Incidence of autoimmune pulmonary alveolar proteinosis estimated using Poisson distribution

Nobutaka Kitamura[1], Shinya Ohkouchi[2], Ryushi Tazawa[3], Haruyuki Ishii[4], Toshinori Takada[5], Takuro Sakagami[6], Takahiro Tanaka[1] and Koh Nakata[1]

Affiliations: [1]Clinical and Translational Research Center, Niigata University Medical and Dental Hospital, Niigata, Japan. [2]Dept of Respiratory Medicine and Dept of Occupational Health, Tohoku University Graduate School of Medicine, Sendai, Japan. [3]Bioscience Medical Research Center, Niigata University Medical and Dental Hospital, Niigata, Japan. [4]Dept of Respiratory Medicine, Kyorin University School of Medicine, Tokyo, Japan. [5]Uonuma Institute of Community Medicine, Niigata University Medical and Dental Hospital, Minami-Uonuma, Japan. [6]Dept of Pulmonary Medicine, Kumamoto University Hospital, Kumamoto, Japan.

Correspondence: Koh Nakata, Clinical and Translational Research Center, Niigata University Medical and Dental Hospital, 1-754 Asahimachi-dori, Niigata 951-8520, Japan. E-mail: radical@med.niigata-u.ac.jp

ABSTRACT   The incidence and prevalence of autoimmune pulmonary alveolar proteinosis in Japan were previously estimated to be 0.49 and 6.2 per million, respectively. Thereafter, an increase in serological diagnosis forced a re-estimation of the incidence based on more contemporaneous data using more robust methods.

Sera of 702 patients were positive for granulocyte-macrophage colony-stimulating factor autoantibody during the 2006–2016 period (group A). Of these patients, 43 were actively surveyed in Niigata prefecture (group B) for estimation of the incidence. To estimate the survival period, 103 patients (group C) were investigated retrospectively for the 1999–2017 period using restricted mean survival time.

In group A, the number of patients diagnosed in each prefecture was closely correlated with the corresponding population, indicating no regional integration of onset. In group B, a total of 43 patients were diagnosed, the annual number followed a Poisson distribution and the incidence was thus estimated to be 1.65 per million. In group C, the retrospective cohort study revealed the mean survival period to be 16.1 years. Taken together, the prevalence was estimated to be 26.6 per million, indicating that the previous data for incidence and prevalence was an underestimation.

@ERSpublications

The latest epidemiologic study of autoimmune pulmonary alveolar proteinosis revealed the incidence and prevalence, estimated using Poisson distribution, to be 1.65 and 26.6 per million, respectively http://ow.ly/Wyyd30n4IYP

This article has supplementary material available from openres.ersjournals.com

Introduction

Pulmonary alveolar proteinosis is characterised by the abnormal accumulation of surfactant in the alveoli and terminal bronchioli, resulting in varying degrees of hypoxaemic respiratory insufficiency [1, 2]. Acquired pulmonary alveolar proteinosis is a syndrome consisting of two distinct diseases that can be grouped into disorders affecting surfactant clearance or surfactant production. It can be further subdivided into primary pulmonary alveolar proteinosis, which is caused by the disruption of granulocyte-macrophage colony-stimulating factor (GM-CSF) signalling and results in functional defects in alveolar macrophages and neutrophils, and secondary pulmonary alveolar proteinosis, which is caused by the presence of another underlying disease [3].

The discovery of GM-CSF autoantibody (GMAb) in idiopathic pulmonary alveolar proteinosis patients in 1999 [4] promoted world-wide interest in this disease, which led to shifts in our understanding and to the development of the current concepts of pulmonary alveolar proteinosis. This in turn led to the disease becoming known as autoimmune pulmonary alveolar proteinosis (APAP) [5, 6]. APAP is the most studied subgroup of pulmonary alveolar proteinosis and accounts for more than 90% of cases. Our previous national pulmonary alveolar proteinosis registry study [6] reported 223 patients with APAP between 1999 and 2006.

In studies based on national APAP registry data for 1999–2006 [6] we estimated the minimum incidence and prevalence to be 0.49 and 6.2 per million, respectively, as evaluated using two approaches. Approach one encompassed nine non-overlapping regions representing the entirety of Japan and approached two focussed on Niigata Prefecture. Since then we have continued to collect and measure the GMAb concentration of sera from across Japan. As the use of GMAb in diagnosing APAP became more prevalent, the number of cases increased, reaching 80–100 cases annually and 872 cases cumulatively by the end of 2016. We therefore realised that the initial incidence and prevalence rates were underestimates. In the present study, we re-evaluate the incidence and prevalence of APAP based on contemporary data using a more robust epidemiological method.

Based on probability theory and statistics, the incidence of APAP is expected to follow a Poisson distribution. The Poisson distribution reflects the probability of a certain number of disease onsets when the population is large but the likelihood of disease onset is low [7–10]. Importantly, the mean and variance of the Poisson distribution are equal and the distribution is based on only one parameter [7–10]. This parameter is estimated from the number of onsets observed, without the need for any supplementary information from survey targets [7–10].

Therefore, to estimate incidence we applied a Poisson distribution to the number of cases diagnosed annually in Niigata Prefecture, where an active survey was conducted from 2006 to 2016. In addition, to estimate prevalence we investigated the prognosis of 103 cases retrospectively and calculated the mean survival period.

Methods

Study design and patients

Subjects enroled in this study consisted of three groups, group A, group B and group C (figure 1). Since 1999, autoantibody concentrations in sera from across Japan have been measured centrally at the Bioscience Medical Research Center in Niigata University Hospital [11]. By the end of 2016, a total of 952 Japanese cases diagnosed as pulmonary alveolar proteinosis by bronchoalveolar lavage (BAL) or lung biopsy had undergone serological diagnosis and, of these, 872 (91.6%) were positive for GMAb. In the present study we focused on 702 of these cases whose sera were sent to us between 2006 and 2016 (group A). In applying the serological diagnosis we asked the treating doctors to provide the following patient information: age, gender, registration date and name of hospital visited. We then created a nationwide database of cases based on this information. During the same period we conducted an active survey of serological screenings in Niigata Prefecture, making use of personal relationships and communication methods such as phone calls, emails and letters between regional pulmonary physicians, 98% of whom had received their training at Niigata University. Forty-three patients were enrolled by this survey (group B) and group B was thus completely a subset of group A (albeit consisting of cases collected entirely in Niigata prefecture).

To estimate the patient survival period, we enroled all the cases who had ever been diagnosed at Tohoku (n=38), Niigata (n=47) and Kyorin (n=18) University Hospitals, located approximately 350 km from each other (group C). Thirty-nine cases in group B belonged to group C, 84 cases in group C belonged to group A and 19 cases in group C did not belong to group A. We collected the following data for each patient: date of GMAb measurement, GMAb level, age at diagnosis, gender, smoking status, dust exposure, disease severity, symptoms, complications, pulmonary function measures (e.g. arterial oxygen tension
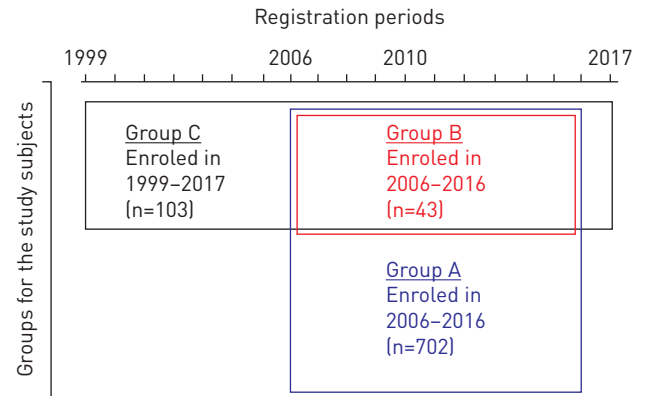
Registration periods

FIGURE 1 Venn diagram for the subjects enrolled in this study. Sera from all subjects were sent to Niigata University Hospital and were judged to be positive for the granulocyte-macrophage colony-stimulating factor autoantibody. Group A consisted of 702 cases whose sera were sent from across Japan between 2006 and 2016. Of this group, 43 cases enroled in Niigata prefecture between 2006 and 2016 were assigned to group B (which was thus entirely a subset of group A and consisted only of cases collected by an active survey in Niigata prefecture). Group C consisted of 103 cases who were diagnosed at Tohoku (n=38), Niigata (n=47) and Kyorin (n=18) University Hospitals, respectively and enroled between 1999 and 2017. Thirty-nine cases in group B belonged to group C, 84 cases in group C belonged to group A and 19 cases in group C did not belong to group A.

(PaO$_2$) and alveolar–arterial oxygen tension difference (PA–aO$_2$, alveolar–arterial oxygen gradient)), serum biomarkers (*e.g.* mucin-like glycoprotein (KL-6) and lactate dehydrogenase (LDH)), survival time to death due to APAP, spontaneous improvement, improvement as a result of treatment and worsening since serological diagnosis. These data were then collated into a database. The study was approved by the Institutional Review Boards of Niigata, Tohoku and Kyorin Universities. All pulmonary physicians agreed to collaborate with us and patient data were anonymised in a linkable manner.

Included patients
APAP diagnosis was based on cytological analysis of bronchoalveolar lavage fluid (BALF) or pulmonary histopathological findings, including both high-resolution computed tomography (HRCT) appearance and positive serum GMAb levels ($\geqslant 1.0\ \mu g \cdot mL^{-1}$) [12].

Excluded patients
Individuals were excluded if they did not have a pathologically proven diagnosis of pulmonary alveolar proteinosis or if serum was unavailable for the analysis of GMAb levels.

Disease severity score
Disease severity score (DSS) was based on the presence of symptoms and the degree of reduction in PaO$_2$ as recorded at registration, the latter being determined while the subject was breathing room air in the supine position, as described previously [6].

Evaluation of prognosis
Death due to APAP was recorded by the treating physicians during the follow-up period. All patient prognostic records were collected by the end of June 2017.

GMAb measurement
Serum samples from each participant were stored at −80 °C and the GMAb concentration was measured using ELISA, as reported previously [13].

Serum biomarker measurements
Serum KL-6 was measured using ELISA with commercial kits (ED046, Eizai, Tokyo, Japan), as described previously [14, 15].

Statistical analysis
The databases included both continuous and qualitative variables. Relationships between continuous variables were determined using Spearman's rank correlation coefficient (ρ). Assuming that the number of

events followed a Poisson distribution, the average number of events in an interval was designated λ (the parameter defining the Poisson distribution) to calculate the mean [7]. The standard error is the square root of the value obtained by dividing λ by the number of events [7]. For the Kaplan–Meier analysis, we used the restricted mean survival time (RMST) method [16–20]. A bootstrap approach with iteration times of 400 was used to assess the validity of our estimation for the parameters. Additional details concerning the statistical analysis are provided in the supplementary material.

Results

Between 2006 and 2016, 702 cases from across Japan were diagnosed with APAP (group A). The hospitals that sent sera from these cases to us for GMAb measurement were mainly located in big cities and were distributed from the subarctic to the subtropical parts of the country, indicating that the occurrence of this disease was unrelated to climate and soil conditions (figure 2a). This was further confirmed by the strong correlation between the number of patients and the population of each prefecture (ρ=0.815, p<0.001; figure 2b). The only exception was Niigata Prefecture, where we conducted an active serological screening survey that capitalised on personal relationships and communication between regional pulmonary physicians, 98% of whom had trained at Niigata University (group B).

Niigata Prefecture is located close to the center of Japan (figure 2a), has an annual population of 2.35 million (approximately 1.9% of the total population of Japan) and enroled 43 patients diagnosed with APAP into the study as group B (table 1). The mean annual number of patients and the incidence of APAP in group B were both significantly larger than those in other prefectures, with mean±sD values of 3.91±2.70 versus 1.35±0.45 and 0.0016±0.0012 versus 0.0005±0.0001, respectively (p=0.013 and p=0.013, as determined by Wilcoxon's signed-rank test; table 1). Thus the number of patients registered in group B was greater than the estimated average in group A.

The total number of patients in group A increased between 2006 and 2016 despite the slight decrease in the Japanese population. The slope for the mean number of patients, as determined using a linear mixed-effect model, was 0.123 (p<0.001, figure 3a). In contrast, the number of registered patients in group B fluctuated annually by between zero and nine patients per year, appearing mostly unchanged and the slope for the mean number of patients in Niigata Prefecture (as determined using the mixed-effect model) was −0.336 (p=0.207, figure 3b). The former slope was significantly larger than the latter slope (p=0.025). These data suggest that the ability to collect patient serum samples for group A screening improved during the 11-year survey period, but the collection and registration in group B seemed to be saturated during the same period.

The number of patients followed the Poisson distribution (p=0.987, as determined using a goodness-of-fit test) and the Poisson parameter was estimated to be 3.9 (95% CI 1.5–10.2) (figure 4). Using the bootstrap
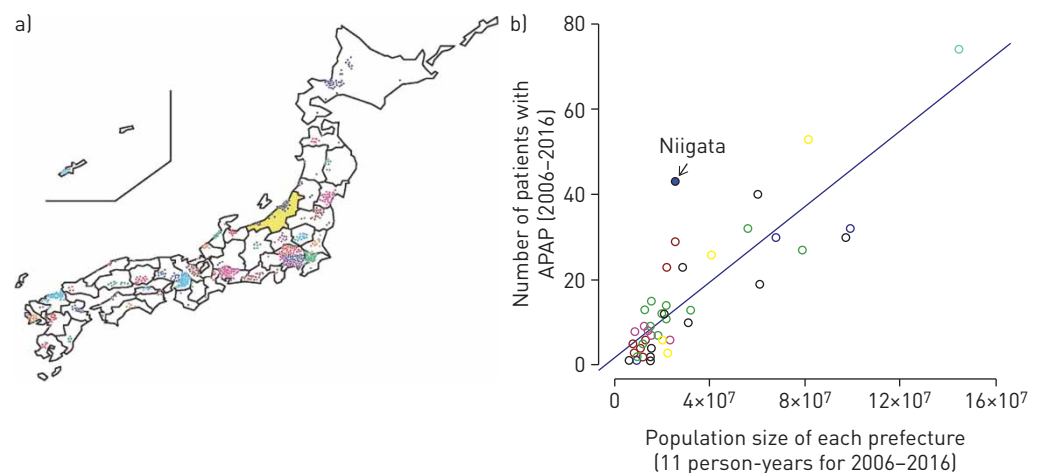
FIGURE 2 Distribution and regional prevalence of autoimmune pulmonary alveolar proteinosis (APAP) in Japan at the time of granulocyte–macrophage colony-stimulating factor autoantibody measurement. a) Geographic locations of the hospitals visited by the patients, for all of the 47 separate prefectures of Japan. Each circle represents one APAP patient and Niigata Prefecture is coloured yellow. b) Correlation of the prevalence of APAP cases with prefecture population size. The apparent incidence for Niigata Prefecture, in which intensive screenings for pulmonary alveolar proteinosis were performed annually, was higher than that for other prefectures. The blue dot indicates Niigata prefecture and the blue line indicates linear regression. The circular markers in the scatter plot are colour-coded by region.

TABLE 1 Newly registered patients[#] from across all of Japan and in Niigata Prefecture between 2006 and 2016

Distribution	Year										
	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
All Japan											
Patients[#]	37	43	50	56	53	39	84	75	83	83	99
Population ×10[6]	128	128	128	127	128	127	127	127	127	127	127
Incidence per 10[6] population	0.3	0.3	0.4	0.4	0.4	0.3	0.7	0.7	0.6	0.7	0.8
Niigata Prefecture											
Patients[#]	3	8	9	2	2	4	5	0	5	2	3
Population ×10[6]	2.4	2.4	2.4	2.4	2.4	2.4	2.3	2.3	2.3	2.3	2.3
Incidence per 10[6] population	1.2	3.3	3.8	0.8	0.8	1.7	2.1	0.0	2.2	0.9	1.3

[#]: newly registered patients were those whose serum was newly diagnosed as positive for granulocyte-macrophage colony-stimulating factor autoantibody.

method, the number of APAP patients in each year was verified to follow the Poisson distribution of parameter $\lambda=3.9$ (95% CI 2.5–5.5) (supplementary figure S1 and supplementary table S1). Dividing 3.9 by the mean population of group B (2.35 million) for 11 years (2006–2016) yielded an estimated incidence of 1.65 per million (95% CI 0.63–4.31). Multiplying this estimate by the mean population of Japan for these 11 years provided an estimated national annual incidence rate of 211.

To estimate the prevalence of APAP in Japan based on this incidence rate, we first had to determine the survival period. To do this we used the 103 cases from group C (Tohoku (n=38), Niigata (n=47) and Kyorin (n=18) University Hospitals) for which we had more detailed information and determined the survival of each patient retrospectively from information provided by their primary care doctors. The demographic data for the registered patients at the time of their GMAb measurements are shown in table 2 and table 3. The 2-year, 5-year and 11-year survival rates were 99.1%, 97.7% and 86.2%, respectively, and the RMST was 16.1 years (95% CI 15.1–17.1) as determined using the Kaplan–Meier method (figure 5a). By the bootstrap method, in the case that the truncation time point from diagnosis was set to the longest observed period (210 months), the RMST of the 103 APAP cases was verified to be approximately 16 years (95% CI 13–17) (supplementary figure S2 and supplementary table S2). From 1999 to 2017, seven patients died of APAP and one died from unknown causes. At the present time, a complete cure for this disease has yet to be defined.

If we exclude the possibility of a complete cure for APAP, the average disease duration should be equal to the average survival period. We therefore calculated APAP prevalence using the following equation: prevalence rate=incidence rate per time period×average disease duration (=0.00000165×16.1=0.0000266) [21–23]. This resulted in an estimate of nationwide disease prevalence of 26.6 per million (95% CI 9−73).
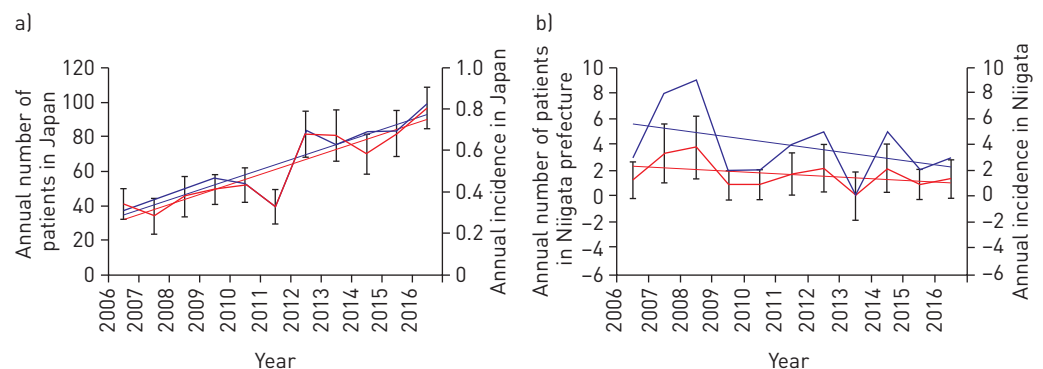

FIGURE 3 Distribution of patients over time. a) Annual number of patients (blue line) and annual incidence (red line) in Japan from 2006 to 2016. b) Annual number of patients (blue line) and annual incidence (red line) in Niigata prefecture from 2006 to 2016. The incidence of patients is defined as the number of new occurrences per million of population in the area. T-bars indicate the standard errors.
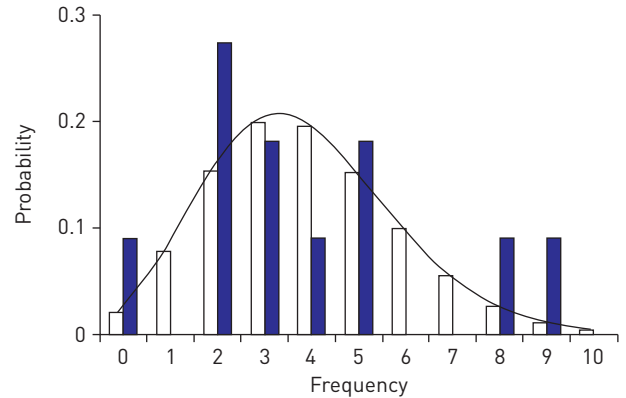
FIGURE 4 Frequency of autoimmune pulmonary alveolar proteinosis (APAP) cases. Blue bars indicate the actual number of new patients registered in group B for each year from 2006 to 2016. White bars and the theoretical curve indicate the number of new patients predicted by the Poisson distribution. The Poisson distribution thus reflects the probability of a given number of onsets per year and the distribution of the number of patients diagnosed with APAP (p=0.987; population mean and variance=3.9 (Kolmogorov–Smirnov test)).

With the recent stabilisation of the Japanese population at 125 million [24], we expect approximately 3300 current cases of APAP, with 211 new cases occurring annually (table 4). This survey has thus revealed that both incidence and prevalence are far higher than previously reported by both ourselves [6] and by others [25].

To determine the factors affecting patient survival, we used a log-rank test to compare the differences in cumulative rates between two or more groups. Using univariate analysis, the probability of death was

TABLE 2 Demographics of participants in the retrospective cohort study (n=103)	
Variable	**Distribution**
Age [#] years (n=98)	
Mean±SD	54.8±12.9
Median (range (IQR))	55.0 (67 (17))
Gender	
Male (n=65)	63±9 (54–72)
Female (n=36)	35±9 (26–44)
Smoking status	
Never smoker (n=34)	33±9 (24–42)
Ex-smoker (n= 39)	38±9 (29–47)
Current smoker (n=16)	16±7 (9–23)
Dust exposure [¶]	
Yes (n=35)	34±9 (25–43)
No (n=55)	53±10 (43–63)
Disease severity score [+]	
1 (n=23)	22±8 (14–30)
2 (n=25)	24±8 (16–32)
3 (n=16)	16±7 (8–23)
4 (n=12)	12±6 (6–18)
5 (n=8)	8±5 (3–13)
Symptoms	
Yes (n=62)	60±10 (50–70)
No (n=28)	27±9 (18–36)
Complications	
Yes (n=35)	34±9 (25–43)
No (n=68)	66±9 (57–75)

Data is presented as %±95% CI (lower–upper) unless otherwise stated. IQR: interquartile range. [#]: age at granulocyte-macrophage colony-stimulating factor autoantibody measurement; [¶]: dust exposure was based on self-reporting by patients/physicians; [+]: disease severity score was defined as described in the Methods section.

Variable	Patients	Measurements	
		Mean±SD	Median (range (IQR))
GMAb mg·mL^{-1}	101	66.2±141.0	34.5 (1194.8 (59.8))
P_{aO_2} mmHg	77	71.4±12.7	71.2 (62.5 (15.3))
P_{aCO_2} mmHg	77	38.5±4.2	38.0 (23.4 (5.3))
P_{A-aO_2} mmHg	80	29.8±15.7	30.1 (78.3 (22.9))
S_{pO_2} %	60	95.0±3.2	96.0 (19 (4))
VC L	70	3.1±0.8	3.2 (3.7 (1.3))
FVC cm^3	66	3.1±0.8	3.3 (3.6 (1.2))
FVC % predicted	64	93.4±19.6	95.8 (100.4 (20.7))
D_{LCO} % predicted	57	75.8±21.4	75.7 (85 (36.8))
D_{LCO}/V_A % predicted	52	78.8±29.8	84.2 (140.9 (26.5))
KL-6 units·L^{-1}	79	4915±5385	3181 (31 143 (5407))
LDH international units·L^{-1}	78	315.3±156.3	254.5 (771 (158.5))

TABLE 3 Laboratory findings for participants in the retrospective cohort study

IQR: interquartile range; GMAb: granulocyte-macrophage colony-stimulating factor autoantibody; P_{aO_2}: arterial oxygen tension; P_{aCO_2}: arterial carbon dioxide tension; P_{A-aO_2}: alveolar–arterial oxygen tension difference (alveolar–arterial oxygen gradient); S_{pO_2}: arterial oxygen saturation measured by pulse oximetry; VC: vital capacity; FVC: forced vital capacity; D_{LCO}: diffusing capacity of the lung for carbon monoxide; V_A: alveolar volume; KL-6: mucin-like glycoprotein; LDH: lactate dehydrogenase.

significantly associated with binary variables for age at diagnosis, arterial carbon dioxide tension (P_{aCO_2}), P_{A-aO_2}, arterial oxygen saturation measured by pulse oximetry (S_{pO_2}), % predicted diffusing capacity of the lung for carbon monoxide (D_{LCO}), KL-6 and LDH, dichotomised at the medians respectively. According to our forward stepwise multivariate analysis, only % predicted D_{LCO} was an independent effect. The probability of death due to APAP significantly increased when % predicted D_{LCO} was less than the median (=75.8%) at diagnosis (p=0.014, figure 5b).

Discussion

Based on the 702 registry data points for GMAb measurements conducted in our laboratory from 2006 to 2016, we initially estimated the incidence of APAP at 1.65 per million, based on the distribution data for patients in group A and the Poisson distribution applied to the annual number of patients in group B for the same time period. We then performed a retrospective cohort study in group C consisting of 103
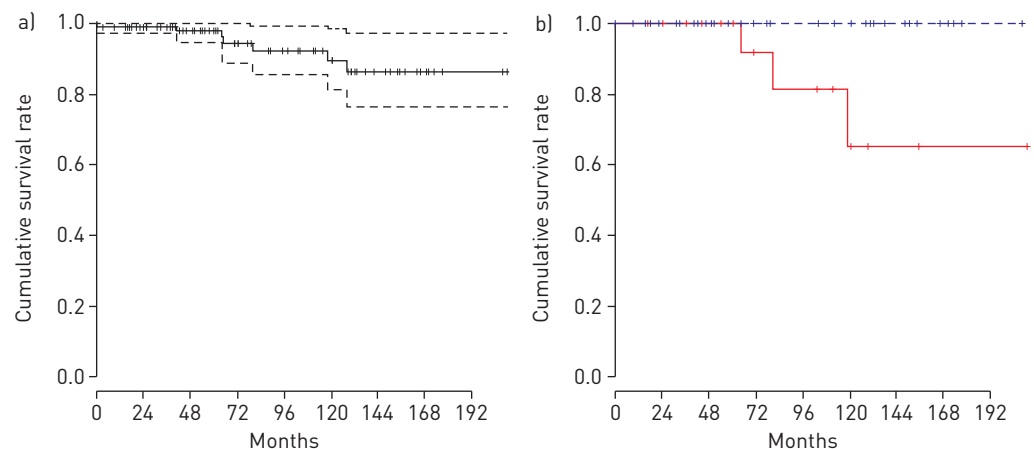

FIGURE 5 Kaplan–Meier survival curves showing survival rates over time (short perpendicular lines indicate the censored data). a) Kaplan–Meier survival curve showing the cumulative survival rate for the 103 cases from group C diagnosed at Tohoku (n=38), Niigata (n=47) and Kyorin (n=18) University Hospitals from 1999 to 2017 (the dotted lines indicate the 95% confidence interval). b) Kaplan–Meier survival curves comparing the cumulative survival rates between two groups separated according to median % predicted diffusing capacity of the lung for carbon monoxide (D_{LCO}) at diagnosis (the blue line indicates a % predicted D_{LCO} at diagnosis that was >75.8% while the red line indicates a % predicted D_{LCO} at diagnosis that was <75.8%).

TABLE 4 Summary of the latest epidemiology for autoimmune pulmonary alveolar proteinosis (APAP) in Japan

Indices	Estimated (95% CI)
Estimated annual onset in Niigata Prefecture	3.9 (1.5–10.2)[#]
Incidence per 10^6 population	1.65 (0.64–4.31)[¶]
Estimated annual cases in all of Japan	211 (81–549)[+]
Estimated survival period years	16.1 (15.1–17.1)[§]
Prevalence per 10^6 population	26.6 (9.6–73.6)[f]
Estimated number of patients in Japan	3387 (1226–9367)[##]

[#]: estimated using the Poisson distribution Z(0.975) as the 97.5% quantile of the standard normal distribution. The lower limit is calculated using the equation $((-1 \times Z(0.975) + (Z(0.975)^2 + 4 \times \text{event number})^{0.5})/2)^2$ and the upper limit using equation $((1 \times Z(0.975) + (Z(0.975)^2 + 4 \times \text{event number})^{0.5})/2)^2$; [¶]: calculated by dividing the Poisson estimate by the population of Niigata Prefecture; [+]: calculated by multiplying the incidence rate per million by the population of Japan; [§]: estimated using the restricted mean survival time; [f]: when the occurrence rate of the disease is stable, the formula prevalence=incidence×mean disease duration (=estimated survival period) is used; [##]: calculated by multiplying prevalence rate per million by the total population of Japan.

patients who were registered at three hospitals between 1999 and 2017 in order to estimate the average survival duration (16.1 years). Excluding the possibility of a complete cure for APAP, we arrived at approximately 3300 cases in Japan. Considering that our cumulative total for GMAb-positive cases since 1999 exceeded 1000 and that not all primary care doctors requested GMAb measurements (either because of the cost or various other reasons), we feel that the present epidemiological estimate is quite reasonable. Therefore, the previous estimates of incidence and prevalence (0.49 and 6.2 per million, respectively [6], are underestimates and should be revised.

We showed previously that there is a close correlation between the number of patients and the population size for each region [6]. The present study confirms the existence of this correlation at the prefecture level, but strongly suggests that there is no regional integration of the onset of APAP. Furthermore, the detection force was similar among prefectures, which also means that we can estimate the nationwide incidence by estimating the incidence in a given prefecture. The fact that the annual number of registrations for serum GMAb testing in group A increased through the study period suggests that the nationwide level of detection must have been improving, with the exception of group B, for the following reasons. First, the number in group A diagnosed serologically increased during 2006–2016, while the number in group B was stable (table 1 and figure 3). Secondly, the point incidence of APAP was much larger in group B than in group A (figure 2b). Thirdly, the Department of Pulmonary Medicine at Niigata University conducted an active survey of pulmonary alveolar proteinosis, taking advantage of personal relationships and communication among 200 regional pulmonary physicians, 98% of whom had received their training at Niigata University. We were therefore able to perform near-perfect screening of APAP in Niigata Prefecture. Together, these facts suggest that it makes sense to focus on the annual number of group B.

In addition, the discovery that the annual number of group B follows the Poisson distribution makes our estimate of the incidence of APAP more persuasive. This means that the probability of any given number of APAP onsets occurring in group B during one year was constant and independent of the amount of time since the last onset. The high probability that the goodness-of-fit test indicated for the Poisson distribution (p=0.987) indicated that using this distribution to estimate APAP incidence was appropriate. The combined conditions of APAP being a rare condition and the population being relatively large (2.35 million) met the requirements for applying the Poisson distribution. A number of other rare-disease studies have also used this distribution to assess incidence and prevalence. For example, GUIDETTI *et al.* [26] used it to report the incidence and prevalence rates of *myasthenia gravis* (1980–1994, n=49). BAMBHA *et al.* [27] reported the incidence and prevalence of primary sclerosing cholangitis in a United States community (1976–2000, n=22), assuming that the number of cases followed a Poisson distribution. However, the usefulness and reliability of fitting the probability of disease onset to the Poisson distribution has not previously been evaluated.

The previous national registry study estimated disease prevalence using an interval estimation method (*i.e.* the total number of patients registered during 1999–2006 was divided by the mean total Japanese population for that period). McCARTHY *et al.* [28] recently used the interval method to estimate the prevalence of pulmonary alveolar proteinosis to be 6.87±0.33 per million for the general population of the United States. They utilised a large health-insurance claims database containing comprehensive data for

approximately 15 million patients in the US over a 15-year period. With the interval estimation method, an insufficient detection force combined with the disease duration exceeding the observation period would result in an underestimated prevalence. In both our previous study and the US study, the detection force must have been insufficient because the GMAb measurement was not generally known by pulmonary doctors in the former and was not performed at all in the latter.

In group C, we assumed that APAP cannot be completely cured (*i.e.* the mean disease duration is equal to the mean survival time). Despite this, some patients did appear to be cured based on chest computed tomography (CT) scan, arterial blood gas analysis, or serum markers. However, the definition of "cure" has not gained consensus among pulmonary physicians. In addition, the observations of several patients were censored during the present retrospective cohort study. However, the RMST method enabled us to estimate the mean survival period of all 103 patients despite this situation [16–20]. Our prevalence estimate of 22.6 per million may therefore have to be revised in the future, when the disease duration can be estimated more accurately.

In the present retrospective cohort study of group C, we recorded seven deaths due to exacerbation of APAP. A univariate analysis applied to the Kaplan–Meier survival curve showed that the probability of death was significantly associated with age at diagnosis, P_{aCO_2}, P_{A-aO_2}, S_{pO_2}, KL-6, LDH and % predicted D_{LCO}. However, a forward stepwise multivariate analysis revealed that only % predicted D_{LCO} affected prognosis independently. Surprisingly, all seven patients who died had a % predicted D_{LCO} of <75.8% at GMAb testing. In future, to minimise the censoring of patient observation, a prospective cohort study is essential for the accurate prognosis of APAP and identification of the risk factors for exacerbation.

In conclusion, based on this study we have revised our own epidemiological knowledge of APAP in Japan. In this regard, we applied a more robust methodology to our GMAb test database. We believe that the updated incidence and prevalence (1.65 and 22.6 per million, respectively) will contribute to patient care, drug development and the administration of rare diseases.

References

1 Rosen SH, Castleman B, Liebow AA. Pulmonary alveolar proteinosis. *N Engl J Med* 1958; 258: 1123e42.
2 Trapnell BC, Whitsett JA, Nakata K. Pulmonary alveolar proteinosis. *N Engl J Med* 2003; 349: 2527–2539.
3 Ishii H, Tazawa R, Kaneko C, *et al.* Clinical features of secondary pulmonary alveolar proteinosis: pre-mortem cases in Japan. *Eur Respir J* 2011; 37: 465–468.
4 Kitamura T, Tanaka N, Watanabe J, *et al.* Idiopathic pulmonary alveolar proteinosis as an autoimmune disease with neutralizing antibody against granulocyte/macrophage colony-stimulating factor. *J Exp Med* 1999; 190: 875–880.
5 Seymour JF, Presneill JJ. Pulmonary alveolar proteinosis: progress in the first 44 years. *Am J Respir Crit Care Med* 2002; 166: 215–235.
6 Inoue Y, Trapnell BC, Tazawa R, *et al.* Characteristics of a large cohort of autoimmune pulmonary alveolar proteinosis patients in Japan. *Am J Respir Crit Care Med* 2008; 177: 752–762.
7 Kahn HA, Sempos CT. Statistical methods in epidemiology. 1st Edn. New York, Oxford University Press USA, 1989.
8 Bailar JC, Ederer F. Significance factors for ratio of a Poisson variable to its expectation. *Biometrics* 1964; 20: 639–643.
9 Hakulinen T, Dyba T. Precision of incidence predictions based on Poisson distributed observations. *Stat Med* 1994; 13: 1513–1523.
10 Flanders WD, Kleinbaum DG. Basic models for disease occurrence in epidemiology. *Int J Epidemiol* 1995; 24: 1–7.
11 Kitamura T, Uchida K, Tanaka N, *et al.* Serological diagnosis of idiopathic pulmonary alveolar proteinosis. *Am J Respir Crit Care Med* 2000; 162: 658–662.
12 Uchida K, Nakata K, Carey B, *et al.* Standardized serum GM-CSF autoantibody testing for the routine clinical diagnosis of autoimmune pulmonary alveolar proteinosis. *J Immunol Methods* 2014; 402: 57–70.
13 Nei T, Urano S, Motoi N, *et al.* IgM-type GM-CSF autoantibody is etiologically a bystander but associated with IgG-type autoantibody production in autoimmune pulmonary alveolar proteinosis. *Am J Physiol Lung Cell Mol Physiol* 2012; 302: L959–L964.
14 Inoue Y, Barker E, Daniloff E, *et al.* Pulmonary epithelial cell injury and alveolar-capillary permeability in berylliosis. *Am J Respir Crit Care Med* 1997; 156: 109–115.
15 Kohno N. Serum marker kl-6/muc1 for the diagnosis and management of interstitial pneumonitis. *J Med Invest* 1999; 46: 151–158.
16 Royston P, Parmar MKB. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol* 2013; 13: 152.
17 Zhao L, Claggett B, Tian L, *et al.* On the restricted mean survival time curve in survival analysis. *Biometrics* 2016; 72: 215–221.

18   Calkins KL, Canan CE, Moore RD, *et al.* An application of restricted mean survival time in a competing risks setting: comparing time to ART initiation by injection drug use. *BMC Med Res Methodol* 2018; 18: 27.

19   Huang B, Kuan PF. Comparison of the restricted mean survival time with the hazard ratio in superiority trials with a time-to-event end point. *Pharm Stat* 2018; 17: 202–213.

20   Liang F, Zhang S, Wang Q, *et al.* Treatment effects measured by restricted mean survival time in trials of immune checkpoint inhibitors for cancer. *Ann Oncol* 2018; 29: 1320–1324.

21   Freeman J, Hutchison GB. Prevalence, incidence and duration. *Am J Epidemiol* 1980; 112: 707–723.

22   Fletcher RH, Fletcher SW, Wagner EH. Clinical epidemiology: the essentials. 3rd Edn. Baltimore, Williams & Wilkins, 1996; pp. 84–85.

23   Deenen JC, Arnts H, van der Maarel SM, *et al.* Population-based incidence and prevalence of facioscapulohumeral dystrophy. *Neurology* 2014; 83: 1056–1059.

24   Statistics Japan. Monthly report: population estimates by age (five-year groups) and sex. Tokyo, Statistics Bureau, Ministry of Internal Affairs and Communications, 2018. www.stat.go.jp/english/data/jinsui/tsuki/index.htm

25   Ben-Dov I, Kishinevski Y, Roznman J, *et al.* Pulmonary alveolar proteinosis in Israel: ethnic clustering. *Isr Med Assoc J* 1999; 1: 75–78.

26   Guidetti D, Sabadini R, Bondavalli M, *et al.* Epidemiological study of myasthenia gravis in the province of Reggio Emilia, Italy. *Eur J Epidemiol* 1998; 14: 381–387.

27   Bambha K, Kim WR, Talwalkar J, *et al.* Incidence, clinical spectrum, and outcomes of primary sclerosing cholangitis in a United States community. *Gastroenterology* 2003; 125: 1364–1369.

28   McCarthy C, Avetisyan R, Carey BC, *et al.* Prevalence and healthcare burden of pulmonary alveolar proteinosis. *Orphanet J Rare Dis* 2018; 13: 129.