Estimating the incidence and diagnosed proportion of HIV infections in Japan: a statistical modeling study

Hiroshi Nishiura Corresp. [1, 2]

[1] Graduate School of Medicine, Hokkaido University, Sapporo, Japan
[2] CREST, Japan Science and Technology Agency, Saitama, Japan

Corresponding Author: Hiroshi Nishiura
Email address: nishiurah@gmail.com

Background. Epidemiological surveillance of HIV infection in Japan involves two technical problems for directly applying a classical backcalculation method, i.e., (i) all AIDS cases are not counted over time and (ii) people diagnosed with HIV have received antiretroviral therapy, extending the incubation period. The present study aimed to address these issues and estimate the HIV incidence and the proportion of diagnosed HIV infections, using a simple statistical model.

Methods. From among Japanese nationals, yearly incidence data of HIV diagnoses and patients with AIDS who had not previously been diagnosed as HIV positive, from 1985 to 2017, were analyzed. Using the McKendrick partial differential equation, general convolution-like equations were derived, allowing estimation of the HIV incidence and the time-dependent rate of diagnosis. A likelihood-based approach was used to obtain parameter estimates.

Results. Assuming that the median incubation period was 10.0 years, the cumulative number of HIV infections was estimated to be 29,613 (95% confidence interval (CI): 29,059, 30,167) by the end of 2017, and the proportion of diagnosed HIV infections was estimated at 80.3% (95% CI: 78.7%, 82.0%). Allowing the median incubation period to range from 7.5 to 12.3 years, the estimate of the proportion diagnosed can vary from 77% to 84%.

Discussion. The proportion of diagnosed HIV infections appears to have not yet reached 90% among Japanese nationals. Compared with the peak incidence from 2005–2008, new HIV infections have clearly been in a declining trend; however, there are still more than 1,000 new HIV infections per year in Japan. To increase the diagnosed proportion of HIV infections, it is critical to identify people who have difficulty accessing consultation, testing, and care, and to explore heterogeneous patterns of infection.

Estimating the incidence and diagnosed proportion of HIV infections in Japan: a statistical modeling study

Hiroshi Nishiura[1,2]

[1] Graduate School of Medicine, Hokkaido University, Sapporo, Japan

[2] CREST, Japan Science and Technology Agency, Saitama, Japan

Corresponding Author:

Hiroshi Nishiura

Graduate School of Medicine, Hokkaido University, Sapporo, Japan

Kita 15 Jo Nishi 7 Chome, Kita-ku, Sapporo-shi, Hokkaido 060-8638, Japan

Email address: nishiurah@med.hokudai.ac.jp

Abstract

Background. Epidemiological surveillance of HIV infection in Japan involves two technical problems for directly applying a classical backcalculation method, i.e., (i) all AIDS cases are not counted over time and (ii) people diagnosed with HIV have received antiretroviral therapy, extending the incubation period. The present study aimed to address these issues and estimate the HIV incidence and the proportion of diagnosed HIV infections, using a simple statistical model.

Methods. From among Japanese nationals, yearly incidence data of HIV diagnoses and patients with AIDS who had not previously been diagnosed as HIV positive, from 1985 to 2017, were analyzed. Using the McKendrick partial differential equation, general convolution-like equations were derived, allowing estimation of the HIV incidence and the time-dependent rate of diagnosis. A likelihood-based approach was used to obtain parameter estimates.

Results. Assuming that the median incubation period was 10.0 years, the cumulative number of HIV infections was estimated to be 29,613 (95% confidence interval (CI): 29,059, 30,167) by the end of 2017, and the proportion of diagnosed HIV infections was estimated at 80.3% (95% CI: 78.7**%**, 82.0**%**). Allowing the median incubation period to range from 7.5 to 12.3 years, the estimate of the proportion diagnosed can vary from 77% to 84%.

Discussion. The proportion of diagnosed HIV infections appears to have not yet reached 90% among Japanese nationals. Compared with the peak incidence from 2005–2008, new HIV infections have clearly been in a declining trend; however, there are still more than 1,000 new HIV infections per year in Japan. To increase the diagnosed proportion of HIV infections, it is critical to identify people who have difficulty accessing consultation, testing, and care, and to explore heterogeneous patterns of infection.

Introduction

Following an infection with human immunodeficiency virus (HIV), development of acquired immunodeficiency syndrome (AIDS) takes about 10 years (Munoz et al., 1997). The long incubation period makes it difficult to directly observe the incidence and prevalence of HIV infections over time. To offer insights into the epidemiology of HIV-infected and -incubating individuals over time, and to evaluate public health control programs, various statistical modeling approaches have been proposed to date (Brookmeyer & Gail, 1994; Donnelly & Cox, 2001; Jewell et al., 1992). Of these, a backcalculation method using a simple integral equation to model AIDS incidence as arising from the HIV incidence convoluted with the independently and identically distributed incubation period allows estimation of the HIV incidence based on epidemiological surveillance data (Brookmeyer & Gail, 1986; Gail & Brookmeyer, 1988). Assuming that the reported number of AIDS cases certainly and accurately captures the actual number of AIDS incidence in industrialized countries, the backcalculation method greatly improves our understanding of the epidemiology of HIV infection, attributing the observed AIDS data to HIV infection events as a function of time.

Understanding the transmission dynamics of HIV using such statistical models is in line with the concept of treatment cascade, introduced by the Joint United Nations Programme on HIV/AIDS (UNAIDS). The so-called care cascade aims to identify and fill gaps in the continuum of services for testing, care, and effective treatment of HIV (UNAIDS, 2014). In relation to this, the UNAIDS report has led to the global initiative “90–90-90” by 2020 that sets out goals in care cascades to achieve the following: 90% of people living with HIV know their HIV status, 90% of people diagnosed with HIV have access to antiretroviral therapy (ART), and

90% of people receiving ART have suppressed viral loads (UNAIDS, 2014). UNAIDS even aims to achieve 95-95-95 at a global level by the year 2030, contributing to successfully controlling HIV and AIDS, as supported by the so-called test-and-treat strategy (Granich et al., 2009; Granich et al., 2017). To quantify the situation of each country, monitoring diagnosed individuals is essential; moreover, estimation of the diagnosed proportion of HIV infections must be supported by firm scientific methods, to estimate the first part of the three 90-90-90 targets, i.e., 90% of HIV-infected people know their HIV status. In Japan, an analysis of blood donors took place in 2017, in which it was estimated that 85.6% of HIV-infected individuals, regardless of nationality, were diagnosed (Iwamoto et al., 2017). Nevertheless, it is known that the analysis of voluntary blood donation data is prone to sampling bias of donors owing to the tendency of people with high risk to repeatedly undertake anonymous laboratory testing through the practice of blood donation (Kihara et al., 2000), even though such screening of a large number of people is very costly. Considering the need to achieve continued monitoring of the diagnosed proportion of HIV-infected individuals, development of a reasonable yet scientifically rigorous method based on other datasets would be beneficial, especially using epidemiological surveillance data (Hsieh et al., 2012; Cuadros & Abu-Raddad, 2016; Hsieh & Lin, 2016; Mumtaz et al., 2018).

Despite the clear need for epidemiological estimation of the number of undiagnosed HIV infections, the surveillance data in Japan possesses two technical problems. First, while the definition of AIDS has remained nearly unchanged over time, reporting AIDS cases that were previously diagnosed as HIV-infected cases has never been mandated (Nishiura, 2007). This makes it impossible to directly apply the simplest convolution equation to the data because the backcalculation method requires the count of all AIDS cases over time. Surveillance in Japan has only consistently counted (i) HIV infections without AIDS at the time of diagnosis and (ii) AIDS

cases without previous diagnosis of HIV infection. Second, ART has been widespread since 1997 and has continuously improved the prognosis of HIV infection. Explicit incorporation of treatment requires us to account for not only the treatment coverage but also the treatment details (e.g., details of combination therapy), adherence, and many other factors. A simple yet tractable estimation method that can overcome these problems is called for.

