Published in final edited form as:

Med Care. 2017 November ; 55(11): 956–964. doi:10.1097/MLR.0000000000000807.

Trajectories of Glycemic Change in a National Cohort of Adults with Previously Controlled Type 2 Diabetes

Rozalina G McCoy, MD MS^{1,2,3}, Che Ngufor, PhD⁴, Holly K Van Houten, BA³, Brian Caffo, PhD⁵, and Nilay D Shah, PhD^{2,6}

¹Division of Primary Care Internal Medicine, Department of Medicine, Mayo Clinic, Rochester, MN 55905

²Division of Health Care Policy & Research, Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905

³Mayo Clinic Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Rochester, MN 55905

⁴Division of Biomedical Statistics & Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905

⁵Department Of Biostatistics, Johns Hopkins University, Baltimore, MD, 21205

⁶OptumLabs, Cambridge MA 02142

Abstract

Background—Individualized diabetes management would benefit from prospectively identifying well-controlled patients at risk of losing glycemic control.

Objectives—To identify patterns of hemoglobin A_{1c} (HbA_{1c}) change among patients with stable controlled diabetes.

Research Design—Cohort study using OptumLabs Data Warehouse, 2001–2013. We develop and apply a machine learning framework that uses a Bayesian estimation of the mixture of

Corresponding author: Rozalina G. McCoy, MD. Division of Primary Care Internal Medicine, Department of Medicine, Mayo Clinic, 200 First Street SW, Rochester, MN 55905. Phone (507) 284-5160; fax (507) 266-0036; mccoy.rozalina@mayo.edu.

AUTHOR INFORMATION

Rozalina G McCoy, MD MS, 200 First Street SW, Rochester, MN 55905, Phone: (507) 284-5160, Fax: (507) 266-0036, mccoy.rozalinamayo.edu

Che Ngufor, PhD, 200 First Street SW, Rochester, MN 55905, Phone: 507-284-5582, Fax: 507-284-1731, ngufor.chemayo.edu

Holly K Van Houten, BA, 200 First Street SW, Rochester, MN 55905, Phone: (507) 284-5136, Fax: (507) 284-1731, vanhouten.hollymayo.edu

Brian Caffo, PhD, 615 N. Wolfe Street, Room E3610, Baltimore, Maryland 21205, Phone: (410) 955-3067, Fax: (410) 955-0958, bcaffowebjhu.edu

Nilay D Shah, PhD, 200 First Street SW, Rochester, MN 55905, Phone: 507-266-5130, Fax: 507-284-1731, shah.nilaymayo.edu

Conflict of Interest: The authors have no relevant conflict of interest to disclose.

Presentation: Parts of this work were presented as a poster presentation at the American Diabetes Association 76th Scientific Sessions in New Orleans, LA in June 2016 and as an oral presentation at the AcademyHealth Research Meeting in Boston, MA in June 2016.

SUPPLEMENTAL DIGITAL CONTENT

Supplemental Digital Content 1.docx

Supplemental Digital Content 2.docx

Supplemental Digital Content 3.docx

Supplemental Digital Content 4.docx

generalized linear mixed effect models to discover glycemic trajectories, and a random forest feature contribution method to identify patient characteristics predictive of their future glycemic trajectories.

Subjects—27,005 U.S. adults with type 2 diabetes, age ≥ 18 years, and stable index HbA_{1c} $< 7.0\%$.

Measures—HbA_{1c} values during 24 months of observation.

Results—We compared models with $k=1,2,3,4,5$ trajectories and baseline variables including patient age, sex, race/ethnicity, comorbidities, medications, and HbA_{1c}. The $k=3$ model had the best fit, reflecting three distinct trajectories of glycemic change: (T1) rapidly deteriorating HbA_{1c} among 302 (1.1%) youngest (mean, 55.2 years) patients with lowest mean baseline HbA_{1c}, 6.05%; (T2) gradually deteriorating HbA_{1c} among 902 (3.3%) patients (mean, 56.5 years) with highest mean baseline HbA_{1c}, 6.53%; and (T3) stable glycemic control among 25,800 (95.5%) oldest (mean, 58.5 years) patients with mean baseline HbA_{1c} 6.21%. After 24 months, HbA_{1c} rose to 8.75% in T1 and 8.40% in T2, but remained stable at 6.56% in T3.

Conclusions—Patients with controlled type 2 diabetes follow three distinct trajectories of glycemic control. This novel application of advanced analytic methods can facilitate individualized and population diabetes care by proactively identifying high risk patients.

Keywords

diabetes mellitus type 2; glycosylated hemoglobin; machine learning; Mixture of Generalized Linear Mixed Effects Model (MGLMM); random forest feature contribution (rffc) method; patient-centered medicine

INTRODUCTION

Clinical guidelines recommend intensifying glucose-lowering therapy to achieve hemoglobin A_{1c} (HbA_{1c}) $< 7.0\%$ for most adults with type 2 diabetes.^{1–3} Patients at highest risk for uncontrolled disease benefit from close monitoring and early intervention. Such efforts usually focus on patients with known risk factors for poor glycemic control: those with currently or recently uncontrolled disease, insulin-treated, younger, or with socioeconomic barriers to care.^{1–16} However, patients with controlled diabetes and without obvious risk factors may also experience glycemic deterioration and benefit from early identification and preemptive management to optimize long-term control.^{17, 18} Historically, such efforts were hindered by difficulty recognizing high-risk individuals, particularly as an individual patient's risk profile may differ from population-level estimates.

The development of advanced analytic methods and growing availability of electronic health-related data paved the way to address methodologic barriers to risk stratification. In particular, machine learning approaches based on data-driven learning algorithms can extract information from complex raw data rather than predefined *a priori* assumptions, but these methods are traditionally used for cross-sectional data, limiting their application to longitudinal clinical problems. In this study, we build on these methods to construct a novel machine learning framework for complex longitudinal data analysis, which would be useful in a wide range of clinical applications. Specifically, we employ a Bayesian estimation of the

generalized linear mixed model (GLMM), with the random effect component replaced by a mixture of normal distributions.¹⁹ This procedure enables data-driven discovery of homogenous longitudinal trajectories characterized by time-dependent covariates. We then apply the model to identify trajectories of HbA_{1c} change among adults with stable controlled type 2 diabetes and predict who will lose glycemic control. To determine which patient characteristics are most predictive of glycemic trajectories, we employ a novel random forest feature contribution method²⁰, which allows for the identification of important features at both the population and individual patient levels. Such information can be leveraged to characterize patient and disease phenotypes, study longitudinal risk factors and outcomes, and facilitate individualized patient-centered diabetes care.

METHODS

Study population

The study cohort consisted of 27,005 adults, age ≥ 18 years, with stable and controlled type 2 diabetes included in OptumLabs Data Warehouse (OLDW), a large administrative claims database of commercially-insured and Medicare Advantage beneficiaries across the U.S.^{21, 22} Diabetes was defined by Healthcare Effectiveness Data and Information Set criteria²³ ascertained using medical and pharmacy claims from 12 months preceding the index HbA_{1c}. Stability of glycemic control prior to cohort entry was ensured by (1) index HbA_{1c} <7.0% between January 1, 2001 and December 31, 2011; and (2) pre-index HbA_{1c} <7.0% immediately preceding the index HbA_{1c} and no earlier than 24 months. To focus on patients at lowest presumed risk of glycemic deterioration, we excluded people using insulin within 120 days prior to the index date; or with history of diabetic ketoacidosis (ICD-9-CM 250.1x), hyperglycemic hyperosmolar state (ICD-9-CM 250.2x), diabetic coma (ICD-9-CM 250.3x), severe hypoglycemia (ICD-9-CM 251.x, 250.8x²⁴), or poisoning by insulin or anti-diabetic drugs (ICD-9-CM 962.3) within one year prior to the index date. Patients were required to have ≥ 2 HbA_{1c} results during 24 months of follow-up and each patient entered the cohort once (see Figure, Supplemental Digital Content 1, which describes in detail each step of cohort assembly).

All study data were accessed after de-identification consistent with Health Insurance Portability and Accountability Act expert de-identification determination. Because this study involved analysis of pre-existing, de-identified data, it was exempt from Institutional Review Board approval.²⁵

Independent variables

Variables used in the analysis were selected based on what is commonly available in the electronic health records, obtained as part of routine care, and not contingent on a concomitant face-to-face clinical encounter. This includes patient age (at the time of the index HbA_{1c}); sex; race/ethnicity; Charlson/Deyo comorbidities²⁶ documented within 12 months of the index date; geographic region; medications filled 120 days before to 24 months after the index date (but only medications prior to and at the time of the index HbA_{1c} were used in the models); and the index, pre-index, and all other HbA_{1c} values during 24 months of follow-up). HbA_{1c} results were identified using LOINC codes 4548-4, 4549-2,

17856-6, 59261-8, 62388-4 and 4547-6. Diabetes medications were grouped into nine classes (see Table, Supplemental Digital Content 2, which categorizes and lists diabetes medications).

We included a measure of HbA_{1c} testing frequency during 24-months of follow-up, categorized as 2, 3–4, or 5 tests per year, as short-term testing and monitoring plans are determined *a priori* and may influence glycemic control.²⁷ Medication changes at the time of the index HbA_{1c} were recorded, classified as no change, class change, intensification (addition of 1 drug or insulin start), or de-intensification (removal of 1 drug). Treatment change during follow up was not ascertained, as this data would not be available at the point-of-care when trajectory class assignment would be made.

Statistical analysis

Data exploration—Basic exploration of longitudinal HbA_{1c} profiles of each patient suggested that the evolution of each profile can be modeled by a linear function over time where the parameters of the function may differ across patients. We therefore selected the GLMM with random intercept and slope on the time variable.

Bayesian Mixture of Generalized Linear Mixed Effects Model (MGLMM)—

Because effect of patient characteristics on HbA_{1c} may vary across time and patients, we employed a Bayesian approximation of the mixture extension of the GLMM where the random effect component is assumed to follow a mixture of normal distributions.¹⁹ The MGLMM specifies that there is unobserved clustering of random effect parameters across an unknown number of population subgroups (trajectories) with different weights. Model parameters of MGLMM were estimated using the Markov Chain Monte Carlo (MCMC) method and implemented using the R package mixAK,¹⁹ which has the advantage that if the computation fails (i.e. does not converge), then model assumptions are violated²⁸ indicating that the proposed model is not appropriate for the data and should be rejected. We assumed weakly informative priors for all fixed effects parameters, set the mixture means to 0, and assumed equal probability for the mixture weights (i.e. equal *a priori* probability of assignments of patients to trajectories). We assessed convergence of the posterior distribution of the parameters of MGLMM in the MCMC simulation through traceplots and autocorrelation plots, and all indicated proper convergence.

Data-driven model validation and selection of the optimal number of HbA_{1c} trajectories—

We assessed robustness and stability of the models and selected the optimal number of trajectories through two “internal” and one “external” validation techniques.²⁹ The internal validation methods are based on information used to generate the models without reference to external information, while the external method uses the derived trajectories. We used the external method to assess the quality of the trajectory classification and not the fit of the MGLMM model, a procedure that is best assessed by the internal methods and the convergence of the MCMC.

The penalized expected deviance (PED) and posterior distribution of deviances were used as internal validation statistics. The PED³⁰ is the most commonly used Bayesian measure to assess model fit, where lower values indicate better model. However, model selection based

on a single criteria such as the PED ignores uncertainty in the comparison, which increases with increasing number of model parameters.^{19, 31} An alternative approach is to assess the full posterior distribution of the deviances of the models. Thus, suppose *Diff* is the difference in deviance between a large and a small model, if the posterior probability $P(\text{Diff} < -2\log(9) = -4.39)$ is 0.9, then there is strong evidence in favor of the larger model.³¹ We computed these statistics by fitting five MGLMM models with the number of mixture components set to $k = 1, \dots, 5$.

For external validation, we used the random forest feature contribution (rfFC) method to predict a patient's trajectory based on new data and assessed how each variable contributes towards this classification. The assumption is that similar patients at baseline will follow the same trajectory, and if trajectory assignments by the MGLMM model are accurate, this will be reflected by high accuracy of the prediction model. Random forest³² is a machine learning method that constructs multiple decision trees on a bootstrap sample of the training data and then combines the predictions by averaging or majority voting. This produces a ranking score for each variable in the model to reflect its importance to the prediction task. This ranking score can help in model interpretation or dimension reduction; however, it is a population average score that is not useful for assessing the importance of individual patient characteristics. A variable may be an important risk factor for the population, but not affect a specific individual or subgroup.

The rfFC²⁰ is a recent extension of the random forest variable importance (or feature contribution [FC]) for each individual in the training data, with the additional benefit that the FC can also be predicted for new data. This allows to further assess the accuracy of the rfFC model and MGLMM trajectory classifications for each patient, as a good match between the average FCs of the training and test sets reflects a high quality model.²⁰ FC=0 indicates that the variable is irrelevant with respect to assigning a patient to a given trajectory, positive values indicate that the variable is influential toward classification to the reference trajectory (in our study, T3), and negative values indicate that the variable is influential towards classification to the other trajectories. We used data at index HbA_{1c} for training and data for 0–3 subsequent months for testing. For each medication, we created an indicator variable to indicate if that medication was filled at least once in the 0–3 months. A 5-fold cross-validation training procedure was implemented and the best model in terms of accuracy was used for prediction on the test set.

We also used the rfFC method to quantify FCs for one randomly selected patient from each trajectory. These *patient-level* FCs were qualitatively compared to the median (or *population-level*) FCs obtained for the whole study sample.

To characterize the trajectory differences in patient baseline characteristics, we used the Pearson's chi-squared test for categorical variables and the Kruskal-Wallis H-test for continuous variables. All statistical tests were performed at 5% level of significance.

RESULTS

The study cohort included 27,005 adults with type 2 diabetes and mean index HbA_{1c} 6.22%. Included patients were, on average, 58.4 years old, 51.1% were male, and 57.6% were white; Table 1. At baseline, 34.4% of patients were treated with lifestyle therapy; 37.6%, with one glucose-lowering medication; 21.4%, with two drugs; and 6.6%, with 3 drugs. The majority of patients were from South U.S. (55.7%), with the remainder from the Northeast (25.6%), Midwest (11.0%), and West (7.7%).

Trajectories of Glycemic Change

Implementation of the Bayesian MGLMM with $k=1,2,3,4,5$ possible mixture components supported a three trajectory model (Figure 1) based on its low PED, lowest median difference in FCs between the training and test datasets, and principles of parsimony and interpretability. Models with $k=1$ and $k=2$ trajectories had much higher PED than models with $k=3,4,5$ trajectories; Appendix Table 1. Models with $k=4$ and $k=5$ trajectories were redundant, with $P(\text{Diff} < -4.39) = 0.9$ and HPD interval of the mean difference in the deviance spanning 0; Appendix Table 2.

With respect to the rFC model, models with $k=4$ and $k=5$ resulted in larger mismatch in FCs between the training and test datasets compared to the $k=3$ model (see Figure, Supplemental Digital Content 3). In addition, the accuracy and positive predictive values for the best $k=3,4,5$ rFC models selected through cross-validation were: 0.88, 0.82, 0.87 and 0.54, 0.41, 0.45 respectively, thus indicating the superiority of the $k=3$ model. It should be noted that the rFC method was used in this study as a validation and interpretation tool, and not for predicting trajectories; predictions are handled by the MGLMM model. As such, these performance measures should be interpreted in terms of selecting the optimal number of trajectories and not for predicting trajectories for new patients.

The first trajectory (T1), comprised of 302 (1.1%) patients, was notable for rapid deterioration of glycemic control with mean HbA_{1c} rising from 6.05% to 8.75% in 24 months. T2 included 903 (3.4%) patients who demonstrated slower deterioration of glycemic control with mean HbA_{1c} rising from 6.53% to 8.40%. T3 was the most common and included 25,800 patients (95.5%) who maintained glycemic control, with mean HbA_{1c} changing slightly from 6.21% to 6.56%.

Patient Phenotypes

Patients following the three trajectories differed significantly in age, sex, and index HbA_{1c} ($p < 0.001$ for all); Table 1. T1 patients were younger (55.2 years), more likely to be male (62.6%), and had lower mean index HbA_{1c} (6.05%) compared to patients in the other two trajectories. In contrast, T3 patients were older (58.5 years), less likely to be male (50.7%), and had mean index HbA_{1c} 6.21%. They were similar in their racial, geographic, and comorbidity profiles.

Baseline medication use differed among patients in the three trajectories. Patients in T1 and T2 were treated with monotherapy (33.8% and 31.6%, respectively), two drugs (27.5% and 30.8%, respectively), or 3 drugs (13.6% and 13.4%, respectively). Patients in T3 were most

likely to be controlled with one (37.9%) or no (34.9%) glucose-lowering medications. Metformin was the most common class of medications in all three trajectories, used by 55.3% of patients in T1, 59.1% of patients in T2, and 48.4% of patients in T3. Sulfonylureas were used most often by patients in T1 (42.1%), while DPP-4 inhibitors and GLP-1 receptor agonists were most commonly used by patients in T2 (12.0% and 3.0%, respectively).

Figure 2 presents the *population-level* FCs in classifying patients into the three trajectories, with T3 as the referent group. Figure 2 also demonstrates FCs for three randomly selected patients (P1, P2, P3), one from each trajectory, in order to demonstrate that *individual* patient FCs may, and do, differ from the *population-level* FCs, and that this analytic method can identify and predict these differences. The *population-level* FCs indicate that patient age, pre-index HbA_{1c}, and metformin and sulfonylurea/glinide use were most influential towards T3, with age and pre-index HbA_{1c} being the strongest predictors toward any trajectory. In contrast, the decision not to alter glucose-lowering therapy at the index date was most influential towards T2. However, when we considered the FCs within individual patients, we saw interesting departures from the *population-level*, demonstrating that risk factors can affect individual patients differently. For example, age was not influential toward P3; and low Charlson comorbidity index and metformin use were influential toward P1 but not T1.

Diabetes Management in Each Trajectory

Treatment regimens were variably intensified among patients in all three trajectories; Figure 3 and (see Table, Supplemental Digital Content 4, which presents the number of medications used over time among patients in the three glycemic trajectories). The fraction of patients taking no glucose-lowering medications was constant in T3 (around 30%), decreased steadily in T2 (24.3% to 16.7%), and decreased but then rose again in T1 (25.2% at index date, 14.5% at 15–18 months, 25.7% at the end). The fraction of patients treated with one-medication regimens decreased steadily in T1 (33.8% to 24.6%) and T2 (31.6% to 25.3%) as patients were intensified to 2 medications, but remained stable in T3 (around 38%). Two-medication regimens increased and then decreased in T1 (27.5% at index date, 35.0% at 15–18 months, 27.3% at the end), but increased steadily in T2 (21.0% to 35.8%). Use of 3 medications increased in all trajectories: 13.6% to 22.5% in T1; 13.4% to 22.3% in T2; and 6.3% to 7.8% in T3.

DISCUSSION

Patient-centered diabetes care is predicated on proactively identifying patients at risk for deterioration of glycemic control. While current HbA_{1c} is an important predictor of future glycemic control, patients with controlled disease may still experience rise in HbA_{1c} requiring clinical attention. In this study, we developed a machine learning framework to identify three distinct trajectories of glycemic change among 27,005 presumably low risk adults with stable controlled type 2 diabetes, 5% of whom experienced marked deterioration of glycemic control with HbA_{1c} approaching 9% within two years. This algorithm has the potential to be used across a wide spectrum of settings and practices as it is based on variables obtained during routine care, available in electronic health records, and not predicated on a concurrent face-to-face encounter.

As expected, the vast majority (95%) of patients maintained glycemic control, with mean HbA_{1c} increasingly negligibly from 6.21% to 6.56% over two years. These patients were more often older, female, and treated with lifestyle therapy at cohort entry. However, a small but clinically important subset of patients experienced marked deterioration of glycemic control. HbA_{1c} rapidly rose from 6.05% to 8.75% in 1.1% of patients who were younger, more likely to be male, and more likely to be treated with multiple glucose-lowering drugs at cohort entry. While the majority of these patients were ultimately taking 2 drugs (this variable was not included in the predictive model as these treatment changes occurred after cohort entry), there was a late rise in the percent of patients taking fewer or no medications despite HbA_{1c} exceeding 8%, suggesting potential non-adherence to therapy. In another 3.3% of patients (T2), HbA_{1c} rose gradually from 6.53% to 8.40%. These patients were less likely than patients in T1 to take 3 diabetes medications despite elevated HbA_{1c}, and were least likely to have their treatment intensified at the time of index HbA_{1c}, potentially reflecting inadequate treatment intensification and/or nonadherence to glucose-lowering polypharmacy. The frequency of HbA_{1c} testing also differed among the groups. Stably controlled patients in T3 were most likely to be tested twice annually, as recommended by most clinical guidelines,¹⁻³ though nearly two-thirds were tested quarterly or more often. Patients losing glycemic control were monitored more frequently, as recommended by the guidelines,¹⁻³ but nearly 7% of rapidly deteriorating patients in T1 still had their HbA_{1c} checked once or twice per year. Time to reassessment is determined *a priori* (e.g. during the index clinical encounter) based on the anticipated likelihood of clinical change based on information available at the time. The lack of close monitoring of patients in T1 and T2, which may be due to patient-, provider-, or system-factors, may have contributed to their poor glycemic outcomes.

Our study confirmed several previously observed characteristics associated with loss of glycemic control, including younger age, male sex, and need for multiple glucose-lower medications.⁵⁻⁸ Such patients may benefit from educational and clinical resources to help manage their diabetes irrespective of their HbA_{1c}, including multidisciplinary programs targeted at medication adherence, education, and self-management.⁴

Administrative data does not allow for ascertainment of the reasons why glycemic control deteriorated, such as failure of treatment intensification, non-adherence, adverse effects or contraindications to intensive therapy, lack of access to healthcare, or competing demands. Importantly, none of the patients in the study had documented severe hypoglycemia (per study exclusion criteria), and there was no difference in comorbidity burden among patients in the three trajectories, so it is unlikely that different treatment goals or contraindications to intensive therapy explain the entirety of differences in glycemic control and treatment. Furthermore, all patients had commercial insurance and thus access to healthcare, though different benefit designs may have placed some patients at greater financial burden than others. Our study dataset and models were further limited by absence of psychosocial information, which has been shown to be important in identifying glycemic trajectories in type 1 diabetes,⁹⁻¹² and may similarly contribute to glycemic patterns among adults with type 2 diabetes. Administrative data also does not capture lifestyle habits and weight changes, which can contribute to glycemic change. Race/ethnicity did not contribute significantly to trajectory assignment, such that commonly observed disparities in glycemic

control⁸ may reflect other important risk factors such as access to healthcare and medications.^{13–15} Our study was conducted among commercially insured and Medicare Advantage beneficiaries, so the results may not fully generalize to uninsured or publicly-insured patients or to situations where social determinants of health preclude adequate access to health care.¹⁶ Nonetheless, the objective of this study was to develop a modeling approach that would be robust even with limited data, and therefore focused on variables routinely collected and electronically catalogued in clinical practice.

Prior studies of longitudinal glycemic control have focused almost exclusively on children and adolescents with type 1 diabetes.^{9–12, 33} One study was conducted among adults with type 2 diabetes, but did not focus on controlled patients.³⁴ By applying a novel analytic framework to administrative and laboratory data from a large and diverse national cohort of adults with type 2 diabetes, our study allowed for data-driven identification of HbA_{1c} trajectories and discovery of potentially meaningful patient- and population-level risk factors that warrant further evaluation. Furthermore, the innovative analytic approach used in this study distinguished between *population-level* FCs, representing population-level risk factors, and *individual* FCs specific to the unique patient. Classical statistical analyses cannot simultaneously provide population and individual risk assessment, yet both are needed for clinical practice. In the current era of proactive population health management and individualized medicine, such analyses can leverage electronic data sources for improved patient care and experience. Evaluation of this approach in clinical practice is needed and is planned for the future.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: Dr. McCoy is supported by the Mayo Clinic Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery and by the National Institute Of Diabetes And Digestive And Kidney Diseases of the National Institutes of Health under Award Number K23DK114497. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Garber AJ, Abrahamson MJ, Barzilay JI, Blonde L, Bloomgarden ZT, Bush MA, et al. Consensus Statement by The American Association Of Clinical Endocrinologists And American College Of Endocrinology On The Comprehensive Type 2 Diabetes Management Algorithm – 2016 Executive Summary. *Endocrine Practice*. 2016; 22(1):84–113. [PubMed: 26731084]
2. American Diabetes Association Standards of Medical Care in Diabetes. *Diabetes Care*. 2016; 39(suppl 1):S39–46. [PubMed: 26696679]
3. National Institute for Health and Care Excellence (NICE). [cited 2015 July 12] Type 2 Diabetes: The Management Of Type 2 Diabetes. 2009. Available from: <http://www.nice.org.uk/guidance/cg87/chapter/1-recommendations#glucose-control-levels>
4. Tricco AC, Ivers NM, Grimshaw JM, Moher D, Turner L, Galipeau J, et al. Effectiveness Of Quality Improvement Strategies On The Management Of Diabetes: A Systematic Review And Meta-Analysis. *The Lancet*. 379(9833):2252–61.

5. Pani LN, Nathan DM, Grant RW. Clinical Predictors Of Disease Progression And Medication Initiation In Untreated Patients With Type 2 Diabetes And A1C Less Than 7%. *Diabetes Care*. 2008; 31(3):386–90. [PubMed: 18083790]
6. Blumenthal KJ, Larkin ME, Winning G, Nathan DM, Grant RW. Changes In Glycemic Control From 1996 To 2006 Among Adults With Type 2 Diabetes: A Longitudinal Cohort Study. *BMC Health Serv Res*. 2010; 10:158. [PubMed: 20534158]
7. Benoit SR, Fleming R, Philis-Tsimikas A, Ji M. Predictors Of Glycemic Control Among Patients With Type 2 Diabetes: A Longitudinal Study. *BMC Public Health*. 2005; 5:36. [PubMed: 15833140]
8. Ayanian JZ, Landon BE, Newhouse JP, Zaslavsky AM. Racial And Ethnic Disparities Among Enrollees In Medicare Advantage Plans. *The New England journal of medicine*. 2014; 371(24): 2288–97. [PubMed: 25494268]
9. Hilliard ME, Wu YP, Rausch J, Dolan LM, Hood KK. Predictors Of Deteriorations In Diabetes Management And Control In Adolescents With Type 1 Diabetes. *The Journal of adolescent health : official publication of the Society for Adolescent Medicine*. 2013; 52(1):28–34. [PubMed: 23260831]
10. Luyckx K, Seiffge-Krenke I. Continuity And Change In Glycemic Control Trajectories From Adolescence To Emerging Adulthood: Relationships With Family Climate And Self-Concept In Type 1 Diabetes. *Diabetes Care*. 2009; 32(5):797–801. [PubMed: 19228859]
11. Helgeson VS, Snyder PR, Seltman H, Escobar O, Becker D, Siminerio L. Brief Report: Trajectories Of Glycemic Control Over Early To Middle Adolescence. *J Pediatr Psychol*. 2010; 35(10):1161–7. [PubMed: 20189951]
12. Rohan JM, Rausch JR, Pendley JS, Delamater AM, Dolan L, Reeves G, et al. Identification And Prediction Of Group-Based Glycemic Control Trajectories During The Transition To Adolescence. *Health Psychol*. 2014; 33(10):1143–52. [PubMed: 24274797]
13. Berkowitz SA, Meigs JB, DeWalt D, Seligman HK, Barnard LS, Bright OJ, et al. Material Need Insecurities, Control Of Diabetes Mellitus, And Use Of Health Care Resources: Results Of The Measuring Economic Insecurity In Diabetes Study. *JAMA Intern Med*. 2015; 175(2):257–65. [PubMed: 25545780]
14. Hu R, Shi L, Rane S, Zhu J, Chen CC. Insurance, Racial/Ethnic, Ses-Related Disparities In Quality Of Care Among Us Adults With Diabetes. *J Immigr Minor Health*. 2014; 16(4):565–75. [PubMed: 24363118]
15. Kim G, Ford KL, Chiriboga DA, Sorkin DH. Racial And Ethnic Disparities In Healthcare Use, Delayed Care, And Management Of Diabetes Mellitus In Older Adults In California. *J Am Geriatr Soc*. 2012; 60(12):2319–25. [PubMed: 23194086]
16. Brown AF, Ettner SL, Piette J, Weinberger M, Gregg E, Shapiro MF, et al. Socioeconomic Position And Health Among Persons With Diabetes Mellitus: A Conceptual Framework And Review Of The Literature. *Epidemiologic Reviews*. 2004; 26:63–77. [PubMed: 15234948]
17. Turner RC, Cull CA, Frighi V, Holman RR. for the United Kingdom Prospective Diabetes Study (UKPDS) Study Group. Glycemic Control With Diet, Sulfonylurea, Metformin, Or Insulin In Patients With Type 2 Diabetes Mellitus: Progressive Requirement For Multiple Therapies (UKPDS 49). *JAMA*. 1999; 281(21):2005–12. [PubMed: 10359389]
18. Best JD, Drury PL, Davis TM, Taskinen MR, Kesaniemi YA, Scott R, et al. Glycemic Control Over 5 Years In 4,900 People With Type 2 Diabetes: Real-World Diabetes Therapy In A Clinical Trial Cohort. *Diabetes Care*. 2012; 35(5):1165–70. [PubMed: 22432105]
19. Komarek A, Komarkova L. Clustering For Multivariate Continuous And Discrete Longitudinal Data. *Annals of Applied Statistics*. 2013; (1):177–200.
20. Palczewska, A., Palczewski, J., Robinson, RM., Neagu, D. Interpreting Random Forest Models Using A Feature Contribution Method. *Information Reuse and Integration (IRI); 2013 IEEE 14th International Conference on; 2013 14–16 Aug.; 2013*. p. 112-9.2013
21. Wallace PJ, Shah ND, Dennen T, Bleicher PA, Crown WH. Optum Labs: Building A Novel Node In The Learning Health Care System. *Health Aff (Millwood)*. 2014; 33(7):1187–94. [PubMed: 25006145]

22. Optum. [cited 2016 May 23] Optum Research Data Assets. 2015 Dec 24. June 2015:[Available from: https://www.optum.com/content/dam/optum/resources/productSheets/5302_Data_Assets_Chart_Sheet_ISPOR.pdf
23. [Accessed on 11/11/2014] National Committee for Quality Assurance (NCQA) Healthcare Effectiveness Data and Information Set (HEDIS) 2013 Diabetes Mellitus Measures. at http://www.ncqa.org/Portals/0/HEDISQM/DM_2013_Measures_9.13.12.pdf. [Website] 2013 [cited 06/18/2013]; Available from: http://www.ncqa.org/Portals/0/HEDISQM/DM_2013_Measures_9.13.12.pdf
24. Ginde AA, Blanc PG, Lieberman RM, Camargo CA Jr. Validation Of ICD-9-CM Coding Algorithm For Improved Identification Of Hypoglycemia Visits. *BMC Endocr Disord*. 2008; 8:4. [PubMed: 18380903]
25. [Accessed on March 16, 2017] Privacy Of Individually Identifiable Health Information: Other Requirements Relating To Uses And Disclosures Of Protected Health Information. 45 CFR 164.514(b)(1). Jun 7. 2013 Available at: http://www.ecfr.gov/cgi-bin/retrieveECFR?gp=1&SID=ff85fb032e00c1b6d135ecf0767f18a0&ty=HTML&h=L&mc=true&r=PART&n=pt45.1.164#se45.1.164_1514. March 14, 2017 [cited March 16, 2017]; Available from:
26. Deyo RA, Cherkin DC, Ciol MA. Adapting A Clinical Comorbidity Index For Use With ICD-9-CM Administrative Databases. *J Clin Epidemiol*. 1992; 45(6):613–9. [PubMed: 1607900]
27. McCoy RG, Van Houten HK, Ross JS, Montori VM, Shah ND. HbA1c Overtesting And Overtreatment Among Us Adults With Controlled Type 2 Diabetes, 2001–13: Observational Population Based Study. *BMJ*. 2015; 351:h6138. [PubMed: 26646052]
28. Gelman, A. Bayesian data analysis. 2. Boca Raton, Fla.: Chapman & Hall/CRC; 2004.
29. Tan, P-N., Steinbach, M., Kumar, V. Introduction to Data Mining. first ed. Pearson Addison-Wesley; Boston, MA, USA: 2006.
30. Plummer M. Penalized Loss Functions For Bayesian Model Comparison. *Biostatistics*. 2008; 9(3): 523–39. [PubMed: 18209015]
31. Aitkin M, Liu CC, Chadwick T. Bayesian Model Comparison And Model Averaging For Small-Area Estimation. *Annals of Applied Statistics*. 2009; (1):199–221.
32. Breiman L. Random Forests. *Mach Learn*. 2001; 45(1):5–32.
33. Clements MA, Lind M, Raman S, Patton SR, Lipska KJ, Fridlington AG, et al. Age At Diagnosis Predicts Deterioration In Glycaemic Control Among Children And Adolescents With Type 1 Diabetes. *BMJ Open Diabetes Res Care*. 2014; 2(1):e000039.
34. Walraven I, Mast MR, Hoekstra T, Jansen AP, van der Heijden AA, Rauh SP, et al. Distinct HbA1c Trajectories In A Type 2 Diabetes Cohort. *Acta diabetologica*. 2015; 52(2):267–75. [PubMed: 25287012]

APPENDIX

TABLE 1
Penalized expected deviance (PED) for models with
 $k=1,2,3,4,5$ components

Smaller (more negative) PED is indicative of better model fit.

Number of trajectories	E(D)	p(opt)	PED
K=1	−89400.99	94.808239	−89306.18
K=2	−91875.42	106.93645	−91768.48
K=3	−92173.83	115.26776	−92058.56
K=4	−92247.88	133.5957	−92114.28
K=5	−92279.32	160.96813	−92118.36

E(D) = posterior mean deviance; p(opt)=penalty term or optimism.

TABLE 2
Posterior summary of the difference in deviances
between models with $k = 1, 2, 3, 4, 5$ components

Higher absolute mean difference in deviances indicates greater divergence between the compared models. The highest posterior density (HPD) is a Bayesian credible interval that is analogous to confidence intervals used in classical statistics. In this analysis, which evaluates for differences between two models, HPD intervals that span 0 are not significantly different from each other.

Comparison of models with k trajectories	P(Diff < -4.39)	Mean	HPD interval	
			Lower	Upper
P₂₁	1.0	-2474.43	-2503.41	-2445.23
P₃₁	1.0	-2772.84	-2801.99	-2743.70
P₄₁	1.0	-2846.89	-2878.56	-2811.72
P₅₁	1.0	-2878.33	-2912.13	-2833.38
P₃₂	1.0	-298.41	-329.50	-267.43
P₄₂	1.0	-372.46	-406.94	-332.62
P₅₂	1.0	-403.91	-439.95	-352.23
P₄₃	1.0	-74.05	-109.29	-35.42
P₅₃	1.0	-105.50	-142.67	-54.93
P₅₄	0.9	-31.44	-67.31	6.54

FIGURE 1. Trajectories of glycemic change over time

Mean HbA_{1c} levels of patients following trajectories T1 (red), T2 (blue), and T3 (green) are shown in bold; individual patient HbA_{1c} levels are presented in the background. Table below shows the mean HbA_{1c} values of patients in each trajectory in 3-months increments and the net change in mean HbA_{1c} from baseline (index HbA_{1c}).

FIGURE 2. Median and individual patient feature contributions

Median feature contributions toward T3 (trajectory of stable glycemic control; green) as opposed to T1 (trajectory of rapidly deteriorating glycemic control; red) or T2 (slowly deteriorating glycemic control; blue) are shown in solid color. Individual feature contributions for three randomly selected patients, one from each trajectory (P1, P2, and P3), are also shown to demonstrate potential differences between median (e.g. population-level) and individual (e.g. patient-level) feature contributions identified by the rFFC method. T3 is the referent group in this analysis; as such, positive feature contributions indicate support for T3 as opposed to T1 or T2, and negative feature contributes indicate support of T1 or T2 as opposed to T3. Feature contribution values have been scaled by 100 to facilitate visualization.

FIGURE 3. Glucose-lowering therapy over time

Change in the number of glucose-lowering medications taken by patients in the three glycemic trajectories over the study period. Medication use was ascertained by pharmacy fill data in 3-months intervals. Data points have been smoothed using locally weighted regression scatterplot smoothing (Loess); shaded areas represent the 95% confidence interval. Loess is a nonparametric technique that provides a graphical summary of the relationship between a dependent and one or more independent variables and is used as a scatterplot smoother.

TABLE 1

Study cohort

Patient characteristics at the time of the index HbA_{1c} test, overall and subset into three trajectories of glycemic change during the 24-months study period following index HbA_{1c}. Data are shown as percent, unless stated otherwise. Per study inclusion criteria, both the index and the pre-index HbA_{1c} test results had to be less than 7.0%.

	Overall	T1	T2	T3	P-value
Patients, N (%)	27,005	302 (1.12%)	902 (3.34%)	25,800 (95.54%)	
HbA_{1c} %, mean					
Index HbA _{1c}	6.22	6.05	6.53	6.21	<0.001
Pre-index HbA _{1c}	6.22	6.27	6.31	6.22	<0.001
Age, years, mean	58.38	55.18	56.54	58.48	
Sex					
Male	51.08	62.58	56.92	50.74	<0.001
Female	48.92	37.42	43.08	49.26	<0.001
Race/ethnicity					
White	57.58	59.27	58.91	57.51	1.00
Black	14.70	16.56	14.06	14.71	1.00
Hispanic	10.98	10.93	11.63	10.95	1.00
Asian	7.49	4.64	6.53	7.56	1.00
Other/unknown	9.25	8.61	8.86	9.27	1.00
Geographic region					
Northeast	25.60	25.50	21.82	25.73	0.83
Midwest	11.01	12.25	12.74	10.94	1.00
South	55.71	55.30	56.59	55.68	1.00
West	7.68	6.95	8.86	7.64	1.00
Frequency of HbA_{1c} testing					
2 times/year	34.77	6.62	2.10	36.24	<0.001
3–4 times/year	58.94	76.16	84.05	57.86	<0.001
5 times/year	6.29	17.22	13.84	5.90	<0.001

	Overall	T1	T2	T3	P-value
Regimen change at the index HbA_{1c} test					
Maintain lifestyle	27.00	16.56	16.06	27.50	<0.001
Maintain drugs	54.42	50.00	60.13	54.27	0.02
Change drugs	0.83	0.66	1.44	0.81	1.00
Intensify	8.41	15.89	11.74	8.21	<0.001
De-intensify	9.35	16.89	10.63	9.22	<0.001
Comorbidities					
Myocardial infarction	1.23	1.32	1.88	1.21	1.00
Heart failure	2.28	3.31	2.88	2.25	1.00
Peripheral vascular disease	7.36	7.62	7.53	7.35	1.00
Stroke	6.51	7.28	6.64	6.50	1.00
Kidney disease	2.08	2.98	3.54	2.02	1.00
Charlson/Deyo index					
0-1	61.20	55.63	56.81	61.41	0.09
2	15.29	14.57	15.28	15.30	1.00
3	13.86	18.21	16.72	13.71	0.10
4	9.65	11.59	11.18	9.58	1.00
Baseline drug class					
Metformin	48.80	55.30	59.14	48.36	<0.001
Sulfonylurea/glinide	23.71	42.05	37.54	23.02	<0.001
Thiazolidinediones	18.54	21.52	23.15	18.35	0.02
DPP-4 inhibitors	7.08	8.94	11.96	6.89	<0.001
GLP-1 receptor analogs	2.42	2.65	2.99	2.40	1.00
Other	0.29	0.66	0.11	0.29	1.00
Baseline regimen					
Lifestyle (no drugs)	34.41	25.17	24.25	34.87	<0.001
1 drug	37.61	33.77	31.56	37.87	<0.001
2 drugs	21.41	27.48	30.79	21.01	<0.001
3 drugs	6.57	13.58	13.40	6.25	<0.001

DPP-4, dipeptidyl peptidase 4; GLP-1, Glucagon-Like Peptide-1.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript