

Original Paper

# Data Mining of a Remote Behavioral Tracking System for Type 2 Diabetes Patients: A Prospective Cohort Study

Noah Wayne<sup>1\*</sup>, PhD, RKin; Nick Cercone<sup>2†\*</sup>, PhD; Jiye Li<sup>2\*</sup>, PhD; Ariel Zohar<sup>1\*</sup>, B BSc (Hons), RKin; Joel Katz<sup>3\*</sup>, PhD (CPsych); Patrick Brown<sup>4,5\*</sup>, PhD; Paul Ritvo<sup>1\*</sup>, PhD (CPsych)

<sup>1</sup>Health Behaviour Change Lab, School of Kinesiology & Health Science, York University, Toronto, ON, Canada

<sup>2</sup>Data Mining Lab, Lassonde School of Engineering, York University, Toronto, ON, Canada

<sup>3</sup>Pain Mechanisms Lab, Department of Psychology, York University, Toronto, ON, Canada

<sup>4</sup>Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

<sup>5</sup>Analytics and Informatics, Prevention and Cancer Control, Cancer Care Ontario, Toronto, ON, Canada

†deceased

\*all authors contributed equally

**Corresponding Author:**

Paul Ritvo, PhD (CPsych)

Health Behaviour Change Lab

School of Kinesiology & Health Science

York University

136 Chemistry Building

4700 Keele Street

Toronto, ON, M3J1P3

Canada

Phone: 1 416 736 2100 ext 22396

Fax: 1 416 736 5774

Email: [pritvo@yorku.ca](mailto:pritvo@yorku.ca)

## Abstract

**Background:** Complications from type 2 diabetes mellitus can be prevented when patients perform health behaviors such as vigorous exercise and glucose-regulated diet. The use of smartphones for tracking such behaviors has demonstrated success in type 2 diabetes management while generating repositories of analyzable digital data, which, when better understood, may help improve care. Data mining methods were used in this study to better understand self-monitoring patterns using smartphone tracking software.

**Objective:** Associations were evaluated between the smartphone monitoring of health behaviors and HbA1c reductions in a patient subsample with type 2 diabetes who demonstrated clinically significant benefits after participation in a randomized controlled trial.

**Methods:** A priori association-rule algorithms, implemented in the C language, were applied to app-discretized use data involving three primary health behavior trackers (exercise, diet, and glucose monitoring) from 29 participants who achieved clinically significant HbA1c reductions. Use was evaluated in relation to improved HbA1c outcomes.

**Results:** Analyses indicated that nearly a third (9/29, 31%) of participants used a single tracker, half (14/29, 48%) used two primary trackers, and the remainder (6/29, 21%) of the participants used three primary trackers. Decreases in HbA1c were observed across all groups (0.97-1.95%), but clinically significant reductions were more likely with use of one or two trackers rather than use of three trackers (OR 0.18,  $P=.04$ ).

**Conclusions:** Data mining techniques can reveal relevant coherent behavior patterns useful in guiding future intervention structure. It appears that focusing on using one or two trackers, in a symbolic function, was more effective (in this sample) than regular use of all three trackers.

(JMIR Diabetes 2016;1(1):e1) doi:[10.2196/diabetes.4506](https://doi.org/10.2196/diabetes.4506)

**KEYWORDS**

diabetes mellitus, type 2; health coaching; mhealth; telehealth; data mining

## Introduction

Diabetes is a cluster of metabolic conditions characterized by dysglycemia from defects in insulin secretion and/or unhealthy behaviors that cause debilitating complications and death [1,2]. In 2013, an estimated 8.3% of the global population lived with diabetes and more than US \$612 billion was spent on treatment [3]. A type 2 diabetes (T2DM) diagnosis, affecting 90-95% of people with diabetes, largely results from genetic predisposition, excess body weight, physical inactivity, and poor diet [1]. Additional studies demonstrate that lower socioeconomic status (SES) populations are at greater risk for developing diabetes [4,5] and often demonstrate poorer disease management, resulting in more frequent (and more expensive) complications and hospitalizations [5].

There is consensus among researchers and clinical professionals that glucose self-monitoring, exercise, and diet-related health behaviors are important in effective T2DM management and well-regulated serum glucose levels [6]. Monitoring health behaviors (ie, exercise, healthy diet, glucose monitoring) may be important in improving glucose control [6]. Mobile tracking technologies can help patients adopt and sustain self-management behaviors and can help health professionals provide better monitoring and support. However, there are challenges in determining the optimal design of self-tracking tools and their optimal use with respect to frequency and duration.

Data mining (DM) refers to analytic approaches useful in detecting coherent patterns in large and complex datasets [7]. The applications of DM methods in analyses of diabetes-related health behaviors are continually being improved, especially in the selection of analytic frameworks that capture key data with sufficient explanatory power [8].

As utilization of electronic health records increases in health care, DM becomes more relevant [9] to chronic disease prevention and management [8]. A recent study utilized descriptive DM algorithms to analyze a dataset with 450 attributes to identify a “short list” of the behavioral correlates of depressive disorder [10]. This study exemplifies use of DM in datasets lacking the uniformity needed for more conventional analyses [11]. DM can help integrate variables of multiple types (eg, diet, exercise, blood pressure, SES, income, geographic location) in better understanding factors affecting diabetes incidence, prevalence, and management [7].

In another diabetes study, DM algorithms were used to construct a model that predicted short-term changes in blood glucose [12]

exemplifying how DM can help identify risk factors for hypoglycemia [13]. Several studies found that information collected on meals, insulin therapy, and physical activity improved the prediction of blood glucose levels [12-14]. Applying DM analyses to self-monitoring data could help identify key indicators in the prevention of life-threatening hypoglycemic events [15]. DM applied to primary medical data could point to useful methods for initiating and maintaining effective T2DM treatment, leading to better resource allocation and treatment personalization [16,17].

