



OPEN

## Diabetes medication recommendation system using patient similarity analytics

Wei Ying Tan<sup>1,2✉</sup>, Qiao Gao<sup>1</sup>, Ronald Wihal Oei<sup>1</sup>, Wynne Hsu<sup>1,3</sup>, Mong Li Lee<sup>1,3</sup> & Ngiap Chuan Tan<sup>4,5</sup>

Type-2 diabetes mellitus (T2DM) is a medical condition in which oral medications avail to patients to curb their hyperglycaemia after failed dietary therapy. However, individual responses to the prescribed pharmacotherapy may differ due to their clinical profiles, comorbidities, lifestyles and medical adherence. One approach is to identify similar patients within the same community to predict their likely response to the prescribed diabetes medications. This study aims to present an evidence-based diabetes medication recommendation system (DMRS) underpinned by patient similarity analytics. The DMRS was developed using 10-year electronic health records of 54,933 adult patients with T2DM from six primary care clinics in Singapore. Multiple clinical variables including patient demographics, comorbidities, laboratory test results, existing medications, and trajectory patterns of haemoglobin A<sub>1c</sub> (HbA<sub>1c</sub>) were used to identify similar patients. The DMRS was evaluated on four groups of patients with comorbidities such as hyperlipidaemia (HLD) and hypertension (HTN). Recommendations were assessed using hit ratio which represents the percentage of patients with at least one recommended sets of medication matches exactly the diabetes prescriptions in both the type and dosage. Recall, precision, and mean reciprocal ranking of the recommendation against the diabetes prescriptions in the EHR records were also computed. Evaluation against the EHR prescriptions revealed that the DMRS recommendations can achieve hit ratio of 81% for diabetes patients with no comorbidity, 84% for those with HLD, 78% for those with HTN, and 75% for those with both HLD and HTN. By considering patients' clinical profiles and their trajectory patterns of HbA<sub>1c</sub>, the DMRS can provide an individualized recommendation that resembles the actual prescribed medication and dosage. Such a system is useful as a shared decision-making tool to assist clinicians in selecting the appropriate medications for patients with T2DM.

### Abbreviations

CPG	Clinical practice guidelines
D3K	Data-driven and domain knowledge
DPP-4 inhibitor	Dipeptidyl peptidase-4 inhibitor
DM	Diabetes mellitus
DHL	Diabetes, hyperlipidaemia and hypertension
EHR	Electronic health records
HbA <sub>1c</sub>	Haemoglobin A <sub>1c</sub>
HDL	Cholesterol high-density lipoprotein
HLD	Hyperlipidaemia
HTN	Hypertension
ICD	International classification of diseases
LDL	Cholesterol high-density lipoprotein
PDD	Prescribed daily dose
SGLT-2 inhibitor	Sodium-glucose cotransporter-2 inhibitor

<sup>1</sup>Institute of Data Science, National University of Singapore, 3 Research Link, #04-06, Singapore 117602, Singapore. <sup>2</sup>Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore. <sup>3</sup>School of Computing, National University of Singapore, Singapore, Singapore. <sup>4</sup>SingHealth Polyclinics, SingHealth, Singapore, Singapore. <sup>5</sup>Family Medicine Academic Clinical Programme, SingHealth-Duke NUS Academic Medical Centre, Singapore, Singapore. ✉email: idstwy@nus.edu.sg

SDL	Standard drug list
T2DM	Type-2 diabetes mellitus

Type-2 diabetes mellitus (T2DM) is a complex, chronic illness with increased risk of premature death and disability<sup>1</sup>. Patients with T2DM differs in terms of their clinical profiles, comorbidities, lifestyles and medical adherence<sup>2–4</sup>. Pharmacotherapy for managing T2DM may involve a combination of medications to achieve optimal glycaemic control. Over the years, increasing prevalence of patients with diabetes worldwide have encouraged the development of new pharmaceutical drugs and expanded the therapeutic treatment options<sup>5,6</sup>. Formulating an individualized treatment plan is increasingly challenging due to heterogeneity in patients' clinical profiles along with a growing list of new medications to prescribe. Numerous clinical practice guidelines (CPGs)<sup>3,7,8</sup> exist to guide clinicians in selecting the most appropriate diabetes treatment options. However, CPGs have broad definitions of patient's groups and a standardized one-size-fits-all treatment approach<sup>9–11</sup>. In addition, most CPGs provide treatment recommendations tailored to a single condition<sup>12</sup> while patients with T2DM often have several comorbidities i.e. hyperlipidaemia (HLD) and hypertension (HTN)<sup>13,14</sup>.

To address these limitations, several artificial intelligence (AI) medication recommendation systems<sup>15–17</sup> have been proposed. While these systems demonstrated the potential to make clinically relevant recommendations, their real-world deployments in clinical practice are rare due to complex logic and the lack of explanation to support recommendations<sup>18</sup>. Previous studies<sup>19–22</sup> have shown how patient similarity analytics can enhance the interpretability of recommendation systems and assist with clinical decision making. In this study, we aim to develop an evidence-based diabetes medication recommendation system (DMRS) underpinned by patient similarity analytics for patients with T2DM. The DMRS takes into consideration the age of a patient, his/her clinical profile, comorbidities, existing medications, and trajectory of their haemoglobin A<sub>1c</sub> (HbA<sub>1c</sub>) results.

## Methods

**Study cohort.** Data used in this study was obtained from the EHR of six primary care clinics in Singapore. The study was approved by SingHealth Centralized Institutional Review Board (Reference Number: 2019/2604) prior to conduct of the study. Requirement of written consent was waived by the SingHealth Centralized Institutional Review Board as it was deemed impracticable while privacy risks were mitigated through the use of de-identified data. All methods were performed in accordance with relevant guidelines and regulations. The study cohort comprises multi-ethnic Asians adult patients, age 21 years or older having diabetes diagnosis under International Classification of Diseases (ICD), 9th or 10th Revision codes (250.90, 250.40, 250.80, E11.9, E11.21, E11.22, E14.31, E14.73 and E11.40) or have at least one diabetes medication in the 10-year period. These patients may also have hypertension (HTN) with ICD codes 401.1, 796.2, I10, or if they were on one or more anti-hypertensive medications. They may also have hyperlipidaemia (HLD) with ICD codes 272.0, E78.5, or if they were on one or more lipid-lowering medications. Patients' demographic, disease history, laboratory test results and medications prescriptions were extracted over a 10-year period from 1 January 2010 to 31 December 2019. In total, the study cohort consisted of 54,933 patients.

