

Genetic Analysis of Gulf War illness: Phenotype Development, GWAS, and Gene-
Environment Interaction

by

Jacqueline M. Vahey

Computational Biology and Bioinformatics Program
Duke University

Date: _____

Approved:

Elizabeth Hauser, Supervisor

Allison Ashley-Koch

Jichun Xie

Li Ma

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Graduate Program in
Computational Biology and Bioinformatics in the
Graduate School of Duke University

2022

ABSTRACT

Genetic Analysis of Gulf War illness: Phenotype Development, GWAS, and Gene-
Environment Interaction

by

Jacqueline M. Vahey

Program in Computational Biology and Bioinformatics
Duke University

Date: _____

Approved:

Elizabeth Hauser, Supervisor

Allison Ashley-Koch

Jichun Xie

Li Ma

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Graduate Program in
Computational Biology and Bioinformatics in the
Graduate School of Duke University

2022

Copyright by
Jacqueline Vahey
2022

Abstract

Veterans who served in the 1990-1991 Gulf War experience debilitating chronic symptoms at extremely high rates. In the 30 years since the Gulf War, many researchers have worked to identify the cause and biological pathway of Gulf War illness (GWI). There is, however, no biomarker, ICD code, or other standardized way to identify veterans with GWI; veterans are told they have GWI based on a clinician's assessment of their unexplained chronic symptoms. There is also little agreement on the causes and potential biological pathways of GWI. This dissertation describes phenotyping efforts, the first genome-wide association study (GWAS) of GWI, and a candidate gene-environment interaction study. First, I describe methods for developing well-documented indicators for complex phenotypes, which have generated GWI indicators that are used for the MVP and GWECB datasets. This is the only tested and published algorithm for defining GWI. This work required extensive exploratory analysis and data cleaning, as it was the first major analysis of the GWECB dataset. The variables generated through both the data cleaning and GWI algorithm have been incorporated into the GWECB. Then, I performed the first GWAS of GWI, which supports prior work in the field and suggests further candidate analyses. Top gene-set associations include response to cadmium ion, regulation of response to interferon gamma, and regulation of autophagosome maturation. Among other top associations, these results indicate association with a neuroimmune response to exposure. GWAS summary statistics will

be made available. Finally, I developed a hypothesis-driven candidate gene-environment interaction study, which replicated a previously published statistically significant association of *rs662*/PB pill exposure with GWI. Future research building off my contributions could help identify the underlying biological pathways and causes of GWI, allowing better treatment of the underlying disease for hundreds of thousands of Gulf War Veterans.

Contents

Abstract.....	vi
List of Tables	xi
List of Appendix Tables	xii
List of Figures	xiii
List of Appendix Figures.....	xiv
Acknowledgements	xvi
1. Introduction	1
Gulf War illness	1
GWI case definitions.....	2
Chronic conditions associated with Gulf War service.....	5
Gulf War experiences and exposures.....	7
Epidemiologic associations of military exposures and GWI.....	10
Available datasets	12
Combining genetics with exposures in GWECB	13
Summary	14
2. Phenotype Identification	16
Research tool for classifying Gulf War Illness using survey responses: Lessons for writing replicable algorithms for symptom-based conditions	18
Introduction	18
Materials and Methods.....	20
Results.....	26

Discussion	32
Conclusions.....	34
Application of GWI Algorithm to CSP585 and CSP2006.....	35
Introduction	36
Results.....	37
Discussion	44
Conclusions.....	48
3. Genome-wide Investigation	49
Genetics of Gulf War illness, a genome-wide association study.....	50
Introduction	50
Materials and Methods.....	52
Results.....	58
Discussion	75
Conclusions.....	80
4. Biologically based Candidate Studies	81
Gene-toxicant Interactions in Gulf War Illness: Differential Effects of the PON1 genotype.....	82
Introduction	82
Materials and Methods.....	85
Results.....	92
Discussion	103
Conclusions.....	107
5. Conclusions.....	108

Overview.....	108
Phenotyping efforts.....	109
Genomic analysis.....	110
Interaction models.....	114
Expanding the genetic model.....	115
Final thoughts.....	117
Appendix A: Phenotype Identification.....	118
Appendix B: Genome-wide investigation.....	127
Appendix C: Biologically based candidate studies.....	128
Appendix D: GWI indicator pseudocode.....	138
Appendix E: GWI indicator SAS code.....	179
<i>Works Cited</i>	223
Biography.....	239

List of Tables

Table 1. Sample characteristics by 1990-1991 Gulf War deployment status.....	37
Table 2. GWI Case Status and Related Components by Deployment Status	41
Table 3. Demographics by CDC Severe GWI status	58
Table 4. Case definitions by deployment.....	60
Table 5. Associations with GWI: top SNPs.....	63
Table 6. Gene-set associations with GWI.....	74
Table 7. Candidate SNPs	91
Table 8. Demographics in deployed veterans, by CDC Severe GWI status	92
Table 9. Exposure Time by CDC Severe GWI Case Status.....	95
Table 10. Tier 1 Candidate SNP results	97
Table 11. Top Results for the Tier 2 SNP/PB and SNP/Pesticide Interactions.....	101

List of Appendix Tables

Appendix Table 1. Kansas GWI case definition symptoms by Gulf War deployment status	120
Appendix Table 2. CDC GWI case definition symptoms by Gulf War deployment status	123
Appendix Table 3. Frequency of veteran-reported exclusionary conditions by 1990–1991 Gulf War deployment status.	125
Appendix Table 4. Demographics by Pesticide Exposure	128
Appendix Table 5. Demographics by PB exposure.	130
Appendix Table 6. Frequency of fulfilling symptom domains and case definitions by exposure to pesticides.....	132
Appendix Table 7. Frequency of fulfilling symptom domains and case definitions by exposure to PB pills.....	133
Appendix Table 8. Symptoms Used to Construct the Kansas and CDC Domains	140
Appendix Table 9. List of Kansas Exclusionary Criteria.....	145
Appendix Table 10. Kansas GWI Symptom Criteria by Symptom Domain Indicators..	154
Appendix Table 11. Does Veteran meet Kansas GWI case definition?	157
Appendix Table 12. CDC GWI Case Status.....	160
Appendix Table 13. Derived Variable List	162
Appendix Table 14. SAS programs to generate case definition indicators.....	173

List of Figures

Figure 1. Kansas GWI Flowchart	26
Figure 2. CDC GWI Flowchart	27
Figure 3. Components of Kansas GWI.....	27
Figure 4. Illustration of Test Cases for Kansas GWI Case Definition Classification Algorithm	29
Figure 5. Association of GWI symptoms with deployment.....	43
Figure 6. MVP case definition graphic.....	44
Figure 7. CDC Severe GWI Manhattan plot and QQ-plot.....	62
Figure 8. LocusZoom rs9937803.....	70
Figure 9. LocusZoom rs993521	71
Figure 10. LocusZoom rs35084682.....	72
Figure 11. Gene-based association: CDC Severe, all	73
Figure 12. Block diagram: individuals in the analytic dataset.....	89
Figure 13. Proportion of individuals fulfilling CDC Severe GWI within each genotype-exposure group.....	99
Figure 14. LocusZoom plots for CDC Severe GWI, Tier 2	100
Figure 15. Kansas Respiratory Domain: PON1 LocusZoom plot.....	103

List of Appendix Figures

Appendix Figure 1. CDC and CDC Severe GWI definition.....	118
Appendix Figure 2. GWI Kansas Case Definition. Left/top: Symptoms in the Symptom Criteria. Right/bottom: Conditions in the Exclusionary Criterion.....	119
Appendix Figure 3. PCA of genetic ancestry in GWECB.....	127
Appendix Figure 4. Estimated heritability of CDC Severe GWI.....	134
Appendix Figure 5. Genotype proportions by exposure time (PB).....	134
Appendix Figure 6. CDC Severe GWI frequency: white, non-Hispanic.....	135
Appendix Figure 7. CDC Severe GWI frequency: Hispanic or Black, non-Hispanic.....	135
Appendix Figure 8. LocusZoom plots for Tier 2 SNPs in interaction with pesticides....	136
Appendix Figure 9. Exposure questions from the CSP585 questionnaire	137
Appendix Figure 10. Construction of Revised Symptom and Symptom Rating Variables	144
Appendix Figure 11. Creating Revised Diagnosis Variables.....	147
Appendix Figure 12. Kansas Symptom Domain Indicators	149
Appendix Figure 13. Kansas Symptom Criteria Indicator, Part 1.....	150
Appendix Figure 14. Kansas Symptom Criteria Indicator, example.....	151
Appendix Figure 15. Kansas Symptom Criteria Indicator, Part 2.....	153
Appendix Figure 16. Kansas GWI Case Definition.....	156
Appendix Figure 17. CDC and CDC Severe Symptom Domain Criteria	159
Appendix Figure 18. CDC and CDC Severe Case Definition Indicators	161
Appendix Figure 19. Pseudocode: symptom and diagnosis cleaning.....	175
Appendix Figure 20. Pseudocode: Kansas GWI.....	176

Appendix Figure 21. Pseudocode: CDC GWI..... 177

Appendix Figure 22. Tree diagram of phenotypes 178

Acknowledgements

The views expressed here are mine, and mine alone. I am grateful for the support of the T32 Training Grant 5T32-GM071340, the Duke Computational Biology and Bioinformatics program, and the Duke University Department of Biostatistics and Bioinformatics, who have financially supported my research. I could not have asked for a more supportive research advisor than Dr. Elizabeth Hauser, whose guidance and support have been invaluable. I thank my committee members, Dr. Allison Ashley-Koch, Dr. Li Ma, and Dr. Jichun Xie, for their patience and guidance over the past four years. I am also grateful for the mentorship and support of the Durham Veterans Affairs Cooperative Studies Program Epidemiology Center (CSPEC-Durham); CSPEC-Durham welcomed me onto the team without hesitation. I would also like to acknowledge Dr. Dawn Provenzale, who was a valued committee member until her passing in 2021. Her mentorship is greatly missed. I also thank the veterans who made this research possible through their participation in the GWECB and in the MVP. Finally, I thank everyone who has made this journey possible, especially my family, who have always supported and encouraged my dreams, and my friends, who have been with me through all the ups and downs. I am especially grateful for my parents, Lisa and Walter Vahey; my sister, Lauren Vahey; my boyfriend, Tim Manganello; and my family of the heart, Siobhan, Fran, and Alma Oca.

1. Introduction

Gulf War illness

Gulf War Illness (GWI) is a debilitating disease experienced by 20-40% of veterans who served in the 1990-1991 Gulf War¹⁻⁴. Beginning shortly after their return stateside and continuing to present day, veterans suffer chronic symptoms such as fatigue, diarrhea, reduced cognitive abilities, irregular moods, and respiratory challenges, at very high rate⁵⁻¹². GWI is significantly associated with deployment to the Gulf during the 1990-1991 Gulf War in every major study^{3,4,13,14}. The heterogeneity of the symptoms reported has made identifying a single consensus definition difficult, let alone identifying any potential genetic susceptibility locus. GWI is defined by an accumulation of chronic symptoms and has no definitive diagnosis as represented by an ICD9 or ICD10 code, making it more difficult to study³. Symptoms-based case definitions built from survey results are the state of the art for research². While military exposures from the Gulf War, such as pyridostigmine bromide (PB, anti-nerve gas pills), pesticides, and burn pits, have been associated with GWI, no single association has proven to be causal in human studies⁴. Many observational studies have been completed in military cohorts, resulting in some common associations but no causal indications, biomarker, or diagnostic tool has been identified^{4,15-17}. Several mouse and rat models have been developed using a combination of these exposures and stress¹⁸. Though many studies have been completed and many manuscripts have been published

over 30 years of research, the field lacks a common consensus on the biological underpinnings of GWI.

GWI case definitions

Gulf War illness has several commonly used research case definitions; there is no uniformly accepted case definition². The two most common are the Kansas⁶ and Centers for Disease Control and Prevention (CDC)⁵ case definitions. The CDC case definition was first described by Fukuda et al, in 1998, and was initially referred to as chronic multisymptom illness (CMI). CMI was initially defined as the endorsement of at least one symptom in at least two of three symptom domains: musculoskeletal, fatigue, and mood/cognitive⁵. Kansas GWI was initially defined as the endorsement of at least one moderate or severe symptom or at least two mild, moderate, or severe symptoms in at least three of the six symptom domains, without endorsing any of the specific exclusionary conditions⁶. Each case definition has a different prevalence, and many individuals would be categorized as having CDC GWI but not Kansas GWI. There are several other case definitions that have been developed and used over the past 30 years, but the Research Advisory Committee for Gulf War research has advised to (1) stop creating new case definitions, and (2) use the CDC GWI and Kansas GWI case definitions². This is due to their consistent positive association with deployment over the past 25-30 years, as well as their use in most Gulf War research.

The CDC GWI case definition, or CMI, as it is called in early literature, is a very broad case definition. Individuals are required to report at least one symptom, at any severity, in two of three symptom domains. The three symptom domains are: fatigue, musculoskeletal, and mood–cognition (see Appendix Figure 1)⁵. Some cohorts report extremely high prevalence of CDC GWI, even in non-deployed individuals; GWECB reports that 80% of non-deployed veterans in the cohort fulfill the CDC GWI case definition^{3,4,19}. This makes research very difficult because the controls are so few. In addition, in current research studies the cases are diluted by those who may be displaying symptoms of aging or of other illnesses. The CDC Severe case definition has the same symptoms arranged into the same symptom domains but requires at least one severe symptom in at least two of the three symptom domains⁵. This case definition has a better balance of cases and controls in modern cohorts, as well as being more highly associated with deployment to the Gulf during the 1990-1991 Gulf War, making it a more specific case definition²⁰. Neither the CDC GWI nor the CDC Severe GWI case definitions consider clinical conditions or age.

Kansas GWI contains both inclusionary and exclusionary criteria: one must fulfill the inclusionary symptom criteria while not fulfilling the exclusionary clinical condition criteria⁶. The exclusionary conditions are somewhat confusingly named, as they are more conditions that would preclude one from being categorized as having GWI. At the time of the development of the Kansas GWI case definition, scientists were working to

prove that GWI was in fact a serious condition, separate from other illnesses. Therefore, those who report clinical conditions that could explain their symptoms are categorized as non-cases⁶. This process is as much an art as a science, requiring deep understanding on a clinical level of both Gulf War illness and other clinical conditions²¹⁻²³. This was designed for extreme specificity among small research cohorts shortly after the Gulf War. Genetic research, however, requires large cohorts for sufficient statistical power and requires a more automated method for identifying Gulf War cases²⁴. The symptom domains for the Kansas GWI case definition are: fatigue and sleep problems; pain; neurologic, cognitive, and mood symptoms; gastrointestinal; respiratory; and skin (see Appendix Figure 1) ⁶. Conditions that might preclude a veteran from fulfilling the Kansas GWI case definition include cancer, diabetes, heart disease, stroke, liver disease, lupus, multiple sclerosis, HIV, tuberculosis, hepatitis C, and mental health disorders that may cause similar symptoms or cause a clinician to question the reported symptoms (e.g. schizophrenia or bipolar disorder) (see Appendix Figure 2) ⁶. It is notable that while some papers publish exactly which symptoms fulfill their version of this case definition, GWI experts still disagree on what exactly each category includes.

Both recommended case definitions depend on survey questions regarding chronic symptoms and the severity of these symptoms, and the Kansas case definition additionally relies on self-reported chronic conditions. Despite this reliance on survey results, there is no validated survey instrument with which to collect these data. As

each research group thinks through how the case definitions could be improved and how the survey questions could be better asked, the questions asked vary slightly. While the design of each new survey does reflect on prior surveys, there is no standard questionnaire or form. Each questionnaire is slightly different: the chronic symptom questions are worded slightly differently, or different symptoms are included. To add to this, there is no published, tested, and validated algorithm for identifying cases for any of the case definitions. To best research GWI, a clear, concise algorithm must be written to operationalize these survey data into a GWI research case definition. This ideal case definition must be agreed upon by researchers within the Department of Veterans Affairs, as consistency across studies allows for replication and comparison of results. Without this consistency, we are unable to compare new and old results, and are unable to determine if prior reported results can be replicated.

Chronic conditions associated with Gulf War service

Several conditions have been previously associated with deployment to the Gulf in 1990-1991. Association with deployment or with deployment-related exposure can lead to bureaucratic decisions that make a difference in the lives of veterans; for example, amyotrophic lateral sclerosis (ALS) has been designated as service connected for Gulf War veterans. Associations between deployment and various conditions have been characterized by the Institute of Medicine as “sufficient evidence of a causal relationship” (PTSD only), “sufficient evidence of an association” (generalized anxiety

disorder, depression, and substance abuse; gastrointestinal symptoms consistent with irritable bowel syndrome and functional dyspepsia, chronic fatigue syndrome, Gulf War illness) and “limited/suggestive evidence of an association” (ALS; fibromyalgia and chronic widespread pain; and self-reported sexual difficulties)⁴. These designations indicate that chronic conditions remain associated with deployment, but do not suggest any biomarkers or biological tests for GWI or any causal pathways for any of these associated diseases. More recently, the United States Department of Veterans Affairs funded an evidence synthesis project *Biological Measures and Diagnostic Tools for Gulf War Illness: A Systemic Review*, which was completed in 2020¹⁵. This report investigated and reported on studies that searched for biological tests to determine GWI case status, including 32 completed studies, 24 ongoing or unpublished studies, and 77 supplemental studies. None of the studies showed any validated biological tests to assess GWI case status. Moreover, biological measures and GWI case definition varied greatly across studies even though many studies show a significant association between a biological measure and GWI. Biological measures included human leukocyte antigen (HLA) alleles, platelet count, monocyte count, squalene antibody status, phospholipid profiles, interferon-gamma plasma concentrations, and brain activation. That GWI and biological measurements continue to be associated, even with varying case definitions and a wide variety of biological measurements, strongly indicates a biological root of GWI.

Gulf War experiences and exposures

Veterans who served in the Gulf War experienced a wide variety of toxic exposures and highly stressful experiences. Troops lived in tents for most of their time deployed to the Gulf, with very little privacy². There were many pests, including snakes, scorpions, and desert flies, leading to not only bites and potential infections, but also widespread pesticide use². Troops were not provided clean uniforms or hot showers very frequently, potentially compounding the toxic effects of pesticide use². Deployment was rapid and there were very few opportunities for relaxation; alcoholic beverages were banned out of respect for local tradition and troops were ordered not to befriend locals². The food provided was packaged and troops drank bottled water; after early contamination of fruits and vegetables, the military did not utilize local produce. The weather was very hot in August and September, when most of the conflict occurred, with sand temperatures reaching 150 degrees Fahrenheit. Wind was strong in both the summer and winter months and troops wore goggles to protect their eyes from the sand. Due to the rapid deployment of troops and the larger than ever portion of National Guard and Reserve troops, many soldiers were unsure that their job would be there for them when they returned stateside. Gulf War illness and post-traumatic stress disorder (PTSD) are correlated and appear to interact, but there is little evidence supporting causation or directionality of this correlation^{10,25,26}.

In addition to the stressful environment, toxic exposures were also common among troops. Oil well fires, burning of trash, sand suppressing petroleum products, and unvented tent heaters, cooking stoves, and generators all caused strong smoke exposure, not all of which is formally recorded. Troops also used pesticides, including dog flea collars, methyl carbamates, organophosphates, pyrethroids, and chlorinated hydrocarbons, in an attempt to combat local insects and rodents³. Sarin attacks, which cause unpreventable seizures and death, were anticipated, and greatly feared. To protect against this nerve gas, troops were given packs of pyridostigmine bromide (PB) pills to temporarily block sarin, with instructions to take pills when ordered or when the chemical attack sensors alarmed. Unfortunately, the chemical attack sensors were very sensitive, alarming frequently in the presence of organic solvents, vehicle exhaust fumes, and insecticides^{2,3}. While the Department of Defense reported that over 5 million doses of PB were fielded and estimated that about 250,000 individuals took at least one dose of PB during the war, pills were self-administered upon orders from the unit commander¹⁴. This lack of documentation forces researchers to depend on self-reported survey results for exposure data¹⁴. In addition to the biological effects of these exposures, one might also consider the psychological effects of the constant state of alarm these troops experienced¹⁰. The sheer number of potential exposures complicates analysis, as does the fact that the Gulf War was over 30 years ago, potentially contributing to recall bias.

Prior work in occupational hazards and in Gulf War research has shown that organophosphates and carbamates, such as pyridostigmine bromide (PB) pills, sarin, and the pesticides used in the Gulf War, bind to acetylcholinesterase, the enzyme that breaks down acetylcholine²⁷⁻³⁰. This binding sterically blocks the acetylcholinesterase active site—in some cases reversibly and in some cases irreversibly^{27,31,32}. Nerve cells release acetylcholine to pass a signal from one cell to the next; the cell receiving the signal detects the presence of acetylcholine through various receptors, stimulating muscle contraction³³. If insufficient free acetylcholinesterase remains, the ensuing accumulation of acetylcholine overstimulates the neuromuscular system, causing systemic long term damage and, in extreme cases, death^{32,34,35}. While irreversible inhibition of acetylcholinesterase is often fatal, even reversible inhibition of acetylcholinesterase temporarily causes this overstimulation³⁶. Sarin is an irreversible inhibitor of acetylcholinesterase, while PB pills are reversible inhibitors, meant to preemptively defend soldiers from sarin attacks. The symptoms of long term damage due to organophosphate poisoning reported in occupational hazards research are similar to the symptoms of GWI: headaches; pain in muscles, joints, or bones; rashes or itchy skin; blurred vision; dizziness; diarrhea; confusion; respiratory and circulatory depression^{34,37}. In addition, rodent models for GWI have been developed using organophosphate pesticides, pyridostigmine bromide (PB), and chronic stress or corticosterone³⁸⁻⁴¹. Coadministration of PB and pesticides, along with stress, led to

delayed cognitive impairment in the rodent models; affected mice had decreased spatial memory and increased depressive behavior⁴². Butyrylcholinesterase, carboxylase, and paraoxonase can modulate the effects of toxic exposure by temporarily replacing acetylcholinesterase in breaking down acetylcholine or by breaking down carbamates and organophosphates⁴³. The concentration of these enzymes in the blood stream, where they can quickly detoxify organophosphates, varies from species to species, making human observational studies vital to understanding the specific human biochemistry at work⁴³. Both organophosphate and carbamate exposures can cause long term neurological damage and were experienced in the Gulf War in non-trivial amounts.

Epidemiologic associations of military exposures and GWI

There is strong evidence of an environmental component to the development of GWI⁴. There is a strong association of GWI symptoms in the literature with deployment, across decades of GWI research^{3-6,44}. An association with deployment indicates that some common experience or exposure specific to deployment to the Gulf in the 1990-1991 Gulf War is likely in the causal pathway. There is also strong evidence of an association (even among just those deployed to the Gulf) between GWI and pyridostigmine bromide (PB) pill use and between GWI and pesticide exposure^{4,9,17,21}. Both PB pill use and pesticide exposure are self-reported through surveys¹⁴. Pesticide exposure alone is significantly associated with GWI in several studies, even across several case definitions of GWI and several populations of veterans^{23,45-49}. This

consistency, especially among the heterogeneous and shifting case definitions and cohorts, indicates that pesticides are likely important to the GWI model. The same is true of self-reported PB pill use: it has been significantly associated with GWI regardless of the shifting case definition and different veteran populations^{23,45-47,49-51}. PB pill and excessive pesticide use are unique exposures to deployment in the 1990-1991 Gulf War and they are both significantly associated with GWI across several studies, even across different populations and case definitions. It is therefore reasonable to consider them as promising candidate exposures for the development of GWI.

Not everyone who is exposed to pesticides or PB pills, even at extremely high doses, develops GWI. In addition, not everyone who is unexposed is entirely protected from developing GWI. This indicates that there must be another component to GWI, such as genetics or epigenetics, that interacts with these exposures. There is no genome-wide association study to date, although several studies examine specific SNPs or genetic variants. Most studies are small (i.e. fewer than 100 people) and examine associations with several different case definitions for GWI¹⁵. Several studies identify genetic components within mouse and rat models, identifying important immune loci^{52,53}. Other studies have identified candidate genetic associations with GWI, such as butyrylcholinesterase²², paraoxonase⁵⁴, and Human Leukocyte Antigen (HLA)⁵⁵⁻⁵⁷, but none have been replicated. There have also been no large genetic studies of GWI that analyze a gene-environment interaction model, even with this strong evidence

suggesting that it may be appropriate¹⁵. Smaller gene-environment interaction work shows association, as does work in related phenotypes^{22,30,58}. A gene-environment candidate gene study may be an appropriate place to begin with this exposure-related work.

Available datasets

The Gulf War Era Cohort and Biorepository (GWECB, funded by the VA Office of Research and Development and maintained by Cooperative Studies Project 585, CSP585) consists of DNA, electronic medical records, and survey data from individuals who served in the United States Armed Forces during the 1990-1991 Gulf War⁵⁹. Data collected includes self-reported exposures, exposure lengths, symptoms, symptom severity ratings, clinical conditions, and deployment locations⁵⁹. Women, Veterans of color, and deployed Veterans were oversampled⁵⁹. Unlike some other VA-sponsored studies of GWI, Veterans did not have to use the Veterans Affairs system for healthcare to be eligible for inclusion⁵⁹. The GWECB received surveys from 1344 individuals⁵⁹. The Million Veteran Program (MVP) also developed a Gulf War cohort, who will have filled out an MVP baseline survey, Gulf War specific survey based on the CSP585, and genetic data (Cooperative Studies Project 2006, CSP2006)⁶⁰. The MVP has developed into the premier genetic analysis cohort for the United States Department of Veterans Affairs, with a diverse population and over 825,000 participants enrolled (Million Veteran Program (MVP) (va.gov))⁶¹. The Gulf War cohort of MVP, CSP2006/MVP029, will allow

for further examination of any associations found in the GWECB, or CSP585, as it has many more individuals (about 40,000 Gulf War Era veterans with GWI survey and genetic data)⁶⁰.

Combining genetics with exposures in GWECB

After considering evidence of both possible genetic and environmental associations with GWI, it is possible to test a gene-environment interaction hypothesis in a candidate gene interaction study⁶². Gene-environment interactions can be challenging, especially as candidate studies^{63,64}. In this case, completing a GWAS as an initial study will be important to be sure that the interaction effects are not simply due to the genotype alone. The GWECB dataset includes SNP data from peripheral blood and self-reported exposure data from the CSP585 survey. The survey asks which exposures each deployed veteran experienced, allowing the veteran to select a length of time that they remember being exposed during (0 days, 1-6 days, 7-30 days, or 31+ days). Because veterans who were not deployed to the Gulf during the 1990-1991 Gulf War were not asked to report on their exposures, we are only able to test this in deployed veterans. This gene-environment interaction can be tested using functional variants in previously associated enzymes from smaller functional studies^{22,54}. With the diversity of ancestry among the GWECB dataset (in survey responses, 65% self-reported as White, not Hispanic, 17% as Black, not Hispanic and 10% as Hispanic²⁰), it is possible to test a wider range of variants than we could test in a dataset with only individuals of European

ancestry. Particularly since Black and Hispanic veterans are classified as having GWI at a higher rate than white veterans in the CSP585 dataset²⁰, this could help elucidate a genetic susceptibility that may be more common in non-European ancestry. Genetic diversity among large cohorts is incredibly important for the progress of genetic research, especially among Veteran populations, as the demographics within the United States shift.

Summary

In summary, current state of the art research suggests that while associations with GWI have been found, especially with toxicants such as nerve gas and pesticides, several open questions in the field must be closed to move forward with GWI research. First, the non-specific nature of the survey instruments and case definitions may obscure or drive associations. Researchers continue to use different criteria to evaluate GWI, which makes comparing results difficult. A reproducible and shared case definition algorithm would move the field forward in this space. Second, no genome-wide study has been completed and much research has been done in very small datasets. A genome-wide association study (GWAS) in the GWECB dataset (n=1061, after data cleaning) would move the field forward into this space. Finally, while several associations between exposures and GWI have been identified, there has been no explanation of why some people who are exposed do not get GWI. A gene-environment interaction candidate SNP study could identify a potential factor that may increase or

decrease penetrance of disease. Contributions of a research case definition algorithm, a GWAS of GWI, and a gene-environment interaction showing genetic susceptibility or protection from disease would open a window through which future researchers could better understand the biological underpinnings of GWI. A biological understanding of GWI could lead to the development of therapies and medication targeting the root of disease rather than the symptoms of disease.

2. Phenotype Identification

The first part of this chapter includes a collaborative manuscript on which I am first author¹. I completed the data management, data cleaning, statistical analysis, table generation, and figure generation, and drafted the manuscript. The second part of this chapter includes modified excerpts from two collaborative manuscripts, the first of which I am second author² and the second of which I am a middle author³. I completed

¹ **J. Vahey**, E. R. Hauser, K. J. Sims, D. A. Helmer, D. Provenzale, and E. J. Gifford,

“Research tool for classifying Gulf War illness using survey responses: Lessons for writing replicable algorithms for symptom-based conditions,” *Life Sci.*, vol. 282, p. 119808, Oct. 2021, doi: 10.1016/J.LFS.2021.119808. s⁵⁹

² E.J. Gifford, **J. Vahey**, E. R. Hauser, K. J. Sims, J. T. Efirid, E. K. Dursa, L. Steele, D. A.

Helmer, and D. Provenzale (2021). “Gulf War illness in the Gulf War Era Cohort and Biorepository: The Kansas and Centers for Disease Control definitions,” *Life Sci.*, 278, p. 119454, Mar. 2021, doi: 10.1016/j.lfs.2021.119454.

³ Radhakrishnan, K., Hauser, E. R., Polimanti, R., Helmer, D. A., Provenzale, D., McNeil,

R. B., Maffucci, A., Quaden, R., Zhao, H., Whitbourne, S. B., Harrington, K. M.,

Vahey, J., Gelernter, J., Levey, D. F., Huang, G.D. Gaziano, J.M., Concato, J.,

Aslan, M. “Genomics of Gulf War Illness in U.S. Veterans Who Served during the

the data management, data cleaning, statistical analysis, table generation, and figure generation for the Gifford et al. manuscript and have rearranged and described in my own words the results, discussion, and conclusions here. I am a middle author on the Radhakrishnan et al. paper and generated the figure presented. I have described in my own words the results, discussion, and conclusions here.

1990–1991 Persian Gulf War: Methods and Rationale for Veterans Affairs
Cooperative Study #2006.” *Brain Sciences* 2021, Vol. 11, Page 845 11(7): 845.

Research tool for classifying Gulf War Illness using survey responses: Lessons for writing replicable algorithms for symptom-based conditions

Introduction

Gulf War illness (GWI), a chronic multi-symptom illness, affects up to one third of Veterans who served in the 1990-1991 Gulf War^{3,5,6,65-68}. There is no associated diagnostic biomarker or objective laboratory criteria defining GWI. GWI is diagnosed by the presence of certain symptoms and the exclusion of other disorders that may explain those symptoms^{1-3,7-10}. While the checklists and scoring algorithms were built through extensive research, published studies sometimes do not contain specific documentation on how the case definitions were applied¹⁵. Authors frequently use various modified versions of the Kansas⁶ and Center for Disease Control and Prevention (CDC)⁵ definitions, the two definitions given special consideration by the Institute of Medicine (IOM), to study the associations and incidence of GWI^{3,21,23}. Teams often use near identical survey questions to assess case status, but comparison of results is difficult without a consistently applied algorithm to identify GWI case status from the questionnaire data.

An under-described feature of replicable research is the computer code that considers symptoms and severity scores to assign disease status in symptoms-based diseases like GWI. Best practices from software engineering, such as developing pseudocode prior to writing the code and developing test cases, are intuitive concepts

that can improve clarity for decision making, reduce coding errors, and reduce research costs. Developing pseudocode prior to writing the code and developing test cases could standardize the application of case definitions to different datasets. For example, the CDC criteria (developed by Fukuda et al.⁵) was applied by Dursa et al.⁶⁹ and Steele et al.²³. Both papers, published in the same year, state that the CDC case definition was applied according to prior publications; however, while Steele et al. applied exclusionary criteria, Dursa et al. did not. A standardized algorithm would promote consistency in the definition, facilitating comparisons across studies. These methods also support principles of replicable research^{70,71}.

This study is built upon research conducted in the VA Cooperative Studies Program (CSP) which has developed two cohort studies of Gulf War Veterans (GWV) with near-identical questionnaires. Using those resources, the aims are to (I) develop an algorithm for GWI case status using best practices of software engineering, (II) implement this algorithm in SAS for the Gulf War Era Cohort and Biorepository (GWECEB)⁵⁹, and (III) apply the SAS code to the Genomics of Gulf War Illness within the Million Veteran Program (MVP)⁶¹ with minimal changes. Aim I will use the IOM-recommended Kansas, CDC, and CDC severe definitions to develop a set of decision rules to determine case status. This set of decision rules is defined as the algorithm. Aim II will implement the algorithm in SAS for GWECEB, applying best practices from software engineering. Aim III will demonstrate the flexibility of the code by applying it

to the MVP dataset with minimal adjustment. The framework described here could have broad applicability to any complex condition.

Materials and Methods

Data Sources and Participants:

The VA Cooperative Studies program launched a pilot study in 2014 to develop a data and specimen repository from veterans who served during the era of the 1990-1991 Gulf War. This study, the Gulf War Era Cohort and Biorepository (GWECB), consisted of a survey questionnaire, blood sample, permission to access medical record data, and all relevant consent forms, and was fully described by Khalil et al.⁵⁹. The Genomics of Gulf War Illness study within MVP is comprised of participants in MVP who were in the military in the First Gulf War (1990-91). This group completed a Gulf War-specific survey that augmented the original baseline MVP survey by collecting data on additional exposures, symptoms and medical conditions related to service in the Gulf War⁶¹. In addition to the survey data, a blood sample and medical record data were collected⁷². Briefly, GWECB served as a pilot for the MVP Genomics of GWI; thus, while the variable names may differ, participants from both cohorts completed near-identical self-report mailed surveys. GWECB has 1,343 participants while the MVP: Genomics of GWI study has 41,077 participants. This paper describes the construction of the code for GWI case status determination; a companion manuscript describes the GWI case results determined by using the algorithm on the GWECB survey responses²⁰.

Measures:

The original CDC, CDC Severe, and Kansas definitions for GWI were used to construct the algorithm. These measures are based on questions from self-administered surveys of GWI symptoms and co-morbid disease diagnoses. In the paper that originally defined the Kansas GWI definition, exclusionary conditions were based on two major concepts: (1) a condition could cause symptoms demonstrated by GWI or a condition that might impair a Veteran's ability to accurately and consistently report symptoms, and (2) a condition does not appear in deployed and non-deployed Veterans at different rates⁶. In some previous applications of this approach, exclusionary conditions were identified from the analyzed dataset^{6,21,23}. In developing this algorithm, we chose to use the same chronic conditions that were reported in the original paper⁶ instead of redefining the exclusionary conditions based on our dataset. The inclusion criteria require 'moderate/severe or multiple symptoms' in each of three of six symptom domains. See Appendix Figure 1 for a more detailed list of the symptom criteria and exclusionary criteria. The CDC definition was based on the original 1998 CMI paper and contains no exclusionary criteria: the inclusion criteria require at least one symptom in each of two of three symptom domains⁵. See Appendix Figure 2 for a detailed list of symptom criteria by symptom domain. The CDC Severe definition requires at least one severe symptom in two of three symptom domains⁵, also shown in Appendix Figure 2.

Design and Implementation of Algorithm

Principles of software construction were applied to systematically implement the GWI case definitions, modeling best practice for reusable and reliable classification algorithms for symptom-based case definitions. The process of translating the description of the case definition into an implemented algorithm involved (1) writing pseudocode to plan the algorithm, (2) displaying the planned algorithm^[4] visually to discuss and determine exact specifications^[5], (3) designing a set of test cases based on these specifications to assess accuracy of the code, implementing^[6] the planned classification algorithm, (4) testing the implemented algorithm using the previously written test cases, and (5) disseminating the code publicly.

^[4] Algorithm in computer science refers to the method by which the computer arrives at the intended answer. It refers to the process, not the actual program or code.

^[5] Specifications in computer science are a list of rules for a piece of code: what is considered a valid input, what is considered a valid output, and which inputs should lead to which outputs.

^[6] Implementing an algorithm is when someone writes the code/program/script to take inputs, follow the algorithm, and deliver the intended outputs. Implementation is the point at which there is computer code in a coding language (ex. Python, R, SAS, Java)

Pseudocode, or plain language instructions upon which the computer code is based, was written by the same team members who implemented the algorithm. This pseudocode reflected the logic presented in the original papers describing the case definitions^{5,6}. Developing the pseudocode was necessary for the programmers to write code which accurately captured the complexity of the definitions and the nuanced decisions of team leaders. Multiple factors drove this complexity. First, a series of decisions were made by the team regarding how to clean the self-reported data, reflecting the numerous ways respondents might approach a paper survey. Second, the Kansas definition includes two principles (rather than specific conditions) for the exclusionary criteria (i.e., (a) having a diagnosis that could explain the symptoms and (b) having a condition which could prevent the Veteran from reliably responding to symptom questions). Because team members vary in their perspectives on what should be included, reasonable individuals disagree. Third, decisions regarding how to handle missing information on symptoms, severity ratings, and diagnoses introduced additional complexity.

Pseudocode was then used to generate flowcharts, which were used to graphically depict details and decision points. These flowcharts helped to ensure that all team members had a shared understanding of the decisions, allowing translation from the big picture aims to the nuanced details⁷³. Pseudocode and flow diagrams can improve communication across the multidisciplinary expertise of the team^{74,75}. This

process defines the algorithm and provides the opportunity to fine tune it to the state-of-the-art understanding of the GWI case definition before coding begins.

To ensure the code accurately reflected the designed algorithm, a thorough testing method was necessary. We employed the test-first programming strategy, which used the pseudocode and flowcharts to write test cases to test every aspect of the designed algorithm⁷⁶. The test cases were designed to represent all possible combinations of survey item inputs and each test case represents one hypothetical person. For each test case, a set of inputs is provided (e.g., symptom ratings and diagnosis indicators, in this example), as well as the gold standard version of the outcome (e.g., GWI status) that is being tested. The code can be tested on these hypothetical individuals to determine if the implemented algorithm was correctly programmed to capture the correct answer in all possible individuals. The implemented algorithm assigns a GWI case status, which is compared to the independently determined gold standard GWI case status. 'Passing' the test case means that the two are identical, while 'failing' indicates that the result output by the implemented algorithm is not the same as the expected result. Test cases were written in Excel, which separated them from the implementation of the algorithm.

The algorithm was implemented using Don't Repeat Yourself (DRY) technique, meaning that repeated pieces of code are written once and are saved as a function or variable⁷⁷. In SAS, these are called macros and macro variables, and work by basically

replacing the chosen variable name with the relevant portion of the previously defined code. This technique is used so that slight modifications can be made with regard to which symptoms or diagnoses are incorporated into the definition, without rewriting the code or editing the code in too many places. It also allows the base case, or smallest repetitive piece of the algorithm, to be implemented just once. This base case is then used several times to implement the full algorithm. For example, the CDC severe definition requires at least one severe symptom from two of three symptom domains. The base case macro takes as an input the symptom domain and the list of symptoms in the domain, then outputs whether or not a severe symptom is in that domain. This base case macro is run once for each symptom domain, then the results are combined to determine CDC GWI Severe case status. Without macro programming, the code in the macro would have needed to be written three separate times, potentially adding typos and other errors. Open-source macros were used to streamline this process⁷⁸. Following the core tenets of software construction to keep the code tight, clean, and modular significantly improves both the reliability and reusability of the case definition algorithm^{76,77}.

The implemented algorithm was tested on the pre-written test cases. Test cases were used by running the code (implemented algorithm) on all of the test cases, then comparing the previously determined outcomes in the test case file with the outcomes generated by the code. Simple two-by-two tables were used to compare the outputs of

the classification algorithm with the outcomes written into the test cases. These comparisons were done for every outcome generated by the designed and implemented algorithm. Code was edited until all test cases passed.

