

Original Paper

Novel Musculoskeletal Hypotheses in the Armed Services Trauma and Rehabilitation Outcome (ADVANCE) Cohort: Development and Application of Sparse Group Factor Analysis Methodology

Fraje C E Watson^{1*}, PhD; Fabio S Ferreira^{2*}, PhD; Balasundaram Kadirvelu², PhD; Alex N Bennett^{1,3}, PhD; Aldo A Faisal^{1,2}, PhD; Neil Graham⁴, PhD; Harriet Kemp⁵, PhD; Paul Cullinan, PhD; Christopher Boos^{1,6}, PhD; Nicola T Fear⁷, PhD; Anthony M J Bull¹, PhD

¹Department of Bioengineering, Imperial College London, London, England, United Kingdom

²Department of Computing, Imperial College London, London, England, United Kingdom

³Defence Medical Rehabilitation Centre, Loughborough, United Kingdom

⁴Department of Brain Sciences, Imperial College London, London, England, United Kingdom

⁵Department of Surgery and Cancer, Imperial College London, London, England, United Kingdom

⁶Health and Social Sciences, Bournemouth University, Bournemouth, England, United Kingdom

⁷King's Centre for Military Research and Academic Department of Military Mental Health, King's College London, London, England, United Kingdom

*these authors contributed equally

Corresponding Author:

Fraje C E Watson, PhD

Department of Bioengineering

Imperial College London

86 Wood Lane, 5th floor, Sir Michael Uren Building

London, England W12 0BZ

United Kingdom

Email: f.watson@imperial.ac.uk

Abstract

Background: Musculoskeletal conditions are a leading global cause of disability, yet the factors influencing long-term musculoskeletal health, particularly following trauma, remain incompletely understood. Machine learning could be applied to identify previously unknown patterns in large-scale, multimodal datasets.

Objective: This study aims to test the ability of a new sparse group factor analysis method to uncover hidden patterns in large-scale multimodal datasets and generate testable, clinically relevant hypotheses.

Methods: This study applies sparse group factor analysis, a hierarchical unsupervised machine learning method, to the Armed Services Trauma and Rehabilitation Outcome (ADVANCE) cohort to identify latent structures in multimodal clinical data. ADVANCE is a prospective longitudinal dataset of 1145 UK military personnel and veterans who served in Afghanistan. Half the cohort sustained combat injuries, and the remainder were frequency matched on deployment, service, rank, role, age, and ethnicity. Study 1 validated the approach by rediscovering known group-level patterns between combat-injured and noninjured participants, including poorer outcomes in pain, mobility, and bone health among those with lower limb loss. Study 2 explored the injured, nonamputee subgroup without prespecified labels to identify new hypothesis-generating clusters that could subsequently be tested using standard hypothesis-testing methods.

Results: The ADVANCE cohort was 34.1 (SD 5.4) years old and 8.3 (SD 2.1) years postinjury or 7.7 (SD 1.9) years since matched deployment. A subgroup of 125 individuals with worse musculoskeletal outcomes was uncovered. This group had greater body mass (mean 92.6, SD 14.7 kg vs mean 88.0, SD 13.4 kg; $P=.002$), higher injury severity (median 12, IQR 5-22 vs median 9, IQR 4-14; $P=.002$), and reduced health-related quality of life with head injury. These findings led to a novel hypothesis that head injury, including potential traumatic brain injury, is associated with long-term musculoskeletal deterioration. This hypothesis is supported by literature in both athletic and military populations and will be tested in follow-up analyses.

Conclusions: Our findings demonstrate how sparse group factor analysis, combined with clinical insight, can uncover hidden patterns in large-scale datasets and generate testable, clinically relevant hypotheses that inform prevention, treatment, and rehabilitation strategies.

J Med Internet Res 2026;28:e91958; doi: [10.2196/91958](https://doi.org/10.2196/91958)

Keywords: machine learning; artificial intelligence; traumatic brain injury; musculoskeletal health; military; cohort study; group factor analysis

Introduction

Musculoskeletal health is an important but complex societal issue. Musculoskeletal disorders are associated with aging [1], genetics [2], occupation [3,4], trauma [5,6], and lifestyle [7] and can be exacerbated by psychosocial factors [8]. Musculoskeletal injury can further compound these challenges [5]. Components associated with increased risk of musculoskeletal injury are multifactorial but include physical fitness, obesity and being underweight [9], age, sex [10], sports such as running [10] and contact sports [11], ergonomics [12], poor neuromuscular control [13], and physically demanding occupations (eg, military) [14,15].

Musculoskeletal conditions are the second-highest cause of global nonfatal disability [16,17]. These conditions affect more than a third of the UK population [18] and 494 million people globally [16]. They result in 2462 disability-adjusted life years per 100,000 globally, which is increasing over time [16,19,20].

A better understanding of the interconnectedness of long-term musculoskeletal outcomes (eg, 2 outcomes influence each other, 1 outcome is associated with a particular injury type) associated with injury could guide preventative, interventional, and rehabilitation efforts, potentially decreasing the disability burden and increasing quality life years. Achieving this better understanding would require large, longitudinal datasets. However, looking beyond the expected to generate novel associations or hypotheses in a large dataset is challenging. Artificial intelligence (AI) could provide the necessary processing power and pattern recognition skills required to overcome this challenge. AI has already shown diagnostic, treatment personalization, and prognostic potential in medicine across a number of diagnoses (eg, detecting breast cancer, personalizing sepsis treatment strategies, predicting progression of neurodegenerative diseases) [21-29].

Unsupervised machine learning, a branch of AI, focuses on uncovering hidden patterns and intrinsic structures within unlabeled datasets, without relying on labeled data [30]. Key techniques within unsupervised learning include clustering, dimensionality reduction, and density estimation. These methods are particularly valuable in exploratory data analysis, as they enhance model interpretability and efficiency by elucidating latent structures and reducing complexity. Here, we use sparse group factor analysis (sparse GFA), a hierarchical probabilistic machine learning model to identify subgroup-specific and subgroup-common multimodal latent disease factors in participant groups [31].

The Armed Services Trauma and Rehabilitation Outcome (ADVANCE) study is collecting prospective, longitudinal physical and psychosocial data from a cohort of 1145 UK servicemen who served in the Afghanistan War [32] with the aim of reporting the long-term physical and psychosocial outcomes of combat-related injury. The baseline dataset of ADVANCE contains over 13 million datapoints. Important, predetermined hypotheses that set the recruitment target numbers for the study have been made and tested using musculoskeletal data [6,33-35]. The application of sparse GFA unsupervised machine learning techniques has the potential to uncover hidden patterns in complex datasets to generate previously unseen hypotheses, thus addressing one of the key limitations of analytical techniques currently in use.

This aim of this paper is to describe the methodology undertaken to identify a novel and clinically relevant hypothesis using the musculoskeletal data of the ADVANCE study. To do this, we present two studies: (1) application of sparse GFA to the ADVANCE dataset and comparison of results to known data patterns and (2) application of sparse GFA to a nested study using a subgroup of heterogeneous and underinvestigated ADVANCE participants to generate a novel hypothesis. We hypothesize that GFA can be applied to a complex musculoskeletal dataset and that new clinically relevant hypotheses on causation/correlation can be inferred *de novo*.

Methods

Ethical Considerations

Ethical approval was granted by the UK Ministry of Defence Research Ethics Committee (357/PPE/12). Informed, written consent was obtained from each participant.

The ADVANCE Cohort

The protocol for the ADVANCE study is described elsewhere [32]. Participants (n=1145) were male serving or veterans of the UK military deployed on combat operations to Afghanistan between 2003 and 2014. Injured participants (n=579, 50.6%) sustained combat injuries requiring aeromedical evacuation to the United Kingdom. Uninjured participants (n=566, 49.4%) were frequency matched by deployment, service, rank, role, and age to injured participants to form a comparison group. The study is limited to men, since fewer than 18 female personnel sustained combat injuries, of whom only 3 had serious or very serious injuries [36].

Baseline data were collected between March 2016 and August 2020, a median of 7.9 (IQR 6.7-9.4) years postinjury or matched deployment.

Comprehensive physical and psychosocial data were collected from all participants across a range of domains, such

as cardiovascular, mental health, respiratory health, musculoskeletal health, proteomics, hearing, and aging. This pilot analysis uses the musculoskeletal-adjacent data described in Table 1.

Table 1. Type, name, and description of Armed Services Trauma and Rehabilitation Outcome (ADVANCE) variables used to assess ADVANCE study participants and entered into the sparse group factor analysis model.

