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.