**Patient variables.** Variables used for patient similarity learning included age, gender, blood pressure, cholesterol, triglycerides, HbA<sub>1c</sub>, disease duration of diabetes (DM), HLD and HTN, medications for DM, HLD and HTN. Medications were grouped by their classes and medication counts were derived by counting the number of medications in each class. In total, there were 18 DM medication types and six DM medication classes, nine HLD medication types and four HLD medication classes and 22 HTN medication types and eight HTN medication classes. All medication dosages except Insulin were expressed in prescribed daily dose (PDD) defined using three intensity level: low (L), moderate (M) and high (H) with reference to the maximum daily dosage (for example, L:  $\leq 1/3$  maximum daily dosage, M:  $\leq 2/3$  maximum daily dosage and H:  $> 2/3$  maximum daily dosage). The list of medication types, classes and their dosage intensity can be found in Supplementary Table S1. To avoid complexity and the risk of overtreatment of insulin with oral DM medications<sup>23</sup>, insulin medications were regarded as binary variable (i.e. 1 if patient was prescribed an Insulin medication, 0 otherwise).

**Recommendation algorithm.** Given a target patient, the proposed DMRS uses patient's latest clinical profiles and HbA<sub>1c</sub> trajectories to retrieve a set of similar patients and use their prescribed medications for the recommendation.

For the clinical profile similarity, we used the data-driven and domain knowledge (D3K) similarity measure proposed by Oei et al.<sup>24</sup> to obtain an initial similarity score. The patient variables consisted of age, gender, systolic and diastolic blood pressure, cholesterol High-Density Lipoprotein (HDL), Low-Density Lipoprotein (LDL), triglyceride, HbA<sub>1c</sub>, HLD and HTN comorbidities, duration of disease(s) in years and count of medication types and classes. These variables were weighted by learning a generalized Mahalanobis measure that maximized the distance between patient pair, ( $P_i, P_k$ ) deemed to be clinically dissimilar while minimizing the distance between patients ( $P_i, P_j$ ) deemed as clinically similar. The personalized score of patient P for variable v, denoted as score (P, v), was determined based on the P's value for v. Given two patients P1, P2, each with D variables, their D3K similarity score was given by

$$D3K\_sim(P1, P2) = \frac{2 \sum_{v=1}^D \min(score(P1, v), score(P2, v))}{(\sum_{v=1}^D score(P1, v) + \sum_{v=1}^D score(P2, v))}$$

For the management of chronic diseases such as diabetes, we observed that besides assessing whether the patient's HbA<sub>1c</sub> value falls within the normal range, clinicians often up or down-titrate medication dosage with

regards to the trend in HbA<sub>1c</sub> test results over the years. We modelled how clinicians typically analysed patient's lab results by mapping patient's HbA<sub>1c</sub> trajectory to a sequence of symbols. The symbols represent whether the results were normal (N), abnormal (A), increasing (U) or decreasing (D). Trajectory mapping were illustrated using the two cases as follows: First, we compared the HbA<sub>1c</sub> values v1 and v2 at every two consecutive months and used  $\tau$  to denote a pre-determined threshold.

Case 1. Difference between v1 and v2  $\leq \tau$ .

If v1 lies within the normal HbA<sub>1c</sub> range, we mapped this to the symbol "N", otherwise it would be mapped to the symbol "A".

Case 2. Difference between v1 and v2  $> \tau$ .

If v1 < v2, we mapped this to the symbol "U" to represent an uptrend trajectory. Otherwise, we mapped to the symbol "D" to represent a downtrend trajectory.

If patients' the HbA<sub>1c</sub> value were missing for that month, linear interpolation was used to estimate the missing value. In the instance where patients' have multiple HbA<sub>1c</sub> values in a month, the average HbA<sub>1c</sub> value will be considered. For example, consider a patient having normal HbA<sub>1c</sub> values for three consecutive months, followed by increasing HbA<sub>1c</sub> values in the next two months and remains in the abnormal range in the last two months. The patient's HbA<sub>1c</sub> trajectory will be illustrated by the sequence, 'NNNUUAA'. Using the HbA<sub>1c</sub> trajectory mapping, we counted the occurrences of each n-grams in the sequence. We used n = 6 since most patients have their HbA<sub>1c</sub> tested every three to six months. There were two 6-grams in the above mapped sequence (i.e. 'NNNUUA' and 'NNUAAA') and both had a count of one.

We let d denote the total number of 6-grams in the mapped sequence of all the patients' HbA<sub>1c</sub> trajectories. A d-dimensional vector was used to represent a patient's trajectory with each dimension corresponds to a 6-gram. The ith entry represents the count of the occurrences of the ith 6-gram. The similarity between two patients P1, P2 with trajectory vectors u and v were computed as follows:

$$traj\_sim(P1, P2) = \frac{1}{1 + \|u - v\|_2}$$

The overall similarity, trajectory-D3K (T-D3K) of patients P1, P2 was determined as follows:

$$sim(P1, P2) = \max(D3K_{sim(P1,P2)}, traj_{sim}(P1,P2))$$

The proposed DMRS used the T-D3K method to generate the candidate set of medications for the target patient by retrieving prescription records of the similar patients (sorted by their overall similarity scores). For patients with the same similarity score, their medication lists were prioritized according to the degree of overlap with the target patient's list of medications and dosages. In the instance of a tie, we ranked the medication lists based on the number of therapies (monotherapy was preferred over higher number of therapy), and dosage (lower dosage was preferred over higher dosage).

