)	162 (21%)			
Royal Marines	36 (2%)	0 (0%)	36 (5%)			
Royal Navy	372 (23%)	281 (34%)	91 (12%)			
Rank				0.626	*	<0.001
Non-officer	1244 (87%)	603 (78%)	641 (98%)			
Officer	186 (13%)	170 (22%)	16 (2%)			

a: *n* (%); missing values are not shown.

b: SMD: standardized mean difference.

c: Pearson's Chi-squared test; Fisher's exact test.

Table 2. Univariate logistic regression predicting being a case

Characteristic	OR	(95% CI)
Gender		
M	Reference	
F	1.72	(1.35–2.20)
Age (binary)		
≥ 66	Reference	
< 66	2.11	(1.74–2.57)
Living condition		
Living with others	Reference	
Living alone	6.03	(4.85–7.54)
Accommodation		
Owner	Reference	
Rented	5.22	(3.29–8.28)
Rank		
Officer	Reference	
Non-officer	11.29	(6.90–19.83)
Service		
Army	Reference	
Royal Air Force	1.04	(0.80–1.28)
Royal Marines	>999	(0.11 – >999)
Royal Navy	0.28	(0.22–0.38)
Dependents		
0	Reference	
≥ 1	0.73	(0.57–0.94)

age (OR 2.11; 95% CI 1.74–2.57), living alone (OR 6.03; 95% CI 4.85–7.54), living in rented accommodation (OR 5.22; 95% CI 3.29–8.28), and not being an officer (OR 11.29; 95% CI 6.90–19.83). Having at least one dependent was protective (OR 0.73; 95% CI 0.57–0.94), while the service was not significantly associated with being a case or a control.

Table 3 shows the results of multivariate logistic regression predicting being a case. Results show that female gender was a risk factor for charity usage, but with more uncertain estimates than in the univariate models (OR 1.44; 95% CI 0.99–2.09). Being in working age (OR 2.99; 95% CI 2.21–4.07), a non-officer (OR 17.53; 95% CI 7.63–51.0) and living alone (OR 11.95; 95% CI 8.25–17.6) stayed associated with the risk of accessing a charity. Having dependents, which was a protective factor in the univariate analysis, became a risk factor (OR 3.73; 95% CI 2.44–5.77).

A random forest model fitted to the final dataset and showed that all variables exhibited significant feature importance in predicting the probability of accessing a charity, as confirmed by the Boruta algorithm (Supplementary Fig. S2). Living condition was identified as the most important predictor (mean decrease in accuracy 30.22). To explore the direction of these associations, partial dependence plots (PDPs) were generated for each predictor, revealing the average predicted probabilities of being classified as a case if all data points assume that predictor value [26]. Results are shown as average predicted probability in Table 3. All variables were found to increase the probability of accessing a charity, accordingly to the multivariate logistic regression, since the predicted probability of being a case of the reference category was lower than the predicted probability of being a case of the category studied. On the contrary, having ≥1 dependents decreased the probability (0.34 vs 0.46) and can therefore be considered a protective factor. The accuracy on the training set was 81.8%, the out-of-bag estimate of error rate was 18.4%, and the accuracy on the test set was 65.8%.

In the post-hoc analysis, logistic regression revealed the role of living condition as a qualitative multiplicative effect measure modifier (Table 4). Specifically, for veterans living with others, having dependents did not correlate with an increased probability of accessing charity services. However, for veterans living alone, the number of dependents emerged as a significant and strong predictor of charity access.

As for the sensitivity analysis, the tables created using different random samples produced results that were comparable to those obtained with the original one. The full results, including all

variations and similarities, are presented in Supplementary Tables S1 and S2.

Discussion

The findings reveal that younger, non-officer veterans living alone are at a substantially higher risk of accessing charity services compared to their older, officer counterparts living with others. This observation aligns with existing literature [13, 20].

The use of multivariate logistic regression allowed for the isolation of the individual contributions of various risk factors to the overall likelihood of charity usage. All factors remained significant except for gender, whose role remains ambiguous. Interestingly, having dependents was found to have a protective effect. Further stratification of the population by living condition revealed that this risk factor is only significant among veterans living alone.

