Journal of Population Therapeutics and Clinical Pharmacology
INCORPORATING FETAL ALCOHOL RESEARCH
Journal de la thérapeutique des populations et de la pharmacologie clinique

Original Research
DOI: 10.22374/1710-6222.24.3.6

DEVELOPMENT AND VALIDATION OF AN INDEX SCORE TO ADJUST FOR HEALTHY USER BIAS IN OBSERVATIONAL STUDIES

By Karly A. Achtymichuk,[1] Jeffrey A. Johnson,[2] Jashu K. Minhas-Sandhu,[2] Mu Lin,[2] Dean T. Eurich[2]

[1]Li Ka Shing, School of Public Health, University of Alberta, Edmonton, Canada. [2]ACHORD, Li Ka Shing Centre for Health Research Innovation, University of Alberta, Edmonton, Canada.

Corresponding author: Dean T. Eurich: deurich@ualberta.ca

Submitted: May 1, 2017. Accepted: November 13, 2017. Published: December 12, 2017.

Abstract

Objectives
To develop a healthy user index to serve as a method of confounding adjustment in future observational studies of preventive therapies.

Methods
A large administrative database of patients with type 2 diabetes was split in half randomly, yielding derivation and validation cohorts. Influenza vaccination was used as a 'prototypical marker' of a healthy user. In our derivation cohort, we fitted a mixed effects logistic regression model, and a points-based system was used to construct the index. The healthy user index was then evaluated in the validation cohort.

Results
Overall, 13% had received the influenza vaccination. In the derivation cohort (n = 914 732), the healthy user index ranged from 0 to 91 with a mean of 41.6 (SD 12.9). When applied to the validation cohort (n = 913 231), the index ranged from 0 to 96 (mean 41.6, SD 12.9) and significantly predicted influenza vaccination with a c-statistic of 0.649 (95% CI = 0.647-0.650).

Conclusion
Our healthy user index combined age, sex, and healthy behaviours to predict healthy users within administrative datasets. This index score may allow for better adjustment of healthy user bias in health services research; however, external validation is further required.

Many observational studies looking at the effectiveness of preventive therapies are prone to healthy user bias.[1–4] Healthy users tend to follow healthy behaviours (diet, exercise, cancer screening, vaccinations, etc.), adhere to their medications and preventive therapies, have higher functional status, and may be prescribed medications and therapies differently than their non-healthy counterparts.[1] The issue of healthy user bias (or healthy vaccinee bias) within observational studies of influenza vaccination are well recognized.[5–7] Although most people are recommended to receive the annual influenza vaccine to reduce their risk of complications from influenza and hospitalization,[8] only a subset of the eligible population is vaccinated. Indeed, patients who either seek out or are prescribed preventive therapies, like influenza vaccination, are often inherently different than their non-healthy counterparts. In the case of influenza vaccination, it has been shown that patients who receive the influenza vaccine tend to be less frail and overall healthier than their non-vaccinated counterparts.[9,10]

Although studies have been able to identify healthy user bias in some cases,[6,7] control for the bias within the analyses has been difficult. It is clear that better methods to control for confounding in these observational studies are needed, which may include using proxy measures (e.g., hormone therapy use has been studied as a marker of healthy users[11]) or predictive scores.[2] Predictive scores are particularly appealing as they combine a large amount of data into an overall score or index which can be used for adjustments in models (e.g., Charlson comorbidity index). However, no index or score has been developed for identification or adjustment of healthy users in research of preventive therapies that we are aware of.

We hypothesized that the healthy user bias can be captured, at least partially, in a prediction score, which can then be used to lessen the impact of confounding within observational studies of preventive therapies. Therefore, the objective of this study was to develop a healthy user index using influenza vaccination receipt as a 'prototypical' example of a healthy user, and to internally validate it in our population of adult patients with type 2 diabetes.

METHODS

Data Sources

We analyzed data from a large US claims and integrated laboratory database (Clinformatics™ Data Mart Database (OptumInsight, Eden Prairie, MN)) which has been used in numerous previous observational studies.[12–15] This database includes de-identified longitudinal data on patients, such as administrative and demographic information, medical service claims, laboratory data and pharmacy claims data. Clinical diagnoses are recorded as ICD-9-CM (International Classification of Diseases- 9th Revision-Clinical Modification) codes and procedure codes are recorded as ICD-9 and CPT-4 (Current Procedural Terminology- 4) codes.