In the present study, the aim was to address the abovementioned issues, estimating the HIV incidence among Japanese nationals, and also to offer statistical estimates of undiagnosed HIV infections and the proportion of diagnosed HIV infections over time.

Materials & Methods

Surveillance data of HIV and AIDS in Japan

The present study investigated the epidemiological surveillance data of HIV and AIDS in Japan, which is publicly reported by the Committee of AIDS Trends (2018), belonging to the Ministry of Health, Labor and Welfare, Japan. Of the reported datasets, our analyses are focused on Japanese nationals because estimation of infection among foreigners requires accounting for human migration, and the decision of migratory behavior (e.g. leaving Japan) is highly dependent on the diagnosis of HIV infection and AIDS. As of the end of 2017, there were 16,663 HIV infections and 7,587 AIDS cases among Japanese nationals (Committee of AIDS Trends, 2018). As mentioned, HIV diagnoses reflect HIV-infected individuals who undertook voluntary diagnostic testing before the onset of AIDS. An AIDS case indicates a patient who has never been diagnosed with HIV infection prior to an AIDS diagnosis and who meets the clinical diagnostic criteria: (i) confirmed HIV infection and (ii) the presence of one of 23 indicator diseases representing opportunistic infections or tumors. According to the Infectious Disease Law, HIV and AIDS are classified as a category V notifiable disease, and once diagnosed,

106 physicians must notify the case within 7 days of diagnosis. In the present study, the yearly
107 incidence of HIV infections and AIDS diagnoses from 1985 to 2017 was used. The data are
108 structured by sex and also by the most likely route of transmission (e.g., heterosexual,
109 homosexual or intravenous drug use, based on a physician's interview of patients). The latter
110 information, i.e., the mode of transmission, is discarded because it is believed that a substantial
111 proportion of men having sex with men do not disclose the actual contact and inform physicians
112 that they acquired infection through heterosexual contact (Inoue et al., 2015). Thus, a stratified
113 estimation by sex was conducted. Although the magnitude of the epidemic in Japan is relatively
114 small compared with that in Western industrialized countries, the incidence of HIV infection in
115 Japan is believed to have steadily increased over time, especially among men who have sex with
116 men and young adults (Kihara et al., 2003; Nemoto, 2004).

117 **Derivation of likelihood using a mathematical model**

118 The proposed statistical model is derived from the following partial differential equation
119 (PDE) model, which is referred to as the McKendrick equation (Nishiura & Inaba, 2011; Ejima
120 et al., 2014). Figure 1 shows a compartmental diagram of the data-generating process. Once
121 infected with HIV, individuals who are either undiagnosed or in the incubation period experience
122 two different hazards, i.e., the force of HIV diagnosis $\alpha(t)$ that depends on calendar time t and the
123 hazard of illness onset $\rho(\tau)$ that depends on the time elapsed since infection τ. Let $h(t,\tau)$ be
124 undiagnosed incubating HIV infections at calendar time t and the time since infection τ, the
125 dynamics of HIV diagnosis and illness onset are described by

126 $$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial s}\right)h(t,s) = -(\alpha(t)+\rho(s))h(t,s), \quad (1)$$

127 with a boundary condition

$$\lambda(t) := h(t,0), \qquad (2)$$

where $\lambda(t)$ represents the HIV incidence (i.e., the number of new HIV infections) at calendar time t. It should be noted that $\rho(\tau)$ constitutes the probability density function of the incubation period as follows:

$$f(s) = \rho(s)exp\left(-\int_0^s \rho(y)dy\right), \quad (3)$$

for $s>0$. It is well known that the McKendrick equation can be solved along the characteristic line, i.e.,

$$h(t,s) = \lambda(t-s)exp\left(-\int_{t-s}^{t} \alpha(x)dx - \int_0^x \rho(y)dy\right), \quad (4)$$

for $t-s>0$. Equations (1) and (4) indicate that the incidence of HIV diagnosis at calendar time t, $u(t)$, is written as

$$u(t) = \int_0^t \alpha(t)\lambda(t-s)exp\left(-\int_{t-s}^{t} \alpha(x)dx - \int_0^x \rho(y)dy\right)ds, \quad (5)$$

and similarly, the incidence of AIDS cases at time t, $a(t)$, is

$$a(t) = \int_0^t \rho(s)\lambda(t-s)exp\left(-\int_{t-s}^{t} \alpha(x)dx - \int_0^x \rho(y)dy\right)ds. \quad (6)$$

Equations (5) and (6) read similarly to the so-called extended backcalculation (Hall et al., 2008), which is derived from a competing risk model (Marschner, 1994; Cui & Becker, 2000). The abovementioned process can be used as the generalization.

Statistical model and estimation

The datasets are reported in a discrete time interval (i.e., year); thus, here I discretized models (5) and (6) as

147 $$u_t = \sum_{s=1}^{t} \lambda_{t-s}\alpha_t \prod_{x=t-s-1}^{t-1}(1-\alpha_x)\prod_{y=1}^{s-1}(1-\rho_y), \quad (7)$$

148 and

149 $$a_t = \sum_{s=1}^{t} \lambda_{t-s}\rho_s \prod_{x=t-s-1}^{t-1}(1-\alpha_x)\prod_{y=1}^{s-1}(1-\rho_y). \quad (8)$$

150 The incidence of HIV infection in year t, λ_t, is modeled as a step function:

151 $$\lambda_t = \begin{cases} \lambda_1 & \text{for } t < 1989, \\ \lambda_2 & \text{for } 1989 \le t < 1993, \\ \vdots & \\ \lambda_9 & \text{for } 2013 \le t, \end{cases} \quad (9)$$

152 such that the yearly incidence can be directly dealt with as the parameter. The yearly probability

153 of diagnosis in year t, α_t, is similarly modeled as

154 $$\alpha_t = \begin{cases} \alpha_1 & \text{for } t < 1989, \\ \alpha_2 & \text{for } 1989 \le t < 1993, \\ \vdots & \\ \alpha_9 & \text{for } 2013 \le t. \end{cases} \quad (10)$$

155 The probability mass function of the incubation period is assumed as known, and in discrete

156 time, this is written as $\rho_s\prod_{y=1}^{s-1}(1-\rho_y)$. As is widely assumed for HIV infection, the incubation

157 period is modeled using the Weibull distribution. Using the property of Weibull distribution with

158 the scale parameter η and shape parameter k, the discrete Weibull model is connected to the

159 continuous version as

160 $$\rho_s = 1 - \frac{\exp\left(-\left(\frac{t+1}{\eta}\right)^k\right)}{\exp\left(-\left(\frac{t}{\eta}\right)^k\right)}, \quad (11)$$

161 and

$$\prod_{y=1}^{t-1}(1-\rho_y) = \exp\left(-\left(\frac{t}{\eta}\right)^k\right). \quad (12)$$

Using the abovementioned model, undiagnosed HIV infections at the end of year t are computed as

$$x_t = \sum_{s=1}^{t} \lambda_{t-s} \prod_{x=t-s-1}^{t-1}(1-\alpha_x) \prod_{y=1}^{s-1}(1-\rho_y). \quad (13)$$

The diagnosed proportion of HIV infections is calculated either as $\Sigma(a+u)/\Sigma(x+a+u)$ or $\Sigma u/\Sigma(x+u)$, taking the summations over time. The former calculates the proportion of diagnosed HIV-positive individuals out of the cumulative number of HIV-positive individuals. This calculation has the drawback of including patients with AIDS who have already died by the year of calculation. As of 2017, it has been reported that a total of 2,321 cases resulted in death (Iwamoto et al., 2017). Alternatively, the latter calculates the fraction of individuals who are HIV positive but have not yet developed AIDS out of the cumulative number of HIV-positive individuals but including undiagnosed individuals, considering that the incubation period in most cases of HIV infection is now considerably extended by ART. The drawback of the latter calculation is that patients with AIDS who have survived and have received ART are excluded; thus, the calculated proportion may not be strictly in line with the target figure in the first goal of the 90-90-90 initiative. Therefore, when estimating the undiagnosed number of HIV infections and the diagnosed proportion at the end of 2017, both calculations are made, and the former is adjusted by subtracting 2,321 AIDS deaths from the cumulative count of AIDS cases.

To quantify the proposed system of equations, we estimate parameters λ_t and α_t by means of the maximum likelihood method. Considering that HIV infections are generated as the nonhomogenous Poisson process, the resulting HIV diagnoses and AIDS cases would also follow Poisson distributions. The likelihood function of HIV diagnoses is

$$L_1 = constant \times \prod_{t=1985}^{2017} E(u_t)^{r_t} exp\,(-E(u_t)), \quad (14)$$

where r_t denotes the reported (observed) number of HIV diagnoses in year t in the surveillance record. Similarly, the likelihood of new AIDS diagnoses is

$$L_2 = constant \times \prod_{t=1985}^{2017} E(a_t)^{w_t} exp\,(-E(a_t)), \quad (15)$$

where w_t denotes the reported number of new AIDS diagnoses in year t. Consequently, the total likelihood L is given by

$$L = L_1 L_2. \quad (16)$$

Maximum likelihood estimates of parameters are obtained by minimizing the negative logarithm of Eq. (16). As mentioned above, the incubation period distribution is assumed as known, and to address the uncertainty, three different estimates are derived from published studies (Boldson et al., 1988; Brookmeyer & Goedert, 1989; Munoz & Xu, 1996). A widely cited estimate by Brookmeyer & Goedert (1989) was derived from the study of patients with hemophilia over 20 years of age with η=11.6 and k=2.5, resulting in a median incubation period of 10.0 years. Boldson et al. (1988) investigated a cohort of AIDS cases in San Francisco with η=14.3 and k=2.5, yielding a median incubation period of 12.3 years. The estimate by Munoz and Xu (1996) was obtained from the Multicenter AIDS Cohort Study with η=10.0 and k=1.3, and the median incubation period is 7.5 years. All three estimates have been used in the present study to address uncertainty with respect to the incubation period.

The 95% confidence interval (CI) of parameters was derived from the profile likelihood. The 95% CI of model estimates (e.g., the number of undiagnosed HIV infections and the proportion diagnosed) was derived using a parametric bootstrap method. In the bootstrapping exercise, model parameters were resampled from a multivariate normal distribution with vectors

of mean **θ** and standard deviation **σ**. The latter vector was derived from the covariance matrix, taking diagonal elements of the inverse Hessian matrix ($\sigma^2 = \mathrm{diag}(H^{-1}(\theta))$). For each set of parameters, the model solution is obtained, and 1,000 times of parameter resampling results in a simulated distribution of model solutions. By taking the 2.5th and 97.5th percentile points of the simulated distribution, the 95% CI is obtained. All statistical data were analyzed using R version 3.1 (Comprehensive R Archive Network) (R Core Team, 2016) and JMP version 12.0.1 statistical software (SAS Institute Inc., Cary, NC, USA).

Ethical considerations

In the present study, the analyzed data are publicly available (Committee of AIDS Trends, 2018). As such, the datasets used in our study are deidentified and fully anonymized in advance, and the analysis of publicly available data with no identifying information does not require ethical approval.

Results

Estimated parameters, i.e., yearly incidence and yearly probability of diagnosis, are shown in Fig. 2. With the assumed median incubation period of 10.0 years, the yearly incidence was the highest from 2005–2008, with an estimated 1,972 (95% CI: 1,829, 2,115) infections per year (Fig. 2A). Subsequently, the incidence began to decline; the yearly estimate in the most recent interval (from 2013–2017) was 1,179 (95% CI: 1,047, 1,293) infections. The yearly probability of diagnosis was monotonously improved over time (Fig. 2B). The estimated diagnosis probability by 1999 was less than 10%, but the latest estimate from 2013–2017 was 15.6% (95% CI: 14.8%, 16.4%). The qualitative patterns of HIV incidence and diagnosis did not vary greatly, even when shorter and longer median incubation periods were used (Figs. 2C and

2D). Figures 2E and 2F show maximum likelihood estimates of the incidence and probability of diagnosis by sex. The incidence in males was the highest from 2005–2008; the latest estimate from 2013–2017 ranged from 1,015 to 1,363 infections per year, assuming a median incubation period from 7.5 to 12.3 years. Similarly, the incidence in females was highest from 1993–1996, ranging from 86 to 97 infections per year; the latest yearly incidence ranged from 31 to 54 infections with a median incubation period of 7.5 to 12.3 years. The yearly probability of diagnosis among males behaved similarly to that of the entire population, but there was no apparent improvement in the frequency of diagnosis among females.

Figure 3 shows a comparison between the observed and predicted number of HIV diagnoses and AIDS cases. All three models with different median incubation periods yielded almost identically good fit to the data (Fig. 3A). Even though the number of diagnosed HIV infections and AIDS cases was relatively small for females, the proposed model successfully captured the observed patterns of HIV diagnoses and AIDS cases by sex (Fig. 3B).

Figure 4 shows the estimated undiagnosed number of HIV infections and the estimated proportion of diagnosed HIV-positive individuals over time, among Japanese nationals. Using the median incubation period of 10.0 years (Fig. 4A), undiagnosed HIV infection was estimated to have peaked in 2009 with 7,532 (95% CI: 6,911, 8,152) infections. In the latest time interval, from 2013–2017, it was estimated that 5,363 (95% CI: 4,809, 5,917) infections remained unrecognized. Varying the median incubation period from 7.5 to 12.3 years, the maximum likelihood estimate of undiagnosed HIV infections in the latest time interval ranged from 4,041 to 6,552 infections. These findings indicate that the cumulative number of HIV infections by the end of 2017 was 29,613 (95% CI: 29,059, 30,167) Japanese nationals, using the median incubation period of 10.0 years, and can range from 28,291 to 30,802 individuals.

Including and excluding AIDS cases, the estimated proportions of diagnosed HIV infections are shown in Figs. 4C and 4D. Including AIDS cases, the diagnosed proportion was estimated at 81.9% (range 78.7% to 85.7%) using the median incubation period of 10.0 (7.5 to 12.3) years. Excluding AIDS cases, the estimate was 75.7% (range 71.8% to 80.5%). Figures 4E and 4F show the estimated number of undiagnosed HIV infections and the diagnosed proportion by sex, excluding AIDS cases. Estimates of undiagnosed HIV infections among males behaved similarly to the entire population of Japanese nationals, whereas those of females peaked in the year 2001. In the latest time interval (2013–2017), it was estimated that 5,150 infections (range 3,881 to 6,287) in males and 210 infections (range 162 to 255) in females remained unrecognized, using the median incubation period of 10.0 (with the range of 7.5 to 12.3) years. The diagnosed proportion of both males and females increased with time, and females tended to yield higher estimates than males. In the latest time interval from 2013–2017, the diagnosed proportion (excluding AIDS cases) was estimated at 75.3% (range 71.4% to 80.2%) among males and 82.1% (range 79.1% to 85.6%) among females.

Figure 4 shows the undiagnosed number of HIV infections and the proportion of diagnosed infections at the end of 2017. The uncertainty bound was greatest with an assumed median incubation period of 12.3 years, with an estimated 6,552 infections (95% CI: 5,632, 7,471). Figure 4B shows the diagnosed proportion, including and excluding AIDS cases, with 95% confidence intervals. Even when AIDS cases were included, the 2,321 deaths known up to that point were subtracted from AIDS cases in advance of the calculation. Assuming that the median incubation period was 10.0 years, the calculation, inclusive of surviving AIDS cases, yielded 80.3% (95% CI: 78.7%, 82.0%); when excluding AIDS cases, the proportion was 75.7% (95% CI: 73.8%, 77.6%).

Discussion

The present study estimated the incidence and diagnosed fraction of HIV infections among Japanese nationals. By the end of 2017, the cumulative number of HIV infections was estimated to be about 30,000 cases, of which 4,000 to 6,000 were considered to have remained undiagnosed. Assuming that the median incubation period was 10.0 years, 80% of infections have ever been diagnosed; accounting for the uncertainty in a median incubation period ranging from 7.5 to 12.3 years, the estimate of the diagnosed proportion can range from 77% to 84%. To the author's knowledge, the present study is the first to offer firm statistical estimates of the incidence and diagnosed proportion of HIV infections based on epidemiological surveillance data in Japan, using an explicit mathematical modeling approach.

There are two take-home messages from the results of this study. First, regardless of whether AIDS cases are included, the proportion of diagnosed HIV infections appears not to have reached 90% among Japanese nationals. Although some estimates exceed 80%, even after subtraction of known deaths owing to AIDS, the findings echo those of a published study that analyzed blood donor data (Iwamoto et al., 2017), which represent a critical problem in Japan for controlling HIV and AIDS. In the present study, the rate of diagnosis was shown to have improved with time, and the trend was particularly apparent among men, mainly comprising men having sex with men. The findings of the present study indicate that there would be a certain number of infected individuals who may not have proper access to consultation, testing, and care with privacy protection. To identify the attributes of such HIV-infected individuals in greater detail, the investigation must be extended to explore heterogeneous patterns, including age-dependence, spatial heterogeneity, and other background characteristics. These are my ongoing research interests.

297 Second, compared with the peak from 2005–2008, the incidence showed a declining
298 trend. Compared with the estimate in 2005–2008, the upper 95% CIs of the next two time
299 periods (2009–2012 and 2013–2017) were significantly lower than those in the peak period. In
300 fact, a declining trend has also been seen in other datasets, including the incidence of counseling
301 and blood testing at local health centers and the proportion of HIV-positive blood donors over
302 time (Committee of AIDS Trends, 2018). The present study results support that these observed
303 declines are partially attributable to actual decreases in the incidence of HIV infection in Japan.
304 The underlying mechanisms of such decreases have yet to be explored using a mathematical
305 model, perhaps requiring modeling of the saturated effect (Heesterbeek & Metz, 1993) together
306 with statistical estimates of the effective reproduction number (Kretzschmar et al., 2013). In
307 addition, it must be remembered that the yearly incidence still remains above 1,000 infections;
308 moreover, such a declining trend is not evident among females.
309 Although the present study was motivated by the need for quantifying the care cascade in
310 Japan, in accordance with the goals of 90-90-90, a few technical issues must be noted to interpret
311 the estimates and apply the present results to the evaluation. First, Japanese estimates of the latter
312 two goals of the 90-90-90 initiative, i.e., access to ART and virus suppression, rest on
313 questionnaire surveys conducted in the prefectures, which do not distinguish between infected
314 individuals who are Japanese nationals and those who are not (Iwamoto et al., 2017). Thus, our
315 estimates of the diagnosed proportion of HIV infections among Japanese nationals alone cannot
316 immediately be compared with subsequent existing proportions as if they were sampled from the
317 same population. Whereas estimation of the HIV incidence among non-Japanese nationals is an
318 ongoing research subject, it is frequently the case that infection with HIV or illness onset of
319 AIDS acts as a trigger for foreigners to leave the country; therefore, incorporation of their

320 involvement in the transmission dynamics of Japan requires that very careful attention be paid to
321 migration events, and ideally, that information is supported by individual-based data. Second, the
322 clinical definition of AIDS in Japan depends on indicator diseases, imposing a certain extent of
323 uncertainty in diagnosis. For instance, Japan has a number of designated AIDS Core Hospitals,
324 and HIV diagnoses in those institutes involve screening of common opportunistic infections
325 upon diagnosis of HIV infection, which sensitively leads to the diagnosis of AIDS. Compared
326 with HIV-infected individuals diagnosed at local health centers, the frequency of AIDS diagnosis
327 may be higher in the designated hospitals, calling for the validation of estimates using other
328 methods. Third, in the present study, we struggled with subtraction of AIDS deaths from the
329 calculation of the diagnosed proportion of infections; this problem essentially stems from the
330 absence of a case registration system in Japan. Once diagnosed, infected individuals are never
331 longitudinally monitored by the government, considerably complicating prevalence estimation.
332 With a registration system of HIV-infected individuals, statistical monitoring of the second and
333 third goals of 90-90-90 can be achieved in real time and in a very reasonable manner.
334 Four technical limitations must be noted. First, the present study did not account for
335 uncertainties other than variations in length of the incubation period. There has been a concern
336 that the incubation period has probably shortened over time (Nakamura et al., 2011), but I did
337 not have substantial data to support this issue. Second, the natural history of HIV infection has
338 yet to be explored in-depth; an explicit proportion of HIV individuals who never develop AIDS
339 over the course of infection is missing. Third, other than sex, the present study accepted
340 homogeneity in the natural course and diagnosis of infection. Our future studies will address
341 several heterogeneities. Fourth, estimates rested on yearly data, and the precision was limited

(e.g., with use of the step function for every 4 years). The use of smoothing with nonparametric back-projection is another of our ongoing studies (Becker, 1997).

Despite these limitations, the present study successfully estimated the incidence of HIV infections, undiagnosed number of infections, and the proportion diagnosed in real time, using limited but readily available epidemiological surveillance data. Improved estimates using age and geographical data, as well as estimates based on other methods, are to follow, which will boost studies of epidemiological estimation in this area in Japan.

Conclusions

In the present study, a statistical modeling method was developed for the estimation of HIV incidence in Japan and estimates made of the undiagnosed number of HIV infections and the proportion of diagnosed HIV infections over time. Using the McKendrick equation, a general convolution-like equation was derived, allowing for joint estimation of the HIV incidence and time-dependent rate of diagnosis. By the end of 2017, the cumulative number of HIV infections was estimated to be about 30,000, and about 80% of infections have ever been diagnosed. Accounting for the uncertainty in the median incubation period ranging from 7.5 to 12.3 years, estimates of the diagnosed proportion of HIV infections can range from 77% to 84%. The proportion of diagnosed HIV infections appears not to have reached 90% among Japanese nationals.

Acknowledgment

We thank Analisa Avila, ELS, of Edanz Group for editing a draft of this manuscript.

References

Becker NG. 1997. Uses of the EM algorithm in the analysis of data on HIV/AIDS and other infectious diseases. *Statistical Methods in Medical Research* 6:24-37.

Boldson JL, Jensen JL, Sogarrd J, Sorensen M. 1988. On the incubation time distribution and the Danish AIDS data. *Journal of the Royal Statistical Society, Series A* 151:42-43.

Brookmeyer R, Gail MH. 1986. Minimum size of the acquired immunodeficiency syndrome (AIDS) epidemic in the United States. *Lancet* 2:1320-1322.

Brookmeyer R, Gail MH. 1994. *AIDS Epidemiology: A Quantitative Approach.* New York: Oxford University Press.

Brookmeyer R, Goedert JJ. 1989. Censoring in an epidemic with an application to haemophilia-associated AIDS. *Biometrics* 45:325-335.

Committee of AIDS Trends, Ministry of Health, Labor and Welfare, Japan. 2018. Annual Report of the Trend of HIV and AIDS in Japan. Tokyo: Ministry of Health, Labor and Welfare. *Available at:* http://api-net.jfap.or.jp/status/ (accessed 10 October 2018).

Cuadros DF, Abu-Raddad LJ. 2016. Geographical Patterns of HIV Sero-Discordancy in High HIV Prevalence Countries in Sub-Saharan Africa. *International Journal of Environmental Research and Public Health* 13:865. DOI: 10.3390/ijerph13090865.

Cui JS, Becker NG. 2000. Estimating HIV incidence using dates of both HIV and AIDS diagnoses. *Statistics in Medicine* 19:1165-1177.

Donnelly CA, Cox DR. 2001. Mathematical biology and medical statistics: contributions to the understanding of AIDS epidemiology. *Statistical Methods in Medical Research* 10:141-154.

386 Ejima K, Aihara K, Nishiura H. 2014. Probabilistic differential diagnosis of Middle East
387 respiratory syndrome (MERS) using the time from immigration to illness onset among
388 imported cases. *Journal of Theoretical Biology* 346:47-53.
389 Gail MH, Brookmeyer R. 1988. Methods for projecting course of acquired
390 immunodeficiency syndrome epidemic. *Journal of the National Cancer Institute* 80:900-
391 911.
392 Granich RM, Gilks CF, Dye C, De Cock KM, Williams BG. 2009. Universal voluntary HIV
393 testing with immediate antiretroviral therapy as a strategy for elimination of HIV
394 transmission: a mathematical model. *Lancet* 373:48-57.
395 Granich R, Williams B, Montaner J, Zuniga JM. 2017. 90-90-90 and ending AIDS: necessary
396 and feasible. *Lancet* 390:341-343.
397 Heesterbeek JAP, Metz J. 1993. The saturating contact rate in marriage-and epidemic
398 models. *Journal of Mathematical Biology* 31:529-539.
399 Hsieh YH, Lin PC. 2016. Current Trends and Future Projection of HIV/AIDS Epidemic in
400 Taiwan: A Modeling Analysis. *Current HIV Research* 14:138-147.
401 Hsieh YH, Ruan Y, Chen CW, Shi W, Li D, Luo F, Shao Y. 2012. HIV prevalence and
402 underreporting of men who have sex with men in Beijing. *International Journal of STD*
403 *& AIDS* 23:606-607.
404 Inoue Y, Takaku Y, Yajima T, Ikushima Y. 2015. A survey comparing the route of HIV
405 transmission reported by patients to their physicians and the self-reported "actual route".
406 *Nihon Koshu Eisei Zasshi* 62:106-116 (in Japanese).
407 Iwamoto A, Taira R, Yokomaku Y, Koibuchi T, Rahman M, Izumi Y, Tadokoro K. 2017.
408 The HIV care cascade: Japanese perspectives. *PLoS One* 12:e0174360.

409 Jewell NP, Dietz K, Farewell VT. 1992. *AIDS Epidemiology: Methodological Issues*. Berlin,
410 Birkhäuser.
411 Kihara M, Imai M, Shimizu M. 2000. Trends of HIV infections among blood donors.
412 *Infectious Agents Surveillance Report* 21:140-141 (in Japanese).
413 Kihara M, Ono-Kihara M, Feldman MD, Ichikawa S, Hashimoto S, Eboshida A, Yamamoto
414 T, Kamakura M. 2003. HIV/AIDS surveillance in Japan, 1984-2000. *Journal of Acquired*
415 *Immune Deficiency Syndromes* 32:S55-S62.
416 Kretzschmar ME, Schim van der Loeff MF, Birrell PJ, De Angelis D, Coutinho RA. 2013.
417 Prospects of elimination of HIV with test-and-treat strategy. *Proceedings of the National*
418 *Academy of Sciences of the United States of America* 110:15538-15543.
419 Marschner IC. 1994. Using time of first positive HIV test and other auxiliary data in
420 back‐projection of AIDS incidence. *Statistics in Medicine* 13:1959-1974.
421 Mumtaz GR, Awad SF, Feizzadeh A, Weiss HA, Abu-Raddad LJ. 2018. HIV incidence
422 among people who inject drugs in the Middle East and North Africa: mathematical
423 modelling analysis. *Journal of the International AIDS Society* 21:e25102.
424 Muñoz A, Sabin CA, Phillips AN. 1997. The incubation period of AIDS. *AIDS* 11:S69-S76.
425 Munoz A, Xu J. 1996. Models for the incubation of AIDS and variations according to age
426 and period. *Statistics in Medicine* 15:2459-2473.
427 Nakamura H, Teruya K, Takano M, Tsukada K, Tanuma J, Yazaki H, Honda H, Honda M,
428 Gatanaga H, Kikuchi Y, Oka S. 2011. Clinical Symptoms and Courses of Primary HIV-1
429 Infection in Recent Years in Japan. *Internal Medicine* 50:95-101.
430 Nemoto T. 2004. HIV/AIDS surveillance and prevention studies in Japan: summary and
431 recommendations. *AIDS Education and Prevention* 16:27-42.

Nishiura H. 2007. Lessons from previous predictions of HIV/AIDS in the United States and Japan: epidemiologic models and policy formulation. *Epidemiologic Perspectives and Innovations* 4:3.

Nishiura H, Inaba H. 2011. Estimation of the incubation period of influenza A (H1N1-2009) among imported cases: addressing censoring using outbreak data at the origin of importation. *Journal of Theoretical Biology* 272:123-130.

R Core Team. 2016. The R Project for Statistical Computing. Vienna, Austria. *Available at: https://www.R-project.org/* (accessed 11 October 2018).

UNAIDS. 2014. 90-90-90: an ambitious treatment target to help end the AIDS epidemic. UNAIDS: Geneva. *Available at: http://www.unaids.org/en/resources/documents/2014/90-90-90* (accessed 11 October 2018).

Figure legends

Figure 1. Data-generating process of HIV infections and AIDS cases in Japan

New HIV infections occur at rate $\lambda(t)$. While going undiagnosed as $h(t,\tau)$, there would be an increase in the time since infection τ. Diagnosis of HIV takes place at a time-dependent rate $\alpha(t)$, and AIDS illness onset occurs at rate $\rho(\tau)$, which depends on the time since infection. Newly diagnosed HIV infections, and AIDS cases that had not been previously diagnosed with HIV, were notified to the surveillance system.

Figure 2. Estimated HIV incidence and rate of diagnosis in Japan.

A. The yearly incidence of HIV infection, assuming that the median incubation period is 10.0 years. The step function for every 4 years was used to model the incidence. The 95% confidence intervals were derived from profile likelihood. B. The yearly rate of diagnosis of HIV infection, assuming that the median incubation period is 10.0 years. C. Maximum likelihood estimates of the yearly incidence with different median incubation periods: 7.5, 10.0, and 12.3 years. D. Maximum likelihood estimates of the yearly rate of diagnosis with different median incubation periods: 7.5, 10.0, and 12.3 years. E. Yearly incidence estimates by sex and different median incubation periods. Maximum likelihood estimates are shown. Note that a common logarithmic scale is used on the vertical axis, to ease comparisons. F. Yearly rate of diagnosis estimates by sex and different median incubation periods. Maximum likelihood estimates are shown.

Figure 3. HIV diagnoses and AIDS cases in Japan, 1985–2017.

A. Comparisons between observed and predicted yearly number of HIV diagnoses and AIDS cases. Different median incubation periods (i.e., 7.5, 10.0, and 12.3 years) were assumed, but predicted values are mostly overlapped. B. Comparisons between observed and predicted values by sex. Circles represent the observed number of HIV diagnoses whereas triangles represent that of AIDS cases. Solid marks represent males; empty marks represent females. A common logarithmic scale is used on the vertical axis.

Figure 4. Undiagnosed number and proportion of HIV infections in Japan, 1986–2017.

A. Estimates of undiagnosed HIV infections, assuming that the median incubation period is 10.0 years. The 95% confidence intervals were derived from profile likelihood. B. Maximum likelihood estimates of undiagnosed HIV infections with different median incubation periods:

475 7.5, 10.0, and 12.3 years. C. Proportion of diagnosed infections out of the cumulative number of
476 HIV infections, inclusive of AIDS cases. D. Proportion of diagnosed infections out of the
477 cumulative number of HIV infections, excluding AIDS cases. E. Maximum likelihood estimates
478 of undiagnosed HIV infections by sex, with different median incubation periods: 7.5, 10.0, and
479 12.3 years. Note that common logarithmic scale is used on the vertical axis. D. Proportion of
480 diagnosed infections out of the cumulative number of HIV infections, excluding AIDS cases, by
481 sex.

482 **Figure 5. Estimated undiagnosed HIV infections and proportion of diagnosed infections at**
483 **the end of 2017.**

484 A. Estimates of undiagnosed HIV infections with different incubation periods. Whiskers extend
485 to lower and upper 95% confidence intervals derived using a parametric bootstrapping method.
486 B. Proportion of diagnosed infections out of the cumulative number of HIV infections, excluding
487 AIDS cases (solid circles) or including AIDS cases but subtracting 2,321 deaths (empty circles).
488 Whiskers extend to lower and upper 95% confidence intervals derived using a parametric
489 bootstrapping method.

490

Figure 1(on next page)

Figure 1. Data-generating process of HIV infections and AIDS cases in Japan

New HIV infections occur at rate $\lambda(t)$. While going undiagnosed as $h(t,\tau)$, there would be an increase in the time since infection τ. Diagnosis of HIV takes place at a time-dependent rate $\alpha(t)$, and AIDS illness onset occurs at rate $\rho(\tau)$, which depends on the time since infection. Newly diagnosed HIV infections, and AIDS cases that had not been previously diagnosed with HIV, were notified to the surveillance system.

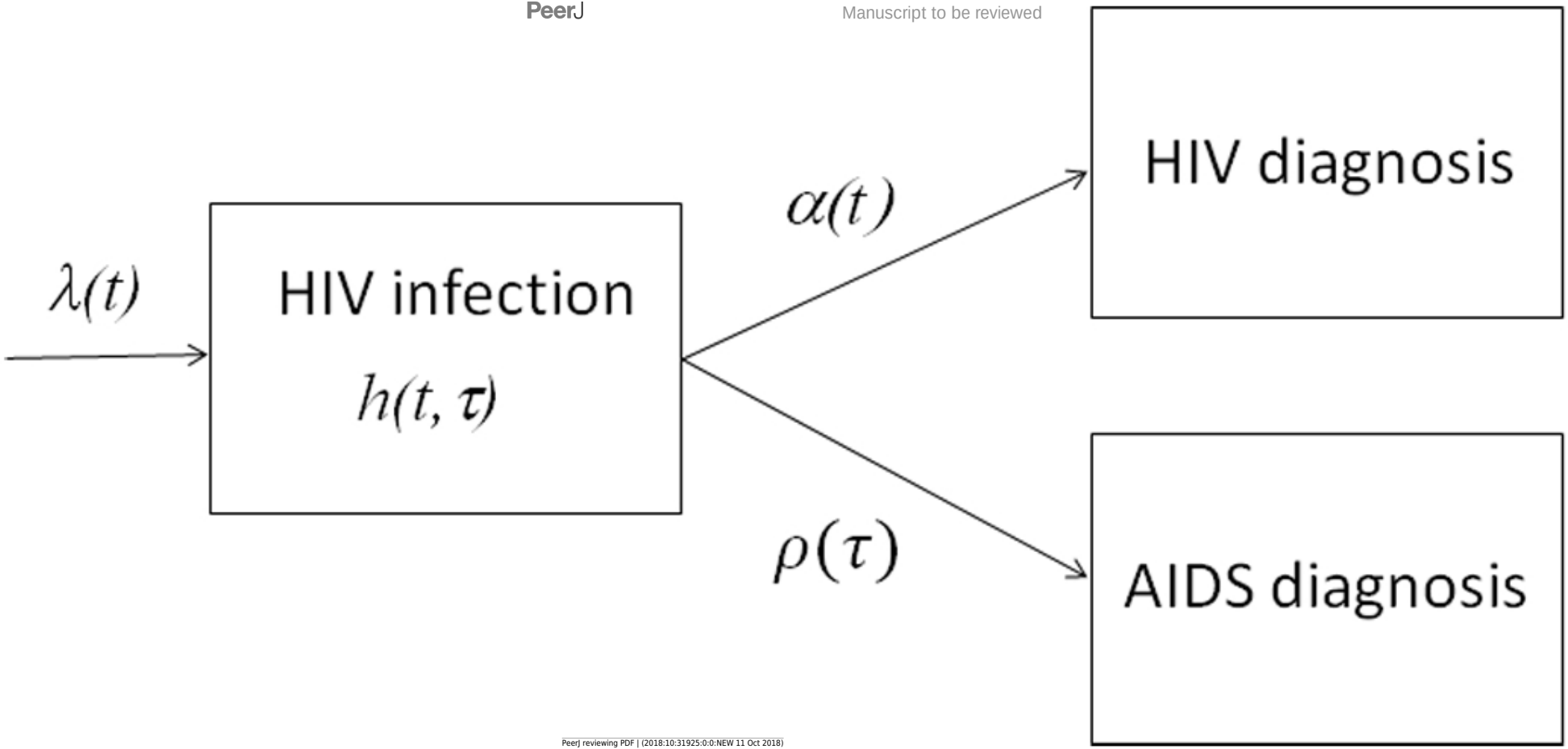
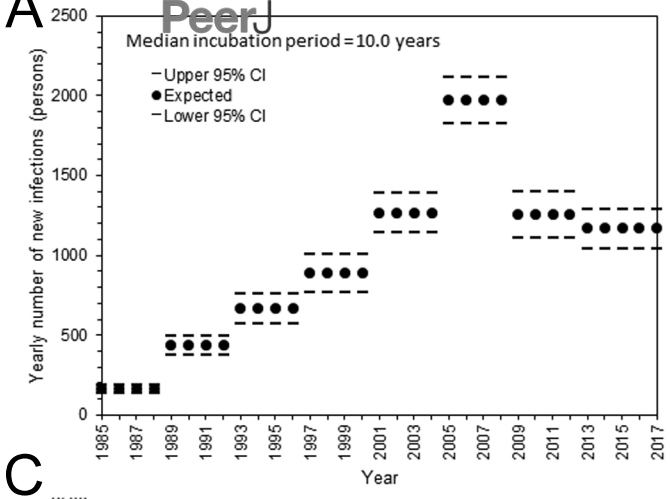
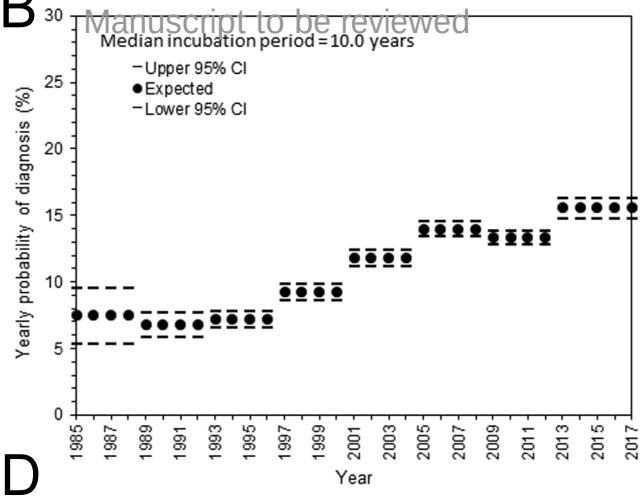

λ(t)
HIV infection
h(t, τ)
α(t)
ρ(τ)
HIV diagnosis
AIDS diagnosis

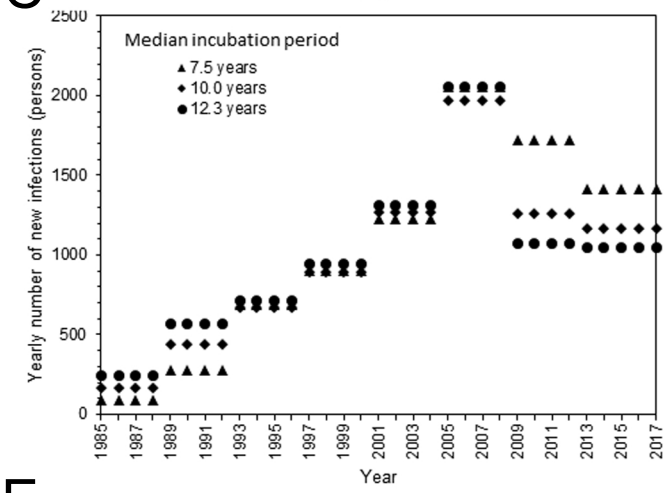
Figure 2(on next page)

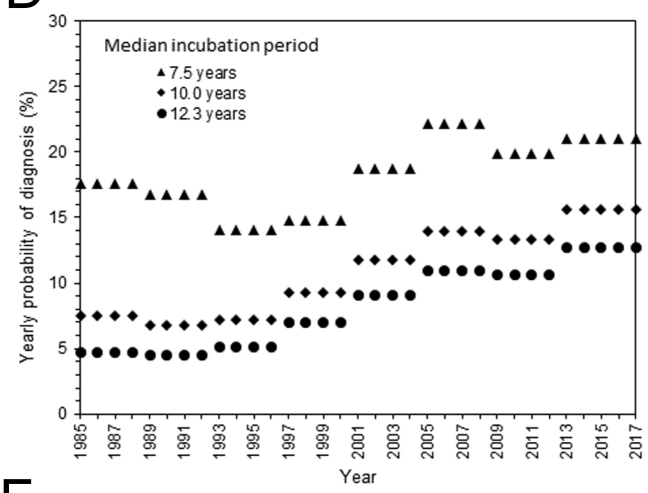
Figure 2. Estimated HIV incidence and rate of diagnosis in Japan.

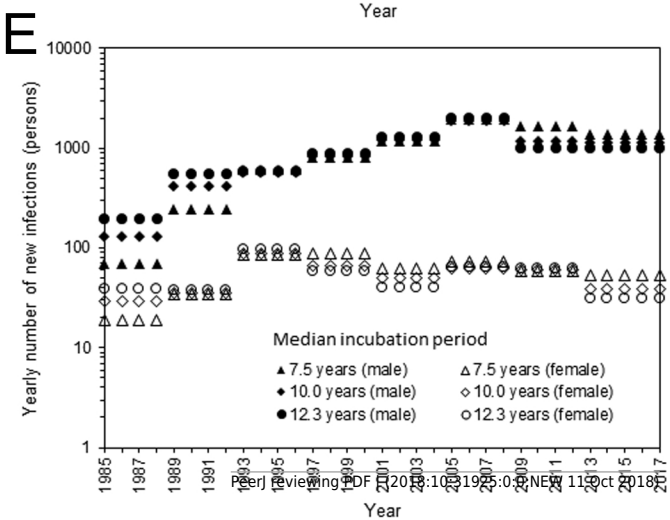
A. The yearly incidence of HIV infection, assuming that the median incubation period is 10.0 years. The step function for every 4 years was used to model the incidence. The 95% confidence intervals were derived from profile likelihood. B. The yearly rate of diagnosis of HIV infection, assuming that the median incubation period is 10.0 years. C. Maximum likelihood estimates of the yearly incidence with different median incubation periods: 7.5, 10.0, and 12.3 years. D. Maximum likelihood estimates of the yearly rate of diagnosis with different median incubation periods: 7.5, 10.0, and 12.3 years. E. Yearly incidence estimates by sex and different median incubation periods. Maximum likelihood estimates are shown. Note that a common logarithmic scale is used on the vertical axis, to ease comparisons. F. Yearly rate of diagnosis estimates by sex and different median incubation periods. Maximum likelihood estimates are shown.











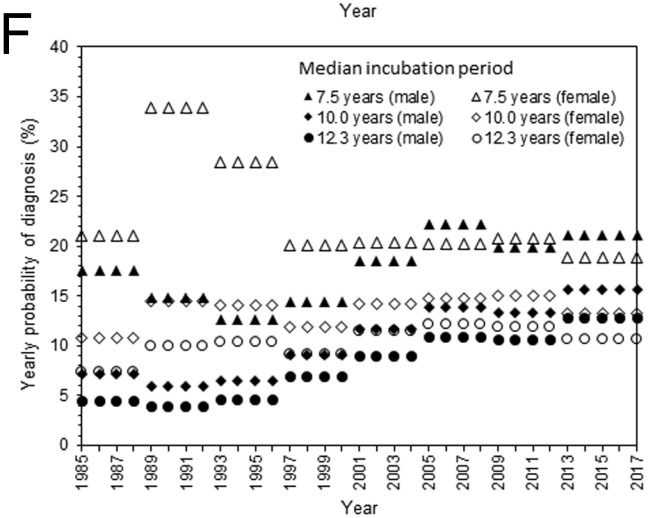
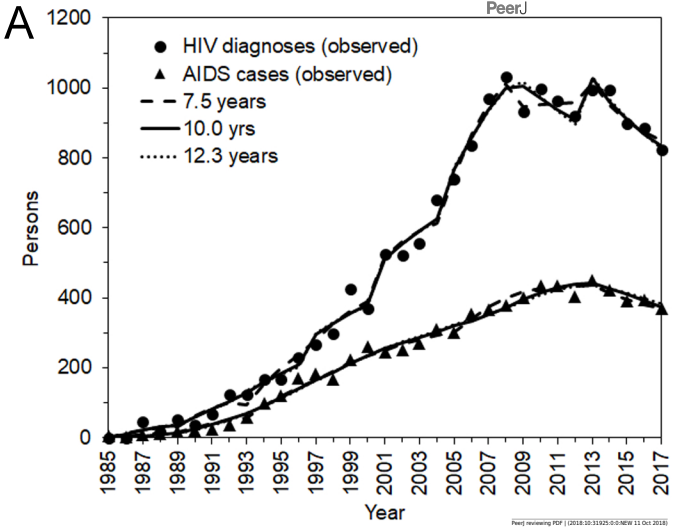


Figure 3(on next page)

Figure 3. HIV diagnoses and AIDS cases in Japan, 1985–2017.

A. Comparisons between observed and predicted yearly number of HIV diagnoses and AIDS cases. Different median incubation periods (i.e., 7.5, 10.0, and 12.3 years) were assumed, but predicted values are mostly overlapped. B. Comparisons between observed and predicted values by sex. Circles represent the observed number of HIV diagnoses whereas triangles represent that of AIDS cases. Solid marks represent males; empty marks represent females. A common logarithmic scale is used on the vertical axis.


A
HIV diagnoses (observed)
AIDS cases (observed)
7.5 years
10.0 yrs
12.3 years
Persons
Year

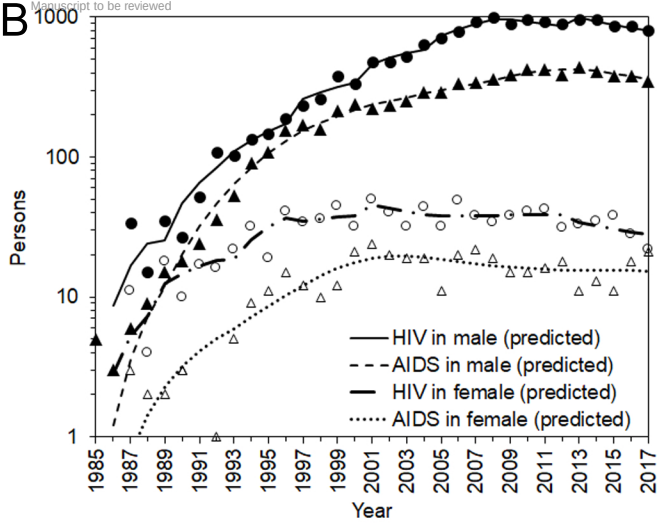

B
HIV in male (predicted)
AIDS in male (predicted)
HIV in female (predicted)
AIDS in female (predicted)
Persons
Year

Figure 4(on next page)

Figure 4. Undiagnosed number and proportion of HIV infections in Japan, 1986–2017.

A. Estimates of undiagnosed HIV infections, assuming that the median incubation period is 10.0 years. The 95% confidence intervals were derived from profile likelihood. B. Maximum likelihood estimates of undiagnosed HIV infections with different median incubation periods: 7.5, 10.0, and 12.3 years. C. Proportion of diagnosed infections out of the cumulative number of HIV infections, inclusive of AIDS cases. D. Proportion of diagnosed infections out of the cumulative number of HIV infections, excluding AIDS cases. E. Maximum likelihood estimates of undiagnosed HIV infections by sex, with different median incubation periods: 7.5, 10.0, and 12.3 years. Note that common logarithmic scale is used on the vertical axis. D. Proportion of diagnosed infections out of the cumulative number of HIV infections, excluding AIDS cases, by sex.

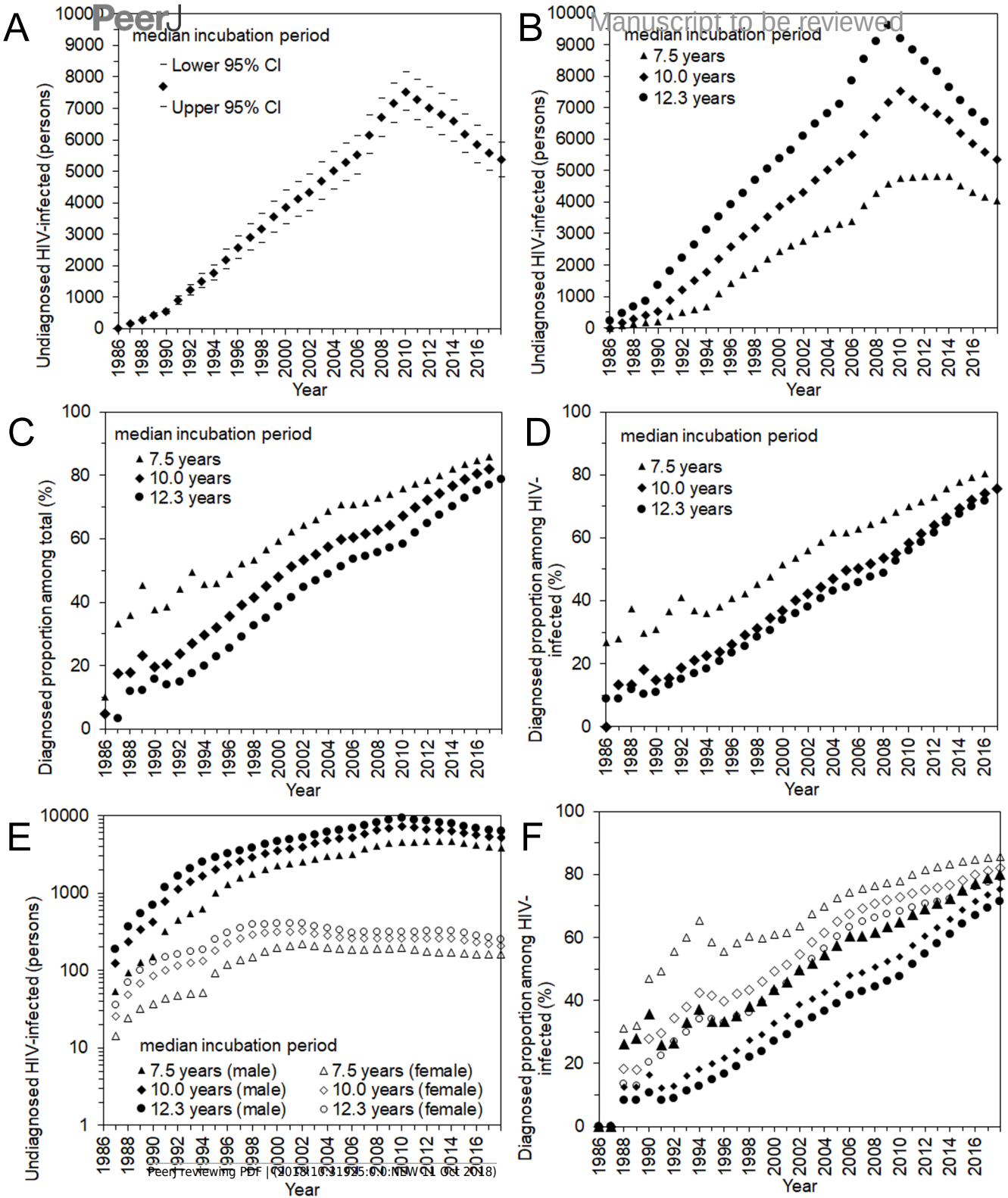
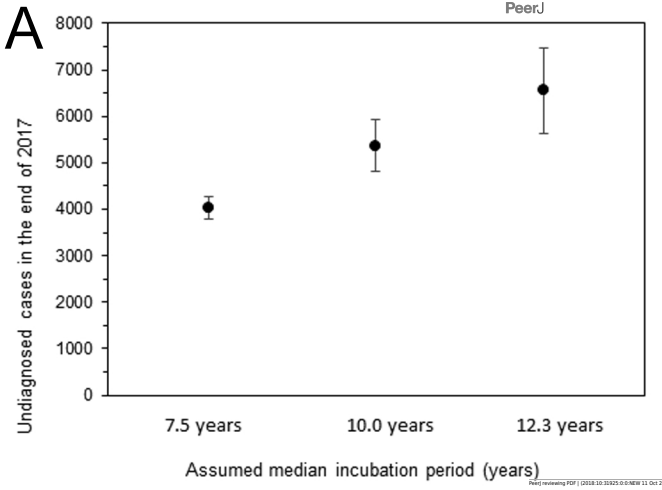
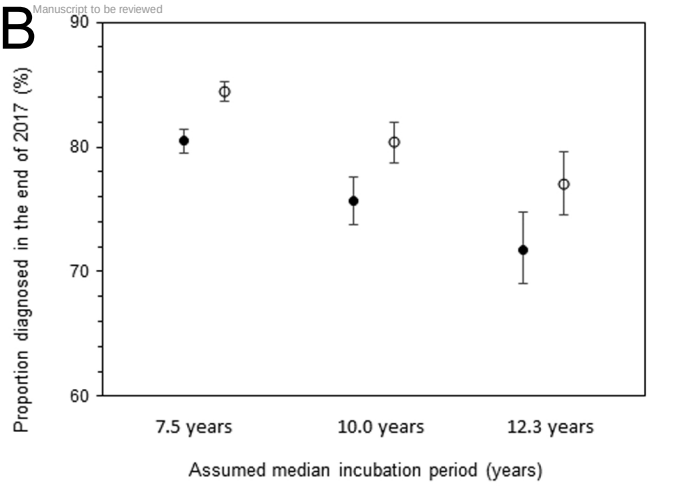


Figure 5(on next page)

Figure 5. Estimated undiagnosed HIV infections and proportion of diagnosed infections at the end of 2017.

A. Estimates of undiagnosed HIV infections with different incubation periods. Whiskers extend to lower and upper 95% confidence intervals derived using a parametric bootstrapping method. B. Proportion of diagnosed infections out of the cumulative number of HIV infections, excluding AIDS cases (solid circles) or including AIDS cases but subtracting 2,321 deaths (empty circles). Whiskers extend to lower and upper 95% confidence intervals derived using a parametric bootstrapping method.


A
8000
7000
6000
5000
4000
3000
2000
1000
0
Undiagnosed cases in the end of 2017
7.5 years
10.0 years
12.3 years
Assumed median incubation period (years)


B
90
80
70
60
Proportion diagnosed in the end of 2017 (%)
7.5 years
10.0 years
12.3 years
Assumed median incubation period (years)