Health coaching is a promising clinical role that stimulates and supports health behavior change in patients with varying SES, health problems and diagnosed chronic diseases. When connected with 24 hour/day/7 day/week mobile phone-based counseling, health coaching is associated with benefits for individuals affected by uncontrolled T2DM [18-20] and chronic obstructive pulmonary disease [17].

## Objective

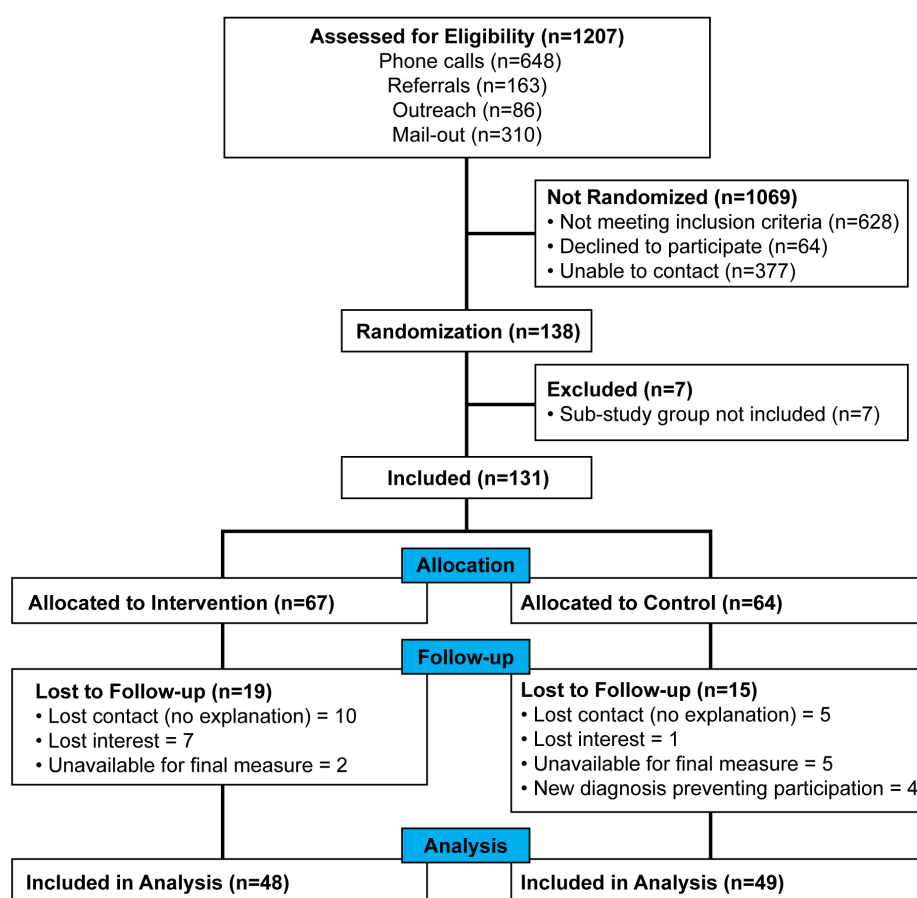
Our primary objective was to evaluate associations between the mobile phone monitoring of health behaviors, within a randomized controlled trial (RCT), and clinically significant reductions in glycated hemoglobin (HbA1c).

## Methods

The RCT protocol was reviewed and approved by the York University research ethics board (Certificate #2012-033), and all patients provided written, informed consent to participate.

The RCT assessed T2DM patients (N=97) assisted by personal health coaches trained in behavior-change theories, practices, and counseling methods (see Figure 1). At baseline, all patients had poor glucose regulation as indicated by glycated hemoglobin (HbA1c)  $\geq 7.3\%$ . In the mobile phone-assisted intervention arm, participants (n=48) were provided a smartphone (Samsung Galaxy Ace 2) with data service and pre-installed health tracking software (NexJ Systems Inc., Connected Wellness Platform [CWP]) enabling the detailed monitoring of two behaviors (exercise and diet) and a risk-related outcome (blood glucose) throughout a 24-week intervention. Health coaches helped patients use the mobile phone software in ways that best fit their daily routines.

CWP use data were extracted from NexJ Systems servers upon trial completion and compiled into .csv files stored on password-protected portable drives. Study participant IDs were matched with software user IDs, as data were anonymized and prepared for analyses.

**Figure 1.** Flowchart of enrollment.

## Data Analysis

### Association Rule Algorithms

To discover useful relationships between self-tracker use and HbA1c outcomes, we employed association-rule algorithms software to find coherent relationships in transactions represented by sets of items, termed “frequent item sets.” For example, when customer A buys bread and cheese, and customer B buys bread, cheese and burgers, bread and cheese appear frequently on both shopping lists. Therefore, bread and cheese are associated and they qualify as a frequent item set [bread, cheese]. Support is a term reflecting the measurement of association frequency, as defined by the percentage of observations to which the item sets belong. In our study, support was defined as the number of times an attribute value (such as *alc 6 month diff*=1.1) or a set of attribute values (such as *glucose\_count*=200 and *food\_count*=150) appear in participant data, divided by the total number of participants, expressed in a percentage (ie, multiplied by 100).

We used association rule algorithms to identify all common attributes in participants. The support threshold was fixed at a minimum of 5%, such that item sets were generated that occurred in at least 5% of the sample. We used the a priori association rule algorithm implemented by C language for the Unix/Linux environment.