**Evaluation.** The proposed DMRS was evaluated on four groups of patients with different comorbidities profile (DM only, DM with HLD, DM with HTN, and DM with HLD and HTN (DHL)). For each group, 100 test patients were randomly selected among those with suboptimal HbA<sub>1c</sub> (i.e. HbA<sub>1c</sub>  $\geq 8$ ) and have at least one DM medication adjustment. For each test patient, the top K candidate medications were retrieved and ranked as described in the "Method" section.

We used the prescribed DM medication type and dose in the EHR records as our ground truth and evaluated the accuracy of our recommendations using three metrics: hit ratio, recall@K, and precision@K. Hit ratio refers to the percentage of patients where at least one of the recommended sets of medication matched the ground truth in both type and dosage.

As described in the previous section, the sets of medications to be recommended for a target patient were obtained from the prescription records of the similar patients. Each target patient can receive up to K unique sets of recommended medication.

We let S denote one set of recommended medication and G denote the ground truth medication. The percentage of matches between the recommended medication in S and that in the ground truth G were measured by recall (S,G) and precision (S,G) as follows:

$$recall(S, G) = \frac{|G \cap S|}{|G|}$$

$$precision(S, G) = \frac{|G \cap S|}{|S|}$$

For each target patient, we computed recall@K as the maximum recall (S,G) and precision@K as the maximum precision (S,G) among the K sets of recommended medication. In addition, to evaluate the quality of the recommendations, we used Mean Reciprocal Rank (MRR)<sup>25</sup>. MRR takes into account the position of the matched recommendation in the list. A high MRR indicates that the matched recommendation was positioned near the top of the recommendation list.

Variables	Overall n = 54,933
Age, years	67.4 ± 11.4
Male (%)	27,018 (49.2)
Race (%)	
Chinese	38,150 (69.4)
Indian	5,429 (9.9)
Malay	8,789 (16.0)
Others	2,565 (4.7)
<b>Comorbidities (%)</b>	
Diabetes (DM only)	567 (1.0)
Diabetes and hyperlipidaemia (DM with HLD)	4,162 (7.6)
Diabetes and hypertension (DM with HTN)	1,505 (2.7)
Diabetes, hyperlipidaemia and hypertension (DHL)	48,699 (88.7)
<b>Disease duration, years</b>	
Diabetes (DM)	5.3 ± 2.1
Hyperlipidaemia (HLD)	5.5 ± 1.9
Hypertension (HTN)	5.5 ± 1.9
Blood pressure systolic (mmHg)	133.0 ± 16.7
Blood pressure diastolic (mmHg)	69.0 ± 9.4
Cholesterol high-density lipoprotein (HDL) (mmol/L)	1.3 ± 0.4
Cholesterol low-density lipoprotein (LDL) (mmol/L)	2.2 ± 0.9
Triglycerides (mmol/L)	1.6 ± 1.1
Haemoglobin A1c level (HbA <sub>1c</sub> ) (%)	7.3 ± 1.3
<b>Medications</b>	
Diabetes (DM) medication count <sup>1</sup>	
One medication	23,127 (42.1)
Two medications	16,587 (30.2)
Three or more medications	15,219 (27.7)
Hyperlipidaemia (HLD) medication count <sup>1</sup>	
No medication	5,838 (10.6)
One medication	43,910 (79.9)
Two medications	5,077 (9.2)
Three or more medications	108 (0.2)
Hypertension (HTN) medication count <sup>1</sup>	
No medication	8,559 (15.6)
One medication	16,819 (30.6)
Two medications	17,498 (31.9)
Three or more medications	12,057 (21.9)

**Table 1.** Characteristics of patients in the study cohort. <sup>1</sup>Medication count denotes count of patients on zero, one, two or three or more medications. Continuous variables were expressed as a mean value ± SD, while categorical variables expressed as the number of patients, n (%).

We compared our T-D3K approach of medication recommender systems with the following methods:

- Euclidean distance on normalized data.
- D3K. This approach retrieves similar patients and their prescribed medications using multiple clinical measurements and medication counts at a single time point.
- Trajectory. This approach retrieves similar patients and their prescribed medications based on HbA<sub>1c</sub> (%) trajectory over multiple time points.

**Ethics approval.** Ethics approval was obtained from SingHealth Centralized Institution Review Board (CIRB) in 2019 (SingHealth CIRB Reference: 2019/2604). Patient consent was not obtained as the analysis was conducted on de-identified data.

## Results

**Patients' characteristics.** Patients' characteristics based on their latest EHR across the 10-year observation period were shown in Table 1. The study cohort consisted of 54,933 diabetic patients with mean disease duration of 5.3 ± 2.1 years. 7.6% of the patients had a comorbidity of hyperlipidaemia (DM with HLD), 2.7% had hypertension (DM with HTN) and 88.7% had both hyperlipidaemia and hypertension (DHL). The mean age of

Patient group	Method	Hit ratio	Recall@10	Precision@10	MRR
DM <sup>1</sup> only	Euclidean	67	0.854	0.877	0.431
	D3K	76	0.923	<b>0.975</b>	0.474
	Trajectory	73	0.938	0.944	0.383
	T-D3K	<b>81</b>	<b>0.946</b>	0.954	<b>0.551</b>
DM with HLD <sup>2</sup>	Euclidean	63	0.849	0.901	0.378
	D3K	82	<b>0.945</b>	<b>0.998</b>	0.455
	Trajectory	70	0.905	0.94	0.367
	T-D3K	<b>84</b>	0.935	0.954	<b>0.526</b>
DM with HTN <sup>3</sup>	Euclidean	65	0.876	0.888	0.316
	D3K	76	0.907	<b>0.990</b>	0.397
	Trajectory	75	0.925	0.947	0.408
	T-D3K	<b>78</b>	<b>0.927</b>	0.945	<b>0.534</b>
DHL <sup>4</sup>	Euclidean	54	0.823	0.862	0.301
	D3K	61	0.879	0.944	0.303
	Trajectory	67	0.902	0.953	0.397
	T-D3K	<b>75</b>	<b>0.940</b>	<b>0.956</b>	<b>0.452</b>
Average of four patient groups	Euclidean	62.25	0.851	0.882	0.357
	D3K	73.75	0.914	<b>0.977</b>	0.407
	Trajectory	71.25	0.918	0.946	0.389
	T-D3K	<b>79.5</b>	<b>0.937</b>	0.952	<b>0.516</b>