Method and type	Details
Clinical assessment and self-reported	
Demographic	Age, height, (adjusted) body mass, NISS ^a , socioeconomic status, BMI, abdominal circumference, waist-to-hip ratio, race, service, role, rank, smoking status, amputation status
DEXA ^b of the whole body, right and left proximal femur, and lumbar spine	
Bone mineral density	t-score for the left and right hip (femoral neck, trochanter, Ward's triangle, intertrochanter, hip total), t-score for the L1-L4 spine, raw bone mineral density for each lumbar vertebra (L1-L4), bone mineral density for the whole body
Body composition	Total lean tissue in the whole body, percentage of fat tissue in the whole body
Posterior-anterior semiflexed knee radiographs	
Radiographic knee osteoarthritis	Kellgren-Lawrence score (0-4) [37], OARSI ^c joint space narrowing score (0-3), OARSI osteophyte score (0-3), OARSI sclerosis score (0-3) [38], minimum medial joint space width, and minimum lateral joint space width for left and right knees
Physical assessment—all participants	
Six-minute walk test	Furthest distance a participant can walk in 6 minutes
Physical assessment—lower limb loss only	
Amputee Mobility Predictor [39]	Ability to perform 21 tasks scored by independent assessor scored 0-47, where a higher score is indicative of better mobility
Questionnaire—all participants	
ASAS ^d definition of inflammatory back pain [40]	Five questions about back pain to identify inflammatory back pain (suspected when ≥ 4 answers are yes)
Health-related quality of life score (EQ-5D-5L ^e) [41]	Health-related quality of life score for individual items; mobility, self-care, usual activities, pain or discomfort, anxiety or depression each scored 1-5, where a higher score is a worse outcome, and overall health scored 0-100, where 100 is the best health.
Back pain	Scored for severity, frequency, and impact 0-10, where 10 is the severe, frequent, or impactful.
ODI ^f [42]	Ten-item questionnaire each scored 0-5 with a total score of 0-100, where 100 is indicative of the most back pain-related disability.
KOOS ^g —pain [43]	Nine-item questionnaire each scored 0-5 for left and right knees, each with a total score of 0-100, where 0 is indicative of the worst pain.
KOOS—symptoms [43]	Seven-item questionnaire each scored 0-5 for left and right knees, each with a total score of 0-100, where 0 is indicative of the worst symptoms.
Knee and Hip Joint Pain scores	Left and right knees and hips scored for severity, frequency, and impact 0-10, where 10 is the severe, frequent, or impactful
NAHS ^h	Twenty-item questionnaire each scored 0-5 for left and right hips, each with a total score of 0-100 for each hip, where 0 is indicative of the worst hip pain, symptoms, activities, or function.
DASH ⁱ —core [44]	Thirty-item questionnaire each scored 0-5 with a total score of 0-100, where 100 is indicative of the most disability
DASH—work [44]	Four-item questionnaire each scored 0-5 with a total score of 0-100, where 100 is indicative of the most disability
DASH—sport [44]	Four-item questionnaire each scored 0-5 with a total score of 0-100, where 100 is indicative of the most disability
IPAQ ^j [45]	Total weekly activities categorized as moderate (“activity that makes you breathe somewhat harder than before”) and vigorous (“activities that take hard physical effort and make you breathe much harder than normal”)
Questionnaire—lower limb loss specific	
SIGAM ^k Mobility Questionnaire [46]	Scored A-F based on self-reported mobility
Socket comfort score [47]	Scored per amputated limb 0-10, where 10 is the most comfortable
Prosthetic satisfaction score [47]	Scored per amputated limb 0-10, where 10 is the most satisfaction
Phantom pain score	Scored for severity, frequency, and impact 0-10, where 10 is the severe/frequent/impactful
Stump pain score	Scored for severity, frequency, and impact 0-10, where 10 is the severe, frequent, or impactful

^aNISS: New Injury Severity Score.

^bDEXA: dual X-ray absorptiometry.

^cOARSI: Osteoarthritis Research Society International.

^dASAS: Assessment of SpondyloArthritis International Society.

^eEQ-5D-5L: EuroQoL 5-Dimensions 5-Level.

^fODI: Oswestry Disability Index.

^gKOOS: Knee Osteoarthritis Outcome Score.

^hNAHS: Non-Arthritic Hip Score.

ⁱDASH: Disability of the Arm, Shoulder and Hand.

^jIPAQ: International Physical Activity Questionnaire.

^kSIGAM: Special Interest Group in Amputee Medicine.

For participants with limb loss, adjusted body mass was calculated depending on body segments missing [48,49] and without wearing prosthetics. The height of participants with lower limb loss was measured using either actual or reported height.

Injuries sustained by the injured group were identified using the Joint Theatre Trauma Registry (JTTR), which details combat injury type and location.

Data Handling

Data Processing

Data were preprocessed with a bespoke Python workflow. For the nested study addressing the second aim, this workflow was applied exclusively to the injured cohort excluding those with lower limb loss ($n=422$). Focusing on this subcohort allowed the assessment of heterogeneity in musculoskeletal outcomes without the confounding influence of limb loss.

All 422 participants were retained through preprocessing. Some records contained item-level missingness before imputation, which affected 231 questionnaire records, 35 anthropometric records, 12 x-ray records, 9 blood records, 7 left hip dual X-ray absorptiometry (DEXA), 5 right hip DEXA, and 3 spinal DEXA records, and 2 pain-location records, in most cases involving only a small number of items per record. Missingness was addressed using joint multivariate iterative imputation fitted over the concatenated multimodal feature matrix (scikit-learn `IterativeImputer` with a decision-tree base learner, 100 iterations), allowing missing values in 1 modality to be informed by observed values in the others. We assumed missingness was conditionally random, as it was driven primarily by logistical factors rather than underlying clinical constructs. We did not perform complete-case exclusion, and the final dataset retained all 422 participants.

Cohort Curation

Participants labeled (injured, nonamputee) in the study metadata were retained. Questionnaire variables were kept at the total- or domain-score level to minimize multicollinearity and reduce the risk of model overfitting; individual item scores were discarded. All musculoskeletal-adjacent variables listed in Table 1 were imported and subjected to range and type checks. Records with missing or implausible values in key demographic or musculoskeletal fields (eg, body mass, New Injury Severity Score [NISS]) were removed.

Modality Construction

Features were organized into separate modality-specific matrices for self-reported outcomes, clinical assessments, and imaging-derived measures, consistent with the sparse GFA framework. Continuous variables were standardized to zero mean and unit variance, and categorical variables were one-hot encoded. Outliers were retained because the hierarchical Bayesian formulation of sparse GFA is robust to extreme observations, and such extremes are clinically informative in trauma cohorts.

To ensure comparable latent-factor estimation across modalities, only participants with complete data for every retained modality were included in the final analysis set. The resulting multimodal matrix constituted the input to the sparse GFA model described in the subsequent section and represents a harmonized, quality-controlled view of musculoskeletal health in the injured nonamputee group.

Data Analysis

Sparse GFA

Sparse GFA identifies latent factors in subgroups that capture the relationships within multimodal data [31]. Sparse GFA is an extension of GFA [50], where regularized horseshoe priors are added over the loading matrices and latent variables. The regularized horseshoe prior [51] is a popular shrinkage prior used in sparse Bayesian regression to ensure that small coefficients are heavily shrunk toward zero, while large coefficients are regularized but remain large. In this way, sparse GFA improves model interpretability and identifies latent variables that are differently expressed across the sample.

In this study, sparse GFA was fitted using a Markov chain Monte Carlo algorithm with 5 sampling chains and 4000 samples (the first 1000 were discarded as warm-up). The inference was randomly initialized 5 times, and the best initialization was selected to maximize the expected log joint posterior density. All sampling chains were initialized with 20 factors. More details about the experiment setup are given in Ferreira et al [31].

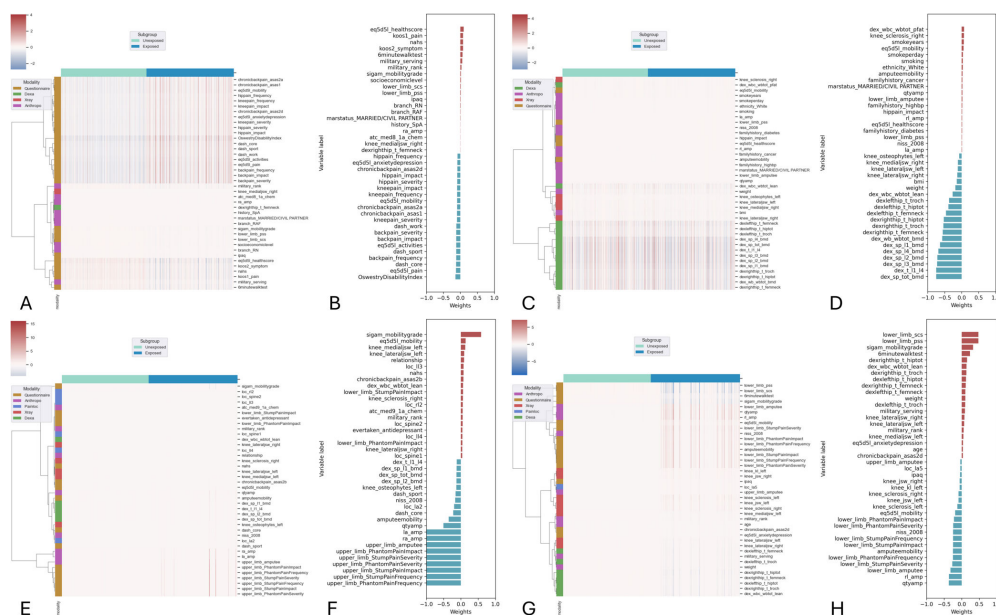
We used the approach proposed by Ferreira et al [31] to select robust factors, that is, the factors are first averaged over the posterior samples within a sampling chain and then compared to the factors obtained in other sampling chains using cosine similarity. Two factors were considered the same if the cosine similarity between them was greater than 0.80, as

previously described [52]. A factor was considered robust if it had been obtained in at least half of the sampling chains.