Results

Aim 1: Design of Algorithm

The algorithm was fully designed before any code was written, allowing for critical decision making prior to the implementation of the algorithm. Plain language pseudocode, developed in Microsoft Word using the original Kansas and CDC case definition papers, was used to highlight key decision points. Flowcharts, designed in Microsoft PowerPoint, accurately and concisely communicated the designed algorithm to the team. The pseudocode and flowcharts were redesigned until the team agreed on the algorithm. The flowcharts facilitated discussion, as it was not necessary to understand the code to contribute to the conversation and decision making. This

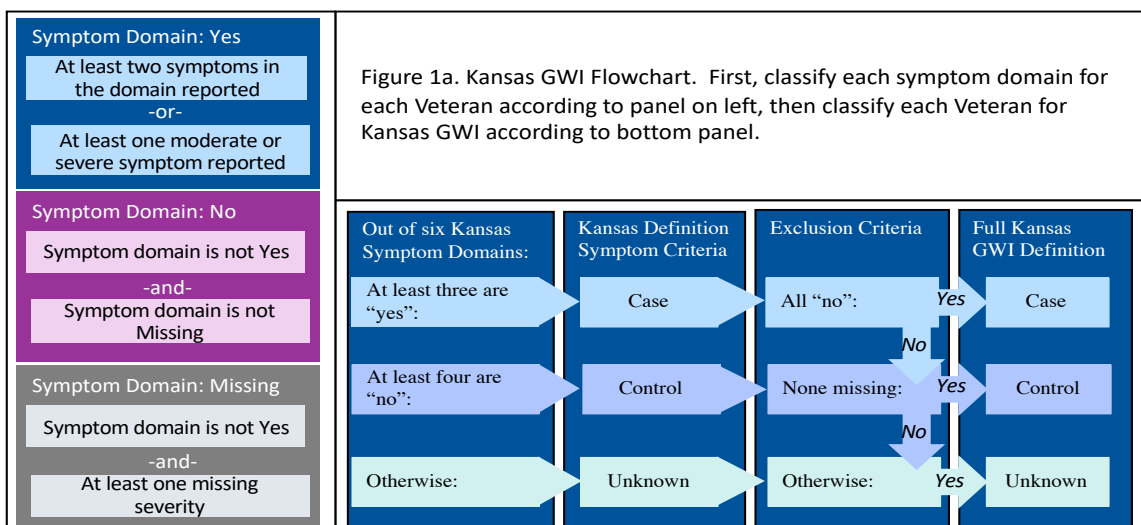


Figure 1. Kansas GWI Flowchart

separation of the design and the implementation of the algorithm permitted important algorithmic decisions to be based on the published case definitions and needs for future projects, rather than ease of coding. Figures 1 and 2 illustrate the flow charts for the Kansas and CDC definitions, respectively. Since the entire team could read and understand the flow charts, everyone was able to contribute to the discussion and decision making. The team-based approach led to a more robust case definition based on the contributions of team members with diverse backgrounds and a sense of ownership over the decisions that were made, engendering confidence in the final algorithm.

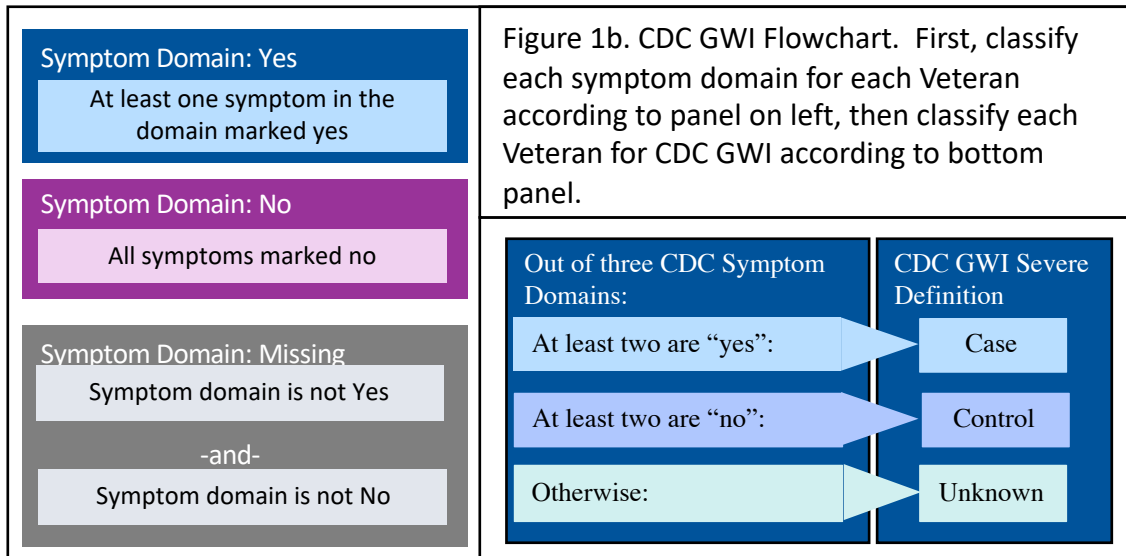


Figure 3. CDC GWI Flowchart

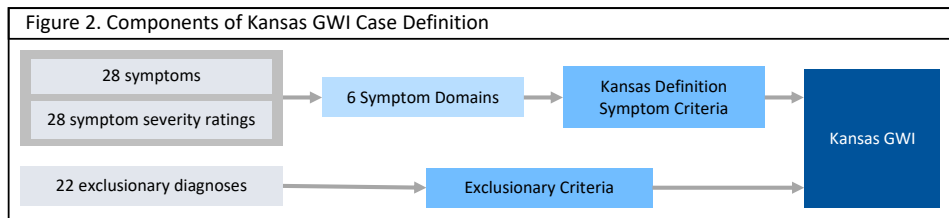


Figure 2. Components of Kansas GWI

The test cases added significant time and effort to the design and implementation process, but without them, it is difficult to assess accuracy of the program. Test cases provided clarity on edge cases, or situations where the algorithm may misclassify individuals, and validated that the code behaved as intended. These tests are always written based on the pseudocode and flowcharts, independently from the code itself, as the goal is to be able to detect errors in the implementation of the algorithm⁷⁶. In particular, writing these test cases clarified the degree of missing information the algorithm could support and still assign a legitimate classification. A missing classification option was retained in the case that it was impossible to assign a case definition. As seen in Figure 3, Kansas GWI begins with symptoms, which make up symptom domains. These symptom domains are then combined to generate a symptom criteria indicator, which is combined with the exclusionary criteria to determine Kansas GWI case status. Each symptom domain was tested extensively: for every symptom domain, a test case was written for every combination of valid values for all symptoms in the domain. Figure 4 shows an example of the process to go from the survey question and pseudocode to the set of test cases. In one case, the Kansas Neurological, Mood, or Cognitive symptom domain, the full cartesian product became intractable (12 symptoms with six valid inputs per symptom led to 78 billion possible test cases), so partitioning was used to test that symptom domain. Documentation for the design of the algorithm will be submitted to the Department of Veterans Affairs (VA) Phenomics library⁷⁹, a

database of research phenotypes defined in VA research studies that can be applied by others to their own datasets, with submission of this manuscript. Pseudocode, flow charts, and test cases allowed full determination of the classification algorithm before any code was written, leading to team-wide confidence in the case definition.

Figure 3. Illustration of Test Cases for Kansas GWI Case Definition Classification Algorithm

Six possible ways to answer a symptom question:				Symptom 1		Symptom 2		Outcomes to be Tested				
Symptom	Severity			Yes/No	Rate Sev.	Yes/No	Rate Sev.	Any	Count	Moderate or Severe	Missing Count	Multiple or Moderate/Severe
Unknown	-	Unknown	-	--	--	--	--	0	0	0	2	Missing
No	0	None	0	--	--	0	0	0	0	0	1	Missing
Yes	1	Unknown	-	--	--	1	--	1	1	0	2	Missing
Yes	1	Mild	1	--	--	1	1	1	1	0	1	Missing
Yes	1	Moderate	2	--	--	1	2	1	1	1	1	Yes
Yes	1	Severe	3	--	--	1	3	1	1	1	1	Yes
				0	0	--	--	0	0	0	1	Missing
				0	0	0	0	0	0	0	0	No
				0	0	1	--	1	1	0	1	Missing
				0	0	1	1	1	1	0	0	No
				0	0	1	2	1	1	1	0	Yes
				0	0	1	3	1	1	1	0	Yes
				1	--	--	--	1	1	0	2	Missing
				1	--	0	0	1	1	0	1	Missing
				1	--	1	--	1	2	0	2	Yes

Each symptom domain has 6ⁿ test cases, where n is the number of symptoms in that domain

Indicator for Symptom Domain Status, used in Figure 1, is outlined

This symptom domain has two symptoms, so there are 36 test cases. The first 15 are shown here.

Figure 4. Illustration of Test Cases for Kansas GWI Case Definition Classification Algorithm

Aim 2: Implementation of Algorithm

The implementation of the case definition classification algorithm utilized techniques drawn from software engineering. SAS Macro language was used to best follow Don't Repeat Yourself (DRY)^{76,77} techniques, which shortened the program dramatically. For example, a macro was written to calculate a Kansas symptom domain indicator using an input list of symptoms. This macro was then used to calculate all six

Kansas symptom domain indicators for each individual, using the six lists of symptoms relevant to each respective symptom domain. An identical process was applied to the CDC symptom domain indicators: one macro was written and then was used to calculate all three CDC symptom domain indicators. Macro variables were also used to define the symptom and diagnosis identifiers for each symptom and diagnosis used in the program. These macro variables provided the opportunity for late phase adjustment to be made quickly and accurately; a symptom was removed from the CDC Fatigue domain, and one was added to the Kansas Skin domain. The ability to change the composition of the definition at a late stage of the project without causing significant delays demonstrates the flexibility of the code and anticipates the adjustment of the case definition in future studies. While this type of programming is an example of best practice in software engineering^{76,77}, it is more time consuming to write and design, and can require more time to debug. Using DRY programming techniques cuts down on the amount of code that is copied, pasted, and edited, which reduces silent errors^[7]. The code produced no errors and the only warnings produced were format warnings due to

^[7] Silent errors are the worst type of programming errors—a mistake is made, causing the wrong answer to be generated. It is silent because an answer is still generated and no error or warning is reported. Often there is very little chance in identifying these errors without a thorough independent code review or set of test cases.

variables that were not used in the algorithm. The implemented algorithm ran to completion in GWECB (n=1,343) in 5.6 seconds and the resulting dataset was saved as a SAS dataset. Comments explaining the logic and indicating where changes could be made are spread throughout the code.

The previously described test cases were used to validate the algorithm, leading to the quick identification of several coding mistakes and providing confidence in the outcomes produced. These test cases demonstrated that the implemented algorithm functioned as intended. Test cases and code have been submitted to the VA Phenomics library⁷⁹. Best practices of software engineering, including techniques to limit the use of copy/paste and techniques to check the results produced against an expected result, resulted in code that was flexible, accurate, and reliable.

Aim 3: Adaptation to New Dataset

The use of SAS Macro language to implement the designed classification algorithm allowed quick modification for use in the GW MVP dataset. The code and test cases were repurposed for the GW MVP dataset with changes on only 15 lines of code and 12 corresponding headers for the test cases. These changes represent differences in variable names for the same or similar questions on the symptom questionnaire. The code was 400 lines in length. The modifications were made in in the first 50 lines of code, with comments indicating how to adjust the code for a new dataset. Adapting the code and test cases for the new dataset, as well as running the code and test cases on the new

dataset, were completed in less than a day. When applied to the 41,077 individuals in the GW MVP dataset, the code ran to completion in 32.7 seconds.

Testing for the GW MVP dataset was also straightforward. Twelve headers, corresponding with 12 of the 15 Macro variables adjusted in the code, were revised to represent the GW MVP symptom and diagnosis identifiers. No additional changes were necessary. Testing caught a misspelled symptom identifier, which was corrected before results were reported. All test cases passed. The ease with which the code was applied to a new dataset with different variable names illustrates its flexibility and shows that this code could be used to standardize the GWI case definitions across datasets.

Discussion

Documentation of algorithm

Through the application of software engineering core principles to write code for the case definition classification decision rules, the exact algorithm is clear and can be replicated in any coding language. The documentation available describes the inputs, outputs, and process for computing the outputs from the inputs. This coding technique lowers the cost of writing code, as decision makers can easily understand where the decision points are and how these decisions could affect results, leading to fewer iterations of rewritten code. It also allows other groups to utilize what has already been written, promoting efficiency and reducing the time required to recreate the case definition algorithm. In addition, the inclusion of pseudocode, flowcharts and test cases allow the algorithm to be readily adapted to different software languages, promoting

usability and enhancing reliability and replicability. In this way, even though the code takes longer to design and write when following this method, the code produced is of higher quality, is easier to reuse, and can be quickly adapted to fit future projects.

Limitations

A major limitation of this method as described here is that the same individual developed both the code and test cases. This can result in identical errors in both locations, leading to the same erroneous output in the test cases and in the implemented algorithm. Techniques employed to mitigate the risk of such errors included using separate programs to write the code and test cases, multiple rounds of team review of the algorithm, peer review of code, and time-delayed self-review of code and test cases. Another possible limitation was the amount of time that the development phase required. This, however, is not a true limitation, as the time saved later in the process more than compensated for the high cost of the design phase. The entire process generated significant confidence in the algorithm produced and the test cases, in particular, ensure that the code followed the planned algorithm. The code and test cases produced are reliable, flexible and generalizable to other projects, potentially saving much more time than that spent planning and documenting the algorithm.

Flexible code for future work

The process we outlined led to the development of flexible code. As the understanding of GWI continues to change, the ability to modify the symptoms in each symptom domain and the conditions in the exclusionary criteria will reduce costs in the

future. In addition, the comments and macros used in the code facilitate revisions in the input and other larger adaptations. Thus, the programming method we described can be readily adapted to future projects as the case definition of GWI evolves. Finally, the implementation of a well-defined process resulted in standardized, validated code to assign GWI case status. This algorithm could be applied to other cohorts, facilitating the comparison of results across studies.

Conclusions

Through the application of best practices of software engineering, the authors were able to take advantage of all skillsets on the team to generate a reliable and replicable case definition classification algorithm for Gulf War Illness. The documentation process, including the modularity of the code and the inclusion of test cases, supports the accuracy of our work and facilitates the replication of our work by other groups. Gulf War Illness is a challenging phenotype, in large part due to imperfect and inconsistently applied case definitions. A standard process and algorithm for the GWI phenotype could facilitate comparisons across cohorts and provide a common platform for future, biologically based studies.

Application of GWI Algorithm to CSP585 and CSP2006

Applying this algorithm to produce CDC, CDC Severe, and Kansas GWI case definition indicators in both GWECB/CSP585 and CSP2006/MVP has led to several papers. Here, I present tables and figures from the two published manuscripts in which application of the GWI algorithm allowed for further study (Gifford et al., 2021²⁰ and Radhakrishnan et al., 2021⁶⁰). This work described the GWECB cohort in terms of both demographics and GWI-related measures, compared the GWI-related measures in GWECB with those from the original Steele, 2000⁶ paper, and used the applied case definition to consider future projects for the MVP dataset. I am the second author on the Gifford et al. paper and performed the data management, data cleaning, statistical analysis, and generated the tables and figures. I am a middle author on the Radhakrishnan et al. paper and generated the figure presented.

Introduction

The GWI case definition algorithm manuscript was the product of a collaborative effort to standardize the GWI case definition. Having completed that, the next step was to apply that case definition to both CSP585 and CSP2006. The CSP2006 surveys were distributed and returned later than the CSP585 surveys, providing an opportunity to adjust the research plan based on preliminary results from CSP585. Described here is some of the published collaborative work that came out of these case definition efforts. First, the GWECB had no previously published results discussing GWI case status. Since the cohort is thought to be highly biased due to a 12% survey return rate, there is little confidence in estimating the cohort GWI prevalence, even in comparison to population prevalence⁵⁹. The confidence instilled by our case definition process was able to ameliorate this problem⁸⁰. Second, the MVP and GWECB cohorts are both VA cohorts; streamlining the case definition to compare results and better utilize both cohorts was a large focus of the collaborative work. Finally, with a reliable and well-defined GWI case status indicator, more detailed and hypothesis focused experiments can be planned. This is especially important for the CSP2006 team, which is working with a dataset of about 40,000 individuals. It would not be possible to go through the dataset to identify who has GWI on a case-by-case basis, as might be possible in small clinical trials or smaller observational studies. This algorithm allows progress to be made on the Gulf War projects in the MVP without further delay.

Results

GWECB demographics and GWI-related outcomes

In the GWECB, 76% of respondents report being deployed to the Persian Gulf in support of the 1990-1991 Gulf War. Overall, 23% were female, which is an overrepresentation of female veterans for the time period. Approximately a third of the cohort reports Black, Hispanic, or Other in response to the race/ethnicity survey question⁵⁹. Table 1 describes the cohort by deployment status. Notably, unit component is associated with deployment ($p=0.01$), while sex, age group, race/ethnicity, household income, education level, service branch, VHA utilization, and deployment in post-9/11 wars were not significantly associated with deployment to the Gulf in 1990-1991.

Table 1. Sample characteristics by 1990-1991 Gulf War deployment status

Characteristics		All ¹ n=1116	Deployed ² n=849	Did Not Deploy n=267	p- value from χ^2
		(%)	(%)	(%)	
Sex	Male	76.8	78.0	73.0	0.0953
	Female	23.2	22.0	27.0	
Age Group	40-49 years	38.7	40.1	34.5	0.2538
	50-59 years	36.9	36.3	39.0	
	60 years and over	24.4	23.7	26.6	

Race/ Ethnicity ³	White, Not Hispanic	65.1	65.3	64.4	0.4615
	Black, Not Hispanic	17.2	16.7	18.7	
	Hispanic (any race)	9.5	9.5	9.4	
	Other	6.2	6.0	6.7	
Household income per year	Under \$30,000	11.2	10.7	12.7	0.5835
	\$30,000 - \$59,999	23.1	22.3	25.8	
	\$60,000 - \$99,999	29.0	29.7	27.0	
	\$100,000 or more	29.8	30.3	28.1	
	Unknown	6.9	7.1	6.4	
Highest achieved education level ³	High School diploma/GED or less	9.0	9.5	7.1	0.6373
	Some college to Associate's or Bachelor's degree	68.2	67.8	69.3	
	Master's degree, Professional degree, or Doctorate degree	20.7	20.4	21.7	
Unit Component	Active Duty Only	60.6	58.1	68.5	0.0108
	Reserves Only	14.6	15.8	10.9	
	Both Active Duty and Reserves	24.4	25.6	20.6	

Service Branch	Army Only	45.5	45.1	46.8	0.4698
	Navy Only	16.1	16.7	14.2	
	Air Force Only	11.0	10.3	13.5	
	Marine Corps Only	12.5	13.3	10.1	
	National Guard ⁴ : All	9.8	9.7	10.1	
	Other	5.0	5.0	5.2	
Used VHA health care or hospital in the last year	Yes	44.3	44.9	42.3	0.3848
	No	54.9	54.1	57.7	
Deployed in support of OEF or OIF	Yes	21.7	22.7	18.4	0.1952
	No	76.5	76.2	77.5	
¹ Includes 75 individuals who were deployed in support of the Persian Gulf war but not to the Gulf War Theater or Operations ² Includes Veterans who were deployed to the Gulf War Theater of Operations in support of the Persian Gulf War ³ Percentages may not sum to 100% due to unknown responses or missing values in some categories. ⁴ National Guard is not a service branch but was asked in conjunction with military branch questions.					

* OEF=Operation Enduring Freedom; OIF=Operation Iraqi Freedom

GWl case status-related measures by deployment

GWl case status, symptom domains, and exclusionary criteria are key measures for the GWECB. All GWl-related measures remain significantly associated with deployment except for the CDC mild-moderate musculoskeletal domain and the Kansas exclusionary criteria. The CDC Severe GWl case definition is the most strongly associated with deployment, followed by the Kansas symptom criteria. Although both have an odds ratio lower confidence level between 1.0 and 1.1, the CDC GWl (any severity) case definition and the full Kansas GWl case definition are significantly associated with deployment. The odds ratio for the musculoskeletal domain is the most impacted by severity, with an initial odds ratio of 1.29 and a severe odds ratio of 2.39. This increase is due to a large decrease in severe symptom reporting among the non-deployed: where 86.5% report any musculoskeletal symptom, only 19.1% report a severe musculoskeletal symptom. Among the deployed, that reduction is smaller: from 87.5% to 33.6%. Skin symptoms are the most significantly associated with deployment among the Kansas GWl components.

Table 2. GWI Case Status and Related Components by Deployment Status

Measure	Deployed (n = 849)	Did not deploy (n = 267)	aOR	95% CI
Panel A. GWI case status-related measures				
Kansas symptom criteria	72.0	59.9	2.05	(1.50, 2.80)
Kansas GWI case	39.9	32.6	1.42	(1.05, 1.93)
CDC GWI case	84.2	80.1	1.57	(1.07, 2.29)
CDC GWI severe case	26.9	13.5	2.67	(1.79, 3.99)
Panel B. Kansas GWI components				
Moderate or multiple symptom domain				
Fatigue/sleep problems (4 symptoms)	79.4	68.9	2.14	(1.53, 3.00)
Pain (3 symptoms)	72.9	65.9	1.56	(1.14, 2.14)
Neurologic/cognitive/mood (14 symptoms)	86.3	80.1	1.80	(1.22, 2.65)
Gastrointestinal (3 symptoms)	38.5	30.3	1.57	(1.15, 2.15)
Respiratory (3 symptoms)	35.5	24.0	1.93	(1.39, 2.67)
Skin (2 symptoms)	32.4	18.4	2.41	(1.70, 3.44)
Exclusionary criteria	40.3	40.8	1.05	(0.78, 1.42)
Panel C. CDC GWI components				

Symptom domain				
Fatigue (1 symptom)	69.0	55.1	2.18	(1.61, 2.96)
Musculoskeletal (3 symptoms)	87.5	86.5	1.29	(0.84, 1.97)
Mood–cognition (6 symptoms)	88.7	81.6	2.09	(1.39, 3.13)
Severe symptom domain				
Fatigue (1 symptom)	16.4	6.7	3.09	(1.83, 5.22)
Musculoskeletal (3 symptoms)	33.6	19.1	2.39	(1.68, 3.41)
Mood–cognition (6 symptoms)	34.0	19.9	2.33	(1.64, 3.32)
aOR = adjusted Odds Ratio; CI = confidence interval.				
Note: adjusted for sex, education, income, service branch, unit component, and age.				

All components plotted in Figure 5 from GWECB are either consistent or decreased in odds ratios from those reported by Steele, 2000⁶; ‘other skin conditions’ was not reported in the earlier paper. Our results have smaller confidence intervals than those reported by Steele in 2000. These results are plotted as three groups: all, those who do not fulfill the Kansas exclusionary criteria, and those who do fulfill the Kansas exclusionary criteria. While the “all” group contains the entire cohort, the second and third groups listed are mutually exclusive. Across all symptoms, the confidence intervals overlap between the three GWECB groups. Almost all symptoms are significantly associated with deployment in the “All” group. The symptoms not

significantly associated with deployment in the “All” group are: ‘diarrhea’ and ‘wheezing in chest’.

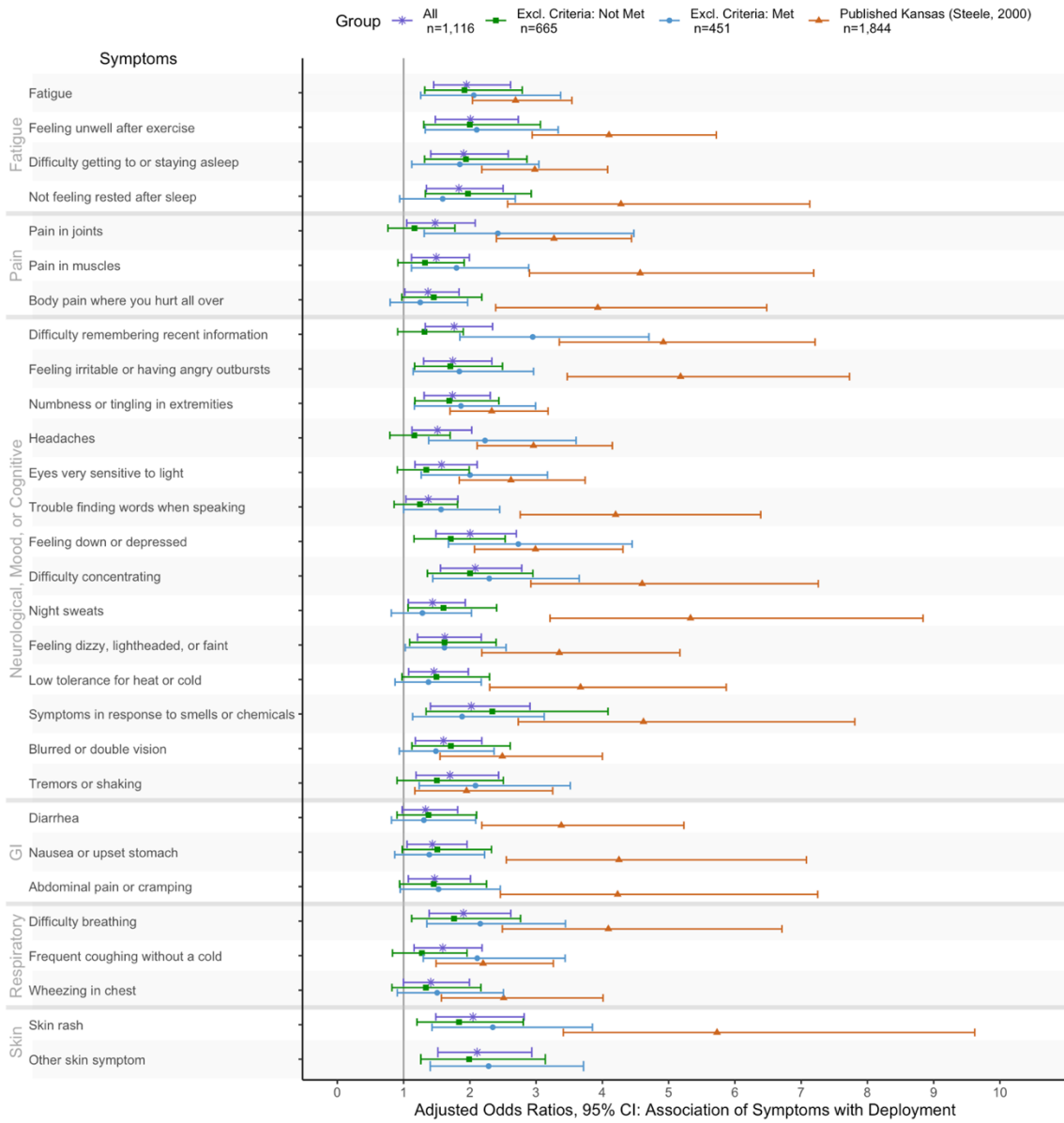


Figure 5. Association of GIWI symptoms with deployment

Application to MVP

The MVP team used the same lists of Kansas symptoms and exclusionary conditions, as well as CDC symptoms, as the GWECB team, with exception of substituting melanoma for general skin cancer (Figure 6)⁶⁰. Using the more specific melanoma indicator was not possible in the GWECB, even though it is more appropriate, because the survey did not differentiate between melanoma and non-melanoma skin cancers²⁰. Once results from the MVP are available, it will be possible to directly compare with GWECB results, as the same algorithm was used to identify cases and non-cases.

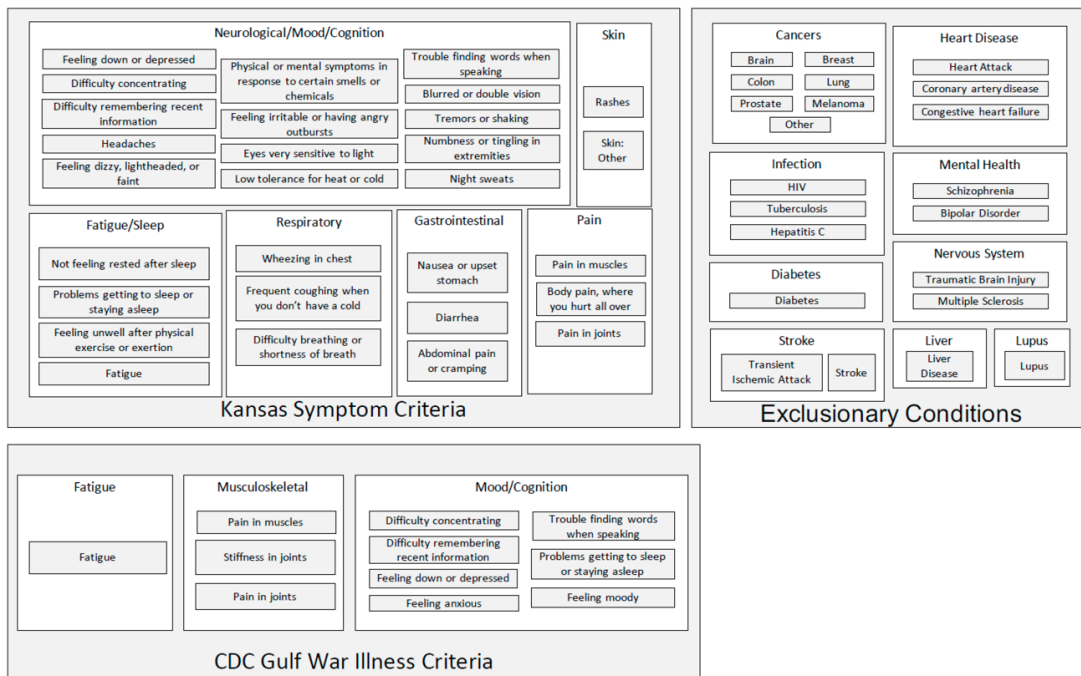


Figure 6. MVP case definition graphic

Discussion

The GWECB dataset is a diverse dataset; compared to the overall population of Gulf War veterans, there is an overrepresentation of women, Black veterans, and

Hispanic veterans. This overrepresentation was planned in the recruitment phase, as an effort to recruit more female and non-white veterans to research studies within the VA⁸¹. The only demographic factor that is associated with deployment is unit component ($p=0.01$, Table 1). This is vitally important for identifying covariates for the GWI association analyses. Importantly, later deployment, measured here with deployment to OEF/OIF (post-9/11 wars in the Gulf region), is not associated with deployment to the Gulf in the 1990-1991 Gulf War. In some military analyses, the healthy warrior effect can severely bias the analysis due to the fact that the military requires a certain level of fitness and health for deployment⁸². In the words of Dr. Haley, who coined the phrase in his 1998 paper, the healthy warrior effect describes, “the selection bias from systematic differences in the health of military personnel who are deployed to a war zone and those who are not deployed due to the selective withholding of chronically ill soldiers from deployment.” We might expect a double bias here, both in deployment to the Gulf in 1990-1991 and in deployment in subsequent wars. We are not seeing an association with deployment in 1990-1991 with deployment in subsequent wars, which would have complicated analysis. The demographic information informs on covariates of interest. With these results and prior results in mind, it was decided that the genetic models of GWI would adjust simply for sex and age, in addition to genetic principal components, as appropriate.

We also do not see an association between Kansas exclusionary criteria and deployment (Table 2), which is vitally important to the full Kansas case definition. When this case definition was first defined, the exclusionary criteria were intended to identify those whose symptoms could be explained by common diseases that are not associated with deployment. The GWECB dataset confirms that these common conditions continue to not be associated with deployment in 1990-1991. Our results also indicate that, with over 40% of both deployed and non-deployed veterans fulfilling the Exclusionary Criteria, they might be too stringent for data collected 25 years after the Gulf War. Further work will be necessary to identify the effects of these Exclusionary Criteria.

The Exclusionary Criteria do not seem to change the association of the symptoms with deployment. In Figure 5, the cohort is stratified by Exclusionary Criteria status, which is not how it is worked into the Kansas case definition. Although the two stratified groups appear to experience the same patterns, it is possible that including those with both GWI symptoms and Exclusionary Criteria in the control group may be contaminating the control group with potential cases. This can be seen in the decrease in odds ratio for the association with deployment between the Kansas symptom criteria and the full Kansas case definition. Further work is necessary, potentially in the MVP dataset, especially as the average age of the cohort increases over time and exclusionary conditions become more common.

The many discussions between the GWECB team and the MVP team streamlined the application of the case definition algorithm to MVP. Since the MVP team was involved in discussions from the beginning, the decisionmakers agreed on the contents and makeup of the case definition. Having seen the test cases catch translation errors between the GWECB and the MVP datasets, there was confidence in the accuracy of the code compared to the agreed-upon algorithm. The application of this case definition to the MVP has been completed, but no results have been published yet.

Confidence in the case definition coding has led to the ability to test more specific hypotheses. The future work described in the Radhakrishnan et al. paper would not be possible without the case definition work. Completing a GWAS and a gene-environment interaction work in a cohort of 40,000 Gulf War era veterans requires an automated case definition. It would not be possible for an individual to go through each person to categorize each veteran as a case or as a control. This will continue to be true as analysis moves forward: work in genetic, methylation, and other high throughput data types will allow for a large dataset, but only if the case definition is applied evenly across the dataset in a consistent and time-efficient manner. This deterministic algorithm has allowed for both consistency and time efficiency, without sacrificing accuracy or confidence in the results. Future work includes a GWAS in both the GWECB and the MVP datasets, as well as gene-environment interaction and linear

mixed model (LMM) approaches to answer specific questions about the relationship between genes, exposures, and GWI outcomes.

Conclusions

Application of the case definition algorithm to the GWECB and MVP datasets has allowed analysis to begin. The results will be comparable between cohorts and features of the cohorts can begin to be observed. The genetic model will be designed using these results. Finally, deep understanding of the cohort and of the case definition will allow for better designed models and better interpretation of results.

3. Genome-wide Investigation

This chapter consists of an unpublished manuscript on which I am first author, with the title and author list noted in the footnote⁸. I completed the data management, data cleaning, statistical analysis, table generation, and figure generation, as well as writing the first draft of the manuscript. I appreciate the mentorship, guidance, and idea generation of several coauthors in the process of drafting the manuscript. Several coauthors contributed to editing, generating the data, and writing grants to support this work. This manuscript has not yet been submitted for publication.

⁸ J. Vahey, X. Qin, A. Stone, W. Carter, L. M. Griffin, S. Pyarajan, G. Turner, E. J. Gifford, K. J. Sims, C. Williams, E. R. Hauser. *Genetics of Gulf War illness, a genome-wide association study*

Genetics of Gulf War illness, a genome-wide association study

Introduction

United States Veterans who served in the 1990-1991 have continuously reported higher rates of chronic multisymptom illness (CMI) than both the general population and other military cohorts³. Out of the 700,000 military personnel who were deployed to the Southwest Asia theater of military operations in the 1991 Gulf War, it is estimated that between 175,000 and 250,000 (25-35%) have chronic multisymptom illness (CMI), in this group defined as Gulf War Illness (GWI)³. This high level of unexplained illness was reported as early as 1995, less than five years after deployment to the Gulf^{11,83}. Many of the symptoms used to define GWI are being reported by veterans deployed more recently to Iraq and Afghanistan^{12,28,84}, suggesting common causes of GWI, most of which remain undiscovered but include both war-related environmental exposures and individual susceptibility. A better understanding of underlying genetic susceptibility to CMI and GWI could help with its treatment or prevention of similar symptoms in other at-risk groups.

The study of GWI poses several challenges the first of which is applying a standard research definition. GWI has several generally accepted research case definitions based on combinations of self-reported chronic symptoms⁸⁰. The two main definitions are the Kansas⁶ and Center of Disease Control (CDC)⁵ definitions. The

Institute of Medicine and the Research Advisory Committee on Gulf War Illness have extensively reviewed research over the past 25-30 years, concluding that the Kansas and CDC case definitions are significantly associated with Gulf War deployment and that these case definitions are preferred for research application^{2,3}. For this analysis we choose the CDC Severe case definition as the primary outcome, as it demonstrates the strongest association with deployment in this dataset²⁰. However, in application to genetic analysis it remains unclear which of these case definitions is the most heritable or even if GWI is heritable at all. Further compounding the lack of information on the heritability of any case definition or its components is the need for exploration of genetic models including genome-wide analysis.

Genome-wide association studies (GWAS) are an established method for gene identification in well-defined diseases^{85,86} and are now being used to study traits with less well-defined genetic architecture, such as suicide^{87,88}, post-traumatic stress disorder (PTSD)⁸⁹, depression⁸⁹, and Chronic Fatigue Syndrome (CFS)⁹⁰. Even small datasets can be a very important starting point when working with a trait that does not yet have any genome-wide genetic work published; the first GWAS of age-related macular degeneration had only 96 cases and 50 controls and launched the GWAS era with its publication in 2005⁹¹. Even where statistical power is low, the results of such GWAS analysis can be very useful for identifying loci of interest and subsequent meta-analyses^{92,93}. There have been many small genetic association studies of GWI, focusing

on candidate loci Human Leukocyte Antigen (HLA)^{25,55}, butyrylcholinesterase (BCHE)²², and paraoxonase (PON1)^{54,94}. These results have been difficult to replicate; in particular, the PON1 finding has been elusive⁹⁵. Genome-wide studies of GWI allow an unbiased screen of the genetic architecture and could identify smaller sets of genes to specifically test in future work. Other work could be done with summary statistics: gene or gene set associations^{96,97}, phenome-wide studies^{98,99}, or linear mixed model (LMM)^{100,101} approaches. These methods require that a GWAS be completed and summary statistics be released. In this study we describe an initial GWAS analysis of the Gulf War Era Cohort and Biorepository (GWECEB), including 1,061 individuals with high-quality genomic data and Gulf War illness status indicators for the IOM/RAC recommended case definitions¹⁰². Summary results will be made available for future work.

Materials and Methods

Gulf War Era Cohort and Biorepository

The Gulf War Era Cohort and Biorepository (GWECEB) was designed as a nationally representative longitudinal cohort of Gulf War era veterans, with survey data, DNA, and electronic medical records⁵⁹. Data collected include demographics, symptoms, symptom severity ratings, clinical conditions, and deployment locations. Use of the Veterans Health Administration healthcare system was not required for inclusion and female veterans, non-white veterans, and deployed Veterans were oversampled⁵⁹. Development of the study, description of the survey data and development of the

analytic dataset have been described^{20,59}. Gifford et al. developed an analytic dataset consisting of 1,116 individuals who had all four GWI indicators, were not missing major covariates (sex or deployment), and were not deployed elsewhere²⁰. GWI-related variables include symptom domains, Kansas exclusion flag, and GWI indicators⁸⁰. In addition, the analytic dataset includes covariates of age, race/ethnicity, highest achieved education level, household income, unit component, service branch, Veterans Health Administration (VHA) use, and deployment to OEF/OIF. Combined with the genetic data cleaning described below, a total of 1061 participants were included in the GWAS analysis.

GWI Case Definitions

GW case status was calculated using a deterministic algorithm based on 32 self-reported chronic symptoms, fully described by Gifford et al., 2021^{20,80}. Symptoms and diagnoses used in the GWI case status indicators are shown in Appendix Figures 1 and 2⁸⁰. Outcomes were chosen based on expert panel recommendations; the Institute of Medicine has recommended the use of the CDC and Kansas definitions for GWI research³. All four case definitions are more prevalent among the deployed Veterans than the non-deployed Veterans. While each case definition has a very different case rate, the case definitions are closely related. All those who fulfill the CDC Severe GWI criteria must also fulfill the CDC GWI criteria. Likewise, all those who fulfill the Kansas GWI criteria must also fulfill the Kansas Symptom Criteria. Many of the symptoms

necessary to fulfill the CDC GWI and Kansas symptom criteria also overlap. Thus, these case definitions are not independent. Our primary outcome was CDC Severe GWI because it is most strongly associated with deployment in the GWECB and is entirely symptoms-based²⁰. Secondary outcomes were CDC GWI, Kansas GWI, and Kansas symptom criteria. The Kansas Exclusionary Criterion, the indicator for self-report of least one exclusionary condition, is fulfilled by 40.6% of the GWECB population, meaning that 40.6% of the study population are ineligible to be considered for Kansas GWI (Table 4, Results). Thus, the Kansas Symptom Criteria was chosen as the primary Kansas case definition for this study. These definitions were developed in conjunction with the Million Veteran Program's Gulf War Illness study to allow comparisons of military cohorts⁸⁰.

Genotyping and genetic data cleaning

Genotyping was done on an Illumina Omni2.5-8v1.4 microarray at the Pharmacogenomics Analysis Laboratory (PAL) in the Central Arkansas Veterans Healthcare System using standard protocols. Plink 1.9 was used for subsequent genetic data cleaning and analysis¹⁰³. A total of 2,372,461 SNPs were genotyped; 82,369 SNPs were removed due to missingness greater than 1% and 512,116 SNPs were removed due to minor allele frequency (MAF) less than 1%, leaving 1,777,976 SNPs. Sample QC replicates were removed, keeping the most informative sample for each individual. Two

individuals were removed due to SNP missingness >1% and 13 individuals were removed due to genetic relatedness (estimated identity-by-descent greater than 12.5%).

GWAS analysis

Genome-wide association studies (GWAS) were completed on 1,061 Veterans, using a logistic regression model, adjusting for age, sex, and ten genetic principal components (PCs). The logistic regression model uses the logit function for each outcome as Y, with a separate regression model for each SNP to test each SNP individually for association. Ten PCs were chosen, to follow prior work; the top 10 genetic PCs represent 87% of the genetic variability¹⁰². The PCA plot (PC1 and PC2) and the scree plot can be found in Appendix Figure 3. GWAS was run for the primary outcome, CDC Severe GWI and three related outcomes: CDC GWI, Kansas GWI, and Kansas GWI symptom criteria. Analyses to evaluate robustness to confounding were performed for the white-only group (n=797) and the deployed-only group (n=810). Principal components were calculated separately for each analysis group. In all a total of 12 GWAS were completed.

Equation 1. GWAS model

$$Y = \beta_1 * SNP + \beta_2 * sex + \beta_3 * age + \beta_4 * PC_1 + \beta_5 * PC_2 + \beta_6 * PC_3 + \beta_7 * PC_4 + \beta_8 * PC_5 + \beta_9 * PC_6 + \beta_{10} * PC_7 + \beta_{11} * PC_8 + \beta_{12} * PC_9 + \beta_{13} * PC_{10}$$

Primary results were filtered to require at least 5% observed alleles in both the affected and unaffected groups. SNPs were ranked by p-value and visualizations for top-ranked SNPs were generated using web-based LocusZoom¹⁰⁴. Results tables were

generated using the output from PLINK and associations reported in published work. Prior associations include eQTL and GWAS associations and were accessed on LDexpress, LDtrait, and the GWAS catalog between 11/22/2021 and 11/30/2021. LDexpress queries GTEx for specific SNPs and displays an output table of results related to that SNP with associated p-values and study descriptions¹⁰⁵. LDtrait queries the GWAS catalog for the requested SNP and displays a table of results related to the queried SNP, with associated p-values and study descriptions¹⁰⁶. The GWAS catalog was also queried directly for top SNPs and for the genes with which the top SNPs were annotated¹⁰⁷. Ancestry-based minor allele frequency reported in tables was gathered from NCBI dbSNP on 12/17/2021^{108,109}.

