

APPENDIX LL

INDEPENDENT CONSULTANT REPORT

BY

SHANNA H. SWAN, PH.D.

Epidemiological Studies of the Reproductive Health of Persian Gulf War Veterans

Report to the Senate Committee on Veterans' Affairs

Shanna H. Swan Ph.D.

Berkeley, CA

December 31, 1997



Table of Contents

Executive Summary	3
Overall Summary	4
Statement of Work	9
Outline	10
Section I: Describing current studies	11
A. Description of methods and results of completed studies	11
B. Description of ongoing studies	15
Section II: Evaluating completed and ongoing studies	21
A. General criteria for evaluating reproductive epidemiology studies	21
B. Evaluation of current studies	27
Section III. Recommendations for further study	29
A. Improved exposure assessment	29
B. Estimating exposure misclassification	30
C. Further study of Goldenhar Syndrome	30
D. Targeted reproductive follow-up for cases of Persian Gulf War Unexplained Illness ...	30
E. Estimating misclassification in birth defect studies	31
F. Estimating delay to conception	31
References	32
Tables	33

Epidemiological Studies of the Reproductive Health of Persian Gulf War Veterans

Report to the Senate Committee on Veterans' Affairs

Shanna H. Swan Ph.D.

December 31, 1997

Executive Summary

This report reviews the epidemiologic studies that have been conducted and are currently underway to assess the reproductive health of personnel who served in the 1990-91 Persian Gulf War (PGW). To this end, I have attempted to review all relevant studies published by December 1, 1997 which evaluate one or more reproductive outcomes in military personnel that deployed to the PGW, as well as proposals, protocols, and questionnaires for ongoing studies on this topic.

This report summarizes the evidence currently available on possible adverse reproductive effects of Gulf War exposures. It is my opinion that existing studies can not adequately evaluate whether or not exposures to PGW have resulted in reproductive harm to exposed veterans. While existing studies indicate some increased risks of menstrual disturbances, sexual discomfort, and the rare complex of congenital anomalies known as Goldenhar Syndrome, numbers of cases are small. Of these findings, only that of an increase in sexual discomfort in exposed veterans reached traditional levels of statistical significance.

With a few notable exceptions (e.g. the Oregon Health Sciences University study and, the Klemm Analysis Group study), these studies are severely limited by their incomplete exposure assessment. Because PGW veterans were potentially exposed to a wide range of chemical, biological, physical and psychological stressors, and because exposure varied with time of deployment, location, service and occupation, the deployment-nondeployment exposure classification used in most of these studies is likely to classify veterans inaccurately with respect to many exposures. As discussed in the report, these limitations are likely to result in underestimates of the risks of PGW exposure. These studies are also quite limited in their statistical power to detect increased risks of rare outcomes. Further, many of these studies are limited by their exclusion of a large proportion of PGW-exposed veterans including those no longer in active service, and National Guard/Reservists. Most of the birth defect studies, in particular, are limited by their exclusion of births in civilian hospitals and of diagnoses made after the birth hospitalization.

I have suggested several additional analyses that can be conducted within the framework of ongoing studies that may remedy several of these limitations. This area of investigation is clearly of considerable public health importance and the extent of the effort to resolve the question of reproductive dysfunction in Gulf War veterans is impressive. Over time the quality of the studies has improved markedly, and it can be expected that the next generation of studies will do much to resolve these issues.

Overall Summary

Introduction

This report reviews the epidemiologic studies that have been, and are being, conducted to assess the reproductive health of personnel that served in the 1990-91 Persian Gulf War (PGW). To this end, I have attempted to review all relevant published studies, as well as proposals, protocols, and questionnaires for ongoing studies. I have included all studies whose results were published in scientific journals or presented at scientific meetings as of December 1, 1997, as well as studies that were in progress as of that date.

With a few notable exceptions (e.g. the Oregon Health Sciences University study, and the Klemm Analysis Group Study), the completed and ongoing studies are severely limited by their incomplete exposure assessment. Because PGW veterans were potentially exposed to a wide range of chemical, biological, physical and psychological stressors, and because exposure varied with time of deployment, location, service and occupation, the deployment-nondeployment exposure classification used in most of these studies is likely to classify veterans inaccurately with respect to many exposures. As discussed in the report, these limitations are likely to result in underestimates of the risks of PGW exposure. These studies are also quite limited in their statistical power to detect increased risks of rare outcomes. Further, many of these studies are limited by their exclusion of a large proportion of PGW-exposed veterans including those no longer in active service, and National Guard/Reservists. Most of the birth defect studies, in particular, are limited by their exclusion of births in civilian hospitals, and diagnoses after the birth hospitalization.

Current studies

Five studies were published by December 1997 which include data on the reproductive health of PGW veterans. These are: Stretch *et al* (1995), Penman *et al* (1996), Iowa Persian Gulf Study Group (1997), Cowan *et al* (1997), and Araneta *et al* (1997). Three of these (Penman, Cowan and Araneta) examined the relationship between birth defects and PGW exposure. In connection with the Araneta publication I also discuss an additional source of case ascertainment for Goldenhar Syndrome, which is the subject of the Araneta study. The remaining two completed studies (Stretch and the Iowa Persian Gulf Study Group) examined PGW exposure and self-reported symptoms, which included one or more reproductive symptoms or conditions.

Stretch *et al* (1995) analyze symptoms self-reported by deployed and non-deployed veterans using questionnaires mailed to 16,167 active duty and reserve personnel in the states of Hawaii and Pennsylvania. Their low response rate (31%) may be due, in part, to the fact that questionnaires were distributed to units rather than to individuals. The only reproductive outcome that was reported in this publication was "menstrual difficulties". Among active duty respondents, rates of this outcome were low and similar in deployed and non-deployed (1.7% and 1.5% respectively). Rates among reservists were higher than those reported by active duty personnel and 34% higher among deployed than non-deployed (3.1% and 2.3% respectively).

Penman *et al* (1996) evaluated birth defects and other health problems among children of veterans of two Mississippi guard units who had served in the PGW. The medical records of all (282) children of these veterans were reviewed. No concurrent control group was utilized; rates were compared to those expected from birth defects surveillance systems and previous surveys. Among 254 (90%) who were interviewed, 54 reported births that were conceived post-

deployment. Medical record review was conducted to ascertain birth defects (major and minor), premature births, low birth weight and other health problems. Five birth defects (three major, two minor), five cases of low birth weight, and no stillbirths or deaths noted. No increased risks were observed compared to rates from surveillance systems. No attempt was made to characterize exposure.

The Iowa Persian Gulf Study Group (1997) estimated the prevalence of self-reported symptoms and illnesses among military personnel deployed during the PGW compared to personnel on active duty at the same time, but not deployed to the PGW (non-PGW). For this purpose, a stratified random sample was used to select a study population of 4,886 Iowa veterans. Each individual was classified as either PGW regular military, PGW National Guard/Reserve, non-PGW regular military and non-PGW National Guard/Reserve. Subjects were interviewed regarding a range of medical and psychiatric conditions. The only reproductive outcome that was reported in this publication was "symptoms of sexual discomfort". The prevalence of sexual discomfort among female partners was approximately doubled among PGW veterans compared to non-PGW veterans (5.0% vs. 2.4% for regular military and 5.4% vs. 2.1% among National Guard Reservists). Both of these comparisons were statistically significant at the 95% level.

Cowan *et al* (1997) studied the relationship between service in the PGW and the overall risk of birth defects for all US veterans. For this purpose the authors accessed live births at 135 military hospitals between 1991 and 1993. During that time, 33,998 infants were born to PGW veterans and 41,463 to non-deployed veterans at these hospitals. Birth defects, as routinely recorded on birth records, were obtained for all live births. Military records were accessed to obtain information on military service and deployment locations. Exposure was defined simply as "deployment to the PGW". While no association between PGW service and birth defects was seen for male service members, among females there was a small, but statistically significant, increase. Using the broadest definition of congenital malformations, malformations were noted in 10.32% of births to deployed veterans versus 9.2% to nondeployed [unadjusted relative risk 1.12, 95% confidence interval (CI) (1.00 to 1.25)]. After adjustment for race, marital status and branch of service the relative risk was reduced to 1.07 (95% CI 0.94-1.22). The risk of a severe birth defect was slightly (and not significantly) lower among children of active duty women than among children of non-deployed (2.0% versus 2.1%), and both were similar to that reported by the CDC (1.9%). Six commonly occurring groups of defects were examined and none were associated with PGW exposure either in men or women. Crude (unadjusted) birth rates were significantly higher in PGW veterans than non-deployed (95.6 per 1,000 versus 93.3 per thousand). The ratio of male to female births was similar in deployed and non-deployed veterans.

The frequency of occurrence of Goldenhar Syndrome, the most severe group of anomalies to form an oculo-auricular-vertebral syndrome was estimated in deployed and non-deployed veterans by Araneta *et al* (1997). The authors ascertained cases diagnosed at birth among infants born to active-duty military personnel in military hospital using a broad screen of hospital discharge diagnoses. Potential cases were identified using 66 ICD-9-CM codes, including the general category "anomaly of skull and face bones", and selected ear anomalies. Medical record review by expert reviewers, blinded to exposure status, was used to identify definite cases of Goldenhar Syndrome among these potential cases. For all the seven cases identified, the father was the parent in the military. Five of these were offspring of PGW veterans (14.7 cases per 100,000) and two were offspring of non-deployed veterans (4.8 cases per 100,000). Thus, the

relative risk was elevated (relative risk = 3.0, 95% CI 0.6 – 20.6) though not statistically significantly. As shown in Table 2, the rate observed in PGW exposed was significantly higher than that reported by either the Hawaii Birth Defects Program or the Metropolitan Atlanta Congenital Defects program (4-5 per 100,000).