Living condition acted as a confounder in the univariate analysis, as it was associated both with the predictor and the outcome. In fact, living with others is associated both with having dependents (i.e. children living at home) and with having less risk of being in crisis, as shown by previous studies [13]. This suggests that living condition plays a crucial role in how dependents influence the need for charity support among veterans. While veterans living with others may have additional support systems that mitigate the need for charity, veterans living alone with dependents face greater challenges, making them more likely to seek charitable assistance.

Additionally, the random forest algorithm confirmed these results and demonstrates that in this context it was a robust alternative to logistic regression. However, while the direction of the association is the same as in logistic regression, the rank of variable importance is not correspondent to the rank of the odds ratios. This can be explained by the fact that both models have a certain level of uncertainty, so the results should be evaluated jointly.

These results confirm once more the necessity of critical thinking in epidemiology and public health research. While statistics provide a framework for understanding relationships between variables, it is critical that these findings are interpreted with care. Without critical reasoning, there is a significant risk of oversimplifying complex interactions or misattributing causality. For example, in this study, the stratification by living condition allowed for the identification of the protective effect of dependents, which is particularly pronounced among veterans living with others, while for veterans living alone, it is a strong risk factor. This reinforces the notion that a one-size-fits-all approach in public health can be misleading.

The use of accurate and targeted data to improve public health, as opposed to a generalized approach, has been defined as 'Precision Public Health', aiming to provide 'the right intervention to the right population at the right time'. This approach is considered by some to be a modernization of traditional epidemiology, leveraging advanced technologies and data analysis to achieve its goals [17–19]. In this study, the use of both logistic regression and random forest algorithms, together with the targeted subgroup analysis, constitutes a practical implementation of precision public health, showing how veteran charity usage data can lead ultimately to precisely targeted interventions.

Strengths and limitations

One of the primary strengths of this study is the use of real-world veteran data from the MONARCH Study, which provides the most comprehensive representation available of the veteran population accessing charities in the UK.

Additionally, this study employs both classic epidemiologic measures and innovative machine learning algorithms. One of the key advantages of the random forest algorithm is its reduced sensitivity to collinearity among predictor variables, which can be a significant issue in logistic regression. This combined methodology allows the

Table 3. Multivariate logistic regression and random forest algorithm predicting being a case

Characteristic	Multivariate logistic regression		Random forest ^a	
	OR	(95% CI)	Averaged predicted probability of being a case ^b	Variable importance ^c
Gender				4.17
M	Reference		0.26	
F	1.44	(0.99–2.09)	0.33	
Age (binary)				11.90
≥ 66	Reference		0.19	
< 66	2.99	(2.21–4.07)	0.35	
Rank				13.86
Officer	Reference		0.11	
Non-officer	17.53	(7.63–51.0)	0.32	
Living condition				30.22
Living with others	Reference		0.13	
Living alone	11.95	(8.25–17.6)	0.70	
Dependents				18.35
0	Reference		0.46	
≥ 1	3.73	(2.44–5.77)	0.34	

a: Number of trees = 100. Out-of-bag estimate of error rate: 18.4%. Seed = 1234.

b: Averaged predicted probabilities of being a case, as resulting from building a partial dependencies plot for the specified predictor or variable.

c: Mean decrease in accuracy was used as variable importance metrics. The higher the value, the higher the variable importance in the model.

Table 4. Multivariate logistic regression predicting being a case by living condition

Characteristic	Multivariate logistic regression		Multivariate logistic regression	
	Among veterans living with others		Living veterans living alone	
	OR	(95% CI)	OR	(95% CI)
Gender				
M	Reference		Reference	
F	1.46	(0.82–2.60)	1.41	(0.80–2.47)
Age (binary)				
≥ 66	Reference		Reference	
< 66	2.02	(1.34–3.06)	5.03	(3.00–8.43)
Rank				
Officer	Reference		Reference	
Non-officer	20.6	(4.96–85.2)	17.7	(4.54–69.1)
Dependents				
0	Reference		Reference	
≥ 1	0.09	(0.04–0.25)	9.20	(5.51–15.4)

cross-validation of the results, leading to a more comprehensive understanding of phenomena than relying on a single method alone.

Another significant strength of this study is the development of a predictive model that, with further testing, could serve as a practical tool for forecasting charity usage.

However, there are several limitations to consider. The quality of the data used in the study may be variable, since cases and controls were collected in two different settings, and this could affect the accuracy of the findings. In addition, inconsistent or incomplete data entries could introduce bias or errors into the analysis.

