Paper 169-2013

SAS® Tools for Transparent and Reproducible Research: Medication History Estimator

SAS® Global Forum 2013

Brian C. Sauer, SLC Veterans Affairs Medical Center, Tao He, University of Utah, Jonathan R. Nebeker, SLC Veterans Affairs Medical Center

ABSTRACT

The Medication History Estimator (MHE) is designed to output data at the course-level; i.e., one row of data per drug course. Course and period proportion of days covered (PDC) are calculated for each medication. Reports that describe the frequency and percent of users for each medication product, average duration of medication courses, medication possession ratios and Kaplan-Meier based persistency curves are automatically generated and formatted for professional reports and journal publications.

INTRODUCTION

Epidemiological and health services research has been criticized as being unreliable.(1) Scientific evidence is strengthened when the study procedures of important findings are transparent, open for review, and easily reproduced by different investigators and in various settings.(2-4) Many research studies have common scientific workflows(5, 6) and general execution engines can be developed to reuse epidemiological software for specific clinical questions. Development of modular SAS programs that can be combined to produce epidemiological pipelines to automate components of the research process will support transparent and rapid response to Nationally important clinical questions.

Estimating each patient's medication exposure is a fundamental component of any study that aims to evaluate the safety or effectiveness of medication therapy in an observational setting. When investigators design studies that evaluate medication exposure or compare medications, they are confronted with a series of decisions concerning how to characterize treatment histories and classify treatment groups. Example decisions include whether to conduct an intention-to-treat analysis or to only evaluate outcomes during medication exposure. Researchers also must decide on criteria for identifying incident or new courses of drug therapy rather than established courses. The approach used to infer a patient's treatment history from medication orders or dispensing data influences cohort identification and treatment group classification. Nevertheless, descriptions of methods are typically not adequately explicit to replicate study procedures directly from the published narrative. This is due to the fact that journals govern the distribution of scientific findings while the task of distributing methods and protocol is customarily relegated to authors.(4)

To be compliant with the basic principles of transparency, reproducibility and reusability(2, 4, 7) - we developed a generalized approach to estimate medication histories that can be used to summarize medication exposure in a population and structure data for epidemiological evaluation. The approach is considered "generalized" because the program can be used to estimate medication histories for any drug therapy and it is flexible enough to allow a vast number of unique parameterizations. This paper describes the features of the SAS program and presents example output data structures and reports.

The *Medication History Estimator* (MHE) is designed to output data at the course-level – i.e., one row per drug course. A course and period proportion of days covered (PDC) is calculated for each medication. Reports that describe the frequency and percent of users for each medication product, average duration of medication courses, medication possession ratios and Kaplan-Meier based persistency curves are automatically generated and formatted for professional reports and journal publications.

MEDICATION HISTORY ESTIMATOR MODULE AND INPUT FILES

The MHE is intended to be a module within a workflow that executes pharmacoepidemiologic processes. The analytic module contains a specification file that allows the user to define the parameter settings of the program, a main program that user defines the file pathways and runs code, execution engine and user document. The MHE also has a user defined input file listing the medications to compute drug courses. To execute this module the user simply needs to specify the input parameters, list the medications of interest, set the file locations in the main program, then "run" the main program.

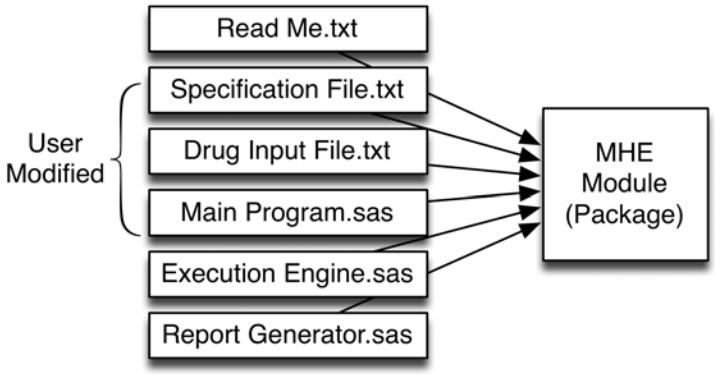

Figure 1 Components of MHE Module

MHE INPUT DATA STRUCTURE

The MHE requires two input data table:

1. A observation period table describing eligibility and enrollment data
2. A drug occurrence table (drug order, dispensing, or administration data).