### Attribute Selection

Among the 97 T2DM study completers, the present analysis considers patients in the experimental group (n=48). As our objective was to evaluate associations between the use frequency of different trackers in the CWP in relation to HbA1c outcomes, we selected the change in HbA1c over 24 weeks (*alc6month\_diff*) and 4 software uses as the attributes for the association rule algorithm. These attributes were use frequency of the blood glucose tracker (*glucose\_count*), use frequency of the food tracker (*food\_count*), use frequency of the exercise tracker (*exercise\_count*), and the total use frequency of all three trackers (*generic\_count*).

### Discretization

We proceeded to discretize the attributes to implement the association rule algorithm. For the four tracker attributes (*glucose\_count*, *food\_count*, *exercise\_count*, and *generic\_count*), data were discretized into categories relevant to typical use during a standard week, with numerical criteria selected to reflect significant adoption levels by patients (see Table 1). Discretization was compiled differently per tracker to align with expected and actual adoption rates per tracked behavior (or outcome). For example, glucose management in T2DM includes self-monitoring of serum-glucose via finger prick, recommended several times daily for poorly managed patients, and less often if there is better gluco-regulation. Use of the blood glucose tracker was discretized based on frequencies of 1-4.9 uses per

week, 5-7.9 uses per week, 8-13.9 uses per week, and 14-21.9 uses per week. For food tracking, a similar discretization pattern was used. Since use frequency fluctuated, discretization was calculated as up to once per week, 1-3.9 uses per week, 4-6.9 uses per week, 7-13.9 uses per week (1 to <2/day), 14-20.9 uses per week (2 to <3/day), and 21+ per week (3+/day). Since the exercise tracker was the least frequently used, discretization

rules were adjusted to once every other week, 0.5-0.9 uses per week, 1-1.9 uses per week, 2-2.9 uses per week, 3-3.9 uses per week, and 4-6 uses per week. The total use of all trackers was discretized to once per week, 1-3.9 times per week, 4-6.9 times per week (0.5-1/day), 7-13.9 times per week (1-2/day), 14-17.9 times per week (2-4/day), 18-34.9 times per week (4-5/day), and 35-50 times per week (5-7/day).

**Table 1.** Discretization of tracker use frequency<sup>a-d</sup>.

Attribute	Group	Range	Meaning
Glucose count	1	0-0	Zero
	2	5-24	Up to once per week
	3	24-120	1-4.9 times/week
	4	120-192	5-7.9 times/week
	5	192-336	8-13.9 times/week
	6	336-520	14-21.9 times/week
Exercise count	1	0-0	Zero
	2	0-5	Minimal (>0-4)
	3	5-12	Once every other week
	4	12-24	0.5-0.9 time/week
	5	24-48	1-1.9 times/week
	6	48-72	2-2.9 times/week
	7	72-96	3-3.9 times/week
	8	96-150	4-6 times/week
Food count	1	0-0	Zero
	2	1-5	>0-5 over 6 months
	3	5-24	Up to 1 time/week
	4	24-96	1-3.9 times/week
	5	96-168	4-6.9 times/week (1/day)
	6	168-336	7-13.9 times/week (1-1.9/day)
	7	336-504	14-20.9 times/week (2-2.9/day)
	8	504-720	21+ times/week (3+/day)
Generic count	1	0-0	Zero
	2	0-24	1 time/week
	3	24-96	1-3.9 times/week
	4	96-168	4-6.9 times/week (0.5-1/day)
	5	168-336	7-13.9 times/week (1-2/day)
	6	336-672	14-17.9 times/ week (2-4/day)
	7	672-840	18-34.9 times/ week (4-5/day)
	8	840-1200	35-50 times/week (5-7/day)

<sup>a</sup>Attribute=type of tracker used.

<sup>b</sup>Group=discretization group.

<sup>c</sup>Range=frequencies of use for allocation to group.

<sup>d</sup>Meaning=frequency of use in terms of use per week/day.

### Preprocessing for HbA1c Attribute

For the *alc6month\_diff* attribute, conceived of as the most important attribute in investigating the desired associations, we preserved its original value to an optimal degree. Therefore, we did not discretize this attribute but created an extension of attributes based on the *alc6month\_diff* attribute. We obtained a set of all possible values for *alc6month\_diff* attribute ( $a$ ), such as  $\{-1.4, -0.8, -0.5, -0.3, -0.1, 0, 0.1, 0.2, 0.5, a_n\}$ . For patients with an *alc6month\_diff* value of  $a_1$  through to  $a_n$ , we created the following new attributes: *alc6\_at\_least\_-1.4*, *alc6\_at\_least\_-0.8*, *alc6\_at\_least\_-0.5*, ..., *alc6\_at\_least\_n*. Accordingly, the patient with the lowest reduction value was associated with only one attribute value (of this kind) while the patient with the highest reduction was associated with all the attribute values established.

### Postprocessing of Associations

The above pre-processing approaches produced redundancies in the input data for the association rule algorithm. Therefore, we applied three approaches to post-processing after the associations were generated. First, if a frequent item set contained only *alc6\_at\_least\_a<sub>n</sub>* attribute values, where  $a_n$  is a float number (from  $a_1$  to  $a_n$ ), we removed such item sets (for example, *alc6\_at\_least\_0.4 alc6\_at\_least\_0.2* and *alc6\_at\_least\_-0.3 alc6\_at\_least\_-0.8 alc6\_at\_least\_-1.4*). The preceding two frequent item sets were thus removed because the extra *alc6\_at\_least* values were added by the pre-processing approach were not relevant unless generated together with the other attributes.

Second, for a frequent item set, if *alc6\_at\_least\_a<sub>n</sub>* the attribute values that appeared multiple times were removed except the *alc6\_at\_least\_a<sub>n</sub>* where  $n$  is the largest value among the “*alc6*” values, (for example, *glucose\_120 alc6\_at\_least\_0.5 alc6\_at\_least\_0.3*).

