

1 **Appendix 2**

2 **1. Data from University of Tsukuba Hospital**

3 We used progress notes, medication data, laboratory test results, and injection
4 records of 245 Japanese patients (female: 48.2%, average age: 60.1 years, standard
5 deviation (SD): 16.2 years). These patients were randomly selected from the 53,246
6 patients who visited one or more of the 11 departments at University of Tsukuba Hospital
7 between January 1, 2013 and September 30, 2018. The progress notes used here were the
8 same as the existing dummy progress notes dataset referred to in Section 2.2. The data
9 were limited to the patients who visited these 11 departments because their basic
10 characteristics were suitable for the existing dummy progress notes dataset. All data were
11 preprocessed to ensure the protection of patient privacy before the analysis was carried
12 out. For more information, please refer to the Appendix 1. In this paper, we refer to these
13 preprocessed progress notes as the “actual progress notes.” The actual progress notes
14 used for this study were written in the Subjective, Objective, Assessment, and Plan
15 (SOAP) format [32]. Notes that did not follow the SOAP format were excluded from this
16 work, because we needed to decompose the progress notes semantically. For more
17 information, please see Section 2.5.

18

19 **2. Existing dummy progress notes dataset**

20 We used existing public dummy progress notes [34] for comparisons and
21 evaluations. We used the notes of only 29 dummy patients, as these followed the SOAP
22 format and could be decomposed semantically for this study. Among these notes, the

assessment (A) and plan (P) were concurrently described as the assessment and plan (A/P) in the progress notes of 22 dummy patients. Therefore, A and P were concurrently treated as A/P in this study for these progress notes and for the progress notes from University of Tsukuba Hospital. In all, 257 dummy progress notes of 29 dummy patients were decomposed into the subjective data (S), objective data (O), or A/P categories, and 514 decomposed dummy progress notes were generated.

Information about the medical departments of medical the doctors who wrote the existing public dummy progress notes was not disclosed in the dataset. One doctor (the first author) evaluated the department for each dummy progress note. As stated in Section 2.1, there were 11 departments included, namely cardiology, respiratory, gastroenterology, nephrology, hematology, collagen diseases, stroke, endocrinology, digestive surgery, oncology, and breast thyroid endocrinology surgery.