Biological Measures and Diagnostic Tools for Gulf War Illness: A Systematic Review

September 2020

Prepared for:
Department of Veterans Affairs
Veterans Health Administration
Health Services Research & Development Service
Washington, DC 20420

Prepared by:
Evidence Synthesis Program (ESP) Center
Portland VA Health Care System
Portland, OR
Devan Kansagara, MD, MCR, Director

Authors:
Principal Investigator:
Emily G. Gean, PhD

Co-Investigators:
Chelsea K. Ayers, MPH
Kara A. Winchell, MA
Michele Freeman, MPH
Ashlyn M. Press, MPH
Robin Paynter, MLIS
Devan Kansagara, MD, MCR
Shannon M. Nugent, PhD

VA | U.S. Department of Veterans Affairs
Veterans Health Administration
Health Services Research & Development Service

PREFACE

The VA Evidence Synthesis Program (ESP) was established in 2007 to provide timely and accurate syntheses of targeted health care topics of importance to clinicians, managers, and policymakers as they work to improve the health and health care of Veterans. These reports help:

- Develop clinical policies informed by evidence;
- Implement effective services to improve patient outcomes and to support VA clinical practice guidelines and performance measures; and
- Set the direction for future research to address gaps in clinical knowledge.

The program is comprised of three ESP Centers across the US and a Coordinating Center located in Portland, Oregon. Center Directors are VA clinicians and recognized leaders in the field of evidence synthesis with close ties to the Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice Center Program and Cochrane Collaboration. The Coordinating Center was created to manage program operations, ensure methodological consistency and quality of products, and interface with stakeholders. To ensure responsiveness to the needs of decision-makers, the program is governed by a Steering Committee comprised of health system leadership and researchers. The program solicits nominations for review topics several times a year via the program website.

Comments on this evidence report are welcome and can be sent to Nicole Floyd, Deputy Director, ESP Coordinating Center at Nicole.Floyd@va.gov.

Recommended citation: Gean E, Ayers CK, Winchell KA, Freeman M, Press AM, Paynter R, Kansagara D, and Nugent SM. Biological measures and Diagnostic Tools for Gulf War Illness - A Systematic Review. Washington, DC: Evidence Synthesis Program, Health Services Research and Development Service, Office of Research and Development, Department of Veterans Affairs. VA ESP Project #05-225; 2020. Available at: https://www.hsrd.research.va.gov/publications/esp/reports.cfm.

This report is based on research conducted by the Evidence Synthesis Program (ESP) Center located at the located at the Portland VA Health Care System, Portland, OR, funded by the Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development. The findings and conclusions in this document are those of the author(s) who are responsible for its contents; the findings and conclusions do not necessarily represent the views of the Department of Veterans Affairs or the United States government. Therefore, no statement in this article should be construed as an official position of the Department of Veterans Affairs. No investigators have any affiliations or financial involvement (*eg*, employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties) that conflict with material presented in the report.

ACKNOWLEDGMENTS

This topic was developed in response to a nomination by Karen Block, PhD, Director of Gulf War Research in the Veterans Affairs (VA) Office of Research and Development (ORD) Gulf War Research Program, for the purpose of informing the planning for a state-of-the-art meeting on Gulf War Research and providing guidance for ORD funding priorities in Gulf War research. The scope was further developed with input from the topic nominators (*ie*, Operational Partners), the ESP Coordinating Center, the review team, and the Technical Expert Panel (TEP).

In designing the study questions and methodology at the outset of this report, the ESP consulted several technical and content experts. Broad expertise and perspectives were sought. Divergent and conflicting opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant systematic review. Therefore, in the end, study questions, design, methodologic approaches, and/or conclusions do not necessarily represent the views of individual technical and content experts.

The authors gratefully acknowledge the following individuals for their contributions to this project:

Operational Partners

Operational partners are system-level stakeholders who have requested the report to inform decision-making. They recommend TEP participants; assure VA relevance; help develop and approve final project scope and timeframe for completion; provide feedback on draft report; and provide consultation on strategies for dissemination of the report to field and relevant groups.

Karen Block, PhD
Office of Research and Development (ORD) (10P9) – Gulf War Research Program
Washington, DC

Drew Helmer, MD, MS
Deputy Director, Center for Innovations in Quality, Effectiveness and Safety (IQuESt)
Houston, TX

Technical Expert Panel (TEP)

To ensure robust, scientifically relevant work, the TEP guides topic refinement; provides input on key questions and eligibility criteria, advising on substantive issues or possibly overlooked areas of research; assures VA relevance; and provides feedback on work in progress. TEP members are listed below:

Mark Helfand, MD, MPH, MS
VA Portland Health Care System
Portland, OR

Arash Javanbakht, MD
Wayne State University
Detroit, MI

Eva Lee, PhD
Georgia Institute of Technology
Atlanta, GA

Kristy B. Lidie, PhD
Department of Defense - Congressionally Directed Medical Research Programs (CDMRP)
Fort Detrick, MD

Peer Reviewers

The Coordinating Center sought input from external peer reviewers to review the draft report and provide feedback on the objectives, scope, methods used, perception of bias, and omitted evidence. Peer reviewers must disclose any relevant financial or non-financial conflicts of interest. Because of their unique clinical or content expertise, individuals with potential conflicts may be retained. The Coordinating Center and the ESP Center work to balance, manage, or mitigate any potential nonfinancial conflicts of interest identified.

TABLE OF CONTENTS

TABLES

FIGURES

ABSTRACT

Background: Gulf War Illness (GWI) is a chronic multisymptom illness comprised of a wide range of systemic symptoms and functional impairments. We conducted a systematic review to catalogue studies (both published and unpublished/ongoing) of validated biological tests for diagnosing GWI and studies of associations between biological measures and GWI for their promise as biomarkers.

Materials and Methods: We searched multiple electronic databases, clinical trial registries, and reference lists through February 2020 for all observational studies of diagnostic tests of GWI and completed or ongoing studies of associations between biological measures and GWI. We abstracted data on study design, demographics, and outcomes. Two reviewers independently assessed the risk of bias of included studies using established methods.

Results: We did not identify any studies validating tests of biomarkers that distinguish cases of GWI from non-cases. We included 32 completed and 24 ongoing or unpublished studies of associations between GWI and biological measures that included comparator groups that provided the most useful information. Studies (n=77) with other comparator groups, no comparator group, or with N<25 were included in a table. Considering all studies, most fell within the central nervous and immune systems and indicated a significant association of the biological measure with GWI case status. Biological measures were heterogeneous across studies.

Conclusion: Our review indicates that there are no existing validated biological tests to determine GWI case status. Many studies have assessed the potential association between a variety of biological measures and GWI, the majority of which pertain to the immune and central nervous systems. More importantly, while most studies indicated a significant association between biological measures and GWI case status, the biological measures across studies were extremely heterogeneous. Due to the great heterogeneity, the focus of the review is to map out what has been examined, rather than synthesize information.

EXECUTIVE SUMMARY

AIM

We conducted a systematic review to catalogue studies (both published and unpublished/ongoing) of validated biological tests for diagnosing GWI and studies of associations between biological measures and GWI for their promise as biomarkers.

METHODS

We searched electronic databases (Ovid MEDLINE, and Ovid PsycINFO, and Ovid EBM Reviews [Cochrane Database of Systematic Reviews and Cochrane Central Register of Controlled Trials]) through February 20, 2020 for all observational studies of diagnostic tests of GWI and completed or ongoing studies of associations between biological measures and GWI. To identify in-progress or unpublished studies, we searched ClinicalTrials.gov and the World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP). We reviewed the bibliographies of relevant articles, contacted experts, and reviewed lists of funded trials from the Department of Defense (DoD) and Veterans Affairs (VA) to identify additional studies.

We included studies of biological measures that have been examined for their promise as diagnostic biomarkers for and/or their association with GWI. We included completed and ongoing/unpublished studies of Veterans with GWI, identified using any GWI diagnostic criteria, in which the comparator population members were Veterans deployed to the Persian Gulf theater during the First Gulf War who did not develop GWI, with or without comorbid conditions (*eg*, posttraumatic stress disorder, anxiety). We excluded studies of Veterans with GWI compared to other populations (*eg*, non-deployed Gulf War Veterans, civilians), and those with insufficient sample sizes (N<25) but did summarize these studies in the Appendix to the full report.

For all included studies, we abstracted data on study design, sample size, population characteristics, case definition, comparator(s), participant inclusion and exclusion criteria, details of the biological measure of interest, and findings for measures of association as available. Data abstraction was confirmed by a second reviewer. Two reviewers independently assessed the risk of bias of included studies using established methods. Discordant ratings were resolved by consensus or an additional reviewer.

RESULTS

We identified no studies of diagnostic tests for GWI. We identified 56 studies of associations between GWI and biological measures (32 completed and 24 ongoing or unpublished).

Key Question 1: Which diagnostic tests (or test combinations) are candidates for distinguishing individuals diagnosed with GWI from individuals without GWI?

We did not identify any studies validating tests of biomarkers that distinguish cases of GWI from non-cases, regardless of the diagnostic criteria used in the study.

Key Question 2: Which biological measures have been examined for their potential association with GWI, and which among them have been shown to be associated with GWI?

We identified 32 studies of biological measures that have been examined for their association with GWI and grouped them broadly into categories under distinct biological systems: 10 studies of immune system biological measures; 10 central nervous system (CNS) studies; 5 studies of autonomic nervous system (ANS) biological measures; 1 of genetic biological measures; and 6 studies of biological measures in other biological systems (See Figure i). We also briefly summarized an additional 77 studies that examined biological measures, but do not include an ideal control group (n=72 studies) or had a sample size less than 25 (n=5 studies; See Appendix D of full report), indicated in the grey areas in Figure i.

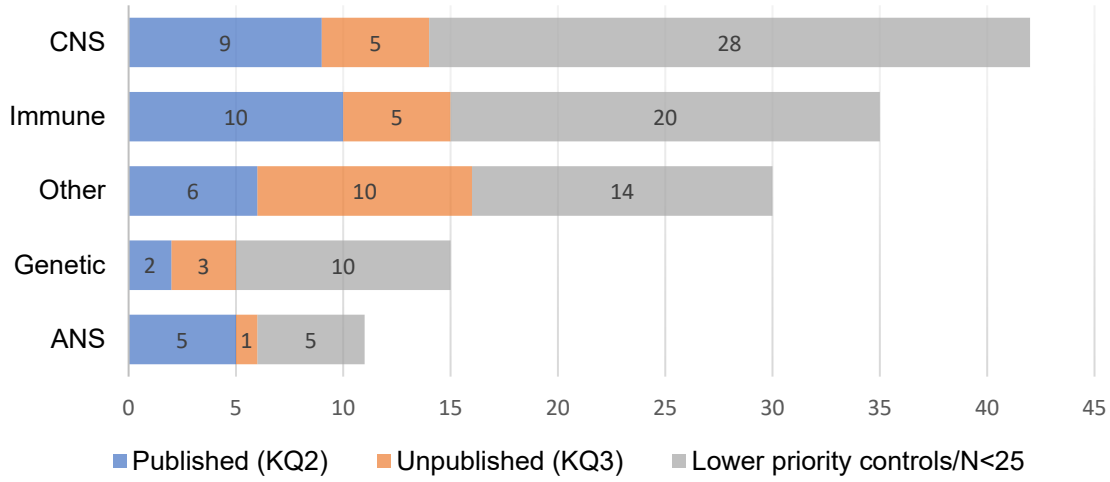
Only 1 included study used a comparator population of deployed GWVs with health conditions other than GWI; in all other studies health conditions were not reported and participants were presumed healthy.

All studies had additional methodological shortcomings. Biological measures were heterogeneous across studies, even those categorized within the same biological system, with the exception of 2 studies that had 1 replication study each.

Key Question 3: Which ongoing or unpublished research studies examine diagnostic tests or biological measures for potential association with GWI?

We did not identify any ongoing/unpublished studies examining diagnostic tests for GWI. We found 24 ongoing or unpublished studies examining biological measures for their potential association with GWI. Similar to the completed studies many are investigating measures of the immune and central nervous system (see Figure i).

Figure i. Number of studies of GWI biological measures by biological system



Abbreviations: ANS=Autonomic Nervous System; CNS=Central Nervous System; KQ=Key Question

CONCLUSION

In the current review, we sought to evaluate studies validating existing diagnostic tests for GWI, and to determine whether biological measurements with promise for further establishment as biomarkers either in completed or ongoing/upcoming studies have been demonstrated. The establishment of biological measures for GWI would allow for increased accuracy in diagnosis and potential mechanisms for treatment.

Our review indicates that there are no existing validated biological tests to determine GWI case status. It is not surprising that no such studies were found, because the case definition for GWI is still debated. In the absence of a gold standard definition or diagnostic test, the determination of biological measures to distinguish a case from a non-case is challenging.

We did identify many studies that have assessed, or are currently assessing, the potential association between a variety of biological measures and GWI. Most of the studies could be characterized as "biomarker discovery" studies and were largely designed to shed light on the potential causes of GWI. Our review indicates that biological measures within the immune and central nervous systems have more often been investigated for their potential relationship with GWI, consistent with some dominant theories of disease etiology and dysfunction, but the literature also suggests other avenues of inquiry in upcoming studies, such as the gut microbiome. More importantly, our review revealed that existing studies are insufficient for determining promising biomarkers due to the extent of heterogeneity in biological measures across studies, inadequate comparator groups, and several other methodological limitations. Future studies that employ ideal control groups, reproduce findings of existing studies, and otherwise apply rigorous methodological practices and reporting specifically appropriate for investigating potential biomarkers would contribute to the establishment of a base of targeted, highly reliable studies from which lines of investigation could grow.

ABBREVIATIONS TABLE

Abbreviation	Definition
AA	Arachidonic Acid or African American
ACE	Angiotensin-Converting Enzyme
ACh	Acetylcholine
ADP	Adenosine Diphosphate
AHI	Apnea-Hypopnea Index
AHRQ	Agency for Healthcare Research and Quality
ANA	Antinuclear Antibody
ANS	Autonomic Nervous System
apoAI	Plasma Apolipoprotein AI
ASA	Anti-Squalene Antibody
ASL	Arterial Spin Labelling
ATP	Adenosine Triphosphate
AUC	Area Under the Curve
BBB	Blood-Brain Barrier
BBRAIN	Boston Biorepository and Integrated Network
BDNF	Brain-Derived Neurotrophic Factor
BIDMC	Beth Israel Deaconess Medical Center
BOLD	Blood-Oxygen-Level-Dependent
BP	Blood Pressure
BPM	Beats-Per-Minute
BRINM	Biomedical Research Institute of New Mexico
BuChE	Butyrylchoninesterase
CASS	Composite Autonomic Severity Score
CBF	Cerebral Blood Flow
CCEP	Comprehensive Clinical Evaluation Program
CDC	Centers for Disease Control and Prevention
CDMRP	Congressionally Directed Medical Research Program
CFS	Chronic Fatigue Syndrome
CI	Confidence Interval
CMAP	Compound Motor Action Potential
CMI	Chronic Multisymptom Illness
CNDP1	Carnosine Dipeptidase 1
CNS	Central Nervous System
CO2	Carbon Dioxide
CpG	Cytidine-Phosphateguanosine
CPTh	Cold Pressor Threshold
CRP	C-Reactive Protein
CSF	Cerebrospinal Fluid
CSP	Cooperative Studies Program
CSRD	Clinical Sciences Research and Development Service

CV	Conduction Velocity
D-GWV	Deployed Gulf War Veteran
DF	Degrees of Freedom
DHA	Docosahexaenoic Acid
DKI	Diffusion Kurtosis Imaging
DLPFC	Dorsolateral Prefrontal Cortex
DNA	Deoxyribonucleic Acid
DoD	Department of Defense
DTI	Diffusion Tensor Imaging
dUTPase	Deoxyuridine Triphosphate Diphosphatase
EBV	Epstein-Barr Virus
ECG	Electrocardiogram
ECL	Electrochemiluminescence
EEG	Electroencephalograph
ELISA	Enzyme-Linked Immunosorbent Assay
EMG	Electromyography
ERP	Event Related Potential
ESP	Evidence Synthesis Program
FM	Fibromyalgia
fMRI	Functional Magnetic Resonance Imaging
GI	Gastrointestinal
GW	Gulf War
GWI	Gulf War Illness
GWIC	Gulf War Illness Consortium
GWIRP	Gulf War Illness Research Program
GWV	Gulf War Veteran
HARDI	High-Angular Resolution Diffusion Imaging
HCMV	Human Cytomegalovirus
HDL	High-Density Lipoprotein
H-FABP	Heart-type Fatty Acid Binding Protein
HHV	Human Herpesvirus
HLA	Human Leukocyte Antigen
HPA	Hypothalamic-Pituitary-Adrenal
HPTh	Heat Pain Threshold
HR	Heart Rate
HRV	Heart Rate Variability
HUT	Head-Up Tilt
IBS	Irritable Bowel Syndrome
ICA	Independent Component Analysis
ICD-9-CM	International Classification of Diseases, Ninth Revision, Clinical Modification
IFN	Interferon
Ig	Immunoglobulin
IL	Interleukin

IoM	Institutes of Medicine
IPF	Immature Platelet Fraction
IQuESt	Center for Innovations in Quality, Effectiveness and Safety
KQ	Key Question
LF	Low Frequency
LLOD	Lowest Level of Detection
LPC	Lysophosphatidylcholines
LPE	Lysophosphatidylethanolamine
LPS	Lipopolysaccharide
MAP	Multi-Analyte Profile
MCD	Mean Consecutive Difference
ME	Myalgic Encephalitis
MEG	Magnetoencephalograph
MiRNA	Micro Ribonucleic Acid
MMP	Matrix Metalloproteinase
MN	Minnesota
MPF	Macromolecular Proton Fraction
MPV	Mean Platelet Volume
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
mRNA	Messenger Ribonucleic Acid
MRS	Magnetic Resonance Spectroscopy
MRSI	Magnetic Resonance Spectroscopic Imaging
MSD	Meso Scale Discovery
MSI	Multisymptom Illness
mt	Mitochondria
mtDNA	Mitochondrial Deoxyribonucleic Acid
MUAP	Motor Unit Action Potential
MUFA	Monosaturated Fatty Acids
MVP	Million Veteran Program
NAA	N-acetyl Aspartate
NCM	Neurocutaneous Melanocytosis
NCT	National Clinical Trial
ND	No Difference
NIH	National Institutes of Health
NPV	Negative Predictive Value
NR	Not Reported
NREM	Non-Rapid Eye Movement
OR	Odds Ratio
ORD	Office of Research and Development
OS	Oxidative Stress
PBMC	Peripheral Mononuclear Cell
PC	Phosphatidylcholine

PCr	Creatine Phosphate
PCR	Polymerase Chain Reaction
PE	Phosphatidylethanolamine
PEF	Post Exertional Fatigue
PEM	Post Exertional Malaise
PET	Positron Emission Tomography
Pi	Inorganic Phosphate
PI	Principal Investigator or Phosphatidylinositol
PICOTS	Population, Intervention, Comparator, Outcome, Timing, Setting, and Study Design
PL	Phospholipid
PON1	Paraoxanase 1
PPV	Positive Predictive Value
PRP	Platelet-Rich Plasma
PSC	Percent Signal Change
PSG	Polysomnograph
PTSD	Posttraumatic Stress Disorder
PUFA	Polyunsaturated Fatty Acids
RCT	Randomized Controlled Trial
REM	Rapid Eye Movement
RERA	Respiratory Effort Related Arousal
RNA	Ribonucleic Acid
ROC	Receiver Operating Characteristic
SD	Standard Deviation
SEM	Standard Error of the Mean
SF-36	Short Form-36 Item
SFA	Saturated Fatty Acids
SFVAFRE	South Florida Veterans Affairs Foundation for Research and Education
SNP	Single-Nucleotide Polymorphism
SPRC	Standardized Partial Regression Coefficient
TCR	T-Cell Receptor
TEP	Technical Expert Panel
TNF	Tumor Necrosis Factor
TPO	Plasma Thrombopoietin
TRAP	Thrombin Receptor Agonist Peptide
UC	University of California
UK	United Kingdom
US	United States
USACHPPM	United States Army Center for Health Promotion and Preventive Medicine
USAMRDC	United States Army Medical Research and Materiel Command
VA	Veterans Affairs
VAMC	Veterans Affairs Medical Center
VHA	Veterans Health Administration
VLF	Very Low Frequency

VO2	Oxygen Volume
WM	White Matter
XMRV	Xenotropic Murine Leukemia Virus-Related Virus

EVIDENCE REPORT

INTRODUCTION

The setting of the 1990-1991 conflict in the Persian Gulf was fraught with potential exposure to multiple toxins and stressors, including environmental and chemical exposures (*eg*, solvents, depleted uranium, excessive heat, oil-well fire smoke, pesticides, nerve agents and their prophylaxis, and vaccines) and psychological stressors associated with a combat setting (*eg*, separation from family and work, uncertainty about presence of chemical and biological agents, and witnessing dead or wounded soldiers and citizens).[1] After the conflict, many Gulf War Veterans began reporting numerous unexplained symptoms. These symptom clusters included fatigue, headaches, joint pain, indigestion, insomnia, dizziness, respiratory disorders, and memory problems.

Clusters of symptoms were categorized into 6 definitions between 1997 and 2009.[2-7] In 2014, the Institutes of Medicine (IoM) determined that, together, 2 definitions best captured the symptoms—the Centers for Disease Control and Prevention (CDC) and Kansas definitions.[8] The common symptoms in all studies reviewed by the committee in the selection of the definition included reports of fatigue, pain, and neurocognitive symptoms. The CDC definition requires at least 1 symptom from 2 of 3 categories (fatigue, mood and cognition, and musculoskeletal) for 6 months or longer.[3] The Kansas approach defines a case as having 1 moderately severe, or 2 or more chronic symptoms in at least 3 of 6 domains (including fatigue or sleep; pain; neurologic, cognitive, or mood; gastrointestinal; respiratory; and skin).[7] The proportion of Gulf War-deployed Veterans who meet case criteria for Gulf War Illness (GWI) is approximately 34% (based on the Kansas case definition) to 50% (based on the CDC case definition).[8]

Yet these definitions remain disputed and outdated, and GWI is still a largely medically unexplained chronic multisymptom illness, due to its widely varying symptoms, reliance on self-reported symptoms for diagnosis, and lack of understanding of its pathophysiology and etiology.[9]

The identification of promising biological markers could help to refine the illness definitions; better detect, predict, or distinguish subgroups of GWI; and ultimately lead to the development of biologically plausible treatments. Identification and treatment of GWI is a top research priority for the VA Office of Research and Development (ORD). While causes of GWI symptomology and areas of biological dysfunction are unknown, investigation of biological systems underlying GWI symptomology has been partially driven by hypotheses involving exposure to acetylcholinesterase-inhibiting toxins – such as nerve agents, Pyridostigmine bromide pills, and pesticides – that are thought to contribute to dysfunction in energy metabolism (mitochondria), leading to dysfunction in cholinergic systems[10-12] and to chronic inflammation that damages the central nervous system (CNS).[13-15] Most studies to date have included biological markers that have focused around the biological systems (*eg,* central nervous system, immune system, and autonomic nervous system, as well as genes that regulate these systems) that are involved in the aforementioned hypothesized GWI mechanisms.

This systematic review provides an overview of the state of biomarker research in GWI. It is not intended to directly guide clinical care but is intended to help future research planning efforts by

summarizing what the field has already examined and by identifying any potentially promising areas of biomarker research. We examined the literature to identify studies of the use of biomarkers as a diagnostic test for GWI, and also to identify studies that evaluated the differential expression of biomarkers among participants with and without GWI.

METHODS

TOPIC DEVELOPMENT

The key research questions for the review were as follows:

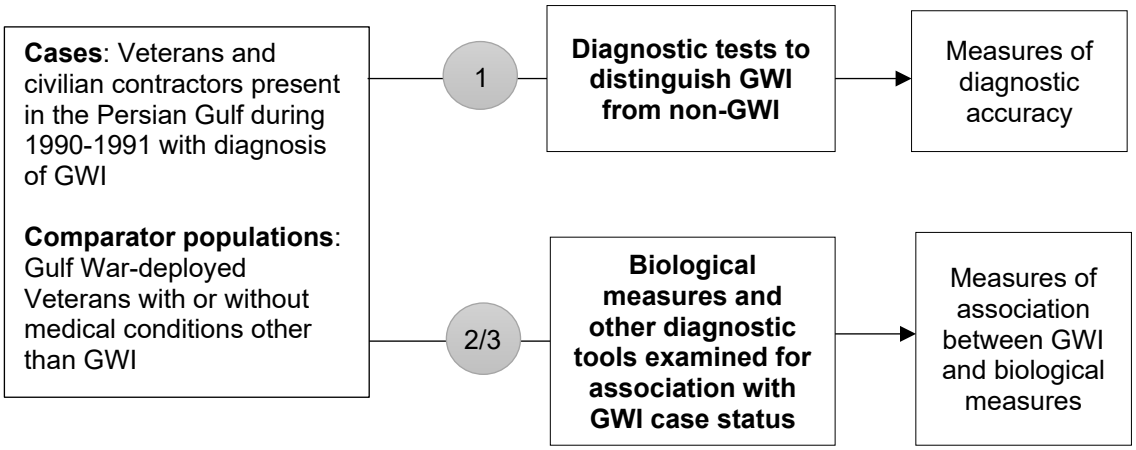
1. Which diagnostic tests (or test combinations) are candidates for distinguishing individuals diagnosed with GWI from individuals without GWI?

2. Which biological measures have been examined for their potential association with GWI, and which among them have been shown to be associated with GWI?

3. Which ongoing or unpublished research studies examine diagnostic tests or biological measures for potential association with GWI?

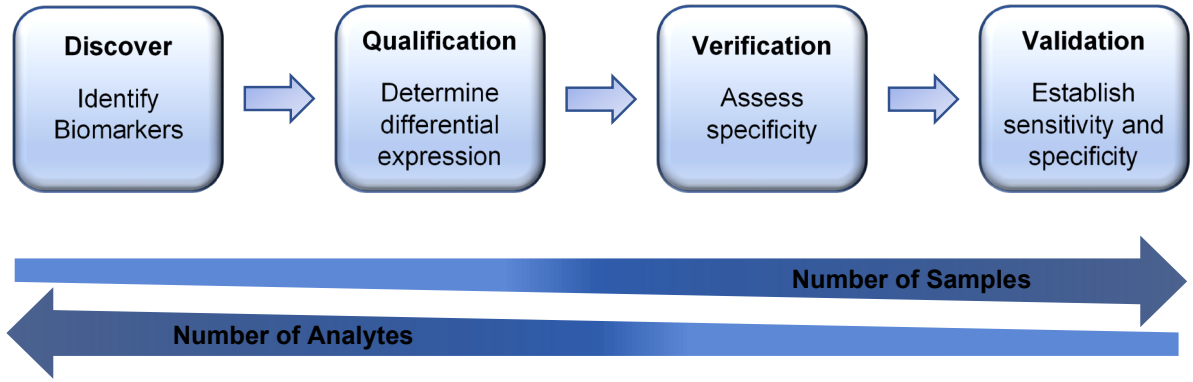
Our approach was guided by our process framework, which we developed in consultation with our operational partners (Figure 1). The protocol was posted to a publicly accessible website prior to commencing the review (PROSPERO registration ID: CRD42020169099).

Figure 1. Process framework



Abbreviations: GWI=Gulf War Illness

Figure 2. Biomarker development Process



Note. Reproduced from Fall et al, 2014[16]

The biomarker development process (Figure 2; reproduced from Fall et al, 2014[16]) provides a framework from which to conceptualize the biomarkers included in our report. Key Question 1, regarding the validity of diagnostic tests, fits in the Verification and Validation stage of the diagram. Key Questions 2 and 3, regarding associations of biological measures with GWI case status, fits into the stages prior to the Verification stage.

SEARCH STRATEGY

We conducted a primary review of the literature by systematically searching, reviewing, and analyzing the scientific evidence as it pertains to the research questions. To identify relevant articles, we searched Ovid MEDLINE, and Ovid PsycINFO, and Ovid EBM Reviews (Cochrane Database of Systematic Reviews and Cochrane Central Register of Controlled Trials) through February 20, 2020. Search strategies were developed in consultation with a research librarian, and peer reviewed by a second research librarian using the instrument for Peer Review of Search Strategies (PRESS).[17] The search strategy included terms to identify Veterans from the Gulf War era (*eg*, Desert Shield, Desert Storm, Kuwait War, Operation GRANBY) combined with past and present terms to identify GWI (*eg*, Chronic Multisymptom Illness, Chronic Fatigue, Gulf War Syndrome). We limited our search to English-language publications but did not limit by publication status or study design. To identify in-progress or unpublished studies, we searched ClinicalTrials.gov and the World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP). We reviewed the bibliographies of relevant articles, contacted experts, and reviewed lists of funded trials from the Department of Defense (DoD) and Veterans Affairs (VA) to identify additional studies.

STUDY SELECTION

Criteria for population, interventions, comparators, outcomes, timing, setting, and study design (PICOTS; Table 1) were developed in collaboration with our operational partners and Technical Expert Panel. Two reviewers independently assessed studies for inclusion based on pre-specified criteria. All discordant results were resolved through consensus or consultation with a third reviewer. Articles meeting eligibility criteria were included for data abstraction.

We defined biomarkers broadly: we included any biological measurement across a broad variety of biological functions or systems. This could include blood tests or even imaging studies. We distinguish biological measurements from other assessments such as symptom questionnaires which were excluded. We only included studies of biomarkers in humans.

For evidence on diagnostic test accuracy (Key Question [KQ] 1), we included all studies that compared a test's classification of GWI diagnosis with any reference standard's classification (*eg*, diagnosis of GWI according to Kansas or CDC/Fukuda 1998 criteria). While we recognize that there are limitations to these reference standards, they are the currently recommended and most current, widely used definitions, and a standard case definition was required to evaluate diagnostic accuracy.

For evidence on biological measures (KQ2), we included all studies that compared the prevalence or quantity of a biological measure in Veterans clinically diagnosed with GWI with Veterans without GWI (we included studies that used any case definition for KQ2). For emerging research (KQ3), we included studies with any of the above designs that are unpublished or in progress. Again, we included cases identified using any GWI diagnostic criteria.

We included studies in which the comparator populations were deployed Gulf War Veterans (GWVs) who were either healthy or had health conditions other than GWI. These groups were prioritized as they would best control for the effects of combat and exposure present in, and associated with, the conflict area. Also, these studies would be further along the biomarker development pathway and would therefore provide results more applicable to situations in which there is diagnostic uncertainty (*ie*, the clinical use of biomarkers would most likely be in participants who had been deployed and therefore had the potential to have GWI).

We were particularly interested in the group of GWVs with other health conditions, as differences in biological measures between this group and the GWI group could potentially help distinguish GWI cases from cases of other health conditions. In other words, it is for the Veteran with symptoms and a history of deployment that the clinical application of a biomarker would be most relevant and, therefore, studies which enrolled GWVs with symptoms (some of whom would have GWI as the cause of symptoms and others who would have symptoms secondary to another illness) are of particular interest.

We also acknowledge that other controls groups, including healthy non-Veterans or non-deployed GWVs, can help provide information about potentially pertinent biomarkers. We therefore identified a group of lower-priority studies, excluded from the main review, that are summarized in a supplementary table (Appendix D; Table 10). These were studies whose comparator groups consisted of non-deployed Gulf War-era Veterans, Veterans who were deployed elsewhere (other than Persian Gulf) during the Gulf War, civilians with other health conditions/conditions with similar symptomology to GWI (*eg*, chronic fatigue syndrome, neurodegenerative disorders, or musculoskeletal problems) and healthy civilian controls. We also considered studies with 25 or fewer participants as lower priority because these were likely to be too small[18] to determine with any confidence whether or not a given measure was associated with GWI.

Table 1. PICOTS by Key Question

Key Question	KQ1. Which diagnostic tests (or test combinations) are candidates for distinguishing individuals diagnosed with GWI from individuals without GWI?	KQ2. Which biological measures have been examined for their potential association with GWI, and which among them have been shown to be associated with GWI?	KQ3. Which ongoing or unpublished research studies examine diagnostic tests or biological measures for potential association with GWI?
Population	*Case definition:* Veterans and civilian contractors who were deployed to the Persian Gulf region between Aug 2, 1990 and Nov 1991, and diagnosed with GWI (*ie*, according to either CDC/Fukuda 1998 or Kansas criteria, or other criteria). Include studies of international Veteran populations if they use any definition of GWI (limited to English-language publications).*		
	Comparator populations: Gulf War-deployed Veterans with or without medical conditions other than GWI		
	Exclude: children and birth outcomes of Gulf War Veterans		
Intervention	Measures of any of the following categories of biological functions/systems that are potential loci of dysfunction: • Genes • Immune activation/inflammation • Neurodegeneration • Autonomic nervous system • Endocrine system • Energy metabolism • General brain activity (central nervous system) • Other Exclude: Assessments that do not include biological measurements (*eg*, questionnaires)		
Comparators	Compares a test's classification of GWI diagnosis with a reference standard's classification (*eg*, diagnosis of GWI according to Kansas or CDC/Fukuda 1998 criteria).	Compares Veterans clinically diagnosed with GWI vs any comparator group (see comparator populations above)	
Outcomes	Measures of diagnostic accuracy: • Sensitivity and specificity • Positive and negative predicative values (PPV, NPV) • Likelihood ratio • The area under the ROC curve (AUC)	Measure of association between biological measurement and GWI	Study objectives, status, outcome measures, and available findings.
Timing	No limits		
Settings	No limits		
Study Design	Non-experimental, cross-sectional study; systematic reviews of diagnostic accuracy studies.	Cross-sectional, cohort, case series, and case-control studies that compare the results of a diagnostic tool, or the prevalence or quantity of a biological measure; or systematic review of such studies.	Cross-sectional study or other comparative observational study design.