Interpretation

For each participant, sparse GFA calculates a distribution across the latent variables, showing how a particular participant is represented in each component. For each component, we also obtain a representation across the features in each data modality. This provides a loading matrix showing the importance of each feature in each component for each participant, which we refer to as *data components*. These data components allow the interpretation of the results (eg, a positive [more red] value for a particular participant can be interpreted as scoring high in a task).

Figure 1. Study 1—top 4 data components identified by sparse group factor analysis in the uninjured and injured participants in the Armed Services Trauma and Rehabilitation Outcome (ADVANCE) dataset and their associated weights (A and B, C and D, E and F, and G and H). Blue represents a value below the sample average, and red represents a value above the sample average. Colors on the left identify up to 5 different modalities or types of data, as listed for each factor.



In study 2, a nested subset of injured participants without lower limb loss injuries was analyzed without an inherent group structure to demonstrate a real-world application of sparse GFA that could potentially generate *de novo* hypotheses on causation or correlation. Loading matrices are analyzed to identify a group structure (eg, subgroup has poor scores across specific outcomes). Next, patterns in the data were investigated and used to develop a novel hypothesis. In this example, the following steps were taken:

1. Group structure identified a subgroup with poor scores across specific outcomes.
2. Participants were manually divided into those with and without poor scores for those specific outcomes.
3. Injury location, demographic information, and injury severity were visually and statistically compared between groups.
4. Overall health EuroQoL 5-Dimensions 5-Level (EQ-5D-5L) scores were visually compared between injury location groups.

In study 1, data components were considered where there was a clear difference between the injured and uninjured groups (eg, Oswestry Disability Index scores were largely red for the injured group and blue for the uninjured group; Figure 1A). The hypothesis regarding this difference was inherent due to the known way the data were split into groups. The injured group contained a subgroup of 157 men who sustained injuries resulting in limb loss. They were not labeled but could be identified through limb loss-specific outcomes (eg, phantom limb pain), or where nonspecific outcomes (eg, low bone mineral density) were shown in combination with limb loss-specific measures.

Statistical Analysis

The JTTR records were unavailable for 41 injured participants who were subsequently excluded from analyses involving injury location or NISS (higher score=greater severity).

In the first study, 1 uninjured participant was excluded due to noncombat-related lower limb loss following matching.

In the second study, where statistical comparisons of continuous variables were made between the groups, the data were visually assessed for normality using histograms, and an independent 2-tailed *t* test, Mann-Whitney *U* test, or chi-squared test was used, as appropriate. Boxplots were compared visually for trends. Descriptive data and statistical analysis were processed using Stata SE version 18.5.

Results

Demographic Information

Across the entire cohort, the mean age was 34.1 (SD 5.4) years, the mean height was 179.1 (SD 6.8) cm, and the mean (adjusted) body mass was 89.0 (SD 13.3) kg. Participants self-described their race as White (n=1036, 90.6%), Black (n=46, 4.0%), Asian (n=41, 3.6%), or Mixed/Other (n=21, 1.8%). Data were collected from injured participants (those who sustained combat injuries in the Afghanistan war requiring aeromedical evacuation to the United Kingdom) at a mean of 8.3 (SD 2.1) years postinjury and from uninjured

participants (frequency-matched participants who did not sustain combat injuries in the Afghanistan war requiring aeromedical evacuation to the United Kingdom) at a mean of 7.7 (SD 1.9) years since matched deployment.

Sparse Group Factor Analysis

Study 1

Fourteen data components were identified (Supplementary Material S1 in [Multimedia Appendix 1](#)), of which the top 4 are shown in [Figure 1](#). [Table 2](#) shows the dominant variables that were different between injured and uninjured groups within each factor and overall interpretation.

Table 2. Data components identified in study 1 analysis, which can be seen in the raw data output in [Figure 1](#).

Group and type	Outcome variable
Injured	
Back pain	• More symptoms associated with spondyloarthritis (Figure 1A) • Worse severity, frequency, and impact of low back pain (Figure 1A) • Worse functional disability (ODI^a score) (Figure 1A)
Health-related quality of life	• Worse EQ-5D-5L^b scores for mobility, depression and anxiety, pain, and ability to carry out their usual activities (Figure 1A) • Worse overall health score (Figure 1A) • Worse 6MWT^c distance (Figure 1A)
Knee pain	• Worse frequency, severity, and impact of knee pain (Figure 1A) • Worse KOOS^d pain and KOOS symptoms scores (Figure 1A)
Hip pain	• Worse frequency, severity, and impact of hip pain (Figure 1A) • Worse NAHS^e scores (Figure 1A)
Upper limb function	• Worse DASH^f scores (Figure 1A)
Lifestyle	• Less often still serving in the military (Figure 1A)
Injured—lower limb loss	
Amputee-specific	• Phantom limb pain (frequency, severity, impact; Figure 1G) • Residuum pain (frequency, severity, impact; Figure 1G) • Poor SIGAM^g grade (Figure 1G)
Mobility	• Poor EQ-5D-5L mobility scores (Figure 1G) • Short 6MWT distances (Figure 1G)
Knee radiographs	• High Kellgren-Lawrence score (Figure 1G) • High score for sclerosis (Figure 1G) • High score for medial joint space narrowing (Figure 1G)
Bone mineral density	• Low bone mineral density in hip (Figure 1G)
Demographic	• High injury severity (NISS^h) (Figure 1G)

^aODI: Oswestry Disability Index.

^bEQ-5D-5L: EuroQoL 5-Dimensions 5-Level.

^c6MWT: Six-Minute Walk Test.

^dKOOS: Knee Injury and Osteoarthritis Outcome Score.

^eNAHS: Non-Arthritic Hip Score.

^fDASH: Disability of the Arm, Shoulder and Hand.

^gSIGAM: Special Interest Group in Amputee Medicine.

^hNISS: New Injury Severity Score.

In [Figure 1C,D](#), bone mineral density was identified as a key component, but it was not explained by a difference between the injured and uninjured groups.

Many of these clusters have already been investigated using hypothesis-driven statistical analyses by the ADVANCE study. For example, we have previously reported higher knee and hip Kellgren-Lawrence, Disability of the Arm, Shoulder and Hand, Oswestry Disability Index, overall pain, and back pain scores in the injured group [6,34,53-55],

and participants with lower limb loss have significantly lower bone mineral density in their amputation-side femoral neck [33].

Study 2

Overview

Fourteen data components were identified (Supplementary Material S2 in [Multimedia Appendix 1](#)), of which the top 4 are shown in [Figure 2](#).

Following the steps described above, a group structure was identified in [Figure 2A](#), which identified a group with high (red) scores across multiple outcomes and low (blue) scores across multiple outcomes.

scores for some other outcomes ([Table 3](#)). The threshold for a variable being red and blue is the sample mean. Next, participants were split into 2 groups along the identified structure where the change from red to blue or blue to red occurred across participants. In this instance, this led to a group of 125 participants with poor musculoskeletal outcomes, and the remaining 297 participants without them. Finally, patterns in the demographic, injury pattern, and overall health outcome data were analyzed to identify possible reasons for worse musculoskeletal outcomes.

Figure 2. Study 2—top 4 data components identified by sparse group factor analysis in the injured, nonamputee participants in the Armed Services Trauma and Rehabilitation Outcome (ADVANCE) dataset and their associated weights (A and B, C and D, E and F, and G and H). Blue represents a value below the sample average, and red represents a value above the sample average. Colors on the left identify up to 5 different modalities or types of data, as listed for each factor.