The method of control selection—via questionnaire distributed to association members—may introduce selection bias, as this group may not be fully representative of the broader veteran population. However, key demographic indicators suggest reasonable representativeness: 59% of participants were over 66 years old, closely aligning with census data indicating 53% in that age group; similarly, the gender distribution in our sample (84% male) is comparable to census figures (86.4%) [29].

Similarly, while simple random sampling ensures our sample reflects the MONARCH dataset, it is possible that the MONARCH dataset, despite its large size, may not fully represent

all UK charity users. However, this does not compromise the validity of our findings, as MONARCH remains the largest and most comprehensive dataset on charity users currently available.

Therefore, it may be assumed that the differences observed between cases and controls in Table 1 are less likely due to selection bias and more plausibly reflect true disparities in access to charitable services. Another limitation is the lack of prospective longitudinal testing. While the models demonstrate good accuracy on the training and test datasets, their performance has not yet been validated with healthy veterans. This limits the generalizability of the findings. Additionally, although the study attempts to control for various confounders, there may still be unmeasured variables that could influence the results. Factors such as mental health status or social support networks were not included in the analysis but could significantly impact charity service usage [30].

The interpretability of machine learning models also poses a challenge. While the random forest algorithm provides robust predictive power, the importance of individual variables in the model is less interpretable compared to logistic regression. This makes it difficult to draw specific conclusions about the role of each predictor. The problem was addressed in this study by comparing the two

algorithms, which provided a more detailed picture of the contribution of each risk factor.

Conclusions

This study makes a significant contribution to understanding the risk factors that lead UK veterans to use charity services, by integrating both traditional logistic regression and an innovative machine learning approach to analyse the data.

Future research should continue to explore the roles of these variables to better inform intervention programs. The predictive models developed in this study hold significant potential for real-world application, enabling the estimation of the likelihood that a healthy veteran will access charity services within a year. This capability is the basis for the implementation of targeted preventive interventions, allowing for proactive measures to be taken to support veterans before they reach a point of needing charity services in a period of crisis. By leveraging these models, policymakers and service providers can allocate resources more efficiently and develop tailored strategies to address the specific needs of at-risk veteran subgroups.

Supplementary data

Supplementary data are available at EURPUB online.

Conflict of interest: None declared.

Funding

MONARCH project: funded by the Armed Forces Covenant Fund Trust (grant n. 3590) and the Forces in Mind Trust (project n. FiMT22/1024NUN).

Data availability

The dataset used for this study is available upon request from the corresponding author, subject to approval of original studies' principal investigator [13, 20].

Key points

- Younger, non-officer, single veterans are more likely to access military charities.
- Predictive models potentially can identify veterans at higher risk of charity use.
- Targeted, data-driven interventions could efficiently allocate resources to specific subgroups of veterans and prevent them from needing charity services.