Heritability analysis

Heritability was calculated using GCTA v1.93.2 using the reml command on autosomes only. REML (restricted maximum likelihood) maximizes the joint likelihood of all error contrasts to estimate the variance components in mixed models¹¹⁰. REML assumes normal distribution of input variables. Minimum minor allele frequency (MAF) was set at 0.01 and covariates of age, sex, and 10 genetic principal components were used, as in the GWAS model. Heritability was calculated for each of the four case definitions across each of the three populations. Comparisons were made between definitions within each population group.

Gene and gene set analysis

Association of genes with GWI was tested using a multiple linear principal components regression model, computing the gene p-value using an F-test, implemented in MAGMA 1.6b⁹⁶. The default of 0.1% of the variance in the SNP data matrix was pruned before the principal components were used as predictors in the linear regression model. The method is fully described by de Leeuw et al., 2015, and utilizes individual-level data to minimize dependence on population-based summary statistics⁹⁶. The genome-wide gene association was based on individual-level data for the CDC Severe outcome. Gene annotations were based on GRCh37.3 and were downloaded from the MAGMA website: <https://ctg.cncr.nl/software/magma>. Age, sex, and 10 genetic PCs were used as covariates for the gene association calculations in MAGMA. Gene set association was computed by converting gene p-values to Z-values using a probit function. Z-values represent the strength of the association between each gene and GWI and are normally distributed. A gene set is associated if the genes in the gene set are jointly associated with GWI, calculated through a linear regression model. Gene set association was also implemented in MAGMA 1.6b. Gene set lists were downloaded from the GSEA website: <http://www.gsea-msigdb.org/gsea/downloads.jsp#msigdb>. Significance levels were adjusted for multiple testing using a Bonferroni correction.

Results

The sample size was 1061 after quality control. Table 3 shows the military and civilian characteristics of the Veterans in this study, stratified by fulfillment of the CDC Severe GWI case definition. Out of the 1,061 individuals, 247 were classified as fulfilling the CDC Severe GWI case definition and 814 were classified as not fulfilling the CDC Severe GWI case definition. Among those who fulfilled the CDC Severe GWI case definition, 72% were male, 45% were under 50 years old, and 19% were at least 60 years old. Black, non-Hispanic and Hispanic populations were overrepresented in the CDC Severe GWI group at 20% and 19%, compared to 17% and 9% in the study population, respectively. The white population is likewise underrepresented in the CDC Severe GWI group, at 46% compared to 65% in the general study population. Those with advanced degrees, incomes over \$100,000, and those in the Reserves only were underrepresented among the CDC Severe GWI group, while those who reported some college through Bachelor's degree, served in the Army, or served in Active Duty only were overrepresented among the CDC Severe GWI group.

Table 3. Demographics by CDC Severe GWI status

		All	CDC Severe GWI	
			Yes	No
	n	1061	247	814
Age at Survey (mean (standard deviation))		54.41 (8.02)	52.62 (6.93)	54.95 (8.25)

Age Group	40-49	399 (37.6)	113 (45.7)	286 (35.1)
	50-59	399 (37.6)	88 (35.6)	311 (38.2)
	60+	263 (24.8)	46 (18.6)	217 (26.7)
Sex	Male	813 (76.6)	177 (71.7)	636 (78.1)
	Female	248 (23.4)	70 (28.3)	178 (21.9)
Race/ Ethnicity	White, not Hispanic	687 (64.8)	113 (45.7)	574 (70.5)
	Black, not Hispanic	183 (17.2)	50 (20.2)	133 (16.3)
	Hispanic	99 (9.3)	47 (19.0)	52 (6.4)
	Other	57 (5.4)	25 (10.1)	32 (3.9)
	Unknown	35 (3.3)	12 (4.9)	23 (2.8)
Education	HS, GED, or less	92 (8.7)	21 (8.5)	71 (8.7)
	Some college to bachelor's degree	713 (67.2)	185 (74.9)	528 (64.9)
	Advanced degree	219 (20.6)	29 (11.7)	190 (23.3)
	Unknown	37 (3.5)	12 (4.9)	25 (3.1)
Household Annual Income	Under \$30,000	116 (10.9)	42 (17.0)	74 (9.1)
	\$30,000-\$59,999	241 (22.7)	70 (28.3)	171 (21.0)
	\$60,000-\$99,999	305 (28.7)	70 (28.3)	235 (28.9)
	\$100,000 or more	316 (29.8)	42 (17.0)	274 (33.7)
	Unknown	83 (7.8)	23 (9.3)	60 (7.4)
Deployed to the Gulf	Yes	810 (76.3)	212 (85.8)	598 (73.5)
	No	251 (23.7)	35 (14.2)	216 (26.5)

Military Branch	Air Force Only	117 (11.0)	15 (6.1)	102 (12.5)
	Army Only	481 (45.3)	136 (55.1)	345 (42.4)
	Marines Only	130 (12.3)	30 (12.1)	100 (12.3)
	Navy Only	173 (16.3)	30 (12.1)	143 (17.6)
	National Guard, Any	107 (10.1)	22 (8.9)	85 (10.4)
	Other or Unknown	53 (5.0)	14 (5.7)	39 (4.8)
Military Component	Active Duty Only	643 (60.9)	163 (66.0)	480 (59.3)
	Reserves and Active Duty	259 (24.5)	64 (25.9)	195 (24.1)
	Reserves Only	154 (14.6)	20 (8.1)	134 (16.6)

The four case definitions examined were CDC GWI, CDC Severe GWI, Kansas GWI, Kansas Symptom Criteria, which have 83.2%, 23.3%, 38.5%, and 69.7% prevalence in GWECB, respectively (Table 4). CDC GWI and Kansas Symptom Criteria both have more cases than controls in GWECB. Table 4 illustrates the frequency of fulfilling each case definition by deployment status.

Table 4. Case definitions by deployment

	Overall N (%)	Deployed N (%)	Not deployed N (%)
n	1061	810	251
Kansas GWI	408 (38.5)	324 (40.0)	84 (33.5)
Kansas Symptom Criteria	739 (69.7)	584 (72.1)	155 (61.8)

Kansas Exclusionary Criteria	431 (40.6)	327 (40.4)	104 (41.4)
CDC GWI	883 (83.2)	685 (84.6)	198 (78.9)
CDC Severe GWI	247 (23.3)	212 (26.2)	35 (13.9)

GWAS for all four of these case definitions under the same model were completed, with no results that show genome-wide significance ($p < 10^{-8}$). Figure 7 shows the QQ-plots and the Manhattan plots for the GWAS of CDC Severe GWI on the entire study population. To assess robustness to additional population stratification and an unmeasured environmental effect, GWAS were completed on each of the case definitions in the entire cohort, in the subgroup who self-identifies on the survey as “white, not Hispanic” and in the deployed subgroup, respectively. QQ-plots and Manhattan plots of the other 11 GWAS are shown in Appendix B.

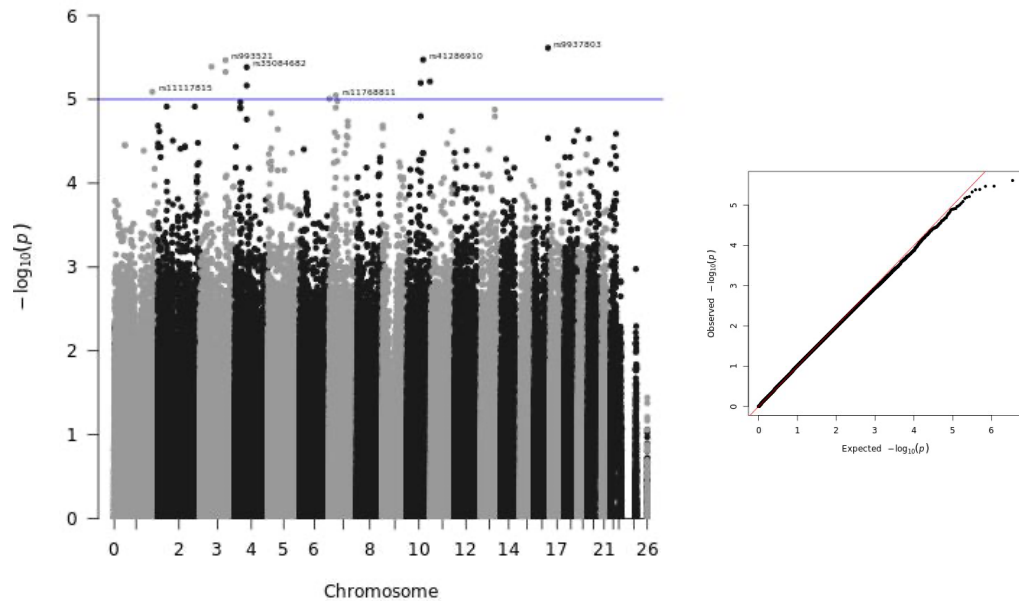


Figure 7. CDC Severe GWI Manhattan plot and QQ-plot

While no Bonferroni-corrected statistically significant SNPs were identified, several SNPs achieved a p-value less than 10^{-5} and were considered the top hits. Table 5 shows the SNPs with suggested significance ($p < 10^{-5}$) in the CDC Severe GWI GWAS among the entire study population, along with their genomic location, ORs, p-values, and prior associations. Top hits that are in high linkage disequilibrium are marked with superscripts. Most allele frequencies were stable across ancestry-based subpopulations, but rs993521, rs993522, rs11768811, and rs4722409 have flipped major/minor alleles in European populations compared to African populations.

Table 5. Associations with GWI: top SNPs

SNP	Chr: BP	European allele frequency	African allele frequency	p- value	OR	Gene	Prior associations
rs993 7803	16: 79, 108, 940	P(A)=0.58, P(G)=0.42	P(A)=0.65, P(G)=0.35	2.4e-06	1.7	WWOX (intron)	WWOX: height, MS, forced expiratory value, platelet count, hemoglobin concentration
rs412 86910	10: 92, 260, 229	P(G)=0.93, P(A)=0.07	P(G)=0.99 8, P(A)=0.00 2	3.4e-06	2.5	LINC02 653 (intron) HTR7	eQTL: Adipose- subcutaneous GWAS LINC02653: protein levels x insomnia interaction, protein QTL (liver), total amyloid

							(SNP×SNP interaction), Alzheimer's disease and age of onset
rs993521 ^a	3: 145, 758, 307	P(T)=0.31, P(G)=0.69	P(T)=0.54, P(G)=0.46	3.4e-06	0.57	LNCSR LR; PLOD2	eQTL: Cells-Cultured Fibroblasts, Thyroid; Muscle- Skeletal, Skin (suprapubic), Esophagus Mucosa GWAS PLOD2: Height, obesity related traits
rs35084682 ^b	4: 69, 363, 682	P(A)=0.69, P(G)=0.31	P(A)=0.74, P(G)=0.26	4.1e-06	1.7	TMPRS S11E	eQTL: Esophagus-Mucosa, GWAS linked SNPs: mean

							<p>corpuscular hemoglobin, triglycerides, triglyceride levels GWAS TMPRSS11E: urinary metabolite levels in chronic kidney disease, cholesterol levels, insulin-like growth factor 1 levels, ApoA1 levels, serum alkaline phosphate levels, LDL levels, APOB levels</p>
rs993 522 a	3: 145, 758, 564	P(C)=0.69, P(A)=0.31	P(C)=0.52, P(A)=0.48	4.7e-06	0.57	LNCSSR LR; PLOC2	eQTL: Cells- Cultured

							Fibroblasts, Thyroid; Muscle- Skeletal, Skin (suprapubic), Esophagus Mucosa GWAS PLOD2: Height, obesity related traits
rs475 0723	10: 1307 2894 0	P(T)=0.80, P(G)=0.20	P(T)=0.83, P(G)=0.17	6.1e-06	0.47	None	None
rs123 55465	10: 77, 893, 830	P(C)=0.58, P(A)=0.42	P(C)=0.62, P(A)=0.38	6.4e-06	1.7	LRMD A (intron)	GWAS LRMDA: balding type 1, male pattern baldness, forced expiratory volume, scoliosis, lung function vital capacity, alopecia

rs260 3152 b	4: 69, 367, 074	P(C)=0.67, P(T)=0.33	P(C)=0.52, P(T)=0.48	6.9e-06	1.7	UGT2B 17; TMPRS S11E	eQTL: TMPRSS11E/ Esophagus- Mucosa eQTL: UGT2B17/ Brain-Spinal cord GWAS SNP: triglycerides, mean corpuscular hemoglobin GWAS UGT2B17: Sex hormone- binding globulin levels adjusted for BMI, Sex hormone-binding globulin levels, Liver enzyme levels (alkaline phosphatase), LDL
--------------------	--------------------------	-------------------------	-------------------------	---------	-----	-------------------------------	---

							cholesterol levels, Serum alkaline phosphatase levels
rs111 17815	1: 217, 403, 176	P(G)=0.72, P(T)=0.28	P(G)=0.51, P(T)=0.49	8.1e-06	1.6	ESRRG	GWAS ESRRG: Photic sneeze reflex, hemoglobin concentration, hematocrit, red blood count
rs117 68811	7: 37, 199, 569	P(C)=0.84, P(T)=0.16	P(C)=0.39, P(T)=0.61	9.0e-06	1.7	None	None
rs472 2409	7: 947, 236	P(T)=0.60, P(C)=0.40	P(T)=0.16, P(C)=0.84	9.8e-06	1.7	ADAP1 ; COX19	eQTL: Adipose- Subcutaneous, Adipose-visceral, brain –cerebellum, breast, heart GWAS ADAP1/COX19: lymphocyte

							counts, ApoA1 levels, HDL levels, lymphocyte percent of white blood cells
--	--	--	--	--	--	--	---

The top ranked SNP, rs9937803, is in an intron of WWOX, close to a chromosomal fragile site (Table 5). The LocusZoom plot for this region (Figure 8) shows the cluster of SNPs in high LD co-located with rs9937803 between loci with high recombination rates. The population minor allele frequency (MAF) estimates are similar in African and European ancestry.

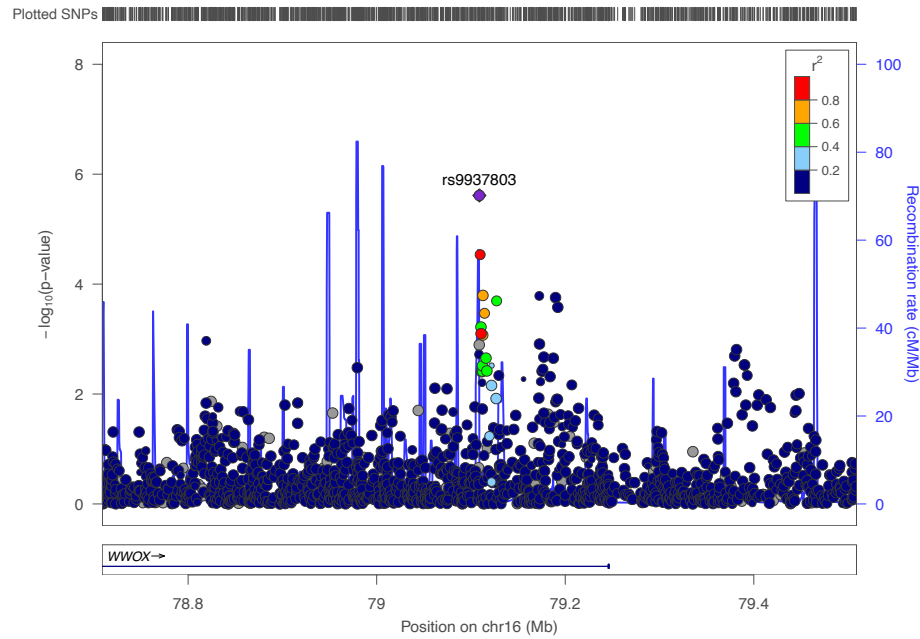


Figure 8. LocusZoom rs9937803

The top ranked SNP rs993521 is in high LD ($r^2=1.0$) with another top ranked SNP (rs993522). These two SNPs have been identified as being significantly associated with gene expression of the lncRNA sorafenib resistance in renal cell-- carcinoma associated (LNCSRLR) gene in cultured fibroblasts ($p=6.1 \times 10^{-12}$) and with expression of the *procollagen-lysine 2-oxoglutarate 5-Dioxygenase 2* gene (*PLOD2*) in skeletal muscle ($p=3.5 \times 10^{-7}$) (Table 5). *PLOD2* has also been previously associated with obesity-related traits. Figure 9 shows the LocusZoom plot for this region, highlighting the high LD region with low p-values close to the end of the *PLOD2* gene, between two loci with high recombination rates.

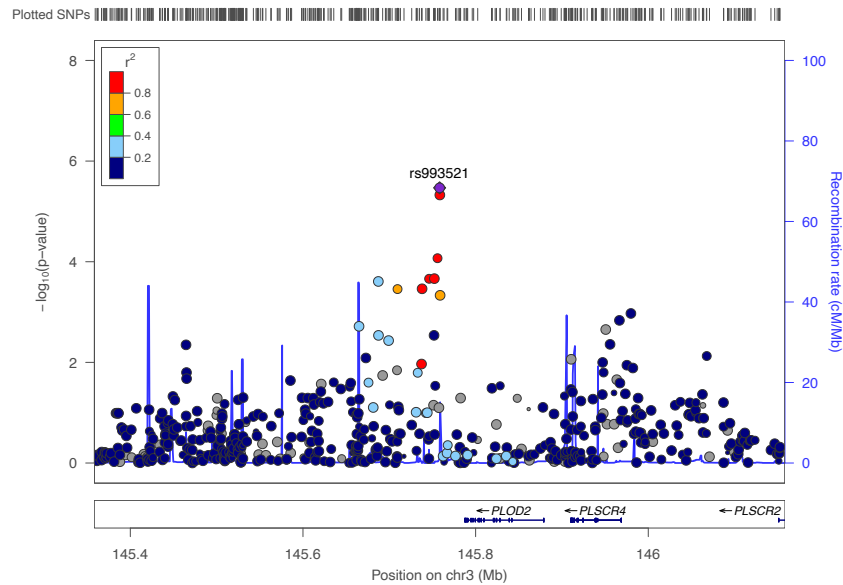


Figure 9. LocusZoom rs993521

Another top SNP identified was rs35084682, which was in high LD with rs2603151 (another top SNP); rs35084682 is identified in GTeX as having a significant association between the expression of *UGT2B17* in Brain-spinal cord (cervical c1) ($p=2 \times 10^{-5}$) cells (Table 5). These two SNPs also were identified in differential gene expression of *TMPRSS11E* in Esophagus Mucosa cells. The LocusZoom for this region is seen in Figure 10.

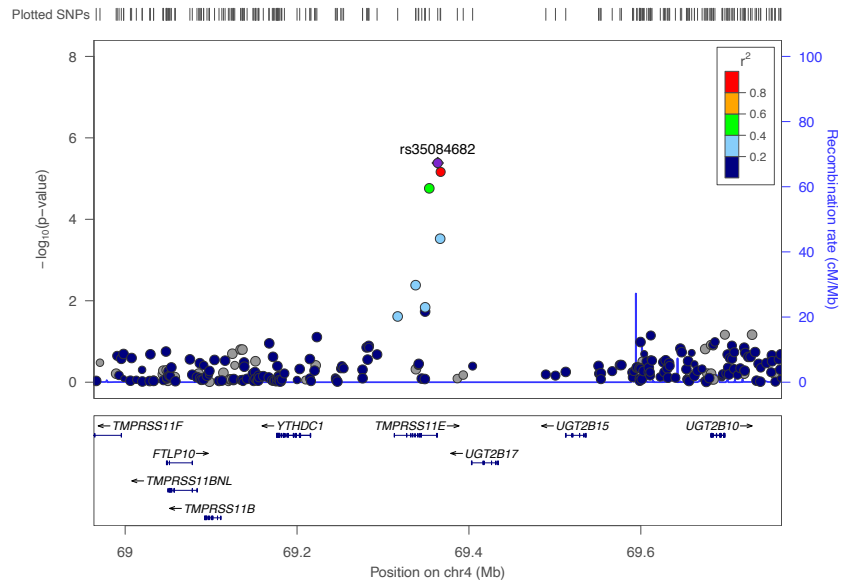


Figure 10. LocusZoom rs35084682

After Bonferroni correction, genome-wide significance was not reached by any gene in the association analysis. The top 10 genes in the gene-based association were SLC22A4, LANCL2, FAM172A, ZNF17, DPT, EPB41L1, ZNF516, ZNF804B, SCFD1, and ZNF394. SLC22A4 is the top associated gene and is an integral membrane transport protein that transports small organic cations. ZNF17, ZNF516, ZNF804B, and ZNF394 are all zinc finger proteins. The results show no clear association spike based on chromosomal location. There are 22 genes with a p-value of less than 0.001, which are labeled in Figure 11.

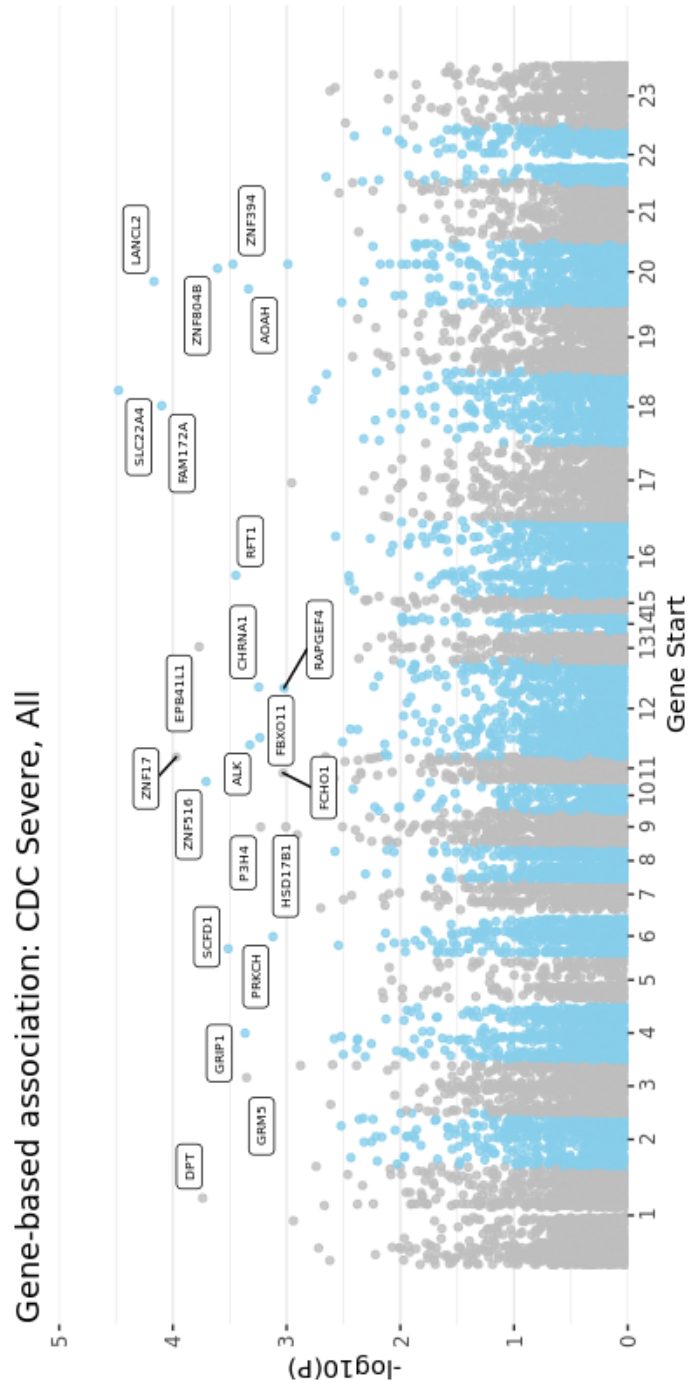


Figure 11. Gene-based association: CDC Severe, all

The gene set analysis produced no Bonferroni-corrected statistically significant results; however, the top ten ranked gene sets reproduce prior work. The top result, response to cadmium ion, has a beta of 0.43 and a p-value of 2.36×10^{-4} , associating the gene set with GWI. Table 6 displays the top ten gene set results, ranked. While some genes do overlap between gene sets (ex. number 7 is entirely encompassed by number 3), most gene sets shown do not contain the same genes. Response to cadmium ion, cranial skeletal system development, and cellular response to inorganic substance have a beta of less than 0.5, while eye pigmentation and positive regulation of autophagosome maturation have a beta over 1. There are no beta values that are extremely large or small, which would indicate potential false positive results due to the small sample size.

Table 6. Gene-set associations with GWI

	Full name	p-value	beta
1	Response to cadmium ion	0.00023594	0.43071
2	Regulation of response to interferon gamma	0.00030829	0.62566
3	Regulation of autophagosome maturation	0.00033461	0.81499
4	Postsynaptic modulation of chemical synaptic transmission	0.00075648	0.81254
5	Response to laminar fluid shear stress	0.00083192	0.84157
6	Eye pigmentation	0.00088436	1.1388
7	Positive regulation of autophagosome maturation	0.00092823	1.114

8	Cranial skeletal system development	0.00096495	0.3742
9	Cellular response to inorganic substance	0.0011139	0.20283
10	Neurotransmitter receptor internalization	0.0016388	0.5939

Discussion

These results are the first unbiased genetic study of GWI and replicate results from prior candidate studies. First, top associated SNPs included one in the *WWOX* gene, a site of high genomic instability, and one in the *LINC02653* gene, which has been previously associated with Alzheimer’s disease. Second, our gene set association study highlighted gene sets intrinsic to several theories regarding the biological basis of GWI. Third, our study had several limitations that may limit the scope of our conclusions. Finally, putting all these analyses together leads to further hypotheses to test in future work.

Of the four case definitions, CDC Severe GWI was the most heritable across the whole cohort. In the group that self-identified as White, not Hispanic, the most heritable definition was Kansas GWI. Appendix Figure 4 shows the heritability values with corresponding standard errors for all three groups. Since the sample size is small, the heritability point estimate may not be reliable, but the comparative measure was informative and solidified the *a priori* decision to prioritize the CDC Severe GWI case definition.

Our SNP-level GWAS results were not statistically significant but the rankings on the results from the three overlapping cohorts (All, White, non-Hispanic, and Deployed) were very similar. The top 11 SNP hits all had p-values less than 1×10^{-5} , which is considered to have suggestive statistical significance. Several of the SNPs were previously associated with differential expression of genes and several of the genes associated with top SNPs have been previously identified in GWAS of other traits, some of which may be related to GWI. The SNP-level and gene-set level results correspond and will be discussed together.

One major theory in the field is that Gulf War illness is caused by extreme immune dysregulation. An early paper comparing immune function in Gulf War veterans with civilians who have chronic fatigue syndrome (CFS) showed that while there is little evidence of immune dysregulation contributing to civilian CFS, Gulf War veterans with similar symptoms had significantly higher levels of IL-2, IL-10, INF- γ and TNF- α ¹¹¹. More recent work identifies immune-related SNPs that are associated with GWI¹¹², gene network remodeling¹¹³, and reduced HLA protection among veterans with GWI⁵⁶. Three of the top ten gene sets in the gene set association analysis are related to immune system regulation: regulation of response to interferon gamma (2nd), regulation of autophagosome maturation (3rd), and positive regulation of autophagosome maturation (7th). Interferon gamma has been found to induce autophagy in some types of immune cells and in cell lines, and has been implicated in the mediation of

autoimmune disease^{114,115}. Several of the top genes were previously implicated in white blood cell count, including SLC22A4 (the top gene) and ZNF516 (the 7th ranked gene).

Increased autophagy and interferon gamma have also been implicated in neuroinflammation in Gulf War illness and in associated neurodegenerative diseases. Interferon-gamma has been implicated in neuroinflammation and neuroinflammatory disease in several studies since 1990, as both protective and causal^{116–120}.

Autophagosomes engulf and deliver damaged cellular components and other organic waste to be recycled. Impaired or dysfunctional autophagosome maturation has been previously associated with neurodegenerative disorders, such as Alzheimer's and ALS^{121–124}. Without functional and properly regulated and matured autophagosomes, toxic proteins can accumulate in the neurons, causing neurodegeneration. In addition, one of the top SNP hits in the GWAS is in an intron in the *LINC02653* gene, which has also been previously associated with Alzheimer's disease. Neurotransmitter-specific gene sets (neurotransmitter receptor internalization and post-synaptic modulation of chemical synaptic transmission) are ranked 10th and 4th, respectively. Cranial skeletal system development, the 8th ranked gene set, could be related to repeated mild traumatic brain injuries (mTBI) have been linked to GWI to other neurodegenerative diseases (ex. ALS) that are experienced more by Gulf War veterans^{125–132}. It is possible that differences in the cranial skeletal system development changes how individuals experience mTBI, associating the gene set with the neurodegenerative symptoms experienced with GWI.

Several earlier papers point to GWI as a neuroimmune disease, bringing together the immune dysregulation and the neurodegenerative symptoms^{25,53,133,134}.

An overarching hypothesis that could bring together immune dysregulation, neurodegenerative symptoms, and response to exposures is genomic instability. Symptoms of genomic instability include immune system dysfunction, neural dysfunction, and gastrointestinal distress^{135,136}. Proposed as a “unifying hypothesis” in several papers between 2016-2021^{122,135,136}, genomic instability is linked to heavy metal exposure, particularly cadmium^{137,138}, and is expressed through a wide range of symptoms. Response to cadmium exposure and cellular response to inorganic substance are the 1st and 9th ranked gene sets and have no overlapping genes. Cadmium is one of the heavy metals that is thought to induce genomic instability, mainly by interfering with cell cycle checkpoints, DNA repair mechanisms, and apoptosis^{137,138}. This genomic instability can be observed as a higher mutation rate during cellular replication and can cause cancer or other diseases¹³⁷. Also directly related to genomic instability is the top SNP hit, rs9937803, which is in an intron of WWOX. WWOX is the site of FRA16D, a chromosomal fragile site that, as a hotspot for somatic and germline mutations, is both a contributor to and a target of genomic instability^{139,140}. Autosomal maturation has also been linked to lowering genomic instability in cancer research^{141,142}, in addition to the earlier-cited connection with neurodegenerative symptoms. Immune

dysregulation, neurodegeneration, and Gulf War-specific exposures coalesce into the genomic instability hypothesis.

There are several limitations to this work. First, the dataset is small for genome-wide association studies. It is, however, a large dataset for GWI and contains agreed-upon GWI indicators. The Million Veterans Program (MVP) has a larger dataset with about 40,000 Gulf War veterans for future work. Second, we are unable to perform replication studies without a second dataset. When a GWAS is completed in MVP, these findings should help separate legitimate from spurious results (e.g., type I errors). We were additionally unable to separate individuals by ancestry for genetic analysis, as the dataset was not large enough. This can also contribute to the generation of type I errors. We have assessed our results for the possibility of population structure driving some results and have mitigated this as much as possible. We have adjusted for genetic principal components and analyzed the allele frequency for all top SNP hits.

A few themes repeat in the GWAS, gene association study, and gene-set association study. First, genomic instability, usually linked with cancer, is known to cause a higher rate of somatic mutation. A higher rate of somatic mutation could account for the diversity of symptoms in GWI and the lack of consistent findings. This unbiased analysis has aggregated support for prior work in this area through the consistently high ranking of SNPs and gene sets identified as contributing to genomic instability. Another common theme has been neurodegeneration. While mTBIs were

more well-discussed in post-9/11 era veterans, many veterans do serve in more than one service era and mTBIs are widely underreported¹⁴³. Links between top results and ALS, Alzheimer's, and dementia are identified again and again. Between susceptibility to chemical exposures, autophagosome activity, autoimmune regulation, and neurotransmitter pathways, neurodegenerative diseases and GWI appear to share some genetic risk factors. These two repeated themes prompt the development of hypotheses to test in future work.

Conclusions

These unbiased genome-wide association results add supporting evidence to prior GWI findings. First, the CDC Severe GWI case definition is both the most heritable and the most highly associated with deployment. Second, genomic instability as a unifying theory seems more plausible than ever, with the added support from our unbiased genomic study. Many top hits appear to support the idea of genomic instability as a biologic underpinning of GWI, and this could also explain the diversity of symptoms across the GWI population. Third, neurodegeneration is a major issue for Gulf War veterans, and it is possible that research on other specific neurodegenerative diseases could help Gulf War veterans. Finally, future directions of research might include more genome-wide work in larger datasets, perhaps in the MVP dataset, to identify more ancestry-specific and GWI subtype-specific results.

4. Biologically based Candidate Studies

This chapter consists of a published manuscript on which I am first author⁹. I completed the data management, data cleaning, statistical analysis, table generation, and figure generation. I also drafted the manuscript. The other authors were responsible for gaining and maintaining project funding, data generation, and editing and revising the manuscript. I also appreciate the mentorship, guidance, and idea generation of several coauthors in the process of drafting the manuscript.

9 **J. Vahey**, E. J. Gifford, K. J. Sims, B. Chesnut, S. H. Boyle, C. Stafford, J. Upchurch, A. Stone, S. Pyarajan, J. T. Efirid, C. Williams, and E. R. Hauser, “Gene-Toxicant Interactions in Gulf War Illness: Differential Effects of the PON1 Genotype”. *Brain Sci* 2021, Vol 11, Page 1558. 2021;11(12):1558. doi:10.3390/BRAINSCI11121558

Gene-toxicant Interactions in Gulf War Illness: Differential Effects of the PON1 genotype.

Introduction

Gulf War Veterans began reporting chronic systemic symptoms almost immediately after returning from deployment to the Gulf in support of the 1990-1991 Gulf War^{3,5,83,144}. These symptoms have only increased, in both severity and frequency, over the past 30 years and have been defined as Gulf War illness (GWI)^{3,6,11,20,21,145,146}. The 2016 Institute of Medicine (IOM) report on GWI indicated great heterogeneity in the expression of GWI across the Veteran population, suggesting heterogeneity in genetic susceptibility, environmental exposures, and/or their interactions⁴. There are still important gaps in our understanding of genetic or environmental risk factors for GWI, especially in how these risk factors may interact to increase susceptibility to GWI. True biological understanding of GWI is necessary to discover and design better treatment options for Veterans with GWI.

Prior work in occupational hazards has shown that organophosphates, such as pyridostigmine bromide (PB) pills and pesticides used in the Gulf War, bind and phosphorylate acetylcholinesterase, the enzyme that breaks down acetylcholine. Nerve cells release acetylcholine to pass a signal from one cell to the next; the cell receiving the signal detects the presence of acetylcholine through various receptors, stimulating muscle contraction³³. When acetylcholinesterase is competitively inhibited by another chemical bound to the active site of the enzyme, not enough free acetylcholinesterase

remains to break down the acetylcholine signal. The ensuing accumulation of acetylcholine causes overstimulation of nerve cells and the neuromuscular system, causing systemic long term damage and, in extreme cases, death^{32,34,35}. While irreversible inhibition of acetylcholinesterase is often fatal, even reversible inhibition of acetylcholinesterase temporarily prevents the breakdown of acetylcholine, causing this overstimulation³⁶. Sarin is an irreversible inhibitor of acetylcholinesterase; PB pills were designed to be reversible inhibitors of acetylcholinesterase to preemptively defend soldiers from sarin attacks. The symptoms of long term damage reported in occupational hazards research into organophosphate poisoning are similar to the symptoms of GWI: headaches; pain in muscles, joints, or bones; rashes or itchy skin; blurred vision; dizziness; diarrhea; confusion; respiratory and circulatory depression^{34,37}. In addition, rodent models for GWI have been developed using organophosphate pesticides, PB, and chronic stress or corticosterone³⁸⁻⁴¹. Coadministration of PB and pesticides, along with stress, led to delayed cognitive impairment in the rodent models; affected mice had decreased spatial memory and increased depressive behavior⁴².

Several other enzymes are capable of either breaking down acetylcholine while acetylcholinesterase is inhibited or detoxifying these organophosphates.

Butyrylcholinesterase, carboxylase, and paraoxonase are among these enzymes and could modulate the effects of toxic exposure⁴³. The concentration of these enzymes in the blood stream, where they can quickly detoxify organophosphates, varies from

species to species⁴³. Due to this variation across species, human observational studies are key to understanding the specific human biochemistry necessary to detoxify these chemicals.

The genetic analysis of these candidate genes has been limited to association studies of a small number of functional polymorphisms in each gene. *ACHE*, *BCHE*, and *PON1* produce acetylcholinesterase, butyrylcholinesterase, and paraoxonase, respectively. *SOD1* is one of the main genes associated with amyotrophic lateral sclerosis (ALS), which has been associated with GW service^{94,147-150}. While there are few common functional variants in human acetylcholinesterase, one SNP, rs1799805, is responsible for the YT blood type. The two versions of acetylcholinesterase generated by the variation of the rs1799805 locus do not differ in catalytic efficiency, they form different antigens that make up this YT blood type. As the largest source of variation in the *ACHE* gene, rs1799805 was chosen to represent *ACHE* in the candidate study. The *BCHE* and *PON1* genes contain common variants across the global population that are known to code for protein variants that modulate the catalytic activity of these two enzymes. The atypical butyrylcholinesterase variant, caused by the genetic variant rs1799807, is known to cause slower metabolism of succinylcholine, a common neuromuscular blockade used in surgery¹⁵¹. Specific to GWI, Steele et al., 2015 show association of butyrylcholinesterase variants with GWI when exposure to PB pills is reported in a small sample of US GW Veterans²². Paraoxonase is known to assist in

organophosphate metabolism, as well as protection against oxidation and inflammation¹⁵². Haley et al., 1998, found an association between low paraoxonase activity and Gulf War related symptoms⁵⁴. The functional variant discussed by Haley et al., rs662, has a high activity phenotype (A allele at this locus leads to the Q version of the protein) and a low activity phenotype (G allele at this locus leads to the R version of the protein). The common SNP rs662 has also been associated with increased risk for cardiovascular disease and ALS^{58,153,154}. *SOD1*, *ACHE*, *BCHE*, and *PON1* are the genes of interest in this gene-by-environment interaction study, along with the exposures of pesticides and PB pills. Our hypothesis is that Gulf War illness case status and features of Gulf War illness and, specifically, the CDC Severe GWI case definition will be associated with the interaction between these genes and PB pill or pesticide exposures.

Materials and Methods

Introduction to GWECB

The Gulf War Era Cohort and Biorepository (GWECB) consists of DNA, electronic medical records, and survey data from individuals who served in the United States Armed Forces during the 1990-1991 Gulf War⁵⁹. Data collected includes self-reported exposures, exposure lengths, symptoms, symptom severity ratings, clinical conditions, and deployment locations. Women, Veterans of color, and deployed Veterans were oversampled, and Veterans did not have to use the Veterans Affairs system for healthcare to be eligible for inclusion.

GWI Phenotypes

Veterans have been classified as meeting or not meeting the case definition using the CDC, CDC severe, Kansas full case definition, and Kansas symptom criteria²⁰. Case status was determined using a deterministic SAS algorithm and basic survey data cleaning was also performed in SAS 9.4, as described in Vahey et al. and Gifford et al.^{20,80}.

Eleven outcomes were used for the regression analysis: the CDC Severe GWI Case Definition, the three CDC Severe symptom domains, Kansas Symptom Criteria, and the six Kansas symptom domains. An additional variable derived from the GWI case algorithm is the Kansas exclusionary conditions indicator. This variable identifies veterans with at least one of 21 exclusionary conditions, such as diabetes, heart disease, and cancer. Fulfilling the Exclusionary Criterion precludes an individual from being considered a GWI case. The CDC Severe and Kansas GWI case definitions are known to be heterogeneous; it is possible that multiple traits with different underlying etiologies make up GWI, complicating analysis. Thus, studying the symptom domains separately may isolate more homogeneous subphenotypes. Each symptom domain in the CDC Severe and Kansas case definitions were considered as individual outcomes, as in prior studies^{22,54}. Previous analyses of deployment in GWECB demonstrated a strong and consistent association of deployment with the CDC Severe Case Definition and thus CDC Severe will serve as the primary outcome for this analysis²⁰.

Gulf War Exposures

Veterans deployed to the Gulf during the 1990-1991 Gulf War were asked to report on their exposures. The survey asks for both a yes/no response and a “number of days exposed” response for 11 exposure questions, which can be found in Appendix B. Ambiguous values were coded based on previous work²⁰: number of days exposed was set to zero for individuals who said they did not experience an exposure, regardless of their secondary answer. Exposures used were “Took pyridostigmine bromide (anti-nerve agent pills)”, “Used pesticide cream or liquid on your skin”, “Wore a uniform treated with pesticides”, and “Used insect baits / no-pest strips in your living area”. The PB exposure variable used was the number of days each individual reported being exposed to PB pills; the options were “No, none”, “Yes, 1-6 days”, “Yes, 7-10 days”, and “Yes, 31 days or more”. The pesticide exposure variable combined the number of days reported for all three pesticide variables listed above, based on the minimum number of days per option (i.e., “Yes, 1-6 days” was counted as 1 day). Less than 31 days combined exposure was a short exposure, 31-60 days combined exposure was a medium exposure, and more than 60 days combined exposure was a long exposure.

