

BMJ Open Can purchasing information be used to predict adherence to cardiovascular medications? An analysis of linked retail pharmacy and insurance claims data

Alexis A Krumme,¹ Gabriel Sanf elix-Gimeno,² Jessica M Franklin,¹ Danielle L Isaman,¹ Mufaddal Mahesri,¹ Olga S Matlin,³ William H Shrank,³ Troyen A Brennan,³ Gregory Brill,¹ Niteesh K Choudhry¹

To cite: Krumme AA, Sanf elix-Gimeno G, Franklin JM, *et al.* Can purchasing information be used to predict adherence to cardiovascular medications? An analysis of linked retail pharmacy and insurance claims data. *BMJ Open* 2016;**6**:e011015. doi:10.1136/bmjopen-2015-011015

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2015-011015>).

Received 4 January 2016
Revised 21 September 2016
Accepted 29 September 2016



CrossMark

For numbered affiliations see end of article.

Correspondence to

Dr Niteesh K Choudhry;
nkchoudhry@bwh.harvard.edu

ABSTRACT

Objective: The use of retail purchasing data may improve adherence prediction over approaches using healthcare insurance claims alone.

Design: Retrospective.

Setting and participants: A cohort of patients who received prescription medication benefits through CVS Caremark, used a CVS Pharmacy ExtraCare Health Care (EHC) loyalty card, and initiated a statin medication in 2011.

Outcome: We evaluated associations between retail purchasing patterns and optimal adherence to statins in the 12 subsequent months.

Results: Among 11 010 statin initiators, 43% were optimally adherent at 12 months of follow-up. Greater numbers of store visits per month and dollar amount per visit were positively associated with optimal adherence, as was making a purchase on the same day as filling a prescription ($p < 0.0001$ for all). Models to predict adherence using retail purchase variables had low discriminative ability (C-statistic: 0.563), while models with both clinical and retail purchase variables achieved a C-statistic of 0.617.

Conclusions: While the use of retail purchases may improve the discriminative ability of claims-based approaches, these data alone appear inadequate for adherence prediction, even with the addition of more complex analytical approaches. Nevertheless, associations between retail purchasing behaviours and adherence could inform the development of quality improvement interventions.

INTRODUCTION

Suboptimal adherence to evidence-based medications for coronary artery disease and other chronic conditions can lead to potentially avoidable morbidity, mortality and health spending.^{1 2} The accurate prediction

Strengths and limitations of this study

- Using a unique data set combining prescription claims data with retail purchase data, this study aims to improve adherence prediction accuracy where traditional claims-based approaches have typically performed poorly.
- This study uses both investigator-specified and empirical predictor selection to evaluate and maximise prediction accuracy.
- This study comprises complete dispensing information for a patient and so minimal misclassification of the outcome and other medication use covariates is expected.
- Members with loyalty cards may not use them at every purchase, and these data do not capture retail purchases patients make at non-CVS Pharmacy locations.

of who will ultimately become non-adherent could help identify which patient groups would most benefit from interventions to maintain adherence.^{3–5} Traditional approaches to prediction generally use insurance reimbursement claims data from commercial or government-sponsored insurance programmes to generate clinical and demographic variables measured at or prior to treatment initiation. Other methods to identify non-adherence rely on patient self-report, pill counts or electronic pill monitors, but are not useful for identifying non-adherence before it occurs.^{6 7} While many of the variables available in insurance databases are correlated with adherence, for example, age and race, they provide weak discrimination between adherers and non-adherers.^{8–12}

Good adherence has repeatedly been linked to other healthy behaviours, such as

greater use of preventive healthcare services, influenza vaccination and compliance with recommended screening programmes through a phenomenon known as the 'healthy user effect'.^{13–15} Thus, the use of richer data sources that capture activities reflecting such healthy attributes and behaviours might improve adherence prediction over approaches using healthcare claims alone. Such rich 'big data' sources have had success in predicting similarly complex individual behaviours. For example, mobile phone data have been used to describe individual mobility patterns and infer friendship networks, while transactional data from credit and debit cards have been used to predict consumption patterns.^{16–20}

The extent to which enrichment of insurance claims data with in-pharmacy purchasing information may improve the prediction of medication filling and refilling behaviours has not been evaluated until now. Using a unique data set combining prescription claims data with retail purchase data of individuals at CVS Pharmacy locations, we explored the association between purchasing behaviours in the year prior to initiation of a statin medication and adherence in the subsequent year. Our approach employed a combination of investigator-specified purchasing predictors and those selected by a high-dimensional propensity score (hd-PS) algorithm.^{21–22} We additionally compared logistic regression results to those enhanced with boosting techniques, a machine-learning method that identifies non-linear associations between predictors and outcome or deep interactions among predictors, which may be useful for modelling behaviour as complex as medication adherence.^{23–24} Enhanced predictive ability from this linked data set could improve targeting of patients who stand to benefit from adherence-improving interventions, while the co-occurrence of in-pharmacy purchasing and filling behaviours could lead to the development of novel interventions.