Study Population

We developed our healthy user index in adults (aged 18 years and older) with type 2 diabetes who were part of the database between January 1, 2003 and Dec 31, 2011. Diabetes patients were chosen as influenza vaccination is universally recommended for this population.[16,17] Moreover, a recent systematic review on influenza vaccination in patients with diabetes has noted that the quality of the studies is low or very low due to major concerns of confounding due to the healthy user.[18,19] Indeed, influenza vaccination is considered one of the prototypical markers of preventative therapies used by healthy users and a large body of scientific work has already evaluated heathy user bias around receipt of influenza vaccinations use.[1–3,6,7,9,18–27] Thus, influenza vaccination was used as our prototypical marker of the healthy user, although other markers were also explored in additional analyses. Overall, within administrative data, coding for influenza vaccination has been shown to have high specificity (96–97%) and positive predictive value (88–91%), moderately high negative predictive value (74–79%), but lower sensitivity (50–56%).[28,29]

As influenza vaccination occurs on an annual cycle, we divided calendar time into years from July 1st to June 30th, as others have done.[6–30] Then, using US national surveillance data,[31] we defined our influenza season as a continuous period with the first to last occurrence of 50 positive isolates per week.[23]

One exception was in the year 2009, with the H1N1 outbreak, where the flu season did not terminate according to our above definition. In this scenario, we truncated the flu season on June 30th, 2009 (which is the end of our calendar year) and started the next flu season on July 1st, 2009.

Identification of Healthy Users

Using influenza vaccination as our proxy for the healthy user,[1–3,6,7,9,18–27] as no gold criterion exists in the literature, within each influenza year, receipt of the influenza vaccine was identified. Influenza vaccine receipt was determined based on Current Procedural Terminology/ Healthcare Common Procedure Coding System (CPT/HCPC) codes (4037F, 4274F, 90470, 90655-90664, 90666-90670, 90724, 90737, 9952, G0008, G8108, G8423, G9141, G9142, Q0034).[32] In addition, as pharmacists can dispense and administer influenza vaccinations, pharmacy codes for 'influenza virus vaccines' was also used to identify recipients of influenza vaccination.

Healthy User Predictor Variables

Predictors included variables that are readily available in many observational databases and that have been postulated to be associated with the healthy user[1–3,6,7,9,18–27] (Appendix 1). Specifically, variables included those identified within the US Medicare preventive services codes[3,32]: (A) cancer screening (including Pap test, pelvic exam, mammography, colorectal screening and prostate screening), (B) cardiovascular disease screening, (C) osteoporosis screening, and (D) medical nutrition therapy.[32–36] In addition, medication adherence (≥80% was considered "adherent" as per convention[3,37–39]) was included and assessed by the medication possession ratios (MPR). Other medications that have been shown or postulated to be related to the healthy user, including hormone replacement therapy,[11] smoking cessation therapy,[1,2] obesity medications as a marker of obesity,[40] statins[3,4] and bone resorption inhibitors,[41] were also considered.

Statistical Analysis

To develop our healthy user index, we first randomly divided our sample into 2 approximately equal sized cohorts: a derivation cohort and a validation cohort. The healthy user index was developed in the derivation cohort and then internally validated within the validation cohort.

To develop the healthy user index, a logistic regression model was used to predict yearly influenza vaccine receipt based on age, sex, and our healthy user predictor variables. If a patient received the influenza vaccine that year, their index date was the day of vaccination. If a patient did not receive the influenza vaccine that year, their index date was the last day of the influenza season as others have done.[6] Then all potential predictor variables were identified for each patient any time prior to their assigned index date for each season. Thus, all predictors and receipt of the influenza vaccine were updated on a yearly basis within the cohort.

We first built a parsimonious model in the derivation cohort, using multivariable mixed effects logistic regression to account for the clustered nature of the data (patients could contribute data for each year they were in the database). To facilitate analyses, certain variables were collapsed together if they measured the same underlying constructs (e.g., bone mineral density screening and bone resorption inhibitor medications). Age and sex were forced into the models and healthy user predictor variables with a $p<0.10$ in univariate analyses were entered into our multivariate model. Then, we further excluded healthy user predictor variables from the multivariate model with a $p>0.05$ (in sensitivity analyses we included all variables irrespective of p-values). The overall model discrimination was assessed with the c-statistic.

