The Impact of Environment and Occupation on the Health and Safety of Active Duty Air Force Members: Database Development and De-Identification

Roger Erich, PhD; Lt Col Melinda Eaton, USAF BSC*; Ryan Mayes, PhD†; Lamar Pierce, PhD‡; Andrew Knight, PhD‡; Paul Genovesi, MS*; James Escobar, MPH*; George Mychalczuk, MS*; Lt Col Monica Selent, USAF BSC**

ABSTRACT Preparing data for medical research can be challenging, detail oriented, and time consuming. Transcription errors, missing or nonsensical data, and records not applicable to the study population may hamper progress and, if unaddressed, can lead to erroneous conclusions. In addition, study data may be housed in multiple disparate databases and complex formats. Merging methods may be incomplete to obtain temporally synchronized data elements. We created a comprehensive database to explore the general hypothesis that environmental and occupational factors influence health outcomes and risk-taking behavior among active duty Air Force personnel. Several databases containing demographics, medical records, health survey responses, and safety incident reports were cleaned, validated, and linked to form a comprehensive, relational database. The final step involved removing and transforming personally identifiable information to form a Health Insurance Portability and Accountability Act compliant limited database. Initial data consisted of over 62.8 million records containing 221 variables. When completed, approximately 23.9 million clean and valid records with 214 variables remained. With a clean, robust database, future analysis aims to identify high-risk career fields for targeted interventions or uncover potential protective factors in low-risk career fields.

INTRODUCTION

We typically define occupational safety by potential environmental and occupational risk factors as well as incidents that occur in the workplace. However, job assignments and associated stress levels of service members can have major implications for safety off duty as well. Individual Air Force workplaces have been examined for environmental and occupational risks; however, a broad examination of all occupations and workplaces across the Air Force has never been conducted. The overall goal of this study is to investigate how environmental and occupational factors affect health outcomes and risk-taking behaviors among active duty Air Force (ADAF) personnel.

All Air Force installation workplaces are evaluated by an Aerospace Medicine team (typically bioenvironmental engineering, public health, and flight surgeons) to determine potential environmental and occupational exposures using nationally identified sources (e.g., Air Force Occupational Safety and Health standards, Occupational Safety and Health Administration expanded standards). Identified exposures lead to implementation and documentation of preventive measures for high-risk workplaces. Aerospace Medicine routinely visits each workplace to ensure utilization of preventative measures and investigates any potential mishaps or accidental exposures. Bioenvironmental Engineering and Public Health offices maintain a record for high-risk workplaces on their installation. In addition, the Air Force Safety Center receives occupationally related injury or illness reports. To date, researchers have not analyzed this information at an Air Force level. For example, the study team can investigate associations between large groups (e.g., location specific, functional units, etc.) and examine broad career field associations with health and safety outcomes.

Along with direct environmental and occupational hazards, military members report higher psychological strain than the general population[1] and significant work-related stress.[2] Stressors may indirectly manifest in service members' health and safety outcomes, such as in diseases (e.g., depression)[3] or injuries incurred while not directly performing official job duties. One potential explanation is altered risk-taking behaviors such as drug use,[4] smoking,[5] or medical issues such as poor diet[5] and obesity.[6] Although a service member's occupation can be a source of stress, it may also have the potential to protect against stress and resulting issues.[7]

Developing a comprehensive and valid database capturing all available demographic, medical, and safety information is essential in examining the link between occupational/environmental factors and health and safety outcomes. Table I provides a broad categorization of these important data elements. A larger, more integrative dataset enables greater generalizability

*Epidemiology Consult Service, Public Health and Preventive Medicine Department, U.S. Air Force School of Aerospace Medicine, 2510 5th Street, Building 840, Wright-Patterson AFB, OH 45433-7913.

†Aeromedical Research Department, U.S. Air Force School of Aerospace Medicine, 2510 5th Street, Building 840, Wright-Patterson AFB, OH 45433-7913.

‡Olin Business School, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO 63130.

doi: 10.7205/MILMED-D-15-00310

TABLE I. Database Content

Key Elements	
Subject Identification	Military Occupational Data
Time Component	Illness/Injury Occurences
Demographics	Health Survey Responses
Location and Workplace	

and allows examination of local contextual factors that might shape these dynamics. In this article, we describe the process to create and deidentify such a database that offers many research opportunities while protecting the subject's privacy. The Air Force Research Laboratory Institutional Review Board reviewed and approved the protocol for this study.