METHODS

Data source

We used a data set linking prescription claims and retail purchase data for individuals receiving prescription medication insurance benefits through CVS Caremark, a pharmacy benefits manager that provides coverage to

more than 65 million members annually making purchases at CVS Pharmacy stores using a CVS Pharmacy 'ExtraCare Health Care' (ECHC) loyalty card. Retail purchasing data consisted of item-level data on each purchase, including date of visit, dollar amount paid, quantity, and whether the product was on sale or a CVS brand item. Data also included information from CVS Pharmacy's internal product hierarchy, a system for classifying products according to merchandise group (eg, consumer health), category (eg, pain relievers) and subcategory (naproxen). Prescription claims data included information on all paid prescription claims including the brand and generic drug name, dose, days supply, patient paid amount and insurer paid amount from all prescription claims, as well as patient gender, age group and region of residence. Data on education level, median income and racial distribution in a zip code, derived from 2010 Census data by CVS Health, were additionally provided.

All purchasing data were linked to prescription claims data based on a unique masked patient identifier. To safeguard the anonymity of the participants included in the analysis, the linked data set used by the research team did not contain any direct patient identifiers. The Institutional Review Board of Brigham and Women's Hospital approved this study.

Study population

The study population consisted of patients who received prescription medication benefits through CVS Caremark, a pharmacy benefits manager that serves over 65 million individual members nationwide who also received a CVS Pharmacy ECHC loyalty card from their health plan. Whereas members are enrolled individually, loyalty cards do not require identification to be used and thus can be shared, typically within a household.

From this cohort, we identified individuals newly initiating a statin or statin combination drug between 1 January 2011 and 31 December 2011. A patient's index date was defined as the date of the first prescription filled during the study window. Patients were required to have continuous insurance enrolment for 180 days before the index date and 1-year postindex (figure 1). New use was defined as not having any statin filled in the 180 days prior to the index date. Finally, we restricted our cohort to patients with at least three

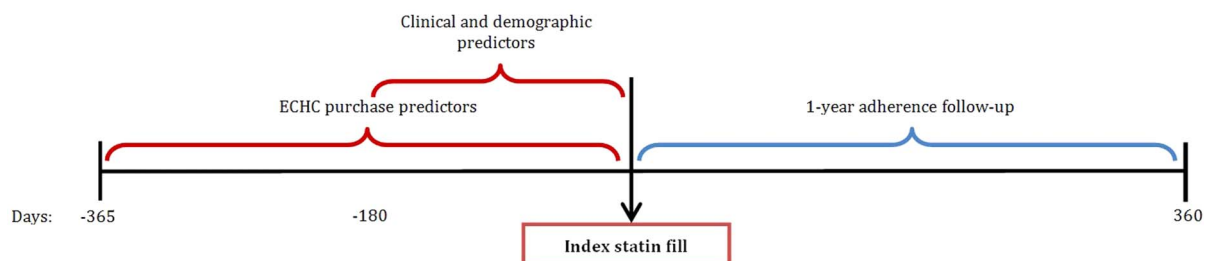


Figure 1 Study design and data sources. ECHC, ExtraCare Health Care.

unique store visits in which the ECHC card was used in the year prior to the first statin prescription to enrich the data set from which purchasing predictors were drawn.

Adherence measures

We measured medication adherence using prescription fill data from prescription claims. By stringing together all paid fills, we created a statin ‘supply diary’ that indicates whether on each day for the year after the index date patients had a statin medication available to them. From this diary, we calculated a proportion of days covered (PDC), defined as the number of days with medication available divided by 365. We defined ‘full adherence’ as a $PDC \geq 0.8$, which corresponds to the level of use above which patients with coronary artery disease derive clinically relevant benefit from statins²⁵ and the threshold employed by most quality measures.^{26 27}

Adherence predictors

Investigator-specified clinical and demographic variables: We defined nine clinical characteristics based on prescription claims incurred during the 180 days before the index date (table 1). Clinical characteristics included the number of unique medications taken and the presence of maintenance medications likely to influence adherence to a cardiovascular medication, including antihypertensive, antiplatelet, oral antidiabetic, chronic obstructive pulmonary disease and asthma, heart failure, antidepressant, antiarrhythmic, anticoagulant, and osteoporosis medications. Characteristics of the index prescription included whether or not it was a generic (as compared with a branded medication) and the dollar amount the patient must pay at the pharmacy (copayment amount). Demographic information included age group and sex. Information at the zip code level included race, education and median household income.