References

- Bollinger MJ, Schmidt S, Pugh JA et al. Erosion of the healthy soldier effect in veterans of US military service in Iraq and Afghanistan. *Popul Health Metr* 2015; 13:8.
- Baillargeon J. Characteristics of the healthy worker effect. *Occup Med* 2001; 16:359–66.
- Amini H, Selaymani-Dodaran M, Moosavi B et al. Long-term health outcomes among survivors exposed to sulfur mustard in Iran. *JAMA Netw Open* 2020; 3:e2028894.
- Rhead R, MacManus D, Jones M et al. Mental health disorders and alcohol misuse among UK military veterans and the general population: a comparison study. *Psychol Med* 2022;52:292–302.
- Penn-Barwell JG, Roberts SAG, Midwinter MJ et al. Improved survival in UK combat casualties from Iraq and Afghanistan: 2003. *J Trauma Acute Care Surg* 2015; 78:1014–20.
- Smith TJ, Marriott BP, Dotson L et al. Overweight and obesity in military personnel: sociodemographic predictors. *Obesity* 2012;20:1534–8.
- Bray RM, Hourani LL. Substance use trends among active duty military personnel: findings from the United States Department of Defense Health Related Behavior Surveys. *Addiction (Abingdon, England)* 2007;102:1092–101.
- Kiernan MD, Rodrigues M, Mann E et al. The map of need: identifying and predicting the spatial distribution of financial hardship in Scotland's veteran community. *BMJ Mil Health* 2022;168:57–63.
- Kiernan MD, Osbourne A, McGill G et al. Are veterans different? Understanding veterans' help-seeking behaviour for alcohol problems. *Health Soc Care Community* 2018;26:725–33.
- Fodeeva A, Tiwari A, Mann E et al. A protocol for developing a complex needs indicator for veterans (CNIV) in the UK. *Public Health Pract* 2022;4:100281.
- Fodeeva A, Tomietto M, Tiwari A et al. Validation of a complex needs indicator for veterans in the UK using a composite indicators' method. *Public Health Pract* 2024; 7:100464.
- Armed Forces Charities—Sector Insights. <https://www.afc.org.uk/wp-content/uploads/2024/04/Sector-Insights-Armed-Forces-Charities-2024-PDF.pdf> (January 2025, date last accessed) (2024).
- Serra G, Tomietto M, McGill A et al. Developing a nationwide registry of UK veterans seeking help from sector charities—a machine learning approach to stratification. *Eur J Public Health* 2024;33:5–10.
- Hoge CW, Castro CA, Messer SC et al. Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. *N Engl J Med* 2004;351:13–22.
- Scotland's Census 2022 - UK Armed Forces Veterans | Scotland's Census. <https://www.scotlandscensus.gov.uk/2022-results/scotland-s-census-2022-uk-armed-forces-veterans/> (January 2025, date last accessed).
- Office for National Statistics. <https://www.ons.gov.uk/peoplepopulationandcommunity/armedforcescommunity/articles/characteristicsofukarmedforcesveteransenglandandwalescensus2021/census2021#:~:text=Of%20the%20total%20just%20over,%20and%2046.5%25%20in%20England> (January 2025, date last accessed) (2023).
- Khouri MJ, Engelgau M, Chambers DA et al. Beyond public health genomics: can big data and predictive analytics deliver precision public health? *Public Health Genomics* 2018;21:244–50.
- Rasmussen SA, Khouri MJ, Del Rio C. Precision public health as a key tool in the COVID-19 response. *JAMA* 2020;324:933–4.
- Naumova EN. Precision public health: is it all about the data? *J Public Health Pol* 2022;43:481–6.
- Johnson A, Serra G, Tomietto M et al. Identifying the risk factors associated with food insecurity in the UK veteran population: a nationwide survey. *J Nutr Sci* 2024; 13:e63.
- Bernad AE. Sampling methods. *J Hum Lact* 2020;36:224–6.
- R. The R Project for Statistical Computing. <https://www.r-project.org/>
- Sjoberg DD, Whiting K, Curry M et al. Reproducible summary tables with the gsummary package. *R Journal* 2021;13:570–80.
- Saleem J, Zakar R, Butt MS et al. Application of the boruta algorithm to assess the multidimensional determinants of malnutrition among children under five years living in Southern Punjab, Pakistan. *BMC Public Health* 2024;24:167.
- Molnar C, Casalicchio G, Bischl B. iml: an R package for interpretable machine learning. *JOSS* 2018;3:786.
- Greenwell BM. pdp: an R package for constructing partial dependence plots. *R Journal* 2017;9:421–36.
- Kursa MB, Rudnicki WR. Feature selection with the boruta package. *J Stat Soft* 2010;36:1–13.
- Liaw A, Wiener M. Classification and regression by randomForest. *R news* 2002; 2:18–22.
- Characteristics of UK Armed Forces Veterans, England and Wales—Office for National Statistics. <https://www.ons.gov.uk/peoplepopulationandcommunity/armedforcescommunity/articles/characteristicsofukarmedforcesveteransenglandandwalescensus2021/census2021> (January 2025, date last accessed).
- Prosencher E, Aase DM, Pasi HM et al. Impact of perceived social support on mental health, quality of life, and disability in post-9/11 U.S. Military veterans. *Armed Forces Soc* 2022;48:115–35. <https://doi.org/10.1177/0095327X20919922>



**Northumbria
University**
NEWCASTLE

Northern Hub for Veterans' & Military Families' Research



MONARCH

Map of Need Aggregated
Research Study

FiMT
forces in mind trust

THE ARMED FORCES
COVENANT FUND TRUST
Funded by HM Government

 **COMMUNITY
FUND**