In this frequent item set, the *alc6\_at\_least* attribute occurred twice. However, when a reduction value for “*alc6*” of at least 0.5 exists, by default, the reduction value of 0.3 must exist. Therefore, *alc6\_at\_least\_0.3* was a redundant attribute value and was therefore removed. The above frequent item set becomes the following after post-processing: *glucose\_120 alc6\_at\_least\_0.5*.

Third, after considering the above two situations, redundant item sets existed in the result. For example, *glucose\_120 alc6\_at\_least\_0.7 alc6\_at\_least\_0.4* and *glucose\_120 alc6\_at\_least\_0.7 alc6\_at\_least\_0.1*. These two frequent item sets both contained redundant “*alc6\_at\_least*” attribute values. After considering the second situation, these two frequent item

sets became *glucose\_120 alc6\_at\_least\_0.7* and *glucose\_120 alc6\_at\_least\_0.7*. Since they were identical and their support values were identical, we removed one of them.

Pre- and postprocessing were implemented by Python using the Linux environment. Attribute selection and categorization were used to determine how system use was associated with change in HbA1c per participant. A minimum clinically significant change approach was used to determine what proportion of participants demonstrated 0.5% (5.5 mmol/mol) or greater reductions of HbA1c and used trackers at variable levels of intensity [21]. Therefore, patient data were extracted only with associations containing the attribute-value pair *alc\_0.5*, representing changes in HbA1c values that were 0.5% or above, which was true for 29 participants.

Last, a Fisher's exact test for count data was applied to determine the statistical relationship between use of all three trackers, use of one or two trackers, and HbA1c reductions. For this test, the 10 intervention participants who used the software but did not have a clinically significant reduction in HbA1c were used as a comparison group. Results were considered significant at the  $P=.05$  level.

## Results

### Usage Data

In total, 48 intervention patients completed the 24-week trial. Only 39 patients used the CWP software to track health behaviors. Of this software software-user sample, 29 reduced their HbA1c measure by clinically significant levels at trial conclusion, defined as 0.5% (5.5 mmol/mol) or greater (see Figure 2). Demographic and baseline characteristics of the clinically significant HbA1c reduced-users are provided in Table 2 and compared with the full intervention sample. While multiple health trackers were available in the CWP, most participants used one or a combination of two-to-three specific trackers (ie, between (1) glucose monitoring, (2) exercise tracking, and (3) food tracking) (see Figures 3-5). These trackers reflect the behaviors often considered relevant in diabetes management and therefore the trackers often recommended by health coaches for patients use. Of the 29 clinically significant HbA1c reduced-user group (selected as a subpopulation of interest [SOI]) (HbA1c reduction  $\geq 0.5\%$ ), 2 singularly used the food tracker and 7 singularly used the glucose tracker, while no subjects singularly used the exercise tracker. Meanwhile, 11 of these subjects used both glucose and exercise trackers, 3 used both glucose and food trackers, while 0 subjects used both food and exercise trackers. Last, 6 subjects used all three trackers.

**Table 2.** Subject demographics.

	SOI (n=29)	Full intervention sample (n=48)
Age in years, mean (range) SD	53.4 (26-68) 10.7	53.1 (26-74) 10.9
<b>HbA1c, % (mmol/mol) SD</b>		
Baseline	8.88% (73.6) 1.30	8.69% (71.5) 1.32
6 months	7.52% (58.7) 0.95	7.88% (62.6) 1.17
Reduction (baseline to 6 months)	1.36% (14.9)	0.81% (8.9)
<b>Sex, n (%)</b>		
Male	11 (33)	17 (35)
Female	18 (66)	31 (65)
<b>Education, n (%)</b>		
Less than high school	5 (17)	10 (21)
High school	14 (48)	17 (35)
College diploma or vocational training	6 (21)	11 (23)
University degree	4 (14)	8 (17)
Did not disclose	0 (0)	2 (4)
<b>Car access, n (%)</b>		
Owns a car	10 (34)	19 (40)
Has access to car	8 (28)	9 (19)
No access to car	11 (38)	19 (40)
Not disclosed	0 (0)	1 (2)
<b>Employment status, n (%)</b>		
Unemployed	10 (35)	16 (33)
Student	1 (3)	3 (6)
Part-time	0 (0)	1 (2)
Full-time	8 (28)	13 (27)
Retired	3 (10)	6 (13)
Self-employed	5 (17)	6 (13)
Work in the home (take care of children)	2 (7)	2 (4)
Not disclosed	0 (0)	1 (2)
<b>Income in CAD\$, n (%)</b>		
0-9999	4 (14)	9 (19)
10,000-25,000	8 (28)	10 (21)
25,000-50,000	7 (24)	12 (25)
50,000-75,000	2 (7)	3 (6)
75,000-up	3 (10)	4 (8)
Did not disclose	5 (17)	10 (21)

As seen in [Table 3](#), the HbA1c reductions of subjects who used single trackers (food-tracker only or glucose-tracker only) did not significantly differ from each other. More subjects (11/48, 22.9%) used the glucose and exercise tracker in combination than the glucose and food tracker in combination (3/48, 6.3%), but there were no significant differences in HbA1c levels in these subjects. The 6 of 48 subjects (12.5%) who used all three

trackers (glucose/food/exercise) achieved a mean HbA1c reduction of 1.55%.

The Fisher's exact test for count data indicated subjects who used the software (n=39) were more likely to achieve a clinically significant reduction in HbA1c if they used one or two trackers than if they used all three trackers (OR 0.18,  $P=.04$ ) (see [Table 4](#)).