Investigator-specified purchasing variables: Using the retail purchasing data, we defined 12 variables that we hypothesised would be associated with adherence. These variables were developed based on a manual review by a team of clinicians and epidemiologists of the purchasing data for a random sample of 100 patients, 50 of whom were fully adherent to their statin and 50 of whom were not. Our team first discussed candidate variables. We subsequently refined potential predictors while blinded to the patients’ observed adherence and then using patients’ observed adherence to calibrate how the variables were defined. In all instances, we used descriptors from the internal product hierarchy, consisting of merchandise group, category and subcategory to assign individual products. After the final list of variables was agreed on, they were generated for the entire cohort for use as potential predictors.

The final list of investigator-specified behaviours is presented in table 2. Behaviours 1, 2 and 3 were based on

Table 1 Cohort baseline characteristics

Demographics	N (%)
Female sex	4345 (39.6%)
Age (mean)	59.3
Median household income in zip	
<\$50 000	3719 (33.8%)
\$50 000–\$99 999	6457 (58.6%)
\$100 000+	834 (7.6%)
Per cent black in zip, mean (SD)	12.3 (17.9%)
Per cent \geq HS education, mean (SD)	86.9 (8.5%)
Region	
Midwest	2122 (19.3%)
Northeast	3053 (27.7%)
South	4750 (43.1%)
West	1085 (9.9%)
Index statin (generic)	
Amlodipine/atorvastatin	56 (0.5%)
Atorvastatin	2474 (22.5%)
Fluvastatin	23 (0.2%)
Lovastatin	436 (4.0%)
Niacin/lovastatin	17 (0.2%)
Niacin/simvastatin	69 (0.6%)
Pitavastatin	96 (0.9%)
Pravastatin	1441 (13.1%)
Rosuvastatin	1760 (16.0%)
Simvastatin	4638 (42.1%)
Copayment index statin prescription, mean (SD)	\$24.70 (\$26.10)
<i>Comorbidity medications</i>	<i>N (%)</i>
Coronary artery disease	391 (3.6%)
Asthma/COPD	1172 (10.6%)
Diabetes	2219 (20.2%)
Congestive heart failure	246 (2.2%)
Hypertension	6105 (55.4%)
Mental health and depression	2074 (18.8%)
Platelet inhibitor	603 (5.5%)
Loop diuretic	545 (5.0%)
Antiarrhythmics	122 (1.1%)
Osteoporosis	481 (4.4%)
Anticoagulant	383 (3.5%)
Total number of medications, mean (SD)	5.3 (4.2)

COPD, chronic obstructive pulmonary disease; HS, higher secondary.

products within a given purchasing category. For example, we classified each product listed in the ‘food’ merchandise group as being ‘healthy’ or ‘unhealthy’ based on product descriptions, hypothesising that unhealthy foods would be associated with worse adherence. Similarly, we classified all products listed in the ‘consumer healthcare’ merchandise group as being for preventive or symptomatic purposes, or both. Behaviours 4 and 11 recorded the number of visits in which any product in the ‘consumer healthcare’ merchandise group was purchased. Behaviours 5–9 used fields outside of the product hierarchy to describe the amount of money spent, whether products were purchased on sale, and the number of monthly store visits with a purchase. Behaviour 10 used retail purchasing

Table 2 Retail purchasing behaviour association with statin adherence

#	Patient-level retail purchasing behaviour	Median (IQR) and proportion above median			Association with optimal statin adherence 12 months after initiation			
		Full cohort (N=11 010)	Optimally adherent (N=4691) (%)	Suboptimally adherent (N=6319) (%)	Univariable OR (95% CI)	p Value	Multivariable OR (95% CI)*	p Value
1	Per cent of visits with ≥ 1 unhealthy food purchase	25 (35)	45.1	50.1	0.996 (0.994 to 0.997)	<0.0001	0.998 (0.996, 1.001)	0.14
2	Per cent of visits with ≥ 1 preventive consumer healthcare product purchased	18 (33)	50.9	48.0	1.005 (1.003 to 1.007)	<0.0001	1.002 (0.999, 1.005)	0.11
3	Per cent of visits with ≥ 1 symptomatic consumer healthcare product purchased	33 (31)	48.4	44.5	1.003 (1.001 to 1.004)	0.001	1.001 (0.997, 1.004)	0.73
4	Per cent of visits with ≥ 1 consumer healthcare product purchased	44 (40)	50.9	46.9	1.003 (1.002 to 1.005)	<0.0001	0.999 (0.995, 1.003)	0.63
5	Per cent of all purchased items that were on sale	35 (25)	49.2	49.1	1.00 (0.998 to 1.002)	0.78	1.000 (0.998, 1.002)	0.92
6	Per cent of all purchased items that were CVS brand	23 (20)	51.1	47.5	1.005 (1.003 to 1.007)	<0.0001	1.003 (1.000, 1.006)	0.02
7	Maximum number of items in a visit	5 (4)	41.1	39.5	1.01 (0.999 to 1.02)	0.07	0.988 (0.973, 1.002)	0.09
8	Dollar amount spent per visit	\$15 (\$11)	48.2	43.4	1.01 (1.007 to 1.014)	<0.0001	1.012 (1.007, 1.016)	<0.0001
9	Number of visits per month	0.9 (1.2)	32.5	27.9	1.07 (1.04 to 1.10)	<0.0001	1.091 (1.004, 1.185)	0.04
10	Ratio of index statin copayment amount vs median income in zip code of residence	0.2 (0.5)	45.8	50.0	0.79 (0.73 to 0.85)	<0.0001	0.795 (0.732, 0.863)	<0.0001
11	Per cent of all consumer healthcare product purchases that were on sale	N (%)						
	No consumer healthcare purchases made	864 (7.8%)	7.0	8.5	REF		REF	
	<25% consumer healthcare purchases were on sale	9540 (86.6%)	87.0	86.4	1.22 (1.06 to 1.41)	0.68	0.995 (0.916, 1.082)	0.91
	$\geq 25\%$ consumer healthcare purchases were on sale	606 (5.5%)	6.0	5.1	1.42 (1.15 to 1.76)	0.01	1.016 (0.884, 1.167)	0.83
12	A prescription fill for any medication occurred on the same day as the store visit, yes vs no	7001 (63.6%)	65.4	62.3	1.14 (1.06 to 1.24)	0.001	1.132 (1.006, 1.274)	0.04

*Includes adjustment for age, sex, region, zip code-level race, household income, and education level, and interaction terms between behaviors #1 and #12 and behavior #9.

data in conjunction with 2010 Census data derived from patient zip code of residence, hypothesising that a higher burden of drug costs relative to neighbourhood income might influence adherence. Behaviour 12 additionally used purchasing and prescription drug data in the 180 days prior to statins initiation to observe whether filling coincided with a retail purchase, hypothesising that patients with co-filling would be more likely to be adherent.

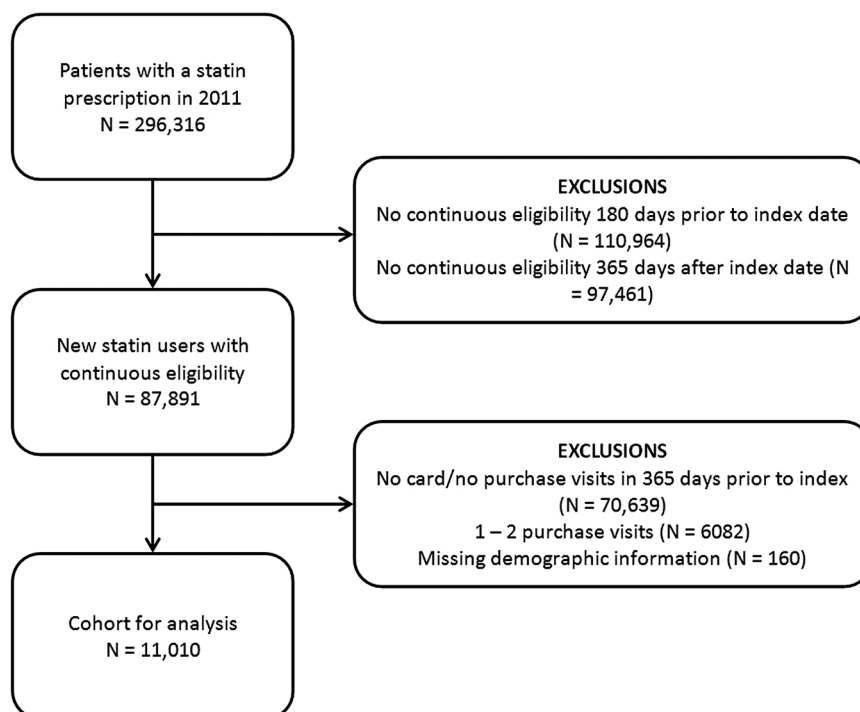
High-dimensional propensity score approach: To more completely use all of the available data for adherence prediction without any a priori hypotheses, we applied a hd-PS variable selection algorithm to the retail purchasing data and prescription claims data incurred during the 180 days prior to statin initiation.^{21 22} When applied to prescription claims, the hd-PS algorithm creates binary variables that indicate the frequency of each unique medication. When applied to purchasing data, the algorithm indicates the frequency of each product subcategory. The hd-PS algorithm can screen thousands of variables for an empirical association with a study exposure, and ranks variables with the expectation that a large number of highly ranked variables can collectively proxy for unmeasured confounders. Equivalently, since our goal was prediction rather than confounder adjustment, we used the hd-PS approach to identify a large collection of variables that could be proxies for underlying behavioural constructs that may be predictive of adherence as has been done previously.²⁸ We ran hd-PS on two sets of data: first, using all prescription claims data available and second, using all products in CVS Pharmacy's retail product hierarchy, in each case selecting the 200 variables with the strongest associations with adherence.

Statistical analyses

We first evaluated associations between the 12 investigator-specified purchasing variables and optimal adherence during follow-up using logistic regression models. We measured univariable associations as well as a multivariable model adjusting for all purchasing variables, two prespecified interaction terms, and demographic covariates to observe the magnitude and direction of each effect after accounting for other characteristics.

To compare the discriminative ability of different groups of adherence predictors, we estimated six models. In models with hd-PS predictors, the models included predictors selected by the algorithm as model covariates. These models were evaluated with respect to their ability to discriminate between patients who were and were not adherent during the year after initiation, as measured by the C-statistic. This measure ranges from 0.5 to 1.0, corresponding to a completely non-informative model and perfect prediction, respectively.²⁹ To avoid the 'overoptimism' bias associated with evaluating model prediction accuracy in the same data that were used to estimate the model, we performed 10-fold cross-validation.³⁰ Each model was predicted using both logistic regression and generalised boosting regression, a non-parametric data-mining technique capable of fitting highly predictive models with deep interactions among predictors.²³ Logistic regression models were estimated using SAS (SAS, V.9.3, Cary, North Carolina, USA); all models were re-estimated using the generalised boosting algorithm, as implemented in the R package gbm.¹¹ To identify the best-performing model(s), we compared the cross-validated C-statistic across all logistic and boosted regression models. Finally, we re-estimated our six

Figure 2 Consort diagram.



prediction models with the addition of retail purchase predictors from the first 3 months after initiation to observe whether changes in retail patterns during this period might provide additional improvement in adherence prediction.

RESULTS

Patient characteristics

The cohort consisted of 11 010 patients initiating a statin during the study period and who used their loyalty card on at least three visits prior to initiation (figure 2). Included versus excluded patients were similar in age and gender composition (59.3 vs 57.8 years, and 40.3% vs 39.6% female), as well as regional distribution, though included patients had a higher proportion of patients from the Northeast (28% vs 18%). Patients tended to be male (60.4%) with an average age of 59 years (table 1). Simvastatin (42.1%) and atorvastatin (22.5%) were the most commonly used statins. More than half of the patients (55.4%) were concomitantly taking an antihypertensive while one-fifth (20.2%) were taking an oral antidiabetic at the time of statin initiation. On average, patients filled prescriptions for more than five unique medications in the year prior to index. At 12 months, 4691 (42.6%) of patients were optimally adherent to their statin medication.

Purchasing patterns

Results from investigator-specified purchasing variables are presented in table 2. The median number of monthly visits was 0.9 (IQR 1.2). During these visits, patients purchased a median of 5 (IQR 4) items, and spent a median of \$15 (IQR \$11). These visits included an unhealthy food product on average 28% of the time, while on average nearly half (45%) included the purchase of a consumer healthcare product. The median patient purchased a symptomatic consumer healthcare product on 33% of visits (IQR 31%), and a preventive consumer healthcare product on 18% of visits (IQR 33%).

Nearly two-thirds (63.6%) of patients made a store purchase on the same day as a prescription fill in the 180 days prior to index. Of all the items purchased during the study period, a median per cent of 35 (IQR 25) were on sale.

Relationship between potential predictors and adherence

In univariable analysis, greater numbers of visits per month and dollar amount per visit were positively associated with adherence, as was making a purchase on the same day as filling a prescription and a higher proportion of visits with preventive and symptomatic healthcare purchases. In contrast, unhealthy purchases and higher copayment amounts relative to median income in zip code of residence were associated with lower adherence; the latter association was additionally observed within strata of median income in zip code

(<\$50 000; \$50 000–\$100 000; >\$100 000). While the proportion of visits with a CVS brand product was significantly associated with adherence, the per cent with an item on sale was not ($p<0.0001$ vs $p=0.78$). Higher statin copayments relative to median income in zip code of residence were associated with worse adherence ($p<0.0001$). Results from multivariable analyses adjusted for demographic and all purchasing behaviour variables, as well several multiplicative interaction terms, were similar.

Prediction models

The results of prediction models are shown in table 3. Both models using purchase variables, one with investigator-specified variables and the other with variables selected by the hd-PS, had similar predictive ability (C-statistic for both: 0.563). Investigator-specified clinical variables achieved a C-statistics of 0.599, while models with both clinical and retail purchase variables had slightly greater discriminative ability (C-statistic: 0.617). Generalised boosting regression improved prediction slightly in five out of six models; the highest C-statistic achieved was 0.621 in the model with clinical and retail purchase variables. Results from logistic regression models incorporating retail purchase data from the 3 months after initiation were similar (see online supplementary appendix A).

Table 3 Model prediction

Adherence predictors	C-statistic	
	Logistic regression	Boosted regression
Investigator-specified retail purchase variables	0.563	0.537
High-dimensional propensity approach for retail purchase data	0.563	0.573
Investigator-specified clinical variables	0.599	0.602
Investigator-specified clinical variables High-dimensional propensity approach for retail purchase data	0.599	0.611
Investigator-specified clinical variables Investigator-specified retail purchase variables	0.607	0.610
Investigator-specified retail purchase variables Investigator-specified clinical variables High-dimensional propensity approach for retail purchase data High-dimensional propensity approach for prescription claims data	0.617	0.621

DISCUSSION

One of the central challenges in predicting adherence is the multifaceted nature of adherence itself. Adherence is the result of complex interactions between patients, providers and the healthcare system, as well as of behavioural characteristics unique to patients themselves.^{8 31–33} It is, therefore, not surprising that efforts to predict adherence using insurance claims data have generally yielded disappointing results.^{12 28 34} Pharmacies also have access to data about retail purchases for goods such as snacks and beverages, beauty products, and over-the-counter medicines that patients make at the ‘front of the store’. Accordingly, we hypothesised that retail purchasing transactions, confidentially linked to prescription claims, could be a rich source of data to capture adherence-related behaviours that are otherwise not observable using routinely collected data. In a cohort of more than 11 000 statin initiators, 43% of whom were optimally adherent at 12 months, we found that several variables generated with these data tend to be associated with long-term medication use. In specific, after multivariable adjustment for all other predictors, more monthly visits, a great dollar amount spent per visit, a greater proportion of items that were store-branded, and making a purchase and filling a prescription on the same visit were significantly associated with optimal adherence to statin medications in the year after initiation. While small in magnitude, these associations nonetheless suggest that greater retail purchase activity, in the form of store visits and amount of money spent, correlates with adherence, as does the purchase of lower priced equivalent store brand products, which could be a manifestation of the well-documented ‘healthy user’ effect.

While the use of retail purchasing data may improve the discriminative ability of traditional claims-based approaches, these data alone appear inadequate for adherence prediction, even with the addition of more complex analytical approaches, including the use of a hd-PS algorithm and boosting techniques. There are several potential reasons for this apparent discrepancy. Most obviously, retail transaction data may not actually sufficiently capture true underlying constructs reflective of medication adherence. Just as adherence behaviours cannot be explained by health system use and disease state information as captured in insurance claims data, retail purchase data that tabulate the frequency, quantity and types of purchases made may not encode patterns or individual behaviours that accurately discriminate adherers from non-adherers.²⁸ Additionally, information on personal behaviours such as diet and exercise, structural barriers such as language and health literacy, and degree of social support, which was not available in our analysis, may be critical to adherence prediction.

Several methodological features may have also impeded our ability to predict adherence. By using purchases made prior to statin initiation, we did not observe how purchasing behaviours may change over time

following statin initiation, which in turn may help predict a patient’s longitudinal adherence pattern. We repeated our analyses including retail purchases made in the first 3 months after initiation as potential predictors and found very similar model discriminative ability. This suggests that postinitiation variables may not be informative. Additionally, the structure of the data set may have reduced our ability to model the strongest predictors of medication adherence. Investigator-specified purchasing variables leveraged only a brief product description available in the data, while the predictors selected by the hd-PS algorithm had a median prevalence of only 1%. A richer data set with greater information on product and pharmacy characteristics including location, combined with an analytical approach that includes time-varying prediction or network analysis, may better capture and characterise hypothesised behaviours and thus achieve better adherence prediction. Whereas some investigator-specified purchasing variables may have had an overlap in the products classified (eg, CVS branded products tend to be mostly healthcare-related), this would not have affected our models’ discriminative ability.

Nevertheless, several of the observed associations between retail purchasing behaviours and adherence could be useful for the development of new quality improvement interventions by pharmacies or payers. For example, the association between greater frequency of visits as well as retail purchasing on the same day as a medication fill and adherence may be markers of convenience of the pharmacy, an individual’s level of comfort in the store, or an underlying routine that adherent patients develop around pharmacy visits. In this way, identifying patients with irregular store visits and with store visits that do not coincide with a medication fill may represent an opportunity to link medication taking with other daily activities such as shopping. Specific interventions could target patients taking medications for chronic disease who have a high frequency of unhealthy food purchases for interventions tailored to adherence and dietary modification.

Several limitations to our study should be acknowledged. Members with loyalty cards may not use them at every purchase, and these data do not capture retail purchases patients make at non-CVS Pharmacy locations. Additionally, other family members may use an individual’s loyalty card. The former would result in an underestimate of store visit frequency, the latter in an overestimate, and both would result in misspecification of other variables in either direction, which, from a prediction perspective, would reduce model discrimination. Finally, behaviours of individuals who have obtained and regularly use a loyalty card may not be generalisable to all individuals using both the retail and pharmacy sections of the store, nor to patients who stopped being covered by Caremark during follow-up.

In summary, we found several retail purchasing behaviours to be associated with adherence to statins in the

year after initiation, some of which might be useful for the development of new quality improvement interventions by pharmacies or payers. In contrast, the use of retail purchasing data similar to those to which we had access appears not to meaningfully add to efforts to discriminate adherers from non-adherers.

Author affiliations

¹Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA

²Center for Public Health Research (CSISP-FISABIO) and REDISSEC, Valencia, Spain

³CVS Health, Woonsocket, Rhode Island, USA

Contributors All authors had access to study results and contributed meaningfully to the analysis. AAK, NKC, GS-G and JMF contributed to the study conception and design, interpretation of results, and manuscript drafting. GB prepared and analysed the data. DLI, WHS, TAB and OSM provided interpretation of results and critical manuscript revisions.

Funding This work was supported by an unrestricted grant from CVS Health to Brigham and Women's Hospital. All data analyses and outcomes assessment were performed independently of the study sponsor. WHS, TAB and OSM are employees of CVS Health and own stock in the company.

Competing interests TAB, OSM and WHS are employees of CVS Health and own stock options in the company.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Technical appendix, study protocol and statistical code available from researchers on request.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

- DiMatteo MR, Giordani PJ, Lepper HS, *et al.* Patient adherence and medical treatment outcomes: a meta-analysis. *Med Care* 2002;40:794–811.
- Roebuck MC, Liberman JN, Gemmill-Toyama M, *et al.* Medication adherence leads to lower health care use and costs despite increased drug spending. *Health Aff (Millwood)* 2011;30:91–9.
- Loewenstein G, Brennan T, Volpp KG. Asymmetric paternalism to improve health behaviors. *JAMA* 2007;298:2415–17.
- Kullgren JT, Troxel AB, Loewenstein G, *et al.* Individual-versus group-based financial incentives for weight loss: a randomized, controlled trial. *Ann Intern Med* 2013;158:505–14.
- Long JA, Jahnle EC, Richardson DM, *et al.* Peer mentoring and financial incentives to improve glucose control in African American veterans: a randomized trial. *Ann Intern Med* 2012;156:416–24.
- Checchi KD HK, Avorn J, Kesselheim AS. Electronic medication packaging devices and medication adherence: a systematic review. *JAMA* 2014;312:1237–47.
- Lam WY, Fresco P. Medication adherence measures: an overview. *Biomed Res Int* 2015;2015:217047.
- Choudhry NK, Fischer MA, Avorn J, *et al.* The implications of therapeutic complexity on adherence to cardiovascular medications. *Arch Intern Med* 2011;171:814–22.
- Gellad WF, Grenard JL, Marcum ZA. A systematic review of barriers to medication adherence in the elderly: looking beyond cost and regimen complexity. *Am J Geriatr Pharmacother* 2011;9:11–23.
- Mann DM, Woodward M, Muntner P, *et al.* Predictors of nonadherence to statins: a systematic review and meta-analysis. *Ann Pharmacother* 2010;44:1410–21.
- Osterberg L, Blaschke T. Adherence to medication. *N Engl J Med* 2005;353:487–97.
- Steiner JF. Can we identify clinical predictors of medication adherence... and should we? *Med Care* 2010;48:193–5.
- Dormuth CR, Patrick AR, Shrank WH, *et al.* Statin adherence and risk of accidents: a cautionary tale. *Circulation* 2009;119:2051–7.
- Jensen MK, Chiuve SE, Rimm EB, *et al.* Obesity, behavioral lifestyle factors, and risk of acute coronary events. *Circulation* 2008;117:3062–9.
- 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.
- De Domenico M, Lima A, Musolesi M. Interdependence and predictability of human mobility and social interactions. *Pervasive Mobile Comput* 2013;9:798–807.
- Krumme C, Llorente A, Cebrian M, *et al.* The predictability of consumer visitation patterns. *Sci Rep* 2013;3:1645.
- González MC, Hidalgo CA, Barabási AL. Understanding individual human mobility patterns. *Nature* 2008;453:779–82.
- Eagle N, Pentland AS, Lazer D. Inferring friendship network structure by using mobile phone data. *Proc Natl Acad Sci U S A* 2009;106:15274–8.
- Eagle N, Macy M, Claxton R. Network diversity and economic development. *Science* 2010;328:1029–31.
- Rassen JA, Glynn RJ, Brookhart MA, *et al.* Covariate selection in high-dimensional propensity score analyses of treatment effects in small samples. *Am J Epidemiol* 2011;173:1404–13.
- Schneeweiss S, Rassen JA, Glynn RJ, *et al.* High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology* 2009;20:512–22.
- Friedman J. Greedy function approximation: a gradient boosting machine. *Ann Stat* 2002;29:1189–536.
- Varian HR. Big data: new tricks for econometrics. *J Econ Perspect* 2014;28:3–28.
- Choudhry NK, Glynn RJ, Avorn J, *et al.* Untangling the relationship between medication adherence and post-myocardial infarction outcomes: medication adherence and clinical outcomes. *Am Heart J* 2014;167:51–8 e5.
- Trends in Part C & D Star Rating Measure Cut Points. 2013. (cited 8 May 2014). http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/Downloads/2014_Trends_CD_StarRating_Measure_CutPoints.pdf
- 2014 HEDIS Measures. 2014. (cited 8 May 2014). http://www.ncqa.org/Portals/0/HEDISQM/HEDIS2014/List_of_HEDIS_2014_Measures.pdf
- Franklin JM, Shrank WH, Lii J, *et al.* Observing versus predicting: initial patterns of filling predict long-term adherence more accurately than high-dimensional modeling techniques. *Health Serv Res* 2016;51:220–39.
- Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–87.
- Steyerberg EW, Harrell FE Jr, Borsboom GJ, *et al.* Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001;54:774–81.
- Lewey J, Shrank WH, Bowry AD, *et al.* Gender and racial disparities in adherence to statin therapy: a meta-analysis. *Am Heart J* 2013;165:665–78. 78 e1.
- Scheurer D, Choudhry N, Swanton KA, *et al.* Association between different types of social support and medication adherence. *Am J Manag Care* 2012;18:e461–7.
- Kesselheim AS, Bykov K, Avorn J, *et al.* Burden of changes in pill appearance for patients receiving generic cardiovascular medications after myocardial infarction: cohort and nested case-control studies. *Ann Intern Med* 2014;161:96–103.
- Chan DC, Shrank WH, Cutler D, *et al.* Patient, physician, and payment predictors of statin adherence. *Med Care* 2010;48:196–202.