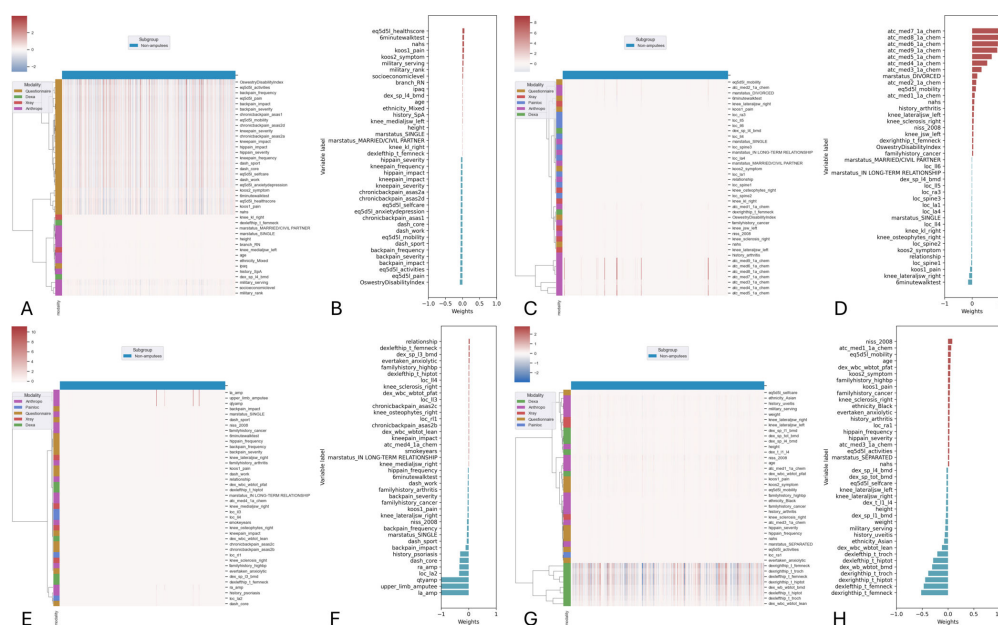


Table 3. Data components identified in round 2 analysis.

Group and type	Outcome variable
Subgroup of 125	
Back pain	- More symptoms associated with spondyloarthritis (Figure 2A) - Worse severity, frequency, and impact of low back pain (Figure 2A) - Worse functional disability (Figure 2A)
Health-related quality of life	- Worse EQ-5D-5L^a scores for mobility, depression and anxiety, pain, and ability to care for themselves (Figure 2A) - Worse overall health score (Figure 2A) - Worse 6MWT^b distance (Figure 2A)
Knee pain	- Worse frequency, severity, and impact of knee pain (Figure 2A) - Worse KOOS^c pain and KOOS symptoms scores (Figure 2A)
Hip pain	- Worse frequency, severity, and impact of hip pain (Figure 2A) - Worse NAHS^d scores (Figure 2A)
Upper limb function	Worse DASH^e scores (Figure 2A)
Demographic	- Less often still serving in the military (Figure 2A) - Lower military rank (Figure 2A)

^aEQ-5D-5L: EuroQoL 5-Dimensions 5-Level.

^b6MWT: Six-Minute Walk Test.

^cKOOS: Knee Injury and Osteoarthritis Outcome Score.

^dNAHS: Non-Arthritic Hip Score.

^eDASH: Disability of the Arm, Shoulder and Hand.

Pattern Identification

There were no significant differences in age (mean 34.5, SD 5.7 y vs 34.2, SD 5.2 y), height (mean 179.1, SD 6.9 cm vs 178.6, SD 6.2 cm), or time since injury (median: 8.5, IQR 7.1-10.0 y vs 8.7, IQR 7.0-10.3 y) between the groups with and without poor musculoskeletal outcomes. The group with poor musculoskeletal outcomes was heavier (mean 92.6, SD 14.7 kg) and had higher NISS scores (median 12, IQR 5-22) than those without (mean weight 88.0, SD 13.4 kg; median NISS 9, IQR 4-14); both comparisons were statistically significant ($P=.002$).

The group with poor musculoskeletal outcomes contained a significantly higher percentage of participants with upper limb ($n=61$, 56.0% vs $n=116$, 42.7%; $P=.02$), abdominal ($n=31$, 28.4% vs $n=46$, 16.9%; $P=.01$), and spinal injuries ($n=23$, 21.1% vs $n=33$, 12.1%; $P=.03$) compared to the group with better musculoskeletal outcomes. For injury

combination, both groups had the same top three: lower extremity injuries only ($n=15$, 13.8% and $n=67$, 24.6%), a combination of upper and lower extremity injuries ($n=8$, 7.3% and $n=20$, 7.4%), and upper extremity injuries only ($n=7$, 6.4% and $n=19$, 7.0%; [Figure 3](#)).

The visual inspection of boxplots further split each group by the presence or absence of an injury in each location and compared overall health score. As expected, scores for those with poor musculoskeletal outcomes were always worse; however, head injury ([Figure 4D](#)) was the only body region to worsen the overall health score further in the group with poor musculoskeletal outcomes. Full results are provided in Supplementary Material S2 in [Multimedia Appendix 1](#).

Pattern identification revealed the following hypotheses regarding worse musculoskeletal outcomes in a subgroup of the injured, nonamputees: greater body mass, worse injury severity, and head injury.

Figure 3. UpSet plot showing injury location combinations for participants in the injured, nonamputee cohort with (A) and without (B) poor musculoskeletal (MSK) outcomes.

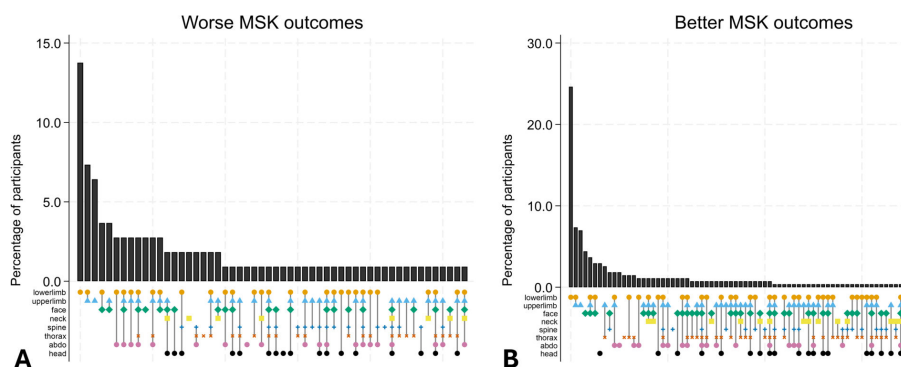
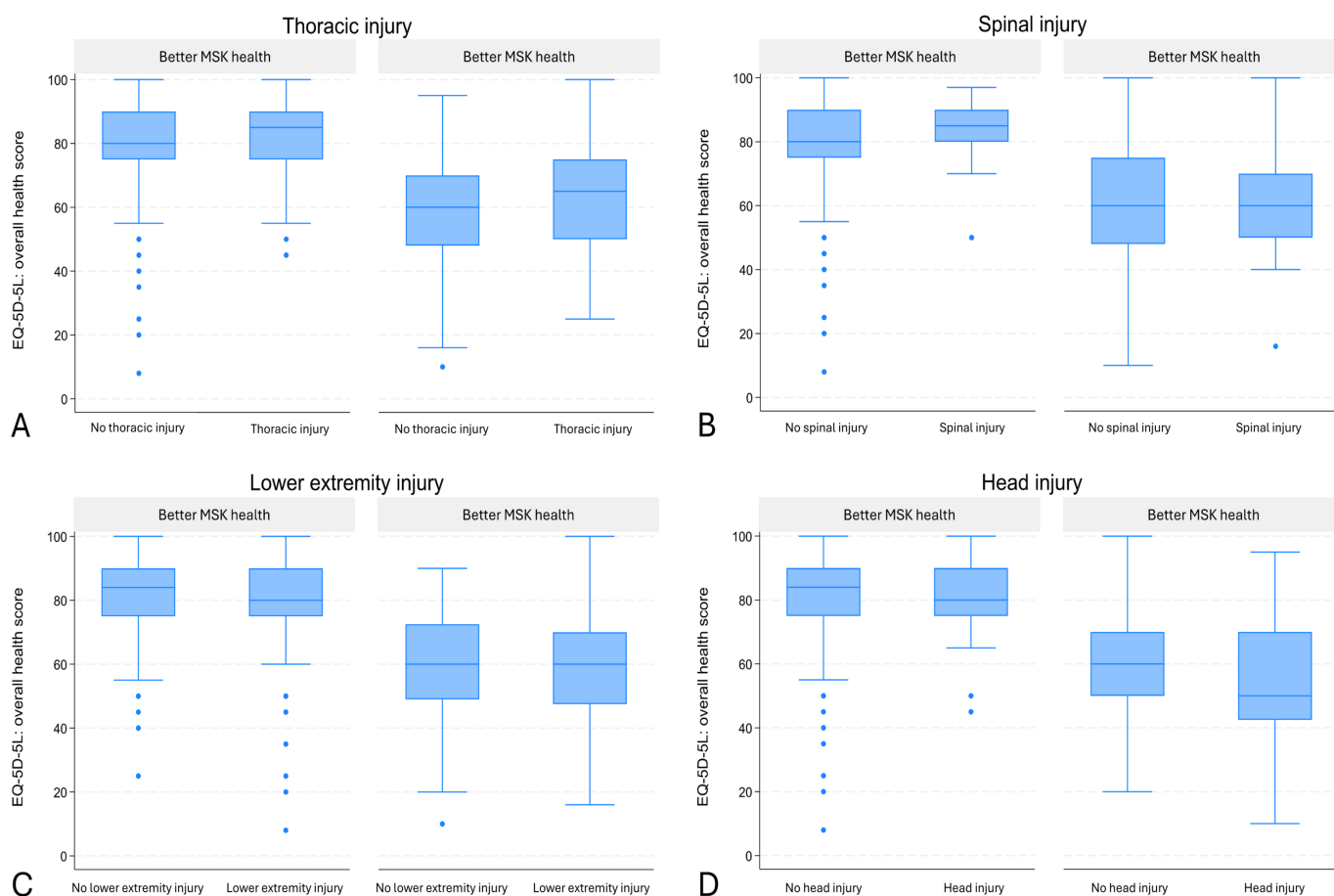


Figure 4. Boxplot of overall health score for injured, nonamputee participants with and without poor musculoskeletal (MSK) outcomes, by presence/absence of a thoracic (A), spinal (B), lower extremity (C), or head (D) injury.