The input data structures are based on the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM).(8, 9) Nevertheless, we have included the option of additional variables, such as drug strength, standardized drug units, drug name, and drug route. The MHE has a feature that computes prescribed and observed dose if drug dose is available and the user requests the output.

Table 1 Observation Table Structure

Variable	Description	Type
PERSON_ID	Unique patient identifier	INTEGER
OBSERVATION_PERIOD_START_DATE	Start date of observation period based on person history data from provider	DATE
OBSERVATION_PERIOD_END_DATE	End date of observation period based on person history data from provider	DATE

Table 2 Pharmacy Data Structure

Variables	Description	Type
PERSON_ID	Unique patient identifier	INTEGER
DRUG_CONCEPT_ID	Unique drug code used to map to standard drug information and concept hierarchy	INTEGER or CHAR
DRUG_EXPOSURE_START_DATE	The activity date (fill date, order date, service date, administration date)	DATE
DRUG_QUANTITY	The number of units ordered or dispensed	INTEGER
DAYS_SUPPLY	Intended duration the order or dispensing should last	INTEGER
DRUG_NAME*	Non-OMOP concept	CHAR
STRENGTH*	Drug dosage (5mg)	INTEGER
UNITS*	Standardized units (e.g., mg, mg/ML)	CHAR
DOSAGE_FORM*	Delivery method (e.g., tablet, injection)	CHAR
*non-OMOP variables		

MHE INPUT USER FILES

The MHE requires two input text files: the user specification file and the drug input file. The user specification file identifies the pharmacy and enrollment tables. The user also identifies the name of the drug list file, the variable name that drugs in the input file are assigned to. The user can specify the drug_concept_id, drug_name or other variables. Values in the drug_name variable can also be used to label drugs identified by the drug_concept_id. Additional features of the MHE are described in detail below.

Table 3 User Modified Specification File

	Description	**User Defined(default)**
1	Name of pharmacy table	*drug_exposure*
2	Enrollment table name	*observation_period*
3	Name of drug list file	*drug_list.txt*
4	Variable name that drugs in list will be assigned to in pharmacy table	*drug_category*
5	Variable type that drug in the list (0= number; 1=character)	*0*
6	Parent drug list or child drug list (0 = child; 1=parent)	*0*
7	Want to use a label for drug_concept_id (0= no; 1= yes)	*1*
8	Variable name with drug label	*drug_name*
9	Standardized dose information in the pharmacy table (0=no; 1=yes)	*1*
10	Variable name for dose strength if you have dose in the pharmacy table	*Strength*
11	Begin date of study period for extracting source data (enter as ddmmyyyy)	*01jan2002*
12	End date of study period for extracting source data (enter as ddmmyyyy)	*31dec2011*
13	Look back period to determine if patients enter study with drug	*0*
14	Extend the dispense day when early fill occurs (0=no, 1=yes)	*0*
15	Gap criteria, in days, that ends a drug course (value less than washout)	*30*
16	Gap criteria, in days, that identifies an incident drug course (washout period)	*90*
17	Cap the proportion of days covered calculation at 1? (0=no, 1=yes)	*0*
18	Run course-level output (0=no, 1=yes)	*1*
19	Need Report (0 = no; 1 = yes)	*1*
20	Full population in the report (default=0-->automatic calculation; or a number)	*10000*
21	Duration of followup time for persistency curves (0= default all days)	*365*
22	User defined tick on survival graph	*25*
23	Assume Period dates are same with enrollment date (0 = no; 1 = yes)	*0*
24	Report file name (rich file format)	*Course_Analysis_output*

Drug Input File

Even though the MHE is designed to operate using the OMOP CDM, we have designed it to be flexible enough to

read and execute parent drug concept unique identifiers (CUI) that aggregate children. The user can also simply put the child CUI or any type of text of numeric identification code in the drug input file and specify the variable name and type in the input user file. The user can also type drug names into the input drug file then identify the variable name in the input user file, i.e., specification file. Examples for each approach are listed below:

Parent/Child	Child CUI	Drug Category
098411=098409	098409	HYDROCODONE
098411=098408	098408	OXYCODONE
098411=098407	098407	ALPRAZOLAM

MHE FEATURES

This program is comprised of 5 macros listed below. The basic concepts of each macro are described. Interested users can download the MHE package and synthetic data to review code and run the module.

1. %Macro obtainDrug
2. %Macro extract
3. %Macro extendHandler
4. %Macro courseGenerator
5. %Macro doReport

%Macro obtainDrug: The obtainDrug macro reads the user defined drug input file into a temporary SAS table and takes the parent drug concept IDs and children drug concept IDs or drug names to query the full drug_occurance table.

%Macro extract: Figure 1 illustrates how the extract macro functions. The user selects the database, the begin and end dates for the study period, the lookback period, and the drug ancestor CUIs.

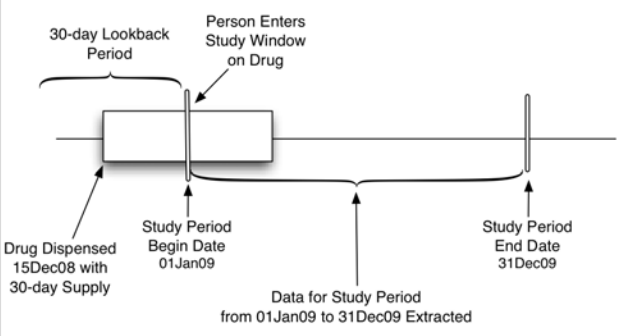

Figure 2 Illustration of the Lookback Feature Designed to Identify Exposure at Study Entry

%Macro extendHandler: The extendHandler macro is a feature that provides the option of extending the medication day supply to account for early refills. This macro operates on dispensing level data. If the extendHandler is set to 1 and an early refill occurs then the current fill is extended to account for the overlap in days supply from the early refill. Figure 3 below demonstrates how the extendHandler extends the second fill to 40-days to account for the refill that occurred 10-days early. This results in a 60-day drug course. If the extendHandler option is set to 0 then the early refill is ignored and the resulting course is only 50-days in duration. The extend handler also calculates the total dispensed dose and the average prescribed daily dose. The dose calculators are “experimental” they require the user to standardize dose and units for the dose calculation to work. Likewise, the MHE assumes the days supply is accurate. Considerable preprocessing is required for medications with atypical dosing schedules, such as methotrexate and Coumadin, and injectable and inhaled products. We have a pharmacy preprocessor that attempts to identify improbably days supply, units dispensed and daily dose. When identified the user can delete or implement several imputation strategies. Please contact us if you are interested in the pharmacy preprocessor module.

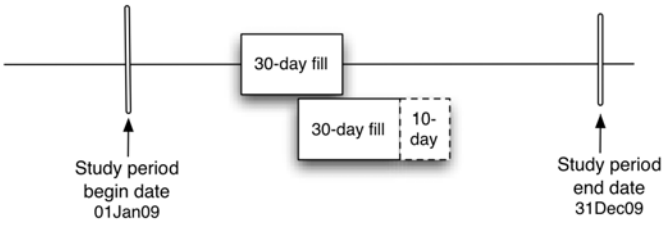

Figure 3 Demonstration of How Early Refills can be Extended to Account for Total Days Supply

The extendHandler macro also determines whether each dispensing meets the requirement for an incident course or not. The first course is determined to be incident if the patient had either a drug free period of user defined length (e.g., 90, 180, or 365-days) from the start of the study period if they were enrolled during this time (Figure 4). If patients were not enrolled at the start of the study period then determination of incidence is based on the drug free criteria from first enrollment during the study window (Figure 5). The patient in Figure 6 entered the study using the medication of interest. This person had a dispensing just prior to the study period as determined using the lookback period. Their first course was considered an established course since it is impossible to use data from within the study period as baseline information. If the user didn't include a lookback period then this patient would not have been identified as being exposed at study onset and the second course of therapy would have been identified as the first course. This person would be identified as having an incident first course rather than established.

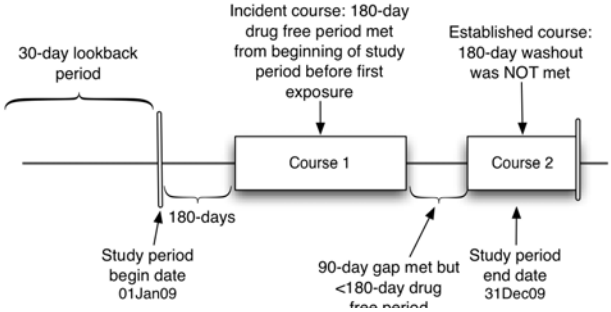

Figure 4 Inferring Incident Courses from Study Onset Based on the Drug Free Criteria of 180-days

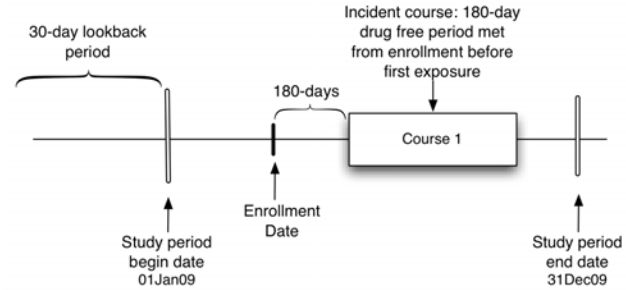

Figure 5 Inferring Incident Courses from Enrollment/Eligibility Data when Not Enrolled in the Health Plan at Study Onset

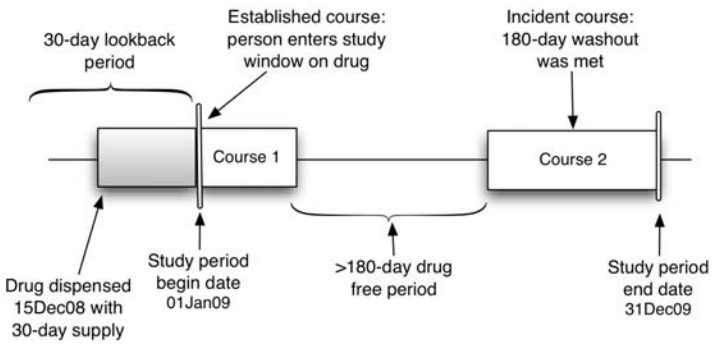

Figure 6 Demonstration of How the Lookback Feature Can Improve Assessment of Incident Courses

% Macro course: The course-level macro organizes the dispensing-level data into drug courses. This macro outputs data structured by drug course – meaning one row of data per person per drug course is generated. To achieve this, dispensing with gaps less than the gap criteria for ending a course are absorbed into the course estimation. Figure 7 shows that the gap criteria for ending a course was set to 90-days in the specification file. Gaps between fills of < 90-days would be absorbed into the course. The proportion of days covered (PDC) is computed to provide an estimate of adherence during the course, within 1-year of starting treatment and during the entire study period. This is the ideal structure for identifying potential "new users" or incident courses for cohort studies using the intention to treat analogue of a RCT. In this case the user would identify the start date for those meeting the inclusion criteria and follow them for the study duration regardless of whether or not the discontinued treatment. In an on-protocol analysis patients can be censored when treatment is discontinued but the analyst must be careful about time-varying confounding using this design.

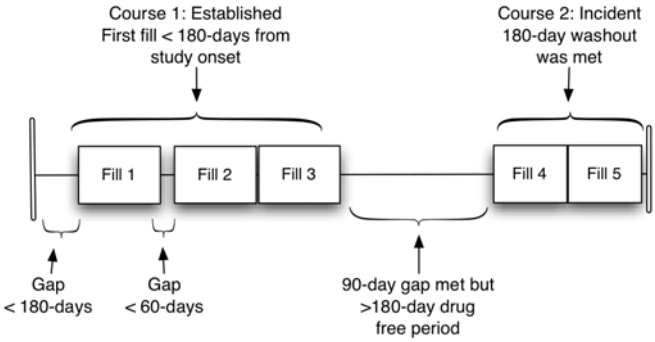

Figure 7 Illustration of How Courses are Constructed from Dispensing Level Data.

% doReport: The doReport macro generates reports describing how the medications were used at the population level. The course level analysis generates tables describing the number of people with an exposure to each medication of interest and the total number of drug courses for each medication. The fraction of the full population exposed to treatment is also presented in the first table. Tables listing the average course duration and standard deviation (SD), PDC by course, year, and study period are also presented. If requested, the report macro will generate a table describing cumulative as well as observed and prescribed daily doses in the population of interest. Kaplan Meier curves are presented to compare discontinuation rates for each treatment and by incident and established courses for each drug class.

MHE OUTPUT

The authors will provide the module and synthetic data upon request. The SAS® code and input files can be downloaded from the sascommunityorgwiki.[1] Briefly, we generated dispensing data for three controlled substances in a population of 10,000 patients. The controlled substances included Alprazolam, Hydrocodone, and Oxycodone. The use patters are similar to what would be expected from a distribution of acute, intermittent and chronic opioid users. The report macro generates table packages for all courses and just the first course. Please note that we are only

[1] http://www.sascommunity.org/wiki/SAS®_Tools_for_Transparent_and_Reproducible_Research:_Medication_History_Estimator

reporting the results from the first observed course for this paper.

Table 4 presents the number of people in the population (total population =10,000), which can be computed from unique people in pharmacy table or defined in the specification input file. Both the number of unique people from each drug class and total number of courses for each class are presented. The percent of the population is based on the number of unique people for each drug class and the total population value.

Table 4 Number of People and Courses in Each Drug Class (1st Course)

Drug Class	Number of People	Percent of Population*	Number of Courses
Alprazolam	5,104	51.040%	5,104
Hydrocodone	9,593	95.930%	9,593
Oxycodone	7,985	79.850%	7,985
TOTAL	9,991	99.910%	22,682
	*Total Population = 10,000		

Table 5 presents the number of courses that are considered incident and established based on the user defined washout criteria. The average course duration in days and standard deviation are also provided.

Table 5 Averages Course Duration and Standard Deviations by Drug Class and Incident Type (1st Course)

Drug Class	Incident Type	Number of Courses	Course Duration Average (days)	Course Duration Std. Dev. (days)
Alprazolam	Incident	3,830	107	248
Alprazolam	Established	1,274	316	609
Hydrocodone	Incident	5,300	50	175
Hydrocodone	Established	4,293	149	426
Oxycodone	Incident	6,835	53	188
Oxycodone	Established	1,150	87	297

Table 6 provides measures of adherence, which include an average course-level Proportion of Days Covered (PDC), as well, an average period and year level PDCs. The course-level begins and ends calculation of the PDC at the beginning and end of the course. The period level PDC calculates the proportion from the start of each course until the end of the study period. The year level PDC starts the calculation at the start of each course and ends after 365-days. The PDC is their observed days with treatment divided by the observation period (i.e., course duration, 1-year, days from course start date to the end of the study period).

$$PDC_i = \frac{\text{Observed treatment days}_i}{\text{Observation period}_i}, \text{ where } i \text{ represents unit (patient or course)}$$

$$\overline{PDC} = \frac{\sum_i PDC_i}{n}, \text{ where } n = \text{total units}$$

Table 6 Proportion of Days Covered

Drug Class	Incident Type	Course PDC Average	Course PDC Std. Dev.	Period PDC Average	Period PDC Std. Dev.	1st Year PDC Average	1st Year PDC Std. Dev.
Alprazolam	Incident	0.919	0.156	0.254	0.220	0.344	0.476
Alprazolam	Established	0.885	0.178	0.348	0.221	0.504	0.617
Hydrocodone	Incident	0.847	0.251	0.155	0.181	0.185	0.361
Hydrocodone	Established	0.807	0.262	0.239	0.205	0.341	0.555
Oxycodone	Incident	0.883	0.214	0.152	0.212	0.200	0.450
Oxycodone	Established	0.845	0.228	0.276	0.247	0.387	0.577

Table 7 presents average cumulative dose within a course, the observed daily dose and the prescribed daily dose. The cumulative dose is the sum of dispensed unit dose multiplied by the number of units dispensed. The observed daily dose is the cumulative course dose divided by the observed course days (early fills would make the observed dose higher on average and late fills would make it lower on average). The prescribed daily dose is cumulative course dose divided by the sum of the recorded days supply.

$$\text{CumDose}_i = \sum_i \text{Number of Units}_i \times \text{Unit Dose}_i$$

$$\text{Observed Daily Course Dose} = \frac{\sum_i \text{Number of Units}_i \times \text{Unit Dose}_i}{\text{Course Days}}$$

$$\text{Prescribed Daily Course Dose} = \frac{\sum_i \text{Number of Units}_i \times \text{Unit Dose}_i}{\text{Prescribed Days}}$$

Table 7 Dose Calculations by Drug Class and Incident Type (1st Course)

Drug Class	Incident Type	Cumulative Course Dose Average	Cumulative Dose Std. Dev.	Observed Daily Course Dose Average	Observed Daily Dose Std. Dev.	Prescribed Daily Course Dose Average	Prescribed Daily Dose Std. Dev.
Alprazolam	Incident	128.3	374.2	1.337	1.120	1.513	1.194
Alprazolam	Established	411.3	1027	1.301	1.192	1.513	1.347
Hydrocodone	Incident	960.7	3652	32.66	53.18	38.68	52.88
Hydrocodone	Established	3016	10101	36.34	458.0	43.37	457.8
Oxycodone	Incident	1838	10951	38.18	28.76	43.16	30.26
Oxycodone	Established	2374	14607	33.44	49.73	38.63	69.93

Table 8 provides counts for the number of courses that were classified as nonpersistent according to the user defined gap criteria. It also presents the numbers who were still exposed at the end of their enrollment or the end of the study period, i.e., censored. The data represented in table 8 is used to generate the Figure 5 Kaplan Meier curves. Persistency curves are generated for each medication and compared between incident and established courses. Medications are compared against each other by incident and established courses.

Table 8 Summary of Persistency Data (1st Course)

Drug Name	Incident Type	Total	Failed	Censored	Percent Censored
Alprazolam	Incident	3830	303	3527	92.09
Alprazolam	Established	1274	761	513	40.27
Hydrocodone	Incident	5300	1327	3973	74.96
Hydrocodone	Established	4293	2638	1655	38.55

Drug Name	Incident Type	Total	Failed	Censored	Percent Censored
Oxycodone	Incident	6835	734	6101	89.26
Oxycodone	Established	1150	759	391	34.00

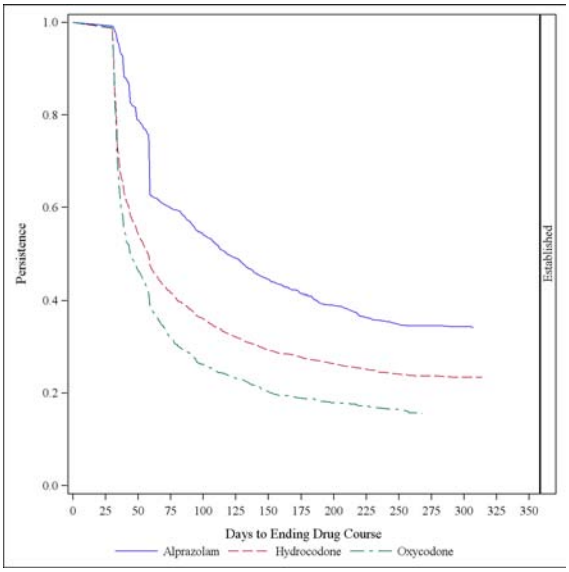

Figure 5 Persistency Curves Comparing Established Medication

CONCLUSION

The MHE is an efficient tool for estimating drug exposures from pharmacy data. The MHE generates reports that describe population-level medication use and it structures the data for further analysis. We believe that the development of transparent and reproducible tools for comparative effectiveness research will improve credibility of nonexperimental data from large healthcare data sources. We hope to establish a culture within the Veterans Affairs Health System where research teams are developing and sharing modules to that string together to execute epidemiological workflows. The MHE is just one such module in our knowledge repository.

REFERENCES

1. Taubes G. Epidemiology faces its limits. Science. 1995 Jul 14;269(5221):164-9. PubMed PMID: 7618077.
2. Peng RD. Reproducible research and Biostatistics. Biostatistics. 2009 Jul;10(3):405-8. PubMed PMID: 19535325. Epub 2009/06/19. eng.
3. Peng RD, Dominici F, Zeger SL. Reproducible epidemiologic research. Am J Epidemiol. 2006 May 1;163(9):783-9. PubMed PMID: 16510544. Epub 2006/03/03. eng.
4. Peng RD, Dominici F, Zeger SL. Reproducible epidemiologic research. American journal of epidemiology. 2006 May 1;163(9):783-9. PubMed PMID: 16510544. Epub 2006/03/03. eng.
5. Curcin V, Bottle A, Molokhia M, Millett C, Majeed A. Towards a scientific workflow methodology for primary care database studies. Stat Methods Med Res. 2010 Aug;19(4):378-93. PubMed PMID: 20442191. Epub 2010/05/06. eng.
6. Curcin V, Ghanem M, Guo Y. The design and implementation of a workflow analysis tool. Philos Transact A Math Phys Eng Sci. 2010 Sep 13;368(1926):4193-208. PubMed PMID: 20679131. Epub 2010/08/04. eng.
7. Peng RD. Reproducible research in computational science. Science. 2011 Dec 2;334(6060):1226-7. PubMed PMID: 22144613. Epub 2011/12/07. eng.
8. Stang PE, Ryan PB, Racoosin JA, Overhage JM, Hartzema AG, Reich C, et al. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. Annals of internal medicine. 2010 Nov 2;153(9):600-6. PubMed PMID: 21041580. Epub 2010/11/03. eng.
9. Reich C, Ryan PB, Stang PE, Rocca M. Evaluation of alternative standardized terminologies for medical conditions within a network of observational healthcare databases. Journal of biomedical informatics. 2012 Aug;45(4):689-96. PubMed PMID: 22683994.

ACKNOWLEDGMENTS

Funding was provided by the Veterans Affairs Health Services Research and Development Award (HIR-08-204). Project tile: VA Informatics and Computing Infrastructure (VINCI). Brian Sauer was also funded by the Veterans Affairs Health Services Research and Development Career Development Award (RCD 06-300-2).

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Brian C. Sauer
Enterprise: SLC VA Medical Center
Address: 500 Foothill Drive
City, State ZIP: SLC, UT 84148
Work Phone: 801 582-1565 x2469
E-mail: brian.sauer@gmail.com
Web: http://medicine.utah.edu/internalmedicine/epidemiology/research_programs/slc-eric.php