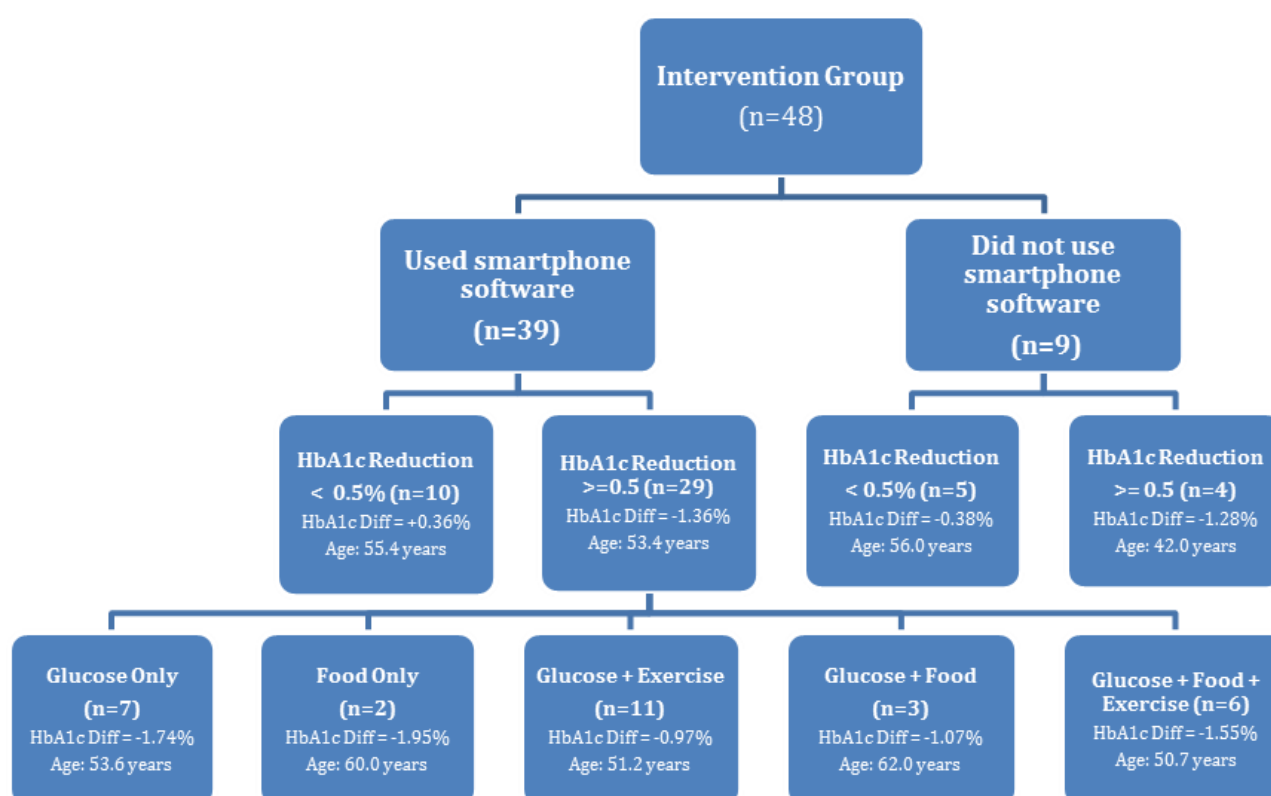


**Table 3.** Tracker usage pattern.

Mobile phone tracker	Users, n	Mean reduction of HbA1c, %	SD
Food only	2	1.95	1.20
Glucose only	7	1.74	1.00
Exercise only	0	0	0
Glucose + exercise	11	0.97	0.38
Glucose + food	3	1.07	0.45
Food + exercise	0	0	0
Glucose + food + exercise	6	1.55	0.49

**Table 4.** Fisher's exact test for count data on three trackers versus less<sup>a</sup>.

	Achieved clinical HbA1c reduction (≥0.5%)	Did not achieve clinical HbA1c reduction (<0.5%)
All 3 trackers	6	6
<3 trackers	23	4

<sup>a</sup>OR 0.18 and *P*=.04.**Figure 2.** Breakdown of smartphone usage group.

**Figure 3.** Food tracker.

Cancel Add Food Save

Tap to Retake Photo

× Remove

What is your portion size?  
Medium

Where was your food prepared?  
Homemade

How healthy is your food?  
High Healthiness

Add Note Set Time

**Figure 4.** Blood glucose tracker.

Cancel Add Blood Glucose Save

Blood Glucose

What is your blood glucose level?

6.2 mmol/L mg/dL

Add Note Set Time



**Figure 5.** Exercise tracker.

Exercise

How intense was your exercise?

Moderate

How long did you exercise?

45 minutes

How many calories did you burn?

calories

Add Note Set Time

### Data Mining

The glucose tracker was the most frequently used CWP function. Altogether, 22.9% (11/48) of subjects tracked their blood glucose with the software 1-4.9 times per week *and* had a clinically significant reduction in HbA1c (>0.5%) (see [Table 5](#)) while an additional 22.9% with >0.5% reduction in HbA1c demonstrated more frequent glucose tracker use (6/48, 12.5%) and used it 5-7.9 times per week while 10.4% (5/48) used it 8-13.9 times per week.

The food tracking function followed a similar pattern, although with slightly less frequent use. We found 9/48 (18.8%) of

participants who achieved at least a 0.5% reduction in HbA1c used the food tracker a minimal amount (one time or less per week), while 7/48 (14.6%) used the system 1-3.9 times per week, and 4/48 (8.3%) used the system 4-6.9 times per week (see [Table 6](#)).

The least used tracker (and the only tracker that was never singularly used) was the exercise tracker, which was used by 7/48 (14.6%) of the intervention participants between 0.5-0.9 times per week ([Tables 7 and 8](#)). When it was paired with glucose tracking, however, every patient who used the exercise tracker had clinically significant reductions in HbA1c (see [Figure 6](#)).

**Table 5.** Glucose tracker use.

HbA1c diff.	Glucose 1-4.9 times/week		Glucose 5-7.9 times/week		Glucose 8-13.9 times/week	
	Users, %	n	Users, %	n	Users, %	n
0.1	29.2	14	12.5	6	12.5	6
0.2	27.1	13	12.5	6	12.5	6
0.3	25	12	12.5	6	12.5	6
0.4	22.9	11	12.5	6	12.5	6
0.5	22.9	11	12.5	6	10.4	5
0.6	20.8	10	10.4	5	10.4	5
0.7	18.8	9	10.4	5	10.4	5
0.8	16.7	8	10.4	5	10.4	5
0.9	16.7	8	10.4	5	10.4	5
1.1	14.6	7	10.4	5	8.3	4
1.2	10.4	5	8.3	4		
1.3	10.4	5	8.3	4		
1.5	8.3	4	8.3	4		

**Table 6.** Food tracker use.

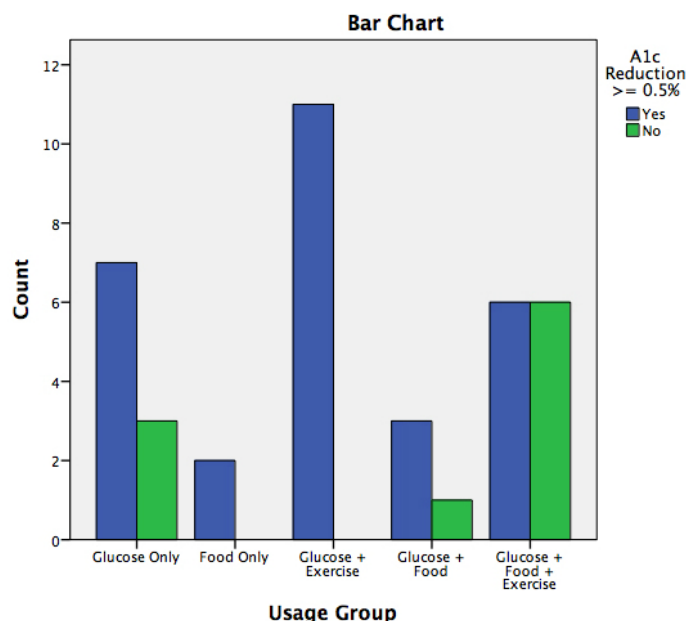
HbA1c diff.	Food up to 1x/wk		Food 1-3.9x/wk		Food 4-6.9x/wk	
	Users, %	n	Users, %	n	Users, %	n
0.1	20.8	10	16.7	8	8.3	4
0.2	18.8	9	16.7	8	8.3	4
0.3	18.8	9	14.6	7	8.3	4
0.4	18.8	9	14.6	7	8.3	4
0.5	18.8	9	14.6	7	8.3	4
0.6	16.7	8	12.5	6	8.3	4
0.7	16.7	8	10.4	5	8.3	4
0.8	16.7	8	10.4	5	8.3	4
0.9	14.6	7	8.3	4	8.3	4
0.11	14.6	7				
0.12	8.3	4				
0.13	8.3	4				

**Table 7.** Exercise tracker use—Part 1.

HbA1c diff.	Exercise up to 1x/wk		Exercise <5x in 6 months		Exercise 1x every other wk	
	Users, %	n	Users, %	n	Users, %	n
0.1	16.7	8	16.7	8	14.6	7
0.2	16.7	8	14.6	7	14.6	7
0.3	14.6	7	14.6	7	14.6	7
0.4	14.6	7	12.5	6	12.5	6
0.5	14.6	7	12.5	6	12.5	6
0.6	12.5	6	12.5	6	8.3	4
0.7	12.5	6	12.5	6	8.3	4
0.8	12.5	6	12.5	6	8.3	4
0.9	12.5	6	10.4	5	8.3	4
0.11	12.5	6	10.4	5	8.3	4
0.13	8.3	4	10.4	5		
0.15			8.3	4		

**Table 8.** Exercise tracker use—Part 2.

HbA1c diff.	Exercise 0.5-0.9x/wk		Exercise 1-1.9x/wk		Exercise 2-2.9x/wk	
	Users, %	n	Users, %	n	Users, %	n
0.1	14.6	7	8.3	4	8.3	4
0.2	14.6	7	8.3	4	8.3	4
0.3	14.6	7			8.3	4
0.4	14.6	7			8.3	4
0.5	14.6	7				
0.6	14.6	7				
0.7	12.5	6				
0.8	12.5	6				
0.9	10.4	5				
0.11	8.3	4				

**Figure 6.** Tracker usage group by clinical HbA1c reduction ( $\geq 0.5\%$ ).

## Discussion

### Principal Findings

In the emergent field of mobile phone-assisted health coaching, a recurring question is whether the expense of integrating mHealth technology justifies the benefits. This highlights the need to further investigate for whom (which subpopulations) these technologies are useful and which technologies are most useful for them. In addressing these questions, this study pilots a method of investigating RCT participants, focusing on those who derived clinically significant benefit from participation and made significant use of the mobile phone software. The study specifies how much use they engaged in, and the associated benefits, using the clinically significant reduction in HbA1c criteria of 0.5% (5.5 mmol/mol) or greater.

