

1 **Appendix 7**

2 **Appendix 7(1)**

3 The number of ICD-10 codes and morphemes and the scores of realities for all four
4 types of medical notes were not normal according to Shapiro–Wilk tests, and the
5 variances of the ICD-10 codes and morphemes and the scores of realities for the
6 three types of notes were found to be unequal with those of the actual medical notes
7 by Bartlett tests. Therefore, we used the nonparametric Brunner–Munzel tests to
8 compare the average of the number of ICD-10 codes and morphemes and the scores
9 of realities between the actual medical notes and the other three types of notes.
10 With regard to readability, the scores for all four types of notes were not normal
11 under the Shapiro–Wilk tests. The variances for the three types of medical notes
12 were found to be equal to that of the actual notes, according to Bartlett tests. Thus,
13 we used Mann–Whitney U tests to compare the average of the readability scores
14 between the actual and other three types of medical notes

15

16

17 **Appendix 7(2)**

18 When $p(x) = 0$ or $q(x) = 0$, 1.0×10^{-5} was inserted to $p(x)$ or $q(x)$.

19

20

21 Appendix 7(3)

22 TABLE. COMPARISON OF THE MEDIAN SCORE OF EACH METRIC FOR THE THREE TYPES OF
23 MEDICAL NOTES AND THE ACTUAL MEDICAL NOTES.

	ICD-10	Morphemes	Readability	Realities
Crowd	<i>p</i> = 0.534 0.39 to 0.56	<i>p</i> = 0.36 0.45 to 0.62	<i>p</i> = 0.536 −6.26e−05 to 1.00+e0	<i>p</i> = 0.001 0.26 to 0.44
Clinicians	<i>p</i> = 0.022 0.20 to 0.48	<i>p</i> = 0.003 0.15 to 0.42	<i>p</i> = 0.0009 1.00 to 4.00	<i>p</i> = 1.8e−05 0.22 to 0.39
Public	<i>p</i> = 1.9e−10 0.22 to 0.34	<i>p</i> = 1.3e−08 0.25 to 0.38	<i>p</i> = 0.028 5.19e−05 to 1.00+e0	<i>p</i> = 6.5e−07 0.20 to 0.37

24 Note: 95% CI means 95% confidence interval. Italics and bold means that P value <
25 0.001. Crowd: Medical notes generated by the crowd workers. Clinicians: Dummy
26 medical notes created by medical doctors. Public: Public dummy medical notes.

27

28