Discussion

Principal Findings

This work has demonstrated a method for generating novel hypotheses in multidimensional data using sparse GFA. In study 1, we showed how sparse GFA was able to identify known clusters in the cohort, thus corroborating their value. Study 2 showed how sparse GFA was able to identify a cluster with particularly poor musculoskeletal outcomes in a cohort with a heterogeneous internal structure. We investigated patterns in the cluster to generate hypotheses about the potential cause. Specifically, in a cluster of participants with worse musculoskeletal outcomes, we identified higher injury severity score, higher body mass, and head injury as potential causes. The novel head injury hypothesis could be investigated further using typical statistical methods.

In study 1 using the full ADVANCE cohort, we were able to identify patterns that had already been addressed, thus confirming the validity and usefulness of the model. Many of these findings were heavily influenced by scores from the participants with lower limb loss, consistent with the ADVANCE study's research, where their musculoskeletal outcomes are often worse [6,33-35,53-55]. Sparse GFA also generated hypotheses on hip pain across the injured and uninjured groups that were already being investigated when this study was performed. Study 1 identified a pattern in bone

mineral density findings that was not explained by injured or uninjured status. Future consideration of that pattern, such as that demonstrated in study 2, could be applied here to formulate a hypothesis.

Having confirmed the validity of sparse GFA, we subsequently analyzed the injured-, nonamputee group separately, as they are a more heterogeneous group without a clear internal structure that has not been investigated. Using one loading matrix as an example, sparse GFA identified a group of 125 participants with similarly poor musculoskeletal outcomes. The exploration of patterns delineating these groups found that those with worse musculoskeletal outcomes were significantly heavier, which is logical as obesity is often linked to worse musculoskeletal outcomes [56]. The group with worse musculoskeletal outcomes had significantly worse injury severity scores, which also explains worse outcomes. Using injury and overall health outcome data, we identified head injury as a possible correlate of worse musculoskeletal outcomes. This methodology has identified a new hypothesis for the ADVANCE study; long-term musculoskeletal outcomes of participants who sustained a head injury are worse than participants who did not. However, this hypothesis requires independent validation before any inferential claims can be made.

A head injury could lead to a traumatic brain injury (TBI) either as a result of direct impact to the head or secondary to a blast [57]. A review of the literature on head injuries,

TBI, or concussion and musculoskeletal injury established that this is a developing field in professional athletes and with existing reports on military personnel. A systematic review and meta-analysis of 8 studies found 2 times higher odds of musculoskeletal injury in athletes following a concussion up to 2 years later [58], with similar findings in a meta-analysis of 27 papers including athletes and military personnel [59]. Military personnel who sustained a concussion (ie, a mild TBI) were 1.84 times more likely to sustain an upper limb musculoskeletal injury in the subsequent year [60]. A long-term outcome study in people with a moderate-to-severe TBI more than 15 years prior reported high rates of musculoskeletal complaints [61]. There is no consensus on the neuropathology of these findings, but some studies suggest impaired neuromuscular control [62] and difficulty with dual-task control [63] following TBI.

Based on the findings from this study, the association between head or facial injury, TBI, and poor musculoskeletal outcomes in military personnel will be investigated in the ADVANCE cohort using hypothesis-driven statistical methods. In the ADVANCE study, incidence of index and lifetime TBI has been defined using the Ohio State University TBI Identification Method, including the clinical assessment of brain magnetic resonance imaging. The novel hypothesis could be tested within a subset of participants with no lifetime TBI and only an index TBI at the point of injury to strengthen the conclusions made about TBI. Regression models adjusted for confounders could compare musculoskeletal scores between groups. Once available for at least 3 timepoints, longitudinal data will be compared over time using a mixed effects ordinal regression model. The density of prospective, longitudinal musculoskeletal outcome data with longer follow-up times than most existing literature will ensure a unique contribution to the literature. These findings will inform guidance on the treatment of TBI, long-term outcomes, and identification of areas for rehabilitation.

Sparse GFA could be applied to a larger, more complex dataset across multiple health care-related modalities such as mental health, cardiovascular health, proteomics, and brain magnetic resonance imaging. Many of the clusters and hypotheses formed in this pilot study had already been addressed because it is likely that different musculoskeletal outcomes will be related to each other. Introducing more data types across specialties could uncover a previously unseen or unknown interconnectedness between data themes. This method blends the high processing power of AI while retaining a rational clinical input to generate novel hypotheses that can be further tested using traditional statistical techniques.

Limitations

This study identified injury locations using the JTTR, which has limitations as the information is documented in a live war

zone and active trauma care setting, which could result in inaccuracies. There is potential that injuries are missed and not documented at this stage, such as head injuries, which could affect the robustness of the hypothesis. Additionally, other injuries could contribute to TBI, not just head injuries, as well as the influence of the NISS. To overcome this, future testing of this new hypothesis will incorporate a thorough and robust analysis of potential TBI to more accurately define head injury status, adjusting for NISS and other potential confounders as required.

The JTTR defines head and face injuries separately, so only head injuries contribute to this hypothesis, but some facial injuries are also linked to TBI [64]. The ADVANCE study is currently working on defining TBI using information from the JTTR, self-reported past medical history, and the Ohio State Questionnaire, with severity defined using the Mayo Classification System. This classification will allow for a comprehensive assessment of TBI and a thorough investigation of the hypothesis.

This study uses data from a male, military cohort, which challenges the direct generalizability of the novel hypothesis it generated to the general population and/or female cohorts. However, presented here is a methodology that could be used in the general population (eg, using UK Biobank data) and/or female military cohorts in future research. The specific novel hypothesis described here could be generalized to male civilians who have been exposed to combat-related trauma, a circumstance that is sadly increasing worldwide. The “overall health today” outcome of the EQ-5D-5L was used to test the effect of injury location as it was not body region specific. Doubts have been cast on the EQ-5D-5L, which has been considered to have a ceiling effect and to be insensitive to change in some populations, despite its strong psychometric properties [65]. Finally, the use of ADVANCE musculoskeletal data only could have biased the results compared to using the full ADVANCE dataset, but this is inherent to the pilot nature of this work and we plan to apply it to the full dataset in the future.

Conclusions

This application of sparse GFA combined with human interpretation can be used to identify hidden patterns and intrinsic structures within complex, multimodality datasets to generate hypotheses that may not have been evident previously. We have demonstrated this method with a pilot portion of musculoskeletal data from the ADVANCE study and generated a novel hypothesis regarding head injury.

Acknowledgments

We wish to thank all of the research and clinical staff at Headley Court, Stanford Hall, and the Defence Medical Services who have helped with the ADVANCE study, including Maria-Benedicta Edwards, Helen Blackman, Melanie Chesnokov, Emma Coady, Sarah Evans, Guy Fraser, Meliha Kaya-Barge, Maija Maskunitty, David Pernet, Helen Prentice, Urszula Pucilowaska,

Stefan Sprinchmoller, Owen Walker, Eleanor Miller, Lajli Varsani, Anna Varey, Molly Waldron, Danny Weston, Tass White, Seamus Wilson, Louise Young, Dan Dyball, Col Ian Gibb, Surg Cmd David Gray, and Lt Col Edward Sellon. We would also like to thank the ADVANCE study participants for their generous and ongoing donation of time.

Generative artificial intelligence was not used to plan, develop, or write any part of this manuscript.

Funding

The ADVANCE Study is funded through the ADVANCE Charity. Key contributors to the charity are the Headley Court Charity (principal funder), HM Treasury (LIBOR Grant), Help for Heroes, Nuffield Trust for the Forces of the Crown, Forces in Mind Trust, National Lottery Community Fund, Blesma—The Limbless Veterans, and the UK Ministry of Defence. The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the manuscript. This study was additionally funded by the Imperial College London Dame Julia Higgins Postdoc Collaborative Research Fund.

Data Availability

The code was adapted from Ferreira et al [31] and is available on GitHub [66]. Given their unusually sensitive nature, the data have not been made widely available. Requests for data will be considered on a case-by-case basis and are subject to UK Ministry of Defence clearance via adv_data_team@imperial.ac.uk.