*We recognize other countries may use different case definitions

Abbreviations: AUC=Area Under the Curve; CDC=Centers for Disease Control and Prevention; GWI=Gulf War Illness; KQ=Key Question; NPV=Negative Predictive Value; PPV=Positive Predictive Value; ROC=Receiver Operating Characteristic

DATA ABSTRACTION

Data from studies meeting inclusion criteria were abstracted by 1 Evidence Synthesis Program (ESP) reviewer and were confirmed by at least 1 additional reviewer. From each study, we abstracted the following where available: study design, sample size, population characteristics, case definition, comparator(s), participant inclusion and exclusion criteria, details of the biological measure of interest, and findings for measures of association (Table 2).

If a study had multiple comparator groups, we abstracted only data that pertained to our comparators of interest: the GWI group and control participants who were deployed to the Gulf War theater in 1990-1991.

QUALITY ASSESSMENT

Two reviewers independently assessed the risk of bias of each study. To assess the risk of bias we used an adapted version of the Newcastle-Ottawa tool for quality assessment[19] along with adapted elements of the risk of bias tool for cross-sectional studies using biological measure data (BIOCROSS).[20] Had we identified any studies of the validation of diagnostic tests, we would have used the Quality Assessment of Diagnostic Accuracy Studies (QUADAS)-2 tool for those studies. Disagreements were resolved by consensus or a third reviewer.

DATA SYNTHESIS

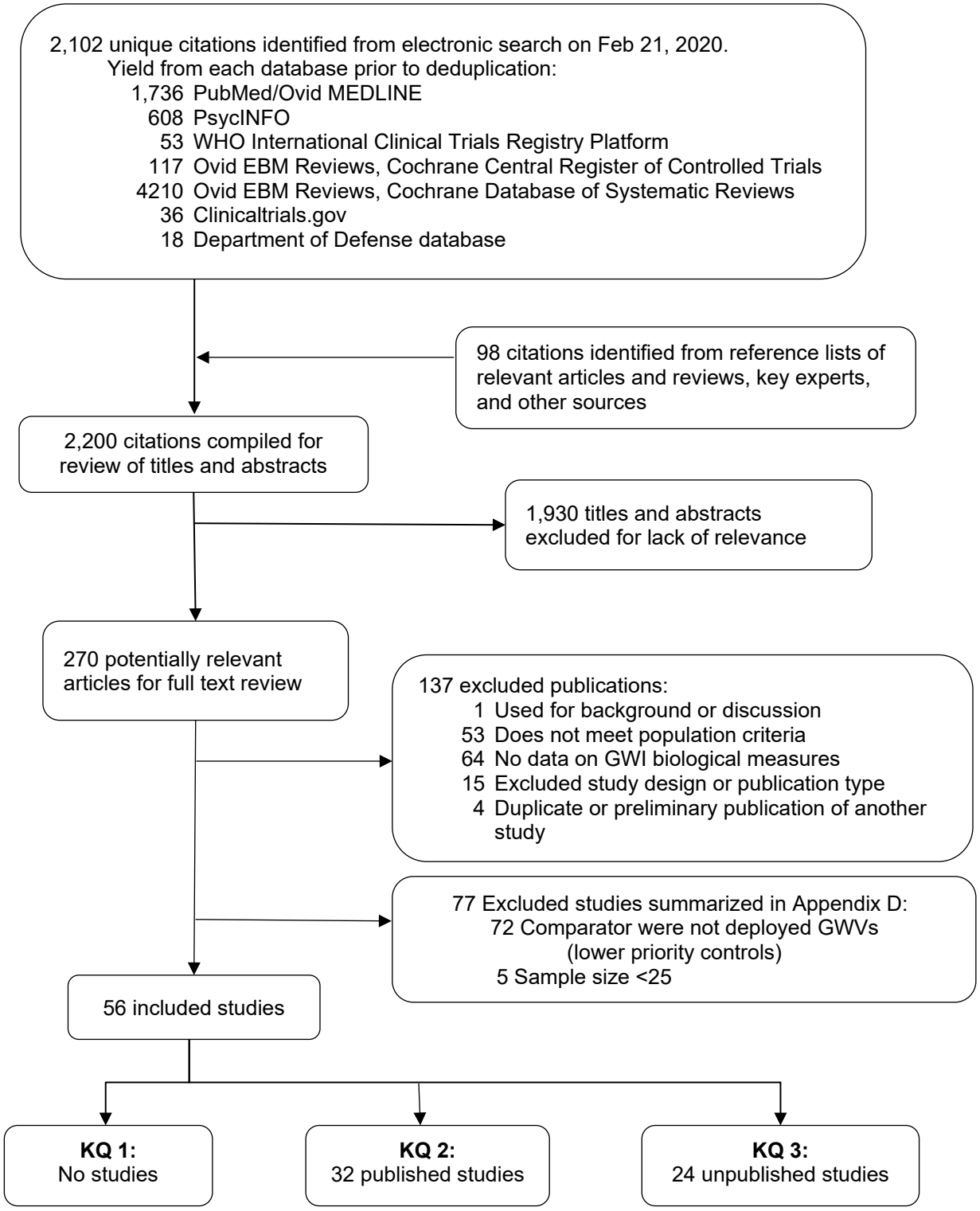
We qualitatively synthesized the evidence and compiled evidence tables of study characteristics and findings for each key question (Table 2). We also briefly summarized information about excluded studies in Appendix D.

We did not perform a formal certainty of evidence assessment due to the heterogeneity among studies and because the intent of the review was largely to identify areas of promise for further investigation of biological measures rather than to assess the body of evidence for clinical recommendations.

RESULTS

We reviewed a total of 2,102 titles and abstracts and identified 270 articles for review at the full-text level. We included a total of 56 studies for KQs 1-3 (Figure 3). However, we also summarized a larger group of lower-priority studies that met our inclusion criteria but did not include a priority comparator group (72 studies), and/or had total sample sizes of less than 25 (5 studies) – these studies are listed in Appendix D.

Figure 3. Literature flow chart



Abbreviations: GWI=Gulf War Illness; GWV=Gulf War Veteran; KQ=Key Question

KEY QUESTION 1: Which diagnostic tests (or test combinations) are candidates for distinguishing individuals diagnosed with GWI from individuals without GWI?

We did not identify any studies that evaluated diagnostic tests (or test combinations) for their ability to distinguish individuals diagnosed with GWI (with CDC or Kansas criteria, or any other criteria) from individuals without GWI (Key Question 1).

In order to identify a biomarker for use as a diagnostic test for GWI, studies would first need to assess the diagnostic accuracy of a given biomarker in identifying people with GWI. In a diagnostic accuracy study, all participants would be subjected to a "gold standard" test in order to distinguish those with the disease from those without the disease. Ideally, there would be diagnostic uncertainty among all participants in a diagnostic accuracy study (*ie*, as opposed to a study which enrolled participants with severe symptoms and participants with no symptoms). The biomarker in question would be assessed in all participants without knowledge of the results of the "gold standard" test, and then measures of diagnostic accuracy such as sensitivity and specificity could be assessed.

KEY QUESTION 2: Which biological measures have been examined for their potential association with GWI, and which among them have been shown to be associated with GWI?

We identified 32 studies of biological measures that have been examined for their association with GWI and grouped them broadly into categories under distinct biological systems: 10 studies of immune system biological measures; 10 central nervous system (CNS) studies; 5 studies of autonomic nervous system (ANS) biological measures; 1 of genetic biological measures; and 6 studies of biological measures in other biological systems (See Figure 4). All cases of GWI in all included studies were in participants who were Gulf War Veterans (*ie*, no studies including civilian contractors present in the Persian Gulf during the same time period were identified). Only 1 study[21] included deployed GWVs with health conditions other than GWI; in all other studies, health conditions were not reported and participants were presumed healthy. Eleven studies used the CDC case definition to identify participants with GWI and 2 used the Kansas definition. Ten studies used the Haley criteria, alone or in combination with CDC criteria. The Haley criteria categorizes cases of GWI into 6 syndromes based on symptom groups, 3 of which are referenced in the included studies: impaired cognition (Syndrome 1); confusion-ataxia (Syndrome 2); and arthro-myo-neuropathy (Syndrome 3).[4] Table 2 details the characteristics of these studies. The detailed study findings are reported according to biological system below.

Further, we identified 77 studies that used a lower-priority comparator group (that is, a comparator group other than deployed GWVs without GWI, with or without other health conditions), no comparator group, or an inadequate sample size (N<25). While not included in this main report, a summary of these studies is provided in Appendix D and Figure 5. Figure 5 indicates that most studies examined biological measures within the central nervous and immune systems, and that the majority of studies in all biological systems showed a statistically significant association of the biological measure with GWI case status.

Table 2. Characteristics of Gulf War Illness Biological Measure Studies

Study *Design* *N=total participants*	**Biological Measure(s) Examined**	**Data Collection**	**Population** *n GWI vs comparator* *GWI case definition* *Population/Sample source*	**Demographics** *GWI vs comparators* *Age: Mean years (SD)* *Female: %* *Race/Ethnicity: %*
IMMUNE SYSTEM				
Asa, 2000[23] Cross-sectional N=50	Squalene antibody status	Serum samples. Anti-squalene Antibody Assay which measures the binding of serum immunoglobulin (IgG) to squalene immobilized on nitrocellulose	**GWI:** 38 **Comparator:** 12 healthy **Case def:** CDC **Source**: Service in the US or UK military during Persian Gulf during 1990-1991	**Age**: NR **Female**: NR **Race**: NR
Butterick, 2019[28] Cross-sectional N=80	IFN-γ, IL-6, IL-8, IL-10, and TNF-α and C-reactive protein	Meso Scale Discovery (MSD) plate-based electrochemiluminescence (ECL) assay platform to quantify plasma concentrations of interferon gamma (IFN-γ), IL-1β, interleukin 2 (IL-2), interleukin 4 (IL-4), IL-6, interleukin 8 (IL-8), interleukin 10 (IL-10), interleukin 12 p70 (IL-12 p70), interleukin 13 (IL-13), and tumor necrosis factor alpha (TNF-α).	**GWI:** 53 **Comparator:** 27 healthy **Case def:** CDC **Source:** Parent study "Biological measures of Gulf War Veterans' Illnesses: Tissue Factor, Chronic Coagulopathy, and Inflammation"	**Age**: NR **Female**: NR **Race**: NR
Emmerich, 2017[24] Cross-sectional N=33	Several phospholipid species in the plasma. PC, LPC, PE, LPE and PI	lipid extracts from plasma were resuspended in isopropanol and separation was achieved using hydrophilic interaction chromatography	**GWI:** 22 **Comparator:** 11 **Case def:** Kansas **Source**: GWI biorepository from Boston and Miami areas	**Age**: 48.4 (6.3) vs 48.5 (7.7) **Female**: 23 vs 18 **Race**: Caucasian 54 vs 45; Hispanic 9 vs 18; Other 5 vs 9
Georgopoulos, 2016[25] Cross-sectional N=82	HLA alleles	DNA isolation, 3 ml of whole blood. High-resolution HLA Sequence-based Typing conducted on purified DNA.	**GWI:** 66 **Comparator:** 16 healthy **Case def:** Kansas or CDC **Source:** VA medical records	**Age**: 50.6 (7.9) vs 50.6 (7.9) **Female**: 3 vs 6 **Race**: NR
James, 2016[27] Cross-sectional N=81	Human leukocyte antigen, brain synchronicity	High-resolution HLA Sequence-based Typing conducted on purified DNA and MRI	**GWI:** 65 **Comparator:** 16 healthy **Case def:** CDC **Source:** Same sample (-1) studied in a previous study[25]	**Age**: 50.8 (7.9) vs 54.9 (10.2) years **Female**: 3 vs 6 **Race**: NR
Johnson, 2013[26] Cross-sectional	Platelet count, immature platelet	CRP and TPO assayed by Multi-Analyte Profiling. Platelet aggregation was measured	**GWI:** 43 **Comparator:** 21 Healthy	**Age**: 49.9 (8.55) vs 50.4 (11.0)

Study *Design* *N=total participants*	**Biological Measure(s) Examined**	**Data Collection**	**Population** *n GWI vs comparator* *GWI case definition* *Population/Sample source*	**Demographics** *GWI vs comparators* *Age: Mean years (SD)* *Female: %* *Race/Ethnicity: %*
N=64	fraction (IPF), plasma thrombopoietin (TPO), C-reactive protein (CRP), platelet aggregation and ATP secretion	by preparing citrated blood at 160g for 25 min, completed within 4 h of venipuncture. Aggregation was evaluated at the platelet count of the PRP without dilution or concentration. Agonists (0.5 mmol/l arachidonic acid, 5mmol/l epinephrine, 5mmol/l ADP, 0.5mmol/lU46619, 1.0mmol/l U46619, 1mg/ml collagen, 10mmol/thrombin receptor agonist peptide 6 (TRAP 6), and 50 mmol/l TRAP 6; final concentrations) were added to 450ml PRP, preincubated at 37°C, and stirred in a Chronolog Lumi-Aggregometer. ATP release was measured by Chronolog Lumi-Aggregometer.	**Case def:** CDC **Source:** Veterans identified through the resources of the US Department of Veterans Affairs as persons who served in Kuwaiti Theater of Operations 1990-1991	**Female**: 7 vs 0 **Race**: NR
Johnson, 2016[29] Cross-sectional N=85	Plasma lymphocytes, monocytes and neutrophils	Proteomic analysis performed by quantitative multiplexed immunoassays using a Multi-Analyte Profile (MAP) platform	**GWI:** 57 **Comparator:** 28 Healthy **Case def:** CDC **Source:** VA Minnesota Gulf War Registry, invited via letter	**Age**: Median – 46 vs 48 **Female**: 5 vs 4 **Race**: White 93 vs 93; Black 4 vs 7; Hispanic 4 vs 0
Lo, 2000[22] Cross-sectional, population-based N=2,951	Mycoplasma fermentans antibodies	Pre- and post-deployment serum was obtained from the DoD Serum Repository	**GWI:** 718 **Comparator:** 2,233 healthy **Case def:** Per medical examination **Source:** Pt selection and serum specimen coding were performed by USACHPPM	**Age** >35 years: 136 (18.9) vs 425 (19.0) **Female**: 15.9 vs 15.7 **Race**: Black/AA 230 (32.0) vs 494 (22.1); White 410 (57.1) vs 1,552 (69.5); Other 78 (10.9) vs 187 (8.4)
Phillips, 2009[30] Case-control N=175	Squalene antibody status	ELISA for antibodies to squalene in human serum	**GWI:** 29 **Comparator:** 146 healthy **Case def:** Unusual fatigue with 3 of 38 additional symptoms **Source:** 3 epidemiologic team visits (late 1994 - early 1995) to each of 2 Navy Seabee Centers	**Age**: 25.9 vs NR **Female**: 0 vs NR **Race**: White 84 vs NR; Black/AA 5 vs NR; Other/unknown 12 vs NR
Skowera, 2004[31] Cross-sectional N=120	Th1/Th2 balance by measuring intracellular production of IFN-γ,	Blood samples were obtained; Flow cytometry was used to measure intracellular cytokine production by CD4 T lymphocytes.	**GWI:** 40 **Comparator:** 80 healthy	**Age**: NR **Female**: NR

Study *Design* *N=total participants*	**Biological Measure(s) Examined**	**Data Collection**	**Population** *n GWI vs comparator* *GWI case definition* *Population/Sample source*	**Demographics** *GWI vs comparators* *Age: Mean years (SD)* *Female: %* *Race/Ethnicity: %*
	IL-2 (Th1), IL-4 (Th2), and IL-10 by CD4 T cells		**Case def:** Score ≤72.2 on SF-36 Physical Functioning subscale **Source:** Random sample of Veterans of the Gulf conflict (1990-91)	**Race**: NR
CENTRAL NERVOUS SYSTEM				
Calley, 2010[32] Cross-sectional N=38	Percent signal change	brain activation during semantic memory task: indicate if 2 features (written words) elicit the retrieval of a memory of a specific object (retrieval), or not (non-retrieval)	**GWI:** 26 (syndromes 1-3: 9;10;7) **Comparator:** 12 Healthy **Case def:** Haley criteria **Source:** Construction Battalion in US Naval Reserve	**Age**: NR **Female**: 0 **Race**: NR
Cooper, 2016[33] Cross-sectional N=67	Regional brain activation	brain activation during a face-name episodic memory task	**GWI:** 57; syndrome 1,2,3=19, 20, 18 **Comparator:** 10 Healthy **Case def:** Haley criteria **Source:** Sampled from nationally representative sample of Gulf War-era Veterans (N=8020)	*GWI (syndromes 1, 2, 3) vs comparators:* **Age**: 49.16 (8.84) vs 49.95 (8.39) vs 51.44 (8.15) vs 45.60 (7.41) **Female**: 32, 30, 17 vs 10 **Race/eth**: Caucasian 68, 80, 72 vs 60; Black/AA 11, 15, 17 vs 30; Hispanic 5, 5, 6 vs 10; Other 16, 0, 6 vs 0
Gopinath, 2012[34] Cross-sectional N=54	Brain activation	brain activation to innocuous and noxious heat stimuli	**GWI (syndromes 1, 2, 3):** 11; 17; 12 **Comparator:** 14 Healthy **Case def:** Haley criteria **Source:** 24th Reserve Naval Mobile Construction Battalion	**Age**: 51 (6); 63 (7); 57 (7) vs 61 (7) **Female**: 0% **Race**: NR
Liu, 2011[35] Cross-sectional N=47	Cerebral blood flow	MRI with an inhibitory cholinergic challenge, physostigmine infusion	**GWI:** syndromes 1,2, 3; 11;12;10 **Comparator:** 14 Healthy **Case Def:** Fukuda/Haley **Source:** Members of the U.S. Naval construction battalion	*GWI (syndromes 1, 2, 3) vs comparators:* **Age**: 51.4 (6.1), 60.9 (6.0), 57.3 (6.7) vs 60.1 (6.3) **Female**: 0 **Race**: NR
Odegard, 2013[36] Cross-sectional N=47	Brain activation	brain activation during face-name associative memory paradigm	**GWI:** 33 **Comparator:** 14 Healthy	*GWI (syndromes 1, 2, 3) vs comparators:* **Age**: 51.30, 61.67, 57.27 vs 60.36

Study *Design* *N=total participants*	**Biological Measure(s) Examined**	**Data Collection**	**Population** *n GWI vs comparator* *GWI case definition* *Population/Sample source*	**Demographics** *GWI vs comparators* *Age: Mean years (SD)* *Female: %* *Race/Ethnicity: %*
			Case def: Haley criteria **Source:** Construction battalion of the U.S. Naval Reserve	**Female**: NR **Race**: NR
Tillman, 2010[37] Cross-sectional N=48	Event-related potential	EEG recordings during a go/no-go task (response inhibition)	**GWI:** 25 **Comparator:** 23 Healthy **Case def:** Presented with major cognitive complaints **Source:** Construction battalion of the U.S. Naval Reserve during 1991 Persian Gulf War	**Age**: 58.4 vs 58.8 **Female**: 0 **Race**: NR
Tillman, 2012[38] Cross-sectional N=28	Event-related potential	EEG recordings during an auditory task: threatening sounds as distractor stimuli, 1000 hz square wave tone as nontarget stimulus, 250-hz square wave tone as the target stimulus	**GWI:** 20 **Comparator:** 8 Healthy **Case def:** Haley criteria **Source:** Construction battalion of the U.S. Naval Reserve during the 1991 Persian Gulf War	*GWI (syndromes 1, 2, 3) vs comparators:* **Age**: 53.17 (5.38), 63.75 (7.05), 53.833 (6.85) vs 61.6 (7.58) **Female**: NR **Race:** NR
Tillman, 2013[39] Cross-sectional N=30	Event-related potentials	EEG recordings during visual task: threatening distractor pics, target stimuli (animals), nonthreatening nontarget stimuli	**GWI:** 22 **Comparator:** 8 Healthy **Case def:** Haley criteria **Source:** Construction battalion in U.S. Naval Reserve during the 1991 Persian Gulf War	**Age**: 57.2 vs 61.6 **Female**: 0 **Race**: NR
Tillman, 2019[40] Cross-sectional N=62	Event-related potential	EEG recordings during an auditory task: threatening distractor sounds, target tone stimulus	**GWI:** 40 **Comparator:** 12 Healthy **Case def:** Haley criteria **Source:** Nationally representative dataset of the Gulf War Veteran population from 4 military branches	*GWI (syndromes 1, 2, 3) vs comparators:* Age: 49.2 (10.4), 52.0 (7.3), 49.7 (7.8) vs 48.4 (7.9) **Female**: 25 vs 14 **Race**: Caucasian: 87 vs NR; Other: 13 vs NR
Weiner, 2011[41] Cross-sectional N=178	N-acetylaspartate, creatine-and choline-containing metabolites	Spectroscopy (MRI, MRS, MRSI)	**GWI:** 81 **Comparator:** 97 healthy **Case def:** CDC and Haley criteria **Source:** N. California VA GWI clinics, fliers at VA hospitals/clinics, and direct mailing to DoD list of GWVs	**Age**: 44.6 (8.8) vs 44.6 (9.9) **Female**: 9 vs 15 **Race/eth:** White 64 vs 61; Black/AA 14 vs 16; Hispanic 11 vs 9; Asian 2

Study *Design* *N=total participants*	**Biological Measure(s) Examined**	**Data Collection**	**Population** *n GWI vs comparator* *GWI case definition* *Population/Sample source*	**Demographics** *GWI vs comparators* *Age: Mean years (SD)* *Female: %* *Race/Ethnicity: %*
				vs 3; Pacific islander 2 vs 4; Other 2 vs 4
AUTONOMIC NERVOUS SYSTEM				
Blanchard, 2019[42] Cross-sectional case-control cohort study N=295	24-hr heart-rate variability, urinary catecholamines and cortisol, hypertension, insulin sensitivity, dyslipidemia, body fat, bone mineral density, and ultrasensitive CRP	Hypothalamic-Pituitary-Adrenal Axis measures: overnight dexamethasone suppression testing ANS Measures: ECG via 24-hour digital monitors, looking at HRV	**GWI:** 73 **Comparator:** 111 healthy **Case def:** CMI – CDC definition **Source:** NR	**Age**: 47.1 (8.1) vs 47.5 (9.0) **Female**: 36 vs 17 **Race**: White 73 vs 83
Davis, 2000[43] Cross-sectional N=27	Heart rate, blood pressure	3-stage tilt-table testing with isoproterenol	**GWI:** 14 **Comparator:** 13 healthy **Case def:** Chronic Fatigue (ICD-9-CM Code 780.7) **Source:** Mostly active duty soldiers who went through DoD CCEP (Evans Army Community Hospital; 1994-1997)	**Age**: 32.1 (1.6) vs 38.9 (1.9) **Female**: 14 vs 8 **Race**: NR
Haley, 2013[44] Cross-sectional N=82	Autonomic function and high-frequency HRV	CASS, measuring the severity of autonomic dysfunction from 0 (no deficit) to 10 (maximal deficit), using sudomotor (range, 0-3), cardiovagal (range, 0-3), and adrenergic (range, 0-4) measurements. 24-hour Holter ECG recordings, performed at home, complexes and artifacts.	**GWI:** 66 (syndromes 1, 2, 3: 21, 24, 21) **Comparator:** 16 healthy **Case def:** CDC and Haley criteria **Source:** Population sample of the US Military Health Survey	*GWI (syndromes 1, 2, 3) vs comparators:* **Age**: 48.2 (8.6), 49.8 (8.0), 51.0 (7.9) vs 47.8 (7.9) **Female**: 33, 29, 19 vs 19 **Race**: Black/AA 14, 17, 14 vs 25
Li, 2014[45] Cross-sectional N=28	BP, HR, tilt-table, nerve conduction, sensory testing, sudomotor axon reflex testing	BP, HR, tilt-table, nerve conduction, sensory testing, sudomotor axon reflex testing	**GWI:** 16 **Comparator:** 12 healthy **Case def:** Post-exertional fatigue: patient-reported recurrent post-exertional fatigue lasting >24 hours, or chronic fatigue >6 months since deployment **Source:** "Health of US Veterans of 1991 Gulf War: A Follow-Up Survey in 10 years" 2005 survey pts in Washington DC and nearby states	**Age**: 48.3 (1.4) vs 48.1 (2.0) **Female**: 19 vs 8 **Race**: NR

Study *Design* *N=total participants*	**Biological Measure(s) Examined**	**Data Collection**	**Population** *n GWI vs comparator* *GWI case definition* *Population/Sample source*	**Demographics** *GWI vs comparators* *Age: Mean years (SD)* *Female: %* *Race/Ethnicity: %*
Nagelkirk, 2003[46] Cross-sectional N=38	Cardiorespiratory and metabolic responses (maximal oxygen uptake, HR, exercise time, workload achieved) to maximal exercise test	Maximal exercise test: BP electronically monitored + verified with concurrent mercury sphygmomanometry; HR monitored by ECG; expired air collected with one-way nonrebreathing valve + analyzed with Max-1 system	**GWI:** 19 **Comparator:** 19 healthy **Case def:** CDC case def for CFS **Source:** Mailed health survey; study site East Orange, NJ VAMC	**Age**: 41.9 (7.8) vs 43.1 (5.1) **Female**: 16 vs 16 **Race**: NR
GENETIC				
Hotopf, 2003[47] Cross-sectional, population-based N=190	Paraoxonase (PON1) activity and genotype	PON1–55 and -192 genotype determined by polymerase chain reaction and restriction enzyme digestion using standard published protocols. Plasma apolipoprotein AI (apoAI) and high-density lipoprotein (HDL) measured with Cobas Mira S autoanalyser with reagents.	**GWI:** 95 **Comparator:** 95 healthy **Case def:** SF-36 (physical functioning subscale) **Source:** Randomly selected cohorts of UK Armed Forces	**Age**: 36.9 (7.3) vs 34.3 (5.4) **Female**: 6 vs 5 **Race**: NR
OTHER BIOLOGICAL SYSTEMS				
Amin, 2011[48] Cross-sectional N=29	Sleep parameters; Sleep-related respiratory parameters	2 nights in sleep lab: full night PSG, surface electromyographic activity (for movement); nasal/oral pressure catheter (nose+mouth airflow); piezoelectric belts (for Thoraco-abdominal movement); oxyhemoglobin saturation monitored at the finger using pulse oximeter. Continuous ECG monitored HR and rhythm	**GWI:** 18 **Comparator:** 11 healthy **Case def:** Modified CDC **Source:** Gulf War Registry	**Age**: 42 (4) vs 41 (6.6) **Female**: 0% **Race**: NR
Haines, 2017[21] Cross-sectional, population-based N=43	Cholinesterase, Serum cytokines	Peripheral venous blood draw	**GWI:** 25 **Comparator**: 4 healthy, 14 PTSD **Case def:** Neurologic factor (4 symptoms): blurred vision, balance problems/dizziness, tremors/shaking, and speech difficulty **Source**: UK military personnel	**Age**: NR **Female**: NR **Race**: NR
Roland, 2000[49] Cross-sectional N=33	Integrity of auditory pathways (inner ear through upper brain stem and vestibulocular reflex)	Audiovestibular testing: Rotary chair for Sinusoidal Harmonic Acceleration; Electronystagmography for ocular motor, positional, and caloric responses; Electrocochleography for auditory-evoked	**GWI:** 23 (Syndrome 1, 2, 3: 5, 13, 5) **Comparator:** 10 healthy **Case def:** Haley criteria **Source:** 24th Reserve Naval Mobile Construction Battalion	**Age**: 46.6 ± 8.6 vs 48.0 ± 6.2 **Female**: NR **Race**: NR

Study *Design* *N=total participants*	**Biological Measure(s) Examined**	**Data Collection**	**Population** *n GWI vs comparator* *GWI case definition* *Population/Sample source*	**Demographics** *GWI vs comparators* *Age: Mean years (SD)* *Female: %* *Race/Ethnicity: %*
		potentials; dynamic platform posturography for balance.		
Sharief, 2002[50] Cross-sectional, population-based N=75	Distal motor latency, amplitude of compound motor action potential (CMAP), motor nerve conduction velocity (CV), and F-wave latency; functions of unmyelinated (C) + small myelinated (A-) fibers	Nerve conduction studies on dominant limbs + symptomatic side if symptoms were unilateral, at skin temperatures of 34 °C. Motor conductions recordings Quantitative sensory and autonomic function tests: thermal thresholds in dominant hand and foot using TSA-2001 machine Concentric needle and single-fiber EMG: On proximal + distal muscles in upper + lower limbs (biceps brachii, first dorsal interosseus, vastus medialis, and tibialis anterior)	**GWI**: 49 **Comparator**: 26 healthy **Case def:** >4 neuromuscular symptoms: fatigue, joint stiffness, muscle weakness, myalgia at rest or after exercise, sensory symptoms, and autonomic symptoms + SF-36 score <72.2 **Source:** UK servicemember database	**Age**: NR **Female**: NR **Race**: NR
Wallace, 1999[51] Cross-sectional, not population-based N=78	(HHVs) HHV6, HHV7, Epstein-Barr virus (EBV), and cytomegalovirus	Blood samples: DNA extraction from PBMC cells; PCR was performed with primers for HHV6, HHV7, EBV, HCMV, TCR-beta	**GWI:** 46 **Comparator**: 32 healthy **Case def:** CDC **Source:** Veterans from 10 states east of the Mississippi River	**Age**: 35.3 vs 32 **Female**: 17% vs NR **Race**: Caucasian 73% vs NR
Zhou, 2018[52] Cross-sectional N=91	Heat pain threshold, cold pressor pain threshold; ischemic pain threshold and ischemic pain tolerance	Experimental pain procedures (3 random pain stimuli)	**GWI:** 53 GWI+GI symptoms; 38 no GI symptoms **Comparator:** 47 healthy **Case def:** Documented GWI +/- GI symptoms **Source:** VAMCs in Cincinnati, OH and Gainesville, FL	*GWI (with, without GI symptoms) vs comparators:* **Age**: 49.1 (3.7), 48.2 (1.2) vs 46.6 (4.1) **Female**: 9, 0 vs 10 **Race/eth**: White 51, 83 vs 21; Black/AA 47, 17 vs 45; Hispanic 0, 0 vs 0; Asian 2, 0 vs 0

Abbreviations: AA=African American; ADP=Adenosine Diphosphate; ANS=Autonomic Nervous System; apoAI=Plasma Apolipoprotein AI; ATP=Adenosine Triphosphate; BOLD=Blood Oxygen Level Dependent; BP=Blood Pressure; CASS=Composite Autonomic Severity Score; CCEP=Comprehensive Clinical Evaluation Program; CDC=Centers for Disease Control and Prevention; CFS=Chronic Fatigue Syndrome; CMAP=Compound Motor Action Potential; CMI=Chronic Multisymptom Illness; CRP=C-Reactive Protein; CV=Conduction Velocity; DNA=Deoxyribonucleic Acid; DoD=Department of Defense; EBV=Epstein-Barr Virus; ECG=Electrocardiogram; ECL=Electrochemiluminescence; EEG=Electroencephalograph; ELISA=Enzyme-Linked Immunosorbent Assay; EMG=Electromyograph; fMRI=Functional Magnetic Resonance Reasoning; GI=Gastrointestinal; GWI=Gulf War Illness; GWV=Gulf War Veteran; HCMV=Human Cytomegalovirus; HDL=High-Density Lipoprotein; HHV=Human Herpesvirus; HLA=Human Leukocyte Antigen; HR=Heart Rate; HRV=Heart Rate Variation; ICD-9-CM=International Classification of Diseases, Ninth Revision, Clinical Modification; IFN=Interferon; IgG=Immunoglobulin; IL=Interleukin; IPF=Immature Platelet Fraction; LPC=Lysophosphatidylcholines; LPE=Lysophosphatidylethanolamine; MAP=Multi-Analyte Profile; MRI=Magnetic Resonance Imaging; MRS=Magnetic Resonance Spectroscopy; MRSI=Magnetic Resonance Spectroscopic Imaging; MSD=Meso Scale Discovery; NR=Not

Reported; PBMC=Peripheral Mononuclear Cell; PC=Phosphatidylcholine; PCR=Polymerase Chain Reaction; PE=Phosphatidylethanolamine; PI=Phosphatidylinositol; PON1=Paraoxonase; PRP=Platelet-Rich Plasma; PSG=Polysomnogram; PTSD=Posttraumatic Stress Disorder; SD=Standard Deviation; SF-36=Short Form-36 Question Questionnaire; TCR=T-Cell Receptor; TNF=Tumor Necrosis Factor; TPO=Plasma Thrombopoietin; TRAP=Thrombin Receptor Agonist Peptide; UK=United Kingdom; US=United States; USACHPPM=United States Army Center for Health Promotion and Preventive Medicine; VA=Veterans Affairs; VAMC=Veterans Affairs Medical Center

Immune System Biological Measure Studies

Table 3 provides results of the 10 studies of immune system biological measures we identified. Identified studies focused on a wide range of immune system-related functions including squalene antibodies (2 studies[23,30]), a variety of inflammatory cytokines (5 studies[24,26,28,29,31]), human leukocyte antigen (HLA) alleles (2 studies[25,27]), and mycoplasma antibodies (1 study[22]).