Second, a points-based system was then used to construct the healthy user index score, which has been extensively described previously.[42,43] These methods are similar to what has been completed in developing the Framingham risk score and mortality risk scores.[42,43] For each predictor retained in the final multivariate model, the estimated regression coefficient was divided by the estimated regression coefficient for age and then rounded to a single integer. Similar to the Framingham risk score and the mortality risk score, we chose age as our constant, which is the variable that determines the number of regression units per point in the scoring system.[42,43] The healthy user index for an individual was then constructed by summing the

following: age-18 years (as our data contained only patients 18 years of age and older), the component for the patient's sex, and the component for each of the predictor variables retained in the final model.

To assess the discrimination ability of the healthy user index, we calculated the healthy user index in our validation cohort based on the values obtained for each predictor variable identified in the derivation cohort. We then completed a univariate mixed effects logistic regression model with our healthy user index as the independent variable and receipt of influenza vaccination as the dependent variable. Overall model discrimination was assessed with the c-statistic. Statistical analysis was conducted with Stata version 12.1 (StataCorp LP, College Station, TX).

SENSITIVITY ANALYSES

To assess whether the healthy user index could be improved, we conducted several sensitivity analyses. First, a number of variables that have been mentioned in the literature as being associated with the healthy user were not included in our final model, such as dementia and having lab work (e.g., cholesterol, creatinine) completed.[23] Thus, we repeated our healthy user index development and forced these variables into the final multivariate model and reevaluated our score. Second, as our index score was developed by dividing each regression coefficient by a constant regression coefficient (age), we also developed a point-scoring system that assigned weights to the predictor variables, which did not account for age or sex. A weight was determined for each predictor variable by multiplying each regression coefficient by 10 and rounding to the nearest integer. A score was then computed by multiplying each predictor variable (1=present; 0=absent) by its estimated weight and summing. This approach is similar to the Charlson Score and the ADG score.[43,44] Last, we changed our 'marker' of the prototypical healthy user from receipt of influenza vaccination to receipt of statin therapies, as statin therapy has also been shown to be a marker of healthy users [4], and repeated the analysis.

RESULTS

Our population consisted of 1 827 963 patients aged 18–88 years. Mean age was 52.7 years (standard deviation (SD) 10.4) and 47.4% were female. Average length of follow-up was 5.5 years (SD 2.0 years). Average prevalence of influenza vaccination was 12.7%, with year over year receipt of vaccination ranging from 6.1% to 20.6% of patients. As expected, vaccination rates were highest in those >=65 years of age (range 16.8% to 21.4% year over year) and lowest in those <40 years of age (2.5% to 12.0% year over year). Our study sample was randomly divided into two approximately equal groups: a derivation cohort (n = 914 732) and a validation cohort (n = 913 231). As would be expected with a random split, characteristics between the two groups were similar (Table 1).

With respect to healthy user predictors, in the derivation cohort, the utilization of statins (39.7%) and cardiovascular screening (27.4%) were the most common, which is not unexpected given the underlying diagnosis of diabetes. Other predictors of the healthy user were less frequent, with hormone replacement therapy (6.0%) and cancer screening (3.8%) being the next most common. Overall, 37.7% of patients in the derivation cohort had a MPR ≥80% during the follow-up (see Table 1).

The scoring of the healthy user index is presented in Table 2. In the derivation cohort, the mean healthy user index was 41.6 (SD 12.9) and scores were normally distributed from a minimum of 0 to a maximum of 91. Points for each component of the healthy user index ranged from a low of 2 (female sex and average MPR ≥80%), to a high of 8 (medical nutrition therapy). Overall model fit, as measured by the c-statistic, in the derivation cohort was of 0.646 (95% CI 0.644-0.647). The estimated regression model was logit(P) = -5.51 + 0.062X; where P represents the probability of receiving the influenza vaccine during the particular flu season and X denotes the patient specific healthy user index score.

When the healthy user index was scored in the validation cohort, the mean healthy user index was 41.6 (SD 12.9), and scores were normally distributed from a minimum of 0 to a maximum of 96. When the healthy user index was regressed on receipt of influenza vaccination in the validation cohort, the c-statistic was 0.649 (95% CI 0.647-0.650), similar to that observed in the derivation cohort.

TABLE 1 Characteristics

	Overall	Derivation Cohort	Validation Cohort
Characteristics	*n* = 1 827 963	*n* = 914 732	*n* = 913 231
Mean age (SD), y	52.69 (10.37)	52.70 (10.37)	52.68 (10.37)
Female (%)	866 653 (47.41)	432 515 (47.28)	434 138 (47.54)
Vaccination			
Influenza Vaccination (%)	232 869 (12.74)	117 188 (12.81)	115 681 (12.67)
Screening			
Any Cancer Screening* (%)	69 378 (3.80)	34 608 (3.78)	34 770 (3.81)
Cardiovascular Disease Screening	500 919 (27.40)	250 854 (27.42)	250 065 (27.38)
Nutrition			
Medical Nutrition Therapy	2 097 (0.11)	1 036 (0.11)	1 061 (0.12)
Medications			
Average MPR>=80% (%)	688 768 (37.68)	344 550 (37.67)	344 218 (37.69)
Hormone replace therapy (%)	111 183 (6.08)	55 143 (6.03)	56 040 (6.14)
Smoking cessation therapy (%)	63 028 (3.45)	31 419 (3.43)	31 609 (3.46)
Obesity medications (%)	4 678 (0.26)	2 340 (0.26)	2 338 (0.26)
Statins (%)	725 051 (39.66)	362 649 (39.65)	362 402 (39.68)
Screening/Medication			
Osteoporosis screening and/or	53 220 (2.91)	26 469 (2.89)	26 751 (2.93)
Bone resorption inhibitors (%)			

MPR = medication possession ratios.
**Pap test, pelvic exam, mammography, colorectal screen, prostate screen.*

TABLE 2 The Healthy User Index: Point-Scoring System

Predictor Variable	Score
Age (for each year above 18 years old)	1
Female Sex	2
Any Cancer Screening*	4
Cardiovascular Disease Screening	6
Medical Nutrition Therapy	8
Average MPR >=80%	2
Hormone Replacement Therapy Prescription	4
Smoking Cessation Therapy Prescription	6
Obesity Medication Prescription	4
Statin Prescription	7
Osteoporosis Screening and/or Bone Resorption Inhibitor	6
Prescription	

MPR = medication possession ratios.
**Pap test, pelvic exam, mammography, colorectal screen, prostate screen*

These results were relatively consistent in both males and females. In the derivation cohort, the c-statistic for males was 0.645 (95% CI 0.643-0.648) and 0.649 (95% CI 0.646-0.651) among females with similar results in the validation cohort (0.648, 95% CI 0.646-0.651 and 0.651, 95% CI 0.648-0.653, respectively). Among those 65 years of age and under, the prediction in the derivation and validation cohorts were also similar to the overall model 0.644 (95% CI 0.642-0.645) and 0.647 (95% CI 0.645-0.648), respectively. Similar results were also observed in those over 65 years of age (0.588, 95% CI 0.583-0.593 in derivation cohort and 0.591, 95% CI 0.586-0.597 in validation cohort).

Sensitivity Analyses

First, inclusion of additional variables (e.g., dementia and having any routine lab work completed) had minimal influence on the discrimination ability of the healthy user index (<1% change in c-statistics in either derivation or validation cohorts). Second, utilization of a weight scoring system that did not account for age or sex also performed similarly. The weighted healthy user index is presented in Supplementary Table 3. Overall model fit as measured by the c-statistic was 0.654 (95%CI 0.652-0.655) in derivation cohort and 0.656 (95%CI 0.655-0.658) in the validation cohort, suggesting similar discrimination relative to our main model. Lastly, use of statin therapy as the dependent variable also performed similarly (Table 4). Overall model fit in the derivation cohort was 0.666 (95%CI 0.665-0.668) and 0.666 (95%CI 0.665-0.668 in the validation cohort).

TABLE 3 The Weighted Healthy User Index

Predictor Variable	**Score**
Any Cancer Screening*	2
Cardiovascular Disease Screening	3
Medical Nutrition Therapy	4
Average MPR >=80%	2
Hormone Replacement Therapy Prescription	2
Smoking Cessation Therapy Prescription	3
Obesity Medication Prescription	1
Statin Prescription	4
Osteoporosis Screening and/or Bone Resorption Inhibitor	5
Prescription	

MPR = medication possession ratios.
**Pap test, pelvic exam, mammography, colorectal screen, prostate screen.*

DISCUSSION

Healthy user bias is present in many observational studies of preventive therapies,[1–4] including influenza vaccination.[6] We built the healthy user index as a point-scoring prediction summary to assist in the control of healthy user bias in studies evaluating preventive medicines and therapies. Overall, the healthy user index score was shown to have moderate discriminating ability with respect to utilization of influenza vaccination, which we used as a prototypical marker of the healthy user. We anticipate this score could be used as a method of confounding adjustment in future observational studies of preventive therapies.

Although we have developed a summary score for controlling of healthy user bias in observational studies of preventive therapies, other approaches have also been suggested. One option is simply to adjust for the predictor variables we have identified separately. This may not be possible in all applications, as observations or outcomes may be low, which precludes the

TABLE 4 The Statin Healthy User Index: Point-Scoring System

Predictor Variable	Score
Age (for each year above 18 years old)	1
Male Sex	11
Any Cancer Screening*	1
Cardiovascular Disease Screening	11
Medical Nutrition Therapy	1
Average MPR >=80%	4
Hormone Replacement Therapy Prescription	4
Smoking Cessation Therapy Prescription	8
Obesity Medication Prescription	1
Osteoporosis Screening and/or Bone Resorption Inhibitor	9
Prescription	

MPR = medication possession ratios.
**Pap test, pelvic exam, mammography, colorectal screen, prostate screen.*

addition of a large number of variables into models. A second option is to adjust for a single variable that serves as a proxy for the healthy user, for example, use of vaccines, mammography or colonoscopy[2]. This approach may be too simple, as a single variable may only partially cover the many factors that are associated with healthy users. Lastly, another option is to build a propensity score around a proxy variable believed to represent healthy users. This was done by a group of researchers who built a propensity score around hormone therapy use.[11] A limitation behind this is that the propensity to use hormone therapy may not completely represent the propensity to be a healthy user. As well, using hormone therapy limits the application of this model to only female subjects. Alternatively, our approach to build a healthy user index score, overcomes many of these issues. First of all, by incorporating a large number of variables into a single summary score, we are not limited by low numbers of observations/events. Using one score for adjustments will preserve degrees of freedom in regression models, as others have previously proposed for the ADG mortality risk score,[43] and also the Charlson comorbidity index.[44] Second, by incorporating a vast number of heathy user markers into an overall summary score, we believe a stronger profile of the healthy user characteristics can be incorporated into the models, as opposed to a single variable that serves as a proxy for the healthy user (e.g., mammography or colonoscopy). Related, although propensity scores have also been tried previously, these models are still limited in that they are attempting to predict a single healthy user attribute. Thus, again, propensity scores may be limited in profiling all major healthy user characteristics.

From a research perspective, beyond adjustment for healthy user bias, our healthy user index could be beneficial in characterizing a population, as it provides an overall summary of the patients' health behaviour attributes and is not reliant on a single 'marker' of the healthy user. Moreover, the healthy user index could also be used to evaluate the consistency of medication/therapy effects in those with low healthy user index scores (i.e., in those with lower probability of being a healthy user) and in those with high healthy user scores (i.e., those with higher probability of being a healthy user) to assist in the identification of healthy use bias within preventive medication/therapy studies. Thus, even if the healthy user index score is not able to fully control for the bias, evaluation of the consistency of study effects among potential subgroups of the index would provide value by helping to identify if healthy user bias may be at play in the results observed in a study of preventive medications/therapies. Importantly, the administrative data required to build the healthy user index is readily available in most jurisdictions. ICD codes for preventive services and procedural codes are routinely captured in administrative data. Further, although some provinces (e.g., Ontario) only fully capture prescription medications for those 65 years and older, the majority of provinces in Canada capture all drugs for all people. As a result, the index should be adaptable to most health jurisdictions.

A strength of this study was the large population and the wealth of information available in the database, including administrative and demographic information, medical service claims, laboratory data and pharmacy claims data. Our study was not without limitations. First, the population we studied was limited to adult patients with type 2 diabetes. The healthy user index will need to be externally validated in the future, including populations with other comorbidities. Moreover, future studies will also need to be completed within administrative databases to evaluate the benefits of the index in controlling for healthy user bias in observational studies. Second, in the index we grouped a number of variables together (e.g., we grouped bone mineral density screening with filling a prescription for a bone-resorption inhibitor medication). This may not be appropriate when applying the score to different populations. Third, we used receipt of influenza vaccination as our prototypical marker of the healthy user. Although a significant amount of literature points to influenza vaccination as a marker of healthy users, this may not be the case. There is no gold standard to identify healthy users in administrative data, and we did do sensitivity analyses with statin use. Fourth, our index was based on variables typically available within administrative databases. Other databases may have access to other

variables that could be incorporated into the index (e.g., smoking status, exercise behaviours, etc.). Furthermore, the healthy predictive variables we used were based on typical administrative coding and may not be fully validated; although the majority of the codes were based on the Medicare recommended preventive services codes.[32] Last, our data was derived from the United States. As such, the rates of preventative services, vaccine use, and other healthy user characteristics we observed may not be directly translated to other jurisdictions including Canada. For example, the rates of influenza vaccination use are lower in our study than those observed in other estimates from Canada during our study period.[45] How this may have impacted the development of the index is uncertain; however, importantly, both the derivation and validation cohorts were established from the same population and thus any bias should be expected to be similar within the two cohorts. Moreover, our results were consistent when statin utilization was used as a prototypical marker and previous research during our time of study has shown statin utilization to be nearly identical between the United States and Canada.[46]

In conclusion, we developed a summary score that combines age, sex and healthy behaviours to predict healthy users within administrative datasets. Our summary score performed modestly well when internally validated, and in the future, it will require refinement and external validation. This index score may allow researchers to better identify and adjust for healthy user bias in observational health services research; however, we acknowledge that observational research will always be limited in the evaluation of effectiveness and perhaps we should focus our efforts on randomized trials.

CONFLICT OF INTEREST

No conflicts of interest exist.

FUNDING

None

REFERENCES

1. Eurich DT, Majumdar SR. Statins and sepsis - Scientifically interesting but clinically inconsequential. J Gen Intern Med 2012;27:268–69.
2. Shrank WH, Patrick AR, Alan Brookhart M. Healthy user and related biases in observational studies of preventive interventions: A primer for physicians. J Gen Int Med 2011;26:546–50.
3. Brookhart MA, Patrick AR, Dormuth C, et al. Adherence to lipid-lowering therapy and the use of preventive health services: An investigation of the healthy user effect. Am J Epidemiol 2007;166:348–54.
4. Dormuth CR, Patrick AR, Shrank WH, et al. Statin adherence and risk of accidents: A cautionary tale. Circulation 2009;119:2051–57.
5. Nelson JC, Jackson ML, Weiss NS, Jackson LA. New strategies are needed to improve the accuracy of influenza vaccine effectiveness estimates among seniors. J Clin Epidemiol 2009:62:687–94.
6. Lau D, Eurich DT, Majumdar SR, Katz A, Johnson JA. Effectiveness of influenza vaccination in working-age adults with diabetes: A population-based cohort study. Thorax 2013;68:658–63.
7. Eurich DT, Marrie TJ, Johnstone J, Majumdar SR. Mortality reduction with influenza vaccine in patients with pneumonia outside "flu" season: Pleiotropic benefits or residual confounding? Am J Respir Crit Care Med 2008;178:527–33.
8. Canadian Communicable Disease Report. Statement on seasonal influenza vaccine for 2013–2014. 2013;39 (ACS-4). [Public Health Agency of Canada website]. October 2013. Available at: http://www.phac-aspc.gc.ca/publicat/ccdr-rmtc/13vol39/acs-dcc-4/assets/pdf/13vol39-acs-dcc4-eng.pdf. Accessed December 9, 2016.
9. Jackson LA, Nelson JC, Benson P, et al. Functional status is a confounder of the association of influenza vaccine and risk of all cause mortality in seniors. Int J Epidemiol 2006;35:345–52.
10. McGrath LJ, Cole SR, Kshirsagar AV, Weber DJ, Stürmer T, Brookhart MA. Hospitalization and skilled nursing care are predictors of influenza vaccination among patients on hemodialysis: Evidence of confounding by frailty. Med Care 2013;51:1106–13.
11. Gleason CE, Dowling NM, Friedman E, Wharton W, Asthana S. Using predictors of hormone therapy use to model the healthy user bias: How does healthy user status influence cognitive effects of hormone therapy? Menopause 2012;19:524–33.
12. Eurich DT, Simpson S, Senthilselvan A, et al. Comparative safety and effectiveness of sitagliptin in patients with type 2 diabetes: retrospective population based cohort study. BMJ 2013;346:f2267.

13. Padwal R, Lin M, Etminan M, et al. Comparative effectiveness of olmesartan and other Angiotensin receptor blockers in diabetes mellitus: retrospective cohort study. Hypertension 2014;63(5):977–83.
14. Weir DL, McAlister FA, Senthilselvan A, et al. Sitagliptin use in patients with diabetes and heart failure: a population-based retrospective cohort study. JACC Heart Fail 2014;2(6):573–82.
15. McAlister FA, Youngson E, Eurich DT. Treatment Deintensification Is Uncommon in Adults With Type 2 Diabetes Mellitus: A Retrospective Cohort Study. Circ Cardiovasc Qual Outcome 2017;10(4) doi: 10.1161/CIRCOUTCOMES.116.003514.
16. Husein N, Woo V. Canadian diabetes association 2013 clinical practice guidelines for the prevention and management of diabetes in Canada: Pharmacologic management of type 2 diabetes. Can J Diabetes [serial online] 2013;37(suppl 1):S93.
17. American Diabetes Association. Standards of medical care in Diabetes—2014. Diabet Care 2014;37(Supplement 1):S14–S80.
18. Remschmidt C, Wichmann O, Harder T. Frequency and impact of confounding by indication and healthy vaccinee bias in observational studies assessing influenza vaccine effectiveness: a systematic review. BMC Infect Dis 2015;15:429. doi: 10.1186/s12879-015-1154-y.
19. Remschmidt C, Wichmann O, Harder T. Vaccines for the prevention of seasonal influenza in patients with diabetes: Systematic review and meta-analysis. BMC Med 2015;13:53.
20. Jackson ML, Yu O, Nelson JC, et al. Further evidence for bias in observational studies of influenza vaccine effectiveness: the 2009 influenza A(H1N1) pandemic. Am J Epidemiol 2013;178(8):1327–36. doi: 10.1093/aje/kwt124.
21. Hottes TS, Skowronski DM, Hiebert B, et al. Influenza vaccine effectiveness in the elderly based on administrative databases: change in immunization habit as a marker for bias. PLoS One 2011;6(7):e22618. doi: 10.1371/journal.pone.0022618.
22. Baxter R, Lee J, Fireman B. Evidence of bias in studies of influenza vaccine effectiveness in elderly patients. J Infect Dis 2010;201(2):186–9.
23. Jackson LA, Jackson ML, Nelson JC, et al. Evidence of bias in estimates of influenza vaccine effectiveness in seniors. Int J Epidemiol 2006;35(2):337–44.
24. Crowcroft NS, Rosella LC. The potential effect of temporary immunity as a result of bias associated with healthy users and social determinants on observations of influenza vaccine effectiveness; could unmeasured confounding explain observed links between seasonal influenza vaccine and pandemic H1N1 infection? BMC Pub Health 2012;12:458. doi: 10.1186/1471-2458-12-458.
25. Achtymichuk KA, Johnson JA, Al Sayah F, et al. Characteristics and health behaviors of diabetic patients receiving influenza vaccination. Vaccine 2015;33(30):3549–55.
26. Hak E, Hoes AW, Nordin J, et al. Benefits of influenza vaccine in US elderly--appreciating issues of confounding bias and precision. Int J Epidemiol 2006;35(3):800–2; author reply 799-800. doi: 10.1093/ije/dyl068.
27. Nichol KL. Challenges in evaluating influenza vaccine effectiveness and the mortality benefits controversy. Vaccine 2009;27(45):6305–11.
28. Schwartz KL, Jembere N, Campitelli MA, et al. Using physician billing claims from the Ontario Health Insurance Plan to determine individual influenza vaccination status: an updated validation study. CMAJ Open 2016;4(3):E463-E70. doi: 10.9778/cmajo.20160009.
29. Kwong JC, Manuel DG. Using OHIP physician billing claims to ascertain individual influenza vaccination status. Vaccine 2007;25(7):1270–4.
30. Neuzil KM, Reed GW, Mitchel EF, Jr., et al. Influenza-associated morbidity and mortality in young and middle-aged women. JAMA 1999;281(10):901–7.
31. Centers for Disease Control and Prevention. Past weekly surveillance reports. [Centers for Disease Control and Prevention web site]. Available at: http://www.cdc.gov/flu/weekly/pastreports.htm. Accessed July 9, 2016.
32. Department of Health and Human Services. Preventive services. [Centers for Medicare and Medicaid Services web site]. Available at: http://www.cms.gov/Medicare/Prevention/PrevntionGenInfo/Downloads/MPS-QuickReferenceChart-1TextOnly.pdf. Accessed October 2016 accessed October 5, 2016.
33. Canadian Chronic Disease Surveillance System Osteoporosis Working Group. Use of administrative data for national surveillance of osteoporosis and related fractures in Canada: results from a feasibility study. Arch Osteoporos 2013;8:143. doi: 10.1007/s11657-013-0143-2.
34. Quan H, Li B, Saunders LD, et al. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. Health Serv Res 2008;43(4):1424–41.

35. Quan H, Parsons GA, Ghali WA. Validity of procedure codes in International Classification of Diseases, 9th revision, clinical modification administrative data. Medical Care 2004;42(8):801–9.
36. Lofters A, Vahabi M, Glazier RH. The validity of self-reported cancer screening history and the role of social disadvantage in Ontario, Canada. BMC Pub Health 2015;15:28. doi: 10.1186/s12889-015-1441-y.
37. Ho PM, Bryson CL, Rumsfeld JS. Medication adherence: its importance in cardiovascular outcomes. Circ 2009;119(23):3028–35.
38. Ismaila A, Corriveau D, Vaillancourt J, et al. Impact of adherence to treatment with tiotropium and fluticasone propionate/salmeterol in chronic obstructive pulmonary diseases patients. Curr Med Res Opin 2014;30(7):1427–36.
39. Meijer WM, Penning-van Beest FJ, Olson M, et al. Relationship between duration of compliant bisphosphonate use and the risk of osteoporotic fractures. Curr Med Res Opin 2008;24(11):3217–22.
40. Hollestein L, Baser Ö, Stricker B, et al. The healthy user and healthy adherer bias: a nested case-control study among statin users in the Rotterdam Study. Arch Pub Health 2015;73(Suppl 1):O6–O6.
41. Silverman SL, Gold DT. Healthy users, healthy adherers, and healthy behaviors? J Bone Miner Res 2011;26(4):681–2.
42. Sullivan LM, Massaro JM, D'Agostino RB, Sr. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. Stat Med 2004;23(10):1631–60.
43. Austin PC, Walraven C. The mortality risk score and the ADG score: two points-based scoring systems for the Johns Hopkins aggregated diagnosis groups to predict mortality in a general adult population cohort in Ontario, Canada. Med Care 2011;49(10):940–7.
44. Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373–83.
45. Canadian Communicable Disease Report. Canadian Community Health Survey Results Influenza Immunization Coverage. Available from: http://www.bccdc.ca/resource-gallery/Documents/Statistics and Research/Statistics and Reports/Immunization/Coverage/CCHS20070820110601.pdf. Accessed July 30, 2017.
46. Jackevicius CA, Tu JV, Ross JS, et al. Use of ezetimibe in the United States and Canada. N Engle J Med 2008;358(17):1819–28.

APPENDIX 1. CODES

Condition	ICD or Procedural codes or AHFS codes
Any Cancer Screening	
Pap Test	ICD9= ′V762′, ′V7647′, ′V7649′, ′V1589′, ′V7231′; proc_cd= ′G0123′,′G0124′, ′G0141′, ′G0143′, ′G0144′, ′G0145′, G0147′, ′G0148′, ′P3000′, ′P3001′, ′Q0091′
Pelvic Exam	ICD9= ′V762′, ′V7647′, ′V7649′, ′V1589′, ′V7231′; proc_cd= ′G0101′
Mammography	ICD9= ′V7611′, ′V7612′; proc_cd= ′77052′, ′77057′, ′G0202′
Colorectal Screen	proc_cd= ′G0104′, ′G0105′, ′G0106′, ′G0120′, ′G0121′, ′G0122′, ′G0328′, ′82270′
Prostate Screen	ICD9= ′V7644′; proc_cd= ′G0102′, ′G0103′
Cardiovascular Screening	ICD9= ′V810′, ′V811′, ′V812′; proc_cd= ′80061′, ′82465′, ′83718′, ′84478′
Medical Nutrition Therapy (as a marker of frailty)	proc_cd= ′97802′, ′97803′, ′97804′, ′G0270′, ′G0271′
Hormone Replacement Therapy Prescription	AHFS= 68:16.04 (Estrogens), 68:16.12 (Estrogen Agonist-Antagonists), 68:32 (Progestins)
Smoking Cessation Therapy Prescription	AHFS= 12:92 (nicotine replacement therapy, varenicline), 28:16.04.92 (bupropion)
Obesity Mediation Prescription	56:92 (orlistat)
Statin Prescription	AHFS=′240608′ or ′24060800′
Osteoporosis Screening and/ or Bone Resorption Inhibitor Prescription	
Osteoporosis Screen	proc_cd= ′76977′, ′77078′, ′77079′, ′77080′, ′77081′, ′77083′, ′G0130′
Bone Resorption Inhibitor Prescription	AHFS= 92:24 (alendronate, denosumab, etidronate, gallium nitrate, ibandronate, pamidronate, risedronate, zoledronic acid)
Dementia	EDC= edcNUR11
Labs Completed	If any of the following were completed in the last year: Albumin, cholesterol, triglycerides, HDL, LDL, A1C, hemoglobin, creatinine