Online supplementary material

The relationship between incidence and prevalence is described as follows [21-23].

Prevalence rate = Incidence per year × Average disease duration (years).

In order to establish the above relationship, it is necessary that the number of new events is stable and we have data on event duration. Therefore, we examined the distribution of the events as described above, and the slope of the number of annual events was analyzed using a linear mixed-effects model. The mean survival time, equivalent to the average disease duration, was estimated using the Kaplan–Meier method, with the endpoints set as death from the disease and the period since the date of GMAb measurement during the follow-up period. If the observation was censored after the longest survival time, we used the restricted mean survival time (RMST) [26-30]. We set the truncation time point for the RMST calculation to a sufficiently long observation period to approximate the average survival time. We used a log-rank test to compare the difference in the cumulative rates between two or more groups.

Continuous variables were dichotomized at the median. Variables that achieved statistical significance in the log-rank test were subsequently included in a multivariate analysis, using a stepwise forward Cox proportional hazard model. The hazard ratio of the group above the median was compared with that of the group below the median. All statistical tests were two-sided, and significance was inferred at $P < 0.05$. We performed analyses using SAS 9.4 (SAS Institute, Cary, NC, USA), SPSS 25 (IBM, Chicago, IL, USA), and R 3.4.2 (Free Software Foundation, Boston, MA, USA).

As byproducts of the survival rates, we were able to evaluate the associations of various parameters at the time of GMAb testing with various prognostic endpoints, such as death, spontaneous improvement, improvement by treatment, and worsening, using univariate and multivariate survival analyses. Univariate analysis showed that the probability of improvement by treatment was significantly associated with dust exposure, disease severity, complications, PaCO2, KL-6, LDH, and %DLCO. According to the forward stepwise multivariate analysis, only PaCO2 had an independent effect ($P = 0.017$).

Specifically, the probability of improvement as a result of treatment for APAP increased when PaCO2 was > 38 mmHg (P = 0.013). The 2-, 5-, and 11-year worsening rates were 20.2%, 35.0%, and 40.0%, respectively, and the restricted-average worsening time was about 138 months (95% CI = 114, 163). Univariate analysis showed that the probability of worsening was significantly associated with disease severity, PaO2, AaDO2, VC, and FVC. According to forward stepwise multivariate analysis, only PaO2 and VC had independent effects. Specifically, the probability of worsening increased when PaO2 was < 71.2 mmHg (P = 0.013) or when VC was < 3.2 L (P = 0.022).

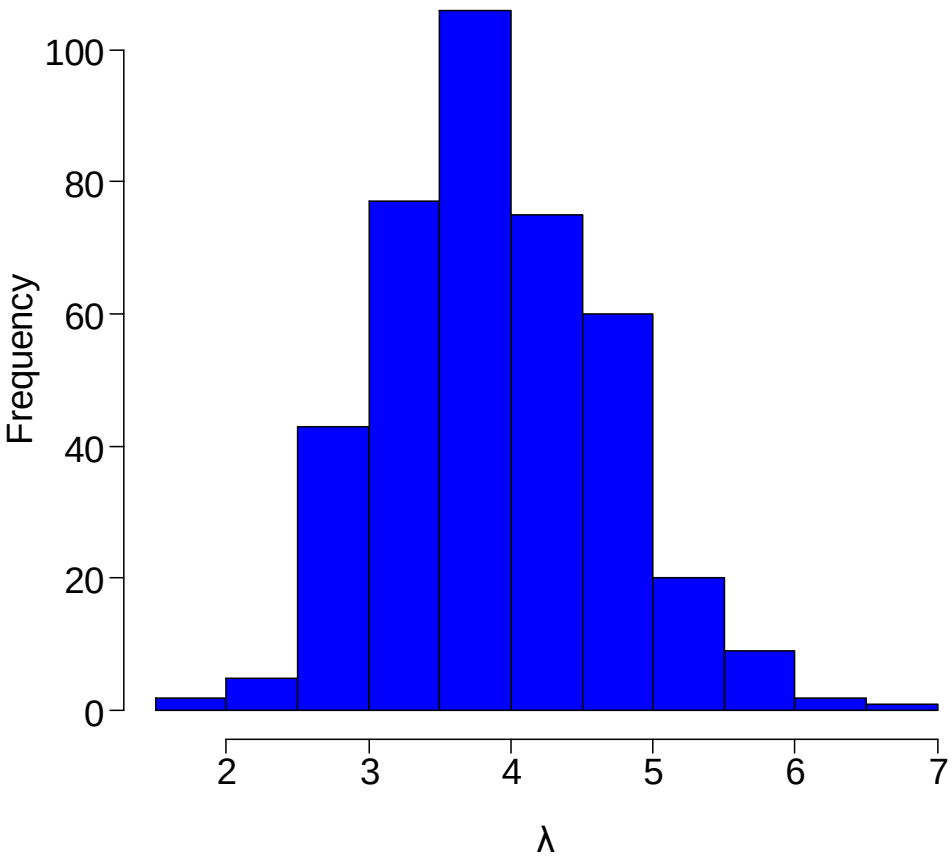

Figure S1. Distribution of annual number of APAP patients in Niigata prefecture from 2006 to 2016 by the bootstrap method with iteration times of 400.

Table S1. Bootstrap estimates for Poisson distribution

				95%CI	
Bootstrap estimate	mean	statistic	P−value	Lower	Upper

Poisson parameter	λ	3.9	0.3	0.4	2.5	5.5

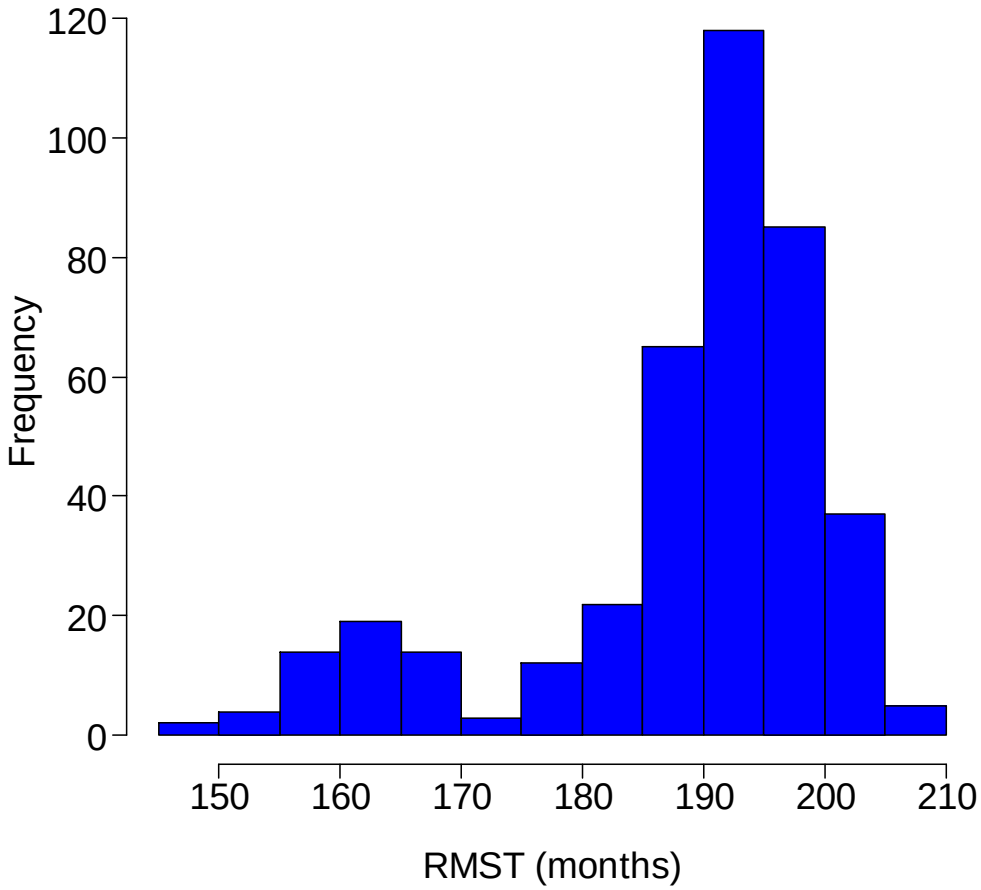

Figure S2. Distribution of RMST after diagnosis of 103 patients registered at 3 university hospitals by the bootstrap method with iteration times of 400.

Table S2. Bootstrap estimates for RMST after diagnosis

		95%CI	

Bootstrap estimate	mean	Lower	Upper
RMST	15.7	13.2	17.0