**Table 2.** Comparison of the Hit ratio, Recall@10, Precision@10 and MRR of the various methods for the four groups of patients. <sup>1</sup>DM denotes patients with T2DM only, DM with HLD<sup>2</sup> denotes T2DM patients with hyperlipidaemia (HLD), DM with HTN<sup>3</sup> denotes T2DM patients with hypertension (HTN), and DHL<sup>4</sup> denotes T2DM patients with hyperlipidaemia and hypertension (DHL). <sup>5</sup>Hit ratio % is the proportion patients with recommendations that have an exact match (both the type and dosage) with prescribed medications. <sup>6</sup>Recall measures the number of common medication type and dose between recommended set (S) and ground truth set (G) over ground truth set (G). <sup>7</sup>Precision measures the number of common medication type and dose between recommended set (S) and ground truth (G) over recommended set (S). [Bold] figures represent the best performance among the four similarity approach for each comorbidities profile.

patients were  $67.4 \pm 11.4$  years with 49.2% being male. The cohort was predominantly Chinese (69.4%) followed by Malay (16.0%), followed by Indian (9.9%), which is similar to the ethnic distribution in the local Asian population. The average systolic and diastolic blood pressure (mmHg) were  $133.0 \pm 6.7$  and  $69.0 \pm 9.4$  respectively. Average cholesterol LDL mmol/L and HbA<sub>1c</sub>% were  $2.2 \pm 0.9$  and  $7.3 \pm 1.3$  respectively. Majority of the patients were on mono (42.1%) and dual diabetic medications (30.2%). More than 90% of the patients have at least one or more DM medication adjustment during their 10 years of visit records.

**Medication recommendations.** Results for the four groups of patients at K = 10 was shown in Table 2. Across the four groups of patients, we found that T-D3K consistently recorded the highest hit ratio and MRR. Averaged hit ratio, recall@10 and MRR for T-D3K were 79.50%, 0.94 and 0.52 respectively. D3K recorded the best averaged precision@10 at 0.98, approximately 2.6% higher than T-D3K (0.95).

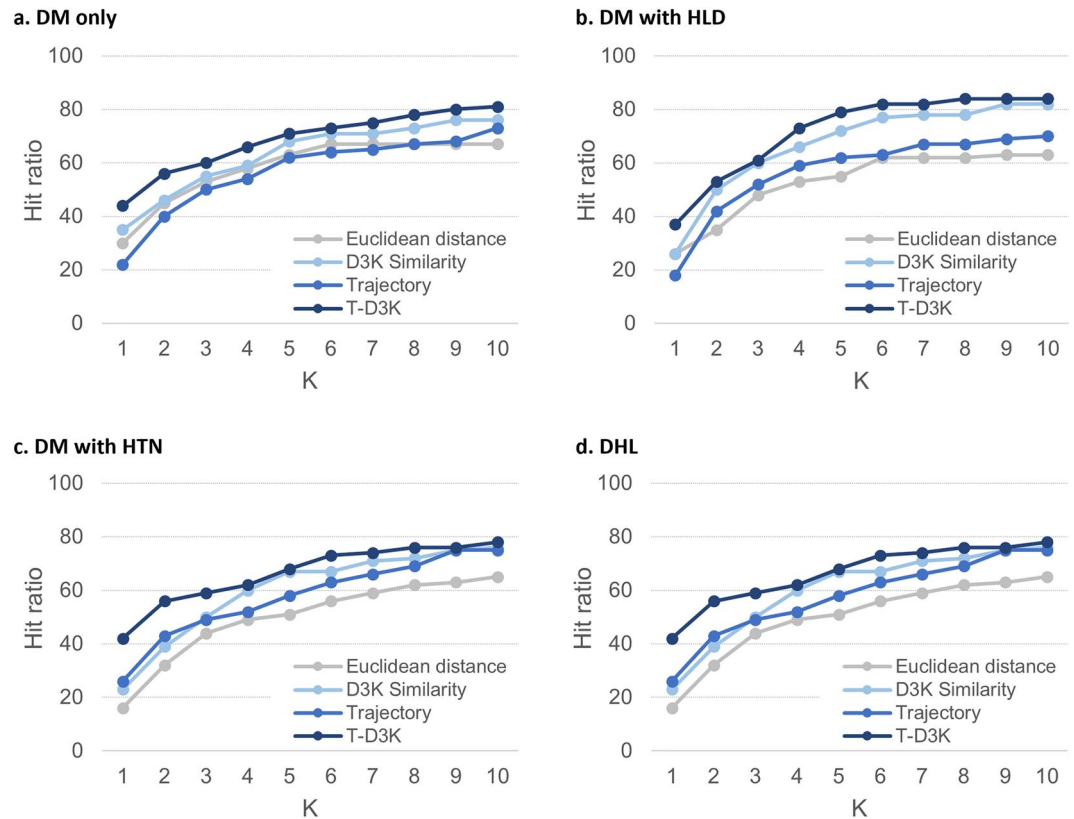
Comparing the performance of T-D3K across the four patient groups, hit ratio was the highest for patient group DM with HLD (84%) followed by patient group with DM only (81%), DM with HTN (78%) and DHL (75%). This suggest that out of the 100-sample patient, T-D3K provided recommendations that matched the actual prescriptions for at least 75 patients. Performance of T-D3K in recall@10 and precision@10 were largely similar across the four patient groups. Recall@10 and precision@10 for T-D3K ranged between 0.93–0.95 and 0.95–0.96 respectively. MRR for T-D3K was above 0.5 for all patient groups except for patient with DHL. A lower MRR indicates that T-D3K tend to make the correct recommendations at mid to bottom position of the recommendations list for the DHL patients group.