Authors' Contributions

Conceptualization: FCEW, FSF, BK, AMJB

Data curation: FCEW, BK

Formal analysis: FCEW, FSF, BK

Funding acquisition: FCEW, FSF, BK, ANB, AAF, HK, PC, CB, NTF, AMJB

Investigation: FCEW, FSF, BK

Methodology: FCEW, FSF, BK

Project administration: FCEW, FSF, BK

Resources: ANB, AAF, AMJB, NTF, PC, HK, CB

Software: FSF, BK

Supervision: ANB, AAF, HK, PC, CB, NTF, AMJB

Validation: FCEW, FSF

Visualization: FCEW, FSF

Writing – original draft: FCEW, FSF, BK

Writing – review & editing: FCEW, FSF, BK, ANB, AAF, NG, HK, PC, CB, NTF, AMJB

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary files containing all the latent variables and matched loadings identified in sparse group factor analysis.

[\[DOCX File \(Microsoft Word File\), 746 KB-Multimedia Appendix 1\]](#)

References

1. Harvey NC, Clegg PD, Dennison EM, et al. UKRI MRC National Musculoskeletal Ageing Network: strategic prioritisation to increase healthy lifespan and minimise physical frailty. *Arch Osteoporos*. Nov 21, 2022;17(1):147. [doi: [10.1007/s11657-022-01172-z](https://doi.org/10.1007/s11657-022-01172-z)] [Medline: [36414807](https://pubmed.ncbi.nlm.nih.gov/36414807/)]
2. Boer CG, Hatzikotoulas K, Southam L, et al. Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. *Cell*. Sep 2, 2021;184(18):4784-4818.e17. [doi: [10.1016/j.cell.2021.07.038](https://doi.org/10.1016/j.cell.2021.07.038)] [Medline: [34450027](https://pubmed.ncbi.nlm.nih.gov/34450027/)]
3. Osborne A, Blake C, Fullen BM, et al. Risk factors for musculoskeletal disorders among farm owners and farm workers: a systematic review. *Am J Ind Med*. Apr 2012;55(4):376-389. [doi: [10.1002/ajim.22001](https://doi.org/10.1002/ajim.22001)] [Medline: [22213399](https://pubmed.ncbi.nlm.nih.gov/22213399/)]
4. Lovalekar M, Hauret K, Roy T, et al. Musculoskeletal injuries in military personnel-descriptive epidemiology, risk factor identification, and prevention. *J Sci Med Sport*. Oct 2021;24(10):963-969. [doi: [10.1016/j.jsams.2021.03.016](https://doi.org/10.1016/j.jsams.2021.03.016)] [Medline: [33824080](https://pubmed.ncbi.nlm.nih.gov/33824080/)]
5. Whittaker JL, Losciale JM, Juhl CB, et al. Risk factors for knee osteoarthritis after traumatic knee injury: a systematic review and meta-analysis of randomised controlled trials and cohort studies for the OPTIKNEE consensus. *Br J Sports Med*. Dec 2022;56(24):1406-1421. [doi: [10.1136/bjsports-2022-105496](https://doi.org/10.1136/bjsports-2022-105496)] [Medline: [36455966](https://pubmed.ncbi.nlm.nih.gov/36455966/)]
6. Behan F, Bennett A, Schofield S, Miller E, Bull A. Osteoarthritis after major combat trauma: the ADVANCE study. *Osteoarthr Cartil*. Apr 2024;32:S265. [doi: [10.1016/j.joca.2024.02.382](https://doi.org/10.1016/j.joca.2024.02.382)]
7. Gwinnutt JM, Wiecek M, Balanescu A, et al. 2021 EULAR recommendations regarding lifestyle behaviours and work participation to prevent progression of rheumatic and musculoskeletal diseases. *Ann Rheum Dis*. Jan 2023;82(1):48-56. [doi: [10.1136/annrheumdis-2021-222020](https://doi.org/10.1136/annrheumdis-2021-222020)] [Medline: [35260387](https://pubmed.ncbi.nlm.nih.gov/35260387/)]

8. Dunn M, Rushton AB, Mistry J, Soundy A, Heneghan NR. The biopsychosocial factors associated with development of chronic musculoskeletal pain. An umbrella review and meta-analysis of observational systematic reviews. *PLoS One*. 2024;19(4):e0294830. [doi: [10.1371/journal.pone.0294830](https://doi.org/10.1371/journal.pone.0294830)] [Medline: [38557647](https://pubmed.ncbi.nlm.nih.gov/38557647/)]
9. Taanila H, Suni JH, Kannus P, et al. Risk factors of acute and overuse musculoskeletal injuries among young conscripts: a population-based cohort study. *BMC Musculoskelet Disord*. May 1, 2015;16:104. [doi: [10.1186/s12891-015-0557-7](https://doi.org/10.1186/s12891-015-0557-7)] [Medline: [25925549](https://pubmed.ncbi.nlm.nih.gov/25925549/)]
10. van Poppel D, van der Worp M, Slabbekoorn A, et al. Risk factors for overuse injuries in short- and long-distance running: a systematic review. *J Sport Health Sci*. Jan 2021;10(1):14-28. [doi: [10.1016/j.jsbs.2020.06.006](https://doi.org/10.1016/j.jsbs.2020.06.006)] [Medline: [32535271](https://pubmed.ncbi.nlm.nih.gov/32535271/)]
11. Evans SL, Owen R, Whittaker G, et al. Non-contact lower limb injuries in rugby union: a two-year pattern recognition analysis of injury risk factors. *PLoS One*. 2024;19(10):e0307287. [doi: [10.1371/journal.pone.0307287](https://doi.org/10.1371/journal.pone.0307287)] [Medline: [39446824](https://pubmed.ncbi.nlm.nih.gov/39446824/)]
12. da Costa BR, Vieira ER. Risk factors for work-related musculoskeletal disorders: a systematic review of recent longitudinal studies. *Am J Ind Med*. Mar 2010;53(3):285-323. [doi: [10.1002/ajim.20750](https://doi.org/10.1002/ajim.20750)] [Medline: [19753591](https://pubmed.ncbi.nlm.nih.gov/19753591/)]
13. Zebis MK, Andersen LL, Brandt M, et al. Effects of evidence-based prevention training on neuromuscular and biomechanical risk factors for ACL injury in adolescent female athletes: a randomised controlled trial. *Br J Sports Med*. May 2016;50(9):552-557. [doi: [10.1136/bjsports-2015-094776](https://doi.org/10.1136/bjsports-2015-094776)] [Medline: [26400955](https://pubmed.ncbi.nlm.nih.gov/26400955/)]
14. Schram B, Orr R, Pope R, Canetti E, Knapik J. Risk factors for development of lower limb osteoarthritis in physically demanding occupations: a narrative umbrella review. *J Occup Health*. Jan 2020;62(1):e12103. [doi: [10.1002/1348-9585.12103](https://doi.org/10.1002/1348-9585.12103)] [Medline: [31828895](https://pubmed.ncbi.nlm.nih.gov/31828895/)]
15. Moen BE, Koefoed VF, Bondevik K, Haukenes I. A survey of occupational health in the Royal Norwegian Navy. *Int Marit Health*. 2008;59(1-4):35-44. [Medline: [19227736](https://pubmed.ncbi.nlm.nih.gov/19227736/)]
16. GBD 2021 Other Musculoskeletal Disorders Collaborators. Global, regional, and national burden of other musculoskeletal disorders, 1990-2020, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. *Lancet Rheumatol*. 2023;5(11):e670-e682. [doi: [10.1016/S2665-9913\(23\)00232-1](https://doi.org/10.1016/S2665-9913(23)00232-1)] [Medline: [37927903](https://pubmed.ncbi.nlm.nih.gov/37927903/)]
17. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. Nov 10, 2018;392(10159):1789-1858. [doi: [10.1016/S0140-6736\(18\)32279-7](https://doi.org/10.1016/S0140-6736(18)32279-7)] [Medline: [30496104](https://pubmed.ncbi.nlm.nih.gov/30496104/)]
18. The state of musculoskeletal health 2024: arthritis and other musculoskeletal conditions in numbers. Versus Arthritis; 2024. URL: <https://www.scribd.com/document/792034138/the-state-of-musculoskeletal-health-2024> [Accessed 2026-07-04]
19. Murray CJL, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. Dec 15, 2012;380(9859):2197-2223. [doi: [10.1016/S0140-6736\(12\)61689-4](https://doi.org/10.1016/S0140-6736(12)61689-4)] [Medline: [23245608](https://pubmed.ncbi.nlm.nih.gov/23245608/)]
20. GBD 2017 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. Nov 10, 2018;392(10159):1859-1922. [doi: [10.1016/S0140-6736\(18\)32335-3](https://doi.org/10.1016/S0140-6736(18)32335-3)] [Medline: [30415748](https://pubmed.ncbi.nlm.nih.gov/30415748/)]
21. Hinton G. Deep learning—a technology with the potential to transform health care. *JAMA*. Sep 18, 2018;320(11):1101-1102. [doi: [10.1001/jama.2018.11100](https://doi.org/10.1001/jama.2018.11100)] [Medline: [30178065](https://pubmed.ncbi.nlm.nih.gov/30178065/)]
22. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. Apr 4, 2019;380(14):1347-1358. [doi: [10.1056/NEJMr1814259](https://doi.org/10.1056/NEJMr1814259)] [Medline: [30943338](https://pubmed.ncbi.nlm.nih.gov/30943338/)]
23. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. Jan 2019;25(1):44-56. [doi: [10.1038/s41591-018-0300-7](https://doi.org/10.1038/s41591-018-0300-7)] [Medline: [30617339](https://pubmed.ncbi.nlm.nih.gov/30617339/)]
24. Ipp E, Liljenquist D, Bode B, et al. Pivotal evaluation of an artificial intelligence system for autonomous detection of referable and vision-threatening diabetic retinopathy. *JAMA Netw Open*. Nov 1, 2021;4(11):e2134254. [doi: [10.1001/jamanetworkopen.2021.34254](https://doi.org/10.1001/jamanetworkopen.2021.34254)] [Medline: [34779843](https://pubmed.ncbi.nlm.nih.gov/34779843/)]
25. Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med*. Jun 2019;25(6):954-961. [doi: [10.1038/s41591-019-0447-x](https://doi.org/10.1038/s41591-019-0447-x)] [Medline: [31110349](https://pubmed.ncbi.nlm.nih.gov/31110349/)]
26. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature*. Jan 2020;577(7788):89-94. [doi: [10.1038/s41586-019-1799-6](https://doi.org/10.1038/s41586-019-1799-6)] [Medline: [31894144](https://pubmed.ncbi.nlm.nih.gov/31894144/)]
27. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med*. Nov 2018;24(11):1716-1720. [doi: [10.1038/s41591-018-0213-5](https://doi.org/10.1038/s41591-018-0213-5)] [Medline: [30349085](https://pubmed.ncbi.nlm.nih.gov/30349085/)]