The Association of Birth Defect Children (ABDC) actively solicits the reporting of birth defects. As part of this activity, 18 cases of Goldenhar Syndrome were identified in veterans; 15 were deployed to the PGW. Since this registry is more likely to obtain case referrals from exposed veterans, it cannot be assumed to include a representative sample of unexposed cases. These cases are summarized and compared to the seven cases reported by Araneta *et al* in Table 3.

Ongoing studies

I have identified eight ongoing studies that should provide additional information on the risk of adverse reproductive outcomes among PGW veterans.

Study 3 is a comparative study of pregnancy outcomes among PGW veterans (male and female) and other active duty personnel. I could not determine whether other outcomes will be examined in this study.

Study 4 is examining differences between PGW veterans and non-deployed veterans with respect to infertility, time to conception and risk of miscarriage. In Phase I of this study a questionnaire was mailed to a random sample of 16,000 couples (8,000 couples for which one or both deployed to the PGW, and 8,000 for which neither deployed). Currently the participation rate is 46%. Phase II will consist of a telephone interview of 5,000 couples to obtain detailed information on exposures and known risk factors for infertility and miscarriage. The following four categories of married couples are included: (1) woman served in the PGW; (2) man served in the PGW; (3) woman served in the military during the PGW, but not in the Gulf area; and (4) man served in the military during the PGW, but not in the Gulf area.

Study 7 is examining the prevalence of congenital anomalies in the seven states that maintain active birth defects surveillance systems. These include all birth defects diagnosed in live births during the first year of life and in still births. This study also proposes to compare rates of preterm birth, low birth weight and still birth between PGW veterans and non-deployed veterans in the seven states. Births between 1989 and 1993 will be included in order to compare conceptions prior to, during, and after the PGW.

The California Birth Defects Monitoring Program (CBDMP) will conduct a feasibility study to determine; (1) whether Department of Defense (DOD) data on births to active duty military personnel are sufficient to allow the CBDMP to locate the medical records of these children during their first year of life; (2) whether hospital record review is possible at DOD facilities, particularly those which may be closed or have incomplete medical record information; (3) whether DOD information about structural congenital anomalies is sufficiently accurate, compared to complete hospital medical record information. This study will also determine if DOD information about the identity of inactive (separated) personnel can be linked to California vital records and CBDMP files, neither of which contains social security information.

The most unusual reproductive tract abnormality reported by PGW veterans and their spouses is the "Gulf War Vaginal Burning Syndrome". In cases of this syndrome, which can be local or systemic, severe vaginal burning and pain are reported to occur immediately on contact with the spouse's seminal fluid. A study being conducted by the University of Cincinnati has, as its first goal to determine whether this syndrome in PGW veterans is due to the same immune responses previously described for cases in the general community. Ten cases in which the husband is an exposed veteran as well as ten unaffected spouses of exposed veterans will be selected for comparison. The second goal of the study is to identify seminal plasma proteins involved in the pathogenesis of this syndrome in spouses of PGW veterans, to determine whether these are the same as the proteins identified in cases in the general population. For this purpose, five ejaculates, collected over five consecutive days will be obtained and used to isolate seminal plasma proteins from each male participant. Women will then be tested for sensitivity to these seminal proteins using skin prick tests. The third study goal is to determine the effects of PGW exposures on human seminal plasma obtained from both PGW-exposed and non-exposed males.

The Oregon Health Sciences University study will identify risk factors for Persian Gulf War

Unexplained Illness (PGWUI) in veterans from the northwestern United States. For this purpose a population-based questionnaire is being mailed to a representative sample of deployed veterans within the following strata; (1) pre-combat (Desert Shield) only; (2) combat (Desert Storm) only; (3) post-combat (desert cleanup) only; and (4) two or more of these. By using a sampling strategy based on period of deployment, the role of potential risk factors such as pyridostigmine bromide, special vaccinations and combat stress can be isolated and analyzed. Respondents to the mailed survey will provide the study population for the clinical case-control phase of the study. In this phase, the nature and pattern of exposures in cases of PGWUI and controls will be compared. A total of 250 cases and controls will be recruited for clinical testing within four months of responding to the survey.

The Veteran's Administration, is conducting a three-phase study which includes a range of reproductive endpoints. In Phase I, a mailed questionnaire was sent to a random sample of 15,000 PGW veterans and a control sample of 15,000 Gulf-era veterans. To validate responses and evaluate effects of a low response rate (50%), in Phase II, 2,000 respondents among the deployed, and 2,000 among the non-deployed are being contacted by phone to obtain permission to review medical records. Further, a random sample of 8,000 non-respondents was selected to compare respondents and non-respondents. In Phase III physical examinations will be conducted on 1,000 veterans randomly selected from each group (deployed and non-deployed) as well as their family members.

The Klemm Analysis Group is conducting a two-year study comparing the health status of 10,000 women who served in the PGW with 10,000 Gulf-era military women. For this purpose a questionnaire has been developed inquiring about symptoms and conditions including adverse reproductive outcomes such as infertility, pre-term births, still births and birth defects. Detailed information on exposures before, during and after the PGW is being elicited.

Recommendations for further study

Most of the studies of the reproductive health of PGW veterans conducted to date include only limited exposure assessment. The most notable exception is the Oregon Health Sciences University Study (OHSU), which can be taken as a model for this purpose. The Klemm Analysis Group questionnaire also includes a strong exposure assessment component. The birth defect studies are particularly weak in this respect, with the exception of the Iowa study, which contains a fairly extensive exposure component. Therefore, I recommend that a nested-case-control study be imbedded in Study 7, and a detailed exposure assessment be conducted, perhaps using the OHSU instrument for consistency and later comparison across studies.

The study of Arenata *et al* documents an increased risk of Goldenhar Syndrome among potentially exposed veterans. However, this increase is not statistically significant, possibly due to small numbers. Therefore, I recommend expanding this study, both to obtain additional cases and to improve the exposure assessment. To this end I recommend first evaluating the possible additional cases of Goldenhar Syndrome which have been identified by the ABDC registry. It should be determined whether any of the 15 exposed cases identified by the ABDC includes cases that should have been identified by the Araneta *et al* study protocol but were inadvertently missed. In other words, were all ten of the additional exposed cases identified by the ABDC ineligible for the Araneta study? Conversely, were all five exposed cases identified in Araneta *et al* included among the ABDC cases? It is also recommended that systematic case ascertainment for Goldenhar Syndrome be expanded in both deployed and nondeployed veterans.

including births to separated personnel and all births to veterans in civilian hospitals. Ascertainment throughout the first year of life, using the full medical records would be ideal. In addition, it is important to obtain detailed exposure information on all cases and a sample of controls, perhaps using the Oregon Health Sciences' questionnaire to obtain exposure information. It is also important to determine whether the cases of Goldenhar were the first live births born to veterans post-deployment. A causal relationship between this syndrome and births after one or more healthy babies seems unlikely.

The Oregon Health Sciences' University is has provided a tentative definition of Persian Gulf War Unexplained Illness (PGWUI), and is ascertaining cases of PGWUI in the Northwest. Since it is still uncertain what exposures are most relevant for reproductive illness in PGW veterans, I recommend looking for an increased incidence of reproductive abnormalities in cases of PGWUI. It is plausible that these veterans, most affected systemically by these exposures, would also exhibit more reproductive dysfunction in connection with PGW exposures. This reproductive assessment should be as complete as possible and should include serum hormone analyses on cases of PGWUI in the Northwest cohort. In addition, it would be valuable to examine semen quality in male cases. To date none of these studies has examined semen quality of veterans. Females could be asked to maintain a detailed dairy recording menstruation, frequency of intercourse and use of contraception that would allow for a precise analysis of time to conception. If daily urine samples were obtained as well, assays would provide information on early fetal loss. (See Tier II analyses, Table 6).

Several sources of misclassification in the birth defect studies conducted or underway are listed above. I recommend that the magnitude of the resulting misclassification be estimated using a sample of births from Study 7. This analysis would probably have to be limited to the five states that have active birth defect surveillance for infants up to one year of age throughout the state (thus excluding Californian and Georgia). This could be done by obtaining as complete an ascertainment of birth defects as possible on the selected sample, and then determining how many of these birth defects would have been missed if; (1) only the birth record had been used; (2) only military hospitals had been used; (3) only active-duty personnel had been included. The degree of underreporting could then be examined as a function of severity of the defect and other covariates.

Statement of Work

This report reviews the epidemiologic studies that have been, and are being, conducted to assess the reproductive health of personnel that served in the 1990-91 Persian Gulf War (PGW). To this end, I have attempted to review all relevant published studies, as well as proposals, protocols, and questionnaires for ongoing studies. Relevant studies for this purpose are studies that evaluate one or more reproductive outcomes in military personnel that deployed to the PGW. Most studies compare reproductive outcomes in PGW veterans and military personnel that did not deploy to the PGW (non-PGW). I have attempted to include all studies whose results were published in scientific journals or presented at scientific meetings as of December 1, 1997, as well as studies that were in progress as of that date. I have summarized each study's design, study population (including the inclusion/exclusion of possible high risk groups, or selected subgroups such as reservists), methodological soundness, and strengths and weaknesses.