Genotypes and genetic data cleaning

Genotyping was done on an Illumina Omni2.5-8v1.4 microarray at the Pharmacogenomics Analysis Laboratory (PAL) in the Central Arkansas Veterans Healthcare System. Plink 1.9 was used for genetic data cleaning and analysis¹⁰³. A total

of 2,372,461 SNPs were genotyped; 82,369 SNPs were removed due to missingness greater than 1% and 512,116 SNPs were removed due to minor allele frequency (MAF) less than 1%, leaving 1,777,976 SNPs. From those 1,777,976 SNPs, 3 SNPs were chosen for Tier 1 candidate SNP testing and 371 SNPs were chosen for Tier 2 candidate SNP testing, based on their proximity to genes of interest. Individuals who returned surveys but who did not provide a specimen for genetic testing were excluded from analysis. Beginning with 1,343 samples, replicates were removed, keeping the most informative sample for each individual. Two individuals were removed because of too many missing genotypes and 13 individuals were removed due to perceived relatedness, leaving 1,260 Veterans for genetic analysis. As only Veterans deployed to the Gulf were asked to report on their exposures, those who were deployed elsewhere and those who were not deployed were removed from the analysis. A total of 810 GW-deployed Veterans were in the final analytic dataset after genotype and phenotype data cleaning, as seen in Figure 12.

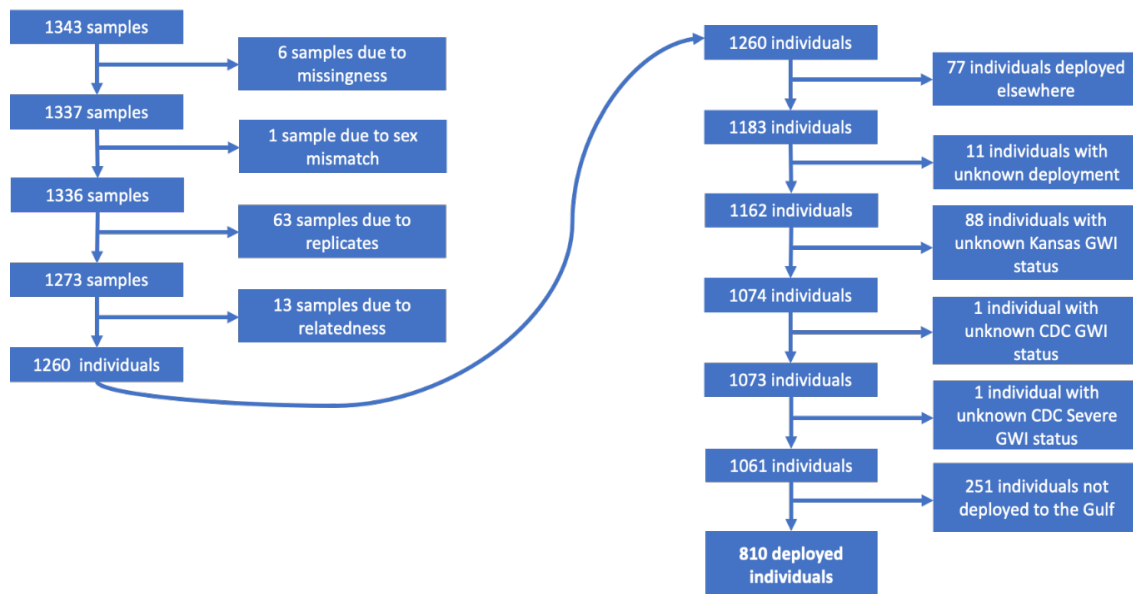


Figure 12. Block diagram: individuals in the analytic dataset

Statistical Analysis

We fit a gene/environment interaction model to estimate the effect and statistical significance of specific candidate gene/toxicant interactions. We used logistic regression as implemented in PLINK 1.9 to test the effect of genetic variants and gene-by-exposure interaction¹⁵⁵. As discussed above, the primary outcome was the CDC Severe phenotype. Analyses of other GWI case phenotypes and the individual symptom domain phenotypes were considered secondary. Given the complexity of the GWI phenotypes and the correlation among individual symptom domains, as well as their relationship to GWI outcomes, we did not adjust for multiple comparisons across the phenotypes but rather reported the p-values and effect sizes for each analysis.

Ten genetic principal components, representing population structure, along with age and sex were used as covariates for the basic model. The fully adjusted model

included age, sex, education, income, and a binary indicator of the Kansas Exclusionary Criterion as an indicator of presence or absence of medical conditions. Because the Kansas exclusionary criterion is part of the full Kansas GWI case definition, the Kansas symptom criteria indicator was used as an outcome rather than the full Kansas GWI case definition in the fully adjusted model. Equation 1 outlines the basic model and Equation 2 outlines the fully adjusted model. We tested B_3 using the standard allele test in an additive genetic model as implemented in PLINK.

Equation 2. Basic additive genetic model

$$Y = \beta_0 + \beta_1 * SNP + \beta_2 * E + \beta_3 * SNP * E + \beta_4 * age + \beta_5 * sex$$

Equation 3. Fully adjusted additive genetic model

$$Y = \beta_0 + \beta_1 * SNP + \beta_2 * E + \beta_3 * SNP * E + \beta_4 * age + \beta_5 * sex + \beta_6 * education + \beta_6 * income + \beta_7 * Kansas_Exclusion$$

Interaction trends were tested for significance using the Cochran-Armitage test for trend, according to the *prop_trend_test()* function in the *rstatix* package in R version 3.61. Locus zoom plots were generated using legacy LocusZoom.org¹⁰⁴. Tier 1 SNPs were tested at $p = 0.05$, while Tier 2 SNPs were tested at $p = 0.00013$ using a Bonferroni correction for the 371 SNPs in Tier 2. SNPs were prioritized based on p-value and representation across phenotypes.

SNP Selection

These models were run only on the SNPs within 50kb of the candidate. Both models were run for each of the 11 outcomes, both exposures, and the basic and

extended covariate lists. For inclusion in these models, we identified two kinds of a priori SNPs based on published evidence for gene-toxicant interactions in GWI. Tier 1 SNPs in Table 7 have been previously identified in genetic analysis of GWI. Tier 2 SNPs, enumerated in Table 7, were chosen to cover all common variation in the genes of interest.

Table 7. Candidate SNPs

Tier 1 SNPs: <i>BCHE</i> , <i>PON1</i> , and <i>ACHE</i> from prior studies				
SNP	Gene	Variant	MAF	Citations
rs1799807	<i>BCHE</i>	Atypical (A); lower catalytic rate; succinylcholine susceptibility	0.020	Zhu et al. 2020 ¹⁵¹ , Goodall 2004 ¹⁵⁶ , Steele et al. 2015 ²²
rs662	<i>PON1</i>	192Q/R; functional variant of Pon1 Modifier for risk of sporadic ALS	0.396	Dardiotis et al. 2018 ¹⁵² , Davies et al. 1996 ¹⁵⁷ , Haley et al. 1999 ⁵⁴ , Verde et al. 2019 ¹⁵³
rs1799805	<i>ACHE</i>	H322N; Yt blood group;	0.037	Shapira et al. 2000 ¹⁵⁸
Tier 2 SNPs: remaining SNPs in <i>BCHE</i> , <i>PON1</i> , <i>ACHE</i> , and <i>SOD1</i>				

All SNPs in GWECB that are within 50kb of the gene	Gene	Location	Number of SNPs
	<i>BCHE</i>	Chr 3. 165490692-165555211 (+/- 50kb)	84
	<i>PON1</i>	Chr 7. 94926988-94953844 (+/- 50kb)	178
	<i>ACHE</i>	Chr 7. 100487615-100494614 (+/- 50kb)	59
	<i>SOD1</i>	Chr 21. 33032006-33041244 (+/- 50kb)	53

Results

Demographics and Outcomes

Veterans who fulfill the CDC Severe GWI criteria and those who do not are similarly distributed by sex, age group, and OEF/OIF deployment. Those who were Black or Hispanic, in the army, have a lower household income, and deployed active duty were more highly represented among those who fulfill the CDC Severe GWI criteria than those who do not fulfill the CDC Severe GWI criteria (Table 8). Those who earned advanced degrees and who were in the reserves were underrepresented among those who fulfill the CDC Severe GWI criteria. Demographics by pesticide and PB exposure times are shown in Appendix Figures 4 and 5, respectively.

Table 8. Demographics in deployed veterans, by CDC Severe GWI status

Demographics	All (n=810)	CDC Severe GWI	
		Yes (n=212)	No (n=598)

Sex	Female	180 (22%)	55 (26%)	125 (21%)
	Male	630 (78%)	157 (74%)	473 (79%)
Age Group	40-49	321 (40%)	98 (46%)	223 (37%)
	50-59	296 (37%)	74 (35%)	222 (37%)
	60+	193 (24%)	40 (19%)	153 (26%)
Race/ Ethnicity	White, non-Hispanic	529 (65%)	101 (48%)	428 (72%)
	Black, non-Hispanic	137 (17%)	43 (20%)	94 (16%)
	Hispanic	76 (9%)	37 (17%)	39 (7%)
	Other or missing	68 (8%)	31 (15%)	37 (6%)
Income	Under \$30,000	84 (10%)	33 (16%)	51 (9%)
	\$30,000-59,999	179 (22%)	60 (28%)	119 (20%)
	\$60,000-99,999	239 (30%)	65 (31%)	174 (29%)
	\$100,000+	246 (30%)	35 (17%)	211 (35%)
Education [^]	High school, GED, or less	75 (9%)	17 (8%)	58 (10%)
	Some college, or associate's or bachelor's degree	544 (67%)	161 (76%)	383 (64%)
	advanced degree	166 (20%)	24 (11%)	142 (24%)
Service	Army only	366 (45%)	114 (54%)	252 (42%)
Branch [^]	Navy only	139 (17%)	28 (13%)	111 (19%)

	Air Force only	80 (10%)	12 (6%)	68 (11%)
	Marine Corps only	105 (13%)	26 (12%)	79 (13%)
	National Guard, any	81 (10%)	21 (10%)	60 (10%)
Military	Active Duty	470 (58%)	136 (64%)	334 (56%)
Component	Both Active Duty and Reserves	208 (26%)	59 (28%)	149 (25%)
	Reserves Only	127 (16%)	17 (8%)	110 (18%)
OEF/OIF deployment	Yes	187 (23%)	38 (18%)	149 (25%)
	No	614 (76%)	168 (79%)	446 (75%)
Kansas Exclusionary Criterion	Yes	327 (40%)	112 (53%)	215 (36%)
	No	483 (60%)	100 (47%)	384 (64%)
rs1799807 (<i>BCHE</i>)	Homozygous reference (AA)	773 (95%)	206 (97%)	567 (95%)
	Heterozygous (AG)	36 (4%)	6 (3%)	30 (5%)
	Homozygous Atypical Variant (GG)	0 (0%)	0 (0%)	0 (0%)
rs1799805 (<i>ACHE</i>)	Homozygous reference (CC)	746 (92%)	199 (94%)	547 (91%)
	Heterozygous (CA)	63 (8%)	13 (6%)	50 (8%)
	Homozygous YT variant (AA)	0 (0%)	0 (0%)	0 (0%)
rs662 (<i>PON1</i>)	Homozygous high function (AA)	329 (41%)	67 (32%)	262 (44%)

	Heterozygous (AG)	340 (42%)	100 (47%)	240 (40%)
	Homozygous low function (GG)	141 (17%)	45 (21%)	96 (16%)
^ Service Branch and Education do not add to 100% as 'missing' or 'unknown' were not presented.				

Across the 810 deployed Veterans, 26% are categorized as fulfilling the CDC Severe GWI criteria. CDC Severe case status is associated with PB and pesticide exposure (Table 9). Table 9 shows that 36% of those who report long PB exposure (n=149) and 33% of those who report long pesticide exposure (n=156) report symptoms consistent with CDC Severe GWI compared to 18% and 19% of those who report no PB or no pesticide exposure, respectively. Across all outcomes except for the Kansas Exclusionary Criterion, a larger proportion of those who were exposed to either pesticides or PB pills fulfilled the symptom criteria compared to those in the overall deployed group. Appendix tables 6 and 7 show the percentage of individuals in each exposure group who fulfill all of the components to both the CDC Severe and Kansas case definitions.

Table 9. Exposure Time by CDC Severe GWI Case Status

Exposure	Length of time	All	CDC Severe GWI	
			Yes	No
All		810	212 (26%)	598 (74%)

PB pill exposure	None	311	55 (18%)	256 (82%)
	Unclear or missing	138	41 (30%)	97 (70%)
	1-6 days	98	27 (28%)	71 (72%)
	7-30 days	114	35 (31%)	79 (69%)
	31+ days	149	54 (36%)	95 (64%)
Pesticide exposure	None	346	67 (19%)	279 (81%)
	Unclear or missing	106	37 (35%)	69 (65%)
	1-30 days	78	19 (24%)	59 (76%)
	31-62 days	124	37 (30%)	87 (70%)
	63+ days	156	52 (33%)	104 (67%)

Tier 1 SNPs

The results for all outcomes for the Tier 1 candidate SNPs are shown for PB pill exposure and pesticide exposure in both the primary outcome (CDC Severe GWI) and the secondary outcomes (Table 10). Candidate SNP results were adjusted for age, sex, Kansas Exclusionary Criterion, education, and income. Only one SNP, rs662, showed a statistically significant interaction with the CDC Severe GWI outcome ($p=0.049$). Among the secondary outcomes, the interaction between rs662 and pesticides has a p-value of 0.029 in the fully adjusted model of the Kansas Gastrointestinal symptom domain. The other Tier 1 interaction below $p=0.05$ was rs1799805 with PB pill exposure for the Kansas skin symptom domain. The SNP rs1799805 is the variant that controls Yt blood type,

which is not thought to affect catalytic efficiency. The *BCHE* and *ACHE* Tier 1 variants had low minor allele frequencies: 0.022 and 0.039, respectively, likely resulting in lower power to detect differences across groups for these two variants.

Table 10. Tier 1 Candidate SNP results

outcome	Exposure	<i>ACHE</i>		<i>BCHE</i>		<i>PON1</i>	
		rs1799805		rs1799807		rs662	
		OR	p	OR	p	OR	p
CDC Severe GWI	PB	1.1 1	0.725	0.9 3	0.84 8	1.2 2	0.049 *
CDC Severe: Fatigue	PB	1.5	0.268	1.6 3	0.25 4	1.2 1	0.105
CDC Severe: Mood/Cognitive	PB	1.0 5	0.856	1.1 3	0.72 1	1.1 3	0.201
CDC Severe: Musculoskeletal	PB	0.9 9	0.953	0.8 6	0.63 9	1.1 3	0.202
Kansas GWI symptom criteria	PB	0.9 3	0.821	1.2 8	0.59 9	0.9 9	0.923
Kansas: Fatigue	PB	0.9 3	0.838	1.0 1	0.97 9	1.0 3	0.814
Kansas: Pain	PB	1.2 6	0.443	1.6 1	0.35 6	1.1 6	0.225
Kansas: Gastrointestinal	PB	0.9 6	0.883	1.1 8	0.61 4	0.9 6	0.663
Kansas: Respiratory	PB	1.1 5	0.581	0.9 5	0.86 8	1.1 8	0.08
Kansas: Mood/Neurological/ Cognitive	PB	0.5 8	0.251	1.0 9	0.85 5	1.2 1	0.278
Kansas: Skin	PB	0.5 9	0.042 *	1.3	0.40 8	1.0 1	0.897
CDC Severe GWI	Pesticides	0.6 7	0.179	0.7 8	0.54 8	1.0 7	0.485
CDC Severe: Fatigue	Pesticides	0.8 9	0.739	1.2 7	0.58 7	1.1 4	0.235
CDC Severe: Mood/Cognitive	Pesticides	0.6 5	0.123	1.1 7	0.67 6	1	0.958

CDC Severe: Musculoskeletal	Pesticides	0.61	0.063	0.95	0.887	1.04	0.684
Kansas GWI symptom criteria	Pesticides	0.6	0.128	2.11	0.189	1.05	0.66
Kansas: Fatigue	Pesticides	0.76	0.484	0.63	0.281	0.97	0.815
Kansas: Pain	Pesticides	0.67	0.151	1.47	0.362	1.17	0.185
Kansas: Gastrointestinal	Pesticides	0.84	0.49	1.11	0.747	0.82	0.029*
Kansas: Respiratory	Pesticides	0.66	0.103	1.39	0.32	1	0.97
Kansas: Mood/Neurological/Cognitive	Pesticides	1.09	0.893	1.72	0.353	1.11	0.562
Kansas: Skin	Pesticides	0.65	0.098	1.28	0.451	1	0.96
* Statistically significant with $p < 0.05$.							

Figure 13 illustrates the unadjusted CDC Severe GWI frequencies across rs662 genotype (AA- high activity allele homozygote, AG- heterozygote and GG- low activity allele homozygote) by PB pill length of exposure group. Significant trends were observed for longer lengths of exposure with higher rates of GWI were seen in the *PON1* rs662 GA and GG genotypes, with p-values for the simple test for trend of 0.0004 and 0.0084, respectively. This trend was not observed in the AA group ($p = 0.179$).

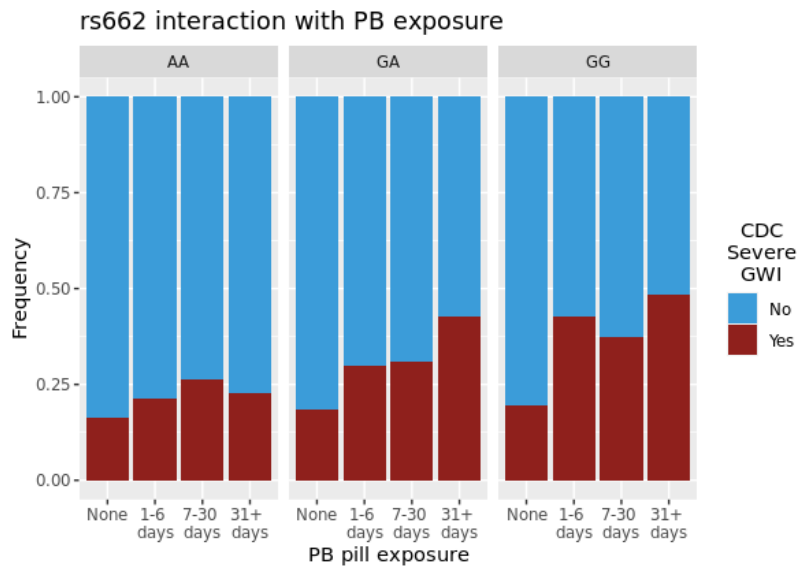


Figure 13. Proportion of individuals fulfilling CDC Severe GWI within each genotype-exposure group

Candidate Genes and Tier 2 SNPs

SNPs in the four candidate genes were plotted by genomic location using locusZoom, highlighting the SNP with the highest interaction $-\log_{10}$ p-value (i.e., lowest p-value) for each gene and outcome. The locusZoom plots for *ACHE*, *BCHE*, *PON1*, and *SOD1* for the primary outcome, CDC Severe GWI, are shown in Figure 14. No results for any of the CDC Severe symptom domains were statistically significant using a Bonferroni correction ($p < 0.00013$). The lowest p-value for the interaction term with PB exposure is seen in *PON1* for the SNP rs2299260 ($p = 0.005$) (Table 11).

CDC Severe: p-values for SNPxPB interaction

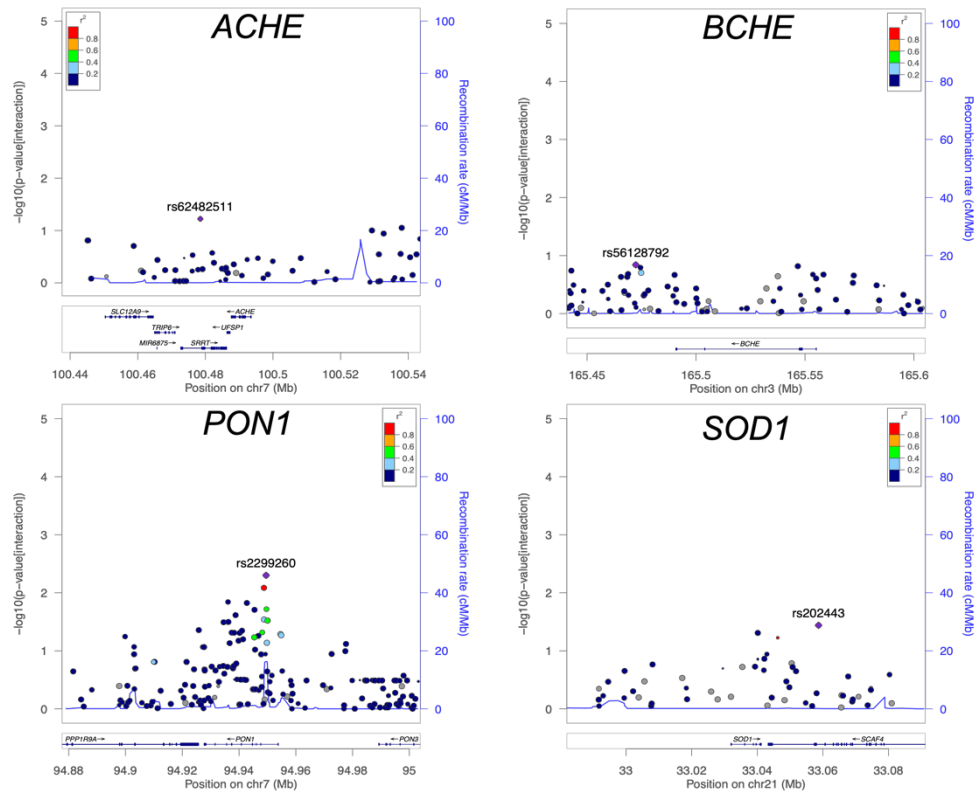


Figure 14. LocusZoom plots for CDC Severe GWI, Tier 2

Across all outcomes, only three interactions were statistically significant: all were interactions with PB exposure, and all three SNPs were in high linkage disequilibrium with one another, across all ancestry groups. All three statistically significant interactions involved SNPs in the *PON1* gene and appeared significant only in associations with the Kansas Respiratory symptom domain. All three top SNPs are in the *PON1* gene (Figure 15 and Appendix Figure 8) and are in high LD ($r^2=1.0$), regardless of ancestry group, so are represented in Table 11 by rs3917545. The other two SNPs presented in Table 11 are the top outcomes in CDC Severe GWI.

Table 11 illustrates top results for the Tier 2 SNP/PB interactions among the 384 tested SNPs across all GWI outcomes ($p = 0.00013$). From left to right, the top SNPs associated with: respiratory symptoms in interaction with PB pills (rs3917545), CDC Severe GWI in interactions with PB pill exposure (rs2299260), and CDC Severe GWI in interactions with pesticide exposure (rs62567349). Significant interactions are starred. Top interactions in each candidate gene are bolded.

Table 11. Top Results for the Tier 2 SNP/PB and SNP/Pesticide Interactions

Outcome	Exposure	<i>PON1</i> rs3917545		<i>PON1</i> rs2299260		<i>PON1</i> rs62467349	
		OR	p	OR	p	OR	p
CDC Severe GWI	PB	1.45	3.2E-02	0.68	5.0E-03	1.39	5.5E-02
	Pesticides	1.45	2.2E-02	0.89	3.7E-01	1.54	8.6E-03
CDC Severe: Fatigue	PB	1.54	2.5E-02	0.64	7.6E-03	1.52	3.1E-02
	Pesticides	1.41	5.7E-02	0.97	8.6E-01	1.38	7.5E-02
CDC Severe: Mood/Cognitive	PB	1.25	1.7E-01	0.82	1.2E-01	1.23	2.1E-01
	Pesticides	1.05	7.7E-01	0.94	5.9E-01	1.07	6.5E-01
CDC Severe: Musculoskeletal	PB	1.44	2.8E-02	0.76	3.0E-02	1.38	5.0E-02
	Pesticides	1.19	2.4E-01	0.86	2.2E-01	1.26	1.4E-01
Kansas GWI symptom criteria	PB	1.23	3.2E-01	0.99	9.7E-01	1.23	3.2E-01
	Pesticides	1.30	1.8E-01	0.85	2.7E-01	1.37	1.2E-01

Kansas: Pain	PB	1.28	2.3E-01	0.73	3.5E-02	1.31	1.9E-01
	Pesticides	0.99	9.5E-01	1.05	7.3E-01	1.05	7.8E-01
Kansas: Gastrointestinal	PB	1.09	5.9E-01	0.79	6.1E-02	1.03	8.6E-01
	Pesticides	0.84	2.2E-01	1.07	5.5E-01	0.83	2.0E-01
Kansas: Mood/Neurological/ Cognitive	PB	2.30	3.3E-02	0.89	5.6E-01	2.18	4.4E-02
	Pesticides	1.41	1.9E-01	0.86	4.3E-01	1.38	2.2E-01
Kansas: Fatigue	PB	1.14	5.7E-01	0.95	7.9E-01	1.13	5.8E-01
	Pesticides	0.99	9.6E-01	0.79	1.4E-01	1.03	8.8E-01
Kansas: Respiratory	PB	2.00	4.1E-05*	0.78	5.5E-02	1.86	2.0E-04
	Pesticides	1.28	1.0E-01	0.98	8.8E-01	1.27	1.1E-01
Kansas: Skin	PB	1.11	5.2E-01	0.92	4.8E-01	1.11	5.1E-01
	Pesticides	1.04	7.9E-01	1.09	4.9E-01	0.99	9.2E-01

The group of interactions that was significant in the Kansas Respiratory domain model remained significant according to both the basic and fully adjusted models, but only in the PB interaction model. Figure 15 shows the interaction p-values for the region plotted on the $-\log_{10}$ scale. The *PON1* candidate SNP rs662 is not in high LD ($r^2 < 0.4$) with rs3917545.

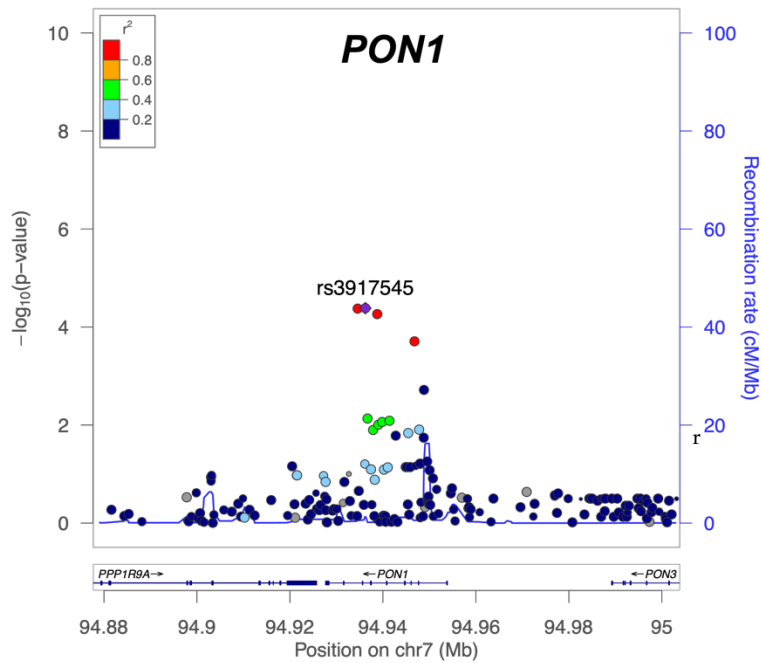


Figure 15. Kansas Respiratory Domain: *PON1* LocusZoom plot

Discussion

We identified a statistically significant association between CDC Severe GWI and the rs662/PB pill interaction, as seen in Figure 13. This figure depicts the primary result as a textbook image of a gene-environment interaction. While this was not the lowest p-value interaction in the study, it is the only signal that appears in our primary outcome out of the Tier 1 SNPs. Significantly higher rates of CDC Severe GWI with increasing exposure to PB pills were detected for GG and GA genotypes. This increase in risk does not appear in the AA genotype, indicating that exposure alone is not responsible for this effect and separating the risk genotypes from the protective genotype. A significant trend among those who reported over 31 days exposure to PB pills suggests an additive risk according to the count of G alleles. This would imply, as previously reported in the

literature, that the R version of the paraoxonase enzyme is less effective in detoxifying most substrates, and therefore results in higher risk with more exposure^{157,159}. Where prior studies have measured catalytic activity, we measured genotype on a chip; this is much higher throughput than protein activity measurements and allows for larger sample sizes. Follow-up analysis to assess enzymatic activity could fully replicate this finding.

A potential biological pathway is highlighted for CDC Severe GWI with the rs662/PB exposure interaction. A functional variant, the rs662 G alternative allele causes a change from a glutamine to an arginine at amino acid 192. As early as 1999, Haley et al. demonstrated a genotype-specific difference in paraoxonase (arylesterase) activity which was strongly associated with GWI⁵⁴. The Q version of the protein was previously found to have higher catalytic activity across most substrates, including sarin, while the R version of the enzyme was found to more efficiently detoxify paraoxone. In individuals who are unable to quickly break down these toxins, less damage can be prevented than in individuals who were able to rapidly metabolize more of these organophosphates. Rs662 has the strongest biological explanation and very clear, statistically significant trends that show a classic gene-by-environment interaction. Interestingly, this variant has also been implicated in several other diseases, including cardiovascular disease, which is part of the Kansas Exclusionary Criterion, and as a modifier of sporadic ALS, which affects Gulf War Veterans at higher rates than even

other comparable military groups¹⁴⁸. Since this variant has also been previously associated with sporadic ALS⁵⁸, this could also explain some of the increased risk of ALS among Gulf War Veterans.

Within the *PON1* candidate gene, a signal was identified in association with the Kansas Respiratory symptom domain outcome. This signal was not identified in any other outcomes and was statistically significant before multiplicity adjustment for the multiple symptom domains tested. We did not further correct for multiple comparisons due to secondary analyses of the multiple phenotypes. These results would no longer be statistically significant with this very conservative adjustment.

The SNPs identified are all in linkage disequilibrium with one another, and thus are likely a single signal. These SNPs are all non-coding variants that have not been identified by other GWI studies, although one of the SNPs, rs3917545 has been identified in the GWAS catalog as associated with blood protein levels of *PON1*¹⁶⁰. The other three candidate genes *BCHE*, *ACHE*, and *SOD1* did not yield any significant interactions for this candidate gene-environment interaction study.

There are some limitations to our study. While our sample size is large relative to other studies of GWI with genetic information, we had limited power to detect rare variants. We were able to prioritize candidate SNPs into two tiers, which allowed us to identify several significant signals, one of which has biological meaning. Future work could include data from the MVP study, which consists of 40,000 Gulf War Era Veterans,

which will allow for greater statistical power. The GWECB survey had a very low return rate and is therefore not representative of the general Gulf War Veteran population. An important assumption in our model is the independence of genotype and exposure. While we did not observe an association between genotype and exposures, we are unable to fully rule out a possible dependency. Genotype proportions by exposure time remained consistent across all exposure categories, indicating that there is not an association between genotype and exposure length, visualized in Appendix Figure 5. One benefit of GWECB is that both women and racial minorities were oversampled, allowing some examination of racial differences despite our modest sample size. The rs662 risk genotype is more common in those who identify as Hispanic or Black, non-Hispanic than in those who identify as White, non-Hispanic. Among White Veterans, risk was higher with increased exposure in the GG and AG groups, but not in the AA group (Appendix Figure 6). The same is true among the Black and Hispanic Veterans in the dataset (Appendix Figure 7), despite the smaller sample size. Finally, we were unable to replicate a statistically significant interaction involving pesticides in our analysis, although prior work indicates that pesticide exposure should also confer signal^{27,39,41,127,161,162} and qualitatively the pesticide results show similar relationships with the SNPs in *PON1*. This is potentially attributable to the heterogeneity in the composition of pesticides experienced. Different pesticides affect

different biological pathways, and we could be seeing this biologic heterogeneity reducing power for the pesticide analysis.

In summary, two significant signals were identified, associating Gulf War illness with the interaction between *PON1* and PB pill exposure. The primary result of the association between CDC Severe GWI and the rs662/PB pill exposure interaction corresponds with prior findings, as the rs662 locus is a functional variant that has been previously identified in smaller studies^{54,58,94,154}. The secondary result of the association between respiratory symptoms and the interaction of PB with a trio of SNPs in high linkage disequilibrium (LD) with one another is a new finding and requires replication in additional, larger studies. The SNPs in these two signals do not appear to be in LD with one another, although both are in *PON1* and both interactions are with PB pill exposure. We did not observe statistically significant interactions or marginal effects for any SNPs in *ACHE*, *BCHE* or *SOD1*.

Conclusions

Aligning with prior work in the area, this study identified an interaction between PB pill exposure and the G allele of rs662. Because this G allele codes for a functional variant, this interaction points to a biological pathway for GWI, identifying a potential biomarker for GWI risk.

5. Conclusions

Overview

Gulf War illness is an incredibly complex phenotype. Self-reported symptoms and conditions accumulate into a research case definition, which is separate from a clinical case definition. In a practical sense, Gulf War veterans looking for treatment require both a diagnosis and a service connection. Specific War-Related Illness and Injuries Study Centers (WRIISCs) have been focusing on both understanding GWI and streamlining clinical diagnosis. The service connection depends on the associations found by researchers between deployment, exposures due to deployment, and disease. From the research side, this work is focused on understanding how GWI might relate to exposures and genetic risk to better understand any biologic underpinnings of GWI. This includes who might be at greatest risk for developing multisymptomatic illness like GWI and how GWI might have developed due to exposures in the Gulf War. Future work could identify not only how GWI develops and works as a disease, but also how it is connected to service in the Gulf War. The research presented here indicates association alone and does not yet indicate or prove causation; more work is necessary to draw those much larger and more impactful conclusions. For Gulf War veterans who have been suffering from GWI for 30 years, identifying these causal pathways could be lifechanging.

Phenotyping efforts

In my work on Gulf War illness, I developed computer code to systematically identify those with GWI among veteran research cohorts⁸⁰. I then published this code, allowing other researchers to utilize not only the method, but the code for the GWI case definitions. As researchers coalesce on a research case definition, GWI research will become more consistent and more replicable. This is a valuable contribution to the field as it is not possible to do genetic work without confidence in the phenotype. My work with the phenotyping of Gulf War illness, specifically in the long discussions with experts in the field to agree on a single algorithm to define GWI in large cohort studies and in the writing of the algorithm in SAS, has already impacted both CSP585 and CSP2006. The algorithm has been applied to the MVP dataset and there are several manuscripts in progress looking at exposure, genetic, and general demographic associations that use my case definition. The process generated ownership across the entire group of the case definition and engendered confidence in the case definition code, making it more likely to be used both within the group and in other groups in the future. The case definition indicators were saved into the cohort datasets and the code has been published on GitHub, which will allow other researchers to easily access and utilize the indicators I organized, wrote, and implemented to push the field forward with a consistent, replicable phenotype.

Applications of the case definition to CSP585 and to CSP2006 have spurred a great deal of in-progress research. Work on associations between GWI and Gulf War exposures is necessary, as is subgrouping efforts related to both phenotype and exposure groups. With such a heterogeneous group, it is likely that there are several different ways to develop GWI; if correlations and associations between specific exposures and specific symptoms can be drawn, we will be closer to understanding a biological basis of GWI. Machine learning methods to cluster the symptoms could help identify subgroups among symptom clusters, where there may be several subtypes of GWI. GWI subtyping has been previously attempted through factor analysis; with the sample size of the MVP dataset and the computational power from a budding collaboration with the Department of Energy at Argonne National Laboratory, machine learning methods such as neural networks are within reach. Even simple machine learning methods such as hierarchical clustering could assist with this effort and could be completed using current VA resources. Further understanding of the phenotype and its potential subtypes could help untangle the heterogeneous nature of GWI.

Genomic analysis

Prior to my work, there were no genome-wide association studies of GWI. The GWAS in the GWECB dataset was an important first step in genomic analysis of GWI, as models should be examined from least to most complex. The genetic model tested in a GWAS assumes that the effects of exposure are negligible compared to the genomic

effects. In this GWAS for GWI in the GWECB dataset, I found no genome-wide significant associations with any of the four tested GWI case definitions in the full cohort, the deployed subgroup, or the white, non-Hispanic subgroup. We acknowledge that the GWECB dataset is underpowered for GWAS relative to the corrections for multiple testing required in a genome-wide setting. However, this lack of genome-wide significance may also suggest that the genetic model is incorrect; prior work suggests that the environmental component may be necessary. I identified candidate pathways that merit further study, perhaps in the MVP cohort. I identified CDC Severe GWI as the most heritable case definition in GWECB and found that the top ten ranked gene sets in the gene set association study supported the top theories of the etiology of GWI in the field. This heritability was calculated among all individuals in the cohort, meaning that deployed and non-deployed were analyzed together and the cohort was not stratified by ancestry group. Future work with a larger dataset could stratify by deployment. We expect the non-deployed veterans to experience a different type of disease than the deployed veterans, under an exposure-dependent model, as they were not exposed to the Gulf War exposures. Therefore, a more specific genetic model would be in the deployed veterans only, removing the presumably entirely unexposed non-deployed veterans. Additional efforts using polygenic risk scores (PRS) for IBS, CFS, or fibromyalgia to isolate the Gulf War-specific disease burden may be appropriate. Further work here could include replication in the MVP. MVP is large enough that it

may be possible to stratify the cohort by genetic ancestry, allowing for more specific analysis to be done in each ancestry group. The heritability analysis could be redone within each ancestry group among deployed veterans. The non-deployed veterans with GWI may provide insight into the genetic risk factors that are related more to civilian multisymptomatic illnesses (such as fibromyalgia or chronic fatigue syndrome) than to GWI or could identify risk factors among military training or preparation.

Support for prior proposed etiologies of GWI was found in the genome-wide analysis, particularly in the gene set analysis. Gene set association analysis, along with a literature review and some biological interpretation, suggests a neuroimmune or genetic instability root to GWI symptoms. A neuroimmune basis for GWI would explain the connection with neurodegenerative disease, as the immune dysregulation, and the heterogenous expression of symptoms. Other areas of interest around this are the links with dementia, ALS, Alzheimer's, and other neurodegenerative diseases. Several of the top gene set results are involved with clearing accumulated cellular waste. With associations between GWI and ALS, in addition to connections between GWI and other neurodegenerative diseases caused by accumulations of plaques, the gene set results are very intriguing. Future work would expand on these topics; for instance, perhaps medication for Alzheimer's could help veterans with GWI. Or perhaps a better understanding of how experiences, exposures, and genetics combine to cause the development of neurological symptoms in GWI could lead to creative ideas to treat

some of these currently incurable neurodegenerative diseases. In either direction, some suggestion towards the biological cause of the symptoms allows researchers to work towards understanding where the pathway can be affected by treatment. This basic science understanding is vital to beginning to work toward real treatments for GWI, rather than temporarily treating symptoms.

My GWAS pathway analysis and my candidate analysis point towards environmental toxicants. The GWAS pathway analysis returned top results related to response to environmental toxicants, response to cellular stress (such as accumulation of waste or misfolded proteins), and resiliency to environmental stress. With these results coming out of a genetics-only model, it is clear that the environmental component must be added to the model. With the genetics-only model, the idea is that the genetic components are strong enough to be detected in a logistic regression model even though the model is thought to be incomplete. In this case, even the genetic results point to environmental components by identifying variation in response to these exposures as the pathways with the highest ranked association with GWI. Preliminary results out of the MVP dataset indicate that even with more individuals, heritability is low in the genetics-only model and there are no genome-wide significant associations¹⁶³. The null results in a dataset that should be large enough to see results if they exist indicates that the model may be incorrect. The addition of exposures that are associated with GWI

and that are unique to service in the Gulf War to the genetic model may increase the power of these studies.

Interaction models

With this need for an environmental component, a gene-environment interaction model was the clear next step. In the GWECB targeted candidate gene study, I found a significantly associated gene-environment interaction with CDC Severe GWI that replicated prior published work. This is the first gene-environment interaction done on a large scale for GWI. Earlier studies that linked exposures and genetic variants were focused on functional protein variants, measured by protein activity. This locus does code for a functional variant but what was measured was the genetic variant as a SNP, not the protein activity. The significant association between GWI and the interaction between the rs662 locus and PB pills is an impactful finding as it is a statistically significant finding that could help explain some of the lack of replication in studies that do not include the interaction term. If it is the interaction that is important, the main effect might not show significant association and could cloud results. In this case, the main effect of the rs662 allele was not significant. The rs662 locus is additionally interesting because it is known to have a variable minor allele frequency across ancestry groups, with the minor allele flipping in European ancestry compared to African ancestry. The diversity of the GWECB dataset has contributed to the possibility of this finding; other studies were unable to gather enough homozygous G genotypes due to

the allele frequency in European populations. While we did test for association between genotype and exposure and found no association, replication in MVP could be stratified by ancestry group, allowing better study of the ancestry-specific nature of this interaction. Isolating specific pesticide chemicals could also expand this work, as we combined many types of pesticides together in the analysis, potentially combining opposite effects. As this interaction was first reported in 1999 through a functional study, it does have significant prior evidence that points to a biological pathway. With further work in this area, especially with parallel phenotyping efforts, a specific subgroup might become evident. Further understanding of this pathway and etiology could identify pathways to target for medication and therapy.