By employing attribute categorizations, tracker-use frequencies were determined per subject in an SOI of significant users who also derived clinically significant glucose reductions. This dataset enabled comparison of subjects who used single trackers versus multiple trackers (two or three trackers). Descriptive results indicated that in subjects achieving a 0.5% HbA1c reduction or better, singular use of the glucose and food monitoring was undertaken, while, in contrast, in the same group, no singular use of the exercise tracker was undertaken. In dual-tracker use, glucose and exercise trackers were employed by more subjects than glucose and food tracking. Interestingly, food and exercise (as dual) trackers were not used by any of the subjects who met criteria for 0.5% HbA1c reduction or greater. Last, all three trackers were used by 12.5% (6/48) of subjects and were associated with a substantive HbA1c mean reduction (1.55% or 16.9 mmol/mol).

The Fisher's exact test for count data indicated that for subjects who used the software ( $n=39$ ) and used all three trackers, there was a significantly lower likelihood of achieving a clinically

significant reduction in HbA1c than for those who used a lesser number (ie, one or two trackers) (OR 0.18,  $P=.04$ ). This finding indicates that it might be advisable for patients with type 2 diabetes to focus on one or two trackers, especially if one tracker assesses blood glucose levels. From a behavioral perspective, these data could influence health coaching recommendations for health behavior tracking.

### Limitations

Data mining is often used to process large amounts of data. One limitation of the pilot application of data mining in this study was the relatively small user sample size. Nonetheless, the association rule algorithm technique offers a foundation with which to study larger datasets of mHealth tracking technologies as they become available. In terms of diabetes intervention, this was an RCT of typical size (48 intervention participants) and, altogether 10,695 uses of the mobile phone app were analyzed (about 62 uses per month per participant who used the software). Although future DM studies may address larger datasets, this pilot demonstrated application in an RCT dataset within which >10,000 app uses were analyzed (averaging ~1 use per day).

### Conclusion

In summary, this study points to a future when the mobile monitoring of health behaviors will increase and provide digital signals representing engagement in discrete behaviors and daily-weekly-monthly outcomes. Whereas previous associations between counseling and outcomes were difficult to obtain and often based on retrospective self-report, mobile phone monitoring offers ongoing records that precisely reflect status improvements, their stability, and fluctuations (eg, relapsing patterns). Altogether, with the increasing collection of wearable data, we may derive a quantifiable perspective on health changes that instructs the patient and health coach in improving chronic disease management.

## Acknowledgments

The authors would like to thank NexJ Systems Inc. for their partnership in this trial and for the use of the Connected Wellness Platform as a clinical research tool. Funding was provided by the Public Health Agency of Canada and the Federal Development Agency of Southern Ontario. We offer special thanks to the staff of the Black Creek Community Health Centre and trial participants from the Jane-Finch community of Toronto, Ontario.

Joel Katz is supported by a Canadian Institutes of Health Research Canada Research Chair in Health Psychology.

The authors would like to acknowledge with sadness the untimely passing of study co-author, colleague, and friend, Dr. Nicholas Cercone. Dr. Cercone's expertise and mentorship on data mining theory and technique was invaluable. His inspiring and supportive presence will be deeply missed.

## Conflicts of Interest

None declared.

## References

1. Seaquist ER. Addressing the burden of diabetes. JAMA 2014 Jun 11;311(22):2267-2268. [doi: [10.1001/jama.2014.6451](https://doi.org/10.1001/jama.2014.6451)] [Medline: [24915253](#)]
2. Riddle MC, Ambrosius WT, Brillon DJ, Buse JB, Byington RP, Cohen RM, Action to Control Cardiovascular Risk in Diabetes Investigators. Epidemiologic relationships between A1C and all-cause mortality during a median 3.4-year follow-up of glycemic treatment in the ACCORD trial. Diabetes Care 2010 May;33(5):983-990 [FREE Full text] [doi: [10.2337/dc09-1278](https://doi.org/10.2337/dc09-1278)] [Medline: [20427682](#)]
3. International Diabetes Foundation. In: Guariguata L, Nolan T, Beagley J, Linnenkamp U, Jacqmain O, editors. IDF Diabetes Atlas (6th Edition). Brussels, Belgium: International Diabetes Federation; 2013.
4. Dinca-Panaitescu S, Dinca-Panaitescu M, Bryant T, Daiski I, Pilkington B, Raphael D. Diabetes prevalence and income: Results of the Canadian Community Health Survey. Health Policy 2011 Feb;99(2):116-123. [doi: [10.1016/j.healthpol.2010.07.018](https://doi.org/10.1016/j.healthpol.2010.07.018)] [Medline: [20724018](#)]
5. Hux J, Booth G, Slaughter P, Laupacis A. Diabetes in Ontario: An ICES Practice Atlas. Toronto: Institute for Clinical Evaluative Science; 2003. URL: <http://www.ices.on.ca/~media/Files/Atlases-Reports/2003/Diabetes-in-Ontario/Full%20report.ashx> [accessed 2015-04-06] [WebCite Cache ID 6XabgpTYm]
6. Canadian Diabetes Association Clinical Practice Guidelines Expert Committee. Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada. Can J Diabetes 2013;37:S1-S212 [FREE Full text]
7. Murdoch TB, Detsky AS. The inevitable application of big data to health care. JAMA 2013 Apr 3;309(13):1351-1352. [doi: [10.1001/jama.2013.393](https://doi.org/10.1001/jama.2013.393)] [Medline: [23549579](#)]
8. Georga E, Protopappas V, Mougiakakou S, Fotiadis D. Short-term vs. long-term analysis of diabetes data: Application of machine learning and data mining techniques. In: IEEE. 2013 Presented at: 13th International Conference on Bioinformatics and Bioengineering (BIBE); November 10-13, 2013; Chania p. 1-4. [doi: [10.1109/BIBE.2013.6701622](https://doi.org/10.1109/BIBE.2013.6701622)]
9. Koh HC, Tan G. Data mining applications in healthcare. J Healthc Inf Manag 2005 Jan;19(2):64-72. [Medline: [15869215](#)]
10. Yoon S, Taha B, Bakken S. Using a data mining approach to discover behavior correlates of chronic disease: a case study of depression. Stud Health Technol Inform 2014;201:71-78. [Medline: [24943527](#)]
11. Yoo I, Alafaireet P, Marinov M, Pena-Hernandez K, Gopidi R, Chang J, et al. Data mining in healthcare and biomedicine: a survey of the literature. J Med Syst 2012 Aug;36(4):2431-2448. [doi: [10.1007/s10916-011-9710-5](https://doi.org/10.1007/s10916-011-9710-5)] [Medline: [21537851](#)]
12. Georga EI, Protopappas VC, Ardigo D, Marina M, Zavaroni I, Polyzos D, et al. Multivariate prediction of subcutaneous glucose concentration in type 1 diabetes patients based on support vector regression. IEEE J Biomed Health Inform 2013 Jan;17(1):71-81. [doi: [10.1109/TITB.2012.2219876](https://doi.org/10.1109/TITB.2012.2219876)] [Medline: [23008265](#)]
13. Zecchin C, Facchinetti A, Sparacino G, Cobelli C. Reduction of number and duration of hypoglycemic events by glucose prediction methods: a proof-of-concept in silico study. Diabetes Technol Ther 2013 Jan;15(1):66-77. [doi: [10.1089/dia.2012.0208](https://doi.org/10.1089/dia.2012.0208)] [Medline: [23297671](#)]
14. Pappada SM, Cameron BD, Rosman PM, Bourey RE, Papadimos TJ, Olorunto W, et al. Neural network-based real-time prediction of glucose in patients with insulin-dependent diabetes. Diabetes Technol Ther 2011 Feb;13(2):135-141. [doi: [10.1089/dia.2010.0104](https://doi.org/10.1089/dia.2010.0104)] [Medline: [21284480](#)]
15. Barrett MA, Humblet O, Hiatt RA, Adler NE. Big Data and Disease Prevention: From Quantified Self to Quantified Communities. Big Data 2013 Sep;1(3):168-175. [doi: [10.1089/big.2013.0027](https://doi.org/10.1089/big.2013.0027)]
16. Marinov M, Mosa ASM, Yoo I, Boren SA. Data-mining technologies for diabetes: a systematic review. J Diabetes Sci Technol 2011 Nov;5(6):1549-1556 [FREE Full text] [Medline: [22226277](#)]
17. Verwey R, van der Weegen S, Spreeuwenberg M, Tange H, van der Weijden T, de WL. A pilot study of a tool to stimulate physical activity in patients with COPD or type 2 diabetes in primary care. J Telemed Telecare 2014 Jan;20(1):29-34. [doi: [10.1177/1357633X13519057](https://doi.org/10.1177/1357633X13519057)] [Medline: [24414397](#)]

18. Wayne N, Ritvo P. Smartphone-enabled health coach intervention for people with diabetes from a modest socioeconomic strata community: single-arm longitudinal feasibility study. *J Med Internet Res* 2014;16(6):e149 [FREE Full text] [doi: [10.2196/jmir.3180](https://doi.org/10.2196/jmir.3180)] [Medline: [24907918](https://pubmed.ncbi.nlm.nih.gov/24907918/)]
19. Wayne N, Perez DF, Kaplan DM, Ritvo P. Health Coaching Reduces HbA1c in Type 2 Diabetic Patients From a Lower-Socioeconomic Status Community: A Randomized Controlled Trial. *J Med Internet Res* 2015;17(10):e224 [FREE Full text] [doi: [10.2196/jmir.4871](https://doi.org/10.2196/jmir.4871)] [Medline: [26441467](https://pubmed.ncbi.nlm.nih.gov/26441467/)]
20. Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. *Diabetes Care* 2011 Sep;34(9):1934-1942 [FREE Full text] [doi: [10.2337/dc11-0366](https://doi.org/10.2337/dc11-0366)] [Medline: [21788632](https://pubmed.ncbi.nlm.nih.gov/21788632/)]
21. Karyekar CS, Frederick R, Ravichandran S. Clinically relevant reductions in HbA1c without hypoglycaemia: results across four studies of saxagliptin. *Int J Clin Pract* 2013 Aug;67(8):759-767 [FREE Full text] [doi: [10.1111/ijcp.12212](https://doi.org/10.1111/ijcp.12212)] [Medline: [23795975](https://pubmed.ncbi.nlm.nih.gov/23795975/)]

## Abbreviations

**CWP:** Connected Wellness Platform  
**DM:** data mining  
**HbA1c:** glycated hemoglobin  
**RCT:** randomized controlled trial  
**SOI:** subpopulation of interest  
**T2DM:** type 2 diabetes mellitus

*Edited by G Eysenbach; submitted 06.04.15; peer-reviewed by TW Ho, M Spreeuwenberg, F Lai; comments to author 29.07.15; revised version received 28.10.15; accepted 08.01.16; published 06.04.16*

### *Please cite as:*

Wayne N, Cercone N, Li J, Zohar A, Katz J, Brown P, Ritvo P

*Data Mining of a Remote Behavioral Tracking System for Type 2 Diabetes Patients: A Prospective Cohort Study*  
*JMIR Diabetes* 2016;1(1):e1

URL: <http://diabetes.jmir.org/2016/1/e1/>

doi: [10.2196/diabetes.4506](https://doi.org/10.2196/diabetes.4506)

PMID:

©Noah Wayne, Nick Cercone, Jiye Li, Ariel Zohar, Joel Katz, Patrick Brown, Paul Ritvo. Originally published in JMIR Diabetes (<http://diabetes.jmir.org>), 06.04.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on <http://mhealth.jmir.org/>, as well as this copyright and license information must be included.