

## SUPPLEMENTARY DATA

Hosomura N, Goldberg SI, Shubina M, Zhang M, Turchin A. Electronic Documentation of Lifestyle Counseling and Glycemic Control in Patients with Diabetes.

**Supplementary Figure 1.** Examples of Levenshtein distance calculation

**Supplementary Figure 2.** Example of normalized Levenshtein distance calculation

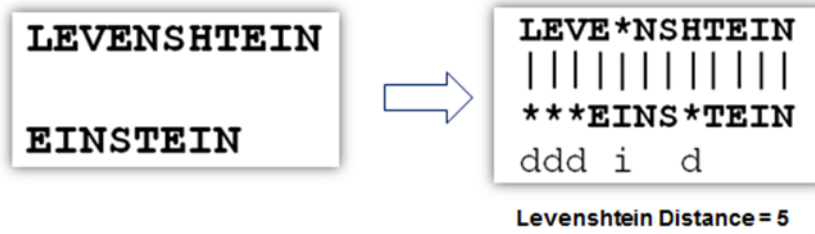
**Supplementary Appendix.** Calculation of documentation heterogeneity

**Supplementary Figure 3.** Sample data and calculation results of documentation heterogeneity and documentation intensity

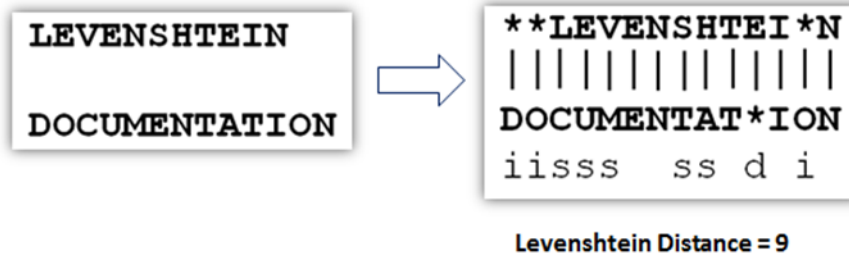
**Supplementary Table.** Results of sensitivity analyses showing hazard ratios for association between documentation characteristics and time to hemoglobin A1c control as estimated by multivariable Cox regression models

**Supplementary Figure 1. Examples of Levenshtein distance calculation**

**Example 1**

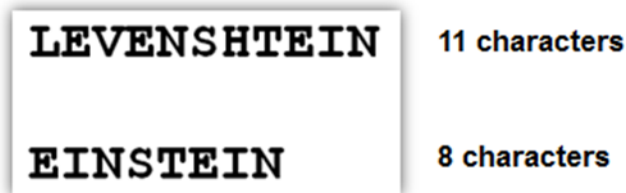


**Example 2**



Editing operations could include insertion of a character, deletion of a character, and/or substitution of a character by another character.

**Supplementary Figure 2. Example of normalized Levenshtein distance calculation**



Levenshtein distance = 5

Normalized Levenshtein distance =  $5/11 = 0.45$

Since Levenshtein distance is affected by the lengths of the two strings, rather than using the raw values, we normalized Levenshtein distance by dividing the value by the length (in characters) of the longer string.

## SUPPLEMENTARY DATA

### Supplementary Appendix 1. Calculation of documentation heterogeneity

We calculated the normalized Levenshtein distance for two consecutive notes by pairwise comparisons of the sentences documenting lifestyle counseling between the two notes. This operation was repeated for all the consecutive notes with counseling documentation for the entire hyperglycemic period. Finally, we took the mean value of the between-note normalized Levenshtein distance values as the documentation heterogeneity for the hyperglycemic period.

### Supplementary Figure 3. Sample data and calculation results of documentation heterogeneity and documentation intensity

#### Hyperglycemic Period 1

NoteID	Sentence
0001	We had a long discussion trying to clarify his diet
0001	Counseled regarding portion control, dietary changes, and need for regular exercise
0002	Reinforced portion control in effort to lose weight
0002	Low-fat, low-cholesterol diet was encouraged
0003	He is monitoring his dietary intake but does have dietary indiscretion
0003	He is well informed regarding portion control and is trying to work on this
0004	Emphasized to him again how critical watching his diet will be
0004	No concentrated sweet, low-carbohydrate diet were encouraged
0005	He has been modifying his diet to avoid sweets and has been feeling overall well
0005	Will stay the course, emphasized diet control

#### Hyperglycemic Period 2

NoteID	Sentence
0006	Reviewed diet and exercise
0007	Reviewed diet and exercise to lose weight
0008	Exercise, weight loss, and low-fat diet were encouraged
0009	Again, exercise, weight loss, and low-fat diet were encouraged
0010	Exercise efforts reinforced

	Documentation Heterogeneity*	Documentation Intensity (characters/note)
Hyperglycemic Period 1	0.81	124
Hyperglycemic Period 2	0.43	42

\*Documentation heterogeneity represented by the normalized Levenshtein distance.

SUPPLEMENTARY DATA

**Supplementary Table 1. Results of sensitivity analyses showing hazard ratios for association between documentation characteristics and time to hemoglobin A1c control as estimated by multivariable Cox regression models**

Multivariable Cox Regression Model Descriptions	Documentation Heterogeneity <sup>a</sup>	Documentation Intensity <sup>b</sup>
Model adjusted for clustering within individual providers	1.16 (1.10-1.23)	1.11 (1.07-1.16)
Model including patients with no income information, no medication information, treatment by endocrinologists, and only transient HbA1c elevation, with provider age and gender included as additional covariates	1.18 (1.14-1.22)	1.09 (1.05-1.12)
Model limited to patients with diagnosed diabetes	1.05 (1.03-1.10)	1.28 (1.24-1.32)
Model using data on or before 12/21/2004	1.06 (1.01-1.12)	1.36 (1.29-1.43)
Model using data after 12/21/2004	1.11 (1.06-1.16)	1.15 (1.11-1.20)

Values in parentheses are 95% confidence intervals. <sup>a</sup>Documentation heterogeneity represented by normalized Levenshtein distance. <sup>b</sup>Documentation intensity represented by characters/note