Figure 1 shows the hit ratio as we vary K for the four groups of patients. Overall, T-D3K approach achieved the highest hit ratio. At K = 10, the hit ratios of T-D3K were 81%, 84%, 78% and 75% for patient groups DM only, DM with HLD, DM with HTN, and DHL respectively.

Figure 2 shows the recall@K as we vary K for all four patient groups. T-D3K had the best performance for all four patient groups except patient group with DM with HLD. D3K recorded the highest recall for patient group DM with HLD at 0.95. Specifically, recall@10 for T-D3K were 0.95, 0.94, 0.93, 0.94 for patient groups DM only, DM with HLD, DM with HTN, and DHL respectively.

Figure 3 shows the precision@K for the four patient groups. T-D3K delivered the best performance for the DHL patient group, while D3K recorded the best precision for patient groups DM only, DM with HLD, DM with HTN. When K = 10, T-D3K had a precision of 0.95 for patient groups DM only, DM with HLD and DM with HTN, and 0.96 for DHL patient group.





**Figure 1.** Hit ratio for the four groups of patients.

Figure 4 shows the MRR results with T-D3K having the highest MRR for all the groups. When  $K = 10$ , MRR for T-D3K were 0.55, 0.53, 0.53 and 0.45 for patient groups DM only, DM with HLD, DM with HTN, and DHL respectively.

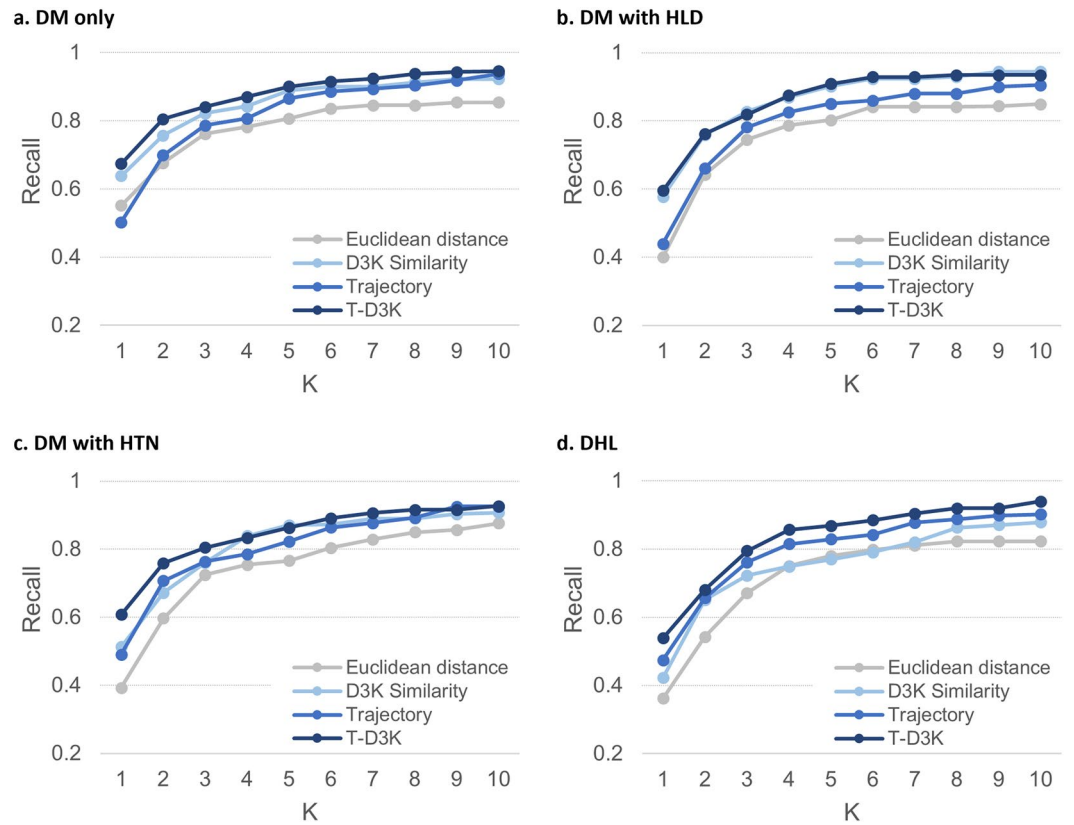
## Discussion

In this work, we proposed a DMRS using patients' clinical profiles and their  $HbA_{1c}$  test results trajectory. Our system used similar patient analytics to personalize the recommendations as we found that patients with clinically similar profile and diabetes disease trajectories were likely to share similar prescribed medication types and dosages. The set of similar patients retrieves also provided some form of explanation for the recommendation, thus enhancing system transparency, interpretability and trust for use in clinical practice.