METHODS

Because of sensitivity and privacy concerns, the study team maintained these data at the U.S. Air Force School of Aerospace Medicine, Epidemiology Consult Service at Wright-Patterson Air Force Base (AFB) on restricted access computers. After deidentifying the database, it was made available to our Washington University in St. Louis research team members through a restrictive data use agreement.

The database building process consisted of several steps. The study team obtained source data from the Air Force Personnel Center (AFPC), the Defense Health Agency (DHA), the Air Force Reportable Events Surveillance System (AFRESS), the Preventative Health Assessment (PHA) data repository, and the Air Force Safety Center. The next step involved cleaning and validating these data followed by linking together the separate databases into one comprehensive, relational database. The final step entailed removing and transforming personally identifiable information to form a Health Insurance Portability and Accountability Act compliant, limited database.

Obtaining Source Data Access

The U.S. Air Force School of Aerospace Medicine, Epidemiology Consult Division (USAFSAM/PHR), has a memorandum of agreement with the data source organization to enable performance under Air Force Instruction and Air Force Medical Service (AFMS) contract. This agreement establishes dual responsibility to support and provide safeguards for the AFMS Military Health System Management Analysis and Reporting Tool Data Store shared between the data source organization and USAFSAM/PHR at a level commensurate with AFMS Information Assurance Program Office and Department of Defense requirements. Thus, source data access was in place before creating the comprehensive database.

USAFSAM/PHR and the data source organization utilized Microsoft SQL Server 2008 R2 to house/structure the databases on a server and a Microsoft SQL Server 2008 Management Studio table backup file to package data transfer files. As a part of the memorandum of agreement, a database administrator liaison at the data source organization assisted with the complex data queries and transfers through coordination with the USAFSAM/PHR data manager.

Source Data Description

AFPC source data consisted of 60 monthly total force snapshots taken on a particular date, labeled as AFPC Import Date (AFPC ID), near the end of each month during the study time frame (January 2006 to December 2010). Each monthly record set contained 25 variables including key identifiers such as Social Security Number (SSN) and date of birth (DOB), time components including date of rank and AFPC ID, and demographics containing rank, sex, race, educational status, and marital status. Air Force Specialty Code (AFSC) variables provide information regarding the subject's military occupation, and installation and unit of assignment provide location and workplace information. Over 30.7 million records were available with an average of over 513,000 records per month.

Data obtained from the Standard Ambulatory Data Record and Outpatient Network/Purchase Care databases, maintained by DHA, were combined to provide illness and injury occurrence information consisting of outpatient diagnoses for each medical treatment facility visit (military or civilian). This combined database contained 19 variables including SSN, date of care, as well as five levels of diagnosis codes in the International Classification of Diseases, 9th Revision (ICD-9) format. Over 19.4 million records with an average of over 324,000 records per month were available.

The Standard Inpatient Data Record and Inpatient Network/Purchase Care databases, which are also maintained by DHA, were combined to provide inpatient diagnoses with similar data structure and variables as the outpatient record database. These data consisted of approximately 50,000 records with an average of 764 records per month.

AFRESS data included key illness information (captured with ICD-9 codes) regarding reportable events such as sexually transmitted diseases, influenza, etc. Thirty-two variables were available, including SSN, case date of onset, date reported, ICD-9 diagnosis codes, as well as other relevant data. This database contained over 41,000 records with an average of 696 records per month.

Health survey response data from the PHA web survey were also included. Each active duty Airman is required to complete this annual survey that captured responses to questions involving the subject's current health and behavior. For this study, we obtained a subset of the complete survey containing question responses about alcohol/tobacco use and reproductive health. Variables included SSN, survey date, question number, and response number among others. Health-related variables contain answers to questions such as "Do you now smoke cigarettes?" and "How often do you have a drink containing alcohol?" The source data contained over 12.4 million records with an average of over 207,000 per month. Each of these records contained only one response to a particular survey question. Limited record counts were available in 2006 and 2007 because of initial, web-based PHA program

testing at several locations in 2006 and Air Force wide implementation in 2007.

The last set of data consisted of safety incident records as reported to the Air Force Safety Center, which included over 16,000 records housed in two separate databases; one for illness-related incidents and the other for injury-related incidents. These data included illness and injury incidents such as hearing loss, toxic fluid/material exposure, needle sticks, motor vehicle mishaps, and broken bones. Thirty-nine variables included SSN, date of incident, hazard type, severity, workdays affected, as well as other relevant data.

