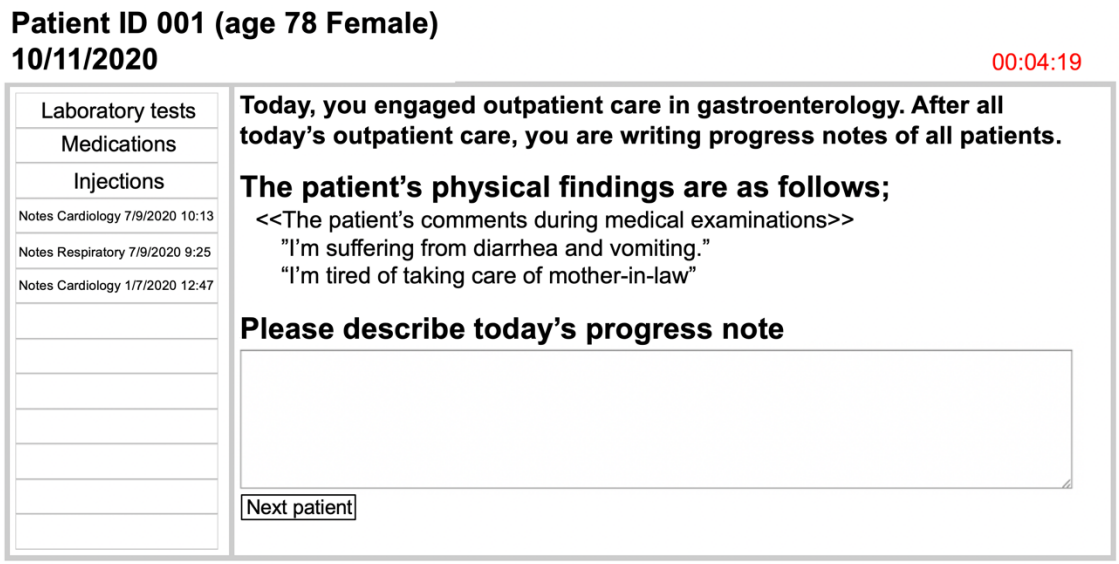
# Appendix 5

## Appendix 5(1)

The experimental computer screen was similar to the one used with Japan's actual electronic health records, where medical doctors write medical notes as a part of their daily clinical practice.

In actual daily clinical practice, progress notes may be written after each physical examination or laboratory test. To make the experimental environment as close to daily clinical practice as possible so that physicians can easily describe the pseudo-progress notes, the information necessary for the description was presented on the experiment screen. The physicians who participated in this study described the pseudo-progress notes based on that information. The experiment screen imitated the actual Electronic Health Records in Japan. The laboratory test results, prescriptions, and injections of the day in this figure and the pseudo-patients' past progress notes cannot be viewed without opening new tabs by clicking the bottom left-hand corner. The number of past progress notes physicians were able to check randomly selected from one, three, five, seven, and ten. In this figure, the number of past progress notes clinicians could check was three. To prevent copy and pasting, the information required for the description was presented in the form of images. A timer was displayed on the top right-hand corner of the screen, prompting the clinician to describe the pseudo-progress notes within five minutes to make the experimental environment close to the hectic daily clinical practice. These were originally written in Japanese but are provided in English for explanation purposes.




Example of the computer screen where clinicians described pseudo-progress notes for evaluation of the generated pseudo-progress notes

## Appendix 5(2)

The number of ICD-10 codes and morphemes and the scores of absolute realities for all four types of medical notes were not normal according to Shapiro–Wilk tests, and the variances of the ICD-10 codes and morphemes and the scores of absolute realities for the three types of notes were found to be unequal with those of the actual medical notes by Bartlett tests. Therefore, we used the nonparametric Brunner–Munzel tests to compare the average of the number of ICD-10 codes and morphemes and the scores of absolute realities between the actual medical notes and the other three types of notes. With regard to readability and relative reality, the scores for all four types of notes were not normal under the Shapiro–Wilk tests. The variances for the three types of medical notes were found to be equal to that of

the actual notes, according to Bartlett tests. Thus, we used Mann–Whitney U tests to compare the average of the readability and relative reality scores between the actual and other three types of medical notes

# Appendix 5(3)

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

| | **ICD-10** | **Morphemes** | **Readability** | **Absolute Realities** | **Relative Realities** |
|---|---|---|---|---|---|
| **Crowd** | *p* = 0.534<br>0.39 to 0.56 | *p* = 0.36<br>0.45 to 0.62 | *p* = 0.536<br>−6.26e−05 to 1.00+e0 | *p* = 0.001<br>0.26 to 0.44 | *p* = 0.151<br>0.42 to 0.92 |
| **Doctors** | *p* = 0.022<br>0.20 to 0.48 | *p* = 0.003<br>0.15 to 0.42 | *p* = 0.0009<br>1.00 to 4.00 | *p* = 1.8e−05<br>0.22 to 0.39 | *p* = 0.035<br>0.51 to 0.97 |
| **Public** | *p* = 1.9e−10<br>0.22 to 0.34 | *p* = 1.3e−08<br>0.25 to 0.38 | *p* = 0.028<br>5.19e−05 to 1.00+e0 | *p* = 6.5e−07<br>0.20 to 0.37 | *p* = 0.017<br>0.56 to 0.99 |

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