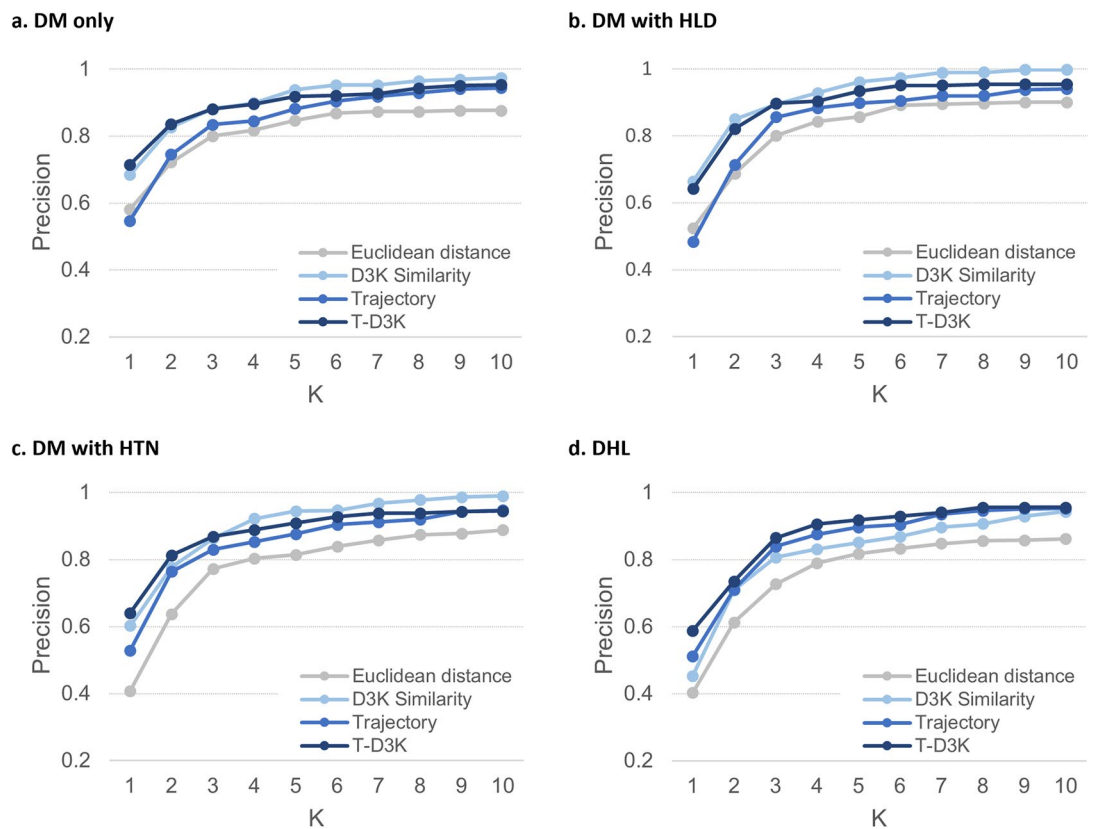
Previous works<sup>26,27</sup> on medicine recommendation system developed using patient similarity approach were mostly derived from a single time point. However, studies<sup>28–30</sup> have shown that  $HbA_{1c}$  trajectories (i.e. the delayed or metabolic memory) of patients with T2DM have important effects on diabetes outcomes. Non-stable  $HbA_{1c}$  trajectories are associated with greater risk of microvascular events and mortality<sup>28</sup>. This study extends existing works by considering patients'  $HbA_{1c}$  trajectories in addition to clinical profile when identifying similar patients. In contrast to studies<sup>20,21</sup> that provide recommendations for medications type only, our study provides both medication type and dosage recommendations. Integrating dosage recommendation increases the complexity of medication recommendation system as titration of dosages need to take into consideration the age of a patient, their clinical profile, and/or other interacting medications<sup>31</sup>.

Among the four patient groups, patients in DHL group recorded the lowest hit ratio and MRR. This result was consistent across the different methods. One possible explanation is that individualizing medication recommendations for patients with multiple, concurrent chronic conditions can be complex and challenging where the treatment for one condition may interfere with the treatment of other conditions<sup>32</sup>.

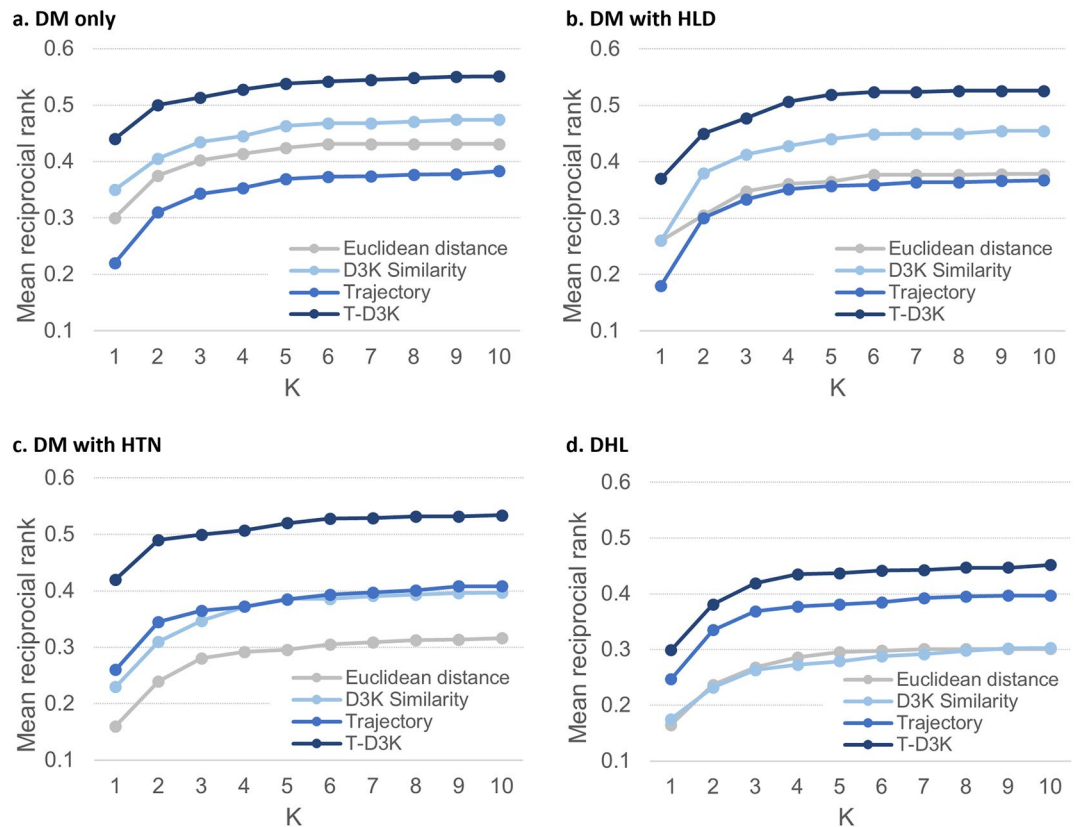
This study has several limitations. First, our study utilized EHR data from a community healthcare system. Hence, diabetes prescriptions were influenced by availability of government subsidies. Only medications classified under the standard drug list (SDL) were eligible for government subsidy<sup>33</sup>. As a result, biguanides and sulfonylureas prescriptions were overrepresented in EHR since they fall under SDL and were used as the first-line treatment for patients with suboptimal  $HbA_{1c}$ <sup>34</sup>. Second, the DMRS was developed using the EHR of six primary care clinics over a 10-year period. This may result in the recommendation of dated medications, e.g., use of Tolbutamide as Sulfonylurea agent. Further, the recommended medications tend not to include new drugs classes such as Dipeptidyl peptidase-4 (DPP-4) inhibitors and Sodium-glucose cotransporter-2 (SGLT-2) inhibitors which may reduce pill burden and improve patient adherence<sup>3</sup>. Third, dosage recommendations in this