Data Preparation

The study team received all source data in text or Microsoft Excel files and imported these data into Microsoft Access for cleaning and organizing. The first step involved establishing the link from the AFPC demographic records to the other databases by pairing the SSN and AFPC ID (i.e., 60 unique monthly dates corresponding to the exact date the AFPC records were archived) identifying each unique, monthly AFPC record (i.e., a person-month record). We assigned an AFPC ID to all other study databases' records based on the applicable record date. This allowed linkage of the pertinent medical, safety, and survey information to the corresponding demographics. To avoid duplication of demographic and occupational data, we considered the AFPC information as the gold standard and removed all duplicate demographic data, besides SSN and AFPC ID, from the other six databases.

The cleaning process began with the AFPC data. Because these data contained total force records, the study team identified and removed Air National Guard and Air Force Reserve records to restrict the dataset to active duty subjects. We also removed Air Force Academy cadet records (143,227) and retiree records (577). Another issue involved inconsistent entry of installation names within the source files (e.g. "Kirtland" vs. "Kirtland AFB"). We resolved this problem by developing mapping/transition matrices to standardize values.

We discovered 62 SSNs with multiple assignments for sex. Review of first name, chronological history/progression through the import files, and results from a sex determination algorithm that assigned probabilities of being female corrected this issue. This algorithm relied on U.S. birth records,[8] which identify first names as either male or female for a given birth year.

Approximately 520 SSNs had multiple DOB assigned. Consequently, age calculations were inaccurate and this compromised valid identification of a unique individual. To resolve this issue, we used the latest DOB listed in the personnel data and, in some cases, cross-checked the DOB with other study databases. Race and ethnicity variables were also problematic (1,005,915 records associated with 20,393 SSNs affected for ethnicity and 562,734 records associated with 11,501 SSNs affected for race). As these subjects' records progressed through the study, race/ethnic codes would change one or more times. To remedy this issue, we used the latest race/ethnic code, for a particular subject, to recode all previous race/ethnic codes for that particular subject. This method assumed that the latest code had the highest probability of being correct, given that it was the most recently reviewed record. Another issue involved missing Primary AFSCs (PAFSC), Duty AFSCs (DAFSC), or both. If only one of these codes was available, the available code replaced the missing code (158,723 records). If both were missing, they both remained blank (127,618 records). We also identified and corrected other minor inconsistencies and errors which we do not mention for sake of brevity.

In addition to cleaning the AFPC database, the study team incorporated several enhancements to allow additional analysis options. To capture each unit's organizational structure, we created new variables to include Squadron, Group, Wing, Numbered Air Force, Other, and Major Command and linked them to each available AFPC record. We conducted extensive research using the Air Force Portal, Air Force unit fact sheets, and various other sources to obtain credible information regarding the unit's organizational structure from 2006 to 2010. For units that changed organizational structure during the study period, we added additional characters to the original unit name to identify the change and provide consistency when performing analysis.

The study team created a Unit Category variable to group similar units based on mission type. For example, the 391st Fighter Squadron was categorized as Fighter and the 366th Communications Squadron was categorized as Communications. Other categories include, but are not limited to Acquisition, Airlift, Civil Engineer, Intelligence Surveillance and Reconnaissance, and All Others. In all, there are 60 distinct unit categories. This variable creates fresh opportunities to examine the type of unit and the relationship with health and safety outcomes.

We also created a variable, based on the subject's duty status code (on leave, temporary duty, etc.) and/or duty title, identifying those who spent time as a prisoner, were under Security Forces custody, were being investigated by the Air Force Office of Special Investigation, or other undetermined legal, judicial, criminal, or punitive-type categories. All subjects meeting these criteria were flagged with the new code regardless of when the status occurred. Some subjects were coded as a prisoner or under legal investigation for the duration of the study. We removed these records since their suspected adverse actions occurred before the study time frame and they were subsequently discharged or retired before leaving confinement or under investigation status.

Other enhancements included calculating age from the difference between AFPC ID and DOB, determining time in grade from the difference between AFPC ID and date of rank, computing cumulative study time spent on a particular installation (permanent assignment situations), and generating a variable to capture marital status changes.