This report includes a summary of the evidence currently available on possible adverse reproductive effects of Gulf War exposures, as well as my opinion as to whether completed and ongoing studies appropriately controlled for biases and risk factors for the adverse reproductive outcomes. I have reviewed the range of possible adverse reproductive outcomes to determine which have, and have not, been addressed in relevant studies to date. Finally, I recommend additional analyses and further research efforts which should allow for a more complete assessment of the extent of reproductive risks from exposures incurred by veterans of the PGW.

Outline

Section I: Describing Current Studies:

A. Description of Methods and Results of Completed Studies

This section summarizes all published epidemiological studies that include data on the reproductive health of PGW veterans. For each, the study goals, study design, study populations, exposure definitions and other methodologic details are discussed. Results of analyses of reproductive outcomes are also summarized.

B. Description of Methods of Ongoing Studies

This section discusses the study proposals, protocols, and questionnaires, when available, for all studies in progress that are collecting data on the reproductive health of PGW veterans. Since these studies have not yet been completed, it is not possible to specify details of study design and methods with certainty. Progress towards study goals is described when this information was available to me.

Section II: Evaluating Current Studies

A. General criteria for evaluating reproductive epidemiology studies

This section discusses commonly utilized criteria for evaluating epidemiologic studies, and reproductive epidemiologic studies in particular. This section includes a discussion of the importance of conducting adequate exposure assessment. It also describes the range of reproductive outcomes that may be relevant. Criteria for assessing whether an observed association should be considered causal (the Hill Criteria) are also discussed briefly.

B. Evaluating current studies

These general criteria are then applied to the current (published and ongoing) studies in order to assess their strengths and weaknesses. For ongoing studies this assessment is somewhat tentative since it must be based on available proposals and protocols, and final analyses may differ.

Section III: Recommendations for future studies

This section makes recommendations for further analyses of existing studies and for additional studies.

Section I: Describing current studies

A. Description of methods and results of completed studies

I have identified five studies that were published by December 1997 which include data on the reproductive health of PGW veterans. These are: Stretch *et al* (1995), Penman *et al* (1996), Iowa Persian Gulf Study Group (1997), Cowan *et al* (1997), and Araneta *et al* (1997). Three of these (Penman, Cowan and Araneta) examined the relationship between birth defects and PGW exposure. In connection with the Araneta publication I also discuss an additional source of case ascertainment for Goldenhar Syndrome, which is the subject of the Araneta study. The remaining two completed studies (Stretch and the Iowa Persian Gulf Study Group) examined PGW exposure and self-reported symptoms, including one or more reproductive symptoms or condition. Only the reproductive outcomes addressed by these latter studies are discussed here. These five studies are discussed below and summarized in Table 1.

Stretch et al (1995)

Stretch *et al* (1995) analyze symptoms self-reported by deployed and non-deployed veterans. For this purpose, questionnaires were mailed to 16,167 active duty and reserve personnel who had been assigned to Army, Navy, Air Force or Marine units in the states of Hawaii and Pennsylvania. Of the total sample of 4,334 veterans who returned the survey (31%), 1,7339 deployed as a result of Operation Desert Shield/Storm and 2,512 did not deploy. This low response rate may be due, in part, to the fact that questionnaires were distributed to units rather than to individuals. Of the 1,739 who deployed, 1,524 (88%) deployed to the Persian Gulf, the remainder deployed to other locations. The active duty sample consists of 715 who deployed (41% of all deployed and 63% of all non-deployed). These respondents were primarily (88%) from the Navy and the Marine Corp. The published paper does not state what proportion of respondents were female.

The only reproductive outcome that was reported in this publication was “menstrual difficulties”. Among active duty respondents, rates of this outcome were low and similar in deployed and non-deployed (1.7% and 1.5% respectively). Rates among reservists were higher than those reported by active duty personnel and 34% higher among deployed than non-deployed (3.1% and 2.3% respectively). With respect to exposure among those deployed, 28% of active duty and 40% of reservists reported exposure to oil fires. No additional exposure information is provided in this publication.

Penman et al (1996)

Penman *et al* (1996) evaluated birth defects and other health problems among children of veterans of two Mississippi guard units who had served in the PGW. The medical records of all children of these veterans were reviewed. In addition to birth defects (major and minor), premature birth, low birth weight and other health problems were ascertained. No concurrent control group was utilized; rates were compared to those expected from birth defects surveillance systems and previous surveys.

The study population consisted of 282 veterans serving in two Mississippi guard units. Of these, 254 were interviewed (90%) and reported 54 births conceived post-deployment. The mother was the veteran for six of these (11%). Medical record review was conducted to ascertain birth defects (major and minor), premature births, low birth weight and other health problems. Five birth defects (three major, two minor), five cases of low birth weight, and no stillbirths or deaths noted. It was not stated which parent was deployed to the PGW among the cases. No attempt was made to characterize exposure.

Iowa Persian Gulf Study Group (1997)

The Iowa Persian Gulf Study Group (1997) estimated the prevalence of self-reported symptoms and illnesses among military personnel deployed during the PGW compared to personnel on active duty at the same time, but not deployed to the PGW (non-PGW). For this purpose, a stratified random sample was used to select a study population of 4,886 Iowa veterans. Military personnel were eligible for inclusion if Iowa was listed as their home of record and they had served in the regular military or activated National Guard/Reserve at some time between August 2, 1990 and July 31, 1991. Each individual was classified as either PGW regular military, PGW National Guard/Reserve, non-PGW regular military and non-PGW National Guard/Reserve. PGW subjects were further subdivided based on the area in which subjects reported spending the most time during the PGW; Iraq, Saudi Arabia or Kuwait; other countries in the Middle East; waters bordering the Middle East.

Subjects were interviewed regarding a range of medical and psychiatric conditions. The only reproductive outcome that was reported in this publication was "symptoms of sexual discomfort". With respect to exposure, a detailed exposure questionnaire was administered inquiring about number of days in the theater, number of vaccinations, number of pyridostigmine tablets used, and exposure to a variety of agents such as solvents, pesticides, chemical warfare agents, lead, ionizing radiation, infectious agents, psychological stressors, and physical trauma.

Overall 76% of eligible study subjects completed a phone interview. This represented 91% of contacted subjects. Symptoms of sexual discomfort (in respondent and respondent's sexual partner) were more frequently reported among PGW military personnel. The prevalence of sexual discomfort among female partners was approximately double among PGW than non-PGW (5.0% vs. 2.4% for regular military and 5.4% vs. 2.1% among National Guard Reservists). Both of these comparisons were statistically significant at the 95% level. While the paper discusses a general relationship of increased symptom reporting with increased exposure reporting, and a higher prevalence of symptoms among veterans deployed to Iraq, Saudi Arabia or Kuwait, it does not discuss these exposure specifically with respect to sexual discomfort.

Cowan et al (1997)

Cowan *et al* (1997) studied the relationship between service in the PGW and the overall risk of

birth defects for all US veterans. For this purpose the authors accessed live births at 135 military hospitals between 1991 and 1993. During that time, 33,998 infants were born to Gulf War veterans and 41,463 to non-deployed veterans at these hospitals. Eligible births to deployed were those born before 10/1/93 with estimated date of conception post-deployment. Eligible births to non-deployed were those born before 10/1/93 with estimated date of conception after 12/31/90. Birth defects, as routinely recorded on birth records, were obtained for all live births at military hospitals. Births paid for by the military at civilian hospitals were used for estimates of number of births and the male to female ratio, but not to ascertain birth defects.

Military records were accessed to obtain information on military service and deployment locations. Exposure was defined simply as "deployment to the PGW". Risk was also estimated as a function of duration of exposure (number of days of deployment). Changes in risk over time were assessed by examining the interval between return from the PGW and birth date of the child. There was no attempt to investigate subgroups with unique exposures.

Complete demographic and military information was available on 578,705 deployed (PGW) veterans (6% female) and 699,954 non-deployed veterans (12% female). As of 10/1/93 (the close of the study period) 56% of deployed males and 57% of non-deployed males were still on active duty, and thus eligible to give birth in military hospitals. (These percents were very similar for females). Of all identified births to wives of veterans during the study period, 58% of those deployed and 57% of nondeployed were in military hospitals. However, nearly all (over 99%) of births to female veterans occurred in military hospitals. Compared to non-deployed, male and female PGW veterans were younger, more likely to be single, black, of enlisted rank and in the army.

While no association was seen between PGW service and birth defects was seen for male service members, among females there was a small, but statistically significant, increase. Using the broadest definition of congenital malformations, malformations were noted in 10.32% of births to deployed veterans versus 9.2% to nondeployed; unadjusted relative risk 1.12, 95% confidence interval (CI) (1.00 to 1.25). Both of these rates are somewhat higher than that reported by the CDC Birth Defects Monitoring Program, using the same definitions (8.4%). This association was seen primarily in black, single women and was not related to time spent in the Persian Gulf. After adjustment for race, marital status and branch of service the relative risk was reduced to 1.07 (0.94-1.22). The risk of a severe birth defect was slightly (and not significantly) lower among children of active duty women than among children of non-deployed (2.0% versus 2.1%), and both were similar to that reported by the CDC (1.9%). Six commonly occurring groups of defects were examined and none were associated with PGW exposure either in men or women. Crude (unadjusted) birth rates were significantly higher in PGW veterans than non-deployed (in males 95.6 per 1,000 versus 93.3 per thousand). The ratio of male to female births was similar in deployed and non-deployed veterans.

This study is part of the Naval Health Research Center (NIIRC) Study 3.

Araneta et al (1997)

Goldenhar Syndrome is considered the most severe group of anomalies to form an oculo-auditory-vertebral syndrome. Estimates of frequency of occurrence range from 4-7 per 100,000 live births in active birth defect surveillance systems (See Table (2)). This syndrome appears somewhat more often in male infants, with a male: female sex ratio of 1:18 (California Birth Defects Monitoring Program). Because of anecdotal reporting of an excess of this rare syndrome

among PGW veterans, Araneta *et al* (1997) examined the prevalence of this syndrome as a function of deployment to the PGW. These analyses, as those of Cowan *et al* above, are part of the Naval Health Research Center (NHRC) Study 3. The authors ascertained cases diagnosed at birth among infants born to active-duty military personnel in military hospital. The study population was almost identical to that utilized by Cowan *et al*, with only slight differences in eligibility based on birth dates. However, unlike Cowan *et al*, Araneta utilized a broad screen of hospital discharge diagnoses (using 66 ICD-9-CM codes, including the general category “anomaly of skull and face bones”, and selected ear anomalies) to ascertain potential cases. Cases were then diagnosed by blinded expert medical record review of the potential cases that had been identified by this broad screen. Exposure was defined by deployment status only.

Seven cases were identified by this study. For all of these, the father was the parent in the military. Five of these were offspring of PGW veterans (14.7 cases per 100,000) and two were offspring of non-deployed veterans (4.8 cases per 100,000). Thus, the relative risk was elevated (relative risk = 3.0, 95% CI 0.6 – 20.6) though not statistically significantly. The rate observed in PGW exposed was significantly higher than that reported by either the Hawaii Birth Defects Program or the Metropolitan Atlanta Congenital Defects program (4-5 per 100,000) and higher than the California Birth Defects Monitoring Program rate of 6.8 per 100,000, although not significantly so (Table (2)). On the other hand, the rate in the unexposed was quite similar to these registry rates.

An additional source of cases of Goldenhar Syndrome was identified. The Association of Birth Defect Children (ABDC), under the direction of Betty Mekdici, actively solicits reporting of birth defects both in the general population and exposed cohorts. As part of this activity, 18 cases of Goldenhar Syndrome were identified in veterans. The father was the veteran in all 18; 15 were deployed to the PGW and three were not. Two of the latter veterans were vaccinated, and the remaining non-deployed veteran handled gear from the PGW and has been diagnosed with Gulf War Syndrome. Since this registry is more likely to obtain case referrals from exposed veterans, it cannot be assumed to include a representative sample of unexposed cases. These cases are summarized and compared to the seven cases reported by Araneta *et al* in Table 3. There are several ways in which these two data sources differed. These differences include; status of veteran (active versus active or separated), study period (three years versus seven), source of ascertainment (all hospitals versus military only), child’s age at diagnosis (diagnosis of Goldenhar or related diagnoses on the birth hospitalization records versus diagnoses at any age). All of these differences, summarized in Table 4, would tend to increase the number of cases in the ABDC registry. Nevertheless, it seems important to ascertain whether all of the 15 exposed cases identified by the ABDC that should have been identified by the Araneta *et al* study protocol were not included among the five exposed cases identified in that study. In other words, were all ten additional exposed cases in the ABDC ineligible for the Araneta study? Conversely, were all five cases identified in Araneta *et al* also present in the ABDC registry? This issue is discussed below under Recommendations for Further Study.

B. Description of ongoing studies

I have identified eight ongoing studies that will provide information on the risk of adverse reproductive outcomes among PGW veterans. Of these, five studies are designed specifically to address reproductive health concerns. Three of these studies are being conducted by the Naval Health Research Center (NHRC); these are referred to as Studies 3,4, and 7. Study 3 examines a

range of birth outcomes; Study 4 examines fertility and miscarriage; Study 7 examines congenital anomalies. The California Birth Defects Monitoring Program is conducting a study to determine the feasibility of identifying birth defects to PGW veterans in this registry. To this end they will attempt to use Department of Defense (DOD) data files to link both active and separated veterans to military hospitals to ascertain birth defects in children of all PGW veterans. In addition, Bernstein (University of Cincinnati) is conducting a clinical study of Seminal Plasma Hypersensitivity Reaction. There are also three studies that examine a range of outcomes that include one or more reproductive endpoints. The Portland Environmental Hazards Research Center of the Oregon Health Sciences University (OHSU) is attempting to define and analyze health effects (including reproductive outcomes) associated with exposure to selected hazards pertinent to military service. The Veteran's Administration is using the National Health Survey to compare the health status of a random sample of PGW veterans and non-deployed veterans. The Klemm Analysis Group is comparing the health status of female PGW veterans and non-deployed female veterans and is examining a broad range of reproductive outcomes in these women.

Study 3: Naval Health Research Center

Study 3 is a comparative study of pregnancy outcomes among PGW veterans (male and female) and other active duty personnel. The papers of Cowan *et al* (1997) and Aranenta *et al* (1997) include results from this study. It is not known whether other analyses are forthcoming from Study 3. However, several of the outcomes of interest, according to the protocol, have not been reported to date. These include prematurity, ectopic and molar pregnancy, and complications of labor and delivery.

Study 4: Naval Health Research Center

Study 4 is examining differences between PGW veterans and non-deployed veterans with respect to infertility, time to conception and risk of miscarriage. This study is being conducted in two phases. Phase I began in spring 1996, when a questionnaire was mailed to a random sample of 16,000 couples (8,000 couples for which one or both deployed to the PGW and 8,000 for which neither deployed). Active duty, National Guard and Reservists are included. A second mailing was completed in fall 1996. Currently the participation rate is 46% and a third mailing is underway. Rate of participation is higher in active military than those who separated (personal communication, R. Calderon). Phase II will consist of a telephone interview of 5,000 couples to obtain detailed information on exposures and known risk factors for infertility and miscarriage. In April 1996 an additional 1,870 women who were pregnant before deployment, or became pregnant during deployment, were added to Phase I.

The following four categories of married couples are included: (1) woman served in the PGW; (2) man served in the PGW; (3) woman served in the military during the PGW, but not in the Gulf area; and (4) man served in the military during the PGW, but not in the Gulf area. The Phase I sample includes 4,000 couples from each of these four groups. No exposure information is collected in Phase I except for the field of operation. No Phase II questionnaire was available, so exposure information collected in that phase could not be evaluated.

Study 7: Naval Health Research Center

Study 7 is examining the prevalence of congenital anomalies in seven states that maintain an

active birth defects surveillance system including all birth defects diagnosed during the first year of life. These states include Arizona, Arkansas, Hawaii, Iowa and Oklahoma and parts of California and Georgia. In addition, still births (fetal deaths after 20 weeks gestation), recorded by the Vital Statistics Registries of these states will be ascertained. This study also proposes to compare rate of preterm birth, low birth weight and still birth between PGW veterans and non-deployed veterans in the seven states. The Department of Defense Manpower Data Center (DMDC) maintains a database that includes demographics, military and hospitalization data. These data will be matched with the Vital Statistics Registry of each state to identify live births and stillbirths. These will then be matched against the birth defect registry in each state. Overall and organ-specific rates will be calculated within each state for PGW veterans and non-deployed veterans and compared.

Births between 1989 and 1993 will be included in order to compare conceptions prior to, during, and after the PGW. As of November 1997, linkage is complete for Hawaii and Arizona and birth defect rates should be available for these states in December 1997 (personal communication, H. Araneta). Active surveillance in Georgia is only conducted in the Metropolitan Atlantic area. Data from California will only be available for 13 counties, accounting for about half the live births in the state. In addition, the California Birth defects Monitoring Program does not routinely access military hospitals. The feasibility of this ascertainment is being explored in a separate study (see below). The protocol estimates that approximately 9,500 live births to PGW veterans will be ascertained by this study (exclusive of Iowa). Power calculations are provided only for all birth defects combined, with an assumed rate to 3%. There is no discussion of exposure assessment in this protocol.

Feasibility study: California Birth Defects Monitoring Program

The goals of this study are to determine whether all participants in the PGW, including National Guard/Reservists, can be identified by linkage to vital statistics and to determine to what extent defects diagnosed in the first year of life can be ascertained by the California Birth Defects Monitoring Program (CBDMP). Specifically, the CBDMP will determine: (1) whether Department of Defense (DOD) data about births to active duty military personnel are sufficient to allow the CBDMP to locate medical records about these children during the first year of life; (2) whether hospital record review is possible at DOD facilities, particularly those which may be closed or have incomplete medical record information; (3) whether DOD information about structural congenital anomalies is sufficiently accurate, compared to complete hospital medical record information. To meet these three objectives, DOD will provide a file with identifying information on all children born to active duty PGW participants posted in California between May 1991 and 1994. From these CBDMP will select a random sample of 500 births (approximately 20%) to trace to military hospitals. Since CBDMP statute does not extend to military hospitals, the Department of Veterans' Affairs must secure access to these medical records.