**Figure 2.** Recall for four groups of patients.



**Figure 3.** Precision for four groups of patients.



**Figure 4.** Mean reciprocal rank (MRR) for four groups of patients.

work were limited to three intensity level (low, medium and high) so as to reduce the computational overhead. Lastly, insulin was regarded as a binary medication variable. Hence, the DMRS is best applied to clinical care of patients before secondary drug failure (i.e. failure of oral medications to maintain optimal glycaemic control).

**Clinical utility.** The DMRS is currently integrated as part of a module along with other diabetes risk stratification and prognostication modules in an easy-to-use web interface known as PERSONalized Diabetes Counselling Tool using Artificial Intelligence (PERDICT.AI)<sup>35</sup>. PERDICT.AI aims to support clinicians in diabetes consultations and to improve medication optimization. To use the DMRS, clinicians can either manually input or enter only the patient's identity number to automatically populate the patient's clinical profile (such as age, gender, blood pressure, cholesterol, HbA<sub>1c</sub> and existing medications). Using patient's latest clinical profiles and HbA<sub>1c</sub> trajectories, the DMRS will run at the backend of the PERDICT.AI interface, to retrieve a set of similar patients and their prescribed medications for diabetes medicine recommendation. The clinical application of PERDICT.AI will be tested at several primary care clinics in Singapore. The pilot study of the PERDICT.AI will induct patients with suboptimal glycaemic control via patient similarity approaches, which allows their understanding of their individual risk profiles, followed by recommendations of suitable pharmacotherapy using the DMRS. The mix-method study and the outcomes will be shared when the study is completed.

## Conclusion

The proposed DMRS is able to provide an individualized recommendation that is close to actual prescribed medication and dosage by taking into consideration patient's clinical profile and glycaemic control trajectories. Such a system is useful as a shared decision-making tool to assist clinicians in selecting the appropriate medications for patients with T2DM.

## Data availability

The datasets analysed in the current study are not publicly available as they contain information that are sensitive to the institution. They may be made available from the corresponding author on reasonable request.