Two studies assessed the associations between the presence of squalene antibodies and GWI status.[23,30] The first was a case-control design that included 43 Veterans with Chronic Multisymptom Illness (CMI) and 536 who did not have CMI, as defined by unusual fatigue accompanied by 3 of a list of symptoms in broad categories such as altered mood, GI issues, pain, dermatologic issues, and generalized malaise (*eg*, sore throat and weakness).[30] Whole blood or sera was collected using standardized approaches and subjects were followed up with 3 years later. There was no significant difference between those who were squalene antibody-positive or -negative with regard to presence of GWI. Limitations to this study included use of a non-standard GWI case definition, lack of comparability between cases and control, and limited data on sample attrition. In contrast, the second study to examine squalene antibodies did find a difference in the presence of squalene antibodies between those with GWI (n=38) and compared to those with idiopathic immune disease or healthy controls (n=12), such that 95% of those with GWI had squalene antibodies, while 0% of healthy deployed GWVs had squalene antibodies. This study was a feasibility study, so no power analyses or inferential statistics were provided.[23]

Many of the other studies focused on measures of inflammation with peripheral blood cytokines[26,28,29,31] and phospholipid species.[24] Two studies focused on peripheral blood cytokines (Th1/Th2 balance measured by intracellular production of Interferon (IFN)-γ, Interleukin (IL)-2 (Th1), IL-4 (Th2), and IL-10 by CD4 T cells) among those who were symptomatic Gulf War Veterans versus non-symptomatic Gulf War Veterans.[28,31] The first found that those who were ill had elevated levels of Th1 immune activation as evidenced by elevated levels of IL-10 by CD4 cells ($p<0.001$), IL-2 ($p=0.01$) and IL-4 ($p=0.001$).[31] Analyses corrected for age, gender, vaccination status, antidepressant use, Beck Depression Inventory score, and history of atopic illness. All other cytokines were normal. This study was limited by not accounting for multiple testing bias, the cases and controls not being drawn from the same population, and a non-standard GWI case definition being used to identify cases (used Short Form [SF]-36 score to identify cases). The other study compared those with GWI (n=53) to healthy, deployed GWVs (n=27) using the CDC case definition. A similar set of cytokines were examined (IL-6, IL-8, IL-10, IFN-γ, tumor necrosis factor (TNF)-α, and C-reactive protein (CRP)) and no statistically significant between group differences were detected based on GWI status.[28] Limitations include sample size and power not being calculated and adjustment for multiple comparisons not being conducted. Johnson and colleagues included 43 men with GWI and 21 men without GWI and compared CRP, platelet counts, platelet aggregation, and several other measures of platelet function and found GWI had elevated platelet counts, spontaneous aggregation, and thrombin receptor agonist peptide (TRAP) 6-induced secretion, but no impairment of platelet counts.[26] Study limitations include multiple testing not adjusted for and non-response rate not reported. Another study examined a similar set of blood inflammatory cytokines measures that also included lymphocyte, monocyte, and neutrophil among 57 individuals with GWI compared to 28 GWVs without GWI. All included inflammatory markers were higher among those with GWI (see Table 3 for specific results).[29] The combinations of lymphocytes, monocytes, and CPR had a

predicted probability of 90% (95% CI=76-90%) for diagnosing GWI when the probability of having GWI was above 70%. Limitations of this study included not adjusting for multiple comparisons or difference between cases and controls, as well as some of the more ubiquitous limitations discussed listed below.

Another study obtained samples of phospholipids, which are inflammatory modulators, from a biorepository (11 controls matched for gender, age, and ethnicity, and 22 who met Kansas criteria for GWI) to examine several different phospholipids (Phosphatidylcholine [PC]; phosphatidylethanolamine [PE]).[24] Peripheral lipids were present in GWI (human and animal models). The study found that multiple species of phospholipids were elevated in humans with GWI, suggesting dysfunction within docosahexaenoic acid and arachidonic acid containing phospholipid (PL) species. This study did correct for multiple testing, but it was unclear whether the cases and controls were derived from the same population and are comparable.

Two additional studies examined HLA alleles.[25,27] The first included 82 Veterans, 66 with and 16 without GWI, and found that the number of copies of the 6 HLA alleles was significantly higher in the control group and correctly classified the GWI status of 84.1% of participants (13/16 control and 56/66 GWI).[25] Limitations included lack of clarity about selection of controls and potential lack of comparability between cases and controls. An additional study examined HLA alleles and brain synchronicity in 81 Veterans (65 with GWI and 16 without GWI). Controls had higher counts of HLA protective alleles than those with GWI (chi-square test=21.9, p=0.000018). There was also an overall strong and significant effect of the HLA-related x neural synchrony interaction on symptom severity (SPRC $|\beta|$=0.274,0.232, and 0.200 for neurocutaneous melanocytosis (NCM), pain, and fatigue, respectively, $p<0.001$ for all 3 coefficients.[27] Limitations included unrepresentativeness of cases and controls, lack of power calculation, lack of comparability between cases and control, and lack of reporting of blinding and non-response rate.

Finally, 1 study examined the presence of antibodies to Mycoplasma (M.) fermentans[22] among 718 Veterans with GWI and 2,233 healthy Gulf War Veterans. There was no difference between rates of seroconversion between cases and controls, indicating that there is no association between GWI and M. fermentans infection.

Limitations across immune system studies included lack of blinding of outcome assessor and/or study team. Additionally, recruitment non-response rate was not reported, giving us little insight into issues of response bias. Sample size and power calculations were also often not reported, and very few studies discussed issues related to data normality and handling of outliers.

Table 3. Results of Gulf War Illness Immune System Biological Measure Studies

Biological Measure/ Outcome Measure *n GWI vs n comparator*	**Detailed Results** *GWI vs control*	**Summary of Findings**	**Study Limitations**
Human leukocyte antigen, brain synchronicity (zero-lag, pairwise cross correlations calculated from pre-whitened, 60-second resting-state MEG recordings)[27] 65 vs 16	**Human leukocyte antigen protective alleles:** Cases had higher counts than healthy GWI chi-square test=21.9 (p=0.000018) **Brain synchronicity:** Significant effect of the HLA-related x neural synchrony interaction on symptom severity (SPRC |β|=0.274,0.232, and 0.200 for NCM, pain, and fatigue, respectively, p<0.001 for all t coefficients). HLA-related effects spared the right anterior temporal lobe and were widespread, vs non-HLA-related effects focused on anterior temporal lobe, medial prefrontal cortex, posterior parietal, and occipital cortex.	HLA antigens are higher in those with GWI; HLA related x neural synchrony is related to symptom severity.	Lack of representativeness of cases and controls, lack of power calculation, lack of comparability between cases and control, lack of blinding, lack of reporting of non-response rate.
Mycoplasma (M.) fermentans antibodies[22] 718 vs 2,233	**Positive for M. fermentans-specific antibodies, (%):** *Pre-deployment:* 34/718 (4.8%) vs 116/2233 (5.2%) *Post-deployment seroconversion from negative to positive:* 8/718 (1.1%) vs 26/2233 (1.2%) *Seropositive at both pre- and post-deployment:* 17 (2.4%) vs 54 (2.4%); OR=0.97 (95% CI: 0.56 to 1.69) *Positive on pre-deployment and negative and post-deployment:* 7 (2.4%) vs 62 (2.8%); OR=0.85 (95% CI: 0.49 to 1.46)	There was no difference between rates of seroconversion between cases and controls, indicating that there is no association between GWI and M. fermentans infection.	Lack of comparability of cases and controls, power calculation, non-response rate not reported, lack of blinding of outcome assessors.
Squalene antibody status[23] 38 vs 12	**% participants with positive ASA reactivity:** 95% of those with GWI vs 0% of healthy controls	Squalene antibodies may be associated with GWI.	Lack of comparability between cases and controls; unrepresentativeness of cases and controls.
Cytokines: IFN-γ, IL-6, IL-8, IL-10, and TNF-α and CRP[28] 53 vs 27	**Plasma cytokine concentrations below the lowest level of detection (LLOD):** Plasma concentrations for 4 cytokines were below the LLOD and were excluded. **Included plasma cytokines:** No difference in concentrations of IFN-γ, IL-6, IL-8, IL-10, and TNF-α, but trend toward significance in IL-6 (p=0.08)	No between group difference detected on any of the cytokines measured.	Power calculations not reported; no adjustments made for multiple comparisons.
Phospholipid species in plasma: PC, LPC, PE, LPE and PI[24] 22 vs 11	**Phospholipid levels in GWI vs healthy GWI controls:** No differences in PC, PE, LPE, PI (p>0.05) LPC was 15% greater for GWI (p=0.020) **Unsaturation of PL classes:** no differences in degree of unsaturation for PE and PI	Multiple species of phospholipids were elevated in humans with GWI	Comparability of cases and controls; and non-response rate not reported

Biological Measure/ Outcome Measure *n GWI vs n comparator*	**Detailed Results** *GWI vs control*	**Summary of Findings**	**Study Limitations**
	SFA containing PC species: reduced by 22% in GWI compared to controls (p=0.024) **LPC, SFA, MUFA, and PUFA containing species**: elevated in GWI compared to controls by 16%, 15%, and 23% (p<0.05) **LPE, PUFA containing species**: increase of 50% in GWI compared to controls (p<0.001) **SFA and MUFA containing LPE species:** no difference between GWI and controls **Ether lipids in plasma:** no differences for ePC, eLPC, ePE between GWI and controls **eLPE:** increased in GWI by 43% compared to controls (p<0.001) **AA- and DHA-containing phospholipid species:** AA species within LPC and LPE were increased by 22% in GWI compared to controls (p=0.023) and 40% respectively (p=0.005)		
Platelet count, IPF, TPO, CRP, platelet aggregation and ATP secretion[26] 43 vs 21	**Platelet count, mean platelet volume, immature platelet fraction, C-reactive protein, and thrombopoietin:** mean platelet count and plasma CRP: 4.05 (4.58) vs 1.54 (1.32); p=0.020 **Mean MPV, IPF and TPO platelet aggregation:** no difference **Spontaneous aggregation**: 7.6 (2.2) vs (6.1 (2.5); p=0.017 **Aggregation responses of platelets with each agonist**: no difference **Platelet ATP secretion** (stimulated with TRAP 6)**:** 15.46 (4.48) vs 12.42 (2.84); p=0.011	GWI had elevated platelet counts, spontaneous aggregation, and TRAP 6-induced secretion, but no impairment of platelet counts.	Multiple testing not adjusted for, non-response rate not reported
Plasma lymphocytes, monocytes and neutrophils[29] 57 vs 28	**Hematological data:** *Lymphocyte, monocyte, neutrophil, and platelet counts:* higher in GWI **Proteomic analysis:** *plasma CRP, leptin, BDNF, and MMP-9:* higher in GWI H-FABP and MMP-2: lower in blood of GWI **Diagnostic model:** *GWI diagnostic model using 3 biological measures* (lymphocytes, monocytes, and CRP): c-statistic=0.77 (95% CI: 0.67 to 0.88; p=0.05)	Monocyte, and neutrophil and found that all were higher in those with GWI.	Lack of blinding of outcome assessors, lack of correction for multiple comparisons, no adjustment for differences between cases and controls
Squalene antibody status[30]	**Squalene antibodies Negative:** 13/29 (44.8%) vs 71/146 (48.6%) **Squalene antibodies Positive:** 75/146 (51.4%) vs 16/29 (55.2%)	Study did not find an association between	Non-standard GWI case definition used, unclear

Biological Measure/ Outcome Measure *n GWI vs n comparator*	**Detailed Results** *GWI vs control*	**Summary of Findings**	**Study Limitations**
CMI 29 vs 146	No association between squalene antibody status and CMI (p=0.465).	presence of squalene antibodies between those with GWI compared to health controls.	whether cases and controls were comparable, lack of reporting of biological measure outliers
Th1/Th2 balance by measuring intracellular production of IFN-γ , IL-2 (Th1), IL-4 (Th2), and IL-10 by CD4 T cells[31] 40 vs 80	**Immune activation in nonstimulated CD4 cells:** *Mean levels of nonstimulated, IL-4C and IL-2C cells:* higher in GWI (for IL-4, P <0.05; for IL-2,P <0.01) *Memory cell cytokine balance:* level of IL-10 producing CD4 cells higher in GWI (p<0.0001) *Non-stimulated cytokine-positive cells:* higher levels of IFN-γ, and IL-2 in GWI *Polyclonally activated cytokine-positive CD4 cells:* higher levels of IL-10+ cells for GWI	IL -4C, -2C and -10 production by CD4 cells was elevated in those with GWI.	Cases and controls not selected from the same populations, using a non-standard definition of GWI, did not account for multiple testing bias
HLA alleles[25] 66 vs 16	Identified 144 HLA alleles in the sample, of which 6 Class II alleles yielded 84.1% correct classification as GWI vs control. All 6 allele frequencies were lower in GWI group (p=0.002) and all ORs < 1, ln(ω)=−1.792 ± 0.383 (mean ± SEM), t=−4.671, DF=5, p=0.005. Negative relationship between overall symptom severity and number of allele copies (t=−4.148, DF=80, p=0.000083, R^2=0.177)	6 HLA alleles correctly classified GWI vs control. These alleles were significantly less frequent in GWI. The number of allele copies were significantly associated with symptom severity. Authors interpretation: reduced HLA protection (*ie*, genetic susceptibility) in Veterans with GWI.	Non-standard case definition, sampling not population-based, lack of adjustment for multiple comparisons

Abbreviations: AA=Arachidonic Acid; ASA=Anti-Squalene Antibody; ATP=Adenosine Triphosphate; BDNF=Brain-derived Neurotrophic Factor; CI=Confidence Interval; CMI=Chronic Multisymptom Illness; CRP=C-reactive Protein; DF=Degrees of Freedom; DHA=Docosahexaenoic Acid; GWI=Gulf War Illness; H-FABP=Heart-type Fatty Acid Binding Protein; HLA=Human Leukocyte Antigen; IFN=Interferon; IL=Interleuken; IPF=Immature Platelet Fraction; LLOD=Lowest Level of Detection; LPC=Lysophosphatidylcholines; LPE=Lysophosphatidylethanolamine; MEG=Magnetoencephalograph; MMP=Matrix Metalloproteinase; MPV=Mean Platelet Volume; MUFA=Monosaturated Fatty Acids; NCM=Neurocutaneous Melanocytosis; OR=Odds Ratio; PC=Phosphatidylcholine; PE=Phosphatidylethanolamine; PL=Phospholipid; PI=Phosphatidylinositol; PUFA=Polyunsaturated Fatty Acids; SEM=Standard Error of the Mean; SFA=Saturated Fatty Acid; SPRC=Standardized Partial Regression Coefficient; TNF=Tumor Necrosis Factor; TPO=Plasma Thrombopoietin; TRAP=Thrombin Receptor Agonist Profile

Central Nervous System Biological Measure Studies

Table 4 provides results of the 10 studies of central nervous system biological measures we identified. All but 1 of the studies measured brain activation, either with functional magnetic resonance imaging (fMRI), MRI, or electroencephalograph (EEG), during the presentation of specific stimuli/tasks. While there were 2 studies with an accompanying replication study each, the body of evidence was otherwise comprised of studies that employed varied stimuli/tasks. Seven studies, 5 with cognitive/emotional tasks[32,33,37-39] and 2 with physical challenges,[34,35] reported differences in brain activation between GWI and comparator groups, while only 2 studies, which involved cognitive/emotional tasks and were each replication/replicated studies, did not identify any group differences in brain activation.[36,40]

In 4 of the 9 CNS studies, the participants engaged in a cognitive task: 1 semantic memory processing task study,[32] 2 face-name association task studies (the second a replication of the first with a different sample and larger sample size),[33,36] and 1 response inhibition task study.[37] In the first (N=47) of the 2 studies employing the face-name association task, no difference in brain activation was reported between the GWI and control groups,[36] while in the replication study, there was greater brain activation in the putamen of GWI Syndromes 1 and 2, each compared to controls (N=67; t=3.63; t=3.45, respectively).[33] Differences in brain activation between participants with GWI and controls were also reported in the study involving a semantic memory task, including differences in brain activation between GWI and controls in the thalamus and caudate head associated with participants correctly indicating a lack of association between words, and a difference in the relationships between reaction time to memory task stimuli and brain activation in GWI Syndrome 2 versus controls (N=38; p=0.02).[32] In the study involving a response inhibition task, brain activation (ERP P3 amplitude) was dampened in the GWI group versus controls (N=48; $F(1, 46)=6.501$, p=0.0142, $\eta 2=.095$).[37]

In another set of 3 studies, brain activation during the presentation of potentially threatening/emotionally triggering stimuli was recorded. In 2 of these studies (1 of them a replication study drawing from a larger, more representative pool of participants[40]) the stimuli were presented auditorily, and, in a third study, visually.[39,40] There were differences in brain activation between GWI and control groups in the original study involving auditory stimuli, with differences in brain activation (EEG P1 amplitude and latency) between Syndromes 2 and 3 compared to controls and Syndrome 1 (N=28; $F(1, 23)=9.915$, p=0.004; ($F(1, 23)=22.025$, p=0.0001)). There were also differences in EEG P3a amplitude between Syndromes 1 and 2 versus controls and Syndrome 3 (N=28; $F(1, 23)=11.172$, p=0.003), and in EEG P3b amplitudes between GWI and controls (N=28; p=0.003),[38] but there were not group differences in the replication study (N=62).[40] There were also group differences in brain activation in the study in which the visual version of the task was presented, with greater brain activity (P3b amplitude) in the control than GWI groups (N=30; p=0.0004).[39]

In 2 studies, the stimuli presented during the measurement of brain activation were innocuous and noxious heat,[34] and physostigmine (to provide an inhibitory cholinergic challenge),[35] respectively. In the study involving innocuous and noxious heat, Syndromes 1 and 2 had greater brain activation to innocuous heat in ventral anterior cingulate (N=54; $p<0.05$) and less activation to innocuous heat in areas involving heat perception (N=54; $p<0.05$), but greater activation to noxious heat in brain areas involving pain, compared to controls. Syndrome 3 had

greater activation to innocuous heat compared to controls only in the dorsolateral prefrontal cortex.[34] In the study involving a physostigmine challenge, change in cerebral blood flow with the challenge was significantly greater in controls than GWI groups (N=47; p=0.014).[35] The remaining study measured the brain chemical N-acetylaspartate, and reported no difference in concentrations in the basal ganglia or pons between GWI and controls.[41]

The body of evidence was limited in that none of the studies used an optimal control group for identifying biological measures that differentiate GWI cases from similar conditions – namely, deployed GWVs with health conditions other than GWI. Further, the outcome assessor was often not reported to be blinded to group status,[32,36-40] participants in the CNS studies were often selected from the same battalion,[32,34-36] decreasing the representativeness of the sample, and, when multiple comparisons were conducted, it was often not reported if statistical corrections were made.[32-35,37,39]

Also of note was the use of the Haley criteria for case definition in most of these CNS studies. Seven[32-34,36,38-40] of the 9 studies identified cases with the Haley criteria, and another used both the CDC and the Haley criteria.[35] The Haley criteria categorizes cases into 6 syndromes based on clusters of symptoms, as described above.

Table 4. Results of Gulf War Illness Central Nervous System Biological Measure Studies

Biological Measure/ Outcome Measure *n GWI vs n comparator*	**Detailed Results** *GWI vs control*	**Summary of Findings**	**Study Limitations**
Percent BOLD signal change (PSC)[32] 26 (syndromes 1, 2, 3: 9, 10, 7) vs 12	**Percent (BOLD) signal change (PSC):** Thalamus PSC increase while correctly indicating no association between 2 given words in Syndrome 2 GWI vs control. Caudate head PSC decrease while correctly indicating no association between 2 given words in GWI Syndrome 1 vs control. Positive association between reaction time and PSC in GWI Syndrome 2 which differed from that of controls (p=0.02). Thalamus PSC associated with an increase in reaction time in GWI Syndrome 2, trend toward statistical difference from all other groups (p<0.08)	Differences in a semantic memory processing task were associated with differences in BOLD PSC in the thalamus and caudate in GWI vs controls. The relationship between reaction time to task stimuli and brain activation also differed between GWI and controls.	Haley criteria case definition, non-representative sample, small sample size, lack of blinding of outcome assessors
Regional brain activation[33] 57 (syndromes 1, 2, 3: 19, 20, 18) vs 10	**Regional brain activation (fMRI BOLD activity):** Left insula (putamen) activation greater in control vs Syndromes 1 and 2 (t=3.63; t=3.45, respectively). No difference in activation in GWI Syndrome 3 vs control.	Greater brain activation in areas of the brain associated with processes including sensation, perception, and emotion during a face-name associative memory task in GWI Syndromes 1 and 2 vs controls.	Haley criteria case definition, small sample size
Brain activation[36] 33 vs 14	**Brain activation (fMRI BOLD activity):** No group differences.	There were no demonstrated differences in brain activation, as measured with fMRI, between GWI and controls during a face-name associative memory paradigm.	Haley criteria case definition, small sample size, lack of outcome assessor blinding
Event-related potential[37] 25 vs 23	**EEG N2 amplitude during behavioral inhibition task:** No interaction between group and condition (F(1, 46)=2.062, p=0.1578). Trend toward a main effect of group (F(1, 46)=3.373, p=0.0727, p=0.0727) **EEG P3 amplitude during behavioral inhibition task:** Interaction between group and condition (F(1, 46)=6.569, p=0.0137, η2=.017) on P3 amplitude. P3 amplitude showed an effect of group (F(1, 46)=6.501, p=0.0142, η2=.095) amplitude, with the control group greater than GWI.	EEG P3 amplitude in GWI during a task requiring behavioral inhibition was dampened compared to in controls.	Non-standard case definition, small sample size, no blinding of outcome assessor

Biological Measure/ Outcome Measure *n GWI vs n comparator*	**Detailed Results** *GWI vs control*	**Summary of Findings**	**Study Limitations**
Event-related potential[38] 20 vs 8	**EEG P1:** Main effect of illness group on P1 amplitude ($F(3, 23)=3.509$, $MS_{error}=2.021$, $p=0.031$), with only 1 of the contrasts significant, Syndromes 2 and 3 compared to controls and Syndrome 1 ($F(1, 23)=9.915$, $p=0.004$, $p=0.004$). Main effect of syndrome group on P1 latency ($F(3, 23)=7.416$, $MS_{error}=115.830$, $p=0.001$), with only 1 of the contrasts significant, Syndromes 2 and 3 had longer latencies vs controls and Syndrome 1 ($F(1, 23)=22.025$, $p=0.0001$). **EEG P3:** No interaction between distractor stimulus type and group, $p>0.273$. Effect of syndrome group on P3a amplitude ($F(3, 23)=4.700$, $MS_{error}=1.188$, $p=0.011$) with only 1 contrast significant, with controls and Syndrome 3 greater than Syndromes 1 and 2 ($F(1, 23)=11.172$, $p=0.003$). There was no omnibus effect of GW Illness syndrome group on P3a latency ($F(3, 23)=1.775$, $MS_{error}=441.372$, $p=0.18$). Effect of group on P3b amplitudes ($p=0.006$), with greater amplitude of the control group was greater than that of GWI $p=0.003$). No effect of group on P3b latency ($p=0.148$).	EEG P1 and P3a responses to auditory stimuli that were previously demonstrated to elicit hyperarousal responses in GWI differed between GWI and control groups.	Haley criteria case definition, small sample size, lack of blinding of outcome assessor
Event-related potential[40] 40 vs 22	**EEG P1 amplitude:** No difference between deployed and nondeployed controls in P1 amplitude ($p=0.550$) or latency ($p=0.555$); data were collapsed into 1 control group. Interaction between group and condition on P1 amplitude ($p=0.014$), but no significant contrasts including control group comparisons. **EEG P1 latency:** There was neither an effect of GWI syndrome group on P1 latency ($F(3, 58)=0.748$, $MSE=275.714$, $p=0.528$) nor an interaction between group and condition ($F(9, 174)=0.955$, $MS_{error}=75.297$, $p=0.479$).	No differences between GWI and control groups in EEG responses during a task involving the presentation of auditory stimuli previously demonstrated to elicit hyperarousal responses in GWI.	Haley criteria case definition, small sample size, no reported blinding of outcome assessors
Event-related potentials[39] 22 vs 8	**EEG P3 amplitude:** No effect of group on P3a amplitude ($F(3, 26)=.339$, $p=0.7973$). Significant effect of group on P3b amplitude ($F(3, 26)=5.282$, $p=0.0056$, $\eta 2=0.3787$), with control group amplitude higher than ill groups ($p=0.0004$). **EEG P3 latencies:** Latencies of P3a and P3b did not differ significantly by group ($p>0.27$).	EEG P3b amplitude during a task involving the presentation of visual trauma-related stimuli was greater in control groups than in GWI.	Haley criteria case definition on, no reported blinding of outcome assessors, small sample size

Biological Measure/ Outcome Measure *n GWI vs n comparator*	**Detailed Results** *GWI vs control*	**Summary of Findings**	**Study Limitations**
Brain activation[34] 40 vs 14	**Brain activation to innocuous heat:** Syndromes 1 and 2 GWI greater ($p<0.05$) activation to innocuous heat than controls in ventral anterior cingulate and less activation than controls in areas involving heat perception ($p<0.05$). Syndrome 3 greater activation than controls in left DLPFC, but no other areas. **Brain activation to noxious heat:** Syndromes 1-2 hyper-activation ($p<0.05$) compared to controls in regions involving pain. Syndromes 1-3 GWI significant ($p<0.0001$) activation to noxious heat in similar areas as controls. Syndrome 1 higher noxious heat activation than controls in left amygdala, arahippocampal gyrus, thalamus, and right basal ganglia. Syndrome 3 activation was not different from controls ($p>0.05$).	GWI Syndromes 1 and 2 had greater brain activation to innocuous heat in the ventral anterior cingulate than controls, and less activation to innocuous heat in brain areas involving heat perception, but greater activation to noxious heat in brain areas involving pain. Syndrome 3 had greater activation to innocuous heat than controls in only the left DLPFC, and no different activation to noxious heat than controls.	Haley criteria case definition on, non-representative sample, insufficient sample size, incomparability of cases and controls
Cerebral blood flow[35] 33 vs 14	**Cerebral blood flow with physostigmine challenge:** GWI (Syndrome groups combined) and controls CBF change significantly greater in control ($p=0.014$).	Change in cerebral blood flow with an inhibitory cholinergic challenge significantly greater in control than GWI groups, indicating persistent cholinergic deficits in GWI.	Non-representative sample, small sample size
Quantities of N-acetylaspartate in the basal ganglia and pons[41] 81 vs 97	No difference in PON1 genotype or paraoxonase activity between controls and GWI. No group differences in N-acetylaspartate concentrations in the left basal ganglia and pons.	No differences in concentrations of N-acetylaspartate were found in the basal ganglia or pons of GWI vs controls.	Multiple comparisons without correction.

Abbreviations: BOLD=Blood-Oxygen-Level-Dependent; CBF=Cerebral Blood Flow; DLPFC=Dorsolateral Prefrontal Cortex; EEG=Electroencephalogram; GWI=Gulf War Illness; fMRI=functional Magnetic Resonance Imaging; PSC=Percent Signal Change

Autonomic Nervous System

Table 5 provides results of the 5 studies examining various autonomic nervous system biological measures, including cardiovascular measures (using an electrocardiogram [ECG] or tilt-table test)[42-46] and nervous system measures (using nervous conduction/reflex testing).[44,45] There were biological measures in each of the studies that differed significantly between cases and controls.

Three studies examined primarily cardiovascular biological measures. In 1 study (N=184),[42] measurement of heart rate with an ECG over 24 hours indicated increased randomness of beat-to-beat heart rate changes (measured by the short-term fractal scaling exponent [DFA1]) in the CMI group versus control group (1.28±0.16 versus 1.35±0.15; p=0.005). In the second study focused on cardiovascular measurements, (N=27)[43] maximum heart rate variation ($p<0.05$) and maximum heart rate response to tilt ($p<0.05$) in a tilt test plus isoproterenol was greater in cases than controls. A study examining cardiorespiratory and metabolic responses to maximal exercise test found no difference in exercise capacity between GWVs with Chronic Fatigue Syndrome (CFS) and healthy GWVs controls.[46]

Two studies examined both cardiovascular and nervous system biological measures. In a 2013 (N=97) study,[44] 24-hour heart rate variability (HRV) using an ECG was measured in 16 healthy Veteran controls and 3 other groups stratified by symptom domains (n=66): Syndrome 1 (impaired cognition, n=21); Syndrome 2 (confusion/ataxia, n=24); and Syndrome 3 (central neuropathic pain, n=21). High frequency HRV increased normally at night in control group, but not in syndrome groups ($p<0.001$). The syndrome groups had reduced distal postganglionic sudomotor function compared to controls. In the quantitative sensory tests, the confusion/ataxia group (Syndrome 2) had increased cooling detection threshold versus controls ($p<0.05$). A 2014 cross-sectional study (N=28)[45] examined changes in blood pressure and heart rate (using a tilt-table test), nerve conduction, sensory testing, and sudomotor axon reflex testing in self-reported post-exertional fatigue (PEF, also known as post-exertional malaise [PEM]) versus controls. There was no significant difference in blood pressure, but there was in supine (p=0.003) and standing ($p<0.001$) heart rate between groups at baseline. There were no significant differences in thermal or vibration threshold testing on hands and feet, or large fiber nerve variables on sural sensory and peroneal motor nerves.

Only 2[42,46] of the 5 studies used a "gold standard" definition for GWI/CMI/CFS (CDC or Kansas), with the others using ICD-9 codes,[43] self-report[45] for post-exertional fatigue, or another validated (Haley) definition/characterization of GWI[44] as criteria for inclusion into the study. The sample sizes were relatively small: 2 studies had fewer than 30 total participants,[43,45] 1 had fewer than 100,[44] and 1 had fewer than 200 Veterans.[42]

Table 5. Results of Gulf War Illness Autonomic Nervous System Biological Measure Studies

Biological Measure/ Outcome Measure *n GWI vs n comparator*	**Detailed Results** *GWI vs control*	**Summary of Findings**	**Study Limitations**
24-hr HRV, urinary catecholamines and cortisol, hypertension, insulin sensitivity, dyslipidemia, body fat, bone mineral density, and ultrasensitive CRP[42] 184 (n=73 with CMI) vs 111	*Adjusted Least Squares Mean difference (95% CI):* **HR:** 1.85 (-0.76 to 4.45). p=0.16 **SDNN (SD of the N-N interval):** -6.64 (-16.64 to 3.35). p=0.19 **SDNN index:** -0.07 (-0.16 to 0.02). p=0.14 **VLF:** -0.15 (-0.32 to 0.03). p=0.10 **LF:** -0.15 (-0.36 to 0.06). p=0.16 **DFA1 (short-term fractal scaling exponent)***:* -0.03 (-0.07 to 0.01). p=0.03 **HPA measures:** Plasma and 24-hour urinary cortisol levels following overnight dexamethasone suppression testing did not differ between groups.	Values for a nonlinear heart-rate-variability parameter (the short-term fractal scaling exponent [DFA1]) were lower in cases than controls, but there were no group differences in HPA measures (cortisol levels following dexamethasone suppression testing).	Sample not representative of whole GW population, distribution of biological measure data not reported, there were significant descriptive differences that were not adjusted for, enrollment non-response rate not reported, unclear if outcome assessor was blinded.
Heart rate, blood pressure[43] 14 vs 13	**Positive tilt response:** not different between the 3 groups (p=0.098). **70-degree head-up tilt:** Maximum heart rate response to tilt was higher in the fatigued GWVs than in each of the control groups (p<0.05). Maximum HRV was highest in the GWVs fatigued group (p<0.05). No group difference in fall in systolic blood pressure with tilt.	Maximum heart rate response and variation to 70-degree tilt was higher in GWVs with fatigue symptoms than in controls.	Non-standard definition for GWI, sample not representative of whole GW population, sample size/power calculation not provided, distribution of biological measure data not reported, there were significant descriptive differences that were not adjusted for, enrollment non-response rate not reported.
Nerve conduction, sensory testing, sudomotor axon reflex testing[45] 16 vs 12	**Cardiovascular measures:** *BP:* no differences at supine, active standing, or 70 degrees tilt. *HR*: greater in GWI group at 5 min supine (P<0.001) and 3 min standing (p=0.003) *Absolute heart rate increment (by HUT):* no difference *HRV to Valsalva maneuver and HRV deep breathing*: no difference	Greater supine and standing heart rates in cases versus controls, but no differences in neurological measures (large fiber nerve variables, or in thermal or vibration thresholds on hands and feet).	Definition for GWI other than CDC/Kansas used, sample not representative of whole GW population, sample size/power calculation not provided, there were significant descriptive differences that were not adjusted for, distribution of biological measure data not reported, enrollment non-response rate not

Biological Measure/ Outcome Measure *n GWI vs n comparator*	**Detailed Results** *GWI vs control*	**Summary of Findings**	**Study Limitations**
	Large and small fiber nerve variables: No group differences for large fiber nerve variables measured on sural sensory (amplitude: p=0.598, latency: p=0.235) and peroneal (amplitude: p=0.178; latency: p=0.462) motor nerves. No difference between groups in averaged sweat volume from the feet. **Quantitative sensory testing:** No group difference for thermal (hands [left, right]: p=0.625, p=0.582; feet: p=0.491, p=0.925) and vibration (hands [left, right]: p=0.076, p=0.402; feet: p=0.500, p=0.031) threshold on hands and feet.		reported, unclear if outcome assessor was blinded.
CASS, and high-frequency heart rate variability from a 24-hour ECG[44] 66 (syndromes 1, 2, 3: 21, 24, 21) vs 16	**CASS:** varied across groups (p=0.45) and was higher in syndrome 2 than controls (p=0.02). Syndrome groups had reduced distal postganglionic sudomotor function in the foot (p=0.02), ankle, and upper leg, but not in the arm, compared to controls. No group differences in tear production, sympathetic adrenergic function, or pupillary measures. **Quantitative Sensory Tests:** Syndrome 2 increased cooling detection threshold vs controls (p<0.05). No group difference on heat pain threshold. **Circadian variation in parasympathetic tone:** High frequency HRV increased normally at night in control group, but not syndrome groups. High frequency HRV during the day for syndrome 1 not different from controls, but syndrome 2 significantly lower, and syndrome 3 significantly higher.	GWI groups had neurological differences from controls, with reduced distal postganglionic sudomotor function in the foot, and increased cooling detection threshold in syndrome 2, as well as cardiovascular differences from controls, with abnormal HRV at night, and lower and higher high frequency HRV during the day for syndromes 2 and 3, respectively.	Definition for GWI other than CDC/Kansas used, sample size/power calculation not provided, distribution of biological measure data not reported, enrollment non-response rate not reported.