The study team also prepared data from the outpatient record database. As with the AFPC data, we removed Air National Guard and Air Force Reserve records as well as

TABLE II. Major Data Issues

Source Data	Data Issues
Personnel Records (AFPC)	• Presence of Guard, Reserve, Retired, and U.S. Air Force Academy Cadet Records • Subjects With Multiple Birth Dates, Ethnic Codes, Race Codes, and/or Sex Codes • Missing Data • Some Data Elements Not Standardized
Outpatient Medical Records (SADR and Outpatient Network/Purchase Care)	• Presence of Guard, Reserve, and Retiree Records • Subjects With Multiple Birth Dates and Sex Codes • Duplicate Records • Records With SSNs Not Present in Personnel Records
Inpatient Medical Records (SIDR and Inpatient Network/Purchase Care)	• Same Issues as Described in Outpatient Medical Records
Safety Incident Records (Reported Safety-Related Injury and Illness Events)	• SSNs Missing or Not Present in Personnel Records • Presence of Guard, Reserve, and Cadet Records • Duplicate Records in Both the Illness and Injury Data
Reportable Event Records (AFRESS)	• Presence of Guard, Reserve, Dependent, Retiree, Other Military Service, and Separated Personnel Records • SSNs Missing, Erroneous, or Not in Personnel Records • Duplicate Records
Health Survey Records (PHA)	• Presence of Guard, Reserve, Dependent, Retiree, Other Military Service, and Separated Personnel Records • SSNs Missing, Erroneous, or Not in Personnel Records • Multiple PHA Surveys per Calendar Year

SADR, Standard Ambulatory Data Record; SIDR, Standard Inpatient Data Record.

corrected multiple DOB and sex. To establish a common variable link to the AFPC data, we converted the beginning date of care for each record to the nearest future AFPC ID within the AFPC database, given that the respective SSNs matched. This ensured assignment of the most current demographics to medical diagnoses. Finally, the team removed duplicate and invalid records (discharged personnel, dependents, civilians, and personnel from other military branches). We prepared the inpatient record data in the same manner.

Safety incident data were cleaned and prepared by removing all nonactive duty members (Guard, Reserve, and cadets) and correcting or removing over 1,000 records with SSNs that were either never assigned (e.g., first three digits were 000, 666, or 900–999 or 00 in the fourth and fifth position), missing, or did not match AFPC records. To correct these SSNs, we cross-checked name and demographics with AFPC records. In addition, we eliminated 14 duplicate records present in both the safety-injury and safety-illness data. From the Safety-Injury data, we removed four 4 illness-related (heat exhaustion) records and from the Safety-Illness data, we removed 10 injury-related records. To link these records to AFPC, we used the mishap date to determine APFC ID assignment.

AFRESS database preparation involved removing non-active duty personnel, correcting or removing records with invalid SSNs, and revising over 700 records with multiple DOB or sex entries. To align the record with the actual event time, we used the case date of onset to determine AFPC ID assignment.

PHA data cleaning consisted of removing nonactive duty members and multiple PHAs present within the same calendar year (up to six PHAs per person in some cases). To handle the multiple PHA issue, we retained only the latest PHA per calendar year for each subject.

As a final step in data preparation, we thoroughly investigated medical, safety, and survey records with no associated AFPC demographic record matching their SSN and AFPC ID. For example, we used Defense Enrollment Eligibility Reporting System data to determine that nonmatching PHA records belonged to members of other service branches. This investigation ensured we did not omit important subgroup data.

All of the previously discussed data issues are generally not unique to this occupational and environmental health effects study. Whenever research is performed utilizing U.S. Air Force personnel records, medical and safety data, etc., errors and other data issues must be addressed before performing meaningful analysis. A brief recap of this study's major data issues is available in Table II.

Database Modifications and Linking

Since the AFPC database is the main focal point of all records, we created a randomly assigned link variable for each record. These random numbers uniquely identify all of the available SSN and AFPC ID combinations and allow efficient database modifications. By utilizing the link variable across the study databases, we will have a complete, relational database. Before assigning this link variable to the other six databases, we modified/restructured each one.

Originally, the PHA database format contained one question response value per record (up to 43 records for each completed PHA per subject and year). We combined all records with responses applicable to a distinct, completed PHA into one record, linked the PHA and AFPC databases together with the SSN/AFPC ID combination, and added the link variable to the PHA database.

TABLE III. Study Database Summary

Database	Initial Records	Final Records	Variables
AFPC	30,794,768	19,920,195	39
Outpatient	19,470,005	3,246,008	55
Inpatient	45,794	44,846	13
AFRESS	41,728	21,934	4
PHA	12,470,061	703,595	45
Safety-Injury	14,366	7,825	28
Safety-Illness	1,765	893	30

The outpatient record database required several modifications before assigning the link variable. The original outpatient record data structure consisted of one record per subject per visit with up to five medical diagnosis ICD-9 codes. There may be several visits per day and/or several visits per month for a study subject. To align these data with the monthly structure of the AFPC database, each day's worth of visits for a subject were combined into one record (duplicate ICD-9 codes were removed and there were up to 52 distinct ICD-9 codes per record). We assigned a time-ordered visit day number to each newly created subject's record to indicate multiple visit days within a given AFPC ID interval (time from the previous AFPC ID up to and including the next AFPC ID). Then we linked the outpatient record and AFPC databases together using the SSN/AFPC ID combination and added the link variable to the outpatient record database. These same procedures apply to the inpatient record database. Each transformed inpatient record contained up to 10 distinct ICD-9 codes.

AFRESS database preparation involved combining a subject's multiple records (up to two) within a particular AFPC ID interval. We created a variable to capture this second ICD-9 code, if applicable. Each modified record represents a subject's reportable event(s) within the AFPC ID interval. We linked the AFRESS and AFPC databases together with the SSN/AFPC ID combination and added the link variable to the AFRESS database.

For the safety databases, we added a variable to capture the number of mishaps or illness issues (maximum of two) for each subject within an AFPC ID interval. After this modification, we linked each safety database and the AFPC database together with the SSN/AFPC ID combination and added the link variable.

The final AFPC database structure contains records for each Airman corresponding to each month of active duty Air Force service. Each record links to a unique row in the databases, listed in Table III, as long as there is a corresponding link value. If there were any illness or injury episodes for a particular subject during the month, the AFPC demographic records attach to the applicable medical/safety records. The next step involved reducing the potential for disclosure of the subject's personal and medical information. To mitigate this issue, we created a limited database as described in the following section.

Creating a Limited Dataset

First, we eliminated unnecessary variables that aid in the subject's identification. Then, we masked several variables with random codes (i.e., replaced SSN with a 6-digit random code and renamed as Subject ID), combined certain demographic groups together (e.g., top two ranks for enlisted and top five ranks for officers), and converted data using a new format (i.e., converted date of rank to time in grade). Table IV displays the work completed for the AFPC database.

TABLE IV. AFPC Database Modifications

Variable	Modification
Rank	Collapsed General Officers and Colonels into a Single Category and CMSgt and SMSgt into a Single Category
Installation	Coded With Random 4-Digit Codes
Installation Kind	Coded With Random 3-Digit Codes
Unit	Coded With Random 4-Digit Codes
SSN	Coded With Random 6-Digit Codes and Renamed as Subject ID
3-Digit PAFSC and DAFSC	Coded Commanders, Generals, CMSAF, First Sergeants, etc. With 888 Identified as Command/Staff
PAFSC and DAFSC Skill Level	Converted G to 0 to Mask General Officers
Date of Rank	Converted This Date to Time in Grade
Aviation Service Code Date	Removed Variable

CMSgt, Chief Master Sergeant; SMSgt, Senior Master Sergeant; CMSAF, Chief Master Sergeant of the Air Force.

Combining the top-level ranks masked those subjects in high-level positions and easily identifiable because of the small number present in a particular unit. Some DAFSC and PAFSC designations directly identified individuals in particular command positions. Recoding these AFSCs was essential to preserve anonymity (skill level recoded for the same reason). We also removed or recoded specific dates to avoid identification. Another key identifier we masked was assignment location (installation). This information, combined with rank, AFSC, marital status, etc., could be identifiable. This is especially true for those assigned to smaller installations or installations with only a few individuals with certain rank and AFSC combinations. This same issue applied to Unit information. We cannot simply exclude installation and unit since they are potential key variables with respect to risk-taking behavior.

RESULTS

After completing the deidentification process, removing invalid records and unnecessary variables, we can perform research on our clean and limited dataset. Table III describes record and variable counts within the resulting databases.

The particular database design permits us to add or modify data as required. For example, we created a new variable allowing analysis of leadership structure and the effect on health and safety. For each subject, there may be an assigned rater who evaluated the subject's performance. We obtained these AFPC data, cleaned the records, and incorporated the information

TABLE V. Range of Injury and Illness Visit Days/Events per Subject During the 5-Year Period

Visit/Event Data Source	Minimum	Maximum
Outpatient	0	680
Inpatient	0	22
Reportable Event	0	7
Safety Incident Event	0	3

into a relational data table for linking to the appropriate AFPC records. When completed, we matched a rater, who was also a study subject, to over 16.6 million records (83.5% match).

There were 510,565 distinct Airmen in the study of which 206,354 (40%) were without an injury or illness visit day (distinct day the subject received care in a medical treatment facility). The range of injury or illness visit days/events for outpatient, inpatient, reportable event, and safety incidents at an individual subject level are shown in Table V.

DISCUSSION

This database gives researchers the ability to study subjects from various perspectives (e.g., career field, installation, age group, etc.) and observe effects on health and safety outcomes such as sexually transmitted diseases, mental disorders, and safety incidents. We can also track/analyze their health changes over time and identify potential issues for action.

One major challenge, when working with outpatient and inpatient medical records, involves determination of whether multiple medical visits are related to one or more injury or illness episodes. We structured medical encounter data to capture visits within a given month. Since the medical visit date was removed during deidentification, we relied on the APFC ID to determine the encounter month and year. This allows estimation of monthly time increments between multiple visits. Illness surveillance case definitions are routinely published by the Armed Forces Health Surveillance Center[9] and define time intervals that define incident cases. Knowing length of time between patient visits allows for differentiation between follow-up visits and incident illnesses. This methodology can also be used to determine incident injury cases.

Our primary focus centers on predicting positive and negative health and safety outcomes for ADAF members by employing regression analysis and longitudinal analysis. The study team will perform multivariable regression analysis to identify specific environmental and occupational exposures relating to higher occupational injury and illness rates. We will analyze regression results to establish/understand the associated factors. The study team will also conduct longitudinal analysis to estimate the likelihood of a negative event as a function of time invariant and changing occupation and environment. Finally, we will examine common variables between the regression and longitudinal models to refine identification of key risk factors. Air Force leadership may utilize these results to form policy recommendations aimed at reducing risk. For instance, if particular career fields are a common element in negative health outcomes, we might recommend screening or prevention programs targeted to those groups. If we find social or occupational patterns are associated with less risk-taking behaviors, a potential recommendation would be to promote these exposures in other groups.

The database will also provide insight into medical care utilization by ADAF members. We will examine direct outcomes involving occupational injury or illness and indirect outcomes, which may manifest as a result of risk-taking behavior and/or additional occupational stressors.

CONCLUSION

This database development allows us to build pathways toward occupation-related human performance improvement. We envision the intended outcome to be specific counseling and/or prevention program implementation geared toward reducing stress and stress-related outcomes within Air Force subgroups. Primary emphasis will be on occupational fields, although the database design allows for inclusion of other areas as well, and we will consider both garrison and deployed environments. The study team will use analysis results, made possible with this comprehensive database, to develop prevention strategies for presentation to Air Force leadership. This may enable Airmen to operate more efficiently and ensure full mission capabilities. We will route these prevention strategy recommendations through the Air Force Surgeon General's office upon study completion.

REFERENCES

1. Bridger RS, Brasher K, Dew A, Kilminster S: Occupational stress and strain in the Royal Navy 2007. Occup Med (Lond) 2008; 58(8): 534–9.
2. Pflanz SE, Sonnek S: Work stress in the military: prevalence, causes and relationship to emotional health. Mil Med 2002; 167(11): 877–82.
3. Mausner-Dorsch H, Eaton WW: Psychosocial work environment and depression: epidemiologic assessment of the demand-control model. Am J Public Health 2000; 90(11): 1765–70.
4. Reed PL, Storr CL, Anthony JC: Drug dependence enviromics: job strain in the work environment and risk of becoming drug-dependent. Am J Epidemiol 2006; 163(5): 404–11.
5. Hellerstedt WL, Jeffery RW: The association of job strain and health behaviours in men and women. Int J Epidemiol 1997; 26(3): 575–83.
6. Schulte PA, Wagner GR, Ostry A, et al: Work, obesity, and occupational safety and health. Am J Public Health 2007; 97(3): 428–36.
7. Grunberg L, Moore S, Anderson-Connolly R, Greenberg E: Work stress and self-reported alcohol use: the moderating role of escapist reasons for drinking. J Occup Health Psychol 1999; 4(1): 29–36.
8. Social Security Administration: Popular baby names: beyond the top 1000 names. Available at http://www.ssa.gov/oact/babynames/limits.html; accessed August 24, 2015.
9. Armed Forces Health Surveillance Center: Case definitions for data analysis and health records. Available at http://www.afhsc.mil/Home/CaseDefinitions; accessed September 4, 2015.