28. Kadirvelu B, Gavriel C, Nageshwaran S, et al. A wearable motion capture suit and machine learning predict disease progression in Friedreich's ataxia. *Nat Med*. Jan 2023;29(1):86-94. [doi: [10.1038/s41591-022-02159-6](https://doi.org/10.1038/s41591-022-02159-6)] [Medline: [36658420](https://pubmed.ncbi.nlm.nih.gov/36658420/)]
29. Ricotti V, Kadirvelu B, Selby V, et al. Wearable full-body motion tracking of activities of daily living predicts disease trajectory in Duchenne muscular dystrophy. *Nat Med*. Jan 2023;29(1):95-103. [doi: [10.1038/s41591-022-02045-1](https://doi.org/10.1038/s41591-022-02045-1)] [Medline: [36658421](https://pubmed.ncbi.nlm.nih.gov/36658421/)]
30. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol*. Jan 2022;23(1):40-55. [doi: [10.1038/s41580-021-00407-0](https://doi.org/10.1038/s41580-021-00407-0)] [Medline: [34518686](https://pubmed.ncbi.nlm.nih.gov/34518686/)]
31. Ferreira FS, Ashburner J, Bouzigues A, et al. Identifying latent disease factors differently expressed in patient subgroups using group factor analysis. *arXiv*. Preprint posted online on Oct 10, 2024. [doi: [10.48550/arXiv.2410.07890](https://doi.org/10.48550/arXiv.2410.07890)]
32. Bennett AN, Dyball DM, Boos CJ, et al. Study protocol for a prospective, longitudinal cohort study investigating the medical and psychosocial outcomes of UK combat casualties from the Afghanistan war: the ADVANCE study. *BMJ Open*. Oct 30, 2020;10(10):e037850. [doi: [10.1136/bmjopen-2020-037850](https://doi.org/10.1136/bmjopen-2020-037850)] [Medline: [33127630](https://pubmed.ncbi.nlm.nih.gov/33127630/)]
33. McMenemy L, Behan FP, Kaufmann J, et al. Association between combat-related traumatic injury and skeletal health: bone mineral density loss is localized and correlates with altered loading in amputees: the Armed Services Trauma Rehabilitation Outcome (ADVANCE) study. *J Bone Miner Res*. Sep 2023;38(9):1227-1233. [doi: [10.1002/jbmr.4794](https://doi.org/10.1002/jbmr.4794)] [Medline: [37194399](https://pubmed.ncbi.nlm.nih.gov/37194399/)]
34. Watson FCE, Kedgley AE, Schofield S, et al. Upper limb function in people with upper and lower limb loss 8 years postinjury: the Armed Services Trauma Outcome Study (ADVANCE) cohort study. *Phys Ther*. Oct 2, 2024;104(10):pzae082. [doi: [10.1093/ptj/pzae082](https://doi.org/10.1093/ptj/pzae082)] [Medline: [38952004](https://pubmed.ncbi.nlm.nih.gov/38952004/)]
35. O'Sullivan O, Stocks J, Schofield S, et al. Association of serum biomarkers with radiographic knee osteoarthritis, knee pain and function in a young, male, trauma-exposed population—findings from the ADVANCE study. *Osteoarthritis Cartilage*. Dec 2024;32(12):1636-1646. [doi: [10.1016/j.joca.2024.07.016](https://doi.org/10.1016/j.joca.2024.07.016)] [Medline: [39103080](https://pubmed.ncbi.nlm.nih.gov/39103080/)]
36. Gender split for UK military and Op Herrick casualty and fatalities statistics October 2001 to December 2014 (Defence Statistics FOI2015/07390). Ministry of Defence; 2015. URL: https://assets.publishing.service.gov.uk/media/5a814a66e5274a2e87dbcd0c/PUBLIC_1442930645.pdf [Accessed 2026-07-04]
37. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthritis. *Ann Rheum Dis*. Dec 1957;16(4):494-502. [doi: [10.1136/ard.16.4.494](https://doi.org/10.1136/ard.16.4.494)] [Medline: [13498604](https://pubmed.ncbi.nlm.nih.gov/13498604/)]
38. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. *Osteoarthritis Cartilage*. 2007;15 Suppl A:A1-A56. [doi: [10.1016/j.joca.2006.11.009](https://doi.org/10.1016/j.joca.2006.11.009)] [Medline: [17320422](https://pubmed.ncbi.nlm.nih.gov/17320422/)]
39. Gailey RS, Roach KE, Applegate EB, et al. The Amputee Mobility Predictor: an instrument to assess determinants of the lower-limb amputee's ability to ambulate. *Arch Phys Med Rehabil*. May 2002;83(5):613-627. [doi: [10.1053/apmr.2002.32309](https://doi.org/10.1053/apmr.2002.32309)] [Medline: [11994800](https://pubmed.ncbi.nlm.nih.gov/11994800/)]
40. Poddubnyy D, van Tubergen A, Landewé R, Sieper J, van der Heijde D. Assessment of SpondyloArthritis international Society (ASAS). Development of an ASAS-endorsed recommendation for the early referral of patients with a suspicion of axial spondyloarthritis. *Ann Rheum Dis*. Aug 2015;74(8):1483-1487. [doi: [10.1136/annrheumdis-2014-207151](https://doi.org/10.1136/annrheumdis-2014-207151)] [Medline: [25990288](https://pubmed.ncbi.nlm.nih.gov/25990288/)]
41. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. Dec 2011;20(10):1727-1736. [doi: [10.1007/s11136-011-9903-x](https://doi.org/10.1007/s11136-011-9903-x)] [Medline: [21479777](https://pubmed.ncbi.nlm.nih.gov/21479777/)]
42. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy*. Aug 1980;66(8):271-273. [Medline: [6450426](https://pubmed.ncbi.nlm.nih.gov/6450426/)]
43. Roos EM, Lohmander LS. The Knee Injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. *Health Qual Life Outcomes*. Nov 3, 2003;1:64. [doi: [10.1186/1477-7525-1-64](https://doi.org/10.1186/1477-7525-1-64)] [Medline: [14613558](https://pubmed.ncbi.nlm.nih.gov/14613558/)]
44. Germann G, Wind G, Harth A. The DASH (Disability of Arm-Shoulder-Hand) Questionnaire—a new instrument for evaluating upper extremity treatment outcome [Article in German]. *Handchir Mikrochir Plast Chir*. May 1999;31(3):149-152. [doi: [10.1055/s-1999-13902](https://doi.org/10.1055/s-1999-13902)] [Medline: [10420282](https://pubmed.ncbi.nlm.nih.gov/10420282/)]
45. Craig CL, Marshall AL, Sjöström M, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc*. Aug 2003;35(8):1381-1395. [doi: [10.1249/01.MSS.0000078924.61453.FB](https://doi.org/10.1249/01.MSS.0000078924.61453.FB)] [Medline: [12900694](https://pubmed.ncbi.nlm.nih.gov/12900694/)]
46. Ryall NH, Eyres SB, Neumann VC, Bhakta BB, Tennant A. The SIGAM mobility grades: a new population-specific measure for lower limb amputees. *Disabil Rehabil*. Aug 5, 2003;25(15):833-844. [doi: [10.1080/0963828021000056460](https://doi.org/10.1080/0963828021000056460)] [Medline: [12851094](https://pubmed.ncbi.nlm.nih.gov/12851094/)]
47. Hanspal RS, Fisher K, Nieveen R. Prosthetic socket fit comfort score. *Disabil Rehabil*. Nov 18, 2003;25(22):1278-1280. [doi: [10.1080/09638280310001603983](https://doi.org/10.1080/09638280310001603983)] [Medline: [14617445](https://pubmed.ncbi.nlm.nih.gov/14617445/)]
48. Colangelo PM, Welch DW, Rich DS, Jeffrey LP. Two methods for estimating body surface area in adult amputees. *Am J Hosp Pharm*. Dec 1984;41(12):2650-2655. [Medline: [6517088](https://pubmed.ncbi.nlm.nih.gov/6517088/)]

49. Haycock GB, Schwartz GJ, Wisotsky DH. Geometric method for measuring body surface area: a height-weight formula validated in infants, children, and adults. *J Pediatr*. Jul 1978;93(1):62-66. [doi: [10.1016/s0022-3476\(78\)80601-5](https://doi.org/10.1016/s0022-3476(78)80601-5)] [Medline: [650346](https://pubmed.ncbi.nlm.nih.gov/650346/)]
50. Virtanen S, Klami A, Khan SA, Kaski S. Bayesian group factor analysis. Presented at: 15th International Conference on Artificial Intelligence and Statistics (AISTATS); Apr 21-23, 2012; La Palma, Canary Islands, Spain. URL: <https://proceedings.mlr.press/v22/virtanen12.html> [Accessed 2026-06-30]
51. Piironen J, Vehtari A. Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electron J Statist*. 2017;11(2):5018-5051. [doi: [10.1214/17-EJS1337SI](https://doi.org/10.1214/17-EJS1337SI)]
52. Bunte K, Leppäaho E, Saarinen I, Kaski S. Sparse group factor analysis for biclustering of multiple data sources. *Bioinformatics*. Aug 15, 2016;32(16):2457-2463. [doi: [10.1093/bioinformatics/btw207](https://doi.org/10.1093/bioinformatics/btw207)] [Medline: [27153643](https://pubmed.ncbi.nlm.nih.gov/27153643/)]
53. Vollert J, Kumar A, Coady EC, et al. Pain after combat injury in male UK military personnel deployed to Afghanistan. *Br J Anaesth*. Jun 2024;132(6):1285-1292. [doi: [10.1016/j.bja.2024.02.019](https://doi.org/10.1016/j.bja.2024.02.019)] [Medline: [38521656](https://pubmed.ncbi.nlm.nih.gov/38521656/)]
54. Watson FCE, N Bennett A, McGregor A, et al. Biopsychosocial factors and low back pain in military personnel with lower limb loss: the ADVANCE study. *BMJ Mil Health*. Jun 20, 2025;20. [doi: [10.1136/military-2025-002962](https://doi.org/10.1136/military-2025-002962)] [Medline: [40541397](https://pubmed.ncbi.nlm.nih.gov/40541397/)]
55. Watson FCE, Agricola R, Bennett A, Bull A. Prevalence, and 3-year progression and incidence of radiographic hip osteoarthritis in a young military population post major trauma: the ADVANCE study. *Osteoarthritis Cartil*. Apr 2025;33(Supplement):S56-S57. [doi: [10.1016/j.joca.2025.02.076](https://doi.org/10.1016/j.joca.2025.02.076)]
56. Blüher M. Obesity: global epidemiology and pathogenesis. *Nat Rev Endocrinol*. May 2019;15(5):288-298. [doi: [10.1038/s41574-019-0176-8](https://doi.org/10.1038/s41574-019-0176-8)] [Medline: [30814686](https://pubmed.ncbi.nlm.nih.gov/30814686/)]
57. Warden D. Military TBI during the Iraq and Afghanistan wars. *J Head Trauma Rehabil*. 2006;21(5):398-402. [doi: [10.1097/00001199-200609000-00004](https://doi.org/10.1097/00001199-200609000-00004)] [Medline: [16983225](https://pubmed.ncbi.nlm.nih.gov/16983225/)]
58. McPherson AL, Nagai T, Webster KE, Hewett TE. Musculoskeletal injury risk after sport-related concussion: a systematic review and meta-analysis. *Am J Sports Med*. Jun 2019;47(7):1754-1762. [doi: [10.1177/0363546518785901](https://doi.org/10.1177/0363546518785901)] [Medline: [30074832](https://pubmed.ncbi.nlm.nih.gov/30074832/)]
59. Reneker JC, Babl R, Flowers MM. History of concussion and risk of subsequent injury in athletes and service members: a systematic review and meta-analysis. *Musculoskelet Sci Pract*. Jul 2019;42:173-185. [doi: [10.1016/j.msksp.2019.04.004](https://doi.org/10.1016/j.msksp.2019.04.004)] [Medline: [31014921](https://pubmed.ncbi.nlm.nih.gov/31014921/)]
60. Roach MH, Aderman MJ, Ross JD, et al. Risk of upper extremity musculoskeletal injury within the first year after a concussion. *Orthop J Sports Med*. May 2023;11(5):23259671231163570. [doi: [10.1177/23259671231163570](https://doi.org/10.1177/23259671231163570)] [Medline: [37197033](https://pubmed.ncbi.nlm.nih.gov/37197033/)]
61. Brown S, Hawker G, Beaton D, Colantonio A. Long-term musculoskeletal complaints after traumatic brain injury. *Brain Inj*. 2011;25(5):453-461. [doi: [10.3109/02699052.2011.556581](https://doi.org/10.3109/02699052.2011.556581)] [Medline: [21401368](https://pubmed.ncbi.nlm.nih.gov/21401368/)]
62. Shumski EJ, Oh J, Schmidt JD, Lynall RC. Trunk and lower extremity biomechanics in female athletes with and without a concussion history. *J Athl Train*. Jul 1, 2024;59(7):751-761. [doi: [10.4085/1062-6050-0259.23](https://doi.org/10.4085/1062-6050-0259.23)] [Medline: [37681680](https://pubmed.ncbi.nlm.nih.gov/37681680/)]
63. Shumski EJ, Schmidt JD, Lynall RC. Cognition uniquely influences dual-task tandem gait performance among athletes with a concussion history. *Sports Health*. 2024;16(4):542-550. [doi: [10.1177/19417381231183413](https://doi.org/10.1177/19417381231183413)] [Medline: [37377161](https://pubmed.ncbi.nlm.nih.gov/37377161/)]
64. Nawi MAA, Noor NFM, Shaari R, et al. The patterns of facial fractures in traumatic brain injury (TBI) patients using ordinal regression: a retrospective study of five years. *AIMS Neurosci*. 2022;9(3):345-357. [doi: [10.3934/Neuroscience.2022019](https://doi.org/10.3934/Neuroscience.2022019)] [Medline: [36329904](https://pubmed.ncbi.nlm.nih.gov/36329904/)]
65. Feng YS, Kohlmann T, Janssen MF, Buchholz I. Psychometric properties of the EQ-5D-5L: a systematic review of the literature. *Qual Life Res*. Mar 2021;30(3):647-673. [doi: [10.1007/s11136-020-02688-y](https://doi.org/10.1007/s11136-020-02688-y)] [Medline: [33284428](https://pubmed.ncbi.nlm.nih.gov/33284428/)]
66. Ferreirafabio80. GitHub. URL: <https://github.com/ferreirafabio80> [Accessed 2026-06-30]

Abbreviations

ADVANCE: The Armed Services Trauma and Rehabilitation Outcome
AI: artificial intelligence
DEXA: dual-energy X-ray absorptiometry
EQ-5D-5L: EuroQoL 5-Dimensions 5-Level
GFA: group factor analysis
JTTR : Joint Theatre Trauma Registry
NISS: New Injury Severity Score
TBI: traumatic brain injury

Edited by Ivan Steenstra; peer-reviewed by Ibukun Abidoye, Oluwaseun Ajayi, Zhao Liu; submitted 23.Jan.2026; final revised version received 23.Jun.2026; accepted 23.Jun.2026; published 21.Jul.2026

Please cite as:

Watson FCE, Ferreira FS, Kadirvelu B, Bennett AN, Faisal AA, Graham N, Kemp H, Cullinan P, Boos C, Fear NT, Bull AMJ

Novel Musculoskeletal Hypotheses in the Armed Services Trauma and Rehabilitation Outcome (ADVANCE) Cohort: Development and Application of Sparse Group Factor Analysis Methodology

J Med Internet Res 2026;28:e91958

URL: <https://www.jmir.org/2026/1/e91958>

doi: [10.2196/91958](https://doi.org/10.2196/91958)

© Fraje C E Watson, Fabio S Ferreira, Balasundaram Kadirvelu, Alex N Bennett, Aldo A Faisal, Neil Graham, Harriet Kemp, Paul Cullinan, Christopher Boos, Nicola T Fear, Anthony M J Bull. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 21.Jul.2026. 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 use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.