The DOD estimates that from 1991-1994 about 60% of PGW participants were still in active duty. This study will determine if DOD information about the identity of inactive (separated) personnel can be linked to California vital records and CBDMP files, neither of which contains social security information. For this purpose, DOD will provide a random sample of 5,500 persons (about 20% of all inactive PGW veterans), including name and birth date of participant and spouse, address at time of entry into the military, address at separation from the military and

address on the IRS tax form. These data will be used to link the inactive veterans to California vital records and the CBDMP files.

Progress of this study has been delayed due to difficulties in obtaining the necessary waiver from the Army Human Subjects Research and Review Board. No data collection had occurred by November 20, 1997 (personal communication, Gary Shaw, CBDMP).

University of Cincinnati: Clinical study of Seminal Plasma Hypersensitivity Reaction

The most unusual reproductive tract abnormality reported by PGW veterans and their spouses is the "Gulf War Vaginal Burning Syndrome". In cases of this syndrome, which can be local or systemic, severe vaginal burning and pain is reported to occur immediately on contact with the spouse's seminal fluid. A disrupted and inflamed vaginal mucosa is typically seen. This study will determine whether this syndrome in PGW veterans is due to the same immune responses previously described for women in the general community who have been diagnosed with seminal plasma hypersensitivity. The first aim of this study is to identify cases of vaginal burning syndrome among spouses of PGW veterans. For this purpose the investigators propose to use a questionnaire which has been previously validated to identify localized and/or systemic seminal plasma hypersensitivity in the general population. Ten cases in which the husband is an exposed veteran will be selected. The timing of the onset of the syndrome will be determined. Clearly onset before PGW exposure cannot be causally related to exposure, and such cases will not be included. Ten unaffected spouses of exposed veterans will be selected for comparison.

The second goal of the study is to identify seminal plasma proteins involved in the pathogenesis of this syndrome in spouses of PGW veterans, to determine whether these are the same as the proteins identified in cases from the general population. For this purpose, five ejaculates, collected over five consecutive days, will be obtained and use to isolate seminal plasma proteins from each male participant. The seminal plasma will be analyzed for protein concentration using the Biorad Protein determination assay. Women will then be tested for sensitivity to these seminal proteins using skin prick tests.

The third study goal is to determine the effect of physical and chemical factors, related to PGW exposure, on human seminal plasma obtained from both PGW-exposed and non-exposed males. The investigators suggest that in cases of this syndrome, seminal plasma protein structures have been altered by an infectious agent, chemical agent or physical factor. To examine this hypothesis, protein fractions from seminal plasma proteins from PGW exposed and non-exposed males will be characterized before and after *in vitro* exposure to a range of PGW exposures including; pyridostigmine, pesticides used in the PGW, extreme heat, and vaccines used to immunize soldiers before combat in the PGW.

No progress reports are yet available since this study was initiated in 1997.

The Portland Environmental Hazards Research Center

The objects of this study is to identify risk factors for Persian Gulf War Unexplained Illness (PGWUI) in veterans from the northwest United States. The investigators seek to determine why some of the veterans who served in Desert Shield, Desert Storm and desert cleanup are healthy, while others have symptoms of PGWUI. For this purpose a population-based questionnaire is being mailed to a representative sample of veterans who were deployed to the PGW during the one-year period following August 1990. The sample was selected utilizing databases provided by the DOD. Sampling was conducted within strata; (1) pre-combat (Desert Shield) only; (2)

combat (Desert Storm) only; (3) post-combat (desert cleanup only); and (4) two or more of these. By using a sampling strategy specifically based on periods of deployment, the role of potential risk factors such as pyridostigmine bromide, special vaccinations and combat stress can be isolated and analyzed.

Respondents to the mailed survey will provide the study population for the clinical case-control phase of the study. In this phase, the nature and pattern of exposures in cases of PGWUI and controls will be compared. A total of 250 cases and controls will be recruited for clinical testing within four months of responding to the survey.

The survey instrument contains extremely detailed exposure information including: military service history, duties, rank, dates and locations in the Gulf and exposures during the PGW including: smoke, petroleum products, insecticides, sand, heat, vaccines, vectors, anti-chemical warfare agents, diet, water source, living conditions and other stressors including perception of danger. It also includes several questions on the reproductive health on the respondent and/or partner, including: infertility, menstrual function, pregnancy history, painful intercourse and other genital tract symptoms.

The questionnaire was developed after interviewing veterans about their exposures and symptoms. A sample of 150 veterans who did not list Oregon or Washington as their home at the time of deployment, and are therefore not eligible for this study, were used to pilot the questionnaire.

National Health Survey of Persian Gulf Veterans

The Veteran's Administration, is conducting a three-phase study which includes a range of reproductive endpoints. In Phase I, a mailed questionnaire was sent to a random sample of 15,000 PGW veterans and a control sample of 15,000 Gulf-era veterans. In this phase, which is now completed, the response rate was low, due primarily to difficulty in locating study subjects. Questionnaires were returned by only 60% of deployed and 40% of non-deployed. A complete pregnancy history was obtained at this time. Phase II is being conducted, in part because of this low response, and to validate a sample of responses using medical records. Phase II has two parts. In the first, 2,000 respondents among the deployed, and 2,000 among the non-deployed were contacted by phone to obtain permission to review medical records. Permission was sought to validate all self-reported office visits and hospitalizations. This part of the study is now complete and data are being abstracted. On the second part of this phase, a random sample of 8,000 non-respondents was selected (without regard to deployment status) in order to compare demographics, exposures and health outcomes between respondents and non-respondents. Approximately 4,200 of these phone interviews are now complete (53%). Once subjects are contacted by phone, 80% agree to be interviewed, but contact has been difficult because of the mobility of this young population.

In Phase III physical examinations will be conducted on 1,000 veterans randomly selected from each group (deployed and non-deployed) as well as their family members. Up to twenty medical centers will be utilized for this phase. In order to obtain a truly random sample, selected participants will be flown to the nearest participating center, if necessary. Each veteran and all immediate family members will receive a detailed medical and psychological examination that will require up to one and one-half days to complete. The medical history will cover all pregnancies and outcomes, including birth defects. In addition blood and urine samples will be

obtained. Serum hormones will be measured in males and females and serum archived for future analyses. The Phase III protocol has been approved and is awaiting funding.

Exposure information will be based on self-report. Detailed information is requested on 24 specific exposures (including: smoke from oil wells, depleted uranium, nerve gas, chemical protective gear and so on) and length of each exposure, as well as vaccinations. In addition, DOD files will be used to provide dates and locations of deployment.

Klemm Analysis Group Inc; Persian Gulf Women's Health Linkage Study

This two-year study is comparing the health status of 10,000 women who served in the PGW with 10,000 Gulf-era military women. For this purpose a 45-minute health survey was developed inquiring about symptoms and conditions including adverse reproductive outcomes such as infertility, pre-term births, still births and birth defects. Detailed information about each pregnancy is elicited. Adverse health effects or reproductive events will be validated (with subjects' permission) by medical record review. Questions will also be included to identify cases of Gulf War Vaginal Burning Syndrome. Exposure assessment includes a detailed set of questions on 33 specific exposures and vaccinations (similar to the NHS questionnaire). However, this group is unique in ascertaining length of exposure to these substances before, during and after the PGW. Study materials are currently under review by the DOD (Personal communication, Rebecca Klemm).

Section II: Evaluating completed and ongoing studies

A. General criteria for evaluating reproductive epidemiology studies

The following criteria are commonly used to evaluate epidemiologic studies which have been conducted in order to assess the strength of association between an exposure (such as lead, pesticides, vaccinations or other exposures experienced by PGW veterans) and an outcome (such as infertility, spontaneous abortion or other adverse reproductive outcome).

Study goals

Are the goals of the study clearly stated? In general, a study with a clearly defined and specific goal (or set of goals) has a far greater chance of success than a study with vague or unstated goals; "Any study with poorly conceptualized objectives is ipso facto a poor study" (Rothman, 1986, p. 77). For example, in the protocol for the study designed by Bernstein (University of Cincinnati) to compare proteins in the seminal fluid of a sample of PGW veterans and civilians diagnosed with human seminal plasma hypersensitivity the investigators have stated a well-defined goal.

Appropriateness of study design

Is the study design selected appropriate to address the goal(s) of the study? For example, while a cohort design may be appropriate to examine all birth defects (without attempting to examine specific and rare defects), in order to assess the association between PGW exposures and specific rare defects (such as Goldenhar Syndrome), it may be more appropriate to utilize a case-control design.

Statistical precision (power)

In addition to stating the study goal, the investigator must state the precision with which the

association under investigation will be estimated. That is, once the study is complete, and a measure of association, such as the relative risk, estimated, how precise does the investigator require that estimate to be? In particular, if little or no association is found for the sample observed (e.g. the observed relative risk is near 1.0) how certain does the investigator need to be that no association of public health concern has been missed? Or, stated in another way, is the estimate of the relative risk sufficiently precise to meet the goals of the study? Further, if an association is found between the exposure and the outcome, is the study able to address such important issues as the presence of a dose-response, or whether the association varies across subsets of the population?

Sample size is not sufficient to guarantee statistical precision (or power). In particular, if in a cohort study there are far more unexposed than exposed subjects, or in a case control study, far more controls than cases, the effective sample size will actually be determined by the smaller of the two groups being compared. Thus, an efficient study is one in which precision is maximized for a given total sample size.

Exposure assessment

An epidemiologic study measures the strength of association between a disease and an exposure. There are many levels on which that exposure can be characterized. I will use an example of a non-PGW exposure to illustrate this point. Suppose we suspect that some constituent of tap water is associated with an increased risk of spontaneous abortion. Assume further that, in fact, only women consuming large quantities of water that has been chlorinated and which is high in one particular chlorination by-product (dichlorobromomethane, or DCBM) are actually at increased risk of spontaneous abortion.

What are alternative definitions of exposure, and how would the choice of exposure definition affect the ability of the study to find a positive association? The least precise definition of exposure would define women who drank any tap water (rather than other drinks, including bottled water) as exposed. This study would measure the association very imprecisely and would be unlikely to find an association. Taking the exposure assessment to one higher level of precision, defining exposure as any consumption of chlorinated water might be able to detect a weak association, but would underestimate the strength of the association. Becoming more specific and defining exposure as consumption of water that is high in DCBM would provide a more precise estimate of the association. The next level of precision would quantify the daily intake of water (via questionnaire) and define exposure as consumption of large amounts (e.g. five or more glasses per day) of water high in DCBM. Still more precise exposure assessment would be obtained by restricting the exposure to the first trimester of pregnancy, the time period in which most spontaneous abortions occur. That is, how much tap water did the woman drink in early pregnancy? Was this water she drank at home? If so, one should also consider her water consumption at work and elsewhere. Which of these locations should be used to assign the woman's first trimester exposure to DCBM? How were the levels of chlorination by-products obtained? Most water utilities must report the level of (some) chlorination by-products to regulatory agents, but these are based on quarterly samples at a few points in a potentially large water system. Thus, these levels may not be strongly related to the levels of DCBM the woman actually consumes. Ideally one might want to sample water from her home (and office) taps and test this for DCBM. Finally, women may receive this compound, which is volatile, from other sources such as showering and bathing. Ideally then, one would want to obtain a biological

marker of exposure, using DCBM levels in serum or urine in early pregnancy to define exposure.

This discussion of the assessment of exposure to a specific water constituent is intended to convey the difficulty of conducting a precise exposure assessment. It is based on an actual series of analyses (Waller *et al*, Epidemiology, 1998).

As difficult as exposure assessment for a chlorination by-product may be, the issues of exposure assessment for PGW exposures are still more complex. Gulf War veterans were exposed to a wide range of reproductive toxicants, including lead, diesel fuel, pesticides, ionizing radiation and biological agents. The degree to which each study collected data that allows for precise assessment of exposure to one or more of these agents is discussed below. Exposure assessment issues for studies of PGW veterans are discussed in detail in the report of McDiarmid to this committee (November 1997).

Definition of study outcome(s)

In the above discussion of spontaneous abortion in relation exposure to a chlorination by-product, a single outcome was studied and it was well defined. This outcome was selected for study because an association had been suggested by prior studies. The selection of study outcomes is often suggested by case reports (e.g. diethylstilbestrol and genital tract cancer), as well as prior studies. The subject studied by Araneta *et al* (1997) was suggested by reports of Goldenhar Syndrome in association with PGW exposure. In addition, outcomes may be selected for study on the basis of toxicological concerns about a particular exposure, as is the case with PGW exposures. Just as the ability of a study to detect an association increases with the precision of the exposure assessment; it also increases with the precision with which the study outcome is specified. For example, the association between heavy prenatal alcohol consumption and the syndrome that became identified as fetal alcohol syndrome would not have been identified in a study which examined the association between alcohol consumption and all major congenital anomalies considered as a single outcome. Table 6 contains frequently studied reproductive outcomes in men, women and children.

Misclassification bias

The above discussion illustrates the importance of correctly assigning exposure and outcome status. The failure to correctly classify a subject in an epidemiological study results in misclassification and bias.

Exposure misclassification can occur in two ways. To make this concrete, suppose we are studying the association between exposure to lead in the PGW and the occurrence of spontaneous abortion. Misclassification with respect to exposure in the direction of a false negative would occur if the subject were classified as unexposed because he was not deployed, when, in fact, he was exposed to lead from a source other than the PGW. Conversely, misclassification with respect to exposure in the direction of a false positive would occur if the subject was classified as exposed (perhaps because he was deployed), when, in fact, his assignment did not bring him in contact with lead.

Similarly, the outcome under study can be misclassified. This occurs either when the study subject is designated as a case (for example, of Goldenhar Syndrome) when she does not meet the case definition, or conversely.

Misclassification of exposure or outcome is a serious concern for all epidemiologic studies, and particularly those for which the definitions of exposure and outcome are difficult, and

possibly subjective. If errors occur in classifying subjects with respect to exposure, and the chance of these errors is equal in cases and controls (termed nondifferential misclassification), these errors will not “cancel each other out”. Rather the results of the study will be biased in a predictable and potential serious, way. This kind of error makes it more difficult to detect a positive association (the bias “towards the null”) if it exists (Rothman, 1986, p.87). Similarly, if the classification of subjects as cases or controls (for example as cases of Goldenhar Syndrome or not) is equally in error for exposed and non-exposed, the result will be similarly be biased, so that detecting an association is more difficult. On the other hand, if exposed subjects are more likely to be classified as cases than unexposed subjects are, this can artificially inflate the estimate of the relative risk, resulting in a bias in the other direction, or an overestimate of the association.

Selection bias

The selection of subsets of veterans may bias study results or decrease statistical power. For example, the studies of Cowan *et al* and Araneta *et al* do not include National Guard/Reserve veterans. How might these exclusions affect study results? Data from the Iowa Persian Gulf Study Group (1997) suggest that the National Guard/Reserve veterans were more affected by their PGW exposures. The prevalence difference between exposed and unexposed was greater among National Guard/Reserve veterans than among the regular military veterans for sixteen of the eighteen symptoms and conditions reported in this study. In addition, National Guard/Reserve veterans included a higher proportion reporting exposure than did regular military veterans for nine of the eleven PGW agents reported in this study. Thus, ascertaining outcomes only in the regular military may underestimate both exposure and the strength of the association between PGW exposure and outcome. In addition, the failure to ascertain outcomes in a large proportion of the exposed population decreases the study’s statistical power, and thus its ability to detect positive associations. These limitations affect all studies that restrict the population under study to the regular military.

Failure to ascertain cases in civilian hospitals may also bias study results in the direction of “no effect”. If infants with serious medical conditions, such as major birth defects are transferred to (or delivered in) more specialized civilian hospitals, studies such as Cowan and Araneta, which ascertained cases only in military hospitals will underestimate the risk of these outcomes. Even though the birth is in a civilian hospital, the veteran will still be included as a study subject. When no birth defect is ascertained the child will be considered a non-case, resulting in a misclassification of outcome for this child. As discussed above, this will create a bias in the direction of “no effect” (“bias towards the null”). This will also decrease the statistical power of the study, since the number of cases available for study will be decreased.

Utilizing only computerized hospital discharge diagnoses to identify cases also leads to underascertainment of cases, and thus misclassification of outcome. In a study of US Army veterans it was estimated that 49% of birth defects (28% of major and 66% of minor defects) were missed when the computerized discharge diagnosis was used compared to use of the complete birth record (Calle and Khoury 199(1). Similarly, identifying birth defects and other adverse birth outcomes only through the birth record will lead to underascertainment of cases, and thus misclassification of outcome, relative to those diagnoses made within the first year of life.

Confounding

Risk factors for a disease, other than the exposure being studied, such as race, smoking or age, may differ between exposed and unexposed. These factors (confounders) may mask an association, or create one spuriously. For example, the significant association between exposure in female veterans and all birth defects, reported by Cowan *et al* (1997) was reduced and no longer significant after controlling for race, marital status and branch of the military. These factors were confounders of the association.

The Hill Criteria

There are several criteria that epidemiologists frequently employ to evaluate a body of epidemiologic evidence. While none of these are themselves necessary or sufficient to establish causation, as a group they provide guidelines to help assess whether an association is causal. These are known as the Hill Criteria, after Sir Bradford Hill, who suggested these, somewhat tentatively and with many caveats (Hill 1965). The first criterion is strength of the association (often measured by the relative risk). Hill argues that strong associations are less easily explained by bias and confounding. However, this is certainly not a necessary condition, since the magnitude of the relative risk depends, in part, on how common the disease is in the unexposed population.

The second criterion is consistency, referring to the repeated observation of the association in different populations under different conditions. However, lack of consistency cannot rule out causation, since no studies will be exactly replicated. The third criterion is specificity, requiring that a cause result in a specific effect. This criterion is seldom met in epidemiologic studies; for example, cigarette smoking causes respiratory disease and heart disease, in addition to lung and other cancers. The fourth criterion, temporality, requiring that the cause precede the disease is seldom disputed. The fifth criterion, biologic gradient refers to the presence of a dose response curve, or some other gradient (such as the relationship between the timing of an exposure and the outcome in reproductive studies). This criterion is useful and one which is frequently sought. For example, Cowan explored the relationship between birth defect rates and time spent in the theatre. However, the dose response may non-linear, or not even detectable with available technology. The sixth criterion, biologic plausibility, is an important concern. For example, occupational studies demonstrating that lead and ionizing radiation are reproductive toxicants increase the plausibility of an association between these exposures and reproductive outcomes in PGW veterans. The seventh criterion, coherence, refers to the consistency of the association with the known natural history and biology of the disease under study. For example, the fact that seminal plasma hyperreactivity syndrome may be initiated by allergic reactions to other chemical stimuli is coherent with an association with chemical exposure in the PGW. Experimental evidence, the eighth criterion, is difficult to obtain in humans. However, animal studies can provide such evidence, as can cessation of symptoms upon removal of a subject from exposure. The final criterion, analogy, is difficult to apply. However, it is one rationale for looking for structural similarity between a known chemical agent and the chemical agent under study.

B. Evaluation of current studies

Exposure assessment

The evaluation of risks to the reproductive health of PGW-exposed veterans is made extremely difficult because neither the exposure nor the outcomes have been well specified. Most of the

studies conducted to date attempt to evaluate a huge range of exposures, chemical, biological, physical and psychological, by subsuming these all under the classification “PGW–deployed” versus “non-deployed”. Quantifying risk in terms of days in the theatre, or time since return from the theatre, does not avoid this difficulty. The latter strategy merely looks for a dose-response to the umbrella exposure “PGW-deployed”. Thus, the studies that do not have the capability of evaluating exposure to specific agents (and quantifying these) are likely to grossly underestimate any risk associated with a specific exposure, unless that exposure is highly correlated with the overall exposure “PGW-deployed”. To give a hypothetical example, assume that anthrax vaccine were the agent causally related to seminal plasma hypersensitivity, and assume further that only veterans deployed in a narrow time window, in a single branch of the service, received this vaccination. In this case, the majority of deployed veterans would actually be unexposed to the relevant agent, though classified as exposed using “PGW-deployed” to define exposure. The result of this misclassification could be a severe underestimation of the relative risk (see Misclassification above). In my opinion, this is the most severe shortcoming of these studies, and one that can only be remedied by the collection of additional personal exposure information. Less serious, but also important, is the fact that for most of these studies, deployment status was assigned on the basis of the unit to which the veteran was assigned. In this case, every person in a deployed unit was classified as deployed and conversely. Apparently 7% of males and a higher percentage of females were misclassified, even with respect to deployment status (personal communication, H. Araneta). Along the same lines, the study of Stretch *et al* (1995) noted that of those deployed, only 88% were actually deployed to the PGW, the remainder deployed to other locations. This too would misclassify subjects with respect to exposure.

Ascertainment of birth defects

The studies which have (or are currently) examining birth defects in relation to PGW exposure suffer from outcome misclassification, as well as the exposure misclassification just discussed. As discussed above, under outcome misclassification, the following factors lead to an underascertainment of cases, and subsequent misclassification and bias.

Excluding births in civilian hospitals (32% of all births were not in military hospitals)

Excluding children born after their parents left active duty (43% of veterans were not on active duty at the end of the study period in Cowan *et al* and Araneta *et al*)

Excluding children born to National Guard/Reserve personnel, those whom in the Iowa Study at least, were more severely affected and more heavily exposed.

Including only those birth defects identified through birth records (may miss 49% of birth defects)

Failure to include abnormalities in fetuses or still births (including those identified by prenatal screening)

Linkage studies unable to identify births for whom father is not included on the birth certificate

Collapsing birth defect categories

Studying all birth defects as a single outcome category is unlikely to identify an association, unless it is extremely strong. This is because a teratogen is unlike to increase the risk of all birth defects. Even thalidomide, perhaps the most potent teratogen known, only increased the risk of isolated rare defects (limb reductions, ear anomalies, heart anomalies). A rare birth defect

occurring, for example, in one in 10,000 births will account for only 2-3% of all birth defects. Therefore, even if its risk is significantly increased by an exposure, this could be missed by a study that collapsed all birth defect categories into one outcome. For example, the study of Araneta *et al* is of particular interest because it focuses carefully on a single rare defect. Let us assume, hypothetically, that the observed association of a three-fold increase in exposed veterans was found in a larger sample, so that it was statistically significant. Further assume that no association exists for any classes of more frequent birth defects. A simple calculation shows that the three-fold increase in risk of Goldenhar Syndrome observed by Araneta *et al* (from 5 per 100,000 to 15 per 100,000 live births) would not be detectable in an analysis that combines all birth defects, all major birth defects or even large groups of less common defects. Therefore, the apparently “negative” findings of Penman *et al* and Cowan *et al* would not be inconsistent with this hypothetical increase (which is, in fact, not unreasonable given the data in Araneta *et al*).

Statistical Power

Related to the issue of collapsing birth defect categories, these studies have only limited statistical power to examine rare birth defects or rare outcomes. The power calculations that were available only estimated power for all birth defects or all major birth defects (as a single class). Thus, none of these studies can, properly, be considered “negative”, since they were not able to rule out excess risks of public health concern.

Section III. Recommendations for further study

A. Improved exposure assessment

Most of the studies of the reproductive health of PGW veterans conducted to date include only limited exposure assessment. The most notable exception is the Oregon Health Sciences University Study (OHSU), which can be taken as a model for this purpose. The birth defect studies are particularly weak in this respect, with the exception of the Iowa study, which contains a fairly extensive exposure component. This classification utilizes dates the soldier was in the field, on the assumption that different deployment times correspond to different tasks, and thus different exposures. The extent to which this occurred must, of course, be checked to the extent possible, utilizing available records. I recommend that a nested-case-control study be imbedded in Study 7, and a detailed exposure assessment be conducted. If the OHSU questionnaire appears to assign exposures that correspond to historical records, perhaps this instrument could be used for consistency and later comparison across studies. I also recommend that duration of exposure to specific agents be asked with respect to the three time periods before, during and after the PGW (as in the Klemm Analysis group questionnaire).

B. Estimating exposure misclassification

It is also possible to assess the degree of exposure misclassification in, for example, the Iowa or Cowan studies by recontacting a sample of respondents, administering the OHSU questionnaire and comparing exposure classifications. Conversely, it would be possible to “collapse” categories from the OHSU exposure assessment questionnaire and compare relative risk estimates for a broad exposure (deployed-nondeployed) to those obtained for specific exposures. This should be most informative for exposures for which an association is observed using the more detailed questionnaire.

C. Further study of Goldenhar Syndrome

The Araneta study documents an increased risk of Goldenhar Syndrome among potentially exposed veterans. However, this increase is not statistically significant, possibly due to small numbers. Therefore, I recommend expanding this study, both to obtain additional cases and to improve the exposure assessment. To this end I recommend first evaluating the possible additional cases of Goldenhar Syndrome which have been identified by the ABDC registry. It should be determined whether any of the 15 exposed cases identified by the ABDC that should have been identified by the Araneta *et al* study protocol were not included among the five exposed cases identified. In other words, were all ten additional exposed cases in the ABDC ineligible for the Araneta study? Conversely, were all five cases identified in Araneta *et al*? It is also recommended that systematic case ascertainment for Goldenhar Syndrome be expanded in both deployed and nondeployed veterans, including births to separated personnel and all births to veterans in civilian hospitals. Ascertainment throughout the first year of life, using the full medical records would be ideal. In addition, it is important to obtain detailed exposure information on all cases and a sample of controls, perhaps using the Oregon Health Sciences’ questionnaire to obtain exposure information. It is also important to determine whether the cases of Goldenhar were the first live births born to veterans post-deployment. A causal relationship between this syndrome and birth after one or more healthy babies seems unlikely.

D. Targeted reproductive follow-up for cases of Persian Gulf War Unexplained Illness

The Oregon Health Sciences’ University has provided a tentative definition of Persian Gulf War Unexplained Illness (PGWUI), and is ascertaining cases of PGWUI in the Northwest. Since it is still uncertain what exposures are most relevant for reproductive illness in PGW veterans, I recommend looking for an increased incidence of reproductive abnormalities in cases of PGWUI. It is plausible that these veterans, most affected systemically by these exposures, would also exhibit more reproductive dysfunction in connection with PGW exposures. This reproductive assessment should be as complete as possible and should include serum hormone analyses on cases of PGWUI in the Northwest cohort. In addition, it would be valuable to examine semen quality in male cases. To date none of these studies has examined semen quality of veterans. Females could be asked to maintain a detailed diary recording menstruation, frequency of intercourse and use of contraception that would allow for a precise analysis of time to conception. If daily urine samples were obtained as well, assays would provide information on early fetal loss. (See Tier II analyses, Table 6).

E. Estimating misclassification in birth defect studies

Several sources of misclassification in the birth defect studies conducted or underway are listed

above. I recommend that the magnitude of the resulting misclassification be estimated using a sample of births from Study 7. This analysis would probably have to be limited to the five states that have active birth defect surveillance for infants up to one year of age throughout the state (thus excluding Californian and Georgia). This could be done by obtaining as complete an ascertainment of birth defects as possible on the selected sample, and then determining how many of these birth defects would have been missed if ascertainment had included; (1) birth record only; (2) military hospitals only; (3) active-duty personnel only. The degree of underreporting from these restrictions could then be examined as a function of severity of the defect and other covariates.

F. Estimating delay to conception

Delayed time to conception is thought to be the most sensitive measure of subfertility. This is not being addressed by any study except the Klemm Analysis Group. However, this questionnaire only asks whether or not the time till conception was longer than six months, and does not ask the length of time. I recommend that this important variable be added.

References

Agency for Toxic Substance and Disease Registry, Standardized assessment of birth defects and reproductive disorders in environmental health field studies. (Ed. G. Terracciano, GK Lemasters, RW Amler), 1996, NTIS (Publication number PB96-199609), Springfield VA.

Araneta MRG, Moore CA, Olney RS, Edmonds L, Karcher J, McDonough C, Hiliopoulos K, Schlangen K, Gray GC. Goldenhar Syndrome among infants born in military hospitals to Persian Gulf War veterans. *7*; 1997;56:244-251.

Calle EE, Khoury MJ. Completeness of the discharge diagnoses as a measure of birth defects recorded in the hospital birth record. *AJE* 1991;134:69-77.

Cowan DN, DeFraités RF, Gray GC, Goldenbaum MB, Wishik SM. The risk of birth defects among children of Persian Gulf War veterans. *New Engl J Med*. 1997;336:1650-6.

Hill AB. The environment and disease: Association or causation? *Proc R Soc Med* 1965;58:295-300.

Penman AD, Tarver RS. No evidence of increase in birth defects and health problems among children born to Persian Gulf War Veterans in Mississippi. *Military Medicine* 1996;161:1-6.

Rothman K.J. Modern Epidemiology. 1986, Little Brown and Co. Boston/Toronto.

The Iowa Persian Gulf Study Group. Self-reported illness and health status among Gulf War veterans. *JAMA* 1997;277:238-245.

Waller K, Swan SH, DeLorens G. Spontaneous abortion in relation to exposure to trihalomethanes in drinking water. *Epidemiology* (In press. To appear March 1998).

Tables

Table 1: Completed epidemiological studies of the reproductive health of PGW veterans

Author (year)	Exposed cohort (N)	Control cohort (N)	Response rate	Study area(s)	Reproductive outcome
Stretch (1995)	Deployed (active duty and reserve) (N=1,739) ⁽¹⁾	Non-deployed (2,512)	31%	Pennsylvania Hawaii	Menstrual difficulties
⁽¹⁾ 1,524 deployed to Persian Gulf					
Penman (1996)	2 National Guard units: (282 veterans with 54 births conceived post-deployment)	No internal control	90% ⁽¹⁾	Jackson, Mississippi	Major/minor birth defects ⁽²⁾ Prematurity/low birthweight Hyperbilirubinemia Stillbirths
⁽¹⁾ Phone interview					
⁽²⁾ Birth defect diagnosed during the first year life (defined by MACDP)					
Iowa Persian Gulf Study Group (1997)	PGW regular military and PGW National Guard/Reserve (n=2,421) ⁽¹⁾	Non-PGW regular military and non-PGW National Guard/Reserve (n=2,465) ⁽¹⁾	76% of eligible; 91% of contacted	IOWA	Symptoms of sexual discomfort
⁽¹⁾ Stratified random sample of those who served between August 2, 1990 and July 31, 1991					

Author (year)	Exposed cohort (N)	Control cohort (N)	Response rate	Study area(s)	Reproductive outcome
Cowan (1997)	33,998 births to PGW veterans ^(1,2,3)	41,463 births to non-deployed veterans ^(1,2,4)	Not relevant	USA	Major/minor birth defects ⁽⁵⁾ 7 common severe defects Live birth rate; sex ratio

⁽¹⁾ Includes active duty only; excludes reservists and National Guard

⁽²⁾ Births only in military hospitals. (Includes about 68% of all births to active duty personnel).

⁽³⁾ Births before 10/1/93 with estimated date of conception post-deployment

⁽⁴⁾ Births before 10/1/93 with estimated date of conception after 12/31/90

⁽⁵⁾ Routinely collected and computerized hospital discharge data on all live births

Araneta (NVRC)	34,069 births to GWV ^(1,2,3)	41,345 births to NDV ^(1,2,4)	Not relevant	USA	Goldenhar Syndrome ⁽⁵⁾
----------------	---	---	--------------	-----	-----------------------------------

⁽¹⁾ Includes active duty only; excludes National Guard/Reservists

⁽²⁾ Births only in military hospitals. (Includes about 68% of all births to active duty personnel).

⁽³⁾ Births before 10/1/93 with estimated date of conception post-deployment

⁽⁴⁾ Births before 10/1/93 with estimated date of conception after 12/31/90

⁽⁵⁾ Ascertained from birth records using 66 ICD-9 diagnostic codes and blinded record review

Table 2: Comparison of estimates of prevalence of Goldenhar from several sources

<u>Source</u>	<u>Years</u>	<u>Diagnosed</u>	<u>Cases/100,000 live births</u>
Hawaii Birth Defects Program (HBDP)	1988-1994	Up to 1 year	4.1
Metropolitan Atlanta Congenital Defect Program (MACDP)	1985-1994	Up to 1 year	5.2
California Birth Defects Monitoring Program (CBDMP)	1983-1991	Up to 1 year	6.8
Araneta <i>et al</i> PGW veterans	1991-1993	In birth record	4.8 (0.8-19.5)
Non-deployed	1991-1993	In birth record	14.7 (5.4-36.4)

Table 3: Comparison of Goldenhar cases in Araneta (1997) and ABDC Data Base

Birth hospital	PGW Veterans			Non-deployed-		
	Military	Civilian	Total	veterans Military	Civilian	Total
Araneta (1997)	4M, 1F	----	4M, 1F	1M, 1F	-----	1M, 1F
ABDC Data Base (as of 11/97)	6M, 1F	5M, 3F	11 M, 4 F	1F ^(1,2) , 1F ⁽²⁾	1F ⁽³⁾	3 F

⁽¹⁾ Born in Wurzburg Germany (assumed to be military hospital)

⁽²⁾ Father vaccinated

⁽³⁾ Father worked with contaminated gear; now ill with Gulf War Syndrome

Table 4: Differences between cases ascertained in Araneta *et al* (1997) and the ABDC (as of 12/1997)

<u>Factor</u>	<u>Araneta <i>et al</i> (1997)</u>	<u>ABDC (as of 12/97).</u>
Ascertainment period	1991-1993	1991-1997
Hospital of birth	Military	Military or civilian
Status at birth	Active	Active or separated
Exposure of ascertained cases	Gulf War and non-deployed	Gulf War or some Gulf-War-related exposure (e.g. vaccination)
Time of diagnosis	Related diagnoses on birth record	Any time in ascertainment period

Table 5: Epidemiological studies of the reproductive health of PGW veterans in progress

Author	Exposed cohort	Control cohort	Response rate	Study area(s)	Reproductive outcomes
Study 3 (NHRC) (Cowan, Araneta)	See Cowan above	See Cowan above	Not relevant	USA	Birth defects Prematurity Ectopic/molar pregnancy Complications of labor and delivery
Study 4 (NHRC) (Calderon)	Female PGW (deployed)	Female non-deployed veterans	46%	USA	Infertility Miscarriage
Study 7 (NHRC) (Araneta)	PGW veterans	Non-deployed	Not relevant	7 states with birth defect registries	Birth defects
Clinical study U. Cincinnati (Bernstein)	PGW veterans (male)	Non-deployed	Not available	Not stated	Seminal Plasma Hypersensitivity Reaction Reaction
Feasibility Study CBDMP (Harris)	PGW veterans	Non-deployed	Not relevant	California	Congenital anomalies (feasibility study)

National Health Survey (Kang)	Random sample of GWV (15,000)	Random sample of NDV (15,000)	50%	USA	Adverse reproductive outcomes in veterans and their families
U. Oregon (McCauley)	GWV deployed 8/90-8/91 (n=3,000)	Not applicable ¹	Not yet available	Oregon Washington	Infertility Menstruation Fetal loss Genital tract symptoms including "burning semen"

¹ All subjects were deployed. Four subgroups with differing exposures are defined: Desert Shield, Desert Storm, Desert clean up, mixed.
Within this cohort, cases of Persian Gulf War Unexplained Illness (PGWUI) will be defined.

Table 6: Suggested assessment methods for reproductive disorders, birth defects and developmental disorders ⁽¹⁾

Female reproduction	Outcome	Assessment methods	
		<u>Tier 1: Basic screening</u>	<u>Tier 2: Specialized studies</u>
	Menstrual dysfunction	Questionnaire, daily diary, basal body temperature	Salivary progesterone Urinary hormones (estrogen, progesterone, LH)
	Infertility/time to conception	Questionnaire, daily diary	Salivary progesterone Urinary hormones (estrogen, progesterone, LH)
	Pregnancy loss	Questionnaire (spontaneous abortion and still birth)	Urine hCG
	Lactation disorders	Questionnaire	Diary, obstetric and pediatric records
Male reproduction	Infertility/time to conception	Questionnaire	None
	Sexual dysfunction	Questionnaire/diary	Physiologic tests
	Hormonal disorders	None	Serum hormones (LH, FSH, Testosterone, Inhibin-B)
	Semen abnormalities	None	Semen quality (count, morphology, motility)
	Pregnancy/birth abnormalities	Questionnaire	None
Birth defects	Structural anomalies	Review of surveillance data Parent interview	Focused case finding/expert case review
	Genetic changes/mutagenesis	Parent interview Sex ratio from birth records	Focused case finding/expert review Genetic pedigree

484

Developmental	Fetal/infant mortality	Cause-specific review of mortality data	Autopsy and medical record review
	Fetal/neonatal growth (including low birth weight, prematurity, intra-uterine growth retardation)	Birth certificate review Parent interview	Gestational age, birth weight height and growth from delivery and medical records
	Functional deficits	Parent interview	Medical record review, school testing, Psychologic and neurologic test batteries

⁽¹⁾ Adapted from: the report of the Agency for Toxic Substance and Disease Registry, Standardized assessment of birth defects and reproductive disorders in environmental health field studies. (Ed. G. Terracciano, GK Lemasters, RW Amler).