Expanding the genetic model

Diversity in genetic ancestry among the GWECB dataset has allowed for statistical analysis that may not have been possible without representation from individuals traditionally underrepresented in genetic research. In particular, Black and Hispanic individuals have historical reason to be skeptical of scientific research studies and genetic work, as genetics have been long misinterpreted into eugenics. Our results show that the outreach effort to include more individuals with admixed and African ancestry can lead to impactful results that would not be possible with a homogeneous European ancestry group alone. The GWECB dataset had only 9 individuals with Asian ancestry, which is too small for a separate analysis group. Further work could include

more outreach to veterans with Asian, Native American, Pacific Islander, and Middle Eastern ancestry. Continuing efforts to include Black and Hispanic veterans are also vital to this type of research. Both the MVP and GWECB have much more representation of Black and Hispanic veterans than other GWI datasets. However, since genetic ancestry and race are both intertwined and fundamentally different, working towards a cohort that most accurately represents the diverse and genetically admixed population of the United States is incredibly important. True genetic diversity is necessary to identify genetic risk factors that affect the entire population of veterans.

The genetic model could also be expanded in terms of differential exposure models. The locus we identified in the gene-environment interaction codes for two functional variants. Both variants of the Pon1 enzyme have been thoroughly studied by other researchers, with the conclusion that neither is uniformly more catalytically efficient than the other. One variant (our risk variant) has higher catalytic efficiency in breaking down paraoxon while the other variant (our protective variant) has higher catalytic efficiency in breaking down sarin, soman, and diazoxon. It is easy to imagine a similar situation where the protective variant of an enzyme depends on which chemical in particular the veteran was exposed to in the Gulf. To identify these situations, a better understanding (and measurement) of the exposures in the Gulf is necessary. Much like how not all genetic models are simply dominant or recessive, more complex exposure interaction models may be necessary.

Final thoughts

GWV is a complex and heterogeneous illness that could have several driving factors. I have contributed to the field in several ways: contributing a deterministic GWV case definition algorithm, publishing the first GWAS of GWV, and identifying a significant gene-environment interaction associated with GWV. Further, GWV indicators and variables generated through my data cleaning process have been included in the GWECB and MVP datasets, allowing other researchers to use these variables. The GWAS summary statistics will also be made available. Future research building off my work could help identify the underlying biological pathways and causes, allowing better treatment of the underlying disease for hundreds of thousands of Gulf War Veterans. Further research in this area could also lead to a better understanding of biological aging, neurodegenerative disease, and more common multisymptomatic illnesses, such as chronic fatigue syndrome, fibromyalgia, and irritable bowel syndrome.

Appendix A: Phenotype Identification

Mood/Cognition	Fatigue
Difficulty concentrating	Fatigue
Difficulty remembering recent information	Musculoskeletal
Feeling down or depressed	Pain in muscles
Feeling anxious	Pain in joints
Trouble finding words when speaking	Stiffness in joints
Problems getting to or staying asleep	
Feeling moody	

Appendix Figure 1. CDC and CDC Severe GWI definition

<p>Neurological/Mood/Cognition</p> <ul style="list-style-type: none"> Feeling down or depressed Low tolerance for heat or cold Trouble finding words when speaking Feeling dizzy, lightheaded, or faint Physical or mental symptoms in response to smells or chemicals Feeling irritable or having angry outbursts Eyes very sensitive to light Difficulty concentrating Headaches Blurred or double vision Tremors or shaking Night Sweats Numbness or tingling in extremities Difficulty remembering recent information <p>Skin</p> <ul style="list-style-type: none"> Rashes Other Skin 	<p>Fatigue</p> <ul style="list-style-type: none"> Not feeling rested after sleep Fatigue Problems getting to or staying asleep Feeling unwell after exercise <p>Pain</p> <ul style="list-style-type: none"> Pain in muscles Body pain, hurt all over Pain in joints <p>Respiratory</p> <ul style="list-style-type: none"> Wheezing in chest Frequent coughing without a cold Difficulty breathing or shortness of breath <p>Gastrointestinal</p> <ul style="list-style-type: none"> Nausea or upset stomach Diarrhea Abdominal pain or cramping 	<p>Cancer</p> <ul style="list-style-type: none"> Brain Cancer Colon Cancer Breast Cancer Lung Cancer Prostate Cancer Other Cancer <p>Heart Disease</p> <ul style="list-style-type: none"> Heart Attack Coronary Artery Disease Congestive Heart Failure <p>Infectious Disease</p> <ul style="list-style-type: none"> HIV Tuberculosis Hepatitis C <p>Mental Health</p> <ul style="list-style-type: none"> Schizophrenia Bipolar Disorder <p>Neurological</p> <ul style="list-style-type: none"> Traumatic Brain Injury Multiple Sclerosis <p>Stroke</p> <ul style="list-style-type: none"> Transient Ischemic Attack Stroke <p>Diabetes</p> <ul style="list-style-type: none"> Diabetes <p>Liver</p> <ul style="list-style-type: none"> Liver Disease <p>Lupus</p> <ul style="list-style-type: none"> Lupus
--	--	---

Appendix Figure 2. GWI Kansas Case Definition. Left/top: Symptoms in the Symptom Criteria. Right/bottom: Conditions in the Exclusionary Criterion

Appendix Table 1. Kansas GWI case definition symptoms by Gulf War deployment status

Domain	Symptom	Did Deploy (n = 849)	Did not deploy (n = 267)	aOR	95% CI
		(%)	(%)		
Fatigue/sleep problems	Fatigue	69.0	55.1	1.95	(1.45, 2.61)
	Feeling unwell after exercise	43.6	29.6	2.01	(1.48, 2.73)
	Difficulty getting to or staying asleep	75.1	62.2	1.91	(1.41, 2.58)
	Not feeling rested after sleep	76.7	64.8	1.83	(1.34, 2.50)
Pain	Pain in joints	82.4	76.8	1.48	(1.05, 2.08)
	Pain in muscles	64.7	56.2	1.49	(1.12, 1.99)

	Body pain where you hurt all over	43.6	37.1	1.37	(1.02, 1.84)
Neurologic/ cognitive/ mood	Difficulty remembering recent information	58.0	44.6	1.76	(1.33, 2.34)
	Feeling irritable or having angry outbursts	57.0	43.8	1.74	(1.30, 2.33)
	Numbness or tingling in extremities	63.7	50.9	1.74	(1.31, 2.31)
	Headaches	54.2	44.6	1.51	(1.13, 2.03)
	Eyes very sensitive to light	45.5	35.6	1.57	(1.17, 2.11)
	Trouble finding words when speaking	51.8	44.6	1.37	(1.03, 1.82)
	Feeling down or depressed	53.9	39.3	2.00	(1.49, 2.70)
	Difficulty concentrating	57.4	40.4	2.08	(1.56, 2.78)
	Night sweats	43.2	35.2	1.44	(1.07, 1.93)

	Feeling dizzy, lightheaded, or faint	45.2	34.1	1.62	(1.21, 2.17)
	Low tolerance for heat or cold	39.2	32.6	1.46	(1.08, 1.98)
	Symptoms in response to smells or chemicals	27.6	16.5	2.02	(1.41, 2.91)
	Blurred or double vision	38.5	28.8	1.60	(1.18, 2.18)
	Tremors or shaking	26.1	17.6	1.70	(1.19, 2.43)
Gastro-intestinal	Diarrhea	33.7	27.3	1.33	(0.98, 1.82)
	Nausea or upset stomach	36.9	29.2	1.44	(1.05, 1.96)
	Abdominal pain or cramping	34.4	27.0	1.47	(1.07, 2.01)
Respiratory	Difficulty breathing	38.2	25.5	1.90	(1.39, 2.62)
	Frequent coughing without a cold	33.9	25.5	1.59	(1.16, 2.18)

	Wheezing in chest	25.1	19.9	1.41	(1.00, 1.99)
Skin	Skin rash	37.5	23.2	2.05	(1.49, 2.82)
	Other skin problems	35.5	21.7	2.11	(1.52, 2.93)
<p>aOR = adjusted Odds Ratio for Gulf War deployed vs. nondeployed era veterans; CI = confidence interval.</p> <p>Note: OR adjusted for adjusted for sex, education, income, and age.</p>					

Appendix Table 2. CDC GWI case definition symptoms by Gulf War deployment status

Symptom domain	Symptom	Did deploy	Did not deploy	aOR	95% CI
		(n = 849)	(n = 267)		
		(%)	(%)		
Fatigue	Fatigue	69.0	55.1	1.95	(1.45, 2.61)
Musculo-skeletal	Pain in joints	82.4	76.8	1.48	(1.05, 2.08)
	Pain in muscles	64.7	56.2	1.49	(1.12, 1.99)
	Stiffness in joints	78.6	75.3	1.22	(0.87, 1.69)

Mood– cognition	Difficulty remembering recent information or difficulty concentrating	66.7	53.9	1.77	(1.32, 2.36)
	Difficulty remembering recent information	58.0	44.6	1.76	(1.33, 2.34)
	Difficulty concentrating	57.4	40.4	2.08	(1.56, 2.78)
	Trouble finding words when speaking	51.8	44.6	1.37	(1.03, 1.82)
	Feeling moody	58.2	46.4	1.65	(1.24, 2.21)
	Feeling down or depressed	53.9	39.3	2.00	(1.49, 2.70)
	Difficulty getting to or staying asleep ^a	75.1	62.2	1.91	(1.41, 2.58)
	Feeling anxious	52.3	37.8	1.90	(1.41, 2.55)

aOR = adjusted Odds Ratio for Gulf War deployed vs. non-deployed era veterans;

CI = confidence interval.

Note: OR adjusted for adjusted for sex, education, income, and age.

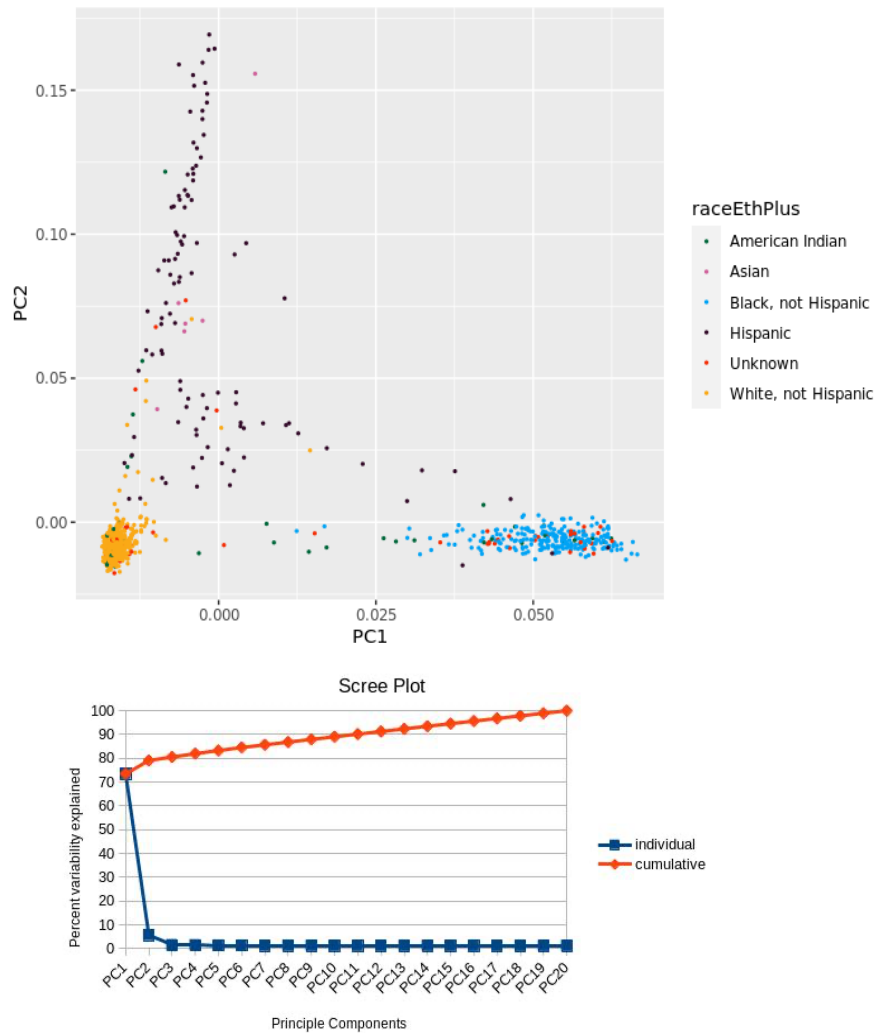
^a This symptom is used in both the Kansas and CDC GWI case criteria; However, it is included in the fatigue domain in the Kansas definition and in mood–cognition domain in the CDC definition.

Appendix Table 3. Frequency of veteran-reported exclusionary conditions by 1990–1991 Gulf War deployment status.

	All (1116)	Deployed (n = 849)	Did not deploy (n = 267)	aOR	95% CI
	(%)	(%)	(%)		
Any exclusionary condition	40.4	40.3	40.8	1.05	(0.78, 1.42)
Cancer	9.2	8.8	10.5	0.96	(0.60, 1.54)
Brain cancer	sup.	sup.	sup.	–	–
Breast cancer	1.0	0.7	sup.	0.41	(0.10, 1.67)
Colon cancer	sup.	sup.	sup.	–	–
Lung cancer	sup.	sup.	sup.	–	–
Prostate cancer	3.0	2.9	3.0	1.43	(0.59, 3.48)
Other cancer	5.4	5.1	6.4	0.84	(0.46, 1.52)
Diabetes	17.0	17.4	15.7	1.24	(0.83, 1.85)
Heart disease	8.7	7.9	11.2	0.81	(0.50, 1.33)
Heart attack	4.2	3.5	6.4	0.66	(0.34, 1.26)
Coronary artery disease	6.4	6.0	7.5	1.03	(0.57, 1.84)
Congestive heart failure	2.4	2.1	3.4	0.73	(0.31, 1.72)
Stroke	3.4	2.8	5.2	0.52	(0.26, 1.04)
Stroke	2.2	1.9	3.4	0.48	(0.20, 1.16)

Transient ischemic attack	1.7	1.3	3.0	0.47	(0.18, 1.21)
Infectious disease	4.7	4.5	5.2	0.85	(0.45, 1.63)
HIV	0.6	0.7	sup.	1.94	(0.22, 17.34)
Tuberculosis	2.3	2.5	sup.	1.35	(0.50, 3.69)
Hepatitis C	2.0	1.5	3.4	0.44	(0.18, 1.09)
Liver disease	2.2	2.0	2.6	0.77	(0.31, 1.93)
Lupus	1.1	1.3	sup.	3.05	(0.36, 25.67)
Mental health	3.6	3.7	3.4	1.17	(0.53, 2.58)
Schizophrenia	0.5	sup.	sup.	0.95	(0.09, 9.71)
Bipolar disorder	3.4	3.4	3.4	1.07	(0.48, 2.38)
Neurological	4.7	4.7	4.9	0.92	(0.47, 1.79)
Multiple sclerosis	0.7	sup.	sup.	0.39	(0.08, 1.81)
Traumatic brain injury	4.0	4.1	3.7	1.08	(0.52, 2.27)
<p>Note: OR adjusted for sex, education, income, service branch, unit component, and age.</p> <p>*All includes 75 individuals who were deployed in support of the Persian Gulf war but not to the Gulf War Theater or Operations.</p> <p>-- indicates OR undefined due to zero cell size.</p>					

Appendix B: Genome-wide investigation



Appendix Figure 3. PCA of genetic ancestry in GWECB

Appendix C: Biologically based candidate studies

Appendix Table 4. Demographics by Pesticide Exposure

		Overall	Pesticide Exposure				
			None	Missing	Short	Medium	Long
Count		810	346	106	78	124	156
Sex	Female	22%	26%	19%	19%	23%	17%
	Male	78%	74%	81%	81%	77%	83%
Age Group	40-49	40%	37%	30%	35%	46%	49%
	50-59	37%	36%	37%	39%	36%	38%
	60+	24%	27%	33%	27%	19%	14%
Race/ Ethnicity	White, non- Hispanic	65%	66%	51%	71%	65%	71%
	Black, non- Hispanic	17%	18%	27%	8%	13%	16%
	Hispanic	9%	11%	12%	9%	9%	5%
	Other	5%	2%	4%	9%	11%	6%
Income	Under \$30,000	10%	8%	16%	8%	10%	13%
	\$30,000-59,999	22%	24%	24%	17%	23%	19%
	\$60,000-99,999	30%	29%	26%	35%	27%	33%

	\$100,000+	30%	30%	26%	33%	34%	30%
Education	High school, GED, or less	9%	12%	9%	5%	4%	10%
	Some college, or associate's or bachelor's degree	67%	66%	70%	74%	73%	60%
	Advanced degree	21%	20%	16%	17%	20%	27%
Service Branch	Army only	45%	35%	50%	46%	52%	59%
	Navy only	17%	27%	11%	13%	10%	8%
	Air Force only	10%	12%	8%	9%	9%	8%
	Marine Corps only	13%	13%	12%	15%	16%	9%
	National Guard, any	10%	8%	14%	10%	10%	11%
Military Component	Active Duty	58%	53%	62%	62%	60%	64%
	Both Active Duty and Reserves	26%	28%	25%	21%	24%	24%

	Reserves Only	16%	18%	13%	18%	15%	12%
OEF/OIF deployment	No	76%	85%	70%	80%	80%	55%
	Yes	23%	15%	24%	21%	19%	45%

Appendix Table 5. Demographics by PB exposure.

		Overall	PB pill exposure				
			None	Missing	Short	Medium	Long
Count		810	311	138	98	114	149
Sex	Female	22%	28%	24%	18%	17%	16%
	Male	78%	72%	76%	82%	83%	84%
Age Group	40-49	40%	38%	37%	40%	43%	42%
	50-59	37%	34%	38%	30%	42%	40%
	60+	24%	27%	25%	31%	15%	18%
Race / Ethnicity	White, non- Hispanic	65%	70%	60%	68%	61%	61%
	Black, non- Hispanic	17%	16%	24%	10%	17%	17%
	Hispanic	9%	7%	10%	11%	11%	11%
	Other	5%	4%	2%	6%	8%	7%
Income	Under \$30,000	10%	11%	14%	6%	9%	10%

	\$30,000-59,999	22%	23%	21%	22%	22%	22%
	\$60,000-99,999	30%	27%	29%	37%	26%	32%
	\$100,000+	30%	30%	32%	30%	32%	30%
Education	High school, GED, or less	9%	11%	7%	8%	7%	11%
	Some college, or associate's or bachelor's degree	67%	65%	73%	68%	71%	62%
	advanced degree	21%	23%	17%	20%	18%	22%
Service Branch	Army only	45%	31%	43%	62%	59%	56%
	Navy only	17%	31%	18%	4%	4%	7%
	Air Force only	10%	15%	10%	5%	6%	5%
	Marine Corps only	13%	8%	15%	16%	15%	18%
	National Guard, any	10%	12%	10%	9%	8%	9%
	Active Duty	58%	54%	59%	48%	66%	66%

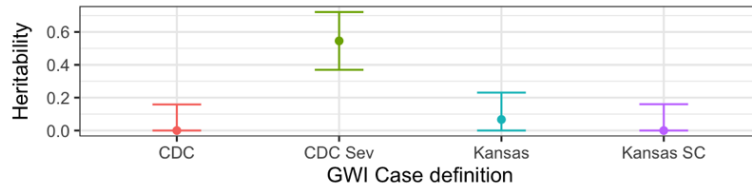
Military Component	Both Active Duty and Reserves	26%	28%	25%	27%	19%	26%
	Reserves Only	16%	17%	15%	26%	15%	7%
OEF/OIF deployment	No	76%	77%	79%	81%	72%	70%
	Yes	23%	23%	15%	19%	27%	30%

Appendix Table 6. Frequency of fulfilling symptom domains and case definitions by exposure to pesticides.

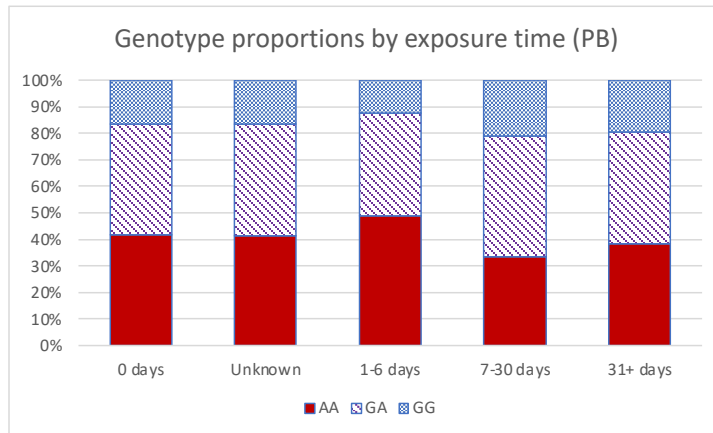
	Overall	Pesticide Exposure				
		None	Missing	Short	Medium	Long
Count	810	346	106	78	124	156
CDC Severe: Fatigue Domain	16%	12%	18%	19%	19%	21%
CDC Severe: Mood Domain	33%	25%	40%	32%	37%	45%
CDC Severe: Musculoskeletal Domain	33%	24%	46%	37%	36%	41%
CDC Severe GWI	26%	19%	35%	24%	30%	33%
Kansas: Fatigue Domain	80%	70%	86%	81%	86%	92%
Kansas: Respiratory Domain	36%	25%	43%	47%	40%	45%
Kansas: Gastrointestinal Domain	39%	30%	41%	47%	36%	55%
Kansas: Pain Domain	73%	62%	84%	80%	76%	83%
Kansas: Skin Domain	33%	22%	35%	42%	40%	42%
Kansas: Mood Domain	87%	80%	90%	87%	95%	93%
Kansas Symptom Criteria	72%	59%	81%	78%	77%	89%
Kansas: Exclusionary Criterion	40%	40%	44%	46%	39%	38%

Appendix Table 7. Frequency of fulfilling symptom domains and case definitions by exposure to PB pills

	Overall	PB Pill Exposure				
		None	Missing	Short	Medium	Long
Count	810	311	138	98	114	149
CDC Severe: Fatigue Domain	16%	9%	20%	22%	17%	24%
CDC Severe: Mood Domain	33%	24%	37%	34%	41%	44%
CDC Severe: Musculoskeletal Domain	33%	24%	36%	38%	34%	48%
CDC Severe GWI	26%	18%	30%	28%	31%	36%
Kansas: Fatigue Domain	80%	71%	80%	84%	89%	89%
Kansas: Respiratory Domain	36%	27%	40%	43%	40%	43%
Kansas: Gastrointestinal Domain	39%	28%	38%	40%	52%	51%
Kansas: Pain Domain	73%	61%	80%	77%	77%	87%
Kansas: Skin Domain	33%	24%	33%	40%	38%	41%
Kansas: Mood Domain	87%	79%	87%	90%	93%	94%
Kansas Symptom Criteria	72%	59%	75%	74%	83%	87%
Kansas: Exclusionary Criterion	40%	38%	46%	42%	39%	41%

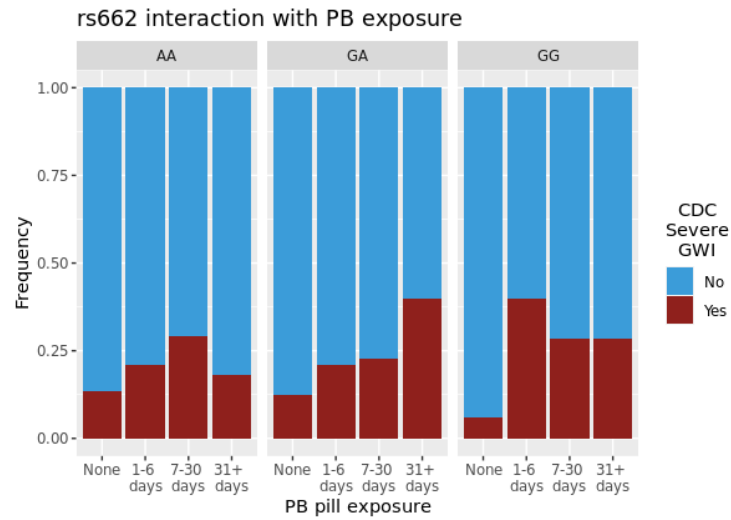


Appendix Figure 4. Estimated heritability of CDC Severe GWI

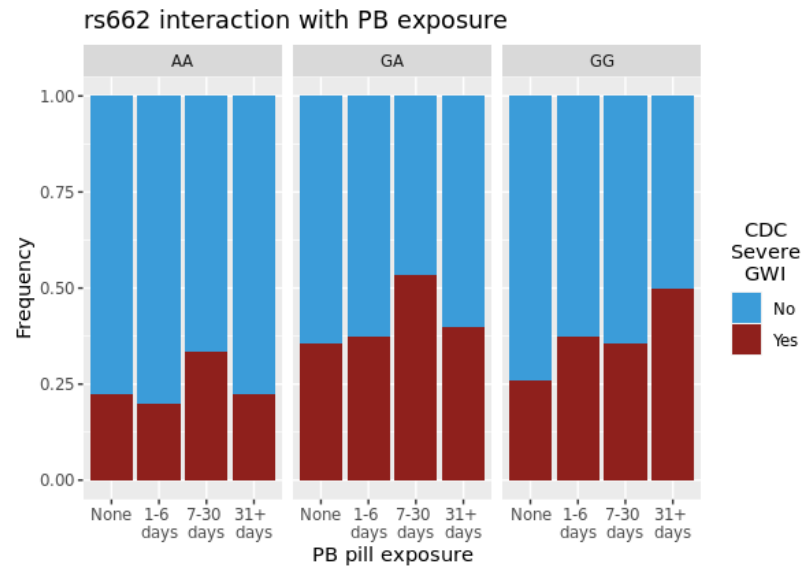


Appendix Figure 5. Genotype proportions by exposure time (PB)

Note: Individuals in Appendix Figures 6 and 7 are stratified by rs662 genotype and PB exposure time.

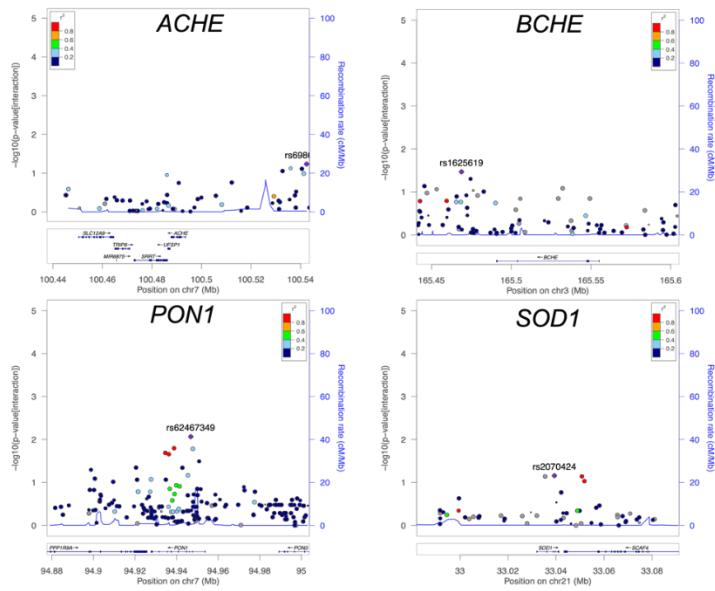


Appendix Figure 6. CDC Severe GWI frequency: white, non-Hispanic



Appendix Figure 7. CDC Severe GWI frequency: Hispanic or Black, non-Hispanic

CDC Severe: p-values for SNPxPesticide interaction



Appendix Figure 8. LocusZoom plots for Tier 2 SNPs in interaction with pesticides

11. While you were in the Gulf region, did you experience any of the following?	Not Sure	No	Yes →	IF YES: About how many days?		
				1-6 days	7-30 days	31 days or more
a. Entered Iraq	<input type="radio"/> Not Sure	<input type="radio"/> No	<input type="radio"/> Yes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b. Entered Kuwait	<input type="radio"/> Not Sure	<input type="radio"/> No	<input type="radio"/> Yes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c. Served on board a ship	<input type="radio"/> Not Sure	<input type="radio"/> No	<input type="radio"/> Yes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d. Close proximity to smoke from oil well fires	<input type="radio"/> Not Sure	<input type="radio"/> No	<input type="radio"/> Yes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e. Directly involved in ground combat	<input type="radio"/> Not Sure	<input type="radio"/> No	<input type="radio"/> Yes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f. Took pyridostigmine bromide (anti-nerve agent pills)	<input type="radio"/> Not Sure	<input type="radio"/> No	<input type="radio"/> Yes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
g. Exposed to chemical or biological warfare agents	<input type="radio"/> Not Sure	<input type="radio"/> No	<input type="radio"/> Yes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
h. Worked with prisoners of war	<input type="radio"/> Not Sure	<input type="radio"/> No	<input type="radio"/> Yes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
i. Used pesticide cream or liquid on your skin	<input type="radio"/> Not Sure	<input type="radio"/> No	<input type="radio"/> Yes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
j. Wore a uniform treated with pesticides	<input type="radio"/> Not Sure	<input type="radio"/> No	<input type="radio"/> Yes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
k. Used insect baits / no-pest strips in your living area	<input type="radio"/> Not Sure	<input type="radio"/> No	<input type="radio"/> Yes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Appendix Figure 9. Exposure questions from the CSP585 questionnaire

Appendix D: GWI indicator pseudocode

Notes and Citation

This code is published on GitHub: <https://github.com/VA-Phenomics-Library-CIPHER/Gulf-War-Illness>

Documentation was co-authored by Jackie Vahey and Beth Gifford.

Glossary

‘symptom’ refers generically to a specific symptom from the list of Kansas and CDC list. Throughout this document, ‘symptom’ implies that the variable name is the same as in the CSP585 dataset.

Rev short for revised

‘diagnosis’ refers generically to a specific diagnosis used in the Kansas definition exclusionary criteria.

Domain refers to a collection of symptoms as defined by the Kansas and CDC studies (see Table 1)

‘domain’ refers generically to a specific domain from the list of Kansas and CDC list. For Kansas, the official name of the domains (and abbreviation) are (1) Fatigue & sleep (fatigue), (2) Pain (pain), (3) Neurological/mood/cognition(mood), (4) Gastrointestinal (GI), (5) Respiratory (resp), and (6) Skin (skin). For CDC the official name of the domains (and abbreviation) are (1) Fatigue (Fatigue), (2) Musculoskeletal

(pain), (3) Mood-Cognition(mood). Part A. Preparing Variables Related to Symptoms and Severity of Symptoms for the Kansas and CDC Definition.

Part A. Prepare Symptoms for Analysis

Step A1. Select Symptoms for Kansas and CDC definitions

Table 1 lists the symptoms from CSP585 used to construct the Kansas and CDC definitions.

Appendix Table 8. Symptoms Used to Construct the Kansas and CDC Domains

Table 1. Symptoms Used to Construct the Kansas and CDC Domains		
Domains	KANSAS	CDC
Fatigue & sleep (KS=4 items CDC=2 items) Called Fatigue by CDC	37D Not feeling rested after you sleep	
	37aFatigue	37aFatigue
	37c Problems getting to sleep or staying asleep ¹	
	37b Feeling unwell after physical exercise or exertion	
Pain (3 items for KS & CDC) Called Musculoskeletal by CDC	37e pain in your joints	37e pain in your joints
		37f stiffness in your joints
	37g. Pain in your muscles	37g. Pain in your muscles
	37h. Body pain where you hurt all over	
Neurological/ mood /cognition (KS=14 items	37cc. Difficulty remembering recent information	37cc. Difficulty remembering recent information
	37ff. Feeling irritable or having angry outbursts	

CDC=7 items) Called Mood- Cognition by CDC	37m. numbness or tingling in your extremities	
	37i. headaches	
	37k. eyes very sensitive to light	
	37dd. Trouble finding words when speaking	37dd. Trouble finding words when speaking
	37ee. Feeling down or depressed	37ee. Feeling down or depressed
	37bb. Difficulty concentrating	37bb. Difficulty concentrating
	37p. night sweats	
	37j. Feeling dizzy, lightheaded or faint	
	37o. low tolerance for heat or cold	
	37q. having physical or mental symptoms in response to certain smells or chemicals	
	37l. blurred or double vision	
	37n. tremors or shaking	
		37c. Problems getting to sleep or staying asleep
		37gg. Feeling moody
	37hh. Feeling anxious	
GI (3 items-KS-only)	37t. Diarrhea	
	37u. Nausea or upset stomach	
	37v. Abdominal pain or cramping	

Respiratory (3 items-KS-only)	37w. Difficulty breathing or shortness of breath	
	37x. Frequent coughing when you don't have a cold	
	37y. wheezing in your chest	
Skin (2 items-KS-only)	37r. rashes	
	37s. other skin	
¹ Note that this symptom is used in the CDC definition but under mood domain rather than the fatigue domain		

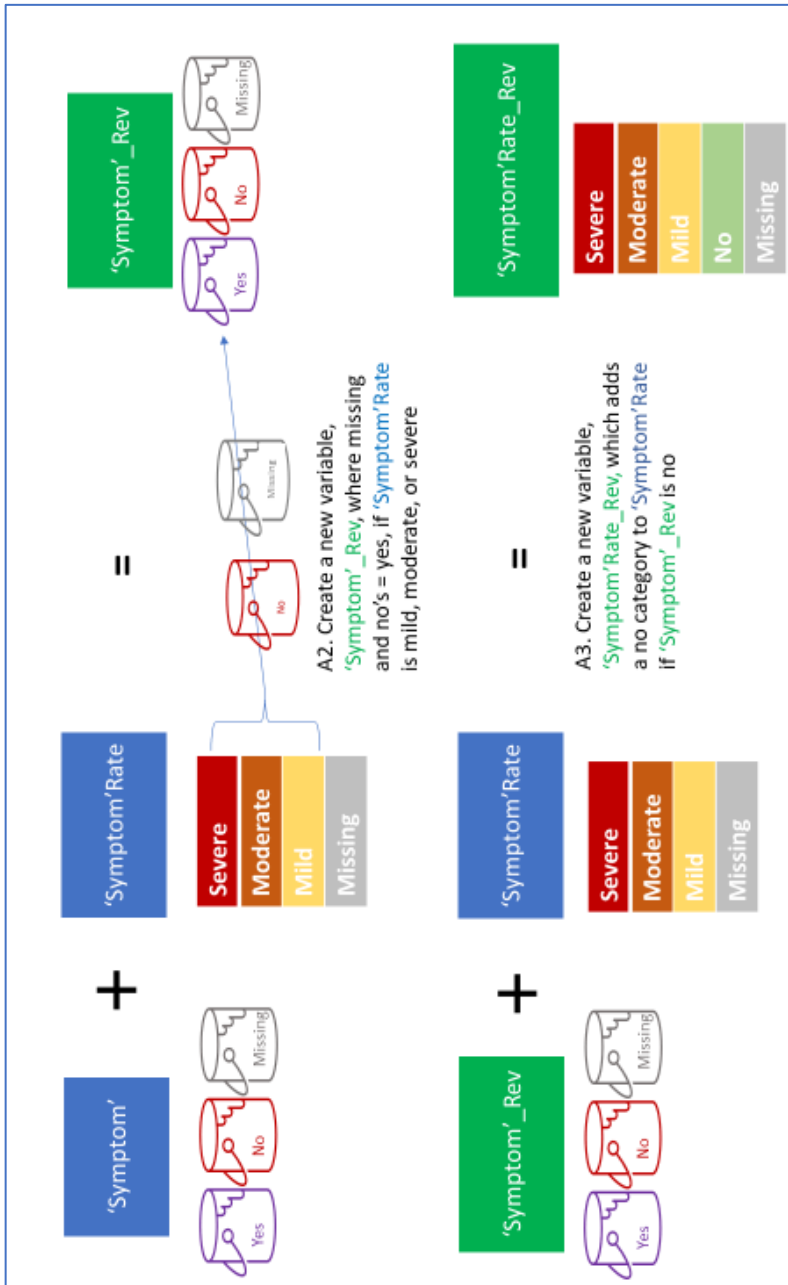
Step A2. Infer a symptom is present if responded rated the severity as mild, moderate or severe

For the symptoms used in (a) Kansas and (b) CDC definition (see Table A1), we create new variables [**'symptom'_Rev**] that revise **'symptom'** responses to account for survey missing response patterns and the MVP standard of updating less specific information when it is in conflict with a more specific answer. If the respondent rated the severity of the symptom as "mild", "moderate", or "severe" (e.g., **'symptom'Rate**) then "missing" & "no" responses to the **'symptom'** become "yes" to indicate that the person experienced the symptom.

These new **'symptom'_Rev** variables are used for all calculations in the Kansas and CDC definitions.

Step A3. Create new variables that revise the severity score from 0 “No” to 3
“Severe”

For the purposes of programming, we create new variables that revise the severity score. Specifically, if the respondent indicated “no” to having a symptom, then we treat the severity rating as “no” and assign the value 0. The new variables are named **‘symptom’Rate_Rev** (figure 1).



Appendix Figure 10. Construction of Revised Symptom and Symptom Rating Variables

Part B. Constructing the Kansas GWI definition

Step B1. Creating the exclusion criteria

The Kansas GWI definition excludes individuals from being eligible for having GWI if they have been previously diagnosed with certain conditions. Table 2 lists the exclusion criteria that were used for the CSP 585 sample.

Step B1a. Determine the categories for exclusion

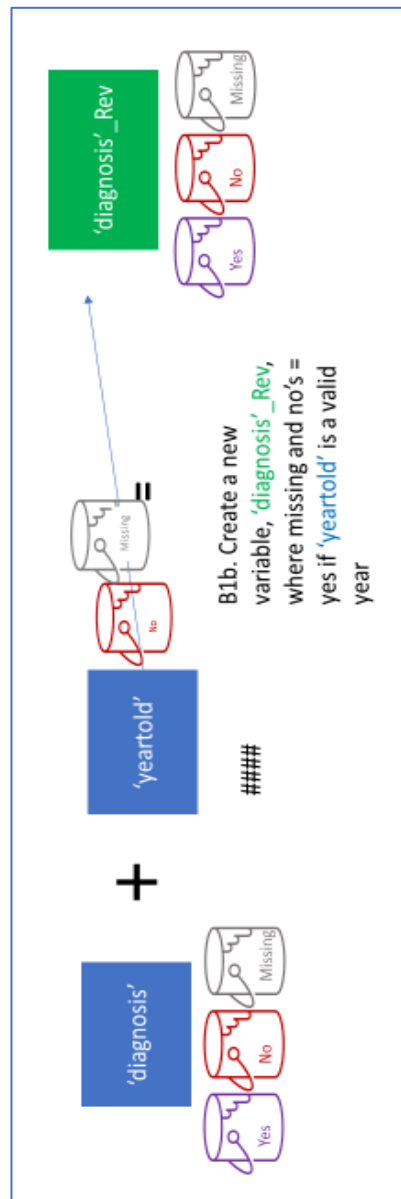
Appendix Table 9. List of Kansas Exclusionary Criteria

Table 2. List of Exclusionary Criteria Used for the Kansas Definition	
Exclusion Criteria	Question #/ Variable name
All Cancers	
Brain	CaBrain
Breast	CaBrst
Colon	CaColon
Lung	CaLung
Prostate	CaPros
Melanoma (where available)	Not available in CSP585.
Other	CaOth
Diabetes	DoDM
Heart Disease	
Heart Attack	CircHrtAtk
Coronary artery disease	CircCAD

Congestive heart failure	CircCHF
Stroke	CircStrk
Transient ischemic attack	CircTIA
Infectious Disease	
HIV	IDHIV
Tuberculosis	IDTB
Hepatitis C	IDHepC
Liver disease	DoLiver
Lupus	DoLupus
Schizophrenia	MHSCZ
Bipolar	MHBPD
Multiple Sclerosis	NSMS
TBI	NSTBI

Step B1b. Create revised diagnosis variables

Create a variable called **'diagnosis'_Rev** for each diagnosis that is 'yes' if the person reported having **'diagnosis'** (figure 2). In addition, if the respondent reported a **'yeartold'** but whose **'diagnosis'** is no or missing, then the **'diagnosis'_Rev** marked yes.



Appendix Figure 11. Creating Revised Diagnosis Variables

Step B1c. Create Kansas Definition Exclusion Indicator Variable

Create a variable called **KS_Excl** as follows:

Yes if the respondent answered “yes” to any of the revised exclusion criteria

‘diagnosis’_Rev (Table 2).

No if the respondent answered “no” to all of the revised exclusion criteria

‘diagnosis’_Rev (Table 2).

Missing If the respondent has not responded “yes” to any of the **‘diagnosis’_Rev** but has at least 1 missing item in the exclusionary criteria.

We also create a variable called **KS_Excl_NumMiss** which is a count of missing items among the revised exclusion criteria **‘diagnosis’_Rev** (table 2).

Step B2. For each domain, determine if the respondent did or did not meet multiple or moderate or severe criteria

This section describes how to construct the 6 variables which document whether or not the veteran met the moderate or multiple criteria of the Kansas Domains (figure 3). Figure 4 demonstrates how to construct the moderate-multiple criteria for the fatigue domain.

For each symptom domain:

Once a person meets the “yes” criteria then they are a yes and they remain in that category. The following criteria qualify a person for “yes”:

Multiple Symptoms: if the respondent answered “yes” to more than 1 **‘symptom’_Rev** in the domain **and/or**

Moderate-to-Severe: at least 1 **‘symptom’Rate_Rev** was coded as “moderate” or “severe”.

Note, in the CSP585 repository, if the veteran marked multiple severity scores (among mild, moderate, several) these were recorded in the data as “81-multiple responses”; In our sas dataset we have used .b as the extended SAS missing code for 81. These should be treated as moderate or severe for this calculation, as it is impossible for a multiple answer options to include neither moderate or severe.