Biological Measure/ Outcome Measure *n GWI vs n comparator*	**Detailed Results** *GWI vs control*	**Summary of Findings**	**Study Limitations**
Cardiorespiratory and metabolic responses to maximal exercise test[46] 19 v 19	Maximal oxygen uptake: 28.9 (6.7) mL/kg/min vs 30.8 (7.1) mL/kg/min; p=0.39 Heart rate: 155.8 (16.1) bpm vs 163.3 (14.9) bpm; p=0.17 Exercise time: 9.6 (1.5) minutes vs 10.2 (1.4) minutes; p=0.26 Workload achieved: 208 (36.7) W for CFS vs 224 +/- 42.9 W for controls; p=0.25 Submaximal intensities: ND (p > 0.05) *W=workload	No difference in exercise capacity between CFS and healthy controls.	Small, non-representative sample, data modeling not addressed, multiple comparisons without adjustment, response rate not reported.

Abbreviations: BP=Blood Pressure; BPM=Beats Per Minute; CASS=Composite Autonomic Severity Score; CDC=Centers for Disease Control and Prevention; CFS=Chronic Fatigue Syndrome; CI=Confidence Interval; CMI=Chronic Multisymptom Illness; CPTh=Cold Pressor Threshold; CRP=C-reactive Protein; ECG=Electrocardiogram; GI=Gastrointestinal; GW=Gulf War; GWI=Gulf War Illness; GWV=Gulf War Veterans; HPA=Hypothalamic-Pituitary-Adrenal Axis; HPTh=Heat Pain Threshold; HR=Heart Rate; HRV=Heart Rate Variability; HUT=Head-Up Tilt; LF=Low Frequency; ND=No Difference; SD=Standard Deviation; VLF=Very Low Frequency

Genetic Biological Measures

Table 6 provides results of the 1 study examining associations between genetic measures and GWI. This study (N=190) of paraoxonase (PON1) genotypes and levels of PON1 activity in serum detected slightly lower levels of PON1 activity among Veterans with GWI, though the difference was not statistically significant (median difference 23.1; 95% CI=-27.7 to 73.9).[47] This study examined genotype variants resulting from amino-acid substitutions at positions 55 and 192, which are associated with different levels of enzymatic activity on various substrates. No difference in the distribution of PON1-192 genotypes was observed (p=0.52). The L to M substitution of PON1-55 occurred more frequently among healthy GWVs (61.9%) compared with Veterans with GWI (41.2%), and the difference was statistically significant (p=0.02). Statistical correction for multiple comparisons was not performed, and the applicability of this study may be further limited by the use of a non-standard case definition for GWI. Non-population-based sampling in this study limits the interpretation of findings. While there was a genetic component in some of the other studies represented in other biological system sections, studies were categorized in the genetic section if the genetic measures were the emphasis of the study. For example, a study of HLA alleles is included with the previous section on immune system biological measures.[25]

Table 6. Results of Gulf War Illness Genetic Biological Measure Studies

Biological Measure/ Outcome Measure *n GWI vs n comparator*	**Detailed Results** *GWI vs control*	**Summary of Findings**	**Study Limitations**
PON1 activity and genotype[47] 95 vs 95/85	**PON1 activity, median (range):** 145.85 (48.3–423) vs 168.92 (59.3–521); p=0.03 GWI had slightly lower PON1 activity, but not statistically significant: Mean Difference=23.1 (95% CI -27.7 to 73.9) **High-Density Lipoprotein (HDL):** lower in GWI; Mean Difference=0.14 (95% CI: 0.04 to 0.25) **PON1 genotype:** No difference in the distribution of PON1-192 genotype (p=0.52). Difference between ill and well groups in PON1-55 genotype (p=0.02), with a higher proportion of D-GWVs healthy with LM genotype compared with GWI.	No statistically significant differences in PON1 activity between GWI and GWVs healthy.	Non-standard case definition (SF-36) and multiple comparisons.

Abbreviations: CI=Confidence Interval; D-GWV=Deployed Gulf War Veteran; GWI=Gulf War Illness; GWV=Gulf War Veteran; HDL=High Density Lipoprotein; PON1=Paraoxonase; SF-36=Short Form-36 Question Questionnaire

Other Biological Systems

We identified 6 studies of other biological measures of GWI not categorized under the preceding biological systems. These included: 1 study of sleep parameters,[48] 1 study examining cholinesterase and serum cytokines,[21] 1 of neurophysiologic markers,[50] a study of audiovestibular function,[49] 1 study of pain tolerance, 1 study of human herpes viruses (HHVs),[51] and 1 study of temperature and pain thresholds.[52] Table 7 contains these study details.

Three studies found significant differences in biological measures assessed between GWVs with GWI and controls. In a study of sleep parameters and sleep-related respiratory parameters in GWVs with GWI versus healthy GWVs, sleep parameters did not differ, but sleep-disordered breathing (measured by sleep respiratory parameters: Apnea-Hypopnea Index [AHI], Respiratory Effort Related Arousal [RERA] index, and percent flow-limited breaths) was greater in those with GWI: RERA index ($p=0.018$); AHI ($p=0.006$); percent flow-limited breaths ($p<0.0001$).[48] There were also significant differences between GWI and controls within all audiovestibular testing domains in a study of audiovestibular function in GWVs with the 3 Haley-defined syndromes and healthy GWVs controls: greater GWI interocular asymmetry of gain in rotational nystagmus; diminished GWI nystagmic velocity after caloric stimulation; greater GWI interaural asymmetry of caloric response; greater GWI saccadic velocity; more frequent GWI pathologic nystagmus; greater GWI interpeak latency differences between ears; lower GWI response strength from right component of platform posturography.[49] One study examined only differences in nervous system functioning, namely pain tolerance and threshold. A 2018 study ($N=91$)[52] examined hot, cold, and ischemic pain tolerances and thresholds in GWVs with "documented" (no definition specified) GWI and gastrointestinal (GI) symptoms (n=53), GWVs with "documented" GWI and no GI symptoms (n=47), and GWVs without GWI or GI symptoms (n=38). Veterans with GWI and GI symptoms showed lower pain thresholds for each stimulus ($p<0.001$) compared with controls, and 20% of Veterans with GWI and GI symptom showed hypersensitivity to all 3 stimuli.

In the remaining 4 studies, there were no differences in biological measures assessed between GWVs with GWI and controls. Notably, 1 of these studies used our ideal comparator group – deployed GWVs with a non-GWI health condition (in this case, posttraumatic stress disorder [PTSD]). Compared to GWVs with PTSD (but no GWI) and 4 healthy GWVs, the GWVs with GWI ("neurologic factor") did not have significantly different activity in the "organophosphate detoxifying" enzymes (Butyrylchoninesterase [BuChE] and PON1), that, according to the authors, suggests that organophosphate exposure is not associated with neurological symptoms of GWI.[21] In the study of neurophysiological markers, GWVs reporting neuromuscular symptoms had no quantitative evidence of neurological disorders and were not significantly different from healthy GWVs.[50] Finally, in a study comparing prevalence of HHVs in GWVs with CFS versus healthy GWVs, there was no difference in prevalence of HHV7 between groups.[51]

All 6 studies suffered from methodological issues. Notably, they had small sample sizes that did not provide sufficient power to detect differences or power calculations were not described, insufficient or no description of data modeling such as normality of data and outlier detection, nor adjustment for multiple comparisons. Additional issues in some studies included unequal or unreported non-response rates and outcome assessor blinding not used or reported. For these

reasons, and due to the fact that there were no 2 studies addressing the same biological system and/or marker, these findings do not provide sufficient evidence for GWI association, or lack thereof, with any of these biological measures.

Table 7. Results of Gulf War Illness Other Biological Systems Biological Measure Studies

Biological Measure/ Outcome Measure *n GWI vs n comparator*	**Detailed Findings** *GWI vs control*	**Summary of Findings**	**Study Limitations**
Sleep parameters and sleep-related respiratory parameters[48] 18 v 11	**Sleep parameters:** Total sleep time: 333 (95) min vs 329 (75) min; p=0.90 Sleep efficiency: 73% (24%) vs 76% (16%); p=0.66 Sleep latency: 13 (15) minutes vs 28 (30); p=0.08 REM latency: 137 (69) minutes vs 136 (65) minutes; p=0.97 NREM 1: 29% (17%) vs 22% (14%); p=0.26 NREM 2: 48% (13%) vs 46% (11%); p=0.74 Slow wave sleep: 12% (9%) vs 1%7 (11); p=0.17 REM: 12% (7%) vs 11% (7%); p=0.73 Arousals/hour: 34 (26) vs 10 (6); p=0.006 Total stage shifts: 39 (12) vs 40 (14); p=0.89 **Sleep-related respiratory parameters:** RERA index: 16 (12) vs 6 (4); p=0.018 AHI: 18 (25) vs 3 (5); p=0.006 % flow-limited: 96 (5) vs 36 (25); p<0.0001	Significantly greater sleep-disordered breathing in Veterans with GWI: RERA index (p=0.018); AHI (p=0.006); % flow-limited (p<0.0001). No difference in sleep parameters except arousals/hour (p=0.006).	Small, non-representative sample, data modeling not adequately addressed, multiple comparisons without adjustment, response rate not reported.
Integrity of auditory pathways (inner ear through upper brain stem and vestibulocular reflex)[49] Syndrome 1 vs 2 vs 3 vs control: 5 vs 13 vs 5 vs 10	**Sinusoidal harmonic acceleration:** Greater interocular asymmetry of gain in rotational nystagmus in GWI vs controls. Asymmetry values of 0.01, 0.02, and 0.04 Hz, differed from controls for Syndrome 1 (p=0.015), Syndrome 2 (p=0.002), but not for Syndrome 3 (p=0.8). In controls, the magnitude of asymmetry decreased monotonically, but not in any GWI groups. **Nystagmic velocity after caloric stimulation:** diminished in Syndrome 3 vs controls for all 4 irrigations (cool right, p=0.02; cool left, p=0.004; warm right, p=0.009; warm left, p=0.004). Interaural asymmetry of caloric response greater in Syndrome 2 than controls (p=0.07). **Asymmetry of saccadic velocity:** greater in Syndrome 2 than controls (p<0.05) **Nystagmus:** pathologic nystagmus in 4 ill Veterans, but none of the controls (p=0.09) **Inter-side asymmetry of wave I to III interpeak latency on auditory brain stem response:** wave 1 to 3 interpeak latency differences between ears greater in cases than controls (p=0.02) - in Syndromes 1 (p=0.005) and 2 (p=0.07), but not 3. Similar number of cases and controls with unilateral latencies or amplitudes exceeding normal limits. **Platform posturography:** Syndrome 3 lower response strength from right and left forward components of the platform than controls (p=0.10) No difference between GWI and controls in somatosensory, visual, and visual preference ratios	Within all audiovestibular testing domains, there were differences between GWI and controls.	Haley criteria for case definition, non-representative sample, inclusion/exclusion not equally applied across groups, small sample size, blinding of outcome assessor not reported.

Biological Measure/ Outcome Measure *n GWI vs n comparator*	**Detailed Findings** *GWI vs control*	**Summary of Findings**	**Study Limitations**
Cholinesterase; Serum cytokines[21] 25 vs 14 vs 4	*Serum enzyme activity:* **PON1:** 577 (81) vs 479 (107) vs 518 (248) µmol/mL/min (no differences) **Arylesterase:** 111 (3) vs 102 (7) vs 116 (8) µmol/mL/min (no differences) **Serum BuChE activity:** 0.63 (0.03) vs 0.64 (0.04) vs 0.65 (0.07) µM/mL/min (no differences) **Serum cytokines (Th1, Th2, and proinflammatory):** both GWI and PTSD groups had serum levels ≥20% higher than highest level expressed by controls	"Neurologic factor" in GWVs not correlated with low activity of the enzymes BuChE and PON1	Not standard case definition, small sample size (+ very small control group), comparability of cases and controls not established, non-response rate and blinding NR
Neurophysiologic assessment: nerve conduction studies; quantitative sensory and autonomic function testing; concentric needle and single-fiber EMG[50] 49 vs 26	**Nerve conduction studies:** No evidence for axonal or demyelinating peripheral neuropathy. Comparisons of motor conduction measurements from median, ulnar, and common peroneal nerves showed no major peripheral nerve abnormalities. No differences. P-values NR throughout. **Quantitative sensory and autonomic assessments:** No specific abnormalities in GWI group. No differences in vibration perception, warm and cool sensory thresholds; thermoregulatory function in face and limbs, or cardiovascular reflexes. **Concentric needle EMG:** Combined mean polyphasic units in the 4 muscles examined: 6.1% vs 7.7% Duration or amplitude of MUAPs: ND Turns analysis: 240±96 vs 299±82 turns/s (ND, p-value NR); No difference in ratio of number of turns/second to mean amplitude **Single-fiber EMG:** no difference in MCD or fiber density values	Results for GWI similar to controls: Peripheral nervous system functional; no chronic denervation or myopathic abnormalities; no impulse blocking. No evidence of peripheral neurological disorders	Not standard case definition; small sample size; unequal non-response rate
HHV6, HHV7, EBV, and cytomegalovirus[51] 46 vs 32	*Prevalence of:* **HHV6 DNA:** Detected in 1 control pt. **HHV7 DNA:** 22/46 (47.8%) vs 14/32 (43.8%); p=0.82 **EBV DNA:** Detected in 1 CFS pt. **HCMV DNA:** None detected	No difference in HHV7 infection rates	No description of sampling strategy, power NR, data modeling NR and no adjustment for multiple comparisons, non-response rate NR
Heat pain threshold, cold pressor pain threshold; ischemic pain threshold and ischemic pain tolerance[52] 100 (GWI+GI symptoms) vs 38 healthy	**Heat Pain Threshold (HPTh):** GWI+GI had lower HPTh compared with GWI no GI symptoms (p<0.01) and controls (p<0.001). No significant differences between controls and GWI no GI symptoms. **Cold Pressor Test (Cold Pressor Threshold [CPTh]):** GWI+GI had lower CPTh compared with GWI-no GI symptoms (p<0.01) and vet controls (p<0.001). No differences between control and GWI-no GI symptoms. **Ischemic Pain Threshold and Ischemic Tolerance Test:** GWI+GI had shorter time to ischemic pain threshold (p<0.001) than controls. GWI-no GI symptoms also had significantly shorter time to ischemic pain threshold (p<0.01) compared to controls. GWI+GI had lower ischemic pain tolerance	Veterans with GWI and GI symptoms showed lower pain thresholds to heat, cold, and tourniquet tests (p<0.001) compared with controls, and 20% of the GWI+GI vets showed hypersensitivity to all 3 stimuli.	Definition for GWI other than CDC/Kansas used, sample not representative of whole GW population, sample size/power calculation not provided, distribution of biological measure data not reported, did not adjust for multiple comparisons, enrollment non-response

Biological Measure/ Outcome Measure *n GWI vs n comparator*	**Detailed Findings** *GWI vs control*	**Summary of Findings**	**Study Limitations**
	compared to controls (p<0.001). GWI-no GI symptoms also had significantly shorter time to tolerance compared to controls (p<0.01).		rate not reported, unclear if outcome assessor was blinded.

Abbreviations: AHI=Apnea-Hypopnea Index; BPM=Beats Per Minute; BuChE= Butyrylchoninesterase; CFS=Chronic Fatigue Syndrome; DNA=Deoxyribonucleic Acid; EBV=Epstein-Barr Virus; EMG=Electromyography; GWI=Gulf War Illness; GWV=Gulf War Veteran; HCMV=Human Cytomegalovirus; HHV=Human Herpesvirus; MCD=Mean Consecutive Difference; MUAP=Motor Unit Action Potential; ND=No Difference; NR=Not Reported; NREM=Non-Rapid Eye Movement; PON1=Paraoxanase; PTSD=Posttraumatic Stress Disorder; REM=Rapid Eye Movement; RERA=Respiratory Effort Related Arousal

KEY QUESTION 3: Which ongoing or unpublished research studies examine diagnostic tests or biological measures for potential association with GWI?

Table 8 provides results of the 24 ongoing research studies examining a wide variety of potential biological measures for GWI. Five studies are looking at how GWI impacts the central nervous system in GWVs.[53-57] Three studies[58-60] plan to look at genetic components associated with GWI (including 1 using data from the Million Veterans Program[59]). Five studies are examining the immune system.[61-65] One study will look at autonomic function testing in Veterans with GWI,[66] and 10 studies are examining biological measures in other biological systems.[67-76]

Of those studies proposing to examine measurements within other biological systems, 3 include measures of mitochondrial dysfunction[67-69] and 2 of the gut microbiome.[70,71] The 3 studies that include measures of mitochondrial dysfunction propose to examine mitochondrial and bioenergetic impairments,[68] peripheral mononuclear cells (PBMCs) to determine mitochondrial function,[67] and mitochondria and peroxisome function and lipids specific to inflammation.[69] Studies of gut microbiomes will look generally at differences in gut microbiomes between GWVs with and without GWI,[70] and at microbes of the small intestine.[71]

The remaining studies will examine plasma proteomics;[72] plasma metabolomes;[73] serum analytes;[74] associations between adrenal, immune, inflammatory and coagulation in Veterans with GWI;[75] and XMRV, a retrovirus.[76]

For all studies with control group information available, the comparator is healthy GWVs. No study uses our ideal comparator of ill GWVs without GWI.

Of note, the VA Cooperative Studies Program (CSP) 585,[77] also known as The Gulf War Era Cohort and Biorepository, has enrolled 1,275 Veterans. These Veterans have submitted to VA researchers blood samples to analyze for research related to health conditions and other related factors. While we were unable to identify published studies using this particular biorepository, we did identify 1 study using data from VA CSP 2006, which analyzes the genomics of Gulf War Illness.[59]

Table 8. Ongoing Studies of Biological Measures for Gulf War Illness

Details: PI *Study design* *Registration/award no.* *Study sponsor* *Setting & status*	**Focus of Investigation**	**Anticipated Participants**	**GWI Case Definition** *Selection criteria/ population*	**Biological Measure/Outcome Measure**
CENTRAL NERVOUS SYSTEM				
Killiany, R.[56] Cross-sectional W81XWH-19-1-0765 CDMRP GWIRP Boston University Medical Campus	Chronic inflammation	20 GWVs: 10 GWI and 10 healthy controls	Case definition: Kansas Participants will be recruited from the BBRAIN for GWI	Astrocyte and microglial activation using PET
Peskind, E. R.[53] Study design: NR Project number: 1I01CX001049-01 NIH VA Puget Sound Health Care System Started: Jan 1, 2014 Anticipated completion: Dec 31, 2017	Brain metabolism, neuronal damage; and abnormalities in central and peripheral systems regulating pain perception, fatigue, and sleep	NR	Case definition: NR Criteria not specified	• Cerebral glucose metabolism in brain regions relevant to cognition (*eg*, medial temporal lobes) using fluorodeoxyglucose-PET • Structural and compositional structural integrity using MRI, diffusion MRI and MPF mapping • Brain regional connectivity among nodes of the ventral and dorsal attention networks on blood oxygen level dependent functional connectivity MRI • CSF biological measures associated with neurodegeneration (decreased Ab42, increased CSF total tau and phosphorylated tau (ptau181) and oxidative damage (increased F2-isoprostanes), and decreases in the neurotrophin, brain-derived neurotrophic factor • Pain sensitivity by Quantitative Sensitivity Testing and impaired activation of endogenous opioids in response to Conditioned Pain Modulation • Abnormalities in neuropeptides, neurotransmitters, hormones, and immune factors associated with pain and fatigue perception and sleep. • Cerebral glucose metabolism in brain regions modulating sensory pain (*eg*, thalamus). • Frequency of the apolipoprotein E (APOE)-e4 - allele, the microtubule associated protein tau

Details: PI *Study design* *Registration/award no.* *Study sponsor* *Setting & status*	**Focus of Investigation**	**Anticipated Participants**	**GWI Case Definition** *Selection criteria/ population*	**Biological Measure/Outcome Measure**
				(MAPT) H1 haplotype, the Met allele of the brain derived neurotrophic factor Val66Met variant, the Val allele of the catechol-O-methyl transferase Val158Met variant, the G allele of the mu opioid receptor 1 A118G single nucleotide polymorphism (*ie*, rs1799971), the Arg allele of the PON1 Gln192Arg variant, and decreased functional activity of PON1. • DNA methylation levels in CpG Islands in the PON1, APOE, MAPT, and BDNF genes.
Steele, L.[55] Case-control W81XWH-14-1-0622 CDMRP Baylor College of Medicine Data collection was scheduled to start in 2016	Nigrostriatal pathway; brainstem and basal ganglia integrity	80 GWVs with GWI; 50 healthy GWVs controls	Case definition: NR "Well-characterized" sample of 1990-91 GWVs	Corticostriatal circuit using high-resolution DTI
Sullivan, K.[54] Longitudinal analysis W81XWH-19-1-0767 CDMRP GWIRP Boston University Medical	Brain health (all major aspects)	N=100, GW Veterans 50 cases and 50 controls	Case definition: Kansas GWVs with GWI And healthy GWVs recruited from BBRAIN	CBF patterns, BBB permeability, and WM microstructural integrity using DKI and HARDI
Younger, J. W.[57] Cross-sectional W81XWH-19-1-0725 CDMRP GWIRP University of Alabama, Birmingham	Neuroinflammation	40 GWVs (deployed): 20 GWI vs 20 healthy controls	Case definition: Kansas Recruited from existing database of individuals interested in research	Whole-brain MRSI scan metabolite concentrations: myoinositol, lactate, choline, NAA, and absolute brain temperature
GENETIC				
Haley, R.[58] Cross-sectional W81XWH-15-1-0672	Whole genome gene-expression	140 GW-era Veterans: 4 clinical groups (the 3 GWI	Case definition: CDC definition of MSI, Factor	Level of gene expression of the messenger RNAs and micro-RNAs in pure suspensions of T

Details: PI *Study design* *Registration/award no.* *Study sponsor* *Setting & status*	**Focus of Investigation**	**Anticipated Participants**	**GWI Case Definition** *Selection criteria/ population*	**Biological Measure/Outcome Measure**
CDMRP GWIRP University of Texas, Southwestern Medical Center at Dallas		variants of the Factor case definition and a control group)	case definition subset (Haley definition) From "samples who participated in our prior studies"	lymphocytes after stimulating aliquots with LPS or ACh
Malanoski, A. P.[60] Case-control CDMRPL-17-0-GW160096 CDMRP GWIRP Naval Research Laboratory	Epigenetics	125 GW Veterans, 75 GWI cases, 50 controls	Case definition: Kansas, CDC case criteria obtained for comparison purposes GWVs with GWI or healthy; participants of Boston GWIC biorepository	DNA methylation, micro RNA expression
Provencale, D.[59] Nested case-control VA CSP #2006 VA MVP	Genome-wide association study	7,500 GWI case participants; 7,500 controls GWVs (estimated)	Case definition: NR 1990-1991 GWVs (both deployed and nondeployed) who participated in the MVP	Potential genetic risk factors for GWI utilizing SNP genotyping
IMMUNE SYSTEM				
Abou Donia, M. B.[61] Cross-sectional + follow-up after acupuncture treatment W81XWH-19-1-0465 CDMRP GWIRP Duke University	Plasma/serum autoantibodies; Neural cells	150 GWVs: 100 with GWI (50 with CFS and 50 without) and 50 healthy GWVs	Case definition: NR NR	IgG-class autoantibodies for CNS markers in the saliva, serum, and plasma compared to correlate brain volumetric and microstructural alterations on brain imaging
James, L.[62] Cross-sectional W81XWH-17-1-0677 USAMRDC	Genes and biological measures of immune system dysfunction	Large sample of GWVs with and without GWI: randomly selected from a DoD list of	Case definition: NR NR	Genes associated with immune system functioning, and biological measures of immune system dysfunction, inflammation, and autoimmunity

Details: PI *Study design* *Registration/award no.* *Study sponsor* *Setting & status*	**Focus of Investigation**	**Anticipated Participants**	**GWI Case Definition** *Selection criteria/ population*	**Biological Measure/Outcome Measure**
University of Minnesota, Twin Cities Recruiting as of Oct 2018		66,229 people in MN who served GW from August 2, 1990-April 11, 1991		
Klimas, N.[63] Double-blind RCT NCT02848417 SFVAFRE, Miami Recruiting as of Mar 2020 Expected: Aug 2020	Intervention study (nutraceuticals) – homeostatic network biological measure response to treatment	75 Veterans with GWI	Case definition: NR Inclusion Criteria: • Veterans with GWI • 35-70 years old • Good health by medical history prior to 1990 • No diagnoses exclusionary for GWI	1. Biological measure response to therapy using a VO2 exercise test 2. Biological measure response to therapy using cytokine panel
Monson, N.[64] Nested case-control (4 groups) W81XWH-17-1-0586 CDMRP GWIRP University of Texas, Southwestern Medical Center, Dallas	Molecular blood biological measures	Development sample (n=142): 96 cases + 46 matched controls Replication sample (n=142): cases and controls group-matched to development sample Longitudinal sample (n~120) Neuroinflammatory sample (n=100)	Case definition: CDC, Factor, and modified Kansas Representative of the GW-era military population in the US Military Health Survey; and a sample of clinical patients with other neuroinflammatory diseases.	Autoantigen arrays, analyte arrays, custom Luminex bead-based arrays to detect neuronal, glial, immune, and neuroinflammatory mediators with high sensitivity in plasma (cytokines, chemokines, autoantibodies, autoantigens, analytes, and other proteins); genome-wide extent of methylation of CpG sites on banked DNA
Shungu, D. C.[65] Case-control W81XWH-15-1-0437 CDMRP GWIRP Weill Cornell Medicine	Neuroinflammation, oxidative stress, and mitochondrial dysfunction	40 GWVs: 20 with GWI vs 20 without GWI	Case definition: NR Criteria not specified	Binding potential of the ligand with PET; glutathione, brain levels of lactate and NAA, ATP, PCr, Pi, and phosphomonoesters and phosphodiesters all with proton MRS; cerebral blood flow with arterial spin-labeling MRI. Complementary: markers of neuroinflammation and oxidative stress will also be measured in the CSF

Details: PI *Study design* *Registration/award no.* *Study sponsor* *Setting & status*	**Focus of Investigation**	**Anticipated Participants**	**GWI Case Definition** *Selection criteria/ population*	**Biological Measure/Outcome Measure**
AUTONOMIC NERVOUS SYSTEM				
Barnes, J. N.[66] Cross-sectional W81XWH-19-1-0381 CDMRP GWIRP University of Wisconsin, Madison	Impaired cerebrovascular and autonomic variables, and their association with neuroimaging biological measures of cognitive decline	NR	Case definition: NR Veterans with GWI compared with age and deployment-matched Veterans	Autonomic function testing and MRI to determine brain structure and intracranial blood flow measurements at rest and in response to physiological stress to identify cerebral circulation vascular dysfunction; impaired neurovascular coupling of blood flow with metabolic demand, and/or autonomic dysregulation
OTHER BIOLOGICAL SYSTEMS				
Abdullah, L.[69] Case-control NCT03082638 DoD Roskamp Institute, Inc. Sarasota, Florida Recruiting as of Jun 2019 Expected completion: Dec 2019	Lipids specific to inflammation and metabolic disturbances	100 GWVs: 50 Cases with GWI, 50 Controls without.	Case definition: Kansas Served in GW 1990-1991, with or without GWI	Lipids specific to inflammation, mitochondria and peroxisome function using mass spectrometry technologies.
Golomb, B.[68] Cross-sectional W81XWH-15-1-0626 CDMRP GWIRP UC San Diego	Mitochondrial/ bioenergetic impairments	54 GWVs: 27 with GWI vs 27 healthy controls (matched 1:1 on age, sex, ethnicity)	Case definition: Kansas and CDC NR	(1) Appearance of mitochondria, including mt density/number; mt networks including fission/fusion or elongation, mt cristae density and patterning, plus capillary #/density (re: energy substrates to mitochondria). (2) Function of mitochondria, specifically energetic function and energetic reserve function, stages of respiratory chain function, in intact/skinned muscle fibers and in isolated mitochondria. (3) Mt Membrane and OS Measures: Mt membrane rigidity-fluidity, integrity, barrier function, vulnerability to calcium induced swelling; mt OS and apoptosis markers. Secondary: (A) Bioenergetic: Basal fasting respiratory exchange ratio. (B) Laboratory: Platelet mt assay. Tests of OS, inflammation. (C)

Details: PI *Study design* *Registration/award no.* *Study sponsor* *Setting & status*	**Focus of Investigation**	**Anticipated Participants**	**GWI Case Definition** *Selection criteria/ population*	**Biological Measure/Outcome Measure**
				Exploratory: Quantitative amino acids. Coagulation activation (subset).
Keating, J. A .[70] Prospective cohort study Pilot Project #CX-001574 VA CSRD NLM5T15LM007359 VAMC, Madison, WI Protocol published May 2019	Gut microbiome	52 deployed GWVs: 26 with GWI and 26 without GWI	Case definition: Modified Kansas Aged 43–75 years. Deployed to Gulf as part of Operations Desert Shield and/or Desert Storm during 1st GW (1990–1991)	Microbiome analyses (weekly): Stool total genomic DNA extraction; Saliva total genomic DNA extraction Blood analyses (2 total; 1 at enrollment & 1 at 8 weeks): C-reactive protein; Flow cytometry
Kokkotou, E.[72] Cross-sectional W81XWH-16-1-0528 CDMRP GWIRP BIDMC, Boston, MA	Plasma proteome	GWI participants matched to healthy individuals and 2 different disease controls	Case definition: NR Existing database of a recently completed RCT testing the effectiveness of acupuncture treatment in GWI (PI: Lisa Conboy).	Plasma proteome; SOMA scan
Lin, H. C.[71] Cross-sectional W81XWH-09-2-0073 CDMRP GWIRP BRINM	Small intestinal microbial community	GWI vs controls (not specified)	Case definition: NR Participants in a GWI antibiotic treatment trial	Quantitative PCR of total microbes on small intestinal mucosal biopsy tissue
Lipkin, W. I.[73] Case-control W81XWH-19-1-0398 CDMRP GWIRP Columbia University	Discovery project to survey plasma metabolomes	100 pts: 50 with GWI and 50 GWVs and civilian controls without GWI	Case definition: NR Criteria not specified	Primary metabolites, biogenic amines, complex lipids, and bioactive oxylipins
Meyer, J. N.[67] Cross-sectional (Aim 2 longitudinal) W81XWH-16-1-0663 CDMRP GWIRP	Mitochondrial dysfunction	152 GWVs: 76 with GWI vs 76 without GWI	Case definition: NR Criteria not specified	Mitochondrial parameters measured in PBMCs: mtDNA copy number and damage

Details: PI *Study design* *Registration/award no.* *Study sponsor* *Setting & status*	**Focus of Investigation**	**Anticipated Participants**	**GWI Case Definition** *Selection criteria/ population*	**Biological Measure/Outcome Measure**
Duke University				
Steele, L.[74] Multiphase case-control W81XWH-12-1-0382 CDMRP GWIRP	Serum analytes	2 samples of 45 GWI vs 30 healthy GWVs Then validation in 3rd sample (90 GWI cases, 60 controls)	Case definition: Kansas 1991 GWVs	Panel of ~190 serum analytes: cytokines, chemokines, growth factors, hormones, hematological measures, and neurotrophic factors
Steele, L.[75] Case-control W81XWH-11-1-0812 USAMRDC; Baylor College of Medicine (Moved to Baylor in 2016 – had not yet started data collection)	Multiple systems	130 GWVs: 80 with GWI; 50 healthy controls	Case definition: NR "well-characterized" sample of GWVs	Neuroimaging (MRI, fMRI, DTI), adrenal function tests, and diverse immune, inflammatory, and coagulation measures
Sutton, R.[76] Cross-sectional W81XWH-11-1-0825 CDMRP GWIRP Yale University	XMRV (retrovirus)	~60 pts: 30 GWI vs 30 matched controls	Case definition: NR "well-characterized" cohort of GWI Veterans and closely matched controls	XMRV in serum samples: in DNA and in XMRV-related mRNA, antibody against XMRV

Abbreviations: ACh=Acetylcholine; ATP=Adenosine Triphosphate; BBB=Blood-Brain Barrier; BBRAIN=Boston Biorepository and Integrated Network; BIDMC=Beth Israel Deaconess Medical Center; BRINM=Biomedical Research Institute of New Mexico; CBF=Cerebral Blood Flow; CDC=Centers for Disease Control and Prevention; CDMRP=Congressionally Directed Medical Research Program; CFS=Chronic Fatigue Syndrome; CNS=Central Nervous System; CpG=Cytidine-phosphate Guanosine; CSF=Cerebrospinal Fluid; CSP=Cooperative Studies Program; CSRD=Clinical Sciences Research and Development Service; DKI=Diffusion Kurtosis Imaging; DNA=Deoxyribonucleic Acid; DoD=Department of Defense; DTI=Diffusion Tensor Imaging; fMRI=functional Magnetic Resonance Imaging; GW=Gulf War; GWI=Gulf War Illness; GWIC=Gulf War Illness Consortium; GWIRP=Gulf War Illness Research Program; GWV=Gulf War Veteran; HARDI=High-Angular Resolution Diffusion Imaging; Ig-G=Immunoglobulin G; LPS=Lipopolysaccharide; MN=Minnesota; MPF=Macromolecular Proton Fraction; MRI=Magnetic Resonance Imaging; mRNA=Messenger Ribonucleic Acid; MRS=Magnetic Resonance Spectroscopy; MRSI=Magnetic Resonance Spectroscopic Imaging; MSI=Multisymptom Illness; mt=Mitochondria; mtDNA=Mitochondrial Ribonucleic Acid; MVP=Million Veteran Program; NAA=N-acetyl Aspartate; NIH=National Institutes of Health; NR=Not Reported; OS=Oxidative Stress; PBMC=Peripheral Mononuclear Cell; PCr=Creatine Phosphate; PCR=Polymerase Chain Reaction; PET=Positron Emission Tomography; Pi=Inorganic phosphate; PI=Principal Investigator; PON1=Paraoxanase; RCT=Randomized Controlled Trial; RNA=Ribonucleic Acid; SFVAFRE=South Florida Veterans Affairs Foundation for Research and Education; SNP=Single-Nucleotide Polymorphism; UC=University of California; USAMRDC=United States Army Medical Research and Materiel Command; VA=Veterans Affairs; VAMC=Veterans Affairs Medical Center; VO2=Oxygen Volume; WM=White Matter; XMRV=Xenotropic Murine Leukemia Virus-related Virus

DISCUSSION

This systematic review provides a broad overview of the state of biomarker research in GWI. We found no studies assessing the diagnostic accuracy of biomarkers in GWI (KQ1). We found 56 studies assessing the potential association of a broad variety of biological measures in participants with and without GWI (KQ2), and 24 ongoing or unpublished studies examining biological measures for their association with GWI (KQ3).