Received: 4 August 2022; Accepted: 16 November 2022

Published online: 03 December 2022



## References

1. The Emerging Risk Factors, C. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *The Lancet* **375**, 2215–2222 (2010). [https://doi.org/10.1016/s0140-6736\(10\)60484-9](https://doi.org/10.1016/s0140-6736(10)60484-9)
2. Pontarolo, R. *et al.* Pharmacological treatments for Type 2 diabetes. *Treat. Type 2*, 147–184. <https://doi.org/10.5772/56456> (2015).
3. *ACE Clinical Guidances (ACGs)*, [https://www.ace-hta.gov.sg/docs/default-source/acgs/oral-glucose-lowering-agents-in-t2dm-\(updated-on-3-august-2017\).pdf](https://www.ace-hta.gov.sg/docs/default-source/acgs/oral-glucose-lowering-agents-in-t2dm-(updated-on-3-august-2017).pdf) (2017).
4. Gokhale, K. *et al.* The clinical profile and associated mortality in people with and without diabetes with Coronavirus disease 2019 on admission to acute hospital services. *Endocrinol. Diabetes Metab.* **5**, e00309. <https://doi.org/10.1002/edm2.309> (2022).
5. Guo, J. & Smith, S. M. Newer drug treatments for type 2 diabetes. *BMJ* <https://doi.org/10.1136/bmj.n1171> (2021).
6. Basu, S. *et al.* Expanding access to newer medicines for people with type 2 diabetes mellitus in low-and middle-income countries: a microsimulation and price target analysis. *The Lancet Diabetes and Endocrinology* (2021).
7. Chaplin, S. NICE guidance on managing type 2 diabetes in adults. *Prescriber* **27**, 22–29. <https://doi.org/10.1002/psb.1450> (2016).
8. Association, A. D. Introduction: Standards of medical care in diabetes—2022. **45**, S1–S2 (2022).
9. Grol, R., Hutchinson, A., Eccles, M. & Grimshaw, J. Clinical guidelines, potential benefits, limitations, and harms of guidelines. *BMJ* **318**, 527–530. <https://doi.org/10.1136/bmj.318.7182.527> (1999).
10. Hayward, R. S. A. Users' guides to the medical literature. *JAMA* **274**, 570–574. <https://doi.org/10.1001/jama.1995.03530070068032> (1995).
11. Graham, R. P., James, P. A. & Cowan, T. M. Are clinical practice guidelines valid for primary care?. *J. Clin. Epidemiol.* **53**, 949–954. [https://doi.org/10.1016/s0895-4356\(99\)00224-3](https://doi.org/10.1016/s0895-4356(99)00224-3) (2000).
12. Franco, J. V. A., Arancibia, M., Meza, N., Madrid, E. & Kopitowski, K. Clinical practice guidelines: Concepts, limitations and challenges. *Medwave* **20**, e7887. <https://doi.org/10.5867/medwave.2020.03.7887> (2020).
13. Rosenblit, P. D. Common medications used by patients with type 2 diabetes mellitus: What are their effects on the lipid profile?. *Cardiovasc. Diabetol.* **15**, 95. <https://doi.org/10.1186/s12933-016-0412-7> (2016).
14. Grant, R. W. *et al.* How doctors choose medications to treat type 2 diabetes: A national survey of specialists and academic generalists. *Diabetes Care* **30**, 1448–1453. <https://doi.org/10.2337/dc06-2499> (2007).
15. Tang, P. C. *et al.* Precision population analytics: Population management at the point-of-care. *J. Am. Med. Inform. Assoc.* **28**, 588–595. <https://doi.org/10.1093/jamia/ocaa247> (2021).
16. Ochoa, J. G. D., Csiszar, O. & Schimper, T. Medical recommender systems based on continuous-valued logic and multi-criteria decision operators, using interpretable neural networks. *BMC Med. Inform. Decis. Mak.* **21**, 186. <https://doi.org/10.1186/s12911-021-01553-3> (2021).
17. Bhoi, S., Li, L. M. & Hsu, W. Premier: Personalized recommendation for medical prescriptions from electronic records. arXiv preprint [arXiv:2008.13569](https://arxiv.org/abs/2008.13569) (2020). <https://doi.org/10.48550/arXiv.2008.13569>
18. Gogleva, A., papa, e., Jansson, E. & De Baets, G. in *Fifteenth ACM Conference on Recommender Systems*. 548–550.
19. Zhang, P., Wang, F., Hu, J. & Sorrentino, R. Towards personalized medicine: Leveraging patient similarity and drug similarity analytics. *AMIA Jt. Summits Transl. Sci. Proc.* **2014**, 132–136 (2014).
20. Ng, K., Kartoun, U., Stavropoulos, H., Zambrano, J. A. & Tang, P. C. Personalized treatment options for chronic diseases using precision cohort analytics. *Sci. Rep.* **11**, 1139. <https://doi.org/10.1038/s41598-021-80967-5> (2021).
21. Oh, S. H., Lee, S. J., Noh, J. & Mo, J. Optimal treatment recommendations for diabetes patients using the Markov decision process along with the South Korean electronic health records. *Sci. Rep.* **11**, 6920. <https://doi.org/10.1038/s41598-021-86419-4> (2021).
22. Mei, J., Liu, H., Li, X., Xie, G. T. & Yu, Y. in *Medinfo*. 300–304.
23. Giugliano, D. *et al.* Simplification of complex insulin therapy: A story of dogma and therapeutic resignation. *Diabetes Res. Clin. Pract.* **178**, 108958. <https://doi.org/10.1016/j.diabres.2021.108958> (2021).
24. Oei, R. W. *et al.* Using domain knowledge and data-driven insights for patient similarity analytics. *J. Pers. Med.* **11**, 699. <https://doi.org/10.3390/jpm11080699> (2021).
25. Craswell, N. Mean reciprocal rank. *Encycl. Database Syst.* [https://doi.org/10.1007/978-0-387-39940-9\\_488](https://doi.org/10.1007/978-0-387-39940-9_488) (2009).
26. Morales, L. F. G., Valdiviezo-Diaz, P., Reátegui, R. & Barba-Guaman, L. Drug recommendation system for diabetes using a collaborative filtering and clustering approach: development and performance evaluation. *J. Med. Internet Res.* **24**, e37233 (2022).
27. Liu, H. *et al.* An efficacy driven approach for medication recommendation in type 2 diabetes treatment using data mining techniques. *Stud. Health Technol. Inform.* **192**, 1071–1071 (2013).
28. Laiterapong, N. *et al.* Ten-year hemoglobin A1c trajectories and outcomes in type 2 diabetes mellitus: The diabetes & aging study. *J. Diabetes Complications* **31**, 94–100. <https://doi.org/10.1016/j.jdiacomp.2016.07.023> (2017).
29. Holman, R. R., Paul, S. K., Bethel, M. A., Matthews, D. R. & Neil, H. A. 10-year follow-up of intensive glucose control in type 2 diabetes. *N. Engl. J. Med.* **359**, 1577–1589. <https://doi.org/10.1056/NEJMoa0806470> (2008).
30. Group & U. P. D. S., Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *The Lancet* **352**, 837–853. [https://doi.org/10.1016/s0140-6736\(98\)07019-6](https://doi.org/10.1016/s0140-6736(98)07019-6) (1998).
31. Powell, J. R., Cook, J., Wang, Y., Peck, R. & Weiner, D. Drug dosing recommendations for all patients: A roadmap for change. *Clin. Pharmacol. Ther.* **109**, 65–72. <https://doi.org/10.1002/cpt.1923> (2021).
32. Vogeli, C. *et al.* Multiple chronic conditions: Prevalence, health consequences, and implications for quality, care management, and costs. *J. Gen. Intern. Med.* **22**(Suppl 3), 391–395. <https://doi.org/10.1007/s11606-007-0322-1> (2007).
33. Ministry of Health, S. Drugs subsidies and schemes, <https://www.moh.gov.sg/cost-financing/healthcare-schemes-subsidies/drug-subsidies-schemes> (2022).
34. Ministry of Health, S. *Clinical Practice Guidelines - Diabetes Mellitus*, [https://www.moh.gov.sg/docs/librariesprovider4/guidelines/cpg\\_diabetes\\_mellitus-booklet---jul-2014.pdf](https://www.moh.gov.sg/docs/librariesprovider4/guidelines/cpg_diabetes_mellitus-booklet---jul-2014.pdf) (2014).
35. JARVIS DHL: Transforming Chronic Care for Diabetes, Hypertension and hyperLipidemia with AI, <https://ids.nus.edu.sg/aiinhalth.html> (2019).

## Author contributions

W.Y.T., W.H. and M.L.L. conceptualized and formulated the method. W.Y.T., R.W.O. and Q.G. implemented the methods. W.Y.T. performed the formal data analysis. W.Y.T. and Q.G. wrote the original draft. W.H. and M.L.L. supervised, reviewed, and edited the manuscript. N.C.T. reviewed and edited the manuscript. All authors reviewed and approved the manuscript.

## Funding

This research is supported by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG-GC-2019-001-2A). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of National Research Foundation, Singapore. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-24494-x>.

**Correspondence** and requests for materials should be addressed to W.Y.T.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022