To distinguish between “No” and “Missing”:

No: all three of the below bullets must be true:

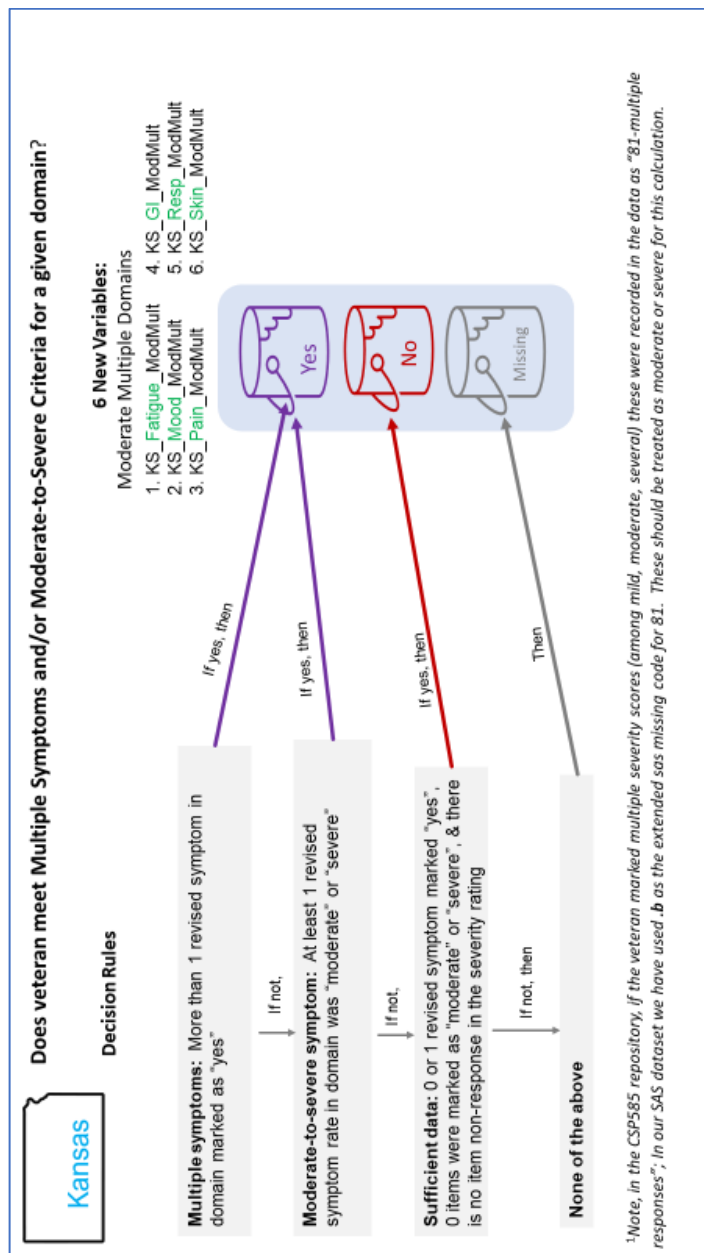
Zero or one ‘**symptom**’_Rev marked “yes”

Zero revised ‘**symptom**’Rate_Rev were marked as “moderate” or “severe”

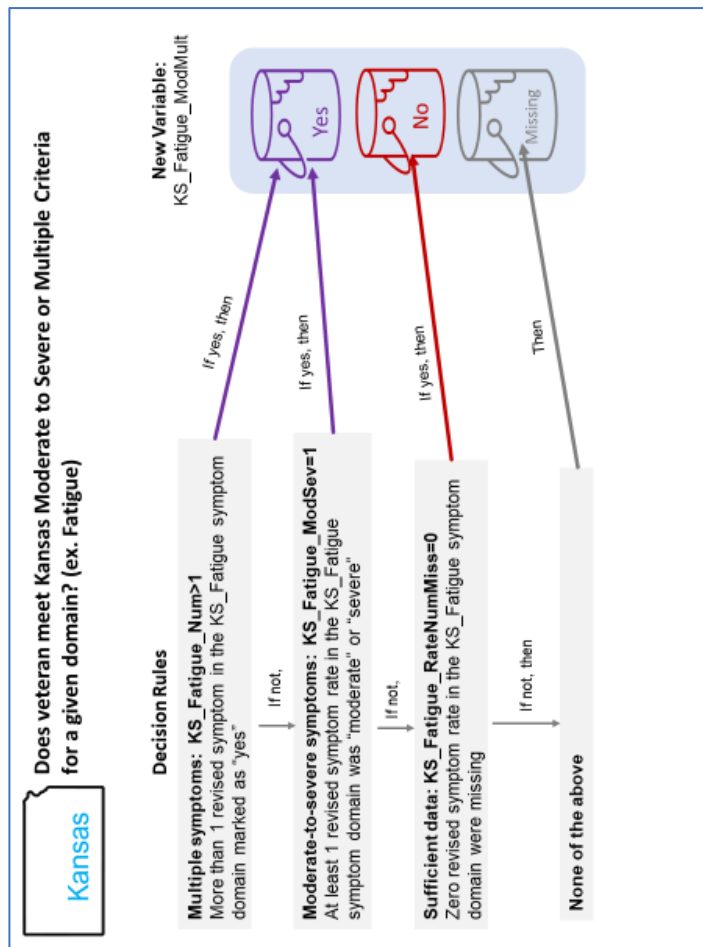
There were zero items missing in ‘**symptom**’Rate_Rev

Missing: 1 or more ‘**symptom**’Rate_Rev items had missing values. *This is because a single “yes, moderate” or “yes, severe” is enough to change a domain from “no” to “yes”.*

Appendix Figure 12. Kansas Symptom Domain Indicators



Appendix Figure 13. Kansas Symptom Criteria Indicator, Part 1



Appendix Figure 14. Kansas Symptom Criteria Indicator, example

Step B3. Determine if the respondent did or did not meet Kansas symptom criteria

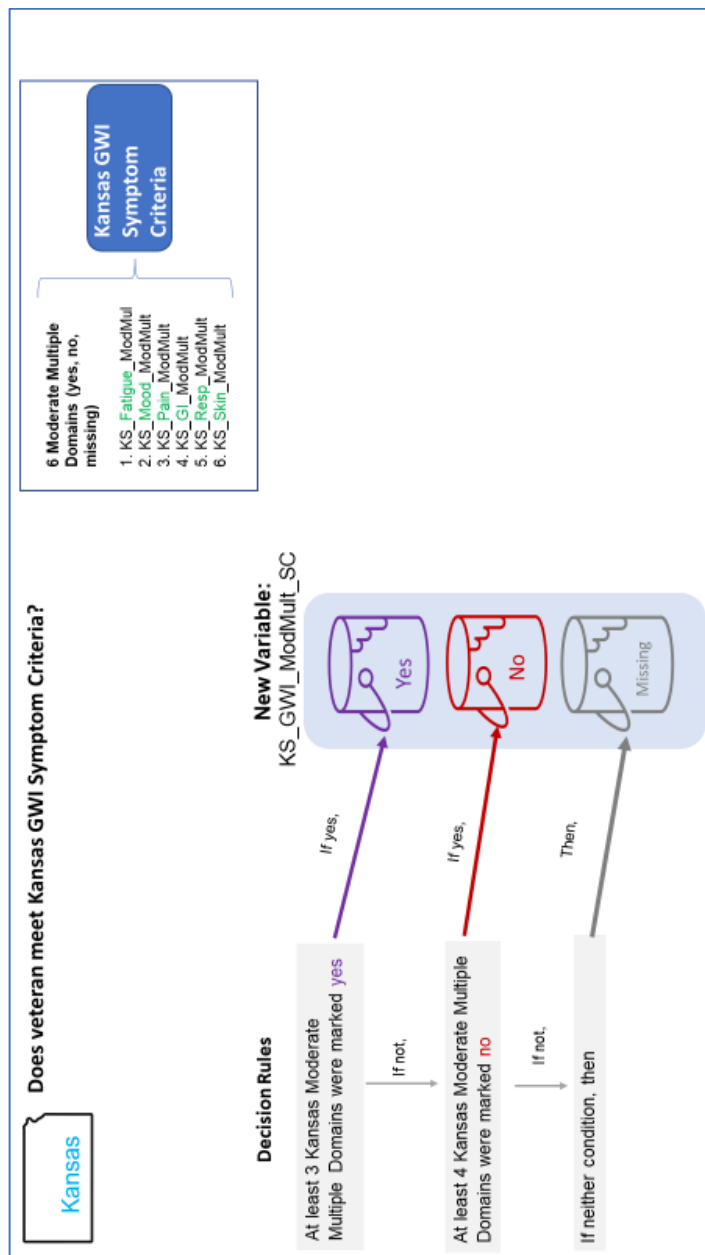
To denote whether or not the veteran met the *symptom criteria* for the Kansas GWI definition, we constructed a variable called $KS_GWI_ModMult_SC$ (see figure 5). Table 3 lists all of the possible combinations of the 6 Kansas domains to determine whether or not the veteran met the symptom criteria.

In summary, this variable is:

Yes: If 3 or more multiple/moderate-to-severe symptom domains are “yes”, then respondent met Kansas Definition Symptom Criteria

No: If 4 or more multiple/moderate-to-severe symptom domains are ”no”, the respondent does not meet Kansas Definition Symptom Criteria.

Missing: If neither 3 or more multiple/moderate-to-severe symptom domains are “yes” or 4 or more multiple/moderate-to-severe symptom domains are “no”, then the Kansas Definition Symptom Criteria is missing. “Missing” Kansas Definition Symptom Criteria indicates that the true (but unknown) value of the information that is missing from the survey questionnaire could change the value of the Kansas Definition Symptom Criteria indicator.



Appendix Figure 15. Kansas Symptom Criteria Indicator, Part 2

Appendix Table 10. Kansas GWI Symptom Criteria by Symptom Domain Indicators

Number of Kansas Domains Marked Yes	Number of Kansas Domains Marked No	Number of Kansas Domains Marked Missing	Does Veteran Meet Kansas Symptom Criteria?
6	0	0	Yes
5	0	1	Yes
5	1	0	Yes
4	0	2	Yes
4	1	1	Yes
4	2	0	Yes
3	0	3	Yes
3	1	2	Yes
3	2	1	Yes
3	3	0	Yes
2	0	4	Missing
2	1	3	Missing
2	2	2	Missing
2	3	1	Missing
2	4	0	No
1	0	5	Missing
1	1	4	Missing
1	2	3	Missing

1	3	2	Missing
1	4	1	No
1	5	0	No
0	0	6	Missing
0	1	5	Missing
0	2	4	Missing
0	3	3	Missing
0	4	2	No
0	5	1	No
0	6	0	No

Step B4. Determine if the respondent met the criteria for Kansas Gulf War Illness

This section combines information from the Kansas symptom criteria with the exclusionary criteria to determine whether or not the veteran met the Kansas GWI definition (see figure 6). Table 4 provides all of the possible combinations of Kansas GWI symptom criteria and the exclusion criteria and how these combinations are reflected in the Kansas GWI case status.

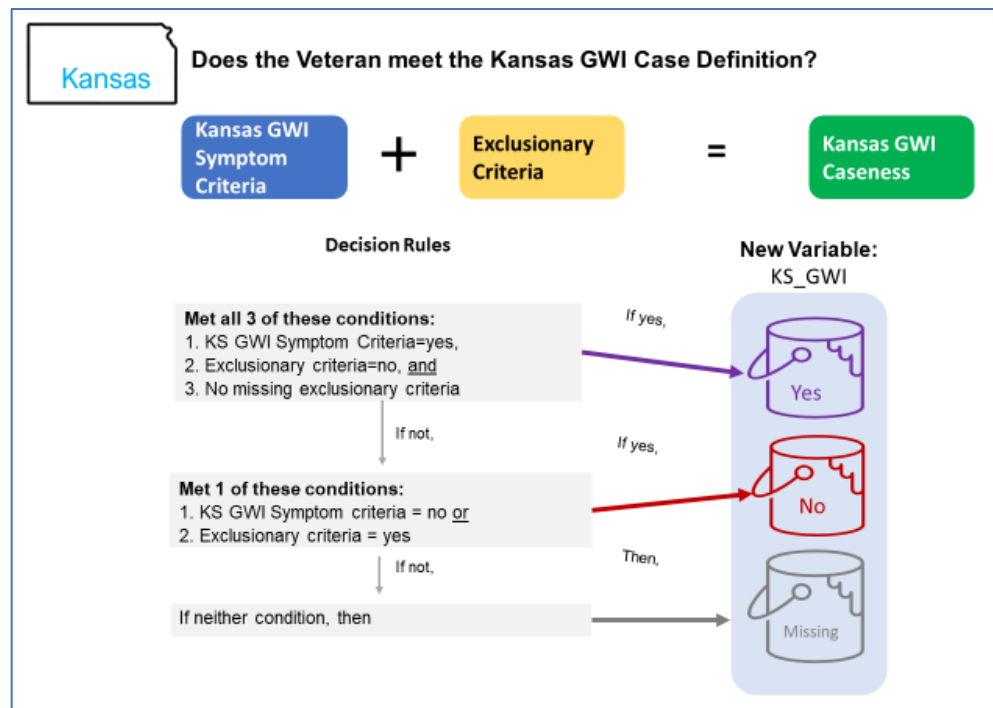
Yes: Respondents who met Kansas Definition Symptom Criteria **and** who did not meet the Exclusionary Criteria **and** had no missing Exclusionary Criteria.

No:

Respondents who did not meet Kansas Definition Symptom Criteria, **or**

Respondent who met exclusionary criteria

Missing: Respondents who were classified as missing Kansas Definition Symptom Criteria **or** who were missing exclusionary criteria.



Appendix Figure 16. Kansas GWI Case Definition

Veteran meets Kansas symptom criteria	Veteran endorsed yes on at least one exclusionary condition	Kansas Definition Case Status
Yes	No	Yes
Yes	Yes	No
No	Yes	No
No	No	No
Yes	Missing	Missing
No	Missing	Missing
Missing	Missing	Missing
Missing	Yes	Missing
Missing	No	Missing

Appendix Table 11. Does Veteran meet Kansas GWI case definition?

Part C. Constructing the CDC GWI definition

The CDC GWI definition is composed of 3 domains: fatigue; mood and cognition; and musculoskeletal (see table 1 for symptoms that compose each domain). Figure 7 depicts the logic for combining possible combinations of the symptoms to determine whether or not the veteran (a) meets the domain criteria and (b) meets the “severe” domain criteria.

Step C1. For each domain, determine if the respondent did or did not meet the domain criteria

Construct 3 variables 1 for each domain named **CDC_‘domain’_Any** which are scored as follows:

“Yes” if the respondent answered yes to any **‘symptom’_Rev** within the domain

“No” if respondent answered no to each **‘symptom’_Rev** within the domain

“Missing” if respondent did not answer yes to any **‘symptom’_Rev** within the domain and at least one **‘symptom’_Rev** is missing

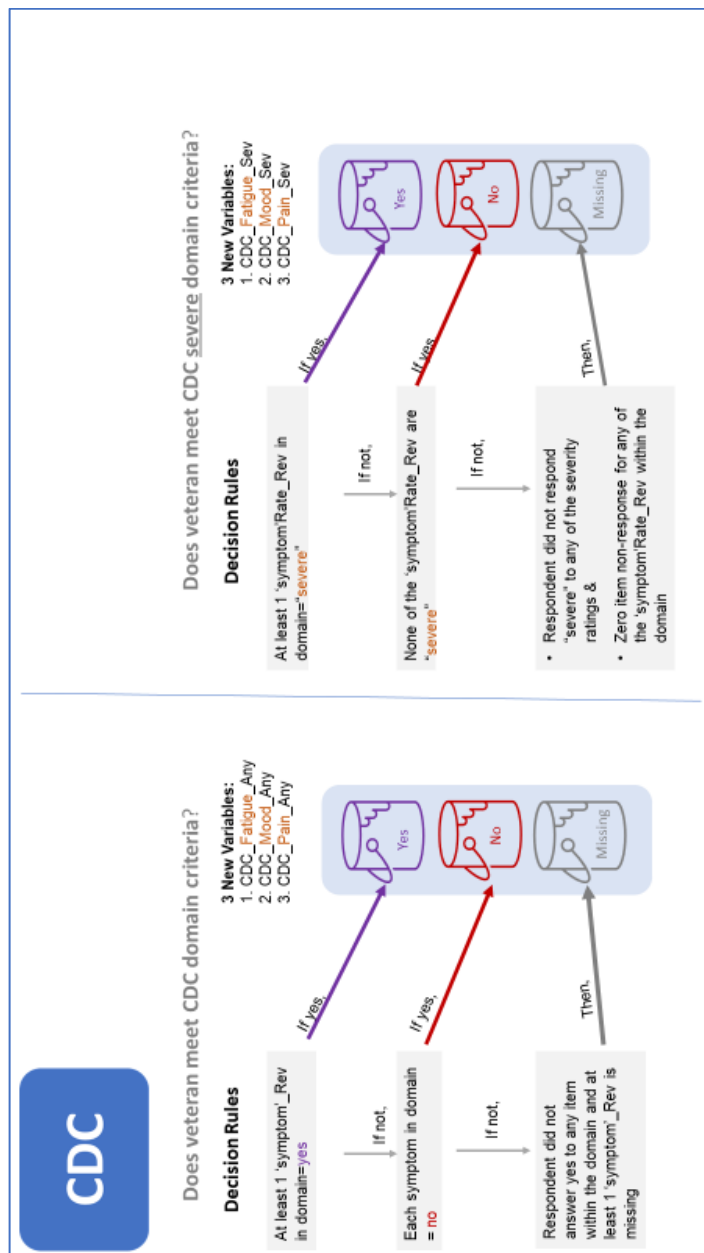
Step C2. For each domain, determine if the respondent did or did not meet the severe criteria

Construct three variables (one for each domain) named **CDC_‘domain’_sev** which are scored as follows:

“Yes” if the respondent answered “severe” to any of the **‘symptom’Rate_Rev** within the domain

“No” if none of the **‘symptom’Rate_Rev** within the domain are “severe” *and* none of the **‘symptom’Rate_Rev** within the domain are missing

“Missing” if none of the **‘symptom’Rate_Rev** are “severe” *and* none of the **‘symptom’Rate_Rev** within the domain are missing



Appendix Figure 17. CDC and CDC Severe Symptom Domain Criteria

Step C3. Determine if the respondent met the criteria for CDC Gulf War Illness

Construct two variables (1) CDC_GWI and (2) CDC_GWI_Sev to identify if respondent meet the CDC GWI case definitions (see table 5 and figure 8).

CDC Gulf War Illness Case Status

“Yes” if the respondent met two or three domain criteria

“No” if the respondent met zero or one domain criteria

“Missing” if the respondent did not meet the yes criteria and had fewer than two no’s

Appendix Table 12. CDC GWI Case Status

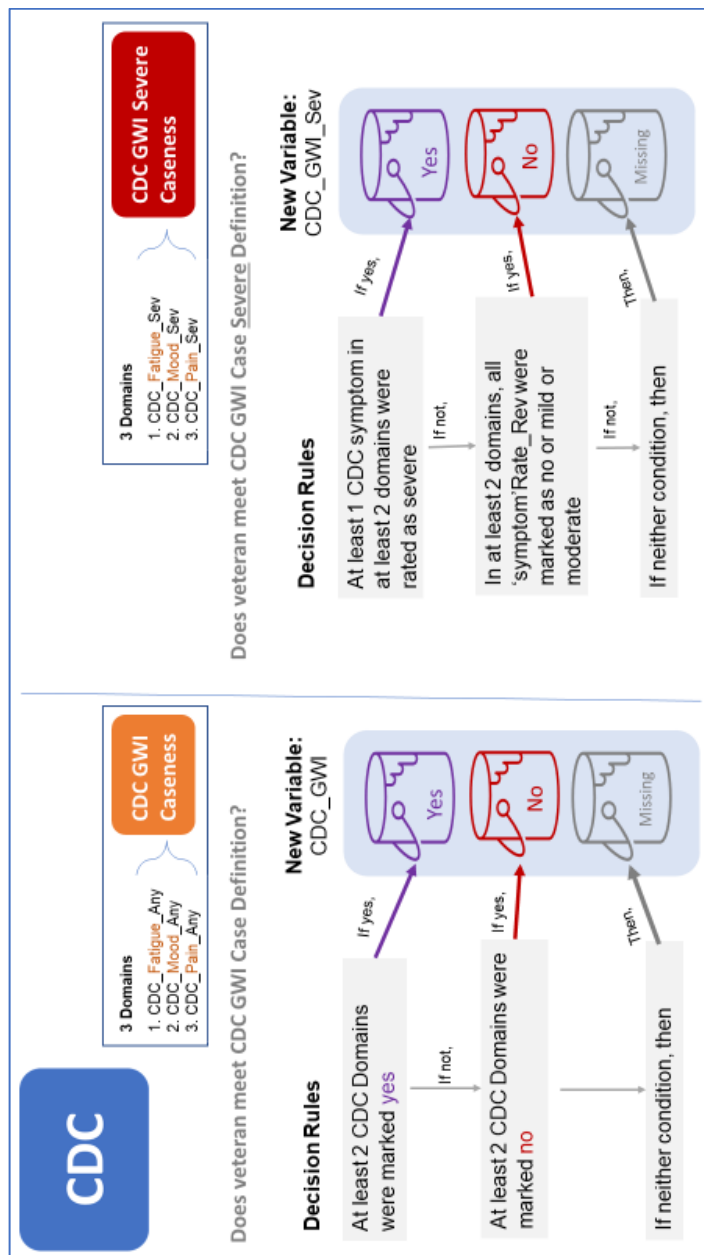
Table 5. Does the Veteran Meet the CDC GWI Case Definition?				
# of domains marked yes	# of domains with no			
	0	1	2	3
0	Insufficient Information to Determine Case Status		No	No
1			No	
2	YES	YES		
3	YES			
Black = Combination not possible because there are only 6 domains Yes = Meets CDC GWI Symptom Criteria No = Does not Meet GWI Symptom Criteria				

CDC Gulf War Illness Severity Case Status

“Yes” if at least one CDC symptom in at least two domains were ranked as severe.

“No” if, in at least two domains, all symptoms were marked as no, mild, or moderate.

“Missing” if neither of the above statements is true.



Appendix Figure 18. CDC and CDC Severe Case Definition Indicators

Summary of variables constructed to make the Kansas and CDC definitions

Appendix Table 13. Derived Variable List

Variable name	Variable Label	Value Label	Calculation	Notes
For each symptom				
'symptom'_Rev	“‘symptom label’ Revised to account for severity information”	1=Yes 0=No . =missing	Generates a new variable that replaces missing values with yes if a severity for that symptom was given	‘symptom’ refers to the original variable name
'symptom'Rate_Rev	“‘symptom’Rate label’ Revised to account for valid skips”	0=No 1=Mild 2=Moderate 3=Severe	Generates a new variable that codes missing values as severity scores as 0 rather than	This is helpful for programming the definitions; Drew has suggested

			missing – reflecting the valid skip pattern	more consideration of this for an analytic variable
For each diagnosis				
'diagnosis'_Rev	“diagnosis label' Revised to account for severity information”	1=Yes 0=No . =missing	Generates a new variable that replaces missing values with yes if a year told for that diagnosis was valid	'diagnosis' refers to the original variable name
For the Kansas Definition				
KS_Excl	“Meets Kansas Exclusion Criteria”	1=Yes 0=No . =missing	Yes= If at least 1 of the exclusionary criteria is yes, then the veteran was considered to have met the	exclusionary criteria are listed in Table 2

			<p>exclusion criteria;</p> <p>.=if the veteran is not yes but has a missing item among the exclusionary criteria then missing exclusion criteria</p> <p>0= if all of the exclusionary criteria are no</p>	
KS_Excl_NumMiss	“Number of missing exclusion criteria”	Integer: 0- (number of exclusionary conditions)	Number of missing items among the exclusionary criteria are	

			listed in Table 2	
KS_Excl_Count	Number of exclusionary criteria endorsed	Integer: 0- (number of exclusionary conditions)	Number of items endorsed yes among the exclusionary criteria (listed in Table 2)	
Kansas Moderate-to-Severe Domain Criteria (At least 1 symptom mild to moderate or more than 1 item is endorsed within domain)				
KS_Fatigue_ModMult	Fulfils the Kansas Fatigue domain criteria	1=Yes 0=No . =missing		
KS_Mood_ModMult	Fulfils the Kansas Mood domain criteria	1=Yes 0=No . =missing		

KS_Pain_ModMult	Fulfils the Kansas Pain domain criteria	1=Yes 0=No .=missing		
KS_GI_ModMult	Fulfils the Kansas GI domain criteria	1=Yes 0=No .=missing		
KS_Resp_ModMult	Fulfils the Kansas Respiratory domain criteria	1=Yes 0=No .=missing		
KS_Skin_ModMult	Fulfils the Kansas Skin domain criteria	1=Yes 0=No .=missing		
Gulf War Symptom Criteria				
KS_GWI_ModMult_S C	Meets the Kansas GWI	1=Yes 0=No		

	Symptom Criteria	.=missing		
KS GWI Definition				
KS_GWI	Kansas GWI case status	1=Yes 0=No . =missing		Individuals who met exclusion criteria are in the denominator
CDC				
CDC_Fatigue_Any	CDC Fatigue domain has any symptoms marked yes	1=Yes 0=No . =missing		
CDC_Mood_Any	CDC Mood domain has any symptoms marked yes	1=Yes 0=No . =missing		
CDC_Pain_Any	CDC Pain domain has any symptoms marked yes	1=Yes 0=No . =missing		

CDC_Fatigue_Sev	CDC Fatigue domain has >=1 symptom marked severe	1=Yes 0=No .=missing		
CDC_Mood_Sev	CDC Mood domain has >=1 symptom marked severe	1=Yes 0=No .=missing		
CDC_Pain_Sev	CDC Pain domain has >=1 symptom marked severe	1=Yes 0=No .=missing		
CDC_GWI	CDC GWI caseness	1=Yes 0=No .=missing		
CDC_GWI_Sev	CDC GWI caseness severe	1=Yes 0=No .=missing		
Other Items: Not in main code				
VHAStatus		1=Yes 0=No .=Missing	Used VA healthcare at least once in the past year	

			or were a patient in a VA healthcare facility overnight or longer at least once in the past year	
Deploy	Deployed during the 1990-1991 Gulf War	1=Yes 0=No . =Missing	Deployed to Gulf or Deployed Elsewhere	Considerable discussion about where to put “Deployed elsewhere”
Income_new		1= Under \$30,000 2= \$30,000-\$59,999 3= \$60,000-\$99,999 4= Over \$100,000	What income category represents the total income of your household from all sources during	

		. = Unknown	the past 12 months	
Education		1= "High School diploma/GE D or less" 2= "Some college to Associate's or Bachelor's Degree" 3= "Master's, Professional, or Doctorate degree" . = "Unknown"	Highest degree or level of school completed	
numConditions		Integer	Number of diagnoses endorsed yes	
BestAge		Integer	Age at time of survey, pulled from mix of sources	

ageGroup			Age group at time of survey	
raceEth		0= White, not Hispanic 1= Black, not Hispanic 2= Hispanic 3= Other 4 = Unknown	Race/Ethnicity	
Cancer_Rev			Fulfill any cancer exclusion	
HeartDis_Rev			Fulfill any heart disease exclusion	
ID_Rev			Fulfill any infectious disease exclusion	
Stroke_Rev			Fulfill any stroke exclusion	

MH_Rev			Fulfill any mental health exclusion	
NS_Rev			Fulfill any neurological exclusion	

Summary of SAS programs used to create KS and CDC Case

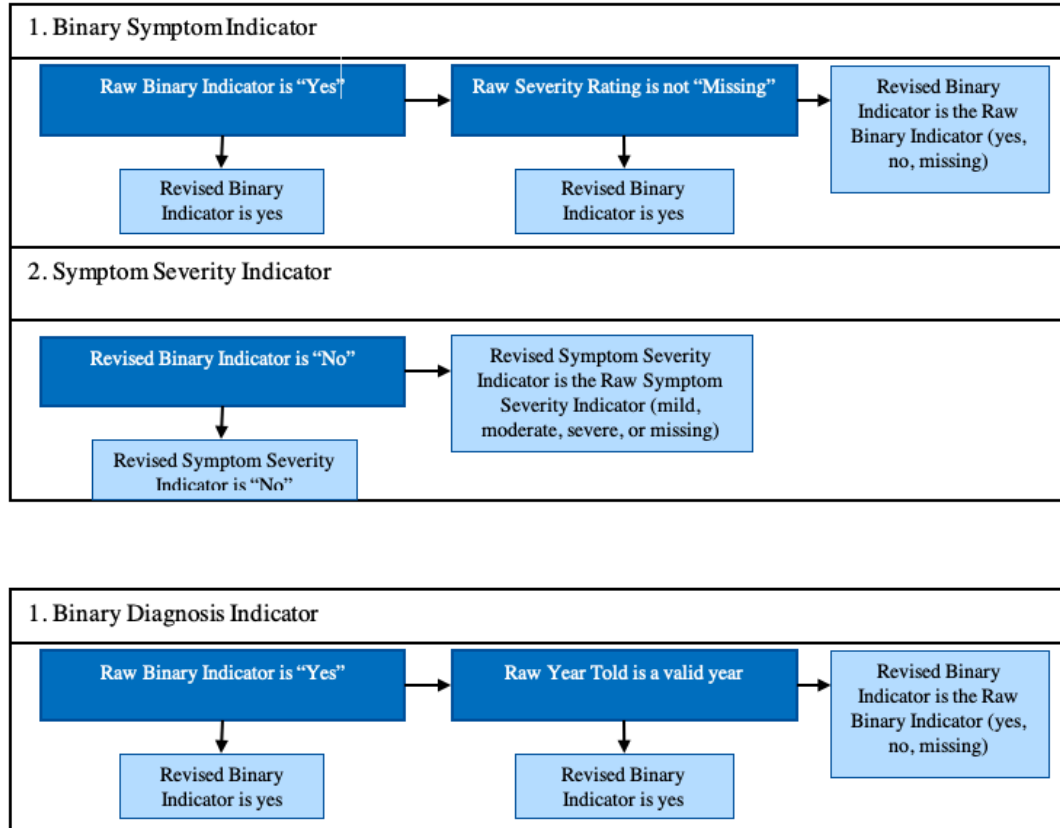
Status Indicators

Appendix Table 14. SAS programs to generate case definition indicators

Program Name	What program does	What datasets it creates
CSP585_GWI_(00)_main	Call all other sas program files for preparing the data Reads in the macros %array, %do_over, and %numlist	
Get data from data warehouse and Basic Data Management (e.g. labeling)		
CSP585_GWI_(01)_Readin Data	Reads in data from CSP_ProdCSP585.Res.vwTeleformSurveyParent_CombinedLogic	Work.CSP585DB
CSP585_GWI_(02)_Labels	Applies label and value formats	Updates Work. CSP585DB
CSP585_GWI_(03)_MissingData	Converts the missing value codes used in the SQL database to sas's missing value codes	Updates Work. CSP585DB
Gulf War Specific Work		
CSP585_GWI_(04)_Setup	Creates formats for yesno, rate, and revised rate	
CSP585_GWI_(05)_PartA	Creates 'symptom'_Rev Creates 'symptom'_Rate_Rev	GWIPartA

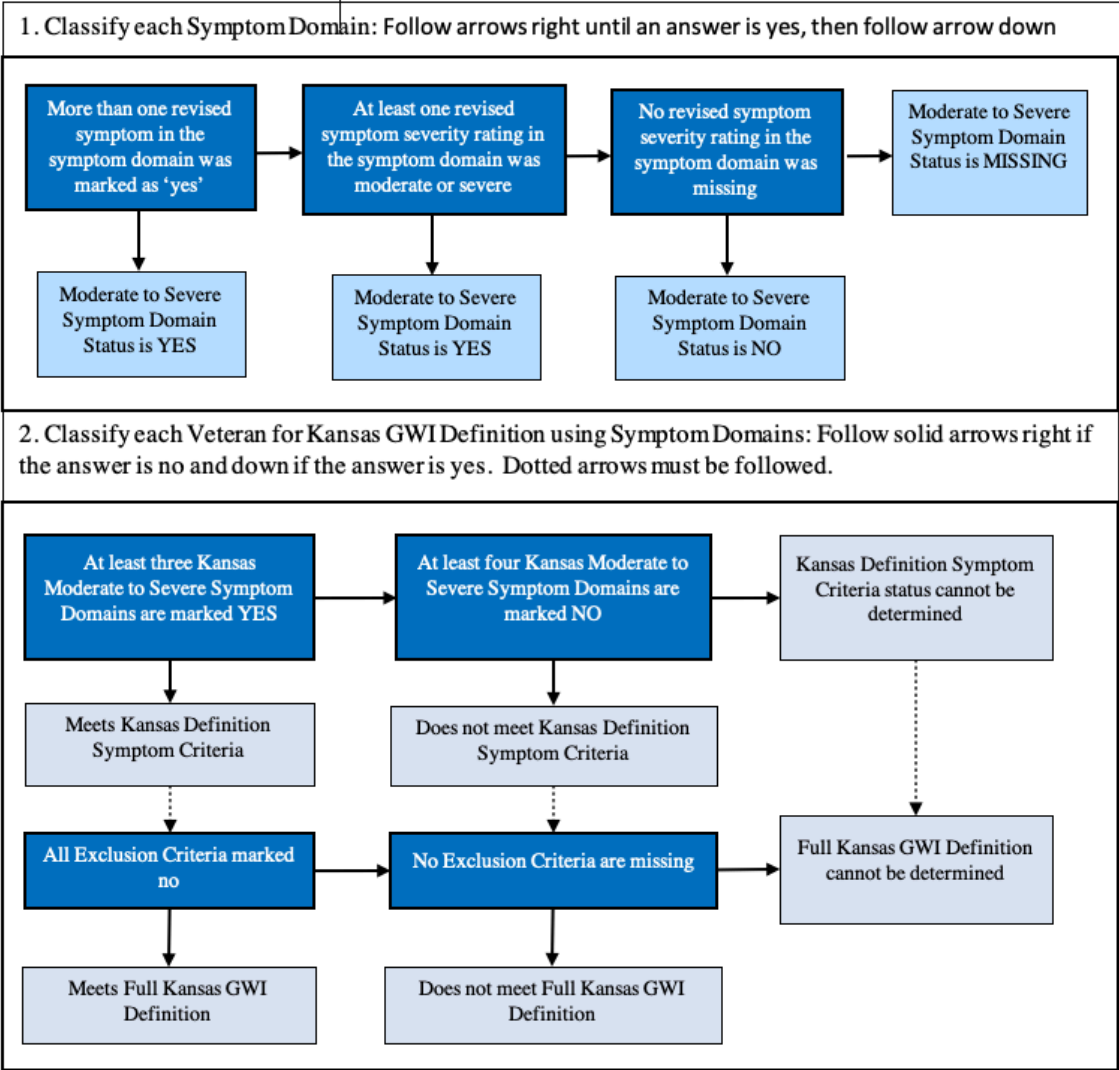
CSP585_GWI_(06)_PartAL abels	Labels 'symptom'_Rev ' & symptom'Rate_Rev	Updates GWI_PartA
Kansas Specific		
CSP585_GWI_(07)_KS_Pa rtB1	Creates KS_Excl variable KS_Excl_Nummiss	Creates: GWI_PartB1 From: GWI_PartA
CSP585_GWI_(08)_KS_Pa rt B2_Domains	Calculates the 6 KS Domains (moderate severity or multiple symptoms)	GWI_PartB2
CSP585_GWI_(09)_KS_Pa rtB3	Creates the symptom criteria	GWI_PartB3
CSP585_GWI_(10)_KS_Pa rtB4	Creates GWI Caseness (Kansas)	Creates: GWI_PartB4
CDC Specific		
CSP585_GWI_(11)_CDC_ PartC1	Calculates the 3 CDC Domains (Any) Calculates the 3 severe CDC Domains (Sev)	Creates: GWI_PartC1
CSP585_GWI_(12)_CDC_ PartC2	Creates GWI caseness (CDC and CDC severe)	Creates: GWI_PartC2

Full Pseudocode Flowchart: Revised Variables



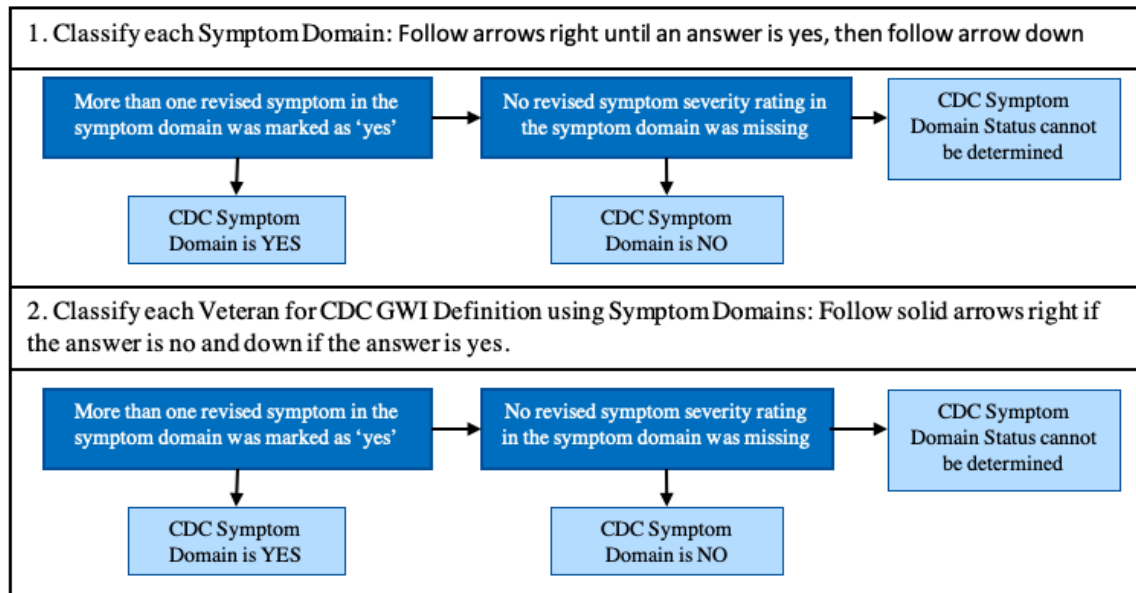
Appendix Figure 19. Pseudocode: symptom and diagnosis cleaning

Full Pseudocode Flow Chart: Kansas Definition



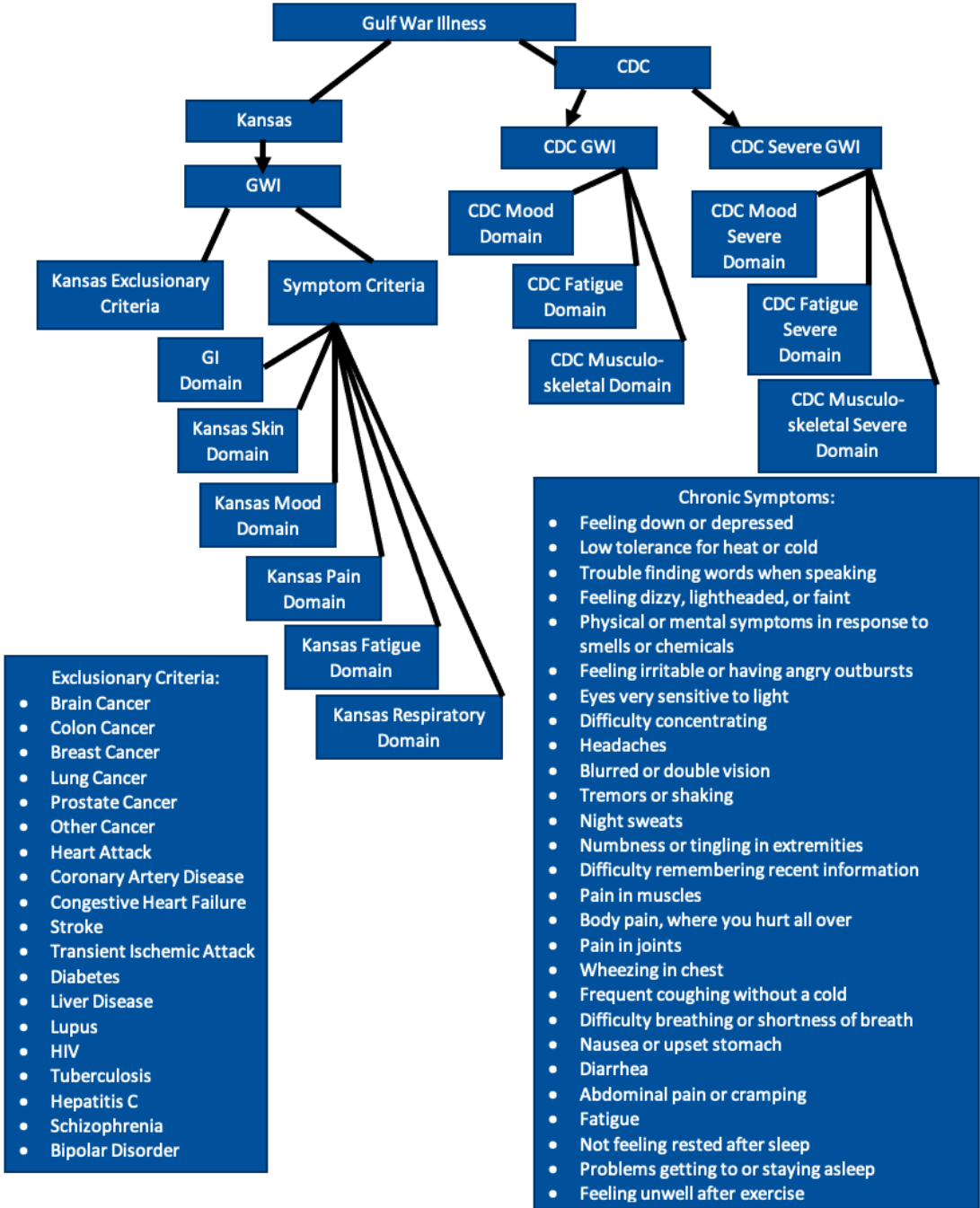
Appendix Figure 20. Pseudocode: Kansas GWI

Full Pseudocode Flow Chart: CDC Definition



Appendix Figure 21. Pseudocode: CDC GWI

Tree Diagram of Phenotypes



Appendix Figure 22. Tree diagram of phenotypes

Appendix E: GWI indicator SAS code

Notes and Citation

This code is published on GitHub: <https://github.com/VA-Phenomics-Library-CIPHER/Gulf-War-Illness>

SAS code was co-authored by Jackie Vahey and Beth Gifford.

CSP585_GWI_(00)_Main

```
/******
```

This code prepares data from CSP585 for analysis

- Replaces missing value codes with SAS missing values
- Labels variables and data
- Calculates GWI status and related information
- Prepares data to be used in analysis

```
*****/
```

```
/*To Use:
```

1. replace path in each filename syntax with the path to the proper files on your machine
2. be sure to have downloaded code for the %ARRAY and %DO_OVER macros

`%Do_Over` and `%Array` Authors: Ted Clay, M.S. and David Katz, M.S.

"Please keep, use and pass on the `ARRAY` and `DO_OVER` macros with this authorship note. "

Full documentation with examples appears in SUGI Proceedings, 2006, "Tight Looping With Macro Arrays" by Ted Clay

3. replace `%let` statements below with corresponding variable names from your dataset

```
*/
```

```
*Macros setup;
```

```
filename dr_macro "R:\CSP 585\AnalysisCode\585  
Manuscripts\Caseness\CODE\SAS\ClayTightLoopingmacros";
```

```
%INCLUDE dr_macro('ARRAY.sas');
```

```
%INCLUDE dr_macro('DO_OVER.sas');
```

```
%INCLUDE dr_macro('NUMLIST.sas');
```

```
*Path setup;
```

```
filename syntax "R:\CSP 585\AnalysisCode\585 Manuscripts\Caseness\CODE\SAS";
```

```
filename syntax2 "R:\CSP 585\AnalysisCode\585
Manuscripts\Caseness\CODE\SAS\PrepareDataForWarehous";

/*Step 0a. Create a place to store final dataset and formats*/

libname out 'R:\CSP 585\AnalysisCode\585
Manuscripts\Caseness\CODE\SAS\Derived SAS'; run; *change this too;
```

/*Step 0b. Edit 'quick change' pieces of the code

- Missing values
- Kansas definition: exclusionary conditions
- Kansas definition: symptoms in domains
- CDC definition: symptoms in domains */

```
*Missing Values;
```

```
%let missing_values= (. .a .b .c .d .e .f .g .h .g);
```

```
%let missing_values_notb = (. .a .c .d .e .f .g .h .g); *this doesn't include
```

'multiple answers', allowing multiple answers to count for moderate to multiple;

```
*Kansas definition: exclusionary conditions;
```

```
%let exclCondList = CaBrain CaBrst CaColon CaLung CaPros CaOth
```

```
DoDM
```

CircHrtAtk CircCAD CircCHF

CircStrk CircTIA

IDHIV IDTB IDHepC

DoLiver DoLupus

MHSCZ MHBPD NSMS NSTBI ;

*all symptoms;

%let allSymptomsList = SMoFatigue SMoUnwel SMoGetSlp SMoRested

SMoPain SMoStiff

SMoMuscl SMoHurtOvr SMoHeadA SMoDizzy SMoSLight

SMoBlur SMoNumb SMoShake SMoLowTol

SMoSweat SMoSmell SMoRash SMoSkin SMoDiarr

SMoNaus SMoAbdom SMoDBreath SMoFCough SMoWheez

SMoSThroat SMoLymph SMoDConct SMoDReme SMoTFind

SMoFDown SMoOutbur SMoMoody SMoAnxio ;

*Kansas definition: symptoms in domains;

%let KSFatigueSymptomList = SMoFatigue SMoGetSlp SMoRested SMoUnwel ;

%let KSPainSymptomList = SMoPain SMoMuscl SMoHurtOvr ;

%let KSMoodSymptomList = SMoDReme SMoOutbur SMoNumb SMoHeadA

SMoSLight SMoTFind SMoFDown SMoDConct SMoSweat SMoDizzy SMoLowTol

SMoSmell SMoBlur SMoShake ;

```

%let KSGISymptomList = SMOdiarr SMOnaus SMOabdom ;

%let KSRespSymptomList = SMOdbreath SMOfcough SMOwheez ;

%let KSSkinSymptomList = SMOrash SMOskin;

*CDC definition: symptoms in domains;

%let CDCFatigueSymptomList = SMOfatigue ;

%let CDCMuscSkelSymptomList = SMOpain SMOmuscl SMOstiff ;

%let CDCMoodSymptomList = SMOdreme SMOtfind SMOfdown SMOdconct
SMOMOODY SMOANXIO SMOGETSLP ;

```

```

/**Step 1. Read in the Data**/

```

```

%INCLUDE syntax('CSP585_GWI_(01)_ReadinData.sas');

```

```

/*Step 2. Label Variables and Labels--This runs Code that is Generated by:

```

```

["R:\CSP 585\AnalysisCode\Beth-Explore\SQL\C-FORREVIEW-
MISSING-VALUE-CODES-SQL-CreateSasFormatting.sql"]*/

```

```

%INCLUDE syntax('CSP585_GWI_(02)_Labels.sas');

```

```

*This creates a file called new;

```



```

*Creates formats for yesno, rate, and revised rate;

    %INCLUDE syntax('CSP585_GWI_(04)_Setup.sas');

run;

*Creates 'symptom'_Rev and 'symptom'Rate_Rev;

*Creates dataset GWI_PartA;

    %INCLUDE syntax('CSP585_GWI_(05)_PartA.sas');

run;

*labels 'symptom'_Rev and 'symptom'Rate_Rev;

*updates dataset GWI_PartA;

    %INCLUDE syntax('CSP585_GWI_(06)_PartALabels.sas');

run;

/*Kansas Specific*/

*Creates KS_Excl and KS_Excl_Nummiss;

*Creates dataset GWI_PartB1 from GWI_PartA;

    %INCLUDE syntax('CSP585_GWI_(07)_KS_PartB1.sas');

run;

*Calculates the 6 KS Domains (Moderate or Multiple);

*Creates dataset GWI_PartB2 from GWI_PartA;

    %INCLUDE syntax('CSP585_GWI_(08)_KS_PartB2_Domains.sas');

run;

```

```

*Creates the KS Symptom Criteria;

*Creates dataset GWI_PartB3 from GWI_PartB2;

    %INCLUDE syntax('CSP585_GWI_(09)_KS_PartB3.sas');

run;

*Calculates KS GWI Caseness;

*Creates dataset GWI_PartB4 from GWI_PartB3;

    %INCLUDE syntax('CSP585_GWI_(10)_KS_PartB4.sas');

run;

/*CDC Specific*/

*Calculates the 3 CDC symptom domains, Any and Severe;

*Creates dataset GWI_PartC1 from GWI_PartB4;

    %INCLUDE syntax('CSP585_GWI_(11)_CDC_PartC1.sas');

run;

*Calculates CDC GWI Caseness: Any and Severe;

*Creates dataset GWI_PartC2 from GWI_PartC1;

    %INCLUDE syntax('CSP585_GWI_(12)_CDC_PartC2.sas');

run;

* run basic tabulations for GWI;

* GWI x (sex, deployment, VA User);

```

```
        %INCLUDE syntax('CSP585_GWI_(13)_demographics.sas');  
  
run;
```

```
/* Save a permanent sas dataset*/  
  
data out.CSP585_GWI;  
  
    set GWI_demographics;  
  
run;
```

CSP585_GWI_(01)_ReadinData.sas

```
*** This prevents errors when bringing in data;
```

```
options nofmterr;
```

```
*** 1. Get list of tables that YOU can access ***;
```

```
*Ideally, I can modify this part of the code to account for the missing values in proc SQL;
```

```
proc sql;
```

```
    connect to odbc(dsn='CSP_ProdCSP585');
```

```
    CREATE TABLE new AS
```

```
        SELECT *
```

```

FROM CONNECTION TO ODBC(
SELECT *
        FROM CSP_ProdCSP585.Res.vwTeleformSurveyParent_CombinedLogic
        ;
);

disconnect from odbc;

quit;

/* load bestAge */

proc sql;

connect to odbc(dsn='CSP_ProdCSP585');

CREATE TABLE getages AS

SELECT *

FROM CONNECTION TO ODBC(

SELECT *

        FROM CSP_ProdCSP585.Res.vwTeleformSurveyAges

        ;

);

disconnect from odbc;

```

```
quit;
```

```
    *This creates a file called new;
```

```
run;
```

CSP585_GWI_(02)_Labels.sas

**/ This code can be found on the GitHub site and is omitted due to space constraints.*

*The code labels each variable and each value for each variable. /**

CSP585_GWI_(03)_MissingData.sas

**/ This code can be found on the GitHub site and is omitted due to space constraints.*

The code identifies all variables that have 4 digit missing value codes,

5 digit missing value codes, and 6-digit missing value codes. Then it converts those

*missing value codes into SAS missing value indicators (".") /**

CSP585_GWI_(04)_Setup.sas

```
proc format;
```

```
value l_yesno
```

```
0= "No"
```

```
1= "Yes"
```

```
. = "Missing"
```

```
.a = "Blank"
```

```
.b = "Multiple Answers"
```

```
.c = "Invalid Value"
```

```
.d = "Below Minimum"
```

```
.e = "Above Maximum"
```

```
.f = "Write In"
```

```
.g = "Invalid Value"
```

```
.h = "Expected Missing"
```

```
;
```

```
value l_Rate
```

```
0="Mild"
```

```
1="Moderate"
```

```
2="Severe"
```

```
. = "Missing"
```

```
.a = "Blank"
```

.b = "Multiple Answers"

.c = "Invalid Value"

.d = "Below Minimum"

.e = "Above Maximum"

.f = "Write In"

.g = "Invalid Value"

.h = "Expected Missing"

;

value 1_RevisedRate

0="No, None"

1="Yes, Mild"

2="Yes, Moderate"

3="Yes, Severe"

. = "Missing"

.a = "Blank"

.b = "Multiple Answers"

.c = "Invalid Value"

.d = "Below Minimum"

.e = "Above Maximum"

.f = "Write In"

```

        .g = "Invalid Value"

        .h = "Expected Missing"

    ;

run;

```

CSP585_GWI_(05)_PartA.sas

```

data GWI_PartA;

    set new;

    /* Step A1. Symptoms from the Kansas and CDC Measures*/

    %Array (symptoms, values = &allSymptomsList);

    array symptom (&symptomsn) %DO_OVER(symptoms, phrase=?);

        *build array of symptoms;

    array symptomRate(&symptomsn) %Do_OVER(symptoms, phrase=?rate);

        *build array of symptomRates;

    array symptom_Rev(&symptomsn) %DO_OVER(symptoms, phrase=?_rev);

        *build array of revised symptoms;

    array symptomRate_Rev(&symptomsn) %DO_OVER(symptoms,
phrase=?rate_rev);    *build array of revised symptomRates;

```

*Step A2. Infer a symptom is present if responded rated the severity as mild, moderate or severe;

```
do i=1 to &symptomsn;
```

```
    symptom_REV[i]=symptom[i];      *revised symptom is the symptom
```

```
unless a symptom rate is recorded;
```

```
    if symptom[i] in(0 . .a .b .c .d .e .f .g .h) & symptomRate[i] in(0 1 2) then
```

```
symptom_Rev[i]=1;
```

```
end;
```

*Step A3. Create new variables that revise the severity score from 0 “No” to 3 “Severe”;

```
do i=1 to &symptomsn;
```

```
    symptomRate_Rev[i]=symptomRate[i];
```

```
    if symptomRate[i] in (0,1,2) then symptomRate_Rev[i] =  
symptomRate[i]+1    ;
```

```
    /*This line of code is trying not to replace values for people who said no  
but then provided a symptom*/
```

```
    if symptomRate[i] in(. .a .b .c .d .e .f .g .h) & symptom_Rev[i]=0 then
```

```
symptomRate_Rev[i]=0;
```

```
end;
```

*Format each type of variable by array;

```

format %DO_OVER(symptoms, phrase=?) l_yesno.;

format %Do_OVER(symptoms, phrase=?rate) l_Rate.;

format %DO_OVER(symptoms, phrase=?_rev) l_yesno.;

format %DO_OVER(symptoms, phrase=?rate_rev) l_RevisedRate.;

run;

```

CSP585_GWI_(06)_PartALabels.sas

```

data GWI_partA;

  set GWI_PartA;

  %let revisedDescript = Revised: Over past 6 months, persistent or recurring problem
with: ;

  %let revisedRateDescript = Revised: Over past 6 months, rate the persistent or
recurring problem with: ;

  label SMoFatigue_rev = "&revisedDescript Fatigue";

  label SMoFatigueRate_rev = "&revisedRateDescript Fatigue";

  label SMoUnwel_rev = "&revisedDescript Feeling unwell after physical exercise or
exertion";

  label SMoUnwelRate_rev = "&revisedRateDescript Feeling unwell after physical
exercise or exertion";

```

label SMoGetSlp_rev = "&revisedDescript Problems getting to sleep or staying asleep";

label SMoGetSlpRate_rev = "&revisedRateDescript Problems getting to sleep or staying asleep";

label SMoRested_rev = "&revisedDescript Not feeling rested after you sleep";

label SMoRestedRate_rev = "&revisedRateDescript Not feeling rested after you sleep";

label SMoPain_rev = "&revisedDescript Pain in your joints";

label SMoPainRate_rev = "&revisedRateDescript Pain in your joints";

label SMoStiff_rev = "&revisedDescript Stiffness in your joints";

label SMoStiffRate_rev = "&revisedRateDescript Stiffness in your joints";

label SMoMuscl_rev = "&revisedDescript Pain in your muscles";

label SMoMusclRate_rev = "&revisedRateDescript Pain in your muscles";

label SMoHurtOvr_rev = "&revisedDescript Body pain, where you hurt all over";

label SMoHurtOvrRate_rev = "&revisedRateDescript Body pain, where you hurt all over";

label SMoHeadA_rev = "&revisedDescript Headaches";

label SMoHeadARate_rev = "&revisedRateDescript Headaches";

label SMoDizzy_rev = "&revisedDescript Feeling dizzy, lightheaded, or faint";

label SMoDizzyRate_rev = "&revisedRateDescript Feeling dizzy, lightheaded, or faint";

label SMoSLight_rev = "&revisedDescript Eyes very sensitive to light";

label SMoSLightRate_rev = "&revisedRateDescript Eyes very sensitive to light";

label SMoBlur_rev = "&revisedDescript Blurred or double vision";

label SMoBlurRate_rev = "&revisedRateDescript Blurred or double vision";

label SMoNumb_rev = "&revisedDescript Numbness or tingling in your extremeties";

label SMoNumbRate_rev = "&revisedRateDescript Numbness or tingling in your extremeties";

label SMoShake_rev = "&revisedDescript Tremors or shaking";

label SMoShakeRate_rev = "&revisedRateDescript Tremors or shaking";

label SMoLowTol_rev = "&revisedDescript Low tolerance for heat or cold";

label SMoLowTolRate_rev = "&revisedRateDescript Low tolerance for heat or cold";

label SMoSweat_rev = "&revisedDescript Night sweats";

label SMoSweatRate_rev = "&revisedRateDescript Night sweats";

label SMoSmell_rev = "&revisedDescript Having physical or mental symptoms in response to certain smells or chemicals";

label SMoSmellRate_rev = "&revisedRateDescript Having physical or mental symptoms in response to certain smells or chemicals";

label SMoRash_rev = "&revisedDescript Skin rashes";

label SMoRashRate_rev = "&revisedRateDescript Skin rashes";

label SMoSkin_rev = "&revisedDescript Other skin problems";

label SMoSkinRate_rev = "&revisedRateDescript Other skin problems";

label SMoDiarr_rev = "&revisedDescript Diarrhea";

label SMoDiarrRate_rev = "&revisedRateDescript Diarrhea";

label SMoNaus_rev = "&revisedDescript Nausea or upset stomach";

label SMoNausRate_rev = "&revisedRateDescript Nausea or upset stomach";

label SMoAbdom_rev = "&revisedDescript Abdominal pain or cramping";

label SMoAbdomRate_rev = "&revisedRateDescript Abdominal pain or cramping";

label SMoDBreath_rev = "&revisedDescript Difficulty breathing or shortness of
breath";

label SMoDBreathRate_rev = "&revisedRateDescript Difficulty breathing or shortness
of breath";

label SMoFCough_rev = "&revisedDescript Frequent coughing when you don't have a
cold";

label SMoFCoughRate_rev = "&revisedRateDescript Frequent coughing when you
don't have a cold";

label SMoWheez_rev = "&revisedDescript Wheezing in your chest";

label SMoWheezRate_rev = "&revisedRateDescript Wheezing in your chest";

label SMoSThroat_rev = "&revisedDescript Sore throat";

label SMoSThroatRate_rev = "&revisedRateDescript Sore throat";

label SMoLymph_rev = "&revisedDescript Tender lymph nodes in your neck or
armpits:";

label SMoLymphRate_rev = "&revisedRateDescript Tender lymph nodes in your neck
or armpits";

label SMoDConct_rev = "&revisedDescript Difficulty concentrating";

label SMoDConctRate_rev = "&revisedRateDescript Difficulty concentrating";

label SMoDReme_rev = "&revisedDescript Difficulty remembering recent
information";

label SMoDRemeRate_rev = "&revisedRateDescript Difficulty remembering recent
information";

label SMoTFind_rev = "&revisedDescript Trouble finding words when speaking";

label SMoTFindRate_rev = "&revisedRateDescript Trouble finding words when
speaking";

label SMoFDown_rev = "&revisedDescript Feeling down or depressed";

label SMoFDownRate_rev = "&revisedRateDescript Feeling down or depressed";

label SMoOutbur_rev = "&revisedDescript Feeling irritable or having angry outbursts";

label SMoOutburRate_rev = "&revisedRateDescript Feeling irritable or having angry
outbursts";

label SMoMoody_rev = "&revisedDescript Feeling moody";

label SMoMoodyRate_rev = "&revisedRateDescript Feeling moody";

label SMoAnxio_rev = "&revisedDescript Feeling anxious";

label SMoAnxioRate_rev = "&revisedRateDescript Feeling anxious";

run;

CSP585_GWI_(07)_KS_PartB1.sas

```
data GWI_PartB1;
```

```
    set GWI_PartA;
```

```
    *If exclusionary diagnoses change, edit this list and the labels at the
```

```
bottom of the page;
```

```
    %Array(exclVars, values = &exclCondList);
```

```
    array Array_KS_Excl (&exclVarsn) %DO_OVER(exclVars, phrase=?);
```

```
    array Array_KS_Excl_yrdx (&exclvarsn) %DO_OVER(exclvars,  
phrase=?YrDx);
```

```
    array Array_KS_Excl_Rev (&exclVarsn) %DO_OVER(exclVars,  
phrase=?_Rev);
```

```
    do i=1 to &exclVarsn;
```

```
        Array_KS_Excl_Rev[i]=Array_KS_Excl[i];    *revised exclusion is  
the original unless a dx year is recorded;
```

```
        if Array_KS_Excl_yrdx[i] not in &missing_values then  
Array_KS_Excl_Rev[i]=1;
```

```
    end;
```

```

KS_Excl=0;

label KS_Excl = "Has 1 or more exclusion criteria";

KS_Excl_Nummiss=0;

label KS_Excl_Nummiss = "Number of missing exclusion items";

KS_Excl_Count=0;

label KS_Excl_Count = "Number of exclusion items endorsed as yes";

do i=1 to &exclVarsn;

    if Array_KS_Excl_Rev[i]=1 then KS_Excl=1;

    if Array_KS_Excl_Rev[i]=1 then KS_Excl_Count = KS_Excl_Count
+ 1;

    if Array_KS_Excl_Rev[i] in( &missing_values) then

KS_Excl_Nummiss= KS_Excl_Nummiss+1;

end;

if KS_Excl=0 & KS_Excl_Nummiss>0 then KS_Excl=.;

*will need to change these labels if the exclusionary diagnoses change;

%let revisedDescript = Revised: Healthcare provider told you that you
have: ;

label CaBrain_Rev = "&revisedDescript Brain Cancer";

label CaBrst_Rev = "&revisedDescript Breast Cancer";

label CaColon_Rev = "&revisedDescript Colon Cancer";

```

```
label CaLung_Rev = "&revisedDescript Lung Cancer";  
label CaPros_Rev = "&revisedDescript Prostate Cancer";  
label CaSkin_Rev = "&revisedDescript Skin Cancer";  
label CaOth_Rev = "&revisedDescript Other Cancer";  
label DoDM_Rev = "&revisedDescript Diabetes";  
label CircHrtAtk_Rev = "&revisedDescript Heart Attack";  
label CircCAD_Rev = "&revisedDescript Coronary Artery Disease";  
label CircCHF_Rev = "&revisedDescript Congestive Heart Failure";  
label CircStrk_Rev = "&revisedDescript Stroke";  
label CircTIA_Rev = "&revisedDescript Transient Ischemic Attack (TIA)";  
label IDHIV_Rev = "&revisedDescript HIV";  
label IDTB_Rev = "&revisedDescript Tuberculosis";  
label IDHepC_Rev = "&revisedDescript Hepatitis C";  
label DoLiver_Rev = "&revisedDescript Liver Condition";  
label DoLupus_Rev = "&revisedDescript Lupus";  
label MHSCZ_Rev = "&revisedDescript Schizophrenia";  
label MHBPD_Rev = "&revisedDescript Bipolar Disorder";  
label NSMS_Rev = "&revisedDescript Multiple Sclerosis";  
label NSTBI_Rev = "&revisedDescript Traumatic Brain Injury";  
format %DO_OVER(exclVars, phrase=?_Rev) l_yesno;
```

run;

CSP585_GWI_(08)_KS_PartB2_Domains.sas

```
%macro calcKSDomain(KSdomain= , symptomList = , domainDescript= );
```

```
    &KSdomain._Any=0;
```

```
        label &KSdomain._Any= " &domainDescript Domain: Endorsed at least 1  
criteria";
```

```
    &KSdomain._Num=0;
```

```
        label &KSdomain._Num= " &domainDescript Domain: # of Items  
endorsed"    ;
```

```
    &KSdomain._ModSev=0;
```

```
        label &KSdomain._ModSev= " &domainDescript Domain: Endorsed at  
least 1 criteria as moderate or severe";
```

```
    &KSdomain.Rate_NumMiss=0; /* This variable is to store the number of missing  
items on severity*/
```

```
        label &KSdomain.Rate_NumMiss= " &domainDescript Domain: # of  
items with missing severity rating";
```

```
    &KSdomain._ModMult=0;
```

```

label &KSdomain._ModMult = " &domainDescript Domain: Meets
moderate to severe or multiple criteria";

/* do the work to calculate accurate values for these variables */

%array (domain_vars, values=&symptomList);

array &KSdomain._items(&domain_varsn) %DO_OVER(domain_vars,
phrase=?_rev);

array &KSdomain._rates(&domain_varsn) %DO_OVER(domain_vars,
phrase=?rate_Rev);

/* Step1 a. Replace if any item was endorsed*/

do i=1 to &domain_varsn;

    if &KSdomain._items[i]=1 then &KSdomain._Any=1;

/* Step1 b. Count number of items endorsed*/

    if &KSdomain._items[i]=1 then &KSdomain._Num=
&KSdomain._Num + 1;

end;

/*Step2 Identify if any items were endorsed as moderate or severe*/

do i=1 to &domain_varsn;

    if &KSdomain._rates[i] in (.b 2 3) then &KSdomain._ModSev=1;

```

```

        if &KSdomain._rates[i] in (&missing_values_notb) then
&KSdomain.Rate_NumMiss= &KSdomain.Rate_NumMiss+1;

        end;

/* Step 3. Does person meet Kansas moderate or multiple Domain */

if &KSdomain._Num >1 then &KSdomain._ModMult=1;

        else if &KSdomain._ModSev=1 then &KSdomain._ModMult=1;

        else if &KSdomain.Rate_NumMiss=0 then &KSdomain._ModMult=0;

        else &KSdomain._ModMult=.;

%mend;

data GWI_PartB2;

        set GWI_PartB1;

        %calcKSDomain(KSdomain= KS_Fatigue, symptomList =
&KSFatigueSymptomList, domainDescript= Fatigue);

        %calcKSDomain(KSdomain= KS_Pain, symptomList = &KSPainSymptomList,
domainDescript= Pain);

        %calcKSDomain(KSdomain= KS_Mood, symptomList = &KSMoodSymptomList,
domainDescript= Neurological Mood and Cognition);

        %calcKSDomain(KSdomain= KS_GI, symptomList = &KSGISymptomList,
domainDescript= Gastrointestinal);

```

```

        %calcKSDomain(KSdomain= KS_Resp, symptomList = &KSRespSymptomList,
domainDescript= Respiratory);

        %calcKSDomain(KSdomain= KS_Skin, symptomList = &KSSkinSymptomList,
domainDescript= Skin) ;

run;

```

CSP585_GWI_(09)_KS_PartB3.sas

```

data GWI_PartB3; set GWI_PartB2;

        KS_GWI_ModMult_SC=.;

        label KS_GWI_ModMult_SC = "Meets the Kansas GWI Symptom
Criteria";

        %let KS_domains = KS_Fatigue_ModMult, KS_Pain_ModMult,
KS_Mood_ModMult, KS_GI_ModMult, KS_Resp_ModMult, KS_Skin_ModMult ;

        KS_domains_Yes = SUM(&KS_domains);    *number of domains that are
endorsed as moderate/severe or multiple;

        if KS_domains_Yes = . then KS_domains_Yes = 0; *if all are missing, the sum will
be missing, but we want it to be zero;

        label KS_domains_Yes = "Numer of Kansas domains endorsed as yes";

```

```
KS_domains_Miss = NMISS(&KS_domains); *number of domains that could not  
be determined as Y/N for moderate/severe or multiple;
```

```
label KS_domains_Miss = "Numer of Kansas domains that cannot be determined  
as yes or no";
```

```
KS_domains_No = 6 - KS_domains_Yes - KS_domains_Miss; *Number of  
domains that are endorsed as not moderate/severe or multiple;
```

```
label KS_domains_No = "Numer of Kansas domains endorsed as no";
```

```
if KS_domains_Yes >= 3 then KS_GWI_ModMult_SC=1; * IFF 3 domains marked  
yes then GWI caseness is yes;
```

```
else if KS_domains_No >= 4 then KS_GWI_ModMult_SC = 0; *IFF  
there are at least 4 No domains, the GWI caseness is no;
```

```
else KS_GWI_ModMult_SC = .; *otherwise we don't know enough to  
decide;
```

```
run;
```

CSP585_GWI_(10)_KS_PartB4.sas

```
data GWI_PartB4; set GWI_PartB3;
```

```
KS_GWI=0;
```

```
label KS_GWI = "Kansas Gulf War Illness Caseness Indicator";
```

```

format KS_GWI1_yesno.;

if KS_Excl = 0 & KS_GWI_ModMult_SC = 1 then KS_GWI = 1;      *Veterans
with no exclusionary criteria who fit the symptoms criteria are marked 1;

else if KS_Excl = 1 & KS_GWI_ModMult_SC = 1 then KS_GWI = 0;

*Veterans with exclusionary criteria who fit the symptoms criteria are marked 0;

else if KS_Excl = 1 & KS_GWI_ModMult_SC = 0 then KS_GWI = 0;

*Veterans with exclusionary criteria who don't fit the symptoms criteria are marked 0;

else if KS_Excl = 0 & KS_GWI_ModMult_SC = 0 then KS_GWI = 0;

*Veterans with no exclusionary criteria who don't fit the symptoms criteria are marked
0;

else KS_GWI = . ;

      *All other veterans are marked missing;

run;

```

CSP585_GWI_(11)_CDC_PartC1.sas

```

%macro calcCDCDomain(CDCdomain= , symptomList = , domainDescript= );

&CDCdomain._Any=0;

label &CDCdomain._Any= " &domainDescript Domain: Endorsed at least
1 criteria";

```

```

format &CDCdomain._Any 1_yesno.;

&CDCdomain._Sev=0;

label &CDCdomain._Sev= " &domainDescript Domain: Endorsed at least
1 criteria as severe";

format &CDCdomain._Sev 1_yesno.;

&CDCdomain.Rate_NumMiss=0; /* This variable is to store the number of
missing items on severity*/

label &CDCdomain.Rate_NumMiss= " &domainDescript Domain: # of
items with missing severity rating";

/* do the work to calculate accurate values for these variables */

%array (domain_vars, values=&symptomList);

array &CDCdomain._items(&domain_varsn) %DO_OVER(domain_vars,
phrase=?_rev);

array &CDCdomain._rates(&domain_varsn) %DO_OVER(domain_vars,
phrase=?rate_Rev);

/* Step1 Replace if any item was endorsed*/

do i=1 to &domain_varsn;

    if &CDCdomain._items[i]=1 then &CDCdomain._Any=1;

end;

/*Step2 Identify if any items were endorsed as severe*/

```

```

do i=1 to &domain_varsn;

    if &CDCdomain._rates[i]=3 then &CDCdomain._Sev=1;

    if &CDCdomain._rates[i] in (&missing_values) then

&CDCdomain.Rate_NumMiss= &CDCdomain.Rate_NumMiss+1;

    end;

    if &CDCdomain.Rate_NumMiss > 0 & &CDCdomain._Any=0    then

&CDCdomain._Any= .    ; *Any is missing if nothing is endorsed and something is

blank;

    if &CDCdomain.Rate_NumMiss > 0 & &CDCdomain._Sev=0    then

&CDCdomain._Sev= . ; *Sev is missing if nothing is severe and something is blank;

%mend;

data GWI_PartC1;

    set GWI_PartB4;

    %calcCDCDomain(CDCdomain= CDC_Fatigue, symptomList =

&CDCFatigueSymptomList , domainDescript= Fatigue);

    %calcCDCDomain(CDCdomain= CDC_MuscSkel, symptomList =

&CDCMuscSkelSymptomList , domainDescript= Musculoskeletal);

    %calcCDCDomain(CDCdomain= CDC_Mood, symptomList =

&CDCMoodSymptomList , domainDescript= Mood and Cognition);

run;

```

CSP585_GWI_(12)_CDC_PartC2.sas

```
data GWI_PartC2; set GWI_PartC1;

    *Set up variables;

    CDC_GWI=.;

        label CDC_GWI = "Meets the CDC GWI Definition";

        format CDC_GWI 1_yesno.;

    CDC_GWI_Sev=.;

        label CDC_GWI_Sev = "Meets the CDC GWI Definition, severe";

        format CDC_GWI_Sev 1_yesno.;

    *Calculate CDC_GWI;

    %let CDC_domains = CDC_Fatigue_Any, CDC_MuscSkel_Any,
    CDC_Mood_Any    ;

        CDC_domains_Yes = SUM(&CDC_domains);    *number of domains that are
    endorsed as yes;

        label CDC_domains_Yes = "Number of CDC domains endorsed as yes";

        CDC_domains_Miss = NMISS(&CDC_domains); *number of domains that could
    not be determined as Y/N ;
```

```

label CDC_domains_Miss = "Number of CDC domains that cannot be
determined as yes or no";

CDC_domains_No = 3 - CDC_domains_Yes - CDC_domains_Miss; *Number of
domains that are endorsed as no;

label CDC_domains_No = "Number of CDC domains endorsed as no";

if CDC_domains_Yes >= 2 then CDC_GWI=1; * IFF 2 domains marked yes then
GWI caseness is yes;

else if CDC_domains_No >= 2 then CDC_GWI = 0; *IFF there are at least
2 No domains, the GWI caseness is no;

else CDC_GWI = .; *otherwise we don't know enough to decide;

*Calculate CDC_GWI_Sev;

%let CDC_domains_sev = CDC_Fatigue_Sev, CDC_MuscSkel_Sev,
CDC_Mood_Sev ;

CDC_domains_Yes_Sev = SUM(&CDC_domains_sev); *number of domains
that are endorsed as severe;

label CDC_domains_Yes_Sev = "Number of CDC severe domains
endorsed as no";

CDC_domains_Miss_Sev = NMISS(&CDC_domains_sev); *number of domains
that could not be determined as Y/N for severe;

```

```

label CDC_domains_Miss_Sev = "Number of CDC severe domains that
cannot be determined as yes or no";

CDC_domains_No_Sev = 3 - CDC_domains_Yes_Sev - CDC_domains_Miss_Sev;
*Number of domains that are endorsed as not severe;

label CDC_domains_No_Sev = "Number of CDC severe domains
endorsed as no";

if CDC_domains_Yes_Sev >= 2 then CDC_GWI_Sev=1; * IFF 2 domains marked
yes then GWI caseness is yes;

else if CDC_domains_No_Sev >= 2 then CDC_GWI_Sev = 0;      *IFF
there are at least 2 No domains, the GWI caseness is no;

else CDC_GWI_Sev = .; *otherwise we don't know enough to decide;

run;

```

CSP585_GWI_(13)_demographics.sas

```

*uncomment below to see all the variables in the data;

*proc contents data=GWI_PartC2 order=varnum;

%INCLUDE syntax('CSP585_zz_formats.sas');

run;

```

Data GWI_table_setup;

set GWI_PartC2;

** Define VHAstatus based on Q39 Percentage healthcare at VA in last year and

Q40a Overnight in Va facility;

** Q39 ;

if VAUse = 0 then VAUse_use = 0; * Non-

user;

else if VAUse in (1 2 3 4 5) then VAUse_use = 1; * User;

else

VAUse_use = 2; * Not answered;

** Q40a ;

if VAHosp = 0 then VAHosp_use = 0; * Non-

user;

else if VAHosp in (1 2 3 4) then VAHosp_use = 1; * User;

else

VAHosp_use = 2; * Not answered;

** STATUS Q39 OR Q40a ;

If (VAUse_use = 0 AND VAHosp_use = 0) then VHAstatus = 0; *

Non-user;

```

else If (VAUse_use = 1 OR VAHosp_use = 1)      then VHAstatus = 1; *
User;

else VHAstatus = .; * Not answered;

format VHAstatus 1_yesno.;

label VHAstatus = "Used VA health care or hospital in the last year";

label KS_GWI = "Meets the full Kansas GWI Definition";

*identify conditions;

%Array(conditions, values = CircHTN      CircStrk CircTIA      CircHrtAtk
CircCAD      CircPVD      CircChol
      CircClot      CircCHF      CircOth      SkMsOA      SkMsRA
SkMsOthArth SkMsGout      SkMsOP      SkMsOth      MHAnxPan
      MHADHD      MHBPD      MHPTSD      MHDep      MHEatDo
MHPersDo      MHSCZ      MHSocPh      MHOth      HVCat HVGlauc
HVMD
      HVBlind      HVTin HVHear      IDTB      IDHepC      IDHIV
KDNoDial      KDDial      KDAcute      GIReflux      GIUlcer
      GIObstGIPolyp      GIIBS      GIUC      GICrohn      GICeliac
CaBrain      CaBrst CaColon      CaLung      CaProsCaSkin

```

```

NSMigrn    NSOthHd    NSMemLoss  NSDem    NSLOC
NSTBI NSSCI NSEpilNSPD NSALSNSMS NSOth
DoAsth     DoLung    DoDM DoBPH    DoLiver
DoApnea    DoLupus   CaOth GIOth DoThy IDOth DoSkin);
array Array_conditions (&conditionsn) %DO_OVER(conditions, phrase=?);
array Array_conditions_yrdx (&conditionsn) %DO_OVER(conditions,
phrase=?YrDx);
array Array_conditions_Rev (&conditionsn) %DO_OVER(conditions,
phrase=?_Rev);    *these are what we want to keep;
do i=1 to &conditionsn;
    Array_conditions_Rev[i]=Array_conditions[i];    *revised condition is
the original unless a dx year is recorded;
    if Array_conditions_yrdx[i] not in (. .a .b .c .d .e .f .g .h) then
Array_conditions_Rev[i]=1;
    end;
label sex="Sex";
format Education l_educatation.;
if NoDplyGulf = 1 AND DplyGulf = 0 then deploy = 0; * Did not deploy;
else if NoDplyGulf = 0 AND DplyGulf = 1 then deploy = 1; * Deployed to Gulf;
else if DplyElsewhere = 1    then deploy = 2; * Deployed elsewhere ;

```

```

else deploy = .; * unknown;

label deploy = "Deployed to the Gulf the 1990-1991 Gulf War";

format deploy l_deploy.;

if income IN (0, 1, 2)           then income_new = 1; *Under 30,000;
    else if income IN (3, 4, 5) then income_new = 2; *$30,000 - $59,999;
    else if income IN (6, 7)    then income_new = 3; *$60,000 - $99,999;
    else if income IN (8, 9)    then income_new = 4; *$100,000 or more;
                                else income_new = 5;

*Unknown;

label income_new="Household income per year";

format income_new l_income.;

if educat IN (0, 1)           then Education = 1; *Less than Bachelor's
degree;
    else if educat IN (2, 3, 4) then Education = 2; *Bachelor's degree,
associates deg;
    else if educat IN (5, 6) then Education = 3; *Master's, Professional, or
Doctorate's degree;
                                else Education =
4;*Unknown;

```

```

label Education = "Highest achieved education level";

format Education l_education.;

if MltySrvTypeAD = 1 AND MltySrvTypeRO = 0                                then
MltyComp = 0; *Active duty only;

    else if MltySrvTypeAD = 1 AND MltySrvTypeRO = 1                        then
MltyComp = 1; *both active duty and reserves;

    else if MltySrvTypeAD = 0 AND MltySrvTypeRO = 1                        then
MltyComp = 2; *Reserves only;

    else MltyComp = .;

label MltyComp = "Military Component";

format MltyComp l_mltycomp.;

*set up raceEth;

if RaceW = 1 and RaceAA = 0 and RaceAI = 0 and RaceAsCh = 0 and RaceAsJp = 0
    and RaceAsIn = 0 and RaceAsOt = 0 and RaceAsFp = 0 and RacePI
= 0 and RaceOt=0

    and ethnic1 = 1

        then raceEth = 0; *raceEth is white, nonhispanic;

    else if RaceAA=1 and raceW=0 and RaceAI = 0 and RaceAsCh = 0 and
RaceAsJp = 0

```

and RaceAsIn = 0 and RaceAsOt = 0 and RaceAsFp = 0 and RacePI
= 0 and RaceOt=0

and ethnic1 = 1

then raceEth=1; *raceEth is black, nonhispanic;

else if ethnic1=0

then raceEth=2; *raceEth is hispanic;

else if raceW in (. .a .b .c .d .e .f .g .h)

then raceEth = 4; *raceEth is missing/unknown;

else if ethnic1 in (. .a .b .c .d .e .f .g .h)

then raceEth = 4; *raceEth is missing/unknown;

else raceEth = 3; *raceEth is other;

label raceEth = "Race/ Ethnicity";

format raceEth l_raceEth.;

SvcBranch = 6; *other;

if mltyBrArmy = 1 then SvcBranch = 1; *Army only;

if mltyBrNavy = 1 then SvcBranch = 2; *Navy only;

if mltyBrAF = 1 then SvcBranch = 3; *Air Force only;

if mltyBrMC = 1 then SvcBranch = 4; *Marines only;

if mltyBrNG=1 then svcBranch = 5; *National Guard, any;

```

if sum(mltyBrArmy, mltyBrNavy, mltyBrAF, mltyBrMC)>1 then SvcBranch = 6;

*other;

format SvcBranch 1_svcBr.;

*Some combined variables;

Cancer_Rev = 0;      *This 0 could represent no or missing;

if CaBrain_rev=1 then Cancer_Rev=1;

if CaBrst_rev = 1 then Cancer_Rev=1;

if CaColon_rev = 1 then Cancer_Rev=1;

if CaLung_rev = 1 then Cancer_Rev=1;

if CaPros_rev =1 then Cancer_Rev=1;

if CaOth_rev =1 then Cancer_Rev=1;

HeartDis_Rev = 0; *This 0 could represent no or missing;

if CircHrtAtk_rev=1 then HeartDis_Rev=1;

if CircCAD_rev =1 then HeartDis_Rev=1;

if CircCHF_rev =1 then HeartDis_Rev=1;

ID_Rev=0; *This 0 could represent no or missing;

if IDHIV_rev=1 then ID_Rev=1;

if IDTB_rev=1 then ID_Rev=1;

if IDHepC_rev=1 then ID_Rev=1;

Stroke_Rev=0;*This 0 could represent no or missing;

```

if CircStrk_rev=1 then Stroke_Rev=1;

if CircTIA_rev=1 then Stroke_Rev=1;

MH_Rev=0; *This 0 could represent no or missing;

if MHSCZ_Rev=1 then MH_Rev=1;

if MHBPD_rev=1 then MH_Rev=1;

NS_Rev=0; *This 0 could represent no or missing;

if NSTBI_rev=1 then NS_Rev=1;

if NSMS_rev=1 then NS_Rev=1;

SMoDRemeConct_rev = .; *this combined variable is used in the original CDC

and is created for comparison;

if SMoDReme_rev=1 | SMoDConct=1 then SMoDRemeConct_rev = 1;

if SMoDReme_rev=0 & SMoDConct=0 then SMoDRemeConct_rev = 0;

format KS_Fatigue_ModMult KS_Pain_ModMult KS_GI_ModMult

KS_Mood_ModMult

KS_Resp_ModMult KS_Skin_ModMult CDC_Fatigue_Any CDC_Mood_Any

CDC_MuscSkel_Any

CDC_Fatigue_Sev CDC_Mood_Sev CDC_MuscSkel_Sev CDC_GWI

CDC_GWI_Sev KS_GWI

KS_Excl KS_GWI_ModMult_SC l_yesno.;

```

numConditions = sum(CircHTN_Rev,      CircStrk_Rev, CircTIA_Rev,
CircHrtAtk_Rev,    CircCAD_Rev,      CircPVD_Rev, CircChol_Rev,
CircClot_Rev,

CircCHF_Rev, CircOth_Rev, SkMsOA_Rev, SkMsRA_Rev, SkMsOthArth_Rev,
SkMsGout_Rev,      SkMsOP_Rev, SkMsOth_Rev,
MHAnxPan_Rev,     MHADHD_Rev,      MHBPD_Rev, MHPTSD_Rev,
MHDep_Rev, MHEatDo_Rev,      MHPersDo_Rev,      MHSCZ_Rev,
MHSocPh_Rev,      MHOth_Rev, HVCat_Rev,  HVGlauC_Rev,
HVMD_Rev,  HVBlind_Rev, HVTin_Rev,  HVHear_Rev, IDTB_Rev,
IDHepC_Rev, IDHIV_Rev,  KDNODial_Rev,      KDDial_Rev,
KDAcute_Rev,      GIReflux_Rev, GIUlcer_Rev,  GIObst_Rev,
GIPolyp_Rev, GIIBS_Rev,  GIUC_Rev,   GICrohn_Rev, GICeliac_Rev,
CaBrain_Rev, CaBrst_Rev,  CaColon_Rev, CaLung_Rev,
CaPros_Rev,  CaSkin_Rev,  NSMigrn_Rev,      NSOthHd_Rev,
NSMemLoss_Rev,  NSDem_Rev, NSLOC_Rev, NSTBI_Rev,  NSSCI_Rev,
NSEpil_Rev,  NSPD_Rev,  NSALS_Rev, NSMS_Rev,  NSOth_Rev,
DoAsth_Rev, DoLung_Rev, DoDM_Rev,  DoBPH_Rev,
DoLiver_Rev, DoApnea_Rev,      DoLupus_Rev,      CaOth_Rev,
GIOth_Rev,  DoThy_Rev,  IDOth_Rev,  DoSkin_Rev);

inData = 1 ; *This is for table formatting;

```

```

        label inData = "Table formatting only. 1 for all";

run;

*bring in the ages from getages table pulled from SQL in file 01;

PROC SQL;

    CREATE TABLE GWI_demographics AS

    SELECT *

    FROM GWI_table_setup , getages

    WHERE upcase(GWI_table_setup.PIDSurvey) = upcase(getages.PIDSurvey);

QUIT;

Data GWI_demographics;

set GWI_demographics;

    ageGroup=0; *no ages were missing and further checks indicate no ageGroup=0;

    if ageAtSurvey>40 then ageGroup=1;

    if ageAtSurvey>50 then ageGroup=2;

    if ageAtSurvey>60 then ageGroup=3;

run;

```

Works Cited

1. Research Advisory Committee. *Gulf War Illness and the Health of Gulf War Veterans.*; 2008. https://www.va.gov/RAC-GWVI/docs/Committee_Documents/GWlandHealthofGWVeterans_RAC-GWVIReport_2008.pdf.
2. Research Advisory Committee. *Updated Scientific Findings and Recommendations Research Advisory Committee on Gulf War Veterans' Illnesses.*; 2014. <https://www.va.gov/RAC-GWVI/RACReport2014Final.pdf>. Accessed October 26, 2021.
3. Institute of Medicine. *Chronic Multisymptom Illness in Gulf War Veterans.* Washington, D.C.: National Academies Press; 2014. doi:10.17226/18623
4. Institute of Medicine, National Academies of Sciences Engineering and Medicine. *Gulf War and Health: Volume 10: Update of Health Effects of Serving in the Gulf War.* Vol 10. (Cory-Slechta D, Wedge R, eds.). Washington, D.C.: National Academies Press; 2016. doi:10.17226/21840
5. 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.
6. 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. doi:10.1093/aje/152.10.992
7. Porter B, Long K, Rull RP, Dursa EK. Prevalence of Chronic Multisymptom Illness/Gulf War Illness over Time among Millennium Cohort Participants, 2001 to 2016. *J Occup Environ Med.* 2020;62(1):4-10. doi:10.1097/JOM.0000000000001716
8. Coughlin SS, McNeil RB, Provenzale DT, Dursa EK, Thomas CM. Method Issues in Epidemiological Studies of Medically Unexplained Symptom-based Conditions in Veterans. *J Mil Veterans Health.* 2013;21(2):4. /pmc/articles/PMC3968909/. Accessed February 2, 2022.
9. Zundel CG, Kregel MH, Heeren T, et al. Rates of Chronic Medical Conditions in 1991 Gulf War Veterans Compared to the General Population. *Int J Environ Res Public Health.* 2019;16(6):949. doi:10.3390/ijerph16060949
10. Kang HK, Natelson BH, Mahan CM, Lee KY, Murphy FM. Post-traumatic stress disorder and chronic fatigue syndrome-like illness among gulf war veterans: A

- population-based survey of 30,000 veterans. *Am J Epidemiol*. 2003.
doi:10.1093/aje/kwf187
11. Blanchard MS, Eisen SA, Alpern R, et al. Chronic Multisymptom Illness Complex in Gulf War I Veterans 10 Years Later. *Am J Epidemiol*. 2006;163(1):66-75.
doi:10.1093/aje/kwj008
 12. Gwini SM, Forbes AB, Sim MR, Kelsall HL. Multisymptom illness in gulf war veterans a systematic review and meta-analysis. *J Occup Environ Med*. 2016;58(7):659-667. doi:10.1097/JOM.0000000000000755
 13. Institute of Medicine. *Gulf War and Health: Volume 8: Update of Health Effects of Serving in the Gulf War*. Washington, D.C.: National Academies Press; 2010.
doi:10.17226/12835
 14. Fulco CE, Liverman CT, Sox HC. *Gulf War and Health: Volume 1. Depleted Uranium, Sarin, Pyridostigmine Bromide, Vaccines*. Washington, D.C.: National Academies Press; 2000. doi:10.17226/9953
 15. Gean EG, Ayers CK, Winchell KA, et al. Biological measures and diagnostic tools for Gulf War Illness – A systematic review. *Life Sci*. 2021;275:119360.
doi:10.1016/j.lfs.2021.119360
 16. Jeffrey MG, Kregel M, Kibler JL, et al. Neuropsychological Findings in Gulf War Illness: A Review. *Front Psychol*. 2019;10(SEP):2088. doi:10.3389/fpsyg.2019.02088
 17. Nettleman M. Gulf War Illness: Challenges Persist. *Trans Am Clin Climatol Assoc*. 2015;126:237-247. <https://www.ncbi.nlm.nih.gov/pubmed/26330683>.
 18. Ribeiro ACR, Deshpande LS. A review of pre-clinical models for Gulf War Illness. *Pharmacol Ther*. 2021;228:107936. doi:10.1016/j.pharmthera.2021.107936
 19. Maule AL, Janulewicz PA, Sullivan KA, et al. Meta-analysis of self-reported health symptoms in 1990–1991 Gulf War and Gulf War-era veterans. *BMJ Open*. 2018;8(2). doi:10.1136/BMJOPEN-2017-016086
 20. Gifford EJ, Vahey J, Hauser ER, et al. Gulf War illness in the Gulf War Era Cohort and Biorepository: The Kansas and Centers for Disease Control definitions. *Life Sci*. 2021;278:119454. doi:10.1016/j.lfs.2021.119454
 21. White RF, Steele L, O’Callaghan JP, et al. Recent research on Gulf War illness and other health problems in veterans of the 1991 Gulf War: Effects of toxicant

- exposures during deployment. *Cortex*. 2016;74:449-475.
doi:10.1016/j.cortex.2015.08.022
22. 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 Heal*. 2015;14(1):4. doi:10.1186/1476-069X-14-4
 23. 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. doi:10.1289/ehp.1003399
 24. Gauderman J, Ph D, Morrison J. QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies,. <http://hydra.usc.edu/gxe>. 2006. doi:10.1007/88-470-0474-8_3
 25. Georgopoulos A, James L, Carpenter A, Engdahl B, Leuthold A, Lewis S. Gulf War illness (GWI) as a neuroimmune disease. *j*. 2017;235(10). doi:10.1007/s00221-017-5050-0
 26. Miller MW, Sperbeck E, Robinson ME, et al. 5-HT2A Gene Variants Moderate the Association between PTSD and Reduced Default Mode Network Connectivity. *Front Neurosci*. 2016;10(JUN):299. doi:10.3389/fnins.2016.00299
 27. Golomb BA. Acetylcholinesterase inhibitors and Gulf War illnesses. *Proc Natl Acad Sci*. 2008;105(11):4295-4300. doi:10.1073/pnas.0711986105
 28. DeBeer BB, Davidson D, Meyer EC, Kimbrel NA, Gulliver SB, Morissette SB. The Association Between Toxic Exposures and Chronic Multisymptom Illness in Veterans of the Wars of Iraq and Afghanistan. *J Occup Environ Med*. 2017;59(1):54-60. doi:10.1097/JOM.0000000000000922
 29. AC P. Gene-environmental interactions and organophosphate toxicity. *Toxicology*. 2010;278(3):294-304. doi:10.1016/J.TOX.2010.02.007
 30. Lee BW, London L, Paulauskis J, Myers J, Christiani DC. Association Between Human Paraoxonase Gene Polymorphism and Chronic Symptoms in Pesticide-Exposed Workers. *J Occup Environ Med*. 2003;45(2):118-122. doi:10.1097/01.jom.0000052953.59271.e1
 31. Dodd-Butera T, Broderick M. Pyridostigmine. *Encycl Toxicol Third Ed*. January 2014:1162-1164. doi:10.1016/B978-0-12-386454-3.00053-1

32. Naughton SX, Terry A V. Neurotoxicity in acute and repeated organophosphate exposure. *Toxicology*. 2018;408:101-112. doi:10.1016/j.tox.2018.08.011
33. Pereira CD, Martins F, Marques F, Sousa JC, Rebelo S. Beyond Brain Signaling. *Tissue-Specific Cell Signal*. 2020:1-32. doi:10.1007/978-3-030-44436-5_1
34. Storm JE, Rozman KK, Doull J. Occupational exposure limits for 30 organophosphate pesticides based on inhibition of red blood cell acetylcholinesterase. *Toxicology*. 2000;150(1-3):1-29. doi:10.1016/S0300-483X(00)00219-5
35. Mileson BE, Chambers JE, Chen WL, et al. Common mechanism of toxicity: A case study of organophosphorus pesticides. *Toxicol Sci*. 1998;41(1):8-20. doi:10.1006/toxs.1997.2431
36. Michalovicz LT, Kelly KA, Sullivan K, O'Callaghan JP. Acetylcholinesterase inhibitor exposures as an initiating factor in the development of Gulf War Illness, a chronic neuroimmune disorder in deployed veterans. *Neuropharmacology*. 2020;171:108073. doi:10.1016/j.neuropharm.2020.108073
37. Strong LL, Thompson B, Coronado GD, Griffith WC, Vigoren EM, Islas I. Health symptoms and exposure to organophosphate pesticides in farmworkers. *Am J Ind Med*. 2004;46(6):599-606. doi:10.1002/ajim.20095
38. Abdullah L, Evans JE, Montague H, et al. Chronic elevation of phosphocholine containing lipids in mice exposed to Gulf War agents pyridostigmine bromide and permethrin. *Neurotoxicol Teratol*. 2013;40:74-84. doi:10.1016/j.ntt.2013.10.002
39. Miller J V., Lebouf RF, Kelly KA, et al. The Neuroinflammatory Phenotype in a Mouse Model of Gulf War Illness is Unrelated to Brain Regional Levels of Acetylcholine as Measured by Quantitative HILIC-UPLC-MS/MS. *Toxicol Sci*. 2018;165(2):302-313. doi:10.1093/toxsci/kfy130
40. Ashbrook DG, Hing B, Michalovicz LT, et al. Epigenetic impacts of stress priming of the neuroinflammatory response to sarin surrogate in mice: a model of Gulf War illness. *J Neuroinflammation*. 2018;15(1):86. doi:10.1186/s12974-018-1113-9
41. Carpenter JM, Gordon HE, Ludwig HD, et al. Neurochemical and neuroinflammatory perturbations in two Gulf War Illness models: Modulation by the immunotherapeutic LNFPIII. *Neurotoxicology*. 2020;77:40-50. doi:10.1016/j.neuro.2019.12.012

42. 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*. 2017;12(4). doi:10.1371/journal.pone.0176634
43. Lockridge OM, Quinn DM. Comprehensive Toxicology (Second Edition). In: McQueen CA, ed. *Comprehensive Toxicology*. Volume 10. Elsevier Ltd; 2018:277-307.
44. Iannacchione VG, Dever JA, Bann CM, et al. Validation of a Research Case Definition of Gulf War Illness in the 1991 US Military Population. *Neuroepidemiology*. 2011;37(2):129-140. doi:10.1159/000331478
45. Haley RW, Kurt TL. Self-reported Exposure to Neurotoxic Chemical Combinations in the Gulf War: A Cross-sectional Epidemiologic Study. *JAMA J Am Med Assoc*. 1997;277(3):231. doi:10.1001/jama.1997.03540270057027
46. NISENBAUM R, BARRETT DH, REYES M, REEVES WC. Deployment Stressors and a Chronic Multisymptom Illness among Gulf War Veterans. *J Nerv Ment Dis*. 2000;188(5):259-266. doi:10.1097/00005053-200005000-00002
47. Gray GC, Reed RJ, Kaiser KS, Smith TC, Gastañaga VM. Self-reported Symptoms and Medical Conditions among 11,868 Gulf War-era Veterans The Seabee Health Study. *Am J Epidemiol*. 2002;155(11):1033-1044. doi:10.1093/AJE/155.11.1033
48. Cherry N, Creed F, Silman A, et al. Health and exposures of United Kingdom Gulf war veterans. Part I: The pattern and extent of ill health. *Occup Environ Med*. 2001;58(5):291-298. doi:10.1136/OEM.58.5.291
49. Cherry N, Creed F, Silman A, et al. Health and exposures of United Kingdom Gulf war veterans. Part II: The relation of health to exposure. *Occup Environ Med*. 2001;58(5):299-306. doi:10.1136/OEM.58.5.299
50. Wolfe J, Proctor SP, Erickson DJ, Hu H. Risk factors for multisymptom illness in US Army veterans of the Gulf War. *J Occup Environ Med*. 2002;44(3):271-281. doi:10.1097/00043764-200203000-00015
51. Spencer PS, McCauley LA, Lapidus JA, Lasarev M, Joos SK, Storzbach D. Self-reported exposures and their association with unexplained illness in a population-based case-control study of Gulf War veterans. *J Occup Environ Med*. 2001;43(12):1041-1056. doi:10.1097/00043764-200112000-00006
52. Gao J, Xu F, Starlard-Davenport A, et al. Exploring the Role of Chemokine

- Receptor 6 (Ccr6) in the BXD Mouse Model of Gulf War Illness. *Front Neurosci.* 2020;14:818. doi:10.3389/fnins.2020.00818
53. O'Callaghan JP, Michalovicz LT, Kelly KA. Supporting a Neuroimmune Basis of Gulf War Illness. *EBioMedicine.* 2016;13:5-6. doi:10.1016/J.EBIOM.2016.10.037
 54. Haley RW, Billecke S, La Du BN. Association of Low PON1 Type Q (Type A) Arylesterase Activity with Neurologic Symptom Complexes in Gulf War Veterans. *Toxicol Appl Pharmacol.* 1999;157(3):227-233. doi:10.1006/TAAP.1999.8703
 55. 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. doi:10.1016/j.ebiom.2017.11.005
 56. 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. doi:10.1016/j.ebiom.2015.11.037
 57. 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. doi:10.1016/j.ebiom.2016.10.019
 58. Slowik A, Tomik B, Wolkow PP, et al. Paraoxonase gene polymorphisms and sporadic ALS. *Neurology.* 2006;67(5):766-770. doi:10.1212/01.wnl.0000219565.32247.11
 59. Khalil L, McNeil RB, Sims KJ, et al. The Gulf War Era Cohort and Biorepository: A Longitudinal Research Resource of Veterans of the 1990–1991 Gulf War Era. *Am J Epidemiol.* 2018;187(11):2279-2291. doi:10.1093/aje/kwy147
 60. Radhakrishnan K, Hauser ER, Polimanti R, et al. Genomics of Gulf War Illness in U.S. Veterans Who Served during the 1990–1991 Persian Gulf War: Methods and Rationale for Veterans Affairs Cooperative Study #2006. *Brain Sci.* 2021;11(7):845. doi:10.3390/brainsci11070845
 61. Gaziano JM, Concato J, Brophy M, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol.* 2016;70:214-223. doi:10.1016/j.jclinepi.2015.09.016
 62. Ottman R, Rao DC. An epidemiologic approach to gene-environment interaction. *Genet Epidemiol.* 1990;7(3):177-185. doi:10.1002/GEPI.1370070302

63. Dempfle A, Scherag A, Hein R, Beckmann L, Chang-Claude J, Schäfer H. *Gene-Environment Interactions for Complex Traits: Definitions, Methodological Requirements and Challenges*. Vol 16.; 2008:1164-1172. doi:10.1038/ejhg.2008.106
64. Dick DM, Agrawal A, Keller MC, et al. Candidate Gene–Environment Interaction Research: Reflections and Recommendations. *Perspect Psychol Sci*. 2015;10(1):37-59. doi:10.1177/1745691614556682
65. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606-613. doi:10.1046/j.1525-1497.2001.016009606.x
66. Wagner D, Nisenbaum R, Heim C, Jones JF, Unger ER, Reeves WC. Psychometric properties of the CDC Symptom Inventory for assessment of Chronic Fatigue Syndrome. *Popul Health Metr*. 2005;3. doi:10.1186/1478-7954-3-8
67. Blanchard EB, Jones-Alexander J, Buckley TC, Forneris CA. Psychometric properties of the PTSD checklist (PCL). *Behav Res Ther*. 1996;34(8):669-673. doi:10.1016/0005-7967(96)00033-2
68. Wiklund IK, Fullerton S, Hawkey CJ, et al. An irritable bowel syndrome-specific symptom questionnaire: Development and validation. *Scand J Gastroenterol*. 2003;38(9):947-954. doi:10.1080/00365520310004209
69. Dursa E, Barth S, Porter B, Schneiderman A. Gulf War Illness in the 1991 Gulf War and Gulf Era Veteran Population: An Application of the Centers for Disease Control and Prevention and Kansas Case Definitions to Historical Data. *J Mil Veterans Heal*. 2018;26(2). <https://jmvh.org/article/gulf-war-illness-in-the-1991-gulf-war-and-gulf-era-veteran-population-an-application-of-the-centers-for-disease-control-and-prevention-and-kansas-case-definitions-to-historical-data/>. Accessed December 21, 2020.
70. Wilkinson MD, Dumontier M, Aalbersberg IJJ, et al. Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3. doi:10.1038/sdata.2016.18
71. Vandembroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE). *Epidemiology*. 2007;18(6):805-835. doi:10.1097/EDE.0b013e3181577511
72. Helmer DA, Provenzale DT. *Personal Communication*.

73. Bjarnason E, Borg M. Aligning Requirements and Testing: Working Together toward the Same Goal. *IEEE Softw.* 2017;34(1):20-23. doi:10.1109/MS.2017.14
74. Scanlan DA. Structured Flowcharts Outperform Pseudocode: An Experimental Comparison. *IEEE Softw.* 1989;6(5):28-36. doi:10.1109/52.35587
75. Evangelou CE, Karacapilidis N. A multidisciplinary approach for supporting knowledge-based decision making in collaborative settings. *Int J Artif Intell Tools.* 2007;16(6):1069-1092. doi:10.1142/S0218213007003631
76. Goldman M, Miller R. 6.005: Software Construction. <http://web.mit.edu/6.005/www/fa15/>. Accessed October 3, 2020.
77. Hunt A, Thomas D. *The Pragmatic Programmer: From Journeyman to Master.*; 1999. <http://portal.acm.org/citation.cfm?id=320326>. Accessed April 2, 2015.
78. Clay T, Katz D. Tight Looping With Macro Arrays. :1-8. <https://support.sas.com/resources/papers/proceedings/proceedings/sugi31/040-31.pdf>. Accessed January 21, 2021.
79. Knight ^1 KE, Honerlaw J, Danciu I, et al. *Standardized Architecture for a Mega-Biobank Phenomic Library: The Million Veteran Program (MVP)*. Vol 2020. American Medical Informatics Association; 2020. </pmc/articles/PMC7233040/?report=abstract>. Accessed January 6, 2021.
80. Vahey J, Hauser ER, Sims KJ, Helmer DA, Provenzale D, Gifford EJ. Research tool for classifying Gulf War illness using survey responses: Lessons for writing replicable algorithms for symptom-based conditions. *Life Sci.* 2021;282:119808. doi:10.1016/J.LFS.2021.119808
81. Grewe ME, Khalil L, Felder K, et al. Gulf War Era Veterans' perspectives on research: a qualitative study. *Life Sci.* 2021;287:120113. doi:10.1016/j.lfs.2021.120113
82. Haley RW. Point: bias from the "healthy-warrior effect" and unequal follow-up in three government studies of health effects of the Gulf War. *Am J Epidemiol.* 1998;148(4):315-323. doi:10.1093/oxfordjournals.aje.a009645
83. Kipen HM, Hallman W, Fiedler N, et al. Prevalence of Chronic Fatigue and Chemical Sensitivities in Gulf Registry Veterans. *Arch Environ Health.* 1999;54(5):313-318. doi:10.1080/00039899909602493
84. Powell TM, Smith TC, Jacobson IG, et al. Prospective Assessment of Chronic

- Multisymptom Illness Reporting Possibly Associated with Open-Air Burn Pit Smoke Exposure in Iraq. *J Occup Environ Med.* 2012;54(6):682-688. doi:10.1097/JOM.0b013e318255ba39
85. Visscher PM, Wray NR, Zhang Q, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet.* 2017;101(1):5-22. doi:10.1016/j.ajhg.2017.06.005
86. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet.* 2012. doi:10.1016/j.ajhg.2011.11.029
87. Docherty AR, Shabalín AA, DiBlasi E, et al. Genome-wide association study of suicide death and polygenic prediction of clinical antecedents. *Am J Psychiatry.* 2020;177(10):917. doi:10.1176/APPI.AJP.2020.19101025
88. Mullins N, Kang JE, Campos AI, et al. Dissecting the Shared Genetic Architecture of Suicide Attempt, Psychiatric Disorders, and Known Risk Factors. *Biol Psychiatry.* 2022;91(3):313-327. doi:10.1016/J.BIOPSYCH.2021.05.029
89. Stein MB, Levey DF, Cheng Z, et al. Genome-wide association analyses of post-traumatic stress disorder and its symptom subdomains in the Million Veteran Program. *Nat Genet.* 2021;53(2):174-184. doi:10.1038/s41588-020-00767-x
90. Schlauch KA, Khaiboullina SF, De Meirleir KL, et al. Genome-wide association analysis identifies genetic variations in subjects with myalgic encephalomyelitis/chronic fatigue syndrome. *Transl Psychiatry.* 2016;6(2):e730-e730. doi:10.1038/tp.2015.208
91. Klein RJ, Zeiss C, Chew EY, et al. Complement Factor H Polymorphism in Age-Related Macular Degeneration. *Science (80-).* 2005;308(5720):385-389. doi:10.1126/science.1109557
92. Andrews SJ, Fulton-Howard B, Goate A. Interpretation of risk loci from genome-wide association studies of Alzheimer's disease. *Lancet Neurol.* 2020;19(4):326-335. doi:10.1016/S1474-4422(19)30435-1
93. Begum F, Ghosh D, Tseng GC, Feingold E. Comprehensive literature review and statistical considerations for GWAS meta-analysis. *Nucleic Acids Res.* 2012;40(9):3777-3784. doi:10.1093/NAR/GKR1255
94. 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. doi:10.1006/bbrc.2000.3526

95. Mackness M, Mackness B. Lack of Association of Paraoxonase 1 Promoter Polymorphisms with Gulf War Illness. *Toxicol Open Access*. 2015;01(01):1-4. doi:10.4172/2476-2067.1000105
96. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput Biol*. 2015;11(4). doi:10.1371/JOURNAL.PCBI.1004219
97. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8(1):1826. doi:10.1038/s41467-017-01261-5
98. Bush WS, Oetjens MT, Crawford DC. Unravelling the human genome-phenome relationship using phenome-wide association studies. *Nat Rev Genet*. 2016;17(3):129-145. doi:10.1038/nrg.2015.36
99. Pendergrass SA, Brown-Gentry K, Dudek S, et al. Phenome-Wide Association Study (PheWAS) for Detection of Pleiotropy within the Population Architecture using Genomics and Epidemiology (PAGE) Network. *PLoS Genet*. 2013. doi:10.1371/journal.pgen.1003087
100. Moore R, Casale FP, Jan Bonder M, et al. A linear mixed-model approach to study multivariate gene–environment interactions. *Nat Genet*. 2019;51(1):180-186. doi:10.1038/s41588-018-0271-0
101. Runcie DE, Crawford L. Fast and flexible linear mixed models for genome-wide genetics. *PLoS Genet*. 2019;15(2):e1007978. doi:10.1371/journal.pgen.1007978
102. Vahey J, Gifford EJ, Sims KJ, et al. Gene-Toxicant Interactions in Gulf War Illness: Differential Effects of the PON1 Genotype. *Brain Sci* 2021, Vol 11, Page 1558. 2021;11(12):1558. doi:10.3390/BRAINSCI11121558
103. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4(1):7. doi:10.1186/s13742-015-0047-8
104. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26(18):2336-2337. doi:10.1093/bioinformatics/btq419

105. Lin SH, Thakur R, Machiela MJ. LDexpress: an online tool for integrating population-specific linkage disequilibrium patterns with tissue-specific expression data. *BMC Bioinformatics*. 2021;22(1):1-6. doi:10.1186/S12859-021-04531-8/FIGURES/1
106. Lin SH, Brown DW, Machiela MJ. LDtrait: An online tool for identifying published phenotype associations in linkage disequilibrium. *Cancer Res*. 2020;80(16):3443-3446. doi:10.1158/0008-5472.CAN-20-0985/654200/AM/LDTRAIT-AN-ONLINE-TOOL-FOR-IDENTIFYING-PUBLISHED
107. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019;47(D1):D1005-D1012. doi:10.1093/nar/gky1120
108. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001;29(1):308. doi:10.1093/NAR/29.1.308
109. Sherry ST, Ward M, Sirotkin K. dbSNP—Database for Single Nucleotide Polymorphisms and Other Classes of Minor Genetic Variation. *Genome Res*. 1999;9(8):677-679. doi:10.1101/gr.9.8.677
110. Gilmour AR, Thompson R, Cullis BR. Average Information REML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models. *Biometrics*. 1995;51(4):1440. doi:10.2307/2533274
111. Zhang Q, Zhou X Di, 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. doi:10.1128/CDLI.6.1.6-13.1999/ASSET/39753E5C-802A-4EC1-85E0-F9C7268F1ACF/ASSETS/GRAPHIC/CD0190086002.JPEG
112. Coller JK, Tuke J, Wain TJ, et al. Associations of Immune Genetic Variability with Gulf War Illness in 1990–1991 Gulf War Veterans from the Gulf War Illness Consortium (GWIC) Multisite Case-Control Study. *Brain Sci* 2021, Vol 11, Page 1410. 2021;11(11):1410. doi:10.3390/BRAINSCI11111410
113. 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. doi:10.1016/J.BBI.2010.10.011
114. Lees JR. Interferon gamma in autoimmunity: A complicated player on a complex stage. *Cytokine*. 2015;74(1):18-26. doi:10.1016/j.cyto.2014.10.014

115. Li P, Du Q, Cao Z, et al. Interferon-gamma induces autophagy with growth inhibition and cell death in human hepatocellular carcinoma (HCC) cells through interferon-regulatory factor-1 (IRF-1). *Cancer Lett.* 2012;314(2):213-222. doi:10.1016/J.CANLET.2011.09.031
116. Hartung H -P, Schäfer B, Van Der Meide PH, Fierz W, Heininger K, Toyka K V. The role of interferon-gamma in the pathogenesis of experimental autoimmune disease of the peripheral nervous system. *Ann Neurol.* 1990;27(3):247-257. doi:10.1002/ana.410270306
117. Campbell IL, Kay TWH, Oxbrow L, Harrison LC. Essential role for interferon-gamma and interleukin-6 in autoimmune insulin-dependent diabetes in NOD/Wehi mice. *J Clin Invest.* 1991;87(2):739-742. doi:10.1172/JCI115055
118. Ottum PA, Arellano G, Reyes LI, Iruetagoiena M, Naves R. Opposing roles of interferon-gamma on cells of the central nervous system in autoimmune neuroinflammation. *Front Immunol.* 2015;6(OCT):539. doi:10.3389/fimmu.2015.00539
119. Arellano G, Ottum PA, Reyes LI, Burgos PI, Naves R. Stage-specific role of interferon-gamma in experimental autoimmune encephalomyelitis and multiple sclerosis. *Front Immunol.* 2015;6(SEP):492. doi:10.3389/FIMMU.2015.00492/BIBTEX
120. Bae HR, Leung PSC, Tsuneyama K, et al. Chronic expression of interferon-gamma leads to murine autoimmune cholangitis with a female predominance. *Hepatology.* 2016;64(4):1189-1201. doi:10.1002/HEP.28641/SUPPINFO
121. Wong YC, Holzbaur ELF. Autophagosome dynamics in neurodegeneration at a glance. *J Cell Sci.* 2015;128(7):1259-1267. doi:10.1242/jcs.161216
122. Heng HHQ. Challenges and new strategies for Gulf War illness research. *Environ Dis.* 2016;1(4):118. doi:10.4103/2468-5690.198618
123. Nixon RA, Yang DS. Autophagy failure in Alzheimer's disease-locating the primary defect. *Neurobiol Dis.* 2011;43(1):38-45. doi:10.1016/j.nbd.2011.01.021
124. Ramesh N, Pandey UB. Autophagy dysregulation in ALS: When protein aggregates get out of hand. *Front Mol Neurosci.* 2017;10:263. doi:10.3389/fnmol.2017.00263
125. Yee MK, Janulewicz PA, Seichepine DR, Sullivan KA, Proctor SP, Kregel MH. Multiple mild traumatic brain injuries are associated with increased rates of

- health symptoms and Gulf War illness in a Cohort of 1990-1991 Gulf War Veterans. *Brain Sci.* 2017;7(7):79. doi:10.3390/brainsci7070079
126. Mouzon BC, Bachmeier C, Ojo JO, et al. Lifelong behavioral and neuropathological consequences of repetitive mild traumatic brain injury. *Ann Clin Transl Neurol.* 2018;5(1):64-80. doi:10.1002/acn3.510
 127. Janulewicz P, Kregel M, Quinn E, et al. The Multiple Hit Hypothesis for Gulf War Illness: Self-Reported Chemical/Biological Weapons Exposure and Mild Traumatic Brain Injury. *Brain Sci.* 2018;8(11):198. doi:10.3390/brainsci8110198
 128. Lee YK, Hou SW, Lee CC, Hsu CY, Huang YS, Su YC. Increased Risk of Dementia in Patients with Mild Traumatic Brain Injury: A Nationwide Cohort Study. *PLoS One.* 2013;8(5):e62422. doi:10.1371/JOURNAL.PONE.0062422
 129. Gardner RC, Yaffe K. Epidemiology of mild traumatic brain injury and neurodegenerative disease. *Mol Cell Neurosci.* 2015;66(PB):75-80. doi:10.1016/J.MCN.2015.03.001
 130. Lehman EJ, Hein MJ, Baron SL, Gersic CM. Neurodegenerative causes of death among retired national football league players. *Neurology.* 2012;79(19):1970-1974. doi:10.1212/WNL.0b013e31826daf50
 131. Pupillo E, Bianchi E, Vanacore N, et al. Increased risk and early onset of ALS in professional players from Italian Soccer Teams. *Amyotroph Lateral Scler Front Degener.* 2020;21(5-6):403-409. doi:10.1080/21678421.2020.1752250
 132. Beghi E. Are professional soccer players at higher risk for ALS? *Amyotroph Lateral Scler Front Degener.* 2013;14(7-8):501-506. doi:10.3109/21678421.2013.809764
 133. Hernandez S, Fried DE, Grubišić V, McClain JL, Gulbransen BD. Gastrointestinal neuroimmune disruption in a mouse model of Gulf War illness. *FASEB J.* 2019;33(5):6168-6184. doi:10.1096/fj.201802572R
 134. Coughlin SS. A Neuroimmune Model of Gulf War Illness. *J Environ Heal Sci.* 2017;3(3):1-6. doi:10.15436/2378-6841.17.1665
 135. Mothersill C, Seymour C. Fundamental Mechanisms Underlying the Ill Health and Chronic Fatigue Syndrome Suffered by Atomic and Gulf War Veterans: A Unifying Hypothesis. *Genet Evol Radiat Crossing Borders, Interdiscip Leg Nikolay W Timofeeff-Ressovsky.* January 2016:347-356. doi:10.1007/978-3-319-48838-7_29

136. Grant SG, Ibrahim OM, Jin X-LL, Klimas NG, Sullivan K, Latimer JJ. Elevated somatic mutation and evidence of genomic instability in veterans with Gulf War illness. *Life Sci.* 2021;281:119746. doi:10.1016/j.lfs.2021.119746
137. Coen N, Mothersill C, Kadhim M, Wright EG. Heavy metals of relevance to human health induce genomic instability. *J Pathol.* 2001;195(3):293-299. doi:10.1002/PATH.950
138. Filipič M. Mechanisms of cadmium induced genomic instability. *Mutat Res Mol Mech Mutagen.* 2012;733(1-2):69-77. doi:10.1016/J.MRFMMM.2011.09.002
139. Hussain T, Liu B, Shrock MS, Williams T, Aldaz CM. WWOX, the FRA16D gene: A target of and a contributor to genomic instability. *Genes Chromosom Cancer.* 2019;58(5):324-338. doi:10.1002/gcc.22693
140. Ried K, Finnis M, Hobson L, et al. Common chromosomal fragile site FRA16D sequence: Identification of the FOR gene spanning FRA16D and homozygous deletions and translocation breakpoints in cancer cells. *Hum Mol Genet.* 2000;9(11):1651-1663. doi:10.1093/hmg/9.11.1651
141. Vessoni AT, Filippi-Chiela EC, Menck CFM, Lenz G. Autophagy and genomic integrity. *Cell Death Differ.* 2013;20(11):1444-1454. doi:10.1038/cdd.2013.103
142. Ambrosio S, Majello B. Autophagy Roles in Genome Maintenance. *Cancers (Basel).* 2020;12(7):1793. doi:10.3390/cancers12071793
143. Chapman JC, Diaz-Arrastia R. Military traumatic brain injury: A review. *Alzheimer's Dement.* 2014;10(3S):S97-S104. doi:10.1016/j.jalz.2014.04.012
144. Fukuda K, Strauss SE, Hickie I, Sharpe MC, Dobbins JG, Komaroff A. The Chronic Fatigue Syndrome: A Comprehensive Approach to Its Definition and Study. *J Chronic Fatigue Syndr.* 1995;1(2):67-84. doi:10.1300/J092v01n02_06
145. Eisen SA. Gulf War Veterans' Health: Medical Evaluation of a U.S. Cohort. *Ann Intern Med.* 2005;142(11):881. doi:10.7326/0003-4819-142-11-200506070-00005
146. Smith BN, Wang JM, Vogt D, Vickers K, King DW, King LA. Gulf War Illness. *J Occup Environ Med.* 2013;55(1):104-110. doi:10.1097/JOM.0b013e318270d709
147. Saccon RA, Bunton-Stasyshyn RKA, Fisher EMC, Fratta P. Is SOD1 loss of function involved in amyotrophic lateral sclerosis? *Brain.* 2013;136(8):2342-2358. doi:10.1093/brain/awt097

148. Haley RW. Excess incidence of ALS in young Gulf War veterans. *Neurology*. 2003;61(6):750-756. doi:10.1212/WNL.61.6.750
149. Merwin SJ, Obis T, Nunez Y, Re DB. Organophosphate neurotoxicity to the voluntary motor system on the trail of environment-caused amyotrophic lateral sclerosis: the known, the misknown, and the unknown. *Arch Toxicol*. 2017;91(8):2939-2952. doi:10.1007/s00204-016-1926-1
150. Horner RD, Kamins KG, Feussner JR, et al. Occurrence of amyotrophic lateral sclerosis among Gulf War veterans. *Neurology*. 2003;61(6):742-749. doi:10.1212/01.wnl.0000069922.32557.ca
151. Zhu GD, Dawson E, Huskey A, Gordon RJ, Tredici AL Del, Del Tredici AL. Genetic testing for bche variants identifies patients at risk of prolonged neuromuscular blockade in response to succinylcholine. *Pharmgenomics Pers Med*. 2020;13:405-414. doi:10.2147/PGPM.S263741
152. Dardiotis E, Aloizou A-M, Siokas V, et al. Paraoxonase-1 genetic polymorphisms in organophosphate metabolism. *Toxicology*. 2019;411:24-31. doi:10.1016/j.tox.2018.10.012
153. Verde F, Tiloca C, Morelli C, et al. PON1 is a disease modifier gene in amyotrophic lateral sclerosis: association of the Q192R polymorphism with bulbar onset and reduced survival. *Neurol Sci* 2019 407. 2019;40(7):1469-1473. <https://link.springer.com/article/10.1007/s10072-019-03834-2>. Accessed April 5, 2021.
154. Huo X, Guo Y, Zhang Y, Li J, Wen X, Liu J. Paraoxonase 1 gene (Q192R) polymorphism confers susceptibility to coronary artery disease in type 2 diabetes patients: Evidence from case-control studies. *Drug Discov Ther*. 2019;13(2):80-88. doi:10.5582/ddt.2019.01003
155. Chen Y, Meyer JN, Hill HZ, et al. Role of mitochondrial DNA damage and dysfunction in veterans with Gulf War Illness. Kavushansky A, ed. *PLoS One*. 2017;12(9):e0184832. doi:10.1371/journal.pone.0184832
156. Goodall R. Cholinesterase: Phenotyping and genotyping. *Ann Clin Biochem*. 2004;41(2):98-110. doi:10.1258/000456304322879971
157. Davies HG, Richter RJ, Keifer M, Broomfield CA, Sowalla J, Furlong CE. The effect of the human serum paraoxonase polymorphism is reversed with diazoxon, soman and sarin. *Nat Genet*. 1996;14(3):334-336. doi:10.1038/ng1196-334

158. Shapira M. A transcription-activating polymorphism in the ACHE promoter associated with acute sensitivity to anti-acetylcholinesterases. *Hum Mol Genet.* 2000;9(9):1273-1281. doi:10.1093/hmg/9.9.1273
159. Furlong CE, Marsillach J, Jarvik GP, Costa LG. Paraoxonases-1, -2 and -3: What are their functions? *Chem Biol Interact.* 2016;259:51-62. doi:10.1016/j.cbi.2016.05.036
160. Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome. *Nature.* 2018;558(7708):73. doi:10.1038/S41586-018-0175-2
161. Parihar VK, Hattiangady B, Shuai B, Shetty AK. Mood and Memory Deficits in a Model of Gulf War Illness Are Linked with Reduced Neurogenesis, Partial Neuron Loss, and Mild Inflammation in the Hippocampus. *Neuropsychopharmacology.* 2013;38(12):2348-2362. doi:10.1038/npp.2013.158
162. Phillips KF, Deshpande LS. Repeated low-dose organophosphate DFP exposure leads to the development of depression and cognitive impairment in a rat model of Gulf War Illness. *Neurotoxicology.* 2016;52:127-133. doi:10.1016/j.neuro.2015.11.014
163. CSP2006 Personal Communication. *Genome-Wide Association Study of Gulf War Illness in the Million Veteran Program.*

Biography

Jacqueline Vahey got her undergraduate degree, a Bachelor of Science in Computer Science and Molecular Biology, from the Massachusetts Institute of Technology in 2017. She was awarded the James B. Duke Fellowship for her first four years of PhD education at Duke. She has published several journal papers, both as a first author and as a coauthor. Her first author manuscripts, “Gene-Toxicant Interactions in Gulf War Illness: Differential Effects of the PON1 Genotype” and “Research tool for classifying Gulf War Illness using survey responses: Lessons for writing replicable algorithms for symptom-based conditions”, and her coauthor manuscripts, “Gulf War illness in the Gulf War Era Cohort and Biorepository: The Kansas and Centers for Disease Control definitions” and “Genomics of Gulf War Illness in U.S. Veterans Who Served during the 1990–1991 Persian Gulf War: Methods and Rationale for Veterans Affairs Cooperative Study #2006”, are referenced in the above dissertation document.