Most of the GWI biomarker literature to date could be characterized as "discovery" studies (see Figure 2 for biomarker discovery process diagram, and Figures 4 and 5 for summary graphs of studies by category) and were largely designed to shed light on the potential causes of GWI. Most studies fell into biological system categories that have been theorized as pathophysiologic contributors to GWI. We did not find studies assessing the relationship between biomarkers and severity of GWI illness. Studies in the discovery phase are an early step toward diagnostic test development and help identify biomarkers that are likely to be differentially expressed in participants with GWI. We categorized biological measure studies according to biological system and further identified those studies which found statistically significant associations between the biological measure and GWI (Figures 4 and 5). The vast majority of studies identified biological measures that were significantly more commonly expressed in participants with GWI. We found it challenging to identify a particular biomarker or set of biomarkers as potentially promising because nearly all of these studies were "one-off" studies where the findings have not been replicated in additional studies.

One potential way to prioritize areas of inquiry for further study would be to focus on those areas which have already been tested in populations in which there is some diagnostic uncertainty. These studies included comparator groups of deployed GWVs without GWI (either healthy individuals or those with health conditions other than GWI) and are the studies that were reviewed in the body of this report.

In Key Question 1, we searched for studies evaluating diagnostic tests for GWI. An included study would have been an evaluation of a diagnostic test's ability to provide differential results for cases (per Kansas or CDC criteria) versus non-cases of GWI. It is not surprising that no such studies were found, because the case definition for GWI is still debated. In the absence of a gold standard definition or diagnostic test, the determination of biological measures to distinguish a case from a non-case is challenging.

For Key Question 2, the reviewed body of evidence included a range of studies examining relationships between GWI and biological measures. The range of included biological measures was wide, from indicators of immune function in serum samples, to general brain activity measured with EEG. While the biological measures tended to cluster within a few broadly-defined biological systems (*eg*, immune system, genetic, *etc*), there was significant heterogeneity among the specific biological measures within these groups. The marked variability in the types of measures that were studied, the lack of corroborating data from multiple studies, and methodological limitations of many of the included studies limit clinical application of this evidence. Instead, the current review provides (1) an overview of the range of biological measures that have been examined, and (2) a highlight of methodological limitations of current

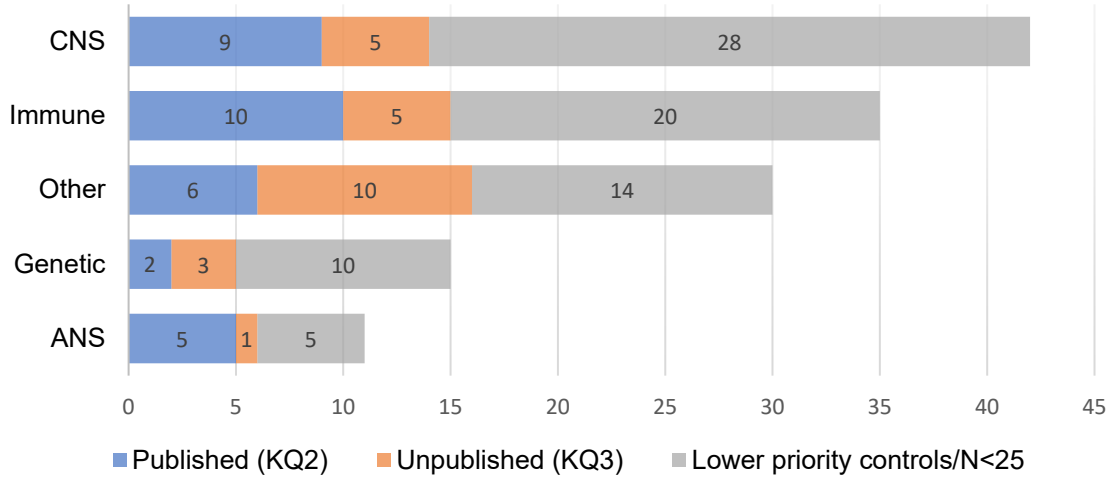
studies that may be adjusted to better position future studies to identify biological measures that distinguish cases of GWI from other conditions.

A majority of the studies we identified were focused on the immune and central nervous systems, with fewer on the autonomic nervous system and genetic markers (Figures 4). The emphasis on immune and central nervous systems is consistent with some hypotheses implicating dysfunction in these systems in GWI.[78] The majority of the immune literature has focused on peripheral blood cytokines (4 studies), with the remainder focused on squalene antibodies (2 studies) mycoplasma fermentans antibodies (1 study), phospholipids (1 study), and human leukocyte antigen alleles (1 study). Most of the studies measuring central nervous system activation have been focused on cognitive/emotional processing, with 7 of 9 of the studies measuring general brain activation in response to cognitive/emotional tasks/stimuli. The studies examining the autonomic nervous system involved both cardiovascular and neurological measurements, with greater emphasis on cardiovascular measures. The studies measuring genetic properties both examined PON1 activity. The remaining studies measured sleep parameters, cardiorespiratory and metabolic responses to maximal exercise, audiovestibular measurements, serum enzymes, neurophysiologic parameters, and presence of viruses.

The lower priority studies (Appendix D) – those that did not include comparator groups of deployed GWI Veterans with or without other health conditions, or that had total sample sizes of <25 – were also heavily focused on the central nervous (28 studies) and immune (20 studies) systems, with several studies in each of the genetic (10 studies), autonomic nervous (5 studies), general nervous system (3 studies), energy metabolism (2 studies), gastrointestinal (2 studies) and skeletal (2 studies) systems, as well as single studies of measures within the respiratory and circulatory systems, and of various measures, biochemical pathways, and bacteria (Figure 5).

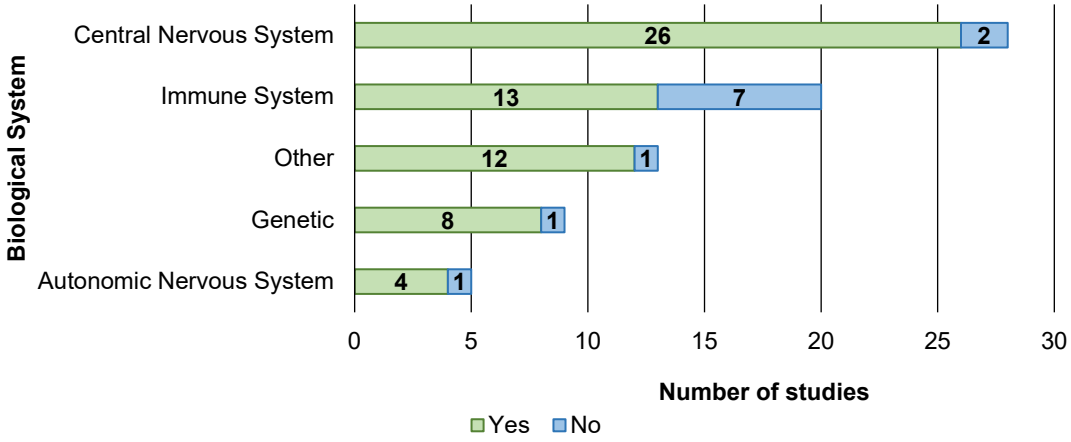
Ongoing and upcoming studies of potential associations between biological measures and GWI (Key Question 3) include an emphasis on studies of measurements in the central nervous system (5 studies[53-57]) and immune system (5 studies[61-65]), which parallels the distribution of biological systems emphasized in the reviewed completed studies. Studies of genetic (3 studies[58-60]) and ANS measures (1 study[66]) are less represented. Additional upcoming studies are diverse in their range of proposed biological measurements, with some emphasis on mitochondrial dysfunction (3 studies[67-69]) and gut microbiome (2 studies[70,71]), which were not areas represented in the completed studies we included.

Figure 4. Number of studies of GWI biological measures by biological system



Abbreviations: ANS=Autonomic Nervous System; CNS=Central Nervous System; KQ=Key Question

Figure 5. Studies of GWI biological measures with lower priority comparator group,* no comparator, or inadequate sample size (N<25) by biological system and promise as a biomarker



* Priority comparator group=. See Appendix D for summary of studies in this figure.
Yes=Indication by statistical significance of association of a biological measure with GWI case status; No=No indication by statistical significance of association of a biological measure with GWI case status

LIMITATIONS

There were significant limitations in the body of evidence in regard to its applicability to the questions posed in this review, including limitations in study design, and heterogeneity in important aspects of the studies. Additionally, we limited inclusion of studies in the review based on the appropriateness of the comparator group for the purpose of identifying potential biological measures, though we tried to mitigate this limitation by providing additional details about the excluded studies (Appendix D). The text below evaluates limitations of the body of evidence.

Study Design

The choice of comparator group was a key limitation of this evidence base. The intent behind Key Question 1 was to identify potential biological measures that identify cases of GWI, and distinguish cases from non-cases, including the distinction between GWI and conditions with overlapping symptomology (*eg*, CFS or depression). To establish a biological metric capable of making this distinction would require biological measures to be compared between cases versus individuals without GWI yet with other health conditions with overlapping symptomology with GWI. The ability of a biologic measure to distinguish GWI when comparing participants with symptoms to healthy participants without symptoms may not translate to its ability to distinguish GWI from another illness in participants presenting with symptoms (which is more typically the context in which a diagnostic test would be used). Therefore, the ideal comparator group to deployed GWVs with GWI would be deployed GWVs without GWI with a condition with overlapping symptoms to GWI. A promising biological measure, then, would have differential results in GWI (loosely defined here) and controls (per the ideal comparator group described above). Studies in the identified body of evidence for the current review almost exclusively employed comparator groups of healthy deployed GWV, with only 1 instance of an ill comparator among deployed GWV.[21] While studies with this composition of comparator groups may contribute to understanding of biological differences between individuals with GWI and healthy individuals, their contribution to establishing biological measures of GWI that are clinically relevant for differential diagnostic purposes is limited.

Still, while the state of the literature may indicate limited studies with the above-described ideal specifications, other types of exploratory studies hold some value in the search for biological measure with promise for development as biomarkers. Thus, the Appendix D table that includes studies that did not include the ideal comparator, as described above, including those studies without a comparator group at all, and studies with n<25, also indicates where there are statistically significant associations of GWI with some biological measure.

In addition to inadequate control groups for the purposes of the current review specifically, there was a general lack of reporting of several features of the study design, which limits the consumer's ability to adequately interpret study results. Many studies did not report power calculations, details regarding how or from where the samples were selected, blinding of investigators to group, recruitment non-response rate, information about the distribution of the data, or whether corrections were made when multiple comparisons were conducted. In the absence of a reported power analysis, it could not be determined whether or not the sample size included in the study was sufficient. Some studies claimed to have insufficient sample sizes but did not provide metrics behind the claim. It was also often not reported how, or from where, the sample was selected, or, in some cases when the selection criteria were reported, all participants were selected from the same battalion, limiting the representativeness of the sample. Further, non-response rate of potential participants during recruitment was also not reported, again limiting the ability to assess the representativeness of the sample.

Blinding of the assessors was also often not reported. When it was reported, it was often the investigators conducting the test who were reported as being blinded to case status and it was very rarely reported that the investigators conducting the data analyses were blinded. Also rarely reported were details of the data analysis including whether or not features of the data, such as

the distribution of the data and outliers in the data, were evaluated, and what, if any, adjustments were made in the face of non-normally distributed data or the presence of outliers, for example. Further, when multiple comparisons were conducted, it was often not reported whether steps were taken to limit false positives due to chance by making corrections to statistical criteria.

Another remarkable feature of study design in the body of evidence was the use of the Haley criteria for case definition.[4] As noted above, the Haley definition categorizes cases of GWI into 6 syndromes based on groupings of related symptoms. In the current body of evidence, the Haley criteria was primarily used in studies of the central nervous system. There were instances of the use of CDC or Kansas criteria for case definition plus stratification of syndrome types using the Haley criteria. Future studies may benefit from the addition of a similar kind of symptom-based stratification, as it is possible that different GWI syndrome types would be associated with different biological measures.

Heterogeneity

In addition to limitations pertaining to study design, the current body of evidence was also limited for the purposes of answering our posed questions due to the heterogeneity of outcome measures, as well as case definition. While there were a few instances of repeated outcome measures – in the case of 2 instances of replication studies, for example – in general the outcome measures were heterogeneous, with only 1 study per outcome measure. Consequently, we were unable to conduct meta-analyses and were unable to otherwise draw conclusions in terms of biological metrics that might hold promise for further establishment as biological measures.

The heterogeneity of case definition in the current body of studies is expected given that the GWI diagnostic criteria is somewhat debated and there are several prevailing case definitions. Namely, a gold standard case definition has not been established and the pursuit of an identification of biological measures is intended to contribute to the development of a gold standard. Consequently, heterogeneity in case definition is expected and we were purposefully inclusive of a wide range of definitions in the current review. One consideration that arose from the current review that might contribute to more effective future studies of potential GWI biological measures is the potential use of stratified clusters of symptoms (*eg*, similar to Haley criteria syndromes). The application of stratified symptom groups may allow for the identification of potential biological measures that are specific to different subsets of GWI, or syndromes. It is possible that GWI itself is too heterogeneous in symptomology to be effectively categorized as 1 cohesive condition, and that stratification may help to organize the condition into more clinically relevant components.

RESEARCH GAPS/FUTURE RESEARCH

While the review of the current body of evidence did not provide insight into promising biological measures, it did provide insight into methodological features of the studies that limit applicability to the questions posed in this review. Future studies examining biological measures of GWI would benefit from using comparator groups composed of deployed GWVs with conditions with overlapping symptomology to GWI and reporting study methods with sufficient detail, including: conducting and reporting a power analysis to determine adequate sample size; providing details regarding how or from where the samples were selected; blinding those collecting and analyzing the data as to group designation, where possible; reporting recruitment

non-response rate; making adjustments for non-normally-distributed data or outliers, and reporting these methods; and potentially stratifying by syndrome sub-category. In addition, for funders of GWI biomarker research, requiring proposed projects to include funding to support the technical expertise, data management, and biostatistical assistance may help to address some of the methodological challenges of these studies. Importantly, the DoD Congressionally Directed Medical Research Program (CDMRP) has implemented a strategic funding mechanism pipeline composed of the following: 1) a discovery stage representing innovative biomarker research that is in the earliest stages of development; 2) a qualification stage representing research already supported by preliminary or published data in the GWI field that is ready for validation through expansion, replication, or comparative studies; 3) a verification stage representing clinical translation (testing in a GW Veteran population) of concepts previously replicated and validated; and 4) a confirmation stage representing large-scale confirmatory and pivotal trials that will transform and revolutionize the clinical management of GWI.[79] This strategy will aid in the translation of the current state of biomarker research into the development of diagnostic and clinical tools.

Additionally, since the outcome measures were diverse, there was an insufficient volume of studies of any given biological measure to draw conclusions about any specific biological measure. Future studies that employ 1) appropriate control groups for GWI biological measure investigation, and 2) methodology that allows for sufficient internal and external validity, would position positive findings of GWI to biological measure associations to be built upon, thus establishing bodies of evidence for promising biological measures. Future research might also consider the complexity between using biological measures to diagnose GWI in the context of the development of other chronic health comorbidities that may impact this aging group of Veterans.

We also identified ongoing or upcoming studies related to GWI biological measures that did not meet our inclusion criteria but are likely to contribute to our knowledge. The Gulf War Illness Consortium has awarded funds to investigate a wide range of metrics: epigenetic DNA changes in Veterans with GWI; effects of exposure to cholinergic compounds; abnormalities in tau, a cytoskeletal protein; and probability scores of case qualification at the individual level using a computerized diagnostic system that incorporates multiple biological measure data.[80] The US Army Medical Research and Development Command is also involved in studies of computer-based diagnostic system for identifying brain-immune interactions in GWI.[81] GWI treatments are being studied simultaneously, including an upcoming trial by the Gulf War Illness Research Program (GWIRP) of an antioxidant, Coenzyme Q10 (CoQ10).[82]

CONCLUSIONS

Gulf War Illness (GWI) is a chronic multisymptom illness comprised of a wide range of systemic symptoms and functional impairments. Hypothesized etiology includes exposure to anticholinergic agents, with suspected dysfunction in cellular energy metabolism with downstream CNS disruption related to inflammation.[78] Thus, much of the literature focused on evaluating biomarkers along this cascade of biological processes. In the current review, we sought to evaluate existing studies validating existing diagnostic tests for GWI, and to determine whether biological measurements with promise for further establishment as biomarkers either in completed or ongoing/upcoming studies have been demonstrated. The establishment of

biological measures for GWI would allow for increased accuracy in diagnosis and potential mechanisms for treatment.

Our review indicates that biological measures within the immune and central nervous systems have more often been investigated for their potential relationship with GWI, consistent with some dominant theories of disease etiology and dysfunction, but the literature also suggests other avenues of inquiry in upcoming studies, such as the gut microbiome. More importantly, our review revealed that existing studies are insufficient for determining promising biomarkers due to the extent of heterogeneity in biological measures across studies, inadequate comparator groups, and several other methodological limitations. Future studies that employ ideal control groups, reproduce findings of existing studies, and otherwise apply rigorous methodological practices and reporting specifically appropriate for investigating potential biomarkers would contribute to the establishment of a base of targeted, highly reliable studies from which lines of investigation could grow.

REFERENCES

1. Anonymous. Gulf War and Health: Volume 10: Update of Health Effects of Serving in the Gulf War, 2016. *Mil Med.* 2017;182(3):1507-1508.
2. Bourdette DN, McCauley LA, Barkhuizen A, et al. Symptom factor analysis, clinical findings, and functional status in a population-based case control study of Gulf War unexplained illness. *J Occup Environ Med.* 2001;43(12):1026-1040.
3. Fukuda K, Nisenbaum R, Stewart G, et al. Chronic multisymptom illness affecting Air Force veterans of the Gulf War. *JAMA.* 1998;280(11):981-988.
4. Haley RW, Kurt TL, Hom J. Is there a Gulf War Syndrome? Searching for syndromes by factor analysis of symptoms. *JAMA.* 1997;277(3):215-222.
5. Kang HK, Li B, Mahan CM, Eisen SA, Engel CC. Health of US veterans of 1991 Gulf War: a follow-up survey in 10 years. *J Occup Environ Med.* 2009;51(4):401-410.
6. Spencer PS, McCauley LA, Joos SK, et al. U.S. Gulf War Veterans: service periods in theater, differential exposures, and persistent unexplained illness. Portland Environmental Hazards Research Centre. *Toxicol Lett.* 1998;102-103:515-521.
7. Steele L. Prevalence and patterns of Gulf War illness in Kansas veterans: association of symptoms with characteristics of person, place, and time of military service. *Am J Epidemiol.* 2000;152(10):992-1002.
8. Institute of Medicine. Chronic Multisymptom Illness in Gulf War Veterans: Case Definitions Reexamined. *National Academies Press (US)* 2014;03(12).
9. US Department of Veterans Affairs. Gulf War Veterans' Medically Unexplained Illnesses. Military Exposures Web site. https://www.publichealth.va.gov/exposures/gulfwar/medically-unexplained-illness.asp. Accessed February 5, 2020.
10. Abou-Donia MB, Conboy LA, Kokkotou E, et al. Screening for novel central nervous system biomarkers in veterans with Gulf War Illness. *Neurotoxicol Teratol.* 2017;61:36-46.
11. Steele L, Sastre A, Gerkovich MM, Cook MR. Complex factors in the etiology of Gulf War illness: wartime exposures and risk factors in veteran subgroups. *Environ Health Perspect.* 2012;120(1):112-118.
12. Koslik HJ, Hamilton G, Golomb BA. Mitochondrial dysfunction in Gulf War illness revealed by 31Phosphorus Magnetic Resonance Spectroscopy: a case-control study. *PLoS ONE [Electronic Resource].* 2014;9(3):e92887.
13. Golomb BA. Acetylcholinesterase inhibitors and Gulf War illnesses. *Proc Natl Acad Sci U S A.* 2008;105(11):4295-4300.
14. Mawson AR, Croft AM. Gulf War Illness: Unifying Hypothesis for a Continuing Health Problem. *International Journal of Environmental Research & Public Health [Electronic Resource].* 2019;16(1):03.
15. Kodali M, Hattiangady B, Shetty GA, Bates A, Shuai B, Shetty AK. Curcumin treatment leads to better cognitive and mood function in a model of Gulf War Illness with enhanced neurogenesis, and alleviation of inflammation and mitochondrial dysfunction in the hippocampus. *Brain Behav Immun.* 2018;69:499-514.
16. Fall DJ, Stessman H, Patel SS, et al. Utilization of Translational Bioinformatics to Identify Novel Biomarkers of Bortezomib Resistance in Multiple Myeloma. *J Cancer.* 2014;5(9):720-727.

17. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *J Clin Epidemiol.* 2016;75:40-46.
18. Hackshaw A. Small studies: strengths and limitations. *Eur Respir J.* 2008;32(5):1141-1143.
19. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155(8):529-536.
20. Wirsching J, Graßmann S, Eichelmann F, et al. Development and reliability assessment of a new quality appraisal tool for cross-sectional studies using biomarker data (BIOCROSS). *BMC Med Res Methodol.* 2018;18(1):122.
21. Haines DD, Ottenweller JE, Dickens BF, Mahmoud FF, Levine PH. Activity of Paraoxonase/Arylesterase and Butyrylcholinesterase in Peripheral Blood of Gulf War Era Veterans With Neurologic Symptom Complexes or Post-Traumatic Stress Disorder. *J Occup Environ Med.* 2017;59(10):1000-1006.
22. Lo SC, Levin L, Ribas J, et al. Lack of serological evidence for Mycoplasma fermentans infection in army Gulf War veterans: a large scale case-control study. *Epidemiol Infect.* 2000;125(3):609-616.
23. Asa PB, Cao Y, Garry RF. Antibodies to squalene in Gulf War syndrome. *Exp Mol Pathol.* 2000;68(1):55-64.
24. Emmerich T, Zakirova Z, Klimas N, et al. Phospholipid profiling of plasma from GW veterans and rodent models to identify potential biomarkers of Gulf War Illness. *PLoS ONE [Electronic Resource].* 2017;12(4):e0176634.
25. Georgopoulos AP, James LM, Mahan MY, Joseph J, Georgopoulos A, Engdahl BE. Reduced Human Leukocyte Antigen (HLA) Protection in Gulf War Illness (GWI). *EBioMedicine.* 2016;3:79-85.
26. Johnson GJ, Leis LA, Slater BC, Bach RR. Elevated platelet count, C-reactive protein and thromboxane analog-induced platelet aggregation in patients with Gulf War veterans' illnesses: evidence of a chronic inflammatory state? *Blood Coagul Fibrinolysis.* 2013;24(7):736-741.
27. James LM, Engdahl BE, Leuthold AC, Georgopoulos AP. Brain Correlates of Human Leukocyte Antigen (HLA) Protection in Gulf War Illness (GWI). *EBioMedicine.* 2016;13:72-79.
28. Butterick TA, Trembley JH, Hocum Stone LL, Muller CJ, Rudquist RR, Bach RR. Gulf War Illness-associated increases in blood levels of interleukin 6 and C-reactive protein: biomarker evidence of inflammation. *BMC Res Notes.* 2019;12(1):816.
29. Johnson GJ, Slater BC, Leis LA, Rector TS, Bach RR. Blood Biomarkers of Chronic Inflammation in Gulf War Illness. *PLoS ONE [Electronic Resource].* 2016;11(6):e0157855.
30. Phillips CJ, Matyas GR, Hansen CJ, Alving CR, Smith TC, Ryan MA. Antibodies to squalene in US Navy Persian Gulf War veterans with chronic multisymptom illness. *Vaccine.* 2009;27(29):3921-3926.
31. Skowera A, Hotopf M, Sawicka E, et al. Cellular immune activation in Gulf War veterans. *J Clin Immunol.* 2004;24(1):66-73.
32. Calley CS, Kraut MA, Spence JS, Briggs RW, Haley RW, Hart J, Jr. The neuroanatomic correlates of semantic memory deficits in patients with Gulf War illnesses: a pilot study. *Brain Imaging Behav.* 2010;4(3-4):248-255.

33. Cooper CM, Briggs RW, Farris EA, Bartlett J, Haley RW, Odegard TN. Memory and functional brain differences in a national sample of U.S. veterans with Gulf War Illness. *Psychiatry Research: Neuroimaging.* 2016;250:33-41.
34. Gopinath K, Gandhi P, Goyal A, et al. FMRI reveals abnormal central processing of sensory and pain stimuli in ill Gulf War veterans. *Neurotoxicology.* 2012;33(3):261-271.
35. Liu P, Aslan S, Li X, et al. Perfusion deficit to cholinergic challenge in veterans with Gulf War Illness. *Neurotoxicology.* 2011;32(2):242-246.
36. Odegard TN, Cooper CM, Farris EA, Arduengo J, Bartlett J, Haley R. Memory impairment exhibited by veterans with Gulf War Illness. *Neurocase.* 2013;19(4):316-327.
37. Tillman GD, Green TA, Ferree TC, et al. Impaired response inhibition in ill Gulf War veterans. *J Neurol Sci.* 2010;297(1-2):1-5.
38. Tillman GD, Calley CS, Green TA, et al. Event-related potential patterns associated with hyperarousal in Gulf War illness syndrome groups. *Neurotoxicology.* 2012;33(5):1096-1105.
39. Tillman GD, Calley CS, Green TA, et al. Visual event-related potentials as markers of hyperarousal in Gulf War illness: evidence against a stress-related etiology. *Psychiatry Res.* 2013;211(3):257-267.
40. Tillman GD, Spence JS, Briggs RW, Haley RW, Hart J, Jr., Kraut MA. Gulf War illness associated with abnormal auditory P1 event-related potential: Evidence of impaired cholinergic processing replicated in a national sample. *Psychiatry Research: Neuroimaging.* 2019;283:7-15.
41. Weiner MW, Meyerhoff DJ, Neylan TC, et al. The relationship between Gulf War illness, brain N-acetylaspartate, and post-traumatic stress disorder. *Mil Med.* 2011;176(8):896-902.
42. Blanchard M, Molina-Vicenty HD, Stein PK, et al. Medical Correlates of Chronic Multisymptom Illness in Gulf War Veterans. *Am J Med.* 2019;132(4):510-518.
43. Davis SD, Kator SF, Wonnett JA, Pappas BL, Sall JL. Neurally mediated hypotension in fatigued Gulf War veterans: a preliminary report. *Am J Med Sci.* 2000;319(2):89-95.
44. Haley RW, Charuvastra E, Shell WE, et al. Cholinergic autonomic dysfunction in veterans with Gulf War illness: confirmation in a population-based sample. *JAMA Neurology.* 2013;70(2):191-200.
45. Li M, Xu C, Yao W, et al. Self-reported post-exertional fatigue in Gulf War veterans: roles of autonomic testing. *Front Neurosci.* 2014;7:269.
46. Nagelkirk PR, Cook DB, Peckerman A, et al. Aerobic capacity of Gulf War veterans with chronic fatigue syndrome. *Mil Med.* 2003;168(9):750-755.
47. Hotopf M, Mackness MI, Nikolaou V, et al. Paraoxonase in Persian Gulf War veterans. *J Occup Environ Med.* 2003;45(7):668-675.
48. Amin MM, Belisova Z, Hossain S, Gold MS, Broderick JE, Gold AR. Inspiratory airflow dynamics during sleep in veterans with Gulf War illness: a controlled study. *Sleep & Breathing.* 2011;15(3):333-339.
49. Roland PS, Haley RW, Yellin W, Owens K, Shoup AG. Vestibular dysfunction in Gulf War syndrome. *Otolaryngology - Head & Neck Surgery.* 2000;122(3):319-329.
50. Sharief MK, Priddin J, Delamont RS, et al. Neurophysiologic analysis of neuromuscular symptoms in UK Gulf War veterans: a controlled study. *Neurology.* 2002;59(10):1518-1525.

51. Wallace HL, 2nd, Natelson B, Gause W, Hay J. Human herpesviruses in chronic fatigue syndrome. *Clin Diagn Lab Immunol.* 1999;6(2):216-223.
52. Zhou Q, Verne ML, Zhang B, Verne GN. Evidence for Somatic Hypersensitivity in Veterans With Gulf War Illness and Gastrointestinal Symptoms. *Clin J Pain.* 2018;34(10):944-949.
53. Peskind ER, Petrie EC. Multimodal Biological Assessment of Gulf War Illness. https://projectreporter.nih.gov/project_info_description.cfm?aid=9278098&icde=0. Published 2014. Accessed 8 May 2020.
54. Defining and Characterizing GWI Pathobiology Using Longitudinal Brain Imaging Biomarkers of White Matter Integrity and Hemodynamic Response. https://cdmrp.army.mil/search.aspx?LOG_NO=GW180099. Accessed 6 May 2020.
55. Little DM. *Assessment of MRI-Based Marker of Dopaminergic Integrity as a Biological Indicator of Gulf War Illness.* Temple, TX: Scott and White Memorial Hospital; October 2016.
56. Clarifying the Role Played by Microglia and Astrocyte Activation in Veterans with Gulf War Illness Using Positron Emission Tomography (PET). https://cdmrp.army.mil/search.aspx?LOG_NO=GW180103. Accessed 6 May 2020.
57. Assessing Neuroinflammation in Gulf War Illness with Whole-Brain Magnetic Resonance Spectroscopy. https://cdmrp.army.mil/search.aspx?LOG_NO=GW180071. Accessed 6 May 2020.
58. An Objective Blood Test from Stimulated Gene Expression for Classification and Outcome Assessment in Clinical Trials of Gulf War Illness. https://cdmrp.army.mil/search.aspx?LOG_NO=GW140158. Accessed 8 May 2020.
59. US Department of Veterans Affairs. CSP #2006: Genomics of Gulf War Illness in Veterans. https://www.vacsp.research.va.gov/CSPEC/Studies/INVESTD-R/CSP-2006-Genomics-Gulf-War-Illness.asp. Published 2018. Accessed 26 June 2020.
60. Identification of Epigenetic Signatures as Biomarkers of Gulf War Illness. https://cdmrp.army.mil/search.aspx?LOG_NO=GW160096. Accessed 6 May 2020.
61. Identifying Objective Diagnostic Markers of Gulf War Illness: Salivary and Plasma Autoantibodies Against Neural Proteins Validated with Brain Imaging. https://cdmrp.army.mil/search.aspx?LOG_NO=GW180121. Accessed 8 May 2020.
62. James LM. *Human Leukocyte Antigen in Gulf War Veterans: Association with Symptoms and Inflammatory Markers.* October 2018.
63. Research SFVAFf, Education. Glutathione vs Curcumin Clinical Trial. In:2016.
64. Next-Generation Biomarkers of Gulf War Illness https://cdmrp.army.mil/search.aspx?LOG_NO=GW160161. Accessed 6 May 2020.
65. Biomarkers and Brain Mechanisms of Gulf War Illness. https://cdmrp.army.mil/search.aspx?LOG_NO=GW140114. Accessed 8 May 2020.
66. Autonomic Dysfunction, Brain Blood Flow, and Cognitive Decline in Veterans with Gulf War Illness. https://cdmrp.army.mil/search.aspx?LOG_NO=GW180103. Accessed 6 May 2020.
67. Mitochondrial Dysfunction and Gulf War Illness. https://cdmrp.army.mil/search.aspx?LOG_NO=GW150184. Accessed 8 May 2020.
68. Muscle Mitochondrial Assessments in Gulf War Illness. https://cdmrp.army.mil/search.aspx?LOG_NO=GW140045. Accessed 8 May 2020.

69. NCT03082638. Identification of Gulf War Illness (GWI) Biomarkers. In. Roskamp Institute I, United States Department of D, trans2017.
70. Keating JA, Shaughnessy C, Baubie K, et al. Characterising the gut microbiome in veterans with Gulf War Illness: a protocol for a longitudinal, prospective cohort study. *BMJ Open.* 2019;9(8):e031114.
71. Small Intestinal Microbial Community in Gulf War Illness. https://cdmrp.army.mil/search.aspx?LOG_NO=GW080151. Accessed 8 May 2020.
72. GWI: Molecular Analysis of Disease Endophenotypes and Response to Acupuncture Treatment https://cdmrp.army.mil/search.aspx?LOG_NO=GW150202. Accessed 27 May 2020.
73. Metabolomics in Gulf War Illness: A Systems Biology Approach to Dissecting Mechanisms of Disease. https://cdmrp.army.mil/search.aspx?LOG_NO=GW180150. Accessed 8 May 2020.
74. Biomarker Discovery in Gulf War Veterans: Development of a War Illness Diagnostic Panel. https://cdmrp.army.mil/search.aspx?LOG_NO=GW110078. Accessed 8 May 2020.
75. Steele L. *Assessment of Diverse Biological Indicators in Gulf War Illness: Are They Replicable? Are They Related?* Waco, TX: Baylor University; October 2012.
76. XMRV and GWI: Is There an Association? https://cdmrp.army.mil/search.aspx?LOG_NO=GW100007. Accessed 8 May 2020.
77. US Department of Veterans Affairs. VA Cooperative Studies Program 585: Information for Veterans. https://www.vacsp.research.va.gov/CSP_585/CSP_585_Information_for_Veterans.asp. Accessed 28 August 2020.
78. Chester JE, Rowneki M, Van Doren W, Helmer DA. Progression of intervention-focused research for Gulf War illness. *Mil Med Res.* 2019;6(1):31.
79. Gulf War Illness Research Program. *The Gulf War Illness Landscape.* March 2020.
80. Chedekel L. Identifying Biomarkers of Gulf War Illness. Boston University School of Public Health,. https://www.bu.edu/sph/2017/03/29/identifying-biomarkers-of-gulf-war-illness/. Published 2017. Accessed 29 June 2020.
81. Koo B-B. *Computer aided decoding of brain-immune interactions in Gulf War Illness (GWI): a joint embedding on brain connectomic and immunogenetic markers.* 2019.
82. Measurement of Biomarkers in Samples Collected in a Coenzyme Q10 Treatment Trial in Gulf War Illness and Control Subjects. https://cdmrp.army.mil/search.aspx?LOG_NO=GW160101. Accessed 29 June 2020.
83. Abou-Donia MB, Conboy LA, Kokkotou E, et al. Screening for novel central nervous system biomarkers in veterans with Gulf War Illness. *Neurotoxicol Teratol.* 2017;61:36-46.
84. Brimacombe M, Zhang Q, Lange G, Natelson B. Immunological variables mediate cognitive dysfunction in gulf war veterans but not civilians with chronic fatigue syndrome. *Neuroimmunomodulation.* 2002;10(2):93-100.
85. Broderick G, Kreitz A, Fuite J, Fletcher MA, Vernon SD, Klimas N. A pilot study of immune network remodeling under challenge in Gulf War Illness. *Brain Behav Immun.* 2011;25(2):302-313.

86. Broderick G, Fletcher MA, Gallagher M, Barnes Z, Vernon SD, Klimas NG. Exploring the Diagnostic Potential of Immune Biomarker Co-expression in Gulf War Illness. *Methods Mol Biol.* 2018;1781:101-120.
87. Diaz-Torne C, Schumacher HR, Yu X, et al. Absence of histologic evidence of synovitis in patients with Gulf War veterans' illness with joint pain. *Arthritis Rheum.* 2007;57(7):1316-1323.
88. Everson MP, Shi K, Aldridge P, Bartolucci AA, Blackburn WD. Immunological responses are not abnormal in symptomatic Gulf War veterans. *Ann N Y Acad Sci.* 2002;966:327-342.
89. Golomb BA, Koslik HJ, Christians U, et al. Depressed prostaglandins and leukotrienes in veterans with Gulf War illness. *Journal of Environmental Science & Health - Part B: Pesticides, Food Contaminants, & Agricultural Wastes.* 2019:1-17.
90. Halpin P, Williams MV, Klimas NG, Fletcher MA, Barnes Z, Ariza ME. Myalgic encephalomyelitis/chronic fatigue syndrome and gulf war illness patients exhibit increased humoral responses to the herpesviruses-encoded dUTPase: Implications in disease pathophysiology. *J Med Virol.* 2017;89(9):1636-1645.
91. Hannan KL, Berg DE, Baumzweiger W, et al. Activation of the coagulation system in Gulf War Illness: a potential pathophysiologic link with chronic fatigue syndrome. A laboratory approach to diagnosis. *Blood Coagul Fibrinolysis.* 2000;11(7):673-678.
92. Khaiboullina SF, DeMeirleir KL, Rawat S, et al. Cytokine expression provides clues to the pathophysiology of Gulf War illness and myalgic encephalomyelitis. *Cytokine.* 2015;72(1):1-8.
93. Klaustermeyer WB, Kraske GK, Lee KG, Klaustermeyer WB, Kurohara ML. Allergic and immunologic profile of symptomatic Persian Gulf War veterans. *Ann Allergy Asthma Immunol.* 1998;80(3):269-273.
94. O'Bryan TA, Romano PJ, Zangwill BC. Human leukocyte antigens in Gulf War veterans with chronic unexplained multiple symptoms. *Mil Med.* 2003;168(12):1015-1018.
95. Parkitny L, Middleton S, Baker K, Younger J. Evidence for abnormal cytokine expression in Gulf War Illness: A preliminary analysis of daily immune monitoring data. *BMC Immunol.* 2015;16:57.
96. Skowera A, Stewart E, Davis ET, et al. Antinuclear autoantibodies (ANA) in Gulf War-related illness and chronic fatigue syndrome (CFS) patients. *Clin Exp Immunol.* 2002;129(2):354-358.
97. Smylie AL, Broderick G, Fernandes H, et al. A comparison of sex-specific immune signatures in Gulf War illness and chronic fatigue syndrome. *BMC Immunol.* 2013;14:29.
98. Tsilibary EC, Souto EP, James LM, Engdahl BE, Georgopoulos AP. Human Immunoglobulin G (IgG) Neutralizes Adverse Effects of Gulf War Illness (GWI) Serum in Neural Cultures: Paving the Way to Immunotherapy for GWI. *Journal of Neurology & neuromedicine.* 2018;3(5):23-28.
99. Vojdani A, Thrasher JD. Cellular and humoral immune abnormalities in Gulf War veterans. *Environ Health Perspect.* 2004;112(8):840-846.
100. Whistler T, Fletcher MA, Lonergan W, et al. Impaired immune function in Gulf War Illness. *BMC Medical Genomics [Electronic Resource].* 2009;2:12.
101. Zhang Q, Zhou XD, Denny T, et al. Changes in immune parameters seen in Gulf War veterans but not in civilians with chronic fatigue syndrome. *Clin Diagn Lab Immunol.* 1999;6(1):6-13.

102. Alshelh Z, Albrecht DS, Bergan C, et al. In-vivo imaging of neuroinflammation in veterans with Gulf War illness. *Brain Behav Immun.* 2020;04:04.
103. Baraniuk JN, Casado B, Maibach H, Clauw DJ, Pannell LK, Hess SS. A Chronic Fatigue Syndrome - related proteome in human cerebrospinal fluid. *BMC Neurol.* 2005;5:22.
104. Chao LL, Mohlenhoff BS, Weiner MW, Neylan TC. Associations between subjective sleep quality and brain volume in Gulf War veterans. *Sleep.* 2014;37(3):445-452.
105. Chao L. Do Gulf War veterans with high levels of deployment-related exposures display symptoms suggestive of Parkinson's disease? *Int J Occup Med Environ Health.* 2019;32(4):503-526.
106. Christova P, James LM, Engdahl BE, Lewis SM, Carpenter AF, Georgopoulos AP. Subcortical brain atrophy in Gulf War Illness. *Exp Brain Res.* 2017;235(9):2777-2786.
107. Clarke T, Jamieson JD, Malone P, et al. Connectivity differences between Gulf War Illness (GWI) phenotypes during a test of attention. *PLoS ONE [Electronic Resource].* 2019;14(12):e0226481.
108. Concato J, Aslan M, Palmisano MM, et al. Acetylcholinesterase activity in veterans of the first Gulf War. *J Investig Med.* 2007;55(7):360-367.
109. Engdahl BE, James LM, Miller RD, et al. A Magnetoencephalographic (MEG) Study of Gulf War Illness (GWI). *EBioMedicine.* 2016;12:127-132.
110. Georgopoulos AP, James LM, Carpenter AF, Engdahl BE, Leuthold AC, Lewis SM. Gulf War illness (GWI) as a neuroimmune disease. *Exp Brain Res.* 2017;235(10):3217-3225.
111. Gopinath KS, Sakoglu U, Crosson BA, Haley RW. Exploring brain mechanisms underlying Gulf War Illness with group ICA based analysis of fMRI resting state networks. *Neurosci Lett.* 2019;701:136-141.
112. Haley RW, Hom J, Roland PS, et al. Evaluation of neurologic function in Gulf War veterans. A blinded case-control study. *JAMA.* 1997;277(3):223-230.
113. Haley RW, Marshall WW, McDonald GG, Daugherty MA, Petty F, Fleckenstein JL. Brain abnormalities in Gulf War syndrome: evaluation with 1H MR spectroscopy. *Radiology.* 2000;215(3):807-817.
114. Haley RW, Spence JS, Carmack PS, et al. Abnormal brain response to cholinergic challenge in chronic encephalopathy from the 1991 Gulf War. *Psychiatry Res.* 2009;171(3):207-220.
115. Hubbard NA, Hutchison JL, Motes MA, et al. Central Executive Dysfunction and Deferred Prefrontal Processing in Veterans with Gulf War Illness. *Clinical Psychological Science.* 2014;2(3):319-327.
116. Jamal GA, Hansen S, Apartopoulos F, Peden A. The "Gulf War syndrome". Is there evidence of dysfunction in the nervous system? *J Neurol Neurosurg Psychiatry.* 1996;60(4):449-451.
117. James LM, Christova P, Engdahl BE, Lewis SM, Carpenter AF, Georgopoulos AP. Human Leukocyte Antigen (HLA) and Gulf War Illness (GWI): HLA-DRB1 13:02 Spares Subcortical Atrophy in Gulf War Veterans. *EBioMedicine.* 2017;26:126-131.
118. Li X, Spence JS, Buhner DM, et al. Hippocampal dysfunction in Gulf War veterans: investigation with ASL perfusion MR imaging and physostigmine challenge. *Radiology.* 2011;261(1):218-225.
119. Menon PM, Nasrallah HA, Reeves RR, Ali JA. Hippocampal dysfunction in Gulf War Syndrome. A proton MR spectroscopy study. *Brain Res.* 2004;1009(1-2):189-194.

120. Moffett K, Crosson B, Spence JS, et al. Word-finding impairment in veterans of the 1991 Persian Gulf War. *Brain Cogn.* 2015;98:65-73.
121. Rayhan RU, Raksit MP, Timbol CR, Adewuyi O, Vanmeter JW, Baraniuk JN. Prefrontal lactate predicts exercise-induced cognitive dysfunction in Gulf War Illness. *American Journal Of Translational Research.* 2013;5(2):212-223.
122. Rayhan RU, Stevens BW, Raksit MP, et al. Exercise challenge in Gulf War Illness reveals two subgroups with altered brain structure and function. *PLoS ONE [Electronic Resource].* 2013;8(6):e63903.
123. Rayhan RU, Stevens BW, Timbol CR, et al. Increased brain white matter axial diffusivity associated with fatigue, pain and hyperalgesia in Gulf War illness. *PLoS ONE [Electronic Resource].* 2013;8(3):e58493.
124. Rayhan RU, Washington SD, Garner R, et al. Exercise challenge alters Default Mode Network dynamics in Gulf War Illness. *BMC Neurosci.* 2019;20(1):7.
125. Tillman GD, Calley CS, Buhl VI, et al. Electrophysiological correlates of semantic memory retrieval in Gulf War Syndrome 2 patients. *J Neurol Sci.* 2017;373:66-72.
126. Turner MP, Hubbard NA, Himes LM, et al. Cognitive Slowing in Gulf War Illness Predicts Executive Network Hyperconnectivity: Study in a Population-Representative Sample. *NeuroImage Clinical.* 2016;12:535-541.
127. Washington SD, Rayhan RU, Garner R, et al. Exercise alters cerebellar and cortical activity related to working memory in phenotypes of Gulf War Illness. *Brain Communications.* 2020;2(1):fcz039.
128. Wylie GR, Genova H, Dobryakova E, et al. Fatigue in Gulf War Illness is associated with tonically high activation in the executive control network. *NeuroImage Clinical.* 2019;21:101641.
129. Falvo MJ, Chen Y, Klein JC, Ndirangu D, Condon MR. Abnormal rheological properties of red blood cells as a potential marker of Gulf War Illness: A preliminary study. *Clin Hemorheol Microcirc.* 2018;68(4):361-370.
130. Fiedler N, Giardino N, Natelson B, et al. Responses to controlled diesel vapor exposure among chemically sensitive Gulf War veterans. *Psychosom Med.* 2004;66(4):588-598.
131. Haley RW, Vongpatanasin W, Wolfe GI, et al. Blunted circadian variation in autonomic regulation of sinus node function in veterans with Gulf War syndrome. *Am J Med.* 2004;117(7):469-478.
132. Peckerman A, LaManca JJ, Smith SL, et al. Cardiovascular stress responses and their relation to symptoms in Gulf War veterans with fatiguing illness. *Psychosom Med.* 2000;62(4):509-516.
133. Stein PK, Domitrovich PP, Ambrose K, et al. Sex effects on heart rate variability in fibromyalgia and Gulf War illness. *Arthritis Rheum.* 2004;51(5):700-708.
134. Baraniuk JN, Shivapurkar N. Exercise-induced changes in cerebrospinal fluid miRNAs in Gulf War Illness, Chronic Fatigue Syndrome and sedentary control subjects. *Sci Rep.* 2017;7(1):15338.
135. Craddock TJ, Harvey JM, Nathanson L, et al. Using gene expression signatures to identify novel treatment strategies in gulf war illness. *BMC Medical Genomics [Electronic Resource].* 2015;8:36.
136. Liu G, Ye CJ, Chowdhury SK, et al. Detecting Chromosome Condensation Defects in Gulf War Illness Patients. *Current Genomics.* 2018;19(3):200-206.

137. Mackness B, Durrington PN, Mackness MI. Low paraoxonase in Persian Gulf War Veterans self-reporting Gulf War Syndrome. *Biochem Biophys Res Commun.* 2000;276(2):729-733.
138. NCT00810225. Study of Gulf War Illness (GWI) by Comparing GWI and Healthy Veterans. In. Georgetown U, Research USAM, Materiel C, trans2008.
139. Steele L, Lockridge O, Gerkovich MM, Cook MR, Sastre A. Butyrylcholinesterase genotype and enzyme activity in relation to Gulf War illness: preliminary evidence of gene-exposure interaction from a case-control study of 1991 Gulf War veterans. *Environ Health.* 2015;14:4.
140. Trivedi MS, Abreu MM, Sarria L, et al. Alterations in DNA Methylation Status Associated with Gulf War Illness. *DNA Cell Biol.* 2019;38(6):561-571.
141. Urnovitz HB, Tuite JJ, Higashida JM, Murphy WH. RNAs in the sera of Persian Gulf War veterans have segments homologous to chromosome 22q11.2. *Clin Diagn Lab Immunol.* 1999;6(3):330-335.
142. Vladutiu GD, Natelson BH. Association of medically unexplained fatigue with ACE insertion/deletion polymorphism in Gulf War veterans. *Muscle Nerve.* 2004;30(1):38-43.
143. Nicolson GL, Nasralla M, Nicolson NL, Haier J. High prevalence of Mycoplasma infections in symptomatic (chronic fatigue syndrome) family members of Mycoplasma-positive Gulf War Illness patients. *Journal of Chronic Fatigue Syndrome.* 2003;11(2):21-36.
144. Naviaux RK, Naviaux JC, Li K, et al. Metabolic features of Gulf War illness. *PLoS ONE [Electronic Resource].* 2019;14(7):e0219531.
145. Falvo MJ, Lindheimer JB, Serrador JM. Dynamic cerebral autoregulation is impaired in Veterans with Gulf War Illness: A case-control study. *PLoS ONE [Electronic Resource].* 2018;13(10):e0205393.
146. Chen Y, Meyer JN, Hill HZ, et al. Role of mitochondrial DNA damage and dysfunction in veterans with Gulf War Illness. *PLoS ONE [Electronic Resource].* 2017;12(9):e0184832.
147. Henry C. Lin M, System NMVH. Bacterial Overgrowth Associated With Chronic Multisymptom Illness Complex. In:2009.
148. Fletcher MA, Rosenthal M, Antoni M, et al. Plasma neuropeptide Y: a biomarker for symptom severity in chronic fatigue syndrome. *Behavioral & Brain Functions [Electronic Resource]: BBF.* 2010;6:76.
149. Khan F, Kennedy G, Spence VA, Newton DJ, Belch JJ. Peripheral cholinergic function in humans with chronic fatigue syndrome, Gulf War syndrome and with illness following organophosphate exposure. *Clin Sci.* 2004;106(2):183-189.
150. Lindheimer JB, Cook DB, Klein-Adams JC, et al. Veterans with Gulf War Illness exhibit distinct respiratory patterns during maximal cardiopulmonary exercise. *PLoS ONE [Electronic Resource].* 2019;14(11):e0224833.
151. Compston JE, Vedi S, Stephen AB, et al. Reduced bone formation in UK Gulf War veterans: a bone histomorphometric study. *J Clin Pathol.* 2002;55(12):897-899.
152. Pessler F, Chen LX, Dai L, et al. A histomorphometric analysis of synovial biopsies from individuals with Gulf War Veterans' Illness and joint pain compared to normal and osteoarthritis synovium. *Clin Rheumatol.* 2008;27(9):1127-1134.

153. NCT00810329. Proteomics of Cerebrospinal Fluid in Chronic Fatigue Syndrome. In. Georgetown University N, National Institute of Environmental Health S, University of South A, trans2008.
154. Broderick G, Ben-Hamo R, Vashishtha S, et al. Altered immune pathway activity under exercise challenge in Gulf War Illness: an exploratory analysis. *Brain Behav Immun.* 2013;28:159-169.
155. Bunegin L, Mitzel HC, Miller CS, Gelineau JF, Tolstykh GP. Cognitive performance and cerebrohemodynamics associated with the Persian Gulf Syndrome. *Toxicol Ind Health.* 2001;17(4):128-137.
156. Haley RW, Fleckenstein JL, Marshall WW, McDonald GG, Kramer GL, Petty F. Effect of basal ganglia injury on central dopamine activity in Gulf War syndrome: correlation of proton magnetic resonance spectroscopy and plasma homovanillic acid levels. *Arch Neurol.* 2000;57(9):1280-1285.
157. Latimer JJ, Alhamed A, Sveiven S, et al. Preliminary Evidence for a Hormetic Effect on DNA Nucleotide Excision Repair in Veterans with Gulf War Illness. *Mil Med.* 2020;185(1-2):e47-e52.
158. Janulewicz PA, Seth RK, Carlson JM, et al. The Gut-Microbiome in Gulf War Veterans: A Preliminary Report. *International Journal of Environmental Research & Public Health [Electronic Resource].* 2019;16(19):04.

APPENDIX A. SEARCH STRATEGIES

Ovid MEDLINE ALL 1946 to February 20, 2020
Date searched: February 21, 2020
1 Persian Gulf Syndrome/ or Gulf War/ (1100)
2 ("Desert Saber" or "Desert Sabre" or "Desert Shield" or "Desert Storm" or "Gulf War" or "Gulf Conflict" or "Gulf Crisis" or "Persian Gulf Syndrome" or "Kuwait War" or "Operation GRANBY" or "Op GRANBY").ti,ab,kf. (2230)
3 (GWI or GWIs or GWVI or GWVIs or "Iowa Persian Gulf Study" or "War Related Illness and Injury Study Center*").ti,ab,kf. (222)
4 ((Kuwait or Iraq or "Persian Gulf" or "Southwest Asia" or "SW Asia") and ("air force" or "armed forces" or army or marines or "military personnel" or "national guard*" or naval or navy or "service members" or servicemembers or soldier* or Veteran*)).ti,ab,kf. (2938)
5 or/1-4 (4955)
6 exp Biological measures/ (727323)
7 (antigen or antigens or autoantibod* or auto-antibod* or antibody or antibodies or bioassay* or bio-assay* or biological measure* or bio-marker* or biopsy or biopsies or blood or coexpress* or co-express* or conduction or "CT scan*" or cytokine or cytokines or diagnos* or dysfunction* or electromyograph* or endoscop* or fluid or fluids or fMRI or genet* or "gene expression" or imaging or inflammat* or marker or markers or MRI or "Magnetic resonance imaging" or mechanism* or neurodegenerat* or neuro-degenerat* or neuroendocrine or neuro-endocrine or neuroimag* or neuro-imag* or neuroinflammat* or neuro-inflammat* or protein or proteins or pulse or receptor or receptors or saliva or scan or scans or scanning or semen or serum or signal* or specimen* or temperature or test or tests or tissue* or tomograph* or ultrasound or urine or "vital signs" or x-ray*).ti,ab,kf. (14445351)
8 (bl or di or dg).fs. (4947785)
9 or/6-8 (16203250)
10 and/5,9 (2503)
11 10 not ((exp animals/ not humans/) or ("animal model" or "animal models" or cat or cats or dog or dogs or marmoset* or mice or mouse or pig or pigs or rat or rats or rodent or sheep or species or swine or bTBI or mTBI or sTBI or PTSD or TBI or "posttraumatic stress" or "post-traumatic stress" or "traumatic brain injury" or "traumatic brain injuries").ti.) (1773)
12 limit 11 to english language (1738)
13 limit 12 to yr="1990 -Current" (1736)

PsycINFO 1806 to February Week 3 2020
Date searched: February 21, 2020
1 ("Desert Saber" or "Desert Sabre" or "Desert Shield" or "Desert Storm" or "Gulf War" or "Gulf Conflict" or "Gulf Crisis" or "Persian Gulf Syndrome" or "Kuwait War" or "Operation GRANBY" or "Op GRANBY").ti,ab. (1135)
2 (GWI or GWIs or GWVI or GWVIs or "Iowa Persian Gulf Study" or "War Related Illness and Injury Study Center*").ti,ab. (64)
3 ((Kuwait or Iraq or "Persian Gulf" or "Southwest Asia" or "SW Asia") and ("air force" or "armed forces" or army or marines or "military personnel" or "national guard*" or naval or navy or "service members" or servicemembers or soldier* or Veteran*)).ti,ab. (2431)
4 or/1-3 (3347)

5 Biological Markers/ (12461)
6 (antigen or antigens or autoantibod* or auto-antibod* or antibody or antibodies or bioassay* or bio-assay* or biological measure* or bio-marker* or biopsy or biopsies or blood or coexpress* or co-express* or conduction or "CT scan*" or cytokine or cytokines or diagnos* or dysfunction* or electromyograph* or endoscop* or fluid or fluids or fMRI or genet* or "gene expression" or imaging or inflammat* or marker or markers or MRI or "Magnetic resonance imaging" or mechanism* or neurodegenerat* or neuro-degenerat* or neuroendocrine or neuro-endocrine or neuroimag* or neuro-imag* or neuroinflammat* or neuro-inflammat* or protein or proteins or pulse or receptor or receptors or saliva or scan or scans or scanning or semen or serum or signal* or specimen* or temperature or test or tests or tissue* or tomograph* or ultrasound or urine or "vital signs" or x-ray*).ti,ab. (1453074)
7 or/5-6 (1453534)
8 and/4,7 (1097)
9 8 not ("animal model" or "animal models" or cat or cats or dog or dogs or marmoset* or mice or mouse or pig or pigs or rat or rats or rodent or sheep or species or swine or bTBI or mTBI or sTBI or PTSD or TBI or "posttraumatic stress" or "post-traumatic stress" or "traumatic brain injury" or "traumatic brain injuries").ti. (625)
10 limit 9 to english language (608)
11 limit 10 to yr="1990 -Current" (608)

EBM Reviews - Cochrane Database of Systematic Reviews 2005 to February 21, 2020
Date searched: February 21, 2020
1 ("Desert Saber" or "Desert Sabre" or "Desert Shield" or "Desert Storm" or "Gulf War" or "Gulf Conflict" or "Gulf Crisis" or "Persian Gulf Syndrome" or "Kuwait War" or "Operation GRANBY" or "Op GRANBY").ti,ab. (0)
2 (GWI or GWIs or GWVI or GWVIs or "Iowa Persian Gulf Study" or "War Related Illness and Injury Study Center*").ti,ab. (0)
3 ((Kuwait or Iraq or "Persian Gulf" or "Southwest Asia" or "SW Asia") and ("air force" or "armed forces" or army or marines or "military personnel" or "national guard*" or naval or navy or "service members" or servicemembers or soldier* or Veteran*)).ti,ab. (0)
4 or/1-3 (0)
5 (antigen or antigens or autoantibod* or auto-antibod* or antibody or antibodies or bioassay* or bio-assay* or biological measure* or bio-marker* or biopsy or biopsies or blood or coexpress* or co-express* or conduction or "CT scan*" or cytokine or cytokines or diagnos* or dysfunction* or electromyograph* or endoscop* or fluid or fluids or fMRI or genet* or "gene expression" or imaging or inflammat* or marker or markers or MRI or "Magnetic resonance imaging" or mechanism* or neurodegenerat* or neuro-degenerat* or neuroendocrine or neuro-endocrine or neuroimag* or neuro-imag* or neuroinflammat* or neuro-inflammat* or protein or proteins or pulse or receptor or receptors or saliva or scan or scans or scanning or semen or serum or signal* or specimen* or temperature or test or tests or tissue* or tomograph* or ultrasound or urine or "vital signs" or x-ray*).ti,ab. (4210)
6 and/4-5 (0)

EBM Reviews - Cochrane Central Register of Controlled Trials January 2020
Date searched: February 21, 2020

1 ("Desert Saber" or "Desert Sabre" or "Desert Shield" or "Desert Storm" or "Gulf War" or "Gulf Conflict" or "Gulf Crisis" or "Persian Gulf Syndrome" or "Kuwait War" or "Operation GRANBY" or "Op GRANBY").ti,ab. (111)
2 (GWI or GWIs or GWVI or GWVIs or "Iowa Persian Gulf Study" or "War Related Illness and Injury Study Center*").ti,ab. (54)
3 ((Kuwait or Iraq or "Persian Gulf" or "Southwest Asia" or "SW Asia") and ("air force" or "armed forces" or army or marines or "military personnel" or "national guard*" or naval or navy or "service members" or servicemembers or soldier* or Veteran*)).ti,ab. (257)
4 or/1-3 (356)
5 (antigen or antigens or autoantibod* or auto-antibod* or antibody or antibodies or bioassay* or bio-assay* or biological measure* or bio-marker* or biopsy or biopsies or blood or coexpress* or co-express* or conduction or "CT scan*" or cytokine or cytokines or diagnos* or dysfunction* or electromyograph* or endoscop* or fluid or fluids or fMRI or genet* or "gene expression" or imaging or inflammat* or marker or markers or MRI or "Magnetic resonance imaging" or mechanism* or neurodegenerat* or neuro-degenerat* or neuroendocrine or neuro-endocrine or neuroimag* or neuro-imag* or neuroinflammat* or neuro-inflammat* or protein or proteins or pulse or receptor or receptors or saliva or scan or scans or scanning or semen or serum or signal* or specimen* or temperature or test or tests or tissue* or tomograph* or ultrasound or urine or "vital signs" or x-ray*).ti,ab. (836972)
6 and/4-5 (197)
7 6 not ("animal model" or "animal models" or cat or cats or dog or dogs or marmoset* or mice or mouse or pig or pigs or rat or rats or rodent or sheep or species or swine or bTBI or mTBI or sTBI or PTSD or TBI or "posttraumatic stress" or "post-traumatic stress" or "traumatic brain injury" or "traumatic brain injuries").ti. (117)

ClinicalTrials.gov
Date searched: February 21, 2020
(EXPAND[Concept] ("Desert Saber" OR "Desert Sabre" OR "Desert Shield" OR "Desert Storm" OR "Gulf War" OR "Gulf Conflict" OR "Gulf Crisis" OR "Persian Gulf Syndrome" OR "Kuwait War" OR "Operation GRANBY" OR "Op GRANBY" OR "GWI" OR "GWIs" OR "GWVI" OR "GWVIs") OR AREA[ConditionSearch] (Gulf AND (illness OR syndrome))) | antigen OR autoantibody OR auto-antibody OR antibody OR bioassay OR bio-assay OR biological measure OR bio-marker OR biopsy OR blood OR coexpression OR co-expression OR conduction OR CT OR cytokine OR diagnosis OR diagnostic OR electromyography OR endoscopy OR fluid OR fMRI OR genetic OR gene OR imaging OR inflammation OR marker OR MRI OR magnetic OR mechanism OR neurodegeneration OR neuro-degeneration OR neuroendocrine OR neuro-endocrine OR neuroimaging OR neuro-imaging OR neuroinflammation OR neuro-inflammation OR protein OR pulse OR receptor OR saliva OR scan OR semen OR serum OR signaling OR specimen OR temperature OR test OR tissue OR tomography OR ultrasound OR urine OR vital or x-ray
(36)

WHO ICTRP
Date searched: February 21, 2020
Condition = "Desert Saber" OR "Desert Sabre" OR "Desert Shield" OR "Desert Storm" OR "Gulf Conflict" OR "Gulf Crisis" OR "Kuwait War" OR "Operation GRANBY" OR "Op

GRANBY" OR (Gulf AND (illness OR syndrome)) OR GWI OR GWIs OR GWVI OR GWVIs
(Without synonyms checked)
Recruitment Status = ALL
(53)

APPENDIX B. STUDY SELECTION

Inclusion codes, code definitions, and criteria

1. Is the full text of the article in English?
   Yes → Proceed to 2.
   No → **Code X1** *(Non-English-language publication).* STOP.

2. Does the population include Veterans with Gulf War Illness?
   Include: Veterans (either U.S. or international) deployed to the Persian Gulf region between Aug 2, 1990 - Nov 1991, defined by the authors as having Gulf War Illness according to a recognized case definition (CDC or Kansas), or defines cases using similar criteria to CDC/Kansas. Also include studies of civilian contractors present during the conflict, if available. Include studies where deployment status and/or time of deployment is unclear.
   Included illness definitions (past and present terms to identify Gulf War Illness): Chronic Multisymptom (or multisystem) Illness (CMI), Irritable Bowel Syndrome (IBS), Chronic Fatigue Syndrome (CFS)/Myalgic Encephalitis(ME), fibromyalgia (FM), Gulf War Syndrome.
   Exclude: children and birth outcomes of Gulf War Veterans.

   Comparator populations may include:
   - Veterans who were deployed elsewhere (other than Persian Gulf) during the Gulf War.
   - Gulf War-deployed Veterans
   - Non-deployed Gulf War era Veterans
   - Civilians with other health conditions/conditions with similar symptomology to GWI (*eg*, chronic fatigue syndrome, neurodegenerative disorders, musculoskeletal problems)
   - Healthy controls

   Yes → Proceed to 3.
   No → Code **X2** *(Excluded population).* STOP.

3. Does the study examine measures of any of the following categories of biological functions/systems that are potential loci of dysfunction:
   - Genes (*eg*, paraoxonase levels, enzyme butyrylcholinesterase)
   - Immune activation/inflammation (*eg*, anti-squalene antibody, natural killer cell activity, humoral immune response, human leukocyte antigen, platelet function, plasma proteins, serum cytokines, peripheral blood lymphocyte factors)
   - Neurodegeneration (*eg*, acetylcholinesterase activity, N-acetylaspartate-to-creatine ratio)
   - Autonomic nervous system (*eg*, feedback regulation of the HPA axis)
   - Endocrine system (*eg*, neuroendocrine-immune signaling)
   - Energy metabolism (*eg*, mitochondrial dysfunction)

- General brain activity (*eg*, synchronous neural interactions, findings from brain imaging (*eg*, fMRI, PET))
- Other

(*Exclude: assessments that do not include biological measurements (eg, questionnaires, symptom inventories)*

Yes → Proceed to 4.
No → Code **X3** *(Not relevant to GWI biological measures).* STOP.

4. Is this study of diagnostic accuracy or a systematic review of such studies?

   Yes →study of diagnostic accuracy. Code **KQ1 diagnostic accuracy [specify test].** STOP.
   Yes → Systematic review. Code **KQ1-SR**. STOP.
   No → Proceed to 5.

5. Is the study a *published* measure of association between biological measures and GWI?
   Yes → Code **Bio-KQ2-[specify biological measure and biological measure category].** STOP.

   No, it is an *unpublished* study that otherwise meets criteria → Code **-KQ3 emerging research [specify biological measure and biological measure category].** STOP.

   No, none of the above → Code **X4.** STOP**.**

Key Questions:
KQ1: Which diagnostic tests (or test combinations) are candidates for distinguishing individuals diagnosed with GWI from individuals without GWI?

KQ2: Which biological measures have been examined for their potential association with GWI, and which among them have been shown to be associated with GWI?

KQ3: Which ongoing or unpublished research studies examine diagnostic tests or biological measures for potential association with GWI?

Exclusion Codes:
X1: Non-English-language publication
X2: Excluded population
X3: Not relevant to GWI biological measures/accuracy of tests
X4: Excluded study design or publication type
X9: Duplicate or preliminary publication of a more recent study
X99: Study terminated

APPENDIX C. QUALITY ASSESSMENT

Table 9. Quality Assessment of Studies of Biological Measures for Gulf War Illness

	Assessment Criteria*										
Study	**#1**	**#2**	**#3**	**#4**	**#5**	**#6**	**#7**	**#8**	**#9**	**#10**	**#11**
Amin, 2011[48]	a (1)	b (0)	a (1)	a (1)	c (0)	a (1)	a (1)	a (1)	b (0)	c (0)	a (1)
Asa, 2000[23]	a (1)	b (0)	a (1)	a (1)	c (0)	d (0)	a (1)	b (0)	d (0)	c (0)	b (0)
Blanchard, 2019[42]	a (1)	d (0)	a (1)	a (1)	a (1)	a (1)	a (1)	a (1)	d (0)	d (0)	b (0)
Butterick, 2019[28]	a (1)	d (0)	a (1)	a (1)	c (0)	d (0)	a (1)	a (1)	d (0)	c (0)	b (0)
Calley, 2010[32]	b (0)	b (0)	a (1)	a (1)	c (0)	a (1)	a (1)	b (0)	c (0)	c (0)	b (0)
Cooper, 2016[33]	b (0)	b (0)	a (1)	b (0)	c (0)	a (1)	a (1)	b (0)	c (0)	c (0)	a (1)
Davis, 2000[43]	b (0)	b (0)	a (1)	a (1)	c (0)	a (1)	a (1)	d (0)	d (0)	c (0)	a (1)
Emmerich, 2017[24]	a (1)	d (0)	a (1)	a (1)	c (0)	a (1)	a (1)	a (1)	a (1)	c (0)	a (1)
Georgopoulos, 2016[25]	d (0)	d (0)	a (1)	a (1)	c (0)	a (1)	a (1)	a (1)	d (0)	c (0)	b (0)
Gopinath, 2012[34]	a (1)	b (0)	a (1)	a (1)	c (0)	a (1)	a (1)	b (0)	a (1)	c (0)	a (1)
Haines, 2017[21]	b (0)	a (1)	a (1)	a (1)	c (0)	d (0)	a (1)	b (0)	c (0)	c (0)	b (0)
Haley, 2013[44]	b (0)	a (1)	a (1)	a (1)	c (0)	a (1)	a (1)	a (1)	a (1)	c (0)	a (1)
Hotopf, 2003[47]	b (0)	a (1)	b (0)	a (1)	a (0)	a (1)	a (1)	a (1)	d (0)	b (0)	a (1)
James, 2016[27]	a (1)	a (1)	a (1)	a (1)	c (0)	b (1)	a (1)	a (1)	c (0)	c (0)	b (0)
Johnson, 2013[26]	a (1)	b (0)	a (1)	a (1)	c (0)	d (0)	a (1)	a (1)	d (0)	c (0)	b (0)
Johnson, 2016[29]	a (1)	b (0)	a (1)	a (1)	c (0)	c (0)	a (1)	a (1)	b (0)	c (0)	b (0)
Li, 2014[45]	b (0)	b (0)	a (1)	a (1)	c (0)	c (0)	a (1)	a (1)	a (1)	b (0)	b (0)
Liu, 2011[35]	a (1)	b (0)	a (1)	b (0)	c (0)	a (1)	a (1)	b (0)	a (1)	c (0)	b (0)
Lo, 2000[22]	b (0)	a (1)	b (0)	a (1)	c (0)	a (1)	a (1)	b (0)	c (0)	c (0)	a (1)
Nagelkirk, 2003[46]	a (1)	d (0)	a (1)	b (0)	c (0)	b (1)	a (1)	b (0)	d (0)	c (0)	a (1)
Odegard, 2013[36]	b (0)	b (0)	a (1)	b (0)	c (0)	a (1)	a (1)	a (1)	d (0)	c (0)	b (0)
Phillips, 2009[30]	a (1)	b (0)	a (1)	b (0)	c(0)	b (1)	a (1)	b (0)	c (0)	c (0)	a (1)
Roland, 2000[49]	b (0)	b (0)	a (1)	b (0)	c (0)	a (1)	a (1)	a (1)	d (0)	c (0)	b (0)
Sharief, 2002[50]	b (0)	a (1)	a (1)	a (1)	c (0)	c (0)	a (1)	a (1)	a (1)	b (0)	a (1)
Skowera, 2004[31]	b (0)	a (1)	a (1)	a (1)	c (0)	a (1)	a (1)	a (1)	d (0)	a (1)	a (1)
Tillman, 2010[37]	b (0)	d (0)	a (1)	a (1)	c (0)	a (1)	a (1)	b (0)	d (0)	c (0)	b (0)
Tillman, 2012[38]	b (0)	b (0)	a (1)	a (1)	c (0)	d (0)	a (1)	a (1)	a (1)	c (0)	b (0)
Tillman, 2013[39]	b (0)	d (0)	a (1)	b (0)	c (0)	a (1)	a (1)	b (0)	d (0)	c (0)	b (0)
Tillman, 2019[40]	b (0)	a (1)	a (1)	a (1)	c (0)	d (0)	a (1)	a (1)	a (1)	c (0)	b (0)
Wallace, 1999[51]	a (1)	d (0)	a (1)	a (1)	c (0)	a (1)	a (1)	b (0)	b (0)	c (0)	a (1)
Weiner, 2011[41]	a (1)	d (0)	a (1)	a (1)	c (0)	b (1)	a (1)	a (1)	d (0)	c (0)	b (0)
Zhou, 2018[52]	b (0)	b (0)	a (1)	a (1)	c (0)	b (1)	a (1)	b (0)	d (0)	c (0)	b (0)

*Quality Assessment Criteria (adapted from Newcastle-Ottawa[19] and BIOCROSS[20]):

1. Is the case definition adequate?
    a. Yes: CDC or Kansas definition (+1)
    b. All other definitions (0)
2. Representativeness of cases and controls:

   a. Truly representative of the population of both GWI+ and GWI- Veterans (*ie,* total pop[census] or random sampling) (+1)
   b. Non-random selection of either GWI+ or GWI- subjects (0)
   c. No description of the sampling strategy (0)
3. Selection of controls: Were D-GWVs controls selected or recruited from the same population as cases (including the same time period)?
   a. Yes (+1)
   b. No (0)
4. Were inclusion and exclusion criteria for being in the study prespecified and applied uniformly to all participants?
   a. Yes (+1)
   b. No/unclear (0)
5. Sample size/power calculation:
   a. Reported having conducted a power analysis, and then used an appropriate sample size based on that analysis (+1)
   b. Reported having conducted a power analysis, but were not able to/did not use an appropriate sample size (0)
   c. Did not report having conducted a power analysis (0)
6. Comparability of cases and controls on the basis of the design or analysis:
   a. Study controls for important confounders like demographics (age, gender, comorbidity, *etc*) through matching participants or statistical adjustment (+1)
   b. Did not match by age or gender, nor adjust for confounders in analysis, but demographic analysis found no statistically significant differences on these variables (+1).
   c. There were significant descriptive differences that were not adjusted for (0)
   d. No matching and/or demographics not reported (0)
7. Were the biological measure measurements (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?
   a. Yes (+1)
   b. No (0)
8. Biological measure data modeling: Was the distribution of biological measure data reported (if non-normal were statistical approaches used to standardize it)? Were methods of outlier detection and handling used? Were any possible errors resulting from measurement inaccuracies discussed?
   a. Any of the above were addressed (+1)
   b. Unclear/did not report (0)
   c. Reported but inadequate (0)
9. If there were multiple comparisons, did they adjust appropriately (*eg*, Bonferroni)?
   a. Yes (1)
   b. No (0)
   c. N/A (no penalty, 0)
   d. Not reported (0)
10. Non-Response rate (for enrollment):

   a. Same rate for both groups, or rate differs but is weighted statistically (+1)
   b. Unequal response rate, non-respondents are described (with no statistical adjustment) (0)
   c. Unclear/not reported (0)
11. Blinding: Were the assessors of the outcome measurement (biological measure) blinded to the (case or control) status of participants?
   a. Yes (+1)
   b. No/not reported (0)

APPENDIX D. SUPPLEMENTAL MATERIAL

Table 10. Studies of Gulf War Illness Biological Measures Using Lower-priority Comparator Groups* or No Comparator

Author, year	Title	Biological Measure/ Outcome Measure	Findings	Findings promising?†
IMMUNE SYSTEM				
Abou-Donia, 2017[83]	Screening for novel central nervous system biological measures in Veterans with Gulf War Illness	Autoantibodies reactive to specified proteins	GWVs with GWI had higher had higher levels of autoantibody reactivity in all proteins examined except S-100B comparted to healthy, non-Veterans with low back pain (GFAP p b 0.001; Tau p b 0.001; MAP p b 0.002; MAG p b 0.001; PNF p b 0.006; Tubulin p b 0.003; MBP p b 0.01; S-100B p = 0.31)	Yes
Brimacombe, 2002[84]	Immunological variables mediate cognitive dysfunction in Gulf War Veterans but not civilians with chronic fatigue syndrome	Patterns of cytokine/symptom relationships	A type 2 cluster of chronic fatigue syndrome plus a T and B cell factor predicted CFS cases for GWVs but not civilians with CFS, which was modulated by reaction time	Yes
Broderick, 2011[85]	Altered immune pathway activity under exercise challenge in Gulf War Illness: an exploratory analysis	Gene expression pathways, cytokines in plasma, lymphocytes, cytotoxicity, with exercise challenge	Mutual information networks linking immune markers in GWI had more abundant connections but were less organized than non-Veteran health controls during and after exercise.	Yes
Broderick, 2018[86]	A pilot study of immune network remodeling under challenge in Gulf War Illness	Immune markers, with exercise challenge	GWI compared to control networks of immune signaling during exercise had more abundant connections but were less organized. NPY, IL-1α, TNF-α and CD2+/CD26+ nodes were better integrated in the GWI network at rest. Under effort (t_1) these differences were replaced by significant restructuring around nodes for CD19+ B cell population, IL-5, IL-6 and soluble CD26 concentrations.	No
Diaz-Torne, 2007[87]	Absence of histologic evidence of synovitis in patients with Gulf War Veterans' illness with joint pain	Synovial biopsy samples	GWVI synovia (synovitis, osteoarthritis, and rheumatoid arthritis scores) did not differ from normal controls.	No
Everson, 2002[88]	Immunological responses are not abnormal in symptomatic Gulf War Veterans	Humoral immune responses	Immune response measures in antigen presenting cells, T cells, type 1-2 T-helper cells, and B cells did not differ between GWI-symptomatic GWVs vs matched controls (asymptomatic Veterans, non-Gulf War Veterans	No
Golomb, 2019[89]	Depressed prostaglandins and leukotrienes in Veterans with Gulf War Illness	Eicosanoids - prostaglandins and leukotrienes	Several plasma eicosanoid levels were lower in GWI vs non-Veteran controls.	Yes

Author, year	Title	Biological Measure/ Outcome Measure	Findings	Findings promising?†
Halpin, 2017[90]	Myalgic encephalomyelitis/chronic fatigue syndrome and Gulf War Illness patients exhibit increased humoral responses to the herpesviruses-encoded dUTPase: Implications in disease pathophysiology	Antibodies against multiple human herpesviruses-encoded dUTPases and/or the human dUTPase	GWI participants had higher levels of antibodies to the HHV-6 and human dUTPases than healthy controls (p=0.0053 and p=0.0036, respectively).	Yes
Hannan, 2000[91]	Activation of the coagulation system in Gulf War Illness: a potential pathophysiologic link with chronic fatigue syndrome. A laboratory approach to diagnosis	Coagulation and platelet activation	More GWVs with GWI (23/33) than healthy controls (0/33) had 2 or more positive scores on the Immune System Activation of Coagulation panel (p<0.001), the laboratory criterion for activation of coagulation.	Yes
Khaiboullina, 2015[92]	Cytokine expression provides clues to the pathophysiology of Gulf War Illness and myalgic encephalomyelitis	77 serum cytokines	A group of 77 cytokines identified myalgic encephalomyelitis (ME) and GWI with sensitivities of 92.5% and 64.9%, respectively. When ME and GWI were compared to healthy controls, the specificity was 33.3%.	No
Klaustermeyer, 1998[93]	Allergic and immunologic profile of symptomatic Persian Gulf War Veterans	Total serum IgE levels	GWVs with allergy symptoms had higher mean IgE level (88.7 I/U/mL) than GWVs without allergy symptoms (47.5 IU/mL)	No
O'Bryan, 2003[94]	Human leukocyte antigens in Gulf War Veterans with chronic unexplained multiple symptoms	Frequency of antigens: HLA-A, -B, -DR, -DQ	Human Leukocyte Antigen-A28 was present in 21.9% of symptomatic Veterans and 6.9% of the healthy population (p=0.01), but not significant when corrected for number of antigens determined.	No
Parkitny, 2015[95]	Evidence for abnormal cytokine expression in Gulf War Illness: A preliminary analysis of daily immune monitoring data	Serum cytokine and chemokine concentrations	No difference in serum cytokine concentrations between GWI and healthy GWV. GWI associated with higher variability in the expression of eotaxin-1 than healthy GWVs (p<0.001).	Yes
Skowera, 2002[96]	Antinuclear autoantibodies (ANA) in Gulf War-related illness and chronic fatigue syndrome (CFS) patients	Antinuclear Autoantibodies	No difference in prevalence of antinuclear autoantibodies between symptomatic GWV, healthy GWV, symptomatic Bosnia and Era Veterans, chronic fatigue syndrome patients, and health control subjects.	No
Smylie, 2013[97]	A comparison of sex-specific immune signatures in Gulf War Illness and chronic fatigue syndrome	Cytokine markers with exercise challenge	No differences between GWI and controls indicated. Differences in cytokine markers by sex.	No
Tsilibary, 2018[98]	Human Immunoglobulin G (IgG) Neutralizes Adverse Effects of Gulf War Illness (GWI) Serum in	Human IgG	Cell spreading was lower in GWI than control ($p=4.4 \times 10^{-34}$). GWI apoptosis was higher than control ($=-6.91 \times 10^{-24}$)	Yes

Author, year	Title	Biological Measure/ Outcome Measure	Findings	Findings promising?†
	Neural Cultures: Paving the Way to Immunotherapy for GWI			
Vojdani, 2004[99]	Cellular and humoral immune abnormalities in Gulf War Veterans	Percentage of immunological markers	Percentage of T cells in symptomatic GWV(sGWV) v. controls not different. More sGWVs had elevated T cells than controls. More B cells in sGWVs v controls. Natural Killer cell activity decreased in patients (24.8 ± 16.5 lytic unit) v controls (37.3 ± 26.4 lytic unit). Immune complexes increased in patients (53.1 ± 18.6, mean ± SD) v controls (34.6 ± 14.3). Autoantibody titers directed against myelin basic protein and striated or smooth muscle greater in sGWVs v control.	Yes
Whistler, 2009[100]	Impaired immune function in Gulf War Illness	Immune cell function with exercise challenge	Differences for 3 Natural Killer cell subsets and Natural Killer cytotoxicity between GWI and controls ($p<0.05$).	Yes
Zhang, 1999[101]	Changes in immune parameters seen in Gulf War Veterans but not in civilians with chronic fatigue syndrome	Lymphocyte subpopulations, cytokine gene expression	Veterans with chronic fatigue syndrome had more total T cells and MHC II^+ T cells and higher percentage of these lymphocyte subpopulations, and lower percentage of Natural Killer cells, than controls. Also had higher levels of IL-2, IL-10, IL-10, IFN-(symbol), and TNF-(alpha symbol) than controls.	Yes
CENTRAL NERVOUS SYSTEM				
Alshelh, 2020[102]	In-vivo imaging of neuroinflammation in Veterans with Gulf War Illness	[11C]PBR28 PET/MRI	GWI had higher cortical [11C]PBR28 PET signal in precuneus, prefrontal, primary motor, and somatosensory cortices compared to both healthy non-Veterans and healthy Veterans. No group differences in inflammatory cytokines.	Yes
Baraniuk, 2005[103]	A Chronic Fatigue Syndrome - related proteome in human cerebrospinal fluid	Proteomes in cerebrospinal fluid	Pooled chronic fatigue syndrome and GWI samples contained proteins in the cerebrospinal fluid not detected in the control sample: α-1-macroglobulin, amyloid precursor-like protein 1, keratin 16, orosomucoid 2 and pigment epithelium-derived factor.	Yes
Chao, 2014[104]	Associations between subjective sleep quality and brain volume in Gulf War Veterans	Cortical, lobar gray matter, and hippocampal volumes	Global Pittsburgh Sleep Quality Index was associated with total cortical and frontal gray matter volume in GWV, and, in the frontal lobe, total Global Pittsburgh Sleep Quality Index was inversely associated with the superior and middle frontal, orbitofrontal, anterior cingulate, and frontal pole volumes.	No
Chao, 2019[105]	Do Gulf War Veterans with high levels of deployment-related exposures display symptoms suggestive of Parkinson's disease?	Total basal ganglia volume	GWI had lower total basal ganglia volume than healthy deployed Veterans.	Yes
Christova, 2017[106]	Subcortical brain atrophy in Gulf War Illness	Subcortical brain atrophy	GWI had subcortical brain atrophy compared to healthy controls.	Yes

Author, year	Title	Biological Measure/ Outcome Measure	Findings	Findings promising?†
Clarke, 2019[107]	Connectivity differences between Gulf War Illness (GWI) phenotypes during a test of attention	Exercise challenge: brain activation (fMRI BOLD response)	Unique brain activation connectivity patterns between control and GWI groups. Controls had an exercise task related network of right dorsolateral and left ventrolateral prefrontal cortex, dorsal anterior cingulate cortex, posterior insulae and frontal eye fields. GWI subgroup with brain stem atrophy and postural tachycardia after exercise had activity in the dorsal anterior cingulate cortex with direct links to basal ganglia, anterior insulae, and right dorsolateral prefrontal cortex notes. GWI subgroup with stress test originated phantom perception had submodules of basal ganglia-anterior insulae, and dorsolateral prefrontal executive control regions.	Yes
Concato, 2007[108]	Acetylcholinesterase activity in Veterans of the first Gulf War	Acetylcholinesterase activity	Acetylcholinesterase activity was similar for Veterans with versus without GWI.	No
Engdahl, 2016[109]	A Magnetoencephalographic (MEG) Study of Gulf War Illness (GWI)	Synchronous neural interactions	Differences in synchronous neural interactions between GWI and healthy controls centered in the cerebellum and frontal cortex.	Yes
Georgopoulos, 2017[110]	Gulf War Illness (GWI) as a neuroimmune disease	Synchronous neural interactions	GWI synchronous neural interactions did not differ from relapse-remitting multiple sclerosis, Sjogren's syndrome, or rheumatoid arthritis, but did differ from control, schizophrenia, Alzheimer's disease, post-traumatic stress disorder, and major depressive disorder.	Yes
Gopinath, 2019[111]	Exploring brain mechanisms underlying Gulf War Illness with group ICA based analysis of fMRI resting state networks	Resting state fMRI	Impaired functional connectivity in GWI between language networks, sensory input networks, motor output networks, between different sensory perception and motor networks, and between different networks in the sensorimotor domain.	Yes
Haley, 1997[112]	Evaluation of neurologic function in Gulf War Veterans. A blinded case-control study	Neurophysiological, audiovestibular, neuroradiological, blood cell count, erythrocyte sedimentation rate	GWI had greater inter-side asymmetry of the wave I to wave III interpeak latency of brain stem auditory evoked potentials, greater interocular asymmetry of nystagmic velocity on rotational testing, increased asymmetry of saccadic velocity, more prolonged interpeak latency of the lumbar-to-cerebral peaks on posterior tibial somatosensory evoked potentials, and diminished nystagmic velocity after caloric stimulation bilaterally.	Yes
Haley, 2000[113]	Brain abnormalities in Gulf War syndrome: evaluation with 1H MR spectroscopy	N-acetyl aspartate-to-creatine ratio, measuring neuronal mass	N-acetyl aspartate-to creatine (NAA/Cr) ratio (functional neuronal mass) was lower in the basal ganglia and brainstem of GWVs than in control participants (p=0.007).	Yes
Haley, 2009[114]	Abnormal brain response to cholinergic challenge in chronic encephalopathy from the 1991 Gulf War	Brain response to cholinergic challenge; normalized regional cerebral blood flow	Baseline normalized regional cerebral blood flow in chronically ill GWVs was lower than controls throughout deep structures.	Yes

Author, year	Title	Biological Measure/ Outcome Measure	Findings	Findings promising?†
Hubbard, 2014[115]	Central Executive Dysfunction and Deferred Prefrontal Processing in Veterans with Gulf War Illness	Brain activation (BOLD fMRI) during working memory task	GWI deferred prefrontal cortex activity from encoding to retrieval for high demand conditions.	Yes
Jamal, 1996[116]	The "Gulf War syndrome". Is there evidence of dysfunction in the nervous system?	Peripheral nerve function	Three measures of peripheral nerve function were abnormal in Veterans compared to controls: cold threshold (p=0.0002), sural nerve latency (p=0.034), and median nerve sensory action potential (p=0.030).	Yes
James, 2017[117]	Human Leukocyte Antigen (HLA) and Gulf War Illness (GWI): HLA-DRB1 13:02 Spares Subcortical Atrophy in Gulf War Veterans	Volume of cerebellar gray matter	Human leukocyte allele DRB1*12:02 spared subcortical brain atrophy in GWVs and subcortical volume was higher in carriers of the allele, and in cerebellar grey matter.	Yes
Li, 2011[118]	Hippocampal dysfunction in Gulf War Veterans: investigation with ASL perfusion MR imaging and physostigmine challenge	Hippocampal regional cerebral blood flow	Decreased hippocampal regional cerebral blood flow with physostigmine challenge in control subjects (p<0.0005) and Veterans with syndrome 1 (impaired cognition) (p<0.05), and increased in syndrome 2 (confusion-ataxia) (p<0.005) and syndrome 3 (central neuropathic pain) (p<0.002).	Yes
Menon, 2004[119]	Hippocampal dysfunction in Gulf War Syndrome. A proton MR spectroscopy study	N-acetyl aspartate to creatine and choline to creatine ratios	The N-acetyl aspartate/creatine ratio of the GWI group was lower that control group.	Yes
Moffett, 2015[120]	Word-finding impairment in Veterans of the 1991 Persian Gulf War	Brain activation (BOLD signal fMRI) during cognitive task	GWI group had reduced activity in the thalamus, putamen, and amygdala, and increased activity in the right hippocampus relative to controls	Yes
Rayhan, 2013[121]	Prefrontal lactate predicts exercise-induced cognitive dysfunction in Gulf War Illness	Brain activation with exercise challenge	GWI who had decreased working memory performance after exercise had elevated prefrontal lactate levels compared to GWI who had increased performance.	Yes
Rayhan, 2013[122]	Exercise challenge in Gulf War Illness reveals 2 subgroups with altered brain structure and function	Brain activation (BOLD fMRI) with exercise challenge	GWI subgroup with orthostatic tachycardia correlated with brainstem atrophy, baseline working and memory compensation in the cerebellar vermis. The other GWI subgroup that developed exercise- induced hyperalgesia was associated with cortical atrophy and baseline working memory compensation in the basal ganglia.	Yes
Rayhan, 2013[123]	Increased brain white matter axial diffusivity associated with fatigue, pain and hyperalgesia in Gulf War Illness	White matter diffusivity properties	GWI had increased axial diffusivity in the right inferior frontal-occipital fasciculus, but not in controls.	Yes
Rayhan, 2019[124]	Exercise challenge alters Default Mode Network dynamics in Gulf War Illness	Brain activation patterns with exercise	GWI had increase in deactivation patterns within the Default Mode Network following exercise that was not seen in controls.	Yes

Author, year	Title	Biological Measure/ Outcome Measure	Findings	Findings promising?†
Tillman, 2017[125]	Electrophysiological correlates of semantic memory retrieval in Gulf War Syndrome 2 patients	Brain activation (ERP) with cognitive task	GWI had an event-related potential difference between memory retrieval and no memory retrieval stimuli at the midline parietal region that had a scalp voltage polarity opposite from that recorded at the left temporal area that was not present in controls.	Yes
Turner, 2016[126]	Cognitive Slowing in Gulf War Illness Predicts Executive Network Hyperconnectivity: Study in a Population-Representative Sample	Brain activation (BOLD fMRI) during cognitive task	Bilateral dorsolateral prefrontal cortex connectivity with task-relevant notes was altered in GWI participants compared to healthy controls during processing speed task.	Yes
Washington, 2020[127]	Exercise alters cerebellar and cortical activity related to working memory in phenotypes of Gulf War Illness	Brain activity with working memory task/ exercise	GWI with stress test associated reversible tachycardia has post-exertional deactivation of cerebellar dentate nucleus and vermis regions associated with working memory. GWI stress tests originated phantom perception had activation of the anterior supplementary motor area .	Yes
Wylie, 2019[128]	Fatigue in Gulf War Illness is associated with tonically high activation in the executive control network	Brain activation (BOLD fMRI) with cognitive challenge	GWI had greater activation than healthy controls in frontal and parietal areas for less difficult cognitive tasks.	Yes
AUTONOMIC NERVOUS SYSTEM				
Falvo, 2018[129]	Dynamic cerebral autoregulation is impaired in Veterans with Gulf War Illness: A case-control study	Cerebral blood flow responses to physostigmine challenge	Greater decreases in cerebral blood flow both a nadir and after standing and during steady state standing in GWI vs controls. Dynamic autoregulation was lower in GWI than controls. Cerebrovascular reactivity was not different between groups.	Yes
Fiedler, 2004[130]	Responses to controlled diesel vapor exposure among chemically sensitive Gulf War Veterans	Responses to diesel vapor exposure: Heart rate, blood pressure, respiration rate, end-tidal CO(2)	GWI had reduced end-tidal CO2 after exposure to diesel and petrochemical fumes compared to controls and were physiologically hyporeactive in response to behavioral tasks administered during, but not before, exposure.	Yes
Haley, 2004[131]	Blunted circadian variation in autonomic regulation of sinus node function in Veterans with Gulf War syndrome	Heart-rate variability by 24-hr electrocardiography, ambulatory blood pressure, Valsalva ratio, sympathetic skin response, sweat imprint test measures	GWI had less increase (1.2-fold) in high-frequency spectral power of heart rate variability during sleep compared to normal increase (2.2-fold) in controls. In GWI, it was lower at night, higher in morning, but no difference from controls during rest of the day. GWI heart rate declined less at night and corrected QT intervals were longer over 24 hours, particularly at night.	Yes
Peckerman, 2000[132]	Cardiovascular stress responses and their relation to symptoms in	Hemodynamic responses to stressors	Veterans with chronic fatigue had diminished blood pressure responses during cognitive stress tests due to unusually small increases in total peripheral resistance. Similar blood pressure	Yes

Author, year	Title	Biological Measure/ Outcome Measure	Findings	Findings promising?†
	Gulf War Veterans with fatiguing illness		responses to cold pressor test in Veterans with chronic fatigue and healthy Veterans.	
Stein, 2004[133]	Sex effects on heart rate variability in fibromyalgia and Gulf War Illness	Heart rate variability	No group differences in heart rate variability.	No
GENETIC				
Baraniuk, 2017[134]	Exercise-induced changes in cerebrospinal fluid miRNAs in Gulf War Illness, Chronic Fatigue Syndrome and sedentary control subjects	MicroRNAs in cerebrospinal fluid	No group differences in microRNAs in cerebrospinal fluid. After exercise, GWI Stress Test Originated Phantom Perception participants had lower miR-22-3p than control and GWI Stress Test Activated Reversible Tachycardia, but higher miR-9-3p than Stress Test Originated Phantom Perception participants.	Yes
Craddock, 2015[135]	Using gene expression signatures to identify novel treatment strategies in Gulf War Illness	Gene Expression Signatures	Found 19 functional modules with significantly altered gene expression patterns in GWI.	Yes
Liu, 2018[136]	Detecting Chromosome Condensation Defects in Gulf War Illness Patients	Chromosome condensation defects	In GWI, 3 subtypes of Defective Mitotic Figures. Another type of condensation defect identified as sticky chromosomes were observed.	Yes
Mackness, 2000[137]	Low paraoxonase in Persian Gulf War Veterans self-reporting Gulf War Syndrome	Paraoxonase	GWVs paraoxon hydrolysis was less than 50%of that found in controls. Serum PON1 concentration was lower in GWV. No group difference in rate of diazoxon hydrolysis.	Yes
NCT00810225, 2008[138]	Study of Gulf War Illness (GWI) by Comparing GWI and Healthy Veterans	CNDP1 gene, cerebrospinal fluid proteome contents	N/A	N/A
Steele, 2015[139]	Butyrylcholinesterase genotype and enzyme activity in relation to Gulf War Illness: preliminary evidence of gene-exposure interaction from a case-control study of 1991 Gulf War Veterans	Butyrylcholinesterase Genotype and Enzyme Activity	No difference between GWI and controls in mean butyrylcholinesterase (BChE) enzyme activity level or BChE genotype.	No
Trivedi, 2019[140]	Alterations in DNA Methylation Status Associated with Gulf War Illness	DNA methylation patterns in peripheral blood mononuclear cells	Global DNA methylation levels not different in GWI v controls. Genome-wide assessment indicated hypermethylation in GWI in 88% of CpG sites across gene regulatory elements and within coding regions.	Yes
Urnovitz, 1999[141]	RNAs in the sera of Persian Gulf War Veterans have segments homologous to chromosome 22q11.2	Amplicons	Genetic alterations in the 22q11.2 region in GWI.	Yes

Author, year	Title	Biological Measure/ Outcome Measure	Findings	Findings promising?†
Vladutiu, 2004[142]	Association of medically unexplained fatigue with ACE insertion/deletion polymorphism in Gulf War Veterans	Frequency of mutant alleles associated with metabolic myopathies or genetic variation associated with physical performance	Increased risk for chronic fatigue syndrome/idiopathic chronic fatigue was associated with alterations of the insertion/deletion polymorphism in the angiotensin-converting enzyme gene in GWV. The I allele frequency was decreased in affected vs unaffected Veterans. The II genotype was decreased 4-fold in affected Veterans DD genotype was increased 2-fold.	Yes
OTHER				
Bacterial				
Nicolson, 2003[143]	High prevalence of Mycoplasma infections in symptomatic (chronic fatigue syndrome) family members of Mycoplasma-positive Gulf War Illness patients	Presence of bacterial infection	Over 80% of GWI who were positive for blood mycoplasma infections had only 1 Mycoplasma spp., M. Fermentans, vs healthy controls with 8.5% incidence of mycoplasma	Yes
Biochemical Pathways				
Naviaux, 2019[144]	Metabolic features of Gulf War Illness	Abnormalities in biochemical pathways, surveyed via broad-spectrum serum metabolomics	GWI, compared to healthy controls, had abnormalities in 8 of 46 biochemical pathways. Lipid abnormalities accounted for 78% of the metabolic impact.	Yes
Circulatory System				
Falvo, 2018[145]	Abnormal rheological properties of red blood cells as a potential marker of Gulf War Illness: A preliminary study	Red blood cell deformability and aggregation	Red blood cells were more deformable in GWI, as indicated by higher elongation indices particularly at higher shear stress values when compared to matched controls.	Yes
Energy Metabolism				
Chen, 2017[146]	Role of mitochondrial DNA damage and dysfunction in Veterans with Gulf War Illness	Mitochondrial DNA damage and dysfunction	Mitochondrial DNA lesion frequency and mitochondrial DNA copy number were elevated in GWI vs controls.	Yes
Koslik, 2014[12]	Mitochondrial dysfunction in Gulf War Illness revealed by Phosphorus Magnetic Resonance Spectroscopy: a case-control study	Calf muscle phosphocreatine	Post-exercise phosphocreatine-recovery time constant was prolonged in GWI vs controls.	Yes
Gastrointestinal				
Lin, 2009[147]	Bacterial Overgrowth Associated with Chronic Multisymptom Illness Complex	Hydrogen and methane in breath	The proportion of Fusobacteria in the was increased in GWI vs controls in the jejunum. In the ileum, Proteobacteria were reduced in GWI vs controls.	Yes

Author, year	Title	Biological Measure/ Outcome Measure	Findings	Findings promising?†
Nervous System				
Fletcher, 2010[148]	Plasma neuropeptide Y: a biological measure for symptom severity in chronic fatigue syndrome	Neuropeptide Y in plasma	Plasma neuropeptide Y elevated in chronic fatigue syndrome participants vs controls and GWI.	Yes
Khan, 2004[149]	Peripheral cholinergic function in humans with chronic fatigue syndrome, Gulf War syndrome and with illness following organophosphate exposure	Skin blood flow responses to iontophoresis of acetylcholine and of methacholine	Response to acetylcholine was higher in participants with chronic fatigue syndrome than controls, but normal in GWI and those exposed to organophosphates. The methacholine response was higher than acetylcholine response in all patient groups compared to controls except for those with chronic fatigue syndrome.	Yes
Respiratory				
Lindheimer, 2019[150]	Veterans with Gulf War Illness exhibit distinct respiratory patterns during maximal cardiopulmonary exercise	Ventilatory variables (minute ventilation, respiratory frequency, tidal volume) in response to maximal cardiopulmonary exercise	Ventilator variables measured during exercise stress test indicated minute ventilation was not different but tidal volume was greater and respiratory frequency was lower in GWI than controls.	Yes
Skeletal				
Compston, 2002[151]	Reduced bone formation in UK Gulf War Veterans: a bone histomorphometric study	Bone measures: cancellous bone area, mineral apposition rate, mean wall width, bone formation rate at tissue level	Measures from iliac crest bone biopsies showed that cancellous bone area was lower in GWVs vs healthy controls, and this was associated with reduced mineral apposition rate, mean wall width, and bone formation rate at the tissue level.	Yes
Pessler, 2008[152]	A histomorphometric analysis of synovial biopsies from individuals with Gulf War Veterans' Illness and joint pain compared to normal and osteoarthritis synovium	Histologic, immunohistochemical, and vascular measures in synovial biopsies	Measures from synovial biopsies indicated no difference between GWI and healthy controls in histologic appearance.	No
Various				
NCT00810329, 2008[153]	Proteomics of Cerebrospinal Fluid in Chronic Fatigue Syndrome	Proteins in cerebrospinal fluid, cerebrospinal pressure, ANS function, pulmonary function, pain threshold, allergic response	N/A - ongoing	N/A - ongoing

* Priority comparator groups=Deployed GWVs without GWI, with or without other health conditions. See main report for studies using priority comparator groups. This table includes studies of biological measures in GWVs with GWI (loosely defined) compared to other groups, or with no comparator group.

† Yes=Indication by statistical significance of association of a biological measure with GWI case status; No=No indication by statistical significance of association of a biological measure with GWI case status

Abbreviations: ACE=Angiotensin-Converting Enzyme; ANA=Antinuclear Antibody; ANS=Autonomic Nervous System; ASL=Arterial Spin Labelling; BOLD=Blood-Oxygen-Level-Dependent; CFS=Chronic Fatigue Syndrome; CNDP1=Carnosine Dipeptidase 1; CO_2=Carbon Dioxide; dUTPase=Deoxyuridine Triphosphate Diphosphatase; DNA=Deoxyribonucleic Acid; ERP=Event Related Potential; fMRI=functional Magnetic Resonance Imaging; GWI=Gulf War Illness; HLA=Human Leukocyte Antigen; ICA=Independent Component Analysis; IgE=Immunoglobulin E; IgG=Immunoglobulin G; MEG= Magnetoencephalograph; miRNA=Micro Ribonucleic Acid; MR=Magnetic Resonance; MRI=Magnetic Resonance Imaging; NCT=National Clinical Trial; PET=Positron Emission Tomography; UK=United Kingdom

Table 11. Gulf War Illness Biological Measure Studies with Insufficient Sample Size (N<25)

Study Author, Year	Title	Biological Measure/ Outcome Measure	Findings	Findings promising?*
IMMUNE SYSTEM				
Broderick, 2013[154]	Exploring the Diagnostic Potential of Immune Biomarker Co-expression in Gulf War Illness	Projection model based on markers of endocrine and immune function	Increases in neuroendocrine-immune signaling and inflammatory activity in GWI with decreased apoptotic signaling associated with exercise stress test.	Yes
CENTRAL NERVOUS SYSTEM				
Bunegin, 2001[155]	Cognitive performance and cerebrohemodynamics associated with the Persian Gulf Syndrome	Middle cerebral artery blood flow velocity with acetone challenge	No difference in pulmonary function tests between GWI and controls breathing clean air or 40 ppm acetone in air. Middle cerebral artery blood flow velocity increases for each of clean air, clean air placebo, and mixture of air and acetone were different between groups.	Yes
Haley, 2000[156]	Effect of basal ganglia injury on central dopamine activity in Gulf War syndrome: correlation of proton magnetic resonance spectroscopy and plasma homovanillic acid levels	Functioning neuronal mass (N-acetyl-aspartate to creatine ratio)	Homovanillic acid: 3-methoxy-4-hydroxyphenlyglycol was inversely associated with functioning neuronal mass in the left basal ganglia but not the right.	Yes
GENETIC				
Latimer, 2020[157]	Preliminary Evidence for a Hormetic Effect on DNA Nucleotide Excision Repair in Veterans with Gulf War Illness	DNA nucleotide excision repair capacity	Total gene expression and nucleotide excision repair differed between GWI and controls.	Yes
OTHER				
Janulewicz, 2019[158]	The Gut-Microbiome in Gulf War Veterans: A Preliminary Report	Gut microbiome patterns	GW controls had more but firmicutes and the GWI plus gastrointestinal symptoms had more phyla bacteroidetes, actinobacteria, euryarchaeota, and proteobacteria, and Bacteroidaceae, Erysipelotrichaceae, and Bifidobacteriaceae. GWI plus gastrointestinal symptoms also showed greater plasma levels of the inflammatory cytokine TNF-RI.	Yes

*Yes=Indication by statistical significance of association of a biological measure with GWI case status; No=No indication by statistical significance of association of a biological measure with GWI case status.
Abbreviations: DNA=Deoxyribonucleic Acid, GWI=Gulf War Illness; TNF-RI=Tumor Necrosis Factor-Receptor 1

APPENDIX E. PEER REVIEW COMMENTS/AUTHOR RESPONSES

Reviewer number	Comment	Author Response
Are the objectives, scope, and methods for this review clearly described?		
1	No - See detailed comments--needs to be clearer conceptually	
4	Yes	
5	Yes	
6	Yes	
Is there any indication of bias in our synthesis of the evidence?		
1	No	
4	No	
5	No	
6	No	
Are there any published or unpublished studies that we may have overlooked?		
1	Yes - I don't know, but I would guess that requiring the 2 IOM-approved definitions might exclude the best studies, which would be done in a very sick group of GWI patients vs controls. It is impossible to get anywhere with GWI using the Kansas and CDC definitions because they include a very diverse, mostly not very sick, very large groups of veterans.	Regardless of case definition restrictions, we found no studies that could answer KQ1, so no studies were excluded based on case definition for KQ1. We agree that the CDC and Kansas definitions include a heterogeneous group of symptoms and do not specify symptom severity. Also, a larger challenge that restricts our review and the GWI research is that CDC and Kansas case definitions are currently recommended for use in research to identify GWI, so the preponderance of studies use one of these as their criteria. We acknowledge this challenge in the discussion.
4	No	
5	Yes • VA Million Veterans Program consisting of biological samples and clinical data from thousands of GW veterans. • VA Cooperative Studies Program 585 - various studies using a repository including blood specimens (serum, buffy coat, DNA) from hundreds of GW veterans.	We identified one study[58] from the VA Million Veterans Program. This study was also part of the VA CSP 2007. Unfortunately, upon reviewing publications and products from the CSP 585 program, we were unable to identify published studies meeting our selection criteria.
6	No	
Additional suggestions or comments can be provided below. If applicable, please indicate the page and line numbers from the draft report.		
1	OVERVIEW Clearly a thorough report and largely a good review of the included studies. However, the report could be improved in several ways, and some of the methods used are not well-	Thank you.

	justified. More detail might help, but the biggest concern is that the report is conceptually weak. Some concerns are:	
	1) using adherence to the CDC and Kansas definitions as a criterion for inclusion; conceptually, if you are after a biomarker related to "symptoms" or severity or course, you would not want to use a general sample of the very nonspecific CDC and Kansas definitions	Our justification for this requirement was that there needs to be a gold standard of case definition to evaluate diagnostic accuracy. We recognize that there are limitations to these case definitions, but they are currently what is recommended and widely used. Because we did not find any studies to include for KQ1, this restriction of case definition did not result in the exclusion of any studies. For KQ2, we were very inclusive of diagnostic criteria.
	2) as above, the investigators do not seem to have a coherent conceptual approach to evaluating biomarkers. The conceptual framework (Figure 1) is not really conceptual--it is merely a graphic saying who the populations, interventions, and measures are.	We have clarified that Figure 1 is not a conceptual model, rather a graphic showing our PICOTS and KQs. We have added an additional figure (Figure 2), which provides an overview of the diagnostic test/biomarker development process, which guided our conceptualization of how our KQs and report fit into the biomarker development pipeline.
	The use of "measures of diagnostic accuracy" as an outcome is out of date; diagnostic tests should be evaluated based on a framework that considers technical, diagnostic, and therapeutic impacts, not just "measures of diagnostic accuracy".	We agree that a diagnostic test should be able to both accurately diagnose a condition as well as give insight to potential therapies to use and therapeutic impacts. The latter two, however, are far beyond the discovery phase of how a given biomarker (or group of biomarkers) are associated with the presence or absence Gulf War Illness. While conceptually including therapeutic impacts would be an important property of a biomarker practically, there were no studies that were far enough along the diagnostic test development pipeline to be able to comment on clinical utility.
	Similarly, "association between symptoms and biological measures" is not a valid basis for evaluating a biomarker. (This sentence is also in the background section, p7) What does "association with symptoms" mean? Association with the severity of illness? Association with the types of symptoms? This is conceptually unclear and incomplete.	To date there have only been studies that have taken a biological measurement and assessed its association with GWI or its symptoms. The literature is not at a stage where we can evaluate the diagnostic and clinical utility of biomarkers for GWI. Furthermore, studies were somewhat broad in how they identified GWI and examined associations. For example, some studies may simply categorize as “symptomatic” GWV. For clarity, we have changed “symptoms” to “GWI” in the report.

Because the report isn't clear about the type of evaluation they are interested in, it is also unclear what an ideal study of the question they have in mind would be. Would it evaluate a biomarker to predict a response to treatments or the risk of a complication? Before criticizing the studies that were included (and excluded) it is important to lay out what the target is, and I could not figure out what the target was.	KQ1: Target was accurate case identification between those with GWI and those with some other illness (which no studies existed, so we could not critique them). We specified in the Study Selection which studies we included which identifies which studies we were interested in. KQ2 : Target was biological markers that are associated with GWI case status and should thus be validated or researched further for potential GWI diagnostic test candidates (many of these studies existed so we could critique them on methodological rigor). The parameters of the included studies for KQ2/3 are also described in the Study Selection section. The inclusion criteria lay out specifically what types of studies we were looking for.
The background section, then, should provide a much clearer description of what the authors are looking for in a biomarker--the sentence "...studies of associations between biological measures and GWI status for potential development of biomarker tests.." should end with at least one possible use of a biomarker other than distinguishing GWI from non-GWI. I think a reasonable goal would be to find a marker that was associated with the severity of illness, its course, or suitability for various treatments, but I cannot tell what the authors had in mind.	Severity of illness, illness course, and suitability for various treatments are all important outcomes, but per the agreed upon a priori KQ?s, we were focused on a test's ability to identify the presence or absence of GWI, not to predict its course.
Conceptually, the report should also distinguish between "discovery" studies and validation studies of biomarkers.	We agree, this is an important distinction and have added language and a figure to help us clarify this difference.
1. p12 line 7 "meeting inclusion criteria" line 45 "initial inclusion criteria" : Why "initial" criteria? Were there additional versions of the criteria? It is unclear what "initial" is meant to convey.	Thank you for pointing this out. We have removed "initial" so that the sentence now reads: "Those studies that met inclusion criteria other than including a priority comparator group (72 studies), and/or…"
2. p14 line 11 "We did not identify any studies that met the criteria for inclusion for Key Question 1." This may be the least informative way to convey the results of search and selection! For anyone but systematic reviewers, this sentence would make more sense if it spelled out what you mean--eg, "We did not find any studies that compared a test's classification of GWI to a reference standard and reported measures of diagnostic accuracy." As it stands, the literature flow chart provides no information about which studies were candidates for KQ1 and why they didn't qualify--the KQs are not distinguished until the	Thank you. We spelled out, as suggested, and indicated that no studies addressing the validity of diagnostic tests, regardless of comparator type, were found. We agree that the lack of an agreed upon gold standard makes finding the ideal comparator group difficult. We did identify one prospective case-control study, all others were cross-sectional. We have also updated the literature flow chart to more clearly portray the studies

last step in the flow diagram. The audience for this report needs to understand whether there were studies that aspired to be about diagnosis but did not meet your criteria, and why. A clearer type of flow for KQ1 would itemize the characteristics of these candidate studies so a reader could see, eg, that among studies in the right population, that evaluated an intervention, how many dropped because of a lack of an appropriate comparator or measure (outcomes). Also, it is concerning that the "comparator" is a disputed reference standard. Where there is no adequate gold standard test or diagnostic criteria, a better "comparator" is what happened to the patient over time.	removed for lack of priority comparator or small sample size.
3. The section about quality assessment lacks important details. It would be helpful for the authors to describe the ideal study for each of the key questions. Then explain or justify the choice of instrument they used. For KQ1, the relevance of the Newcastle-Ottawa tool escapes me. If the plan was to evaluate diagnostic accuracy studies, why wouldn't something like QUADAS-2 be appropriate? Also, BIOCROSS is not a quality appraisal ("risk of bias") tool, it assesses the quality of reporting, not of the science or study itself, so it should not be described as " the quality appraisal tool for cross-sectional studies using biological..." Overall I could not make out how the tool (Newcastle-Ottawa+items from BIOCROSS) could be used to assess the quality of either diagnostic accuracy studies or cross sectional studies of biomarkers. Considering that the end of the report summarizes limitations of the studies, there seems to be a mismatch between the instruments you used to assess the studies and the problems you found with them. The mismatch might be because you included studies that might be described as "discovery" studies but assessed "risk of bias" as if they were clinical studies. Regarding the ideal study, the material on pp54ff describing the problems of the literature would be much stronger if, up front, you described a study that would be strong.	The QUADAS-2 tool is used for diagnostic tests, though none of our studies actually examined a diagnostic test, so the QUADAS-2 would be largely irrelevant. We added to the Quality Assessment section indicating that had we found studies of validity of diagnostic tests, we would have used the QUADAS-2. We modified the language associated with BIOCROSS, as suggested. Most of the studies that were identified were cross-sectional or case-control studies and looked for associations between GWI case status and a specified set of biomarkers. For this reason, we believe the Newcastle-Ottawa items were the most applicable. We did note some limitations with the Newcastle-Ottawa tool, which is why we used this descriptive approach rather than a definitive rating of ROB. We agree with your comment about describing the ideal study and have added a description of what an ideal diagnostic test study would like.
4. The report doesn't give me confidence that what was excluded was not of interest and what was included was of interest; that is, the authors need to show why applying these PICOTS does not exclude material of interest. It is by no means obvious that these PICOTS make sense. Why would studies that don't	For KQ1, we did not find any studies testing validity of diagnostic tests, regardless of the case definition used for the comparator group, so we did not miss any studies by using these criteria. We have added language in the report to describe this as well.

relate to CDC or Kansas be excluded? Wouldn't that close off research that could demonstrate there is a better definition? Those instruments are said to be the "best" for what the IOM was interested in, but case definition, research definition, and other criteria might be best for studying biomarkers. The approach of "included" studies vs "excluded" studies also doesn't serve the purpose of the review very well, at least without more detail about what was excluded. A landscape of the 270 potentially relevant studies could be useful--make a table of how many of these evaluated each biomarker (similar to Figure 3, but for the 270 studies). In a review intended to inform a state of the science conference, it is important to describe what has been studied. You might show, eg, that there were 30 studies of energy metabolism, only 2 of which were included.	For KQ2/3, we have expanded upon the table of 72 studies that were not included in the body of the report either due to n<25, a non-ideal comparator group (i.e., comparator groups other than deployed GWV without GWI and with or without other health conditions. We have added to the table heading this description of what was included.
5. p54 "To establish a biological metric capable of making this distinction would require biological measures to be compared between cases versus individuals without GWI and with other health conditions with overlapping symptomology with GWI. The ability of a biologic measure to distinguish GWI when comparing patients with symptoms to healthy patients without symptoms may not translate to its ability to distinguish GWI from another illness in patients presenting with symptoms (which is more typically the context in which a diagnostic test would be used)." These sentences are confusing. You have 2 goals here--one is to explain when healthy controls are not appropriate, and the other is what to do instead. Again, this section should start with your view of what a good study would look like, then contrast what you found with it. This must be done because putting out there what a good study would look like will establish conceptually what you are measuring the actual studies against. Do healthy controls have any role in evaluation at all? I would say they do--as an early test, a discovery test, it could be useful to see which markers differ from sick and well people. Next, you would want to do a different kind of study, perhaps still retrospective, with comparisons to other illnesses (as you say). Then, if a biomarker passes these phases, the best design is prospective and in a prospective study one doesn't pick cases and controls at all--one identifies a cohort of patients in whom GWI is suspected, and then applies the marker, and then follows up to see who is	We have added to the description of the ideal study. We also added a description of the potential utility of the table of other studies.

	actually diagnosed with GWI (preferably without knowledge of the biomarker result). The people who are not diagnosed with GWI may be diagnosed with something else or may be undiagnosed. So these sentences are really only about studies that pick cases and controls, and they imply that instead of healthy controls, investigators should pick people with (known) other illnesses. That isn't really always the case--it is only a step in the early evaluation of a biomarker.	
	Minor comments 1. p24 lines 13-17. Do phospholipids come in "species"? This may be the right term but it is new to me.	Thank you. Yes, it seems that different phospholipids can be referred to as “species”. This was the terminology used in the study.
	2. The actual writeups of studies on pages 23-24 and pp29-30 is quite good, but some studies, particularly ref 28 and 42, merit a more detailed critique in the text.	For reference 28 and 42 we have added additional detail.
	3. p24 line 42 "reporting' should be 'report" I think	Thank you. We changed “reporting” to “reported” where indicated.
4	1. I'm concerned you threw some of the baby out with the bath water by excluding the 72 studies with non-priority comparator groups. Is there nothing that can be learned by including those studies in this review, perhaps in a separate category and then triangulating the findings with findings from the studies with better comparator groups? This is especially important in the field of GWI research given the paucity of data, the frustration with the lack of progress in its understanding, and the amount of resources expended.	We have expanded on the table of 76 studies to include study findings and indication of whether or not there were statistically significant findings related to associations between GWI and biological measures.
	2. Almost all included studies were cited for not providing adequate power calculations. When differences were reported between groups (KQ2), was there not, empirically speaking, adequate power to detect a difference? When there was no difference, I understand how a power calculation is critical in assessing whether the study contributes to our understanding. Also, is it sometimes possible to calculate the power from the results and methods reported in the publication? If so, did the review team do this?	We request from future studies certain information that would increase the consumer’s ability to determine level of confidence in the findings. Our conclusions were not greatly influenced by lack of methodological information like this. More heavily weighted factors were the great heterogeneity in biological measures and the comparator group.
	3. I concur that subgroup analysis is a vital strategy to better understanding the diverse symptoms afflicting Gulf War Veterans with Gulf War Illness. I do not think the Haley subsyndromes should be promoted as a standard approach for doing this, however. The subsyndromes were developed on a small cohort and have not been replicated. I do not recall exactly at the moment, but I believe the	Thank you. We changed the language to recommend a stratification similar to the Haley categorizations, without recommending Haley categorizations specifically.

sample size was so small that even if randomly selected from the population (which they may not have been), they are likely not representative of the population. The 2014 VA/DoD Clinical Practice Guideline for the Management of Chronic Multisymptom Illness used the labels fatigue-, GI- and pain-predominant CMI which correspond to CFS, IBS, and FM. This labelling has been abandoned in the current draft of the 2020 update to that CPG, but still has clinical relevance. I'm not suggesting this approach, but merely highlight that the subgroups of GWI are also far from settled.	
4. I appreciate the discussion (pp. 54-55) of the need for a comparison group with similar symptoms to GWI, but have some questions. Many such potential comparison groups have biomarkers that would differentiate them from GWI. Why then would a biomarker for GWI established compared to a healthy group, not be of value if it is different from the biomarker for a condition with similar symptoms? For example, if we were to select multiple sclerosis (MS) as the appropriate symptomatic comparison group for GWI, there are already biomarkers that differentiate GWI from MS. If we found a marker for GWI compared to healthy comparators that is different from the markers for MS, we wouldn't confuse GWI and MS with that biomarker any more than we currently do. It would be a remarkable advance in the diagnosis and care of Veterans with GWI. Also, what other conditions would be suitable comparison groups for GWI in general? Using this approach, one would likely only find a biomarker for pain, or fatigue, or cognitive deficits or whatever the symptom of focus is in GWI, not for the constellation of chronic multiple symptoms. Do your findings lead you to conclude that is the best approach?	We agree that a study to examine a healthy control compared to GWI could shed some light on biomarkers that could be researched further for a diagnostic biomarker for GWI. Thus, we have expanded upon the Appendix Table D to provide more information about studies including healthy controls, including findings. Still, we prioritized studies with comparators that would give us the most useful information about a biomarker as a GWI diagnostic tool. Specifically, an ideal biomarker would enable us to differentiate GWI from another illness.
5. On p. 55 there is a discussion of the general lack of information in the included studies about the distribution of the data and outliers and how they were handled. I assume this is a matter of degree, but most of these were peer reviewed; are you holding these studies to too high a standard? Or are you extracting information from studies that focused on reporting other findings and therefore these were not held to a high enough standard? In other fields/conditions, are comparable studies reported in a manner more consistent with the standards you applied?	We hope that our synthesis provides some guidance for how biological marker research might be more transparent in their reporting, that would increase the consumer's ability to determine level of confidence in the findings. Our conclusions were not greatly influenced by lack of methodological information such as the handling of outliers. More heavily weighted factors were the great heterogeneity in biological measures and the comparator groups included in the studies

	6. Finally, are there NO specific areas that seem promising for differentiating, even at a group level, Veterans with GWI from Veterans without GWI after this review? If you had to pick one or two where we should invest resources, which would they be?	The emphasis of our review was to map out what biomarkers have been studied. Had there been strong enough evidence in any one direction, we would highlight that. We attempted to synthesize the larger biological systems in which the majority of the extant research had focused,, but we were unable to identify any specific biomarkers with sufficient strength of evidence.
5	Minor correction on p.51 row 37, NR should be Steele, L.	Thank you. We made the suggested adjustment.
	Minor correction on p.56 row 57, The US Army Medical Research and Material Command should read The US Army Medical Research and Development Command.	Thank you. We made the suggested adjustment.
	General Comments: Included studies were limited to those with 1) a comparator population of deployed healthy or deployed with health conditions other than GWI and 2) greater than 25 participants. The following are concerns with this limitation:	
	1) Because there is currently no objective, evidenced-based case definition of GWI, selection of the reported “ideal” comparator group (GWV without GWI and with a condition with overlapping symptoms) is problematic.	We agree the lack of a gold standard is problematic and has both hindered the field’s ability to develop and validate a biological test for GWI and hindered our ability to comment on the ability of such a biomarker to distinguish those with GWI from those without GWI (KQ1). We were very inclusive of case definitions of GWI for KQ2 and 3. Further, we have expanded upon the table of excluded studies to provide additional information such as findings from studies that included a comparator besides a healthy, deployed GWV.
	2) Participants selected from the same battalion was considered a limitation; however, given reported exposure differences depending on deployment location, branch of service, etc. subgrouping may in fact be a reasonable approach to biomarker research rather than a one size fits all approach to this multi-symptom illness. It is likely there will not be a single diagnostic criteria or tool. Future consideration of further GW Veteran subgrouping, including by molecular characteristics, may facilitate biomarker discovery and would allow for use of smaller sample sizes.	Thank you. We agree that inclusion of individuals from the same battalion, some who developed GWI and sone who did not, could provide important and nuanced insights about how GWI develops, etc. Our rationale for calling this a limitation, was in the context of considering how the findings of one study might be applicable or generalizable to a larger group (in evidence synthesis, this is known as applicability). We do acknowledge in the discussion, the importance of subgrouping this complex and heterogenous illness.
	3) Many clinical biomarker studies in the GWI field are exploratory, as the underlying pathobiology of the illness is still being discovered (primarily in preclinical systems). An evidence-based framework is necessary	We have expanded upon the table of excluded studies to include findings and indication of statistical significance in the association between biological measures and GWI.

	prior to pursuing larger scale clinical validation. Therefore, a goal of many early clinical biomarker investigations is refinement of existing hypotheses rather than testing validity. These investigations provide a platform to generate preliminary data and give direction to future investigations. It would be worthwhile to pursue an evaluation of excluded studies that did not meet the comparator and group size threshold, but that show promise for replication and future validation. This would allow prioritization of the most promising pathways/potential diagnostics going forward.	
	4) The GW Veteran population is limited with respect to recruitment. Unlike disease fields where there are new cases each year, the deployed 1990-1991 Gulf War population is relatively small and difficult to recruit. Compounding effects of aging in this population create additional obstacles. These confounding variables should be taken into consideration when determining an appropriate sample size, particularly for exploratory and pilot translational studies. Again, appropriate subgrouping (and potentially smaller group sizes) may be the most reasonable approach.	We have expanded upon the table of excluded studies to include findings and indication of statistical significance in the association between biological measures and GWI.
	Observation - The lack of outcome assessor blinding may reflect financial shortfalls in technical expertise, database management, and biostatistical assistance. This is an important consideration for GWI research funders.	Thank you for this insight. We have added a comment about this in our discussion.
	The strategic and specific funding mechanism pipeline implemented by the DoD CDMRP GWIRP in FY19 will aid translation of research in this area. A description of this strategy could be considered for the Future Research section. FY20 and beyond, the GWIRP will be continuing this funding pipeline composed of 1) a discovery stage representing innovative biomarker research that is in the earliest stages of development; 2) a qualification stage representing research already supported by preliminary or published data in the GWI field that is ready for validation through expansion, replication, or comparative studies; 3) a verification stage representing clinical translation (testing in a GW Veteran population) of concepts previously replicated and validated; and 4) a confirmation stage representing large-scale confirmatory and pivotal trials that will transform and revolutionize the clinical management of GWI. Objective biomarkers to measure the biological effect of an intervention or predictive/cohort-	Thank you for notifying us of this strategic funding pipeline. We have added this in the future research section of the discussion.

	selective biomarkers are required in the conformation stage. This promotes biomarker and diagnostic assay development and validation simultaneously with testing of new treatments instead of as separate steps in the development process.	
6	This report is a comprehensive review of studies of potential biomarkers of the Gulf War Illness, reviewing studies of GWI patients compared to deployed veterans without GWI. The report is very well written, easy to follow, and detailed and concise enough. Intro and methods look good to me. I just have a few suggestions re results and discussion:	Thank you.
	Results: In table 2, all the fMRI studies measure the same thing: BOLD signal during a task, which is an indicator of changes in the activation in different areas of the brain in response to a task. So for avoiding confusion, “The biological measures examined” should be the same for all those, and can be “brain activation.”	Thank you. We have modified so that the data collection column for all fMRI studies refers to ‘brain activation’.
	Unsure why Zhou et al, examining pain is classified under “ANS”. There are several different mechanisms involved in pain tolerance, and ANS does not seem to be the major one. This work can come under the “Other Biological Systems”	Thank you. We moved the Zhou et al study from the ANS section to the Other Biological Systems section, as suggested.
	Weiner 2011, An spectroscopy study should be under “CNS” category, and not genes.	We have moved Weiner 2011 from the genetic to the CNS category, as suggested.
	Nagelkirk, 2003, could come under ANS.	We have moved Nagelkirk, 2003 into the ANS category.
	Results: It might be helpful to add a brief paragraph in the beginning of each section (or in the intro) about why researching each of these systems (immune, ANS, CNS, etc) sounded reasonable for this illness.	We agree that it is important to put into context the involved biological systems, we included a sentence and some additional language in the introduction about the rationale and hypotheses about the involvement of each of these systems in GWI.
	Similarly, it might be helpful to add a couple sentences about each (or some of the less commonly known by general readers) measure addressed here. For instance: what is squalene antibody? Or what is the function of the candidate genes in genetic studies, and why were they selected?	Due to the extensive heterogeneity of studies, this level of information was not feasible to include.
	Discussion: I understand the studied biomarkers are all over the place and the findings are inconsistent, but it will be helpful to address the consistencies in findings of these limited studies. For instance, HRV and ANS seem	As you point out, there were very few studies of any one biological measure. Because of that, the review is meant to be a map of what has been studied.

	more consistent, although very limited number of studies.	
	Since the ongoing studies are mentioned in this report, it might be helpful to discuss how those studies might be informed by, and following (or not following) trajectory of the published studies. Is there a direction where the research seem to have been following? And what are the new areas which have not been addressed before, and why are those selected? To me there seems to be a heavier focus on neuroinflammation, plus addition of mitochondrial and gut microbiome studies. I think the above is important as an aim of this report is to help inform the future research.	We agree and have tried to synthesize this by listing the respective frequency of extant studies for each of the biological systems, to provide insight to what has and has not been studied thus far.
	I think one of the inherent limitations of such studies, and challenges of this field which might be worth mentioning is the time x illness interaction. The chronic illness over many years since the Persian Gulf War might have led to differences in health behaviors and lifestyle (e.g. chronically reduced activity and exercise due to fatigue, medications side effects, etc.) among GWI veterans compared to healthy controls. These differences could lead to some of the current or future detected differences (e.g. cardiovascular) which could not have been a part of the illness itself, but a consequence. In that sense, it will be important for future studies to consider such confounding variables (BMI, level of activity, comorbid psychiatric and cardiovascular disorder, medications) when researching differences between GWI and control groups.	Thank you. This is a good suggestion; we have added a sentence about this in the future research section. In addition, one of our quality assessment rating questions examined whether important confounders such as the presence of other chronic illnesses were considered during sample selection. For the immune section, there were many exclusions of those who might have chronic health conditions that would muddy the interpretation of the immunological measures included in the study. On the other hand, exclusion of those with comorbid chronic illness might become increasingly challenging as this populations ages and may miss an important part of the diagnostic test utility, which is